

SOLM-2018 machine learning training data set

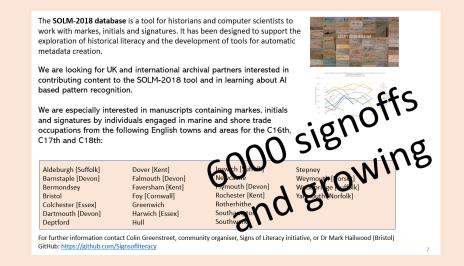
Our vision is a SOLM-2023 database with 1 million markes, initials & signatures from across Europe & North America from the C16th to C18th



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!!!!!



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.

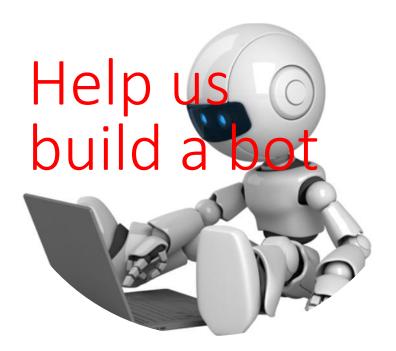
United Kingdom, Netherlands, Sweden, Poland, Germany, France, Spain & North America

Contact:

Colin Greenstreet (community organiser, Signs of Literacy initiative) or Dr Mark Hailwood (Bristol)

GitHub: https://github.com/Signsofliteracy/
Web: http://signsofliteracy.org

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



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

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

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

Volunteers ArchiveBot Machine learning Machine learning archival volunteer teams Machine learning Machine learning + database interrogation via API HTR/Key word detection + NLP HTR/Key word detection + NLP HTR/Key word detection + NLP **Facilitated** HTR/Key word detection + NLP Key word/phrase detection + proximity analysis

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]

Birth place

"borne" [x923]
"hath lived" [x535]

"hee hath lived" [x134]

"lifetime" [x3]

"life tyme" [x63]

"living all his life tyme"

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]

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"



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**

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

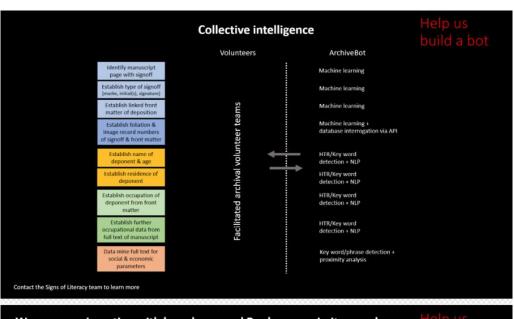
Call for ideas

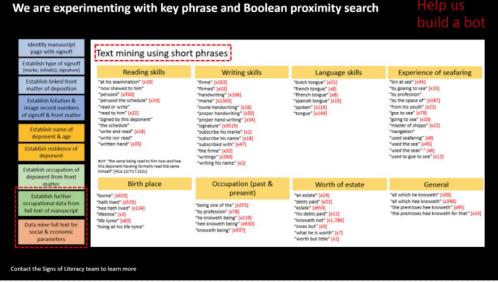


Nesta is offering grants of up to £20,000 for projects that generate new knowledge on how to advance collective intelligence (combining human and machine intelligence) to solve social problems.

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
- . 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
- · Pilating 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
- · Charities and public sector organisations piloting ways to increase their internal/organisational

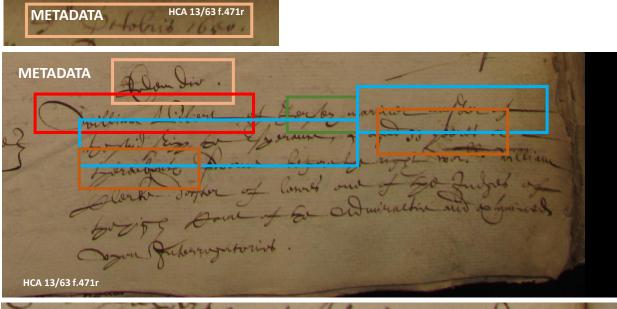


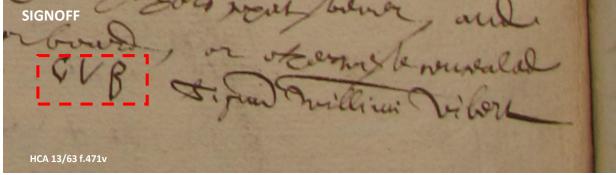


Imagine an ArchiveBot

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







Occupation

Age

Date

Name

Residence



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)

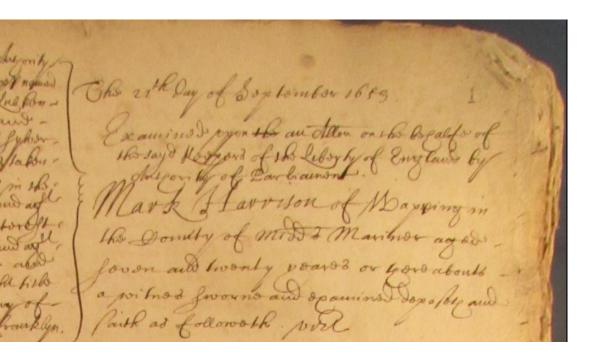
Challenge for a Bot

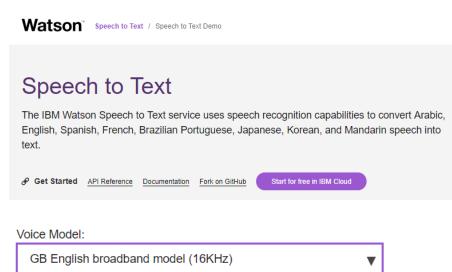
- (1) Find the date of the deposition (it is at the top of the same manuscript page above the preceding deposition)
- (2) Deal with the interlineation in the metadata
- (3) On the next manuscript page, recognise "GVB" are initials not a marke. The "G" is quite tough (easy to mistake for "C"). Yes, they 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

Speech to text recognition







mark Harrison³ of walking in the county of Middlesex mariner aged seven and twenty years or there about mark Harrison³ of walking in the county of Middlesex mariner aged seven and twenty years or there about mark Harrison³ of what happened in the county of Middlesex mariner aged seven and twenty years or there amark Harrison³ or walking in the county of Middlesex mariner aged seven and twenty years or there about mark Harrison³ or walking in the county of Middlesex mariner aged seven and twenty years or the about about

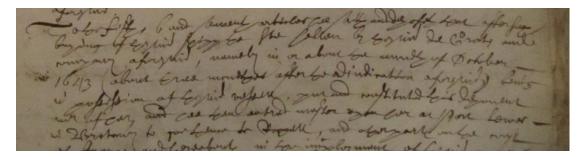
Can we use key word spotting to excavate raw metadata?

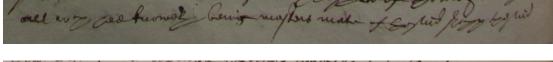
Signs of Literacy initiative

LANGUAGE DENOTING OCCUPATION



Jose for for fat he having bein the for for for fays -





how francis week of the purify and more all for find royage in cer.

"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.674r] [CONCLUSION: Boatswain's mate]

KaggleTestSnippet_HCA_1370_f.546r.PNG	HCA 13/70	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1370_f.571v.PNG	HCA 13/70	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1370_f.596v_One.PNG	HCA 13/70	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1370_f.636r.PNG	HCA 13/70	Signature	Mariner; Principal hoatswain	
KaggleTestSnippet_HCA_1370_f.671v.PNG	HCA 13/70	Marke 🧲	Mariner; Boatswain's mate	
KaggleTestSnippet_HCA_1368_f.631v.PNG	HCA 13/68	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1371_f.27r.PNG	HCA 13/71	thitials	Mariner; Boatswain	
KaggleTestSnippet_HCA_1371_f.27v_One.PNG	HCA-13/71	Initials	Mariner; Boatswain	
KaggleTestSnippet_HCA_1371_f.27v_Two.PNG -	HCA 13/71	Initials	Mariner; Boatswain	
KaggleTestSnippet_HCA_1368_f.640r.PNG	HCA 13/68	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1368_f_687r.PNG - CREATE	HCA 13/68	Signature	Mariner; Boatswain [of the Civi	l Society]
KaggleTestSnippet_HCA_1371_f.77v.PNG	HCA 13/71	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1370_f.378r.PNG	HCA 13/70	Signature	Mariner; Boatswain	
KaggleTestenippet_HCA_1371_f.99r.PNG	HCA 13/71	Signature and	Mariner; Boatswain [of man of	war]
KaggleTestSnippet_HCA_1370_f.484v.PNG	HCA 13/70	Signature	Mariner; Quartermaster; Boatsv	vain
KaggleTestSnippet_HCA_1371_f.139v.PNG	HCA 13/71	Signature	Mariner; Boatswain	
KaggleTestSnippet_HCA_1371_f.167r.PNG	HCA 13/71	Signature	Mariner; Boatswain [of the John	and Mary]
KaggleTestSnippet HCA 1371 f.279r.PNG	HCA 13/71	Signature	Mariner; Boatswain	



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



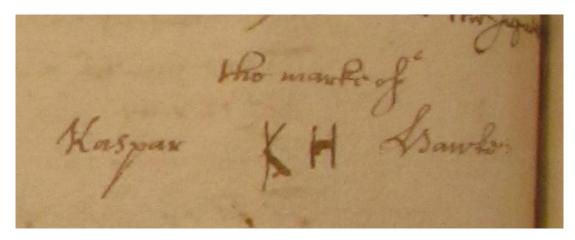
Table 21.1: Extract from occupational data sub-set in SOLM-	-2018 data set		
Data as of 11/09/2018			
		Additional occupations & titles in body of	
Occupations [Synthesised]	Occupations stated in main metadata	deposition	Stated activities in body of text
BOATSWAINS			
Seaman; Commonman [outwards]; Boatswain [return]	Seaman	Commonman [outwards]; Boatswain [return]	
Nauta [mariner]; Coxwain [outwards] [on Thomas and	Nauta [mariner]	Coxwain [outwards] [on Thomas and John]; Boatswain	
John]; Boatswain [return] [on Thomas and John]		[return] [on Thomas and John]	
Nauta [mariner]; Boatswain's mate [of Thomas and	Nauta [mariner]; Boatswain's mate [of Thomas and	[?Part-owner of ship Thomas and John]	
John]; [?Part-owner of ship Thomas and John]	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	Mariner; Boatswain [of the Hopewell]	One of the company [of the Hopewell]	
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	Mariner	Boatswain [of the Liesde] One of the company [of the	
company [of the Liesde]		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	Sailor		Worked on building of ship [the Great Christofer]
Christofer]		<u> </u>	

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

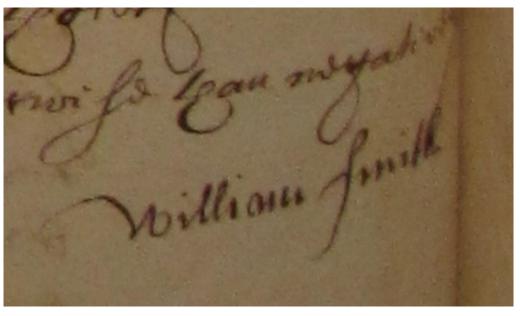
Note information on activities of deponents in main body of text helps clarify and refine the main occupational descriptions

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]

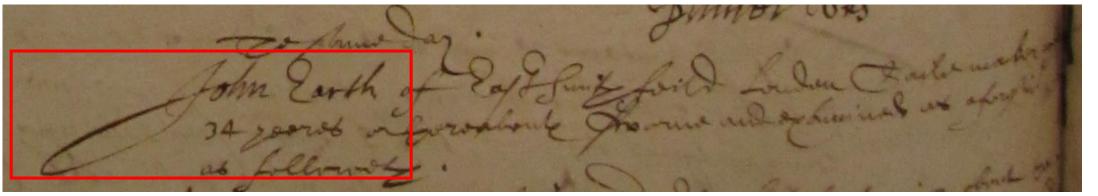


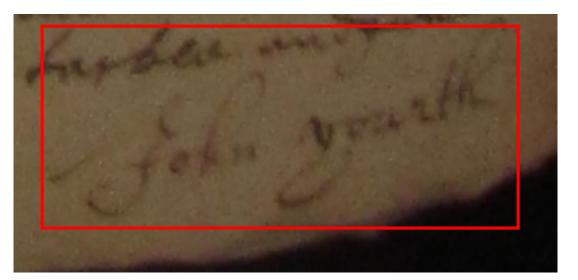
Front matter: William de Smitt

Signature: William Smith

Phonetic difference between versions of signatures in legal documents







Front matter: "John Earth" Signature: "[?John Yearth]"

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

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



Table 21.2: Extract from occupational data sub-set in SOL	
Occupations in front matter associated with "One of the	company" in body of text
Data as of 12/09/2018	
	Aller I a Carl I I I
	Additional occupations & titles in body of
Occupations stated in main metadata	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	One of the company [of the Seven Sisters, a ship
granted a Commission]	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

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

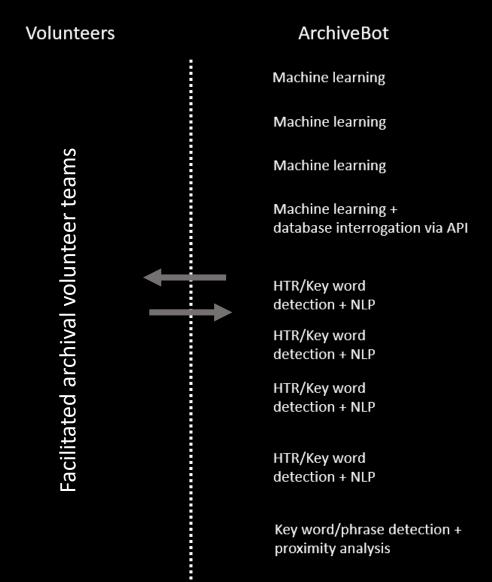
Establish name of deponent & age

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 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]

Birth place

"borne" [x923]
"hath lived" [x535]

"hee hath lived" [x134]

"lifetime" [x3]

"life tyme" [x63]

"living all his life tyme"

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

"subscribe his marke" [x1]
"subscribe his name" [x18]

"subscribed with" [x47]
"subscription" [x80]

g skills

"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]

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"

"well acquainted with"

Full text mining

Text mining using short phrases

Reading skills

"at his examination" [x10]
"to be read"
"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]

Birth place

```
"borne" [x923]
"borne there"
"by birth"
"childhood" [x28]
"from his childhood" [x15]
"hath lived" [x535]
"hee hath lived" [x134]
"lifetime" [x3]
"life tyme" [x63]
"lived all his time" [x10]
"living all his life tyme"
"hee is a native of"
```

Writing skills

"the firme" [x33]

"writeing" [x122]

"writings" [x269]

"usuall manner of" [x5]

"writing his name" [x2]

"the letter"

"the signeing"

"a writing" "firme" [x163] "drawe up" "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]

Language skills

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

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 to goe to sea" [x12]

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]

"fayled or broke"

"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"

Text mining using short phrases

Indirect hints as to occupation

"his booke not being nowe in his custodie" [x1] [hints at boatswain]

[&]quot;his ship" [x151] [weakly hints at master]

[&]quot;his ship" [x450] [weakly hints at master]

Visualising metadata created with aid of ArchiveBots

We need visual metadata, which can be machine processed

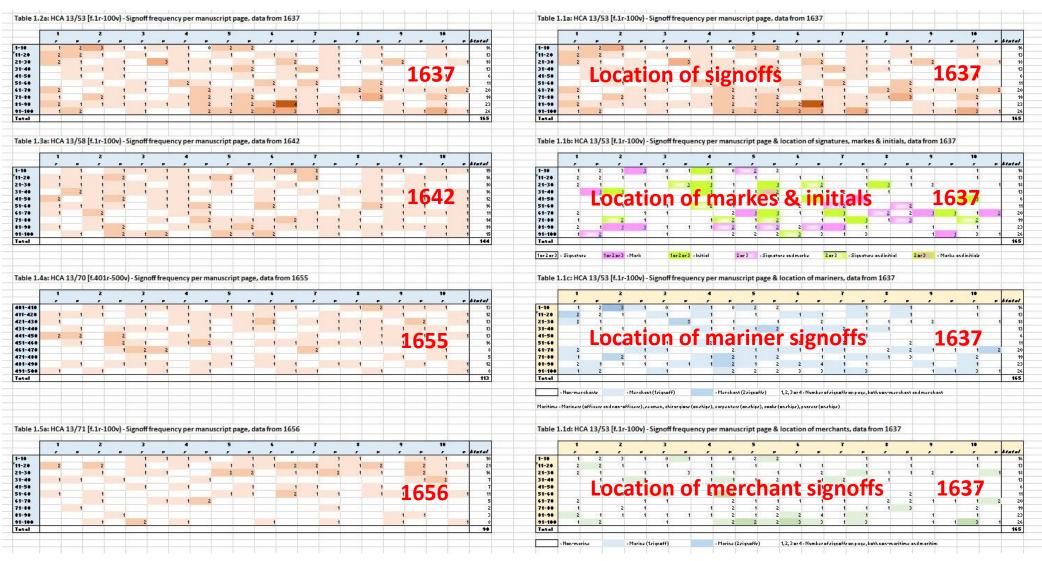


Table	1.2a	EXPA	NDE): HC	4 13/	′53 [f	.1r-3	40v) ·	- Sign	off fr	equer	ncy p	er ma	nusc	ript p	age,	data	from	1637	7	
	1		2 r	v	3 r	v	4	v	5 r	ν	6	ν	7 r	v	8 r	v	9 r	v	10 r	v	Subtotal
1-10	1	. 2	3	1	0	1	1	0	2	2				1		1			1	•	16
11-20	2	2 2	1		1		1		1		1	1		1		1			1		13
21-30	2	2 1		1		3	1	1		1		2		1			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	2		1	1			2	1_	1	1	1	1		2	2	1	1	1		18
71-80	1	l	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	2 1	2	2	1	1	1	2	2	1	1	1	1		2	1	2				23
111-120	1	L		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	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	1		2		1		22
151-160		1					2		1	1	2	1	2	2	1	1	2		1	1	18
161-170		1		2	2	1					1	2	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	1		1	1	2	1	2	2		17
201-210	2	2 2		2		1	1			2	4	3 1		1	2	2	1	2	1	2	24
211-220 221-230	1		- 1	2	1		1	2		1	1	1	4	2	2	2	2	2	2	2	25 25
231-240	4	1	1		1		2	1	1	1	1	2	1	2	2	2	2			1	15
241-250	2	2	1		1		2	1	1		2	1	1	1	2	2	1		2	1	15
251-260	2		2	2	1		1	1		1	1	1	1	1	1	2	1		2	1	15
261-270		1	1	2	1		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			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

