Help us build a bot

ArchiveBots & the SOLM-2018 machine learning data set, Ver 3.8

Monday, October 8th , 2018

SOLM-2018 machine learning training data set

Our vision is a SOLM-2023 database with 1 million markes, initials & signatures from across Europe & North America from the C16th to C18th

Signs of Literacy initiative

The maths

- 3 person/months to create 5,000 signoff SOLM-2018 database consisting of image snippets; boundary boxed snippets on full page images; 5,000 lines x 25 rows of metadata

- 6 person/months to create our targeted 10,000 SOLM-2018 training database

- 20,000 signoff processing per person year

- Target of 1 million signoffs in our database

- 100,000 signoffs per year with 5 people working full time

That's TEN YEARS to achieve our vision
with 50 person years to do it!!!!!

SOLM-2018 database

The **SOLM-2018 database** is a tool for historians and computer scientists to work with markes, initials and signatures. It has been designed to support the exploration of historical literacy and the development of tools for automatic metadata creation.

We are looking for UK and international archival partners interested in contributing content to the SOLM-2018 tool and in learning about AI based pattern recognition

We are especially interested in manuscripts containing markes, initials and signatures by individuals engaged in marine and shore trade occupations from the following English towns and areas for the C16th, C17th and C18th:

Aldeburgh [Suffolk]	Dover [Kent]	Ipswich [Suffolk]	Stonor
Barnstaple [Devon]	Falmouth [Devon]	Newcastle	Weymouth [Dorset]
Bermondsey	Faversham [Kent]	Plymouth [Devon]	Woodbridge [Suffolk]
Bristol	Foy [Cornwall]	Rochester [Kent]	Yarmouth [Norfolk]
Colchester [Essex]	Greenwich	Ramsgate	
Dartmouth [Devon]	Harwich [Essex]	Southampton	
Deptford	Hull	Southwark	

For further information contact Colin Greenstreet, community organiser, Signs of Literacy initiative, or Dr Mark Hailwood (Bristol)
GitHub: https://github.com/Signsofliteracy

6300 signoffs and growing

Our challenge to archivists, computer scientists and historians: Help us develop the tools to create a SOLM-2023 database of 1 mill signoffs with a productivity rate of ten times today's best, at a resource cost of 5 person/years, not 50 person/years, and in half the time

More generally, we need to work together, if we are going to make sense of our digitised manuscript archives – **developing AI tools to process archival images and to identify, extract, read and record metadata**

For more information contact Colin Greenstreet, community organiser of the Signs of Literacy initiative, and Dr Mark Hailwood (Bristol)
https://github.com/Signsofliteracy

We are looking for international partners – archives & archivists, digital publishers,
image & text oriented computer scientists & machine learners,
corpus & historical linguists, and historians.
United Kingdom, Netherlands, Sweden, Poland,
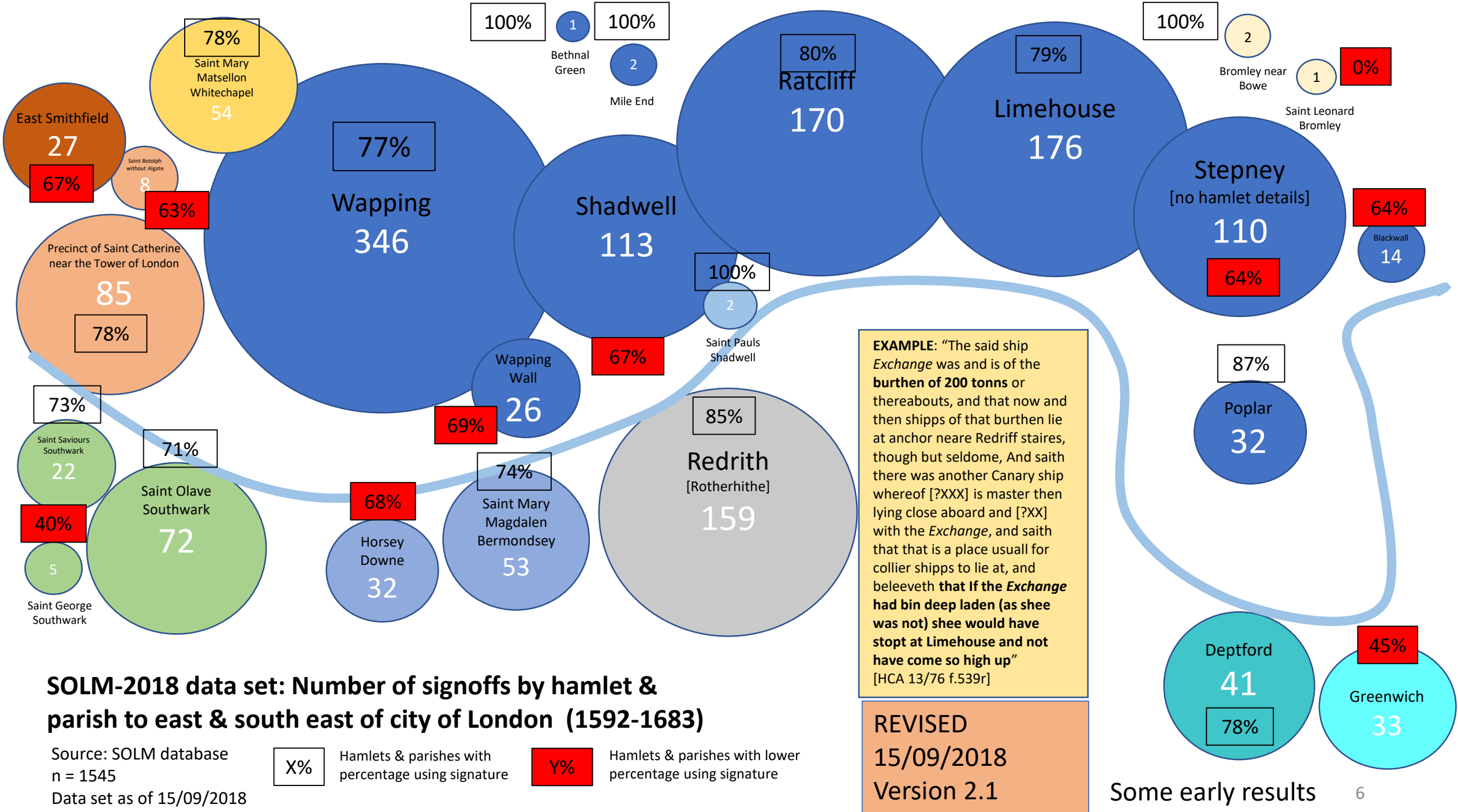Germany, France, Spain & North America

Contact:
Colin Greenstreet (community organiser, Signs of Literacy initiative)
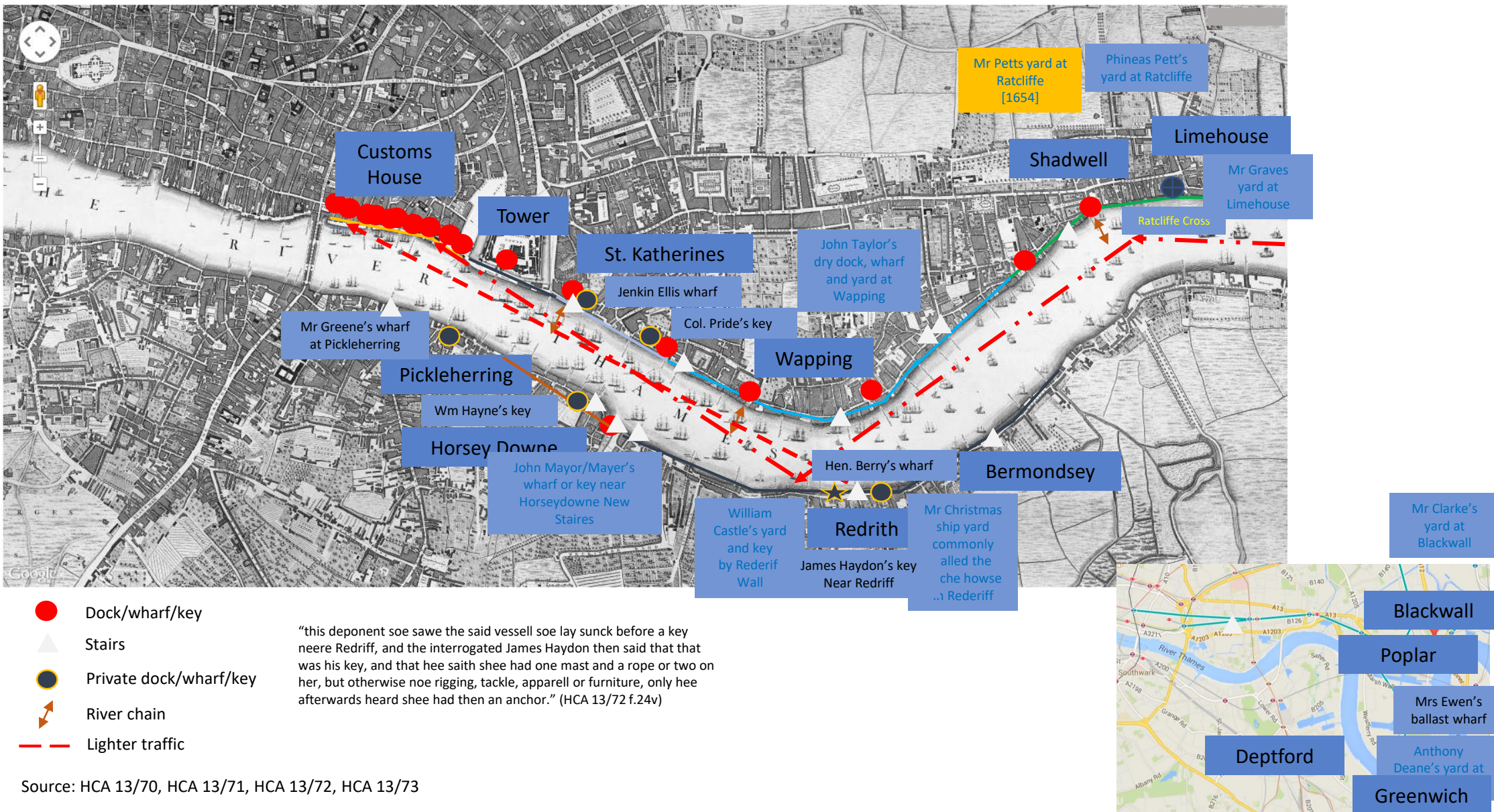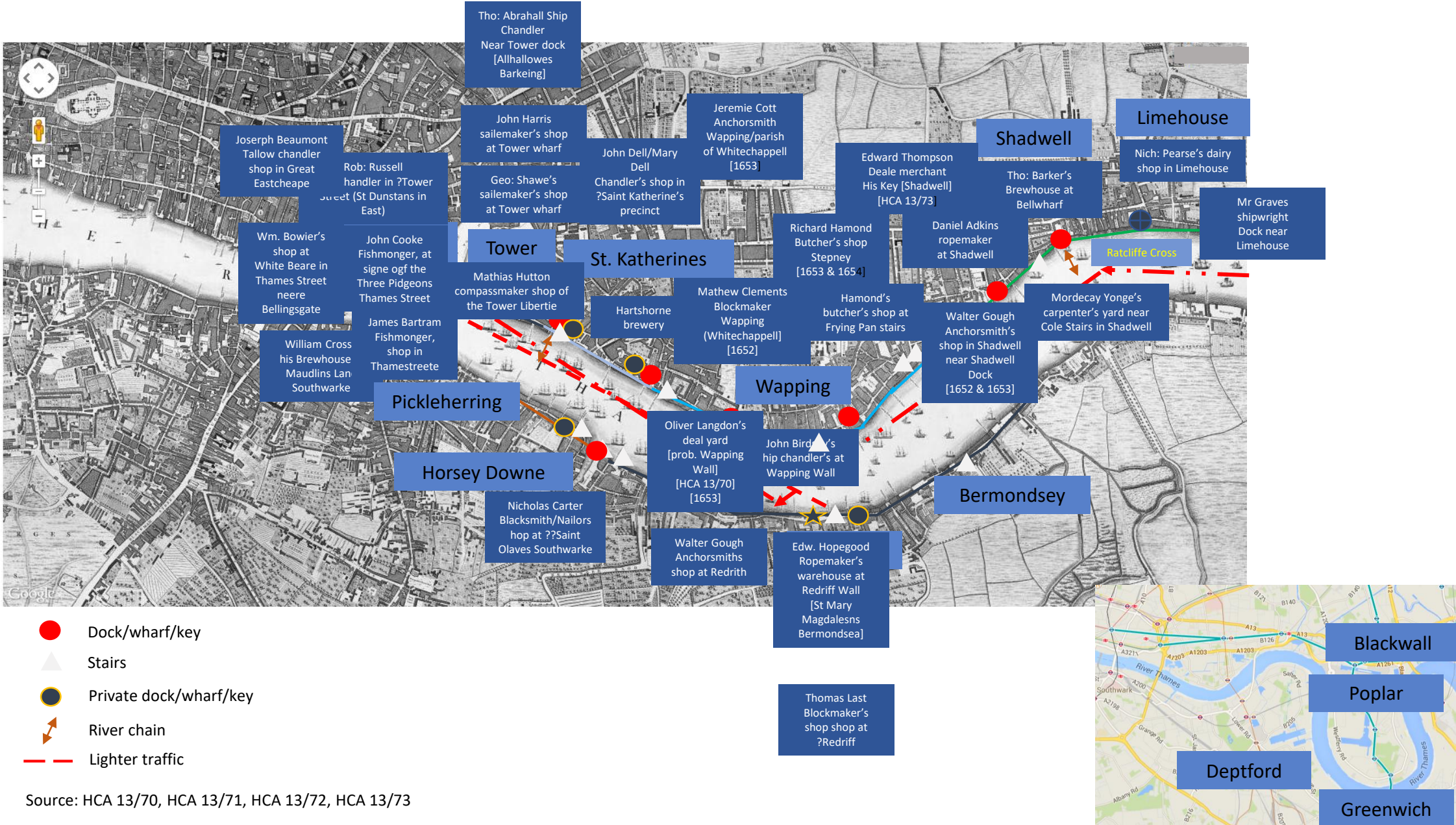or Dr Mark Hailwood (Bristol)
GitHub: https://github.com/Signsofliteracy/
Web: http://signsofliteracy.org

**What can we learn from the
SOLM-2018 database
about C17th historical literacy amongst
mariners and shore trades?
Early results**

SOLM-2018 data set: Number of signoffs by hamlet & parish to east & south east of city of London (1592-1683)

Source: SOLM database
n = 1545
Data set as of 15/09/2018

| X% | Hamlets & parishes with percentage using signature |
| Y% (red) | Hamlets & parishes with lower percentage using signature |

100% — Bethnal Green 1

100% — Mile End 2

78% — Saint Mary Matsellon Whitechapel 54

77% — Wapping 346

80% — Ratcliff 170

79% — Limehouse 176

100% — Bromley near Bowe 2

0% — Saint Leonard Bromley 1

67% — East Smithfield 27

63% — Saint Botolph without Algate 8

78% — Precinct of Saint Catherine near the Tower of London 85

67% — Shadwell 113

100% — Saint Pauls Shadwell 2 / 67%

Stepney [no hamlet details] 110 — 64%

64% — Blackwall 14

87% — Poplar 32

73% — Saint Saviours Southwark 22

40% — Saint George Southwark 5

71% — Saint Olave Southwark 72

69% — Wapping Wall 26

68% — Horsey Downe 32

74% — Saint Mary Magdalen Bermondsey 53

85% — Redrith [Rotherhithe] 159

78% — Deptford 41

45% — Greenwich 33

EXAMPLE: "The said ship *Exchange* was and is of the **burthen of 200 tonns** or thereabouts, and that now and then shipps of that burthen lie at anchor neare Redriff staires, though but seldome, And saith there was another Canary ship whereof [?XXX] is master then lying close aboard and [?XX] with the *Exchange*, and saith that that is a place usuall for collier shipps to lie at, and beleeveth **that If the *Exchange* had bin deep laden (as shee was not) shee would have stopt at Limehouse and not have come so high up**"
[HCA 13/76 f.539r]

REVISED
15/09/2018
Version 2.1

Some early results

6

Private wharves, docks and keys – Tower bridge to Blackwall, mid-C17th

Mr Petts yard at Ratcliffe [1654]

Phineas Pett's yard at Ratcliffe

Limehouse

Mr Graves yard at Limehouse

Shadwell

Customs House

Tower

Ratcliffe Cross

St. Katherines

John Taylor's dry dock, wharf and yard at Wapping

Jenkin Ellis wharf

Col. Pride's key

Mr Greene's wharf at Pickleherring

Wapping

Pickleherring

Wm Hayne's key

Horsey Downe

John Mayor/Mayer's wharf or key near Horseydowne New Staires

Hen. Berry's wharf

Bermondsey

William Castle's yard and key by Rederif Wall

Redrith

James Haydon's key Near Redriff

Mr Christmas ship yard commonly alled the che howse ..ı Rederiff

Mr Clarke's yard at Blackwall

Blackwall

Poplar

● Dock/wharf/key

△ Stairs

◉ Private dock/wharf/key

River chain

– · – Lighter traffic

Mrs Ewen's ballast wharf

Deptford

Anthony Deane's yard at

Greenwich

"this deponent soe sawe the said vessell soe lay sunck before a key neere Redriff, and the interrogated James Haydon then said that that was his key, and that hee saith shee had one mast and a rope or two on her, but otherwise noe rigging, tackle, apparell or furniture, only hee afterwards heard shee had then an anchor." (HCA 13/72 f.24v)

Source: HCA 13/70, HCA 13/71, HCA 13/72, HCA 13/73

Marine suppliers - Tower bridge to Blackwall, mid-C17th

Tho: Abrahall Ship Chandler Near Tower dock [Allhallowes Barkeing]

Joserph Beaumont Tallow chandler shop in Great Eastcheape

Rob: Russell handler in ?Tower Street (St Dunstans in East)

John Harris sailemaker's shop at Tower wharf

Geo: Shawe's sailemaker's shop at Tower wharf

John Dell/Mary Dell Chandler's shop in ?Saint Katherine's precinct

Jeremie Cott Anchorsmith Wapping/parish of Whitechappell [1653]

Edward Thompson Deale merchant His Key [Shadwell] [HCA 13/73]

Shadwell

Limehouse

Nich: Pearse's dairy shop in Limehouse

Tho: Barker's Brewhouse at Bellwharf

Mr Graves shipwright Dock near Limehouse

Wm. Bowier's shop at White Beare in Thames Street neere Bellingsgate

John Cooke Fishmonger, at signe ogf the Three Pidgeons Thames Street

Tower

St. Katherines

Richard Hamond Butcher's shop Stepney [1653 & 1654]

Daniel Adkins ropemaker at Shadwell

Mathias Hutton compassmaker shop of the Tower Libertie

Hartshorne brewery

Mathew Clements Blockmaker Wapping (Whitechappell) [1652]

Hamond's butcher's shop at Frying Pan stairs

Walter Gough Anchorsmith's shop in Shadwell near Shadwell Dock [1652 & 1653]

Ratcliffe Cross

Mordecay Yonge's carpenter's yard near Cole Stairs in Shadwell

James Bartram Fishmonger, shop in Thamestreete

William Cross his Brewhouse Maudlins Lane Southwarke

Pickleherring

Wapping

Horsey Downe

Oliver Langdon's deal yard [prob. Wapping Wall] [HCA 13/70] [1653]

John Bird?n's hip chandler's at Wapping Wall

Bermondsey

Nicholas Carter Blacksmith/Nailors hop at ??Saint Olaves Southwarke

Walter Gough Anchorsmiths shop at Redrith

Edw. Hopegood Ropemaker's warehouse at Redriff Wall [St Mary Magdalesns Bermondsea]

Thomas Last Blockmaker's shop shop at ?Redriff

Legend

- 🔴 Dock/wharf/key
- 🔺 Stairs
- 🟠 Private dock/wharf/key
- ↗ River chain
- - - - Lighter traffic

Blackwall

Poplar

Deptford

Greenwich

Some early results

SOLM-2018 data set: Number of signoffs by parish in city of London [partial]

n = 932 [Partial parish based data set, including all signoffs labelled only "London", without parish data]. Data set as of 03/09/2018

SOLM-2018 data set: Number of signoffs by hamlet & parish to east & south east of city of London (1592-1683)

Source: SOLM database
n = 1545
Data set as of 03/09/2018

| X% | Hamlets & parishes with lower percentage using signature | Y% | Hamlets & parishes with lower percentage using signature |

75% of Admiralty Court deponents outside City of London to south and east are north of the river

North of river =1147
South of river = 398
Parish of Stepney = 63%
[assigning all Wapping to Stepney]

Some early results

Source: SOLM-2018 database

9

The mid-C17th growth of the marine "city" of Stepney depended on large scale migration from outside London and its surrounding parishes, by youths and young men

Residential mobility of mariners and shoretraders in mid-C17th Stepney
n = 14, SOLM-2018, ver1.47, 08/10/2018
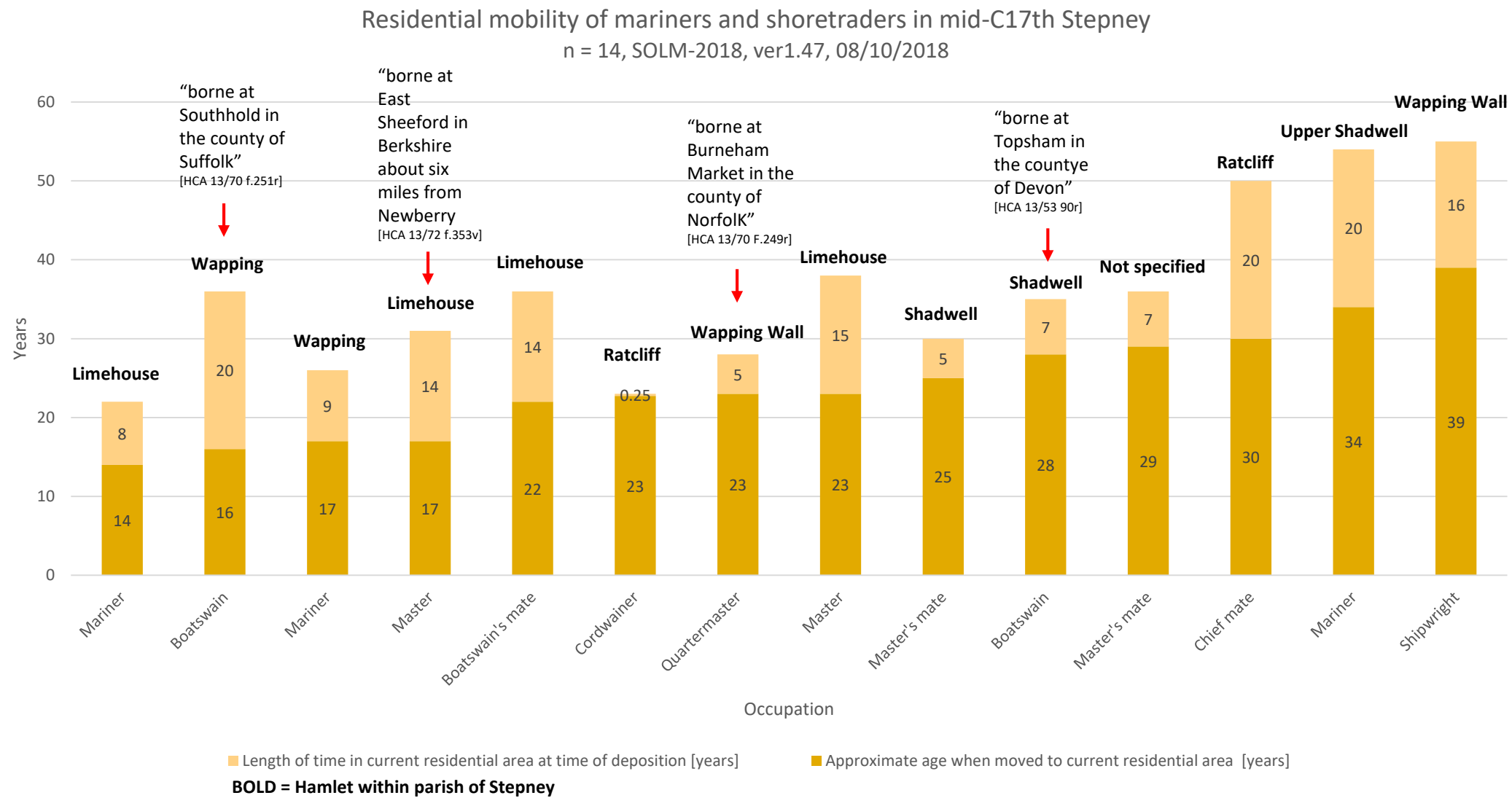
C17th marine occupations with lower signatures as % of all signoffs broadly show higher marke to initial ratios

Markes as a % of total signoffs by occupation
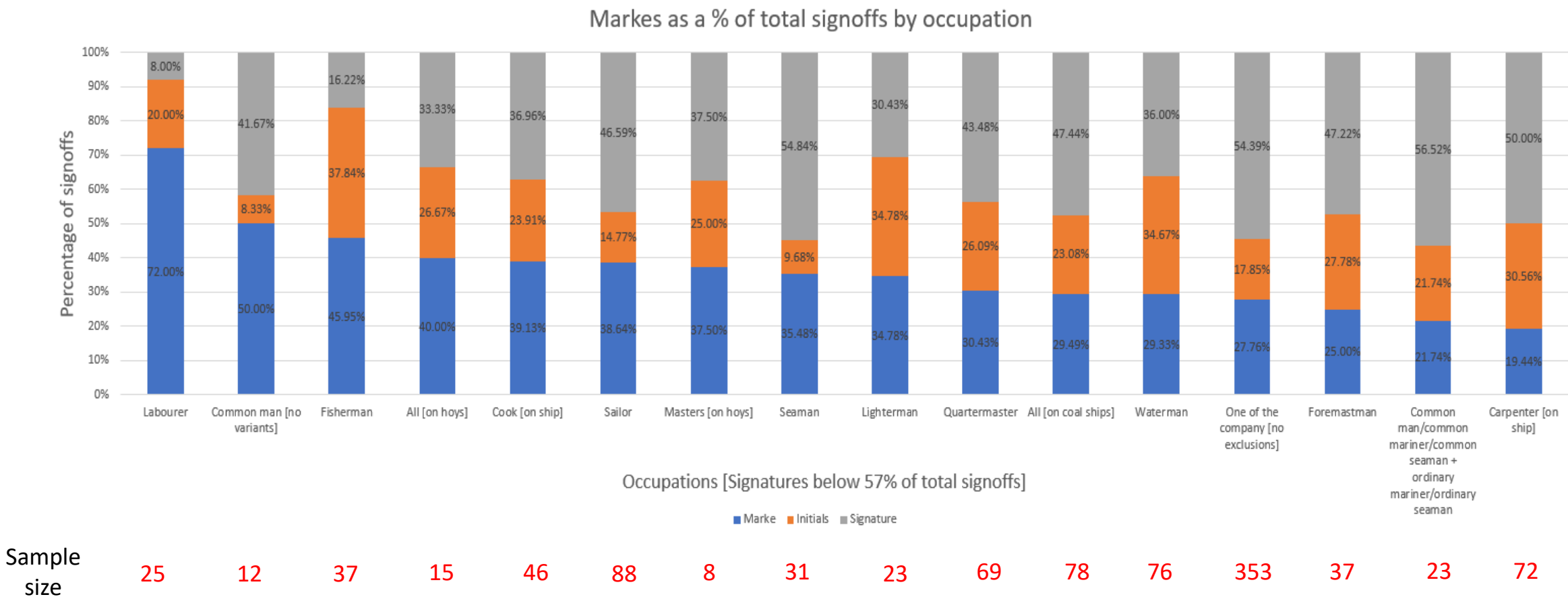
Source: SOLM-2018 database, 28/09/2018; data primarily for 1637-1667

English C17th commercial mariner literacy appears to have been significantly lower on coastal, river and smaller ships than on standard ocean going ships

PROVISIONAL DATA

Signatures as % of all signoffs by mariner office & ship type
n = 1921, SOLM-2018, 27/09/2018

% of signoffs as signatures

| 95.24% | 77.88% | 76.92% | 54.55% | 54.39% | 47.44% | 47.22% | 46.59% | 44.44% | 40.00% | 36.00% | 30.43% | 16.22% |

Master [all ship types] · Mariners [all English mariners, all ships] · Masters [on coal ships] · Common man/common mariner/common seaman + ... · One of the company [no exclusions] · All [on coal ships] · Foremastman · Sailor · Masters [on hoys] · All [on hoys] · Waterman · Lighterman · Fisherman

Mariner and ship type

Legend:
- Masters (green)
- All mariners [on specific types of vessel] (yellow)
- Different names for common men (light blue)
- Watermen, lightermen, fishermen (dark blue)

| Sample size | 273 | 1921 | 13 | 23 | 353 | 78 | 37 | 88 | 8 | 15 | 75 | 23 | 37 |

- Masters of hoys and coal ships significantly less literate than masters of all ships
- All mariners of hoys and coal ships significantly less literate than English mariners on all ships
- Watermen and lightermen possibly less literate than all mariners on hoys
- Fishermen the least literate of the seagoing occupations we have examined
- Hoys, lighters and fishing boats were all much smaller [20 tons to 90 tons burthen] than average commercial ships [200 to 250 tons burthen] & more lightly crewed per ton
- Coal ships were sizeable [200 to 300 tons, but lightly crewed with few officers]

Source: SOLM-2018 database, 28/09/2018; data primarily for 1637-1667

C17th English masters of ships were early starters & showed early career progression

CAUTION: SMALL SAMPLE SIZE

EXAMPLES: Career progression of C17th English mariners
n = 7, SOLM-2018, 28/09/2018
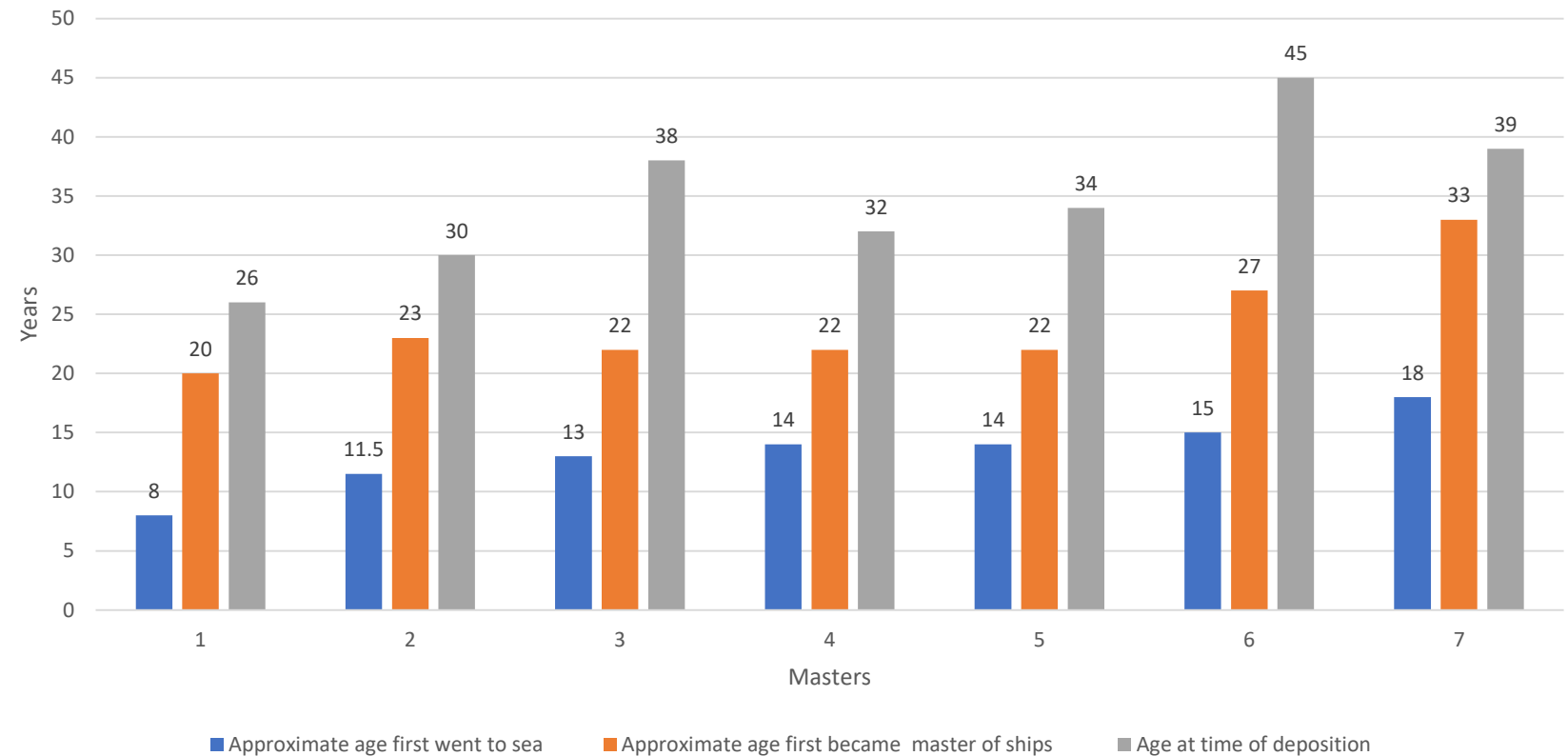
- Average age went to sea is 13.4 years [n = 7]
- Average age went to sea for a larger sample size (for which we have no data on further progression) is 16.7 years [n = 41]
- Average age assumed control of a ship as master is 24 years [n = 7]
- Average age of masters in this small sample is 34.2 years, with an average of ten years as masters
- All masters in this sample signed their depositions with a signature

Residence	Shadwell	Redrith [Rotherhithe]	Wapping	London	Wapping	Ratcliff	Wapping
Ship	Unnamed ship	The Peter			The Mary of London	The Seaven Brothers	The Indian Merchant

Source: SOLM-2018 database, 28/09/2018; data primarily for 1637-1667

Age mariners first went to sea & number of years at sea at time of deposition in English High Court of Admiralty

n = 12 [11 English; 1 Rotterdam]; deposition years 1650 - 1658

"Hee hath used to goe to sea theise 23 yeares, and hath gone under Captaine Ewens and Captaine Cheney in the Parliaments service and twice under Captaine Page to the West Indies, once in the foresaid shipp and once in the *Exchange* of which hee was quarter master, and hath bin in Spaine and many other places, and saith that hee hath bin an officer in ships for 4 yeares last, and that hee is noe carpenter" [HCA 13/63 f.428v]

"Hee hath used to goe to sea by the space of theise 22 yeares and hath bin in Russia Holland Jealy Turkie the West Indies Poland Denmarke ffrance and other places in severall ships and hath bin an officer abord severall ships by the space of theise 15teene yeares, but is noe carpenter" [HCA 13/63 f.430v]

"Hee hath used to goe to sea for theise twelve yeares last and upwards, and saith hee went in the *William and George* (Captaine Bab commander) about seaven yeares since to Virginia, and before that in the shipp the *ffoure Sister*s (Roger Harman master) in a voyage to Greeneland, and since videlicet about 5 yeares since with Mr Sanford of Redriffe in the shipp the *Hopefull* into the Streights, and since that hee went the voyage in question, and before all this hee served an apprenticeshipp at sea under Thomas Lisle of Sole in Suffolke, and hath bin an officer off and on in severall shipps for theise six yeares last" [HCA 13/63 f.426v]

Years

Occupation at time of deposition

■ Approximate age first went to sea [Years]   ■ Years at sea

Source: SOLM-2018, as of 18/09/2018

Signs of Literacy initiative

Age mariners first moved to their current residence in Stepney, Southwark, Redrith [Rotherhithe] or Deptford, and years lived in current residence at time of deposition in English High Court of Admiralty

n = 9; deposition years 1653 - 1669

Southwark

Limehouse (Stepney)

Limehouse (Stepney)

Redrith [Rotherhithe]

Deptford

Redrith [Rotherhithe]

Wapping (Stepney)

Wapping (Stepney)

Shadwell (Stepney)

Years

Occupation	age moved	time in residence
Mariner		29
One of the company	18	6
Boatswain's mate	22	14
Quartermaster	23	5
Master	23	15
Master's mate	25	5
Mariner / Possibly master	25	18
Master	29	5
Sopra cargo	35	5.5

Occupation at time of deposition

Born in Southwark

Born in Burnham Market, Norfolk

Born in Stockholm

Signs of Literacy initiative

■ Approximate age when moved to current residential area [years]   ■ Length of time in current residential area at time of deposition [years]

Year of birth of deponents in SOLM-2018

Chart 1.0: Year of birth of deponent
SOLM-2018, ver.1.45, n = 5647, 04/10/2018
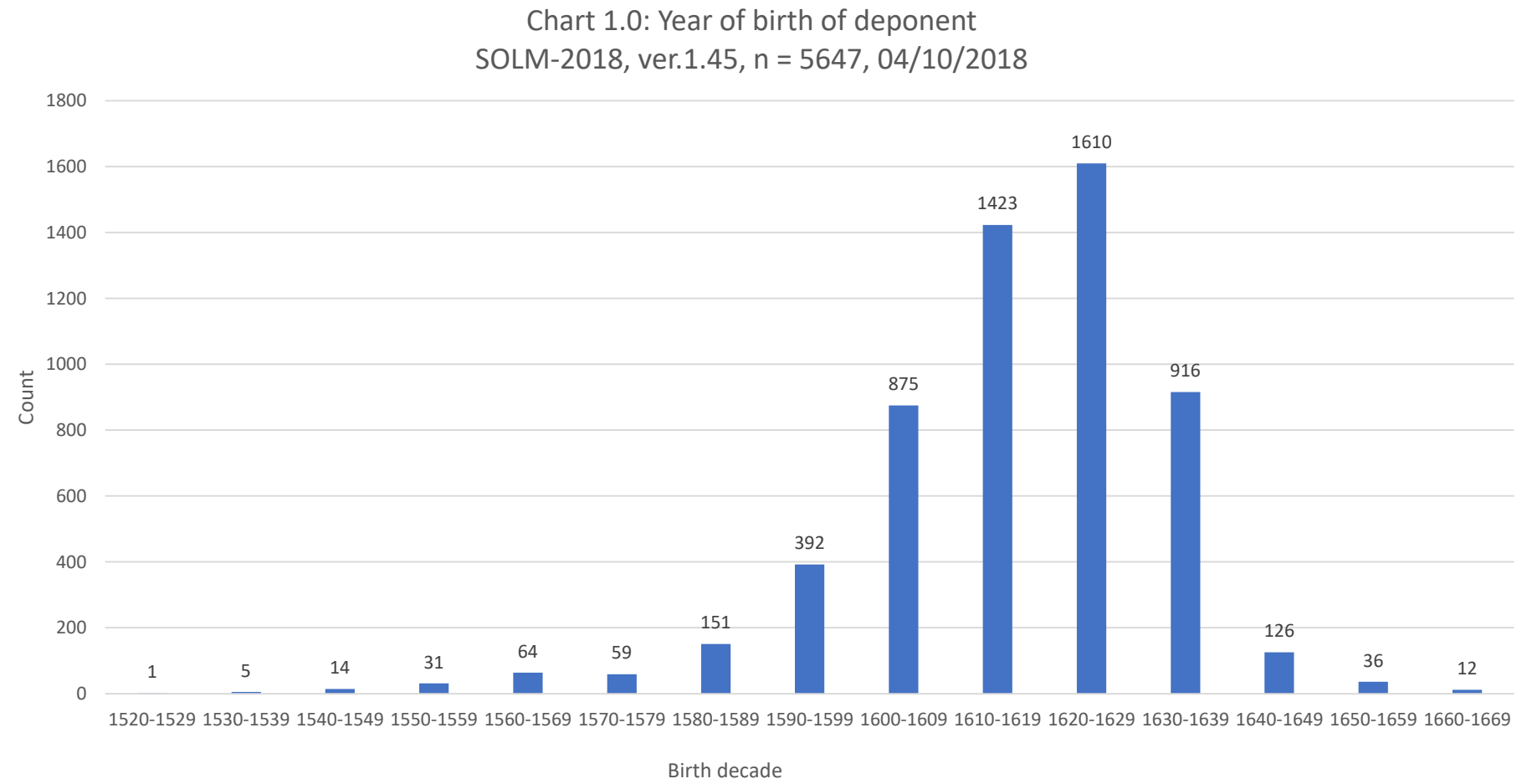
Age of deponent in year of deposition in SOLM-2018

Chart 2.0: Age of deponent in year of deposition
SOLM-2018, ver. 1.46, n = 5,811, 05/10/2018
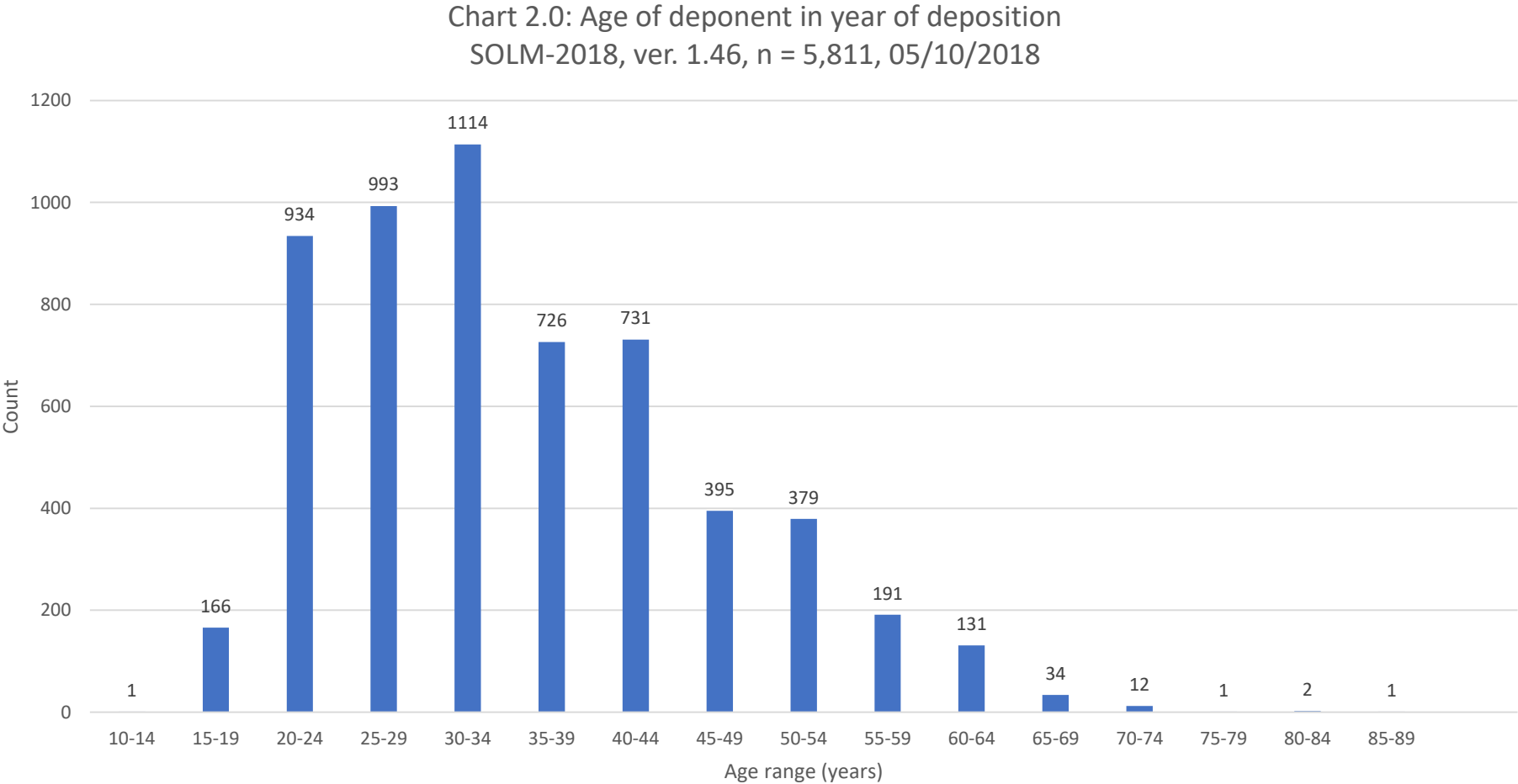
Age early & mid-C17th mariners first went to sea, year of birth 1586 to 1643
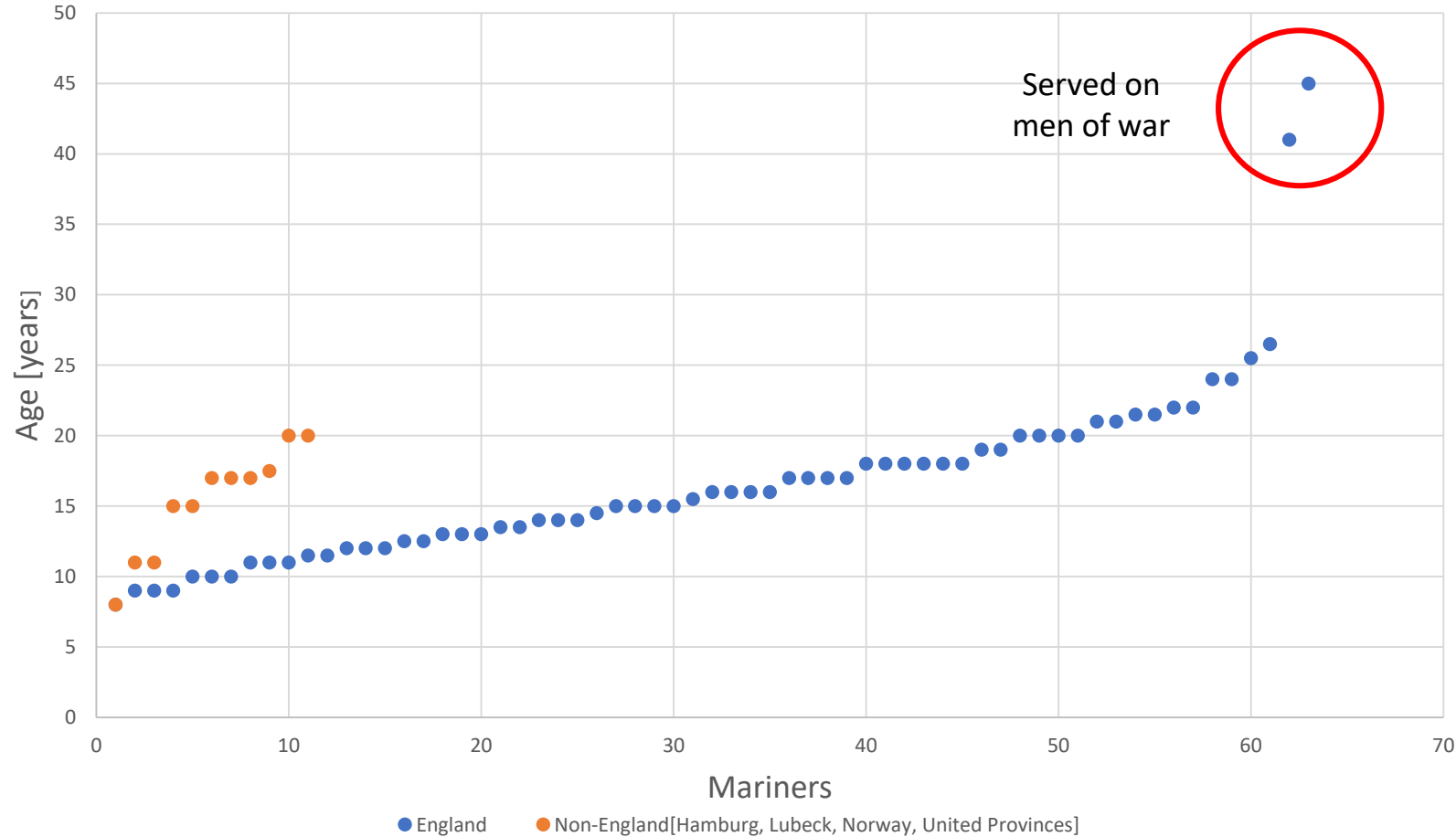
Country of residence	Mean [years]	Median [years]
England [all]	16.7	15.5
England [excluding outliers]	15.8	15.0
Non-England	15.3	17.0

Source: SOLM-2018 database, 05/10/2018; data primarily for depositions from 1637-1667

Proposed project & funding bid

Help us build a bot

Build a quick and dirty Archival Bot in 50 days for further development in 2019

Work has started

Establish a process for machine supported volunteer extraction & synthesis of metadata for the SOLM-2018 machine learning training data set of markes, initials and signatures
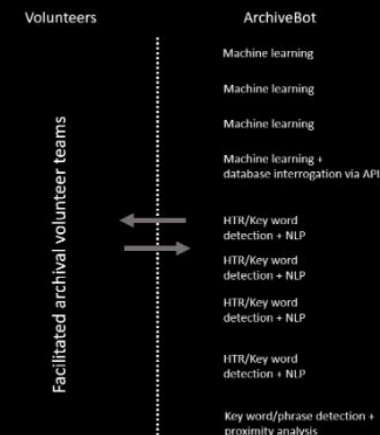
Share first results at the Transkribus users conference in Vienna, November 8th-9th, 2018

Signs of Literacy initiative

Identify manuscript page with signoff

Establish type of signoff [marke, initial(s), signature]

Establish linked front matter of deposition

Establish foliation & image record numbers of signoff & front matter

Establish name of deponent & age

Establish residence of deponent

Establish occupation of deponent from front matter

Establish further occupational data from full text of manuscript

Data mine full text for social & economic parameters

We are looking for a small number of volunteers to work with us as a facilitated remote team – public historians, who can read cursive script, coders & NLP experts - to create a quick and dirty bot as a Proof of Concept to show at the Transkribus Users Conference

We are interested in bidding for Nesta funds to develop our November POC – who is interested in bidding with us?

Expressions of interest to be submitted by Friday, November 11th, 2018

Final proposals submitted, February 2019

Final selection of experiments, March 2019

Experiments to be completed in ten months

Source: ArchiveBots, Ver2.0, 18/09/2018;
https://github.com/Signsofliteracy/ArchiveBots/blob/master/ArchiveBots_Discussion_Document_Ver2.0_18092018.pdf

---

Home / Projects / Collective Intelligence Grants /

Call for ideas

Nesta is offering grants of up to £20,000 for projects that generate new knowledge on how to advance collective intelligence (combining human and machine intelligence) to solve social problems.

Proposals should give an indication of the problem they are looking to solve and how this will be done. Below is a list of potential areas for experimentation. *Please note the examples are not exhaustive, so experiments looking at other areas are also welcome.*

- Increasing inclusion of citizens and public engagement in the participation of collective intelligence online platforms
- Translating data and insights generated through collective intelligence into action
- Harnessing collective intelligence to address the needs of underserved communities in cities or rural areas
- How factors such as the diversity of the "crowd" (socio-economic background, gender balance, ethnicity etc) affect the outcome of the collective intelligence approach
- Testing effective means of incentivising sustained participation in crowd-based collective intelligence
- Piloting effective ways of translating between humans and machines involved in collective intelligence approaches e.g. improving the explainability of data/algorithms used
- Evidence for the added value of using collective intelligence versus other approaches to address social challenges
- Charities and public sector organisations piloting ways to increase their internal/organisational collective intelligence

Collective intelligence

Help us build a bot

Volunteers   **ArchiveBot**

Identify manuscript page with signoff

Establish type of signoff [marke, initial(s), signature]

Establish linked front matter of deposition

Establish foliation & image record numbers of signoff & front matter

Establish name of deponent & age

Establish residence of deponent

Establish occupation of deponent from front matter

Establish further occupational data from full text of manuscript

Data mine full text for social & economic parameters

Machine learning

Machine learning

Machine learning

Machine learning + database interrogation via API

HTR/Key word detection + NLP

HTR/Key word detection + NLP

HTR/Key word detection + NLP

HTR/Key word detection + NLP

Key word/phrase detection + proximity analysis

Facilitated archival volunteer teams

Contact the Signs of Literacy team to learn more

We are experimenting with key phrase and Boolean proximity search

| Identify manuscript page with signoff |
| Establish type of signoff [marke, initial(s), signature] |
| Establish linked front matter of deposition |
| Establish foliation & image record numbers of signoff & front matter |
| Establish name of deponent & age |
| Establish residence of deponent |
| Establish occupation of deponent from front matter |
| Establish further occupational data from full text of manuscript |
| Data mine full text for social & economic parameters |

Text mining using short phrases

Reading skills

"at his examination" [x10]
"now shewed to him"
"perused" [x350]
"perused the schedule" [x14]
"read or write"
"read to him" [x22]
"signed by this deponent"
"the schedule"
"write and read" [x18]
"write nor read"
"written hand" [x35]

BUT: "the same being read to him now and hee this deponent haveing formerly read the same himself" [HCA 13/71 f.232v]

Writing skills

"a writing" "firme" [x163]
"firmed" [x33]
"handwriting" [x166]
"his hand(e)"
"his usuall manner" [x7]
"in writ(e)ing"
"letter of advise"
"marke" [x1365]
"owne handwriting" [x16]
"proper handwriting" [x30]
"proper hand writing" [x34]
"signature" [x3515]
"signed sealed and delivered"
"subscribe his marke" [x1]
"subscribe his name" [x18]
"subscribed with" [x47]
"subscription" [x80]

"the firme" [x33]
"the letter"
"the signeing"
"usuall manner of" [x5]
"writeing" [x122]
"writings" [x269]
"writing his name" [x2]

Language skills

"dutch tongue" [x21]
"french tongue" [x8]
"flemmish" [x29]
"flemmish speech" [x1]
"ffrench tongue" [x8]
"spanish tongue" [x15]
"spoken" [x216]
"tongue" [x144]

Experience of seafaring

"bin at sea" [x45]
"by goeing to sea" [x10]
"by profession"
"by the space of" [x187]
"from his youth" [x21]
"goe to sea" [x79]
"going to sea" [x10]
"master of shipps" [x22]
"navigation"
"used seafaring" [x8]
"used the sea" [x45]
"used the seas" " [x6]
"used to goe to sea" [x12]

Birth place

"borne" [x923]
"hath lived" [x535]
"hee hath lived" [x134]
"lifetime" [x3]
"life tyme" [x63]
"living all his life tyme"

Occupation (past & present)

"being one of the" [x255]
"by profession" [x78]
"he knoweth being" [x119]
"hee knoweth being" [x630]
"knoweth being" [x937]

Worth of estate

"an estate" [x24]
"debts paid" [x32]
"estate" [x653]
"his debts paid" [x11]
"knoweth not" [x1,786]
"owes but" [x3]
"what he is worth" [x7]
"worth but little" [x1]

General

"all which he knoweth" [x30]
"all which hee knoweth" [x346]
"the premisses hee knoweth" [x95]
"the premisses hee knoweth for that" [x14]
"well acquainted with"

Slide 1

Help us build a bot

Signs of Literacy initiative

Build a quick and dirty Archival Bot in 50 days for further development in 2019

Work has started

Establish a process for machine supported volunteer extraction & synthesis of metadata for the SOLM-2018 machine learning training data set of markes, initials and signatures

Share first results at the Transkribus users conference in Vienna, November 8th-9th, 2018

Identify manuscript page with signoff
Establish type of signoff [marke, initial(s), signature]
Establish linked front matter of deposition
Establish foliation & image record numbers of signoff & front matter
Establish name of deponent & age
Establish residence of deponent
Establish occupation of deponent from front matter
Establish further occupational data from full text of manuscript
Data mine full text for social & economic parameters

We are looking for a small number of volunteers to work with us as a facilitated remote team – public historians, who can read cursive script, coders & NLP experts - to create a quick and dirty bot as a Proof of Concept to show at the Transkribus Users Conference

Slide 2

Collective intelligence

Help us build a bot

	Volunteers	ArchiveBot
Identify manuscript page with signoff		Machine learning
Establish type of signoff [marke, initial(s), signature]		Machine learning
Establish linked front matter of deposition		Machine learning
Establish foliation & image record numbers of signoff & front matter		Machine learning + database interrogation via API
Establish name of deponent & age		HTR/Key word detection + NLP
Establish residence of deponent		HTR/Key word detection + NLP
Establish occupation of deponent from front matter		HTR/Key word detection + NLP
Establish further occupational data from full text of manuscript		HTR/Key word detection + NLP
Data mine full text for social & economic parameters		Key word/phrase detection + proximity analysis

Facilitated archival volunteer teams

Contact the Signs of Literacy team to learn more

Slide 3

We are interested in bidding for Nesta funds to develop our November POC – who is interested in bidding with us?

Home / Projects / Collective Intelligence Grants /

Call for ideas

nesta

Nesta is offering grants of up to £20,000 for projects that generate new knowledge on how to advance collective intelligence (combining human and machine intelligence) to solve social problems.

Proposals should give an indication of the problem they are looking to solve and how this will be done. Below is a list of potential areas for experimentation. *Please note the examples are not exhaustive, so experiments looking at other areas are also welcome:*

- Increasing inclusion of citizens and public engagement in the participation of collective intelligence online platforms
- Translating data and insights generated through collective intelligence into action
- Harnessing collective intelligence to address the needs of underserved communities in cities or rural areas
- How factors such as the diversity of the "crowd" (socio-economic background, gender balance, ethnicity etc) affect the outcome of the collective intelligence approach
- Testing effective means of incentivising sustained participation in crowd-based collective intelligence
- Piloting effective ways of translating between humans and machines involved in collective intelligence approaches e.g. improving the explainability of data/algorithms used
- Evidence for the added value of using collective intelligence versus other approaches to address social challenges
- Charities and public sector organisations piloting ways to increase their internal/organisational collective intelligence

Help us build a bot (Signs of Literacy initiative)

Build a quick and dirty Archival Bot in 50 days for further development in 2019

Work has started

Establish a process for machine supported volunteer extraction & synthesis of metadata for the SOLM-2018 machine learning training data set of markes, initials and signatures

Share first results at the Transkribus users conference in Vienna, November 8th-9th, 2018

We are looking for a small number of volunteers to work with us as a facilitated remote team – public historians, who can read cursive script, coders & NLP experts - to create a quick and dirty bot as a Proof of Concept to show at the Transkribus Users Conference

Expressions of interest to be submitted by Friday, November 11th, 2018

Final proposals submitted, February 2019

Final selection of experiments, March 2019

Experiments to be completed in ten months

Source: ArchiveBots, Ver2.0, 18/09/2018;
https://github.com/Signsofliteracy/ArchiveBots/blob/master/ArchiveBots_Discussion_Document_Ver2.0_18092018.pdf

Slide 4

We are experimenting with key phrase and Boolean proximity search

Help us build a bot

Identify manuscript page with signoff
Establish type of signoff [marke, initial(s), signature]
Establish linked front matter of deposition
Establish foliation & image record numbers of signoff & front matter
Establish name of deponent & age
Establish residence of deponent
Establish occupation of deponent from front matter
Establish further occupational data from full text of manuscript
Data mine full text for social & economic parameters

Text mining using short phrases

Reading skills	Writing skills	Language skills	Experience of seafaring
"at his examination" [x10]	"firme" [x163]	"dutch tongue" [x21]	"bin at sea" [x45]
"now shewed to him"	"firmed" [x33]	"french tongue" [x8]	"by goeing to sea" [x10]
"perused" [x350]	"handwriting" [x166]	"ffrench tongue" [x8]	"by profession"
"perused the schedule" [x14]	"marke" [x1365]	"spanish tongue" [x15]	"by the space of" [x187]
"read or write"	"owne handwriting" [x16]	"spoken" [x216]	"from his youth" [x21]
"read to him" [x22]	"proper handwriting" [x30]	"tongue" [x144]	"goe to sea" [x79]
"signed by this deponent"	"proper hand writing" [x34]		"going to sea" [x10]
"the schedule"	"signature" [x3515]		"master of shipps" [x22]
"write and read" [x18]	"subscribe his marke" [x1]		"navigation"
"write nor read"	"subscribe his name" [x18]		"used seafaring" [x8]
"written hand" [x35]	"subscribed with" [x47]		"used the sea" [x45]
	"the firme" [x33]		"used the seas" " [x6]
	"writings" [x269]		"used to goe to sea" [x12]
BUT: "the same being read to him now and hee this deponent haveing formerly read the same himself" [HCA 13/71 f.252v]	"writing his name" [x2]		

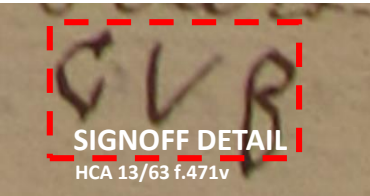| Birth place	Occupation (past & present)	Worth of estate	General
"borne" [x923]		"an estate" [x24]	"all which he knoweth" [x30]
"hath lived" [x535]		"debts paid" [x32]	"all which hee knoweth" [x346]
"hee hath lived" [x134]	"being one of the" [x255]	"estate" [x653]	"the premises hee knoweth" [x95]
"lifetime" [x3]	"by profession" [x78]	"his debts paid" [x11]	"the premises hee knoweth for that" [x14]
"life tyme" [x63]	"he knoweth being" [x119]	"knoweth not" [x1,786]	
"living all his life tyme"	"hee knoweth being" [x630]	"owes but" [x1]	
	"knoweth being" [x937]	"what he is worth" [x7]	
		"worth but little" [x1]	

Contact the Signs of Literacy team to learn more

Imagine an ArchiveBot

Imagine an ArchiveBot extracting metadata automatically from handwritten manuscripts and working with volunteers to finalise

METADATA — HCA 13/63 f.471r

METADATA — HCA 13/63 f.471r

SIGNOFF — HCA 13/63 f.471v
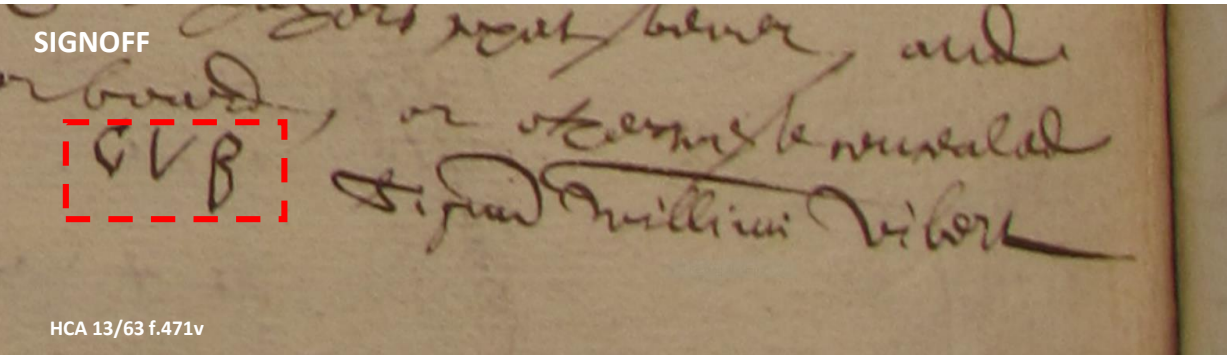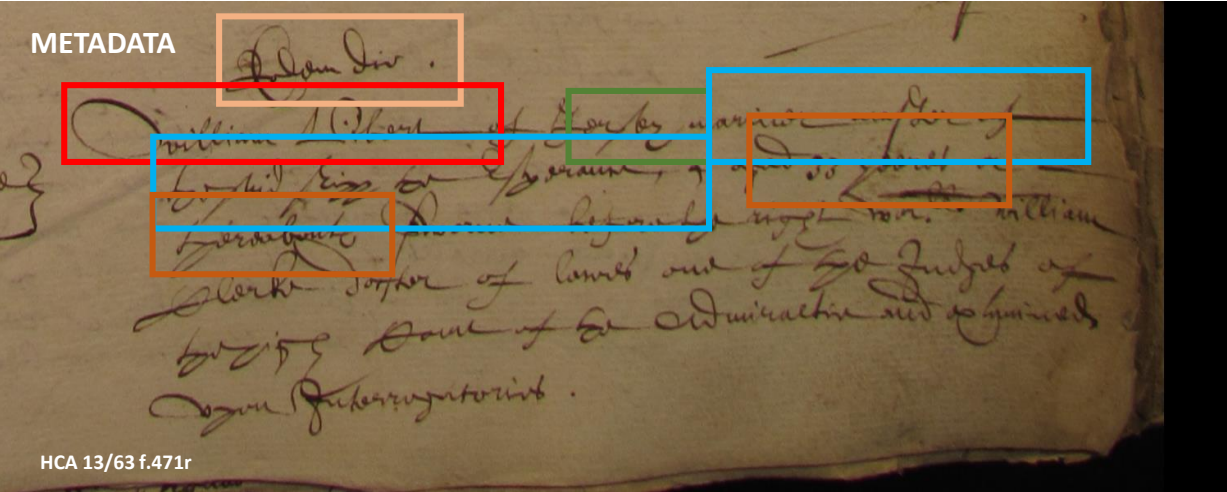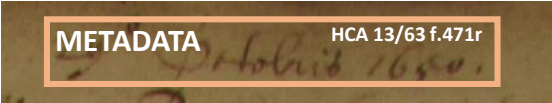
	Date		Occupation
	Name		Age
	Residence		Signoff

SIGNOFF DETAIL
HCA 13/63 f.471v

William Vibert (Guillaume Vibert) was a thirty-three year old mariner from the isle of Jersey, who was master of the ship the *Esperansa*.

He signed his deposition in the English High Court of Admiralty on October 29th 1650, using three not two initials.

The choice of "G", "V" and "B" for initials suggests he thought of himself and pronounced his name as "Guillaume Vee Bert" (though he could not write a full signature)

Challenge for a Bot

(1) Find the date of the deposition (it is at the top of the same manuscript page above the preceding deposition)
(2) Deal with the interlineation in the metadata
(3) On the next manuscript page, recognise "GVB" are initials not a marke. The "G" is quite tough (easy to mistake for "C"). Yes, they are recognisable letters, but the three don't match the two names "William Vibert". Need to know William – Guillaume in French, and that mariners in Jersey may be born in France and French speaking.

Source: http://www.marinelives.org/wiki/HCA_13/63_f.471r_Annotate; http://www.marinelives.org/wiki/HCA_13/63_f.471v_Annotate

26

Speech to text recognition
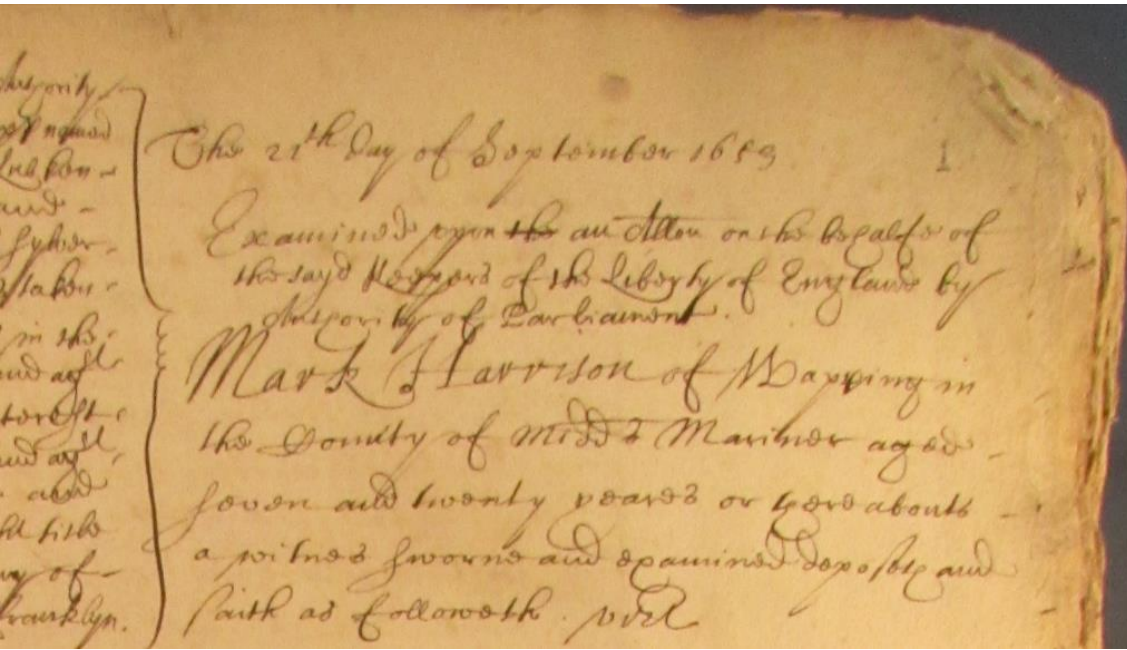
Watson   Speech to Text / Speech to Text Demo

Speech to Text

The IBM Watson Speech to Text service uses speech recognition capabilities to convert Arabic, English, Spanish, French, Brazilian Portuguese, Japanese, Korean, and Mandarin speech into text.

🔗 **Get Started**   API Reference   Documentation   Fork on GitHub   [Start for free in IBM Cloud]

Voice Model:

GB English broadband model (16KHz) ▼

mark Harris[2] and of[4] what happened[7] in[2] the county of Middlesex mariner[8] aged seven and twenty years

mark Harrison[3] of[2] walking[8] in[2] the county of Middlesex mariner[15] aged seven and twenty years or[2] there[4] about

mark[2] Harrison[3] of[2] what[4] happened in the county of Middlesex mariner[10] aged seven and twenty years or[4] there[4]

mark[2] Harrison[3] or[3] walking in the county of Middlesex mariner[8] aged[2] seven and twenty years[3] or[2] the[3] about[5]

Can we use key word spotting to excavate raw metadata?

LANGUAGE DENOTING OCCUPATION

"The premisses hee deposeth being one of the company of the *Bridgewater ffrygott*, and sawe the same soe done" [HCA 13/72 f.90r] [CONCLUSION: One of the company]

"The premisses he deposeth for that he the deponent was not onely for the voyage arlate wherein she was stranded, but in two former voyages stiersman of the sayd ship" [HCA 13/72 f.90v] [CONCLUSION: Steersman]

"after such buying of the said shipp the *Santa Hellen* by the said da [?Groots] and company aforesaid, namely in or about the moneth of October 1643 (about three monethes after the adiudication aforesaid) being in possession of the said vessell, put and constituted this deponent master of her, and hee then entred master upon her at Port Lewes in Bretany" [HCA 13/72 f.95r] [CONCLUSION: Master]

"all which hee knoweth being masters mate of the said shipp the said voyage" [HCA 13/70 f.669v] [CONCLUSION: Master's mate]

"hee knoweth the premisses because hee was boatswaines mate of the said shipp and went all the said voyage in her" [HCA 13/70 f.671r] [CONCLUSION: Boatswain's mate]

Can we refine raw machine generated metadata using a combination of NPL, controlled vocabularies, and programmable decision rules?

LANGUAGE DENOTING OCCUPATION

"The premisses hee deposeth being one of the company of the *Bridgewater ffrygott*, and sawe the same soe done" [HCA 13/72 f.90r] [CONCLUSION: One of the company]

"The premisses he deposeth for that he the deponent was not onely for the voyage arlate wherein she was stranded, but in two former voyages stiersman of the sayd ship" [HCA 13/72 f.90v] [CONCLUSION: Steersman]

"after such buying of the said shipp the *Santa Hellen* by the said da [?Groots] and company aforesaid, namely in or about the moneth of October 1643 (about three monethes after the adiudication aforesaid) being in possession of the said vessell, put and constituted this deponent master of her, and hee then entred master upon her at Port Lewes in Bretany" [HCA 13/72 f.95r] [CONCLUSION: Master]

"all which hee knoweth being masters mate of the said shipp the said voyage" [HCA 13/70 f.669v] [CONCLUSION: Master's mate]

"the premisses because hee was boatswaines mate of the said shipp and went all the said voyage in her" [HCA 13/70 f.671r] [CONCLUSION: Boatswain's mate]

An ArchiveBot would need to be able to read manuscript text from the main body of legal depositions, not just the front matter at the start of a deposition

Table 21.1: Extract from occupational data sub-set in SOLM-2018 data set
Data as of 11/09/2018

Occupations [Synthesised]	Occupations stated in main metadata	Additional occupations & titles in body of deposition	Stated activities in body of text
BOATSWAINS			
Seaman; Commonman [outwards]; Boatswain [return]	Seaman	Commonman [outwards]; Boatswain [return]	
Nauta [mariner]; Coxwain [outwards] [on Thomas and John]; Boatswain [return] [on Thomas and John]	Nauta [mariner]	Coxwain [outwards] [on Thomas and John]; Boatswain [return] [on Thomas and John]	
Nauta [mariner]; Boatswain's mate [of Thomas and John]; [?Part-owner of ship Thomas and John]	Nauta [mariner]; Boatswain's mate [of Thomas and John]	[?Part-owner of ship Thomas and John]	
Nauta [mariner]; Boatswain [of the Samuel and Mary]	Nauta [mariner]; Boatswain [of the Samuel and Mary]		
Nauta [mariner]; Boatswain [of the Mathewe]	Nauta [mariner]; Boatswain [of the Mathewe]		
Mariner; Boatswain [of the Hopewell]; One of the company [of the Hopewell]	Mariner; Boatswain [of the Hopewell]	One of the company [of the Hopewell]	
Mariner; Boatswain [of the Hawke of Stockholm]	Mariner	Boatswain [of the Hawke]	
Mariner; Boatswain [of the herring busse the Liesda]	Mariner; Boatswain [of the herring busse the Liesda]		
Mariner; Boatswain [of the Pearce]	Mariner; Boatswain [of the Peace]		
Mariner; Boatswain [of the Saint John]	Mariner; Boatswain [of the Saint John]		
Mariner; Boatswain [of the Liesde]; One of the company [of the Liesde]	Mariner	Boatswain [of the Liesde]; One of the company [of the Liesde]	
Mariner; Boatswain [of the Recovery]	Mariner; Boatswain [of the Recovery]		
Boatswain [of the Arke]	Boatswain [of the Noahs Arke or Arke]		
RANDOM SELECTION OF OCCUPATIONS			
Waterman	Waterman		Being aboard the Saint Christofer
?	?		Bound in a bond
Mariner; Sopra cargo; Steward [of unnamed ship]	Mariner	Sopra cargo; Steward [of unnamed ship]	Has used the Virginia trade for 16 years
Haberdasher; Citizen	Haberdasher; Citizen		Involved in arbitration
Merchant	Merchant		Kept accounts for Mr William Bewly, merchant of London
Merchant; Servant [to William Brugg]	Merchant	Servant	Kept his bookes [of William Brugg]
Merchant cashier [to merchant]	Merchant cashier		Made entries in books of accounts
Mariner; Master's mate [of the Advantage frigot]	Mariner; Master's mate [of the Advantage frigot]		Present at taking of the Golden Starr
Sailor	Sailor		Put on board captured ship
Cooper [on shore]	Cooper [on shore]		Taking up oiles
Sailor; Worked on building of ship [the Great Christofer]	Sailor		Worked on building of ship [the Great Christofer]

Note four out of thirteen boatswains identified here are described simply as seaman, nauta or mariner in the legal front matter

Note information on activities of deponents in main body of text helps clarify and refine the main occupational descriptions

Source: Extract from SOLM-2018 database, as of 11/09/2018

Initials are not always easy to recognise and confirm
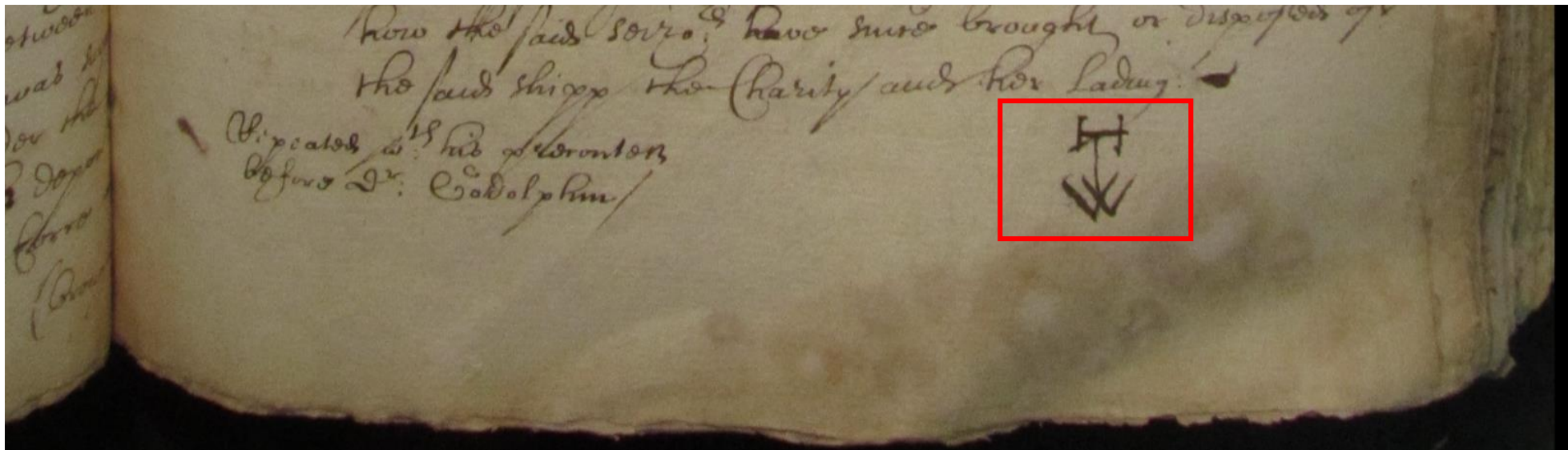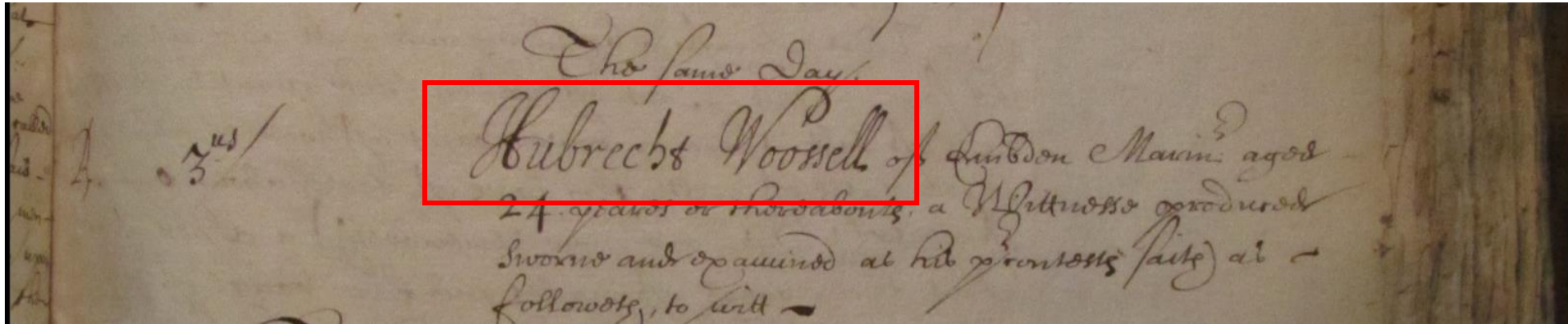
Front matter: "Hubrechs Woossell"
Initials: "H
                  VV"

Source: Extracts from HCA 13/72 f.277r and HCA 13/72 f.278r

Initials are not always easy to recognise and confirm
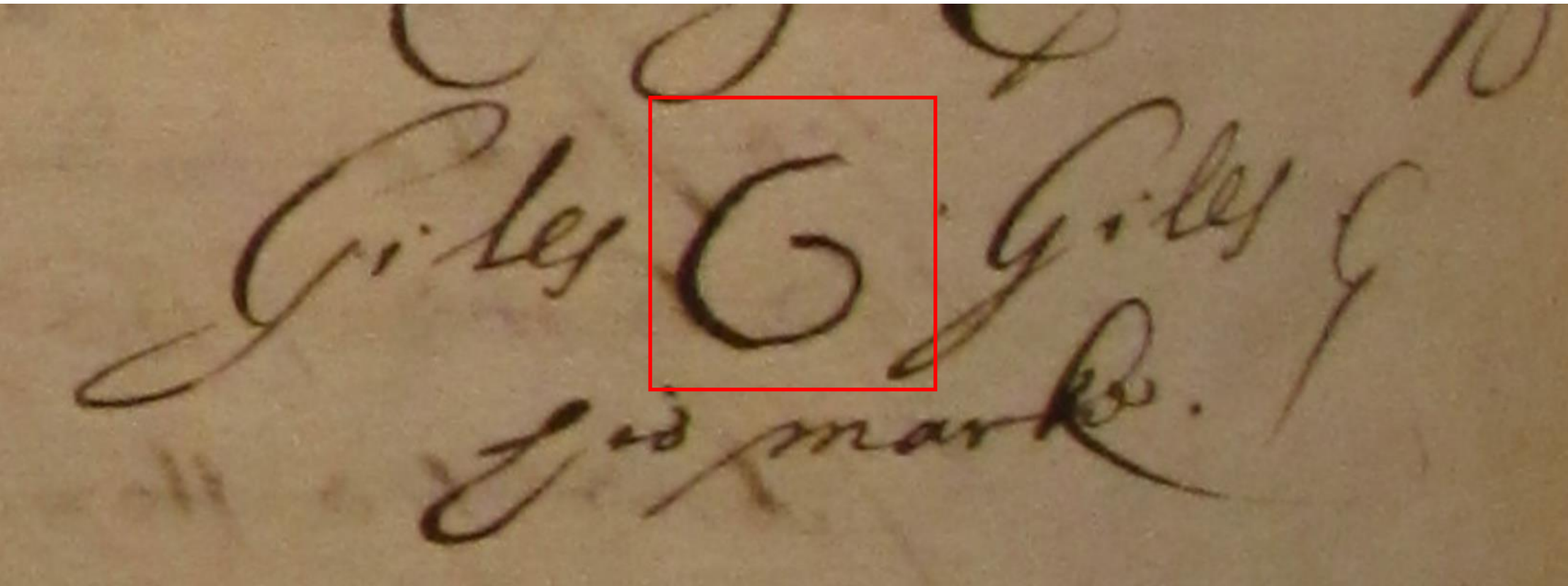
Front matter: "Giles Giles"
Initial or Marke?   "?G"

Source: Extracts from HCA 13/68 f.331v

Mismatch between English rendition of names in front matter and initials and signatures made by non-native English speakers at end of deposition
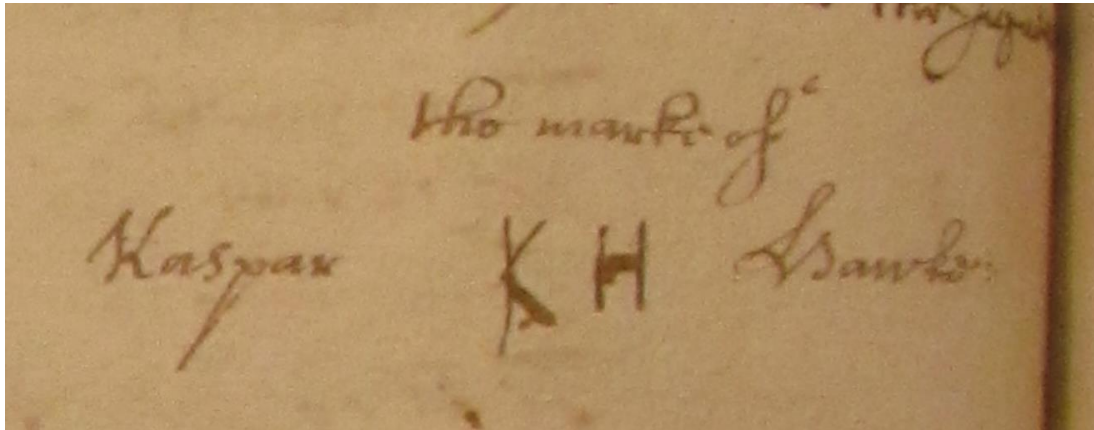
Front matter: Gaspar Hawke
Initials: KH [Kaspar Hawke]
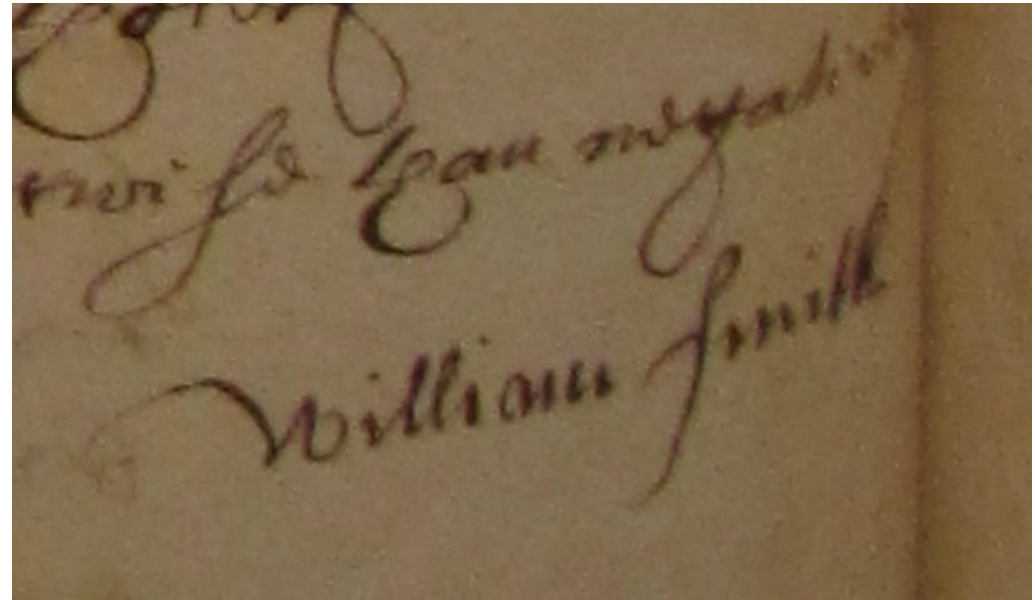
Front matter: William de Smitt
Signature: William Smith

Distinguishing a deponent's signature from goods marked with a deponent's initials
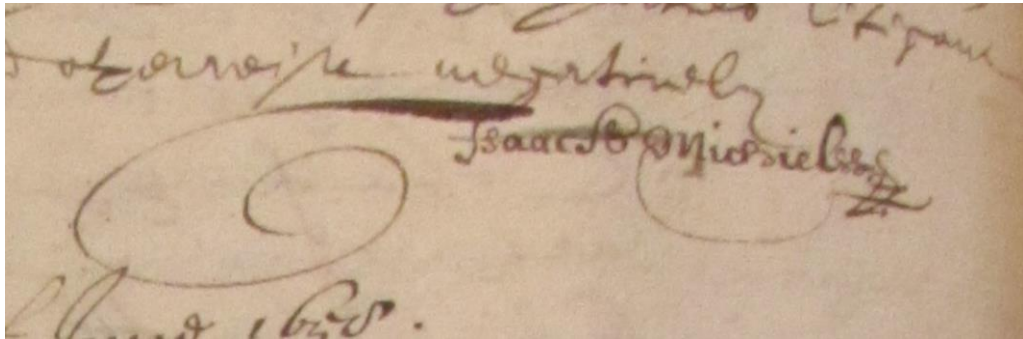
Signature: Isaacs Michsielss

Good marked: IM

Source: Clockwise from LH corner: KaggleTestSnippet_HCA_1372_f.357v.PNG
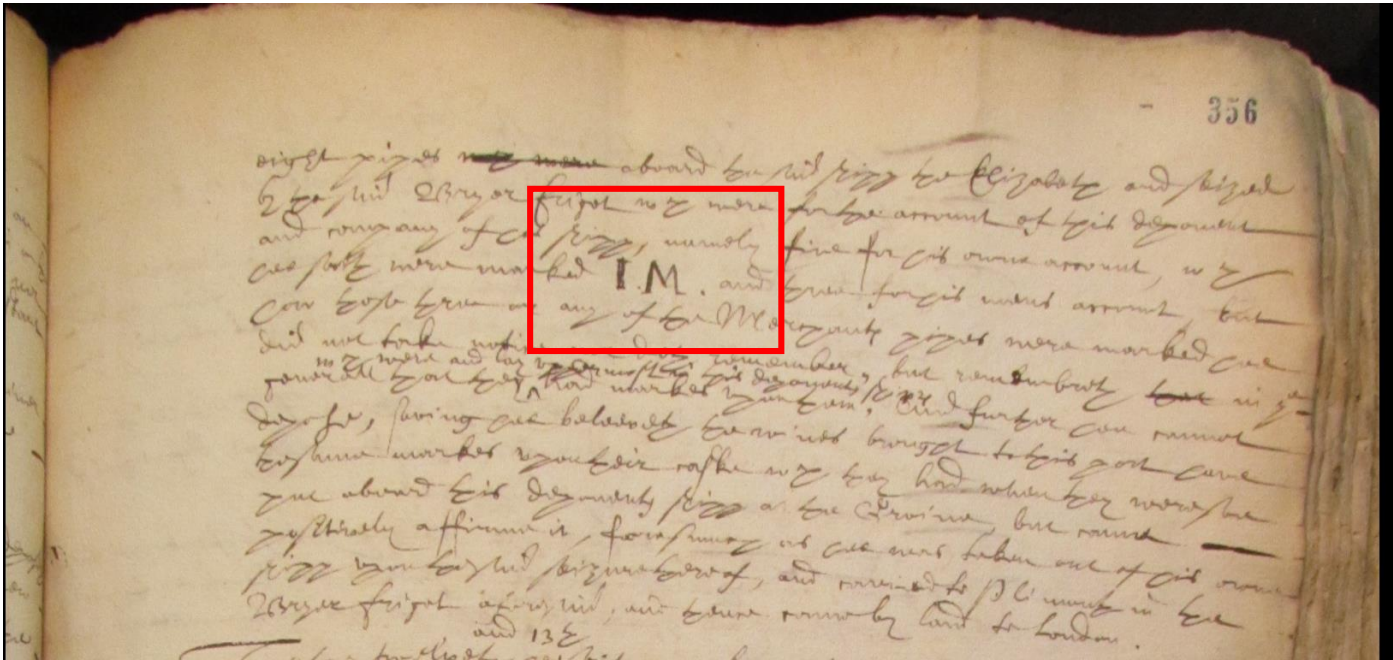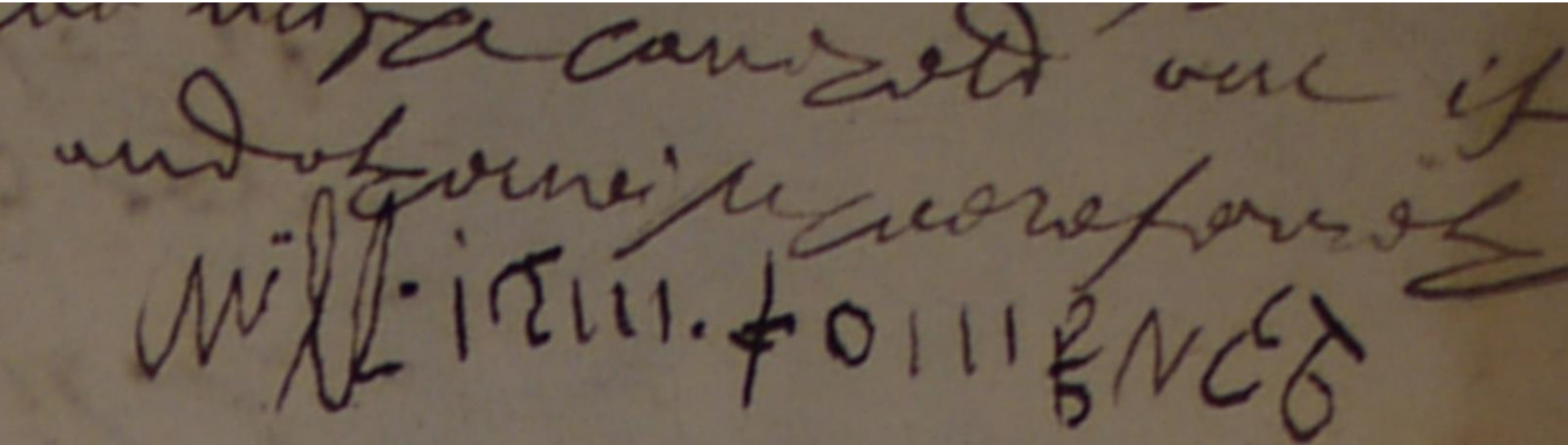
Phonetic difference between versions of signatures in legal documents
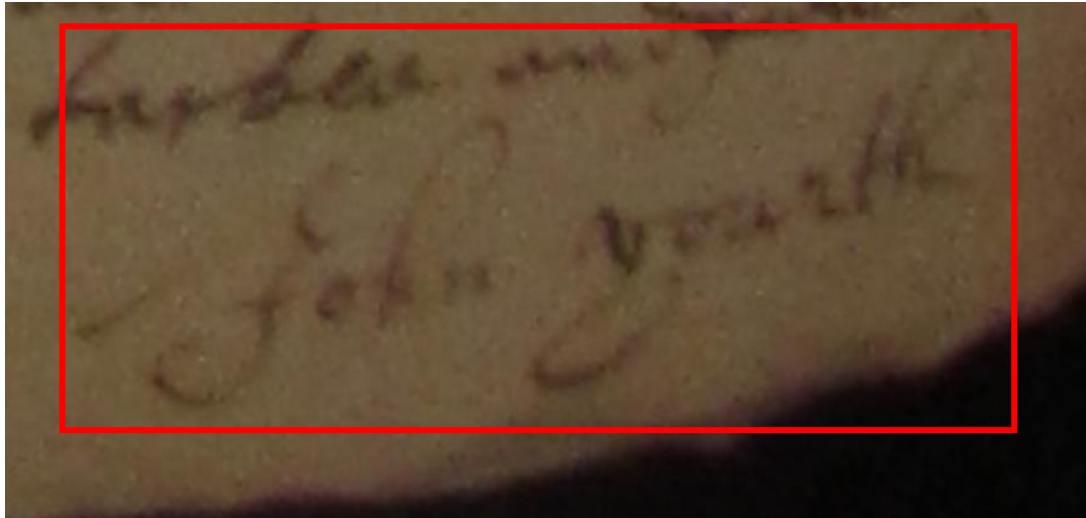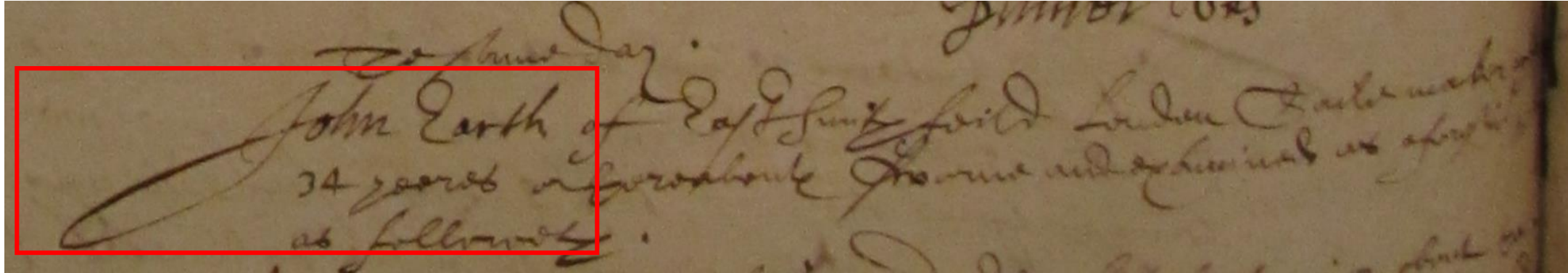
Front matter: "John Earth"
Signature: "[?John Yearth]"

Source: Extracts from HCA 13/68 f.543v

A very basic signature – few ligatures, reversed letters

William Tomkins, a fifty-two year old mariner of Fleet near Southampton. Pilot of the *Prosperous*. Deposed in the English High Court of Admiralty, June 12th, 1656

Source: KaggleTestSnippet_HCA_1371_f.246r.PNG

Occupations in front matter associated with "One of the company" in body of text

Table 21.2: Extract from occupational data sub-set in SOLM-2018 data set

Occupations in front matter associated with "One of the company" in body of text

Data as of 12/09/2018

Occupations stated in main metadata	Additional occupations & titles in body of deposition
Carpenter [on ship] [of the Great Christopher]	One of the company [of the Great Christofer]
Mariner	Boatswain [of the Liesde]; One of the company [of the Liesde]
Mariner	Foremastman; One of the company
Mariner	One of the company
Mariner	One of the company [of the Peter and Jane, burthen = 250 tons]; Prisoner [in Lisbon]
Mariner	One of the company [of the ffreetrade]
Mariner	One of the company
Mariner	One of the company; Cooke [cook]
Mariner	One of the company; Common man of the company
Mariner	One of the company [of the Olive Branch]; Master [of the Olive Branch, on death of its master during voyage]
Mariner	One of the company [of the John]; Steersman [of a Spanish man of war, the Angell, under a Claes Johnson]; Appointed acting master [of the Angell, by Claes Johnson, while he went overland to Saint Domingo]
Mariner	Coxen [of the ffrancis and John]; One of the company [of the ffrancis and John]
Mariner; Boatswain [of the Hopewell]	One of the company
Mariner; Boatswain [of the Seven Sisters, a ship granted a Commission]	One of the company [of the Seven Sisters, a ship granted a Commission]
Sailor	One of the company
Sailor	Cook [of the Saint Peter]; One of the company [of the Saint Peter]
Sailor	One of the company [of the Endeavour of Arundell]
Sailor	One of the company [of the Maidstone frigot in the immediate servce of the English Commonwealth]

Source: Extract from SOLM-2018 database, as of 12/09/2018

Processes for automation

Collective intelligence

Help us build a bot

Volunteers

ArchiveBot

Identify manuscript page with signoff

Establish type of signoff [marke, initial(s), signature]

Establish linked front matter of deposition

Establish foliation & image record numbers of signoff & front matter

Establish name of deponent & age

Establish residence of deponent

Establish occupation of deponent from front matter

Establish further occupational data from full text of manuscript

Data mine full text for social & economic parameters

Machine learning

Machine learning

Machine learning

Machine learning + database interrogation via API

HTR/Key word detection + NLP

HTR/Key word detection + NLP

HTR/Key word detection + NLP

HTR/Key word detection + NLP

Key word/phrase detection + proximity analysis

Facilitated archival volunteer teams

We are experimenting with key phrase and Boolean proximity search

Help us build a bot

Identify manuscript page with signoff

Establish type of signoff [marke, initial(s), signature]

Establish linked front matter of deposition

Establish foliation & image record numbers of signoff & front matter

Establish name of deponent & age

Establish residence of deponent

Establish occupation of deponent from front matter

Establish further occupational data from full text of manuscript

Data mine full text for social & economic parameters

Text mining using short phrases

Reading skills

"at his examination" [x10]
"now shewed to him"
"perused" [x350]
"perused the schedule" [x14]
"read or write"
"read to him" [x22]
"signed by this deponent"
"the schedule"
"write and read" [x18]
"write nor read"
"written hand" [x35]

BUT: "the same being read to him now and hee this deponent haveing formerly read the same himself" [HCA 13/71 f.232v]

Writing skills

"a writing" "firme" [x163]
"firmed" [x33]
"handwriting" [x166]
"his hand(e)"
"his usuall manner" [x7]
"in writ(e)ing"
"letter of advise"
"marke" [x1365]
"owne handwriting" [x16]
"proper handwriting" [x30]
"proper hand writing" [x34]
"signature" [x3515]
"signed sealed and delivered"
"subscribe his marke" [x1]
"subscribe his name" [x18]
"subscribed with" [x47]
"subscription" [x80]

"the firme" [x33]
"the letter"
"the signeing"
"usuall manner of" [x5]
"writeing" [x122]
"writings" [x269]
"writing his name" [x2]

Language skills

"dutch tongue" [x21]
"french tongue" [x8]
"flemmish" [x29]
"flemmish speech" [x1]
"ffrench tongue" [x8]
"spanish tongue" [x15]
"spoken" [x216]
"tongue" [x144]

Experience of seafaring

"bin at sea" [x45]
"by goeing to sea" [x10]
"by profession"
"by the space of" [x187]
"from his youth" [x21]
"goe to sea" [x79]
"going to sea" [x10]
"master of shipps" [x22]
"navigation"
"used seafaring" [x8]
"used the sea" [x45]
"used the seas" " [x6]
"used to goe to sea" [x12]

Birth place

"borne" [x923]
"hath lived" [x535]
"hee hath lived" [x134]
"lifetime" [x3]
"life tyme" [x63]
"living all his life tyme"

Occupation (past & present)

"being one of the" [x255]
"by profession" [x78]
"he knoweth being" [x119]
"hee knoweth being" [x630]
"knoweth being" [x937]

Worth of estate

"an estate" [x24]
"debts paid" [x32]
"estate" [x653]
"his debts paid" [x11]
"knoweth not" [x1,786]
"owes but" [x3]
"what he is worth" [x7]
"worth but little" [x1]

General

"all which he knoweth" [x30]
"all which hee knoweth" [x346]
"the premisses hee knoweth" [x95]
"the premisses hee knoweth for that" [x14]
"well acquainted with"

Full text mining

Text mining using short phrases

Reading skills

"at his examination" [x10]
"to be read" [x8]
"not well reade"
"now shewed to him" [x15]
"perused" [x350]
"perused the schedule" [x14]
"read" [x556]
"read or write" [x2]
"read to him" [x22]
"shewed him" [x34]
"signed by this deponent"
"the schedule"

"write" [x407]
"write and read" [x18]
"write nor read"
"written hand" [x35]

BUT: "the same being read to him now and hee this deponent haveing formerly read the same himself" [HCA 13/71 f.232v]

Writing skills

"a writing"
"drawe up"
"firme" [x163]
"firmed" [x33]
"handwriting" [x166]
"his hand(e)"
"his usuall manner" [x7]
"in writ(e)ing"
"letter of advise"
"marke" [x1365]
"owne handwriting" [x16]
"proper handwriting" [x30]
"proper hand writing" [x34]
"signature" [x3515]
"signed sealed and delivered"
"subscribe his marke" [x1]
"subscribe his name" [x18]
"subscribed with" [x47]
"subscription" [x80]

"the firme" [x33]
"the letter"
"the signeing"
"usuall manner of" [x5]
"writeing" [x122]
"writings" [x269]
"writing his name" [x2]

Language skills

"dutch tongue" [x21]
"french tongue" [x8]
"flemmish" [x29]
"flemmish speech" [x1]
"ffrench tongue" [x8]
"Italian tongue" [x11]
"spanish tongue" [x15]
"speake" [x423]
"speake it" [x3]
"spoken" [x216]
"tongue" [x144]
"understand Spanish"
"understandeth"
"writeing and understanding" [x2]

Experience of seafaring

"bin a seaman"
"bin at sea" [x45]
"bin master" [x123]
"by goeing to sea" [x10]
"by the space of" [x187]
"from his youth" [x21]
"goe to sea" [x79]
"going to sea" [x10]
"gone to sea" [x22]
"master of shipps" [x22]
"navigation" [x99]
"these twenty yeares" [x41]
"used seafaring" [x8]
"used the sea" [x45]
"used the seas" [x6]
"used to goe to sea" [x12]

Birth place

"all his life" [x15]
"always dwelt"
"borne" [x923]
"borne there" [x28]
"by birth" [x172]
"childhood" [x28]
"fixed residencie"
"from his childhood" [x15]
"habitavit" [x?]
"hath lived" [x535]
"hee hath lived" [x134]
"hee is a native of" [x54]
"his aboade"

"lifetime" [x3]
"life tyme" [x63]
"lived all his time" [x10]
"living all his life tyme"
"native of" [x376]
"Stepney" + "borne at" [x25]
"Stepney" + "borne" [x63]
"Wapping" + "borne" [x58]

Occupation (past & present)

"being one of the" [x255]
"by profession" [x78]
"by profession a mariner"
"by trade" [?]
"commander of shipps" [x6]
"he knoweth being" [x119]
"hee knoweth being" [x630]
"knoweth being" [x937]
"master of ships" [x10]
"master of shipps" [x22]
"occupation"

Worth of estate

"an estate" [x24]
"debts paid" [x32]
"estate" [x653]
"fayled or broke"
"his debts paid" [x11]
"owes but" [x3]
"what he is worth" [x7]
"worth but little" [x1]

General

"all which he knoweth" [x30]
"all which hee knoweth" [x346]
"knoweth not" [x1,786]
"the premisses hee knoweth" [x95]
"the premisses hee knoweth for that" [x14]
"well acquainted with"

Source: Phrase searches performed on full text of MarineLives wiki corpus, as of 20/09/2018

Text mining using short phrases

Indirect hints as to occupation

"his booke not being nowe in his custodie" [x1] [hints at boatswain]
"his ship" [x151] [weakly hints at master]

Visualising metadata created with aid of ArchiveBots

We need visual metadata, which can be machine processed

Table 1.2a EXPANDED: HCA 13/53 [f.1r-340v] - Signoff frequency per manuscript page, data from 1637

	1 r	1 v	2 r	2 v	3 r	3 v	4 r	4 v	5 r	5 v	6 r	6 v	7 r	7 v	8 r	8 v	9 r	9 v	10 r	10 v	Subtotal
1-10	1	2	3	1	0	1	1	0	2	2				1		1			1		16
11-20	2	2	1		1		1		1		1	1		1		1			1		13
21-30	2	1		1		3	1	1		1		2		1			2			1	16
31-40		1	1	1			1	1	1	2		1	2	1		1					13
41-50		1		1					1			1		1					1		6
51-60		1		1		2	1			2		2			2						11
61-70	2		1	1			2	1	1	1	1	1		2	2	1	1	1			18
71-80	1		2	1		1	2	1	2		1	1	1	1	3			2			19
81-90	2	1	1	1	1	1	1	2	1	2	2	4	1	1		1		1			23
91-100	1	2		1			2	2	2	3	3	1	3		1	1	3	1			26
101-110	2	1	2	2	1	1	1	2	2	1	1	1	1		2	1	2				23
111-120	1			1			1	1	2	1		2	1	2	2	1	1				16
121-130			1				2		1	1	2			3	1		1				12
131-140	2	3	2	1	1	2			1		1	1	1	2	2	1	2		1		23
141-150	1	1	2	1	2	2	1	2		1	2	1	1	1		2		1			22
151-160		1				2	1	1	2	1	2	2	1	1	2			1			18
161-170		1		2	2	1				1	2	1	1	1	2	1	1	2			17
171-180	1		2		2	1	1	1	1		1						1				11
181-190				2							3	1	2	1	4	4	2	1	1		21
191-200		1		1			1	1	1		1		1		2	1	2	2			17
201-210	2	2			1	1			2	4	3				2	1	2	1	2		24
211-220	1		2	1			1		1		1	4	1	2	3	4	2	2			25
221-230	2	2	1	3			1	2	1	1	1	1	2	2	2	2		2			25
231-240		1	1	1		2	1	1			3		2	2				1			15
241-250	2								2	1	1	1	2	2	1		2	1			15
251-260	2		2	2	1		1	1	1	1	1		2		1			1			15
261-270		1	1		1		1		1	1		1		1				2			11
271-280	2			1			1		1	1	1		1		2	1		1			12
281-290	1		1		2	1		1	1	1		1	1	1	2			1			14
291-300		1		1	1	1	1	2		1	1	1	1		1	1		1		1	15
301-310		1		2					2	1		2	1	1		1			1		12
311-320		1					1						1	2	1						6
321-330		1						1			1					1	3				7
331-340	1	2	2	2		2		1	1	1	1	1	1	1	1		1				18
Total	31	30	24	27	23	17	23	30	19	31	31	39	31	33	30	42	33	16	29	16	**555**

Source: SOLM-2018 database

Archivists, computer scientists and users of all sorts need to work together

Source: SOLM-2018 database

Anticipating changing technology

MarineLives & its progeny: changing technology and processes

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023

Activities

Launched MarineLives ★

Hackathon

Maphackathon event ★

Hackathon

Launched Signs of Literacy ★ (Conferences)

★ Kaggle competition

Technologies + Processes used

Scripto + Wordpress

Wikispot — Semantic Media Wiki

Dropbox

Dropbox

HTR

HTR + Key word spotting

NLP

NLP

Text proximity mining

ArchivalBots

Collaborative processes

Facilitated remote teams with Skype

Facilitated remote teams with ZOOM

Data type & scale

1 mill full text words

1,500 images

Data migration

650 semantic biographies

6 mill full text words

12,000 images

1 mill images

Technology environment

Machine learning

Self-service digital imaging

IIIF API Ver 2.0