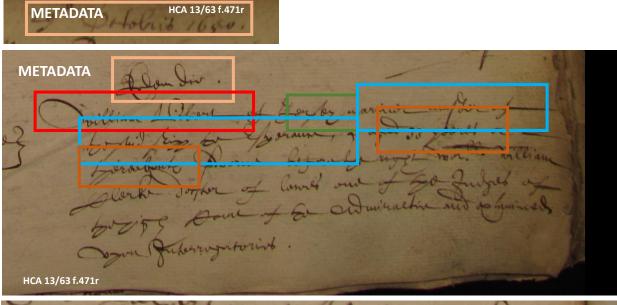
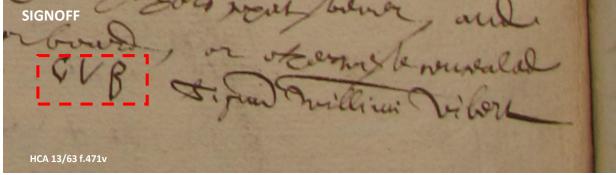


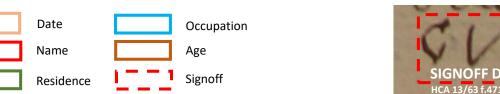
Imagine an ArchiveBot

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











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.

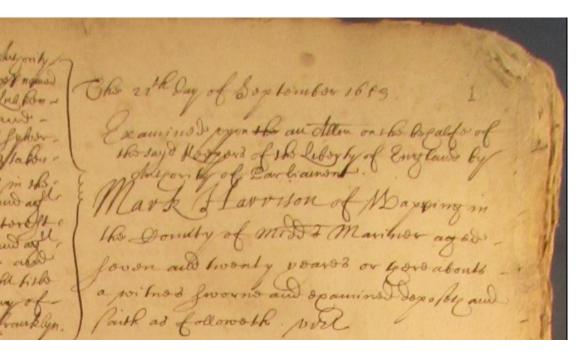
The choice of "G", "V" and "B" for initials suggests he thought of himself and pronounced his name as "Guillaume Vee Bert" (though he could not write a full signature)

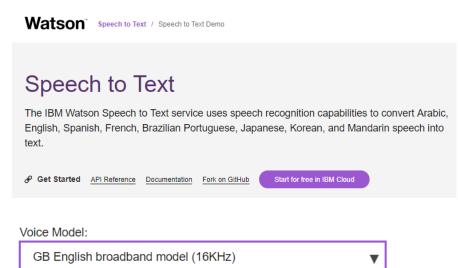
Challenge for a Bot

- (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 Harriso² and of⁴ what happened⁷ in² the county of Middlesex mariner⁸ aged seven and twenty years

mark Harrison³ of² walking⁸ in² the county of Middlesex mariner¹⁵ aged seven and twenty years or² there⁴ about

mark² Harrison³ of² what⁴ happened in the county of Middlesex mariner¹⁰ aged seven and twenty years or⁴ there⁴

mark² Harrison³ or³ walking in the county of Middlesex mariner⁸ aged² seven and twenty years³ or² the³ about⁵

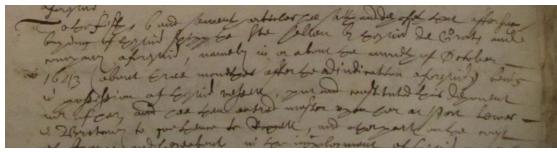
Can we use key word spotting to excavate raw metadata?

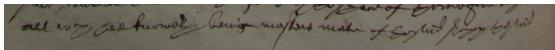
Signs of Literacy initiative

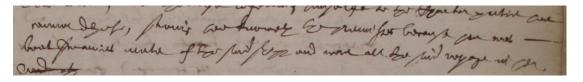
LANGUAGE DENOTING OCCUPATION



for of London for for and Dougany, Chi ships the days of the days for fait he having bein those formen his all you by door for fait he having bein those formen his all you from the formen to form the formen for fait the fays -







"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 Civill 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
KaggleTest8nippet_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; Boatswain
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]		_		

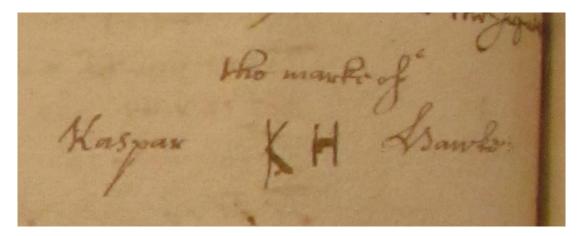
Note four out of thirteen boatswains identified here are described simply as seaman, nauta or mariner in the legal front matter

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

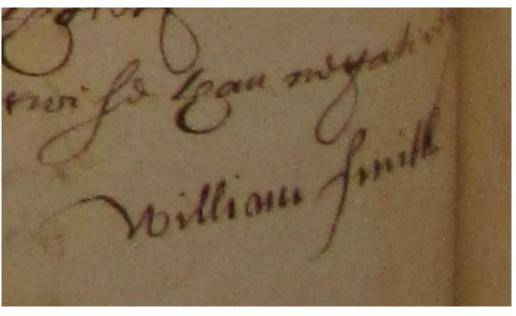
7

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]

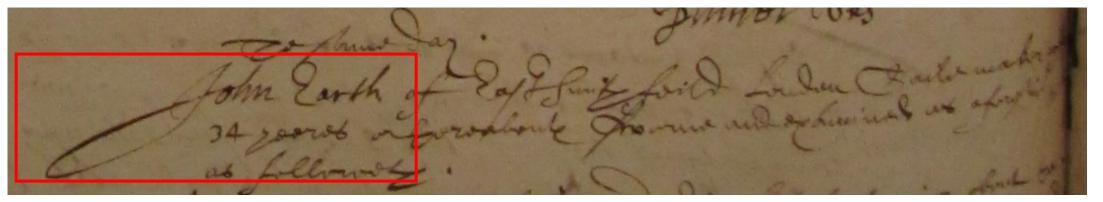


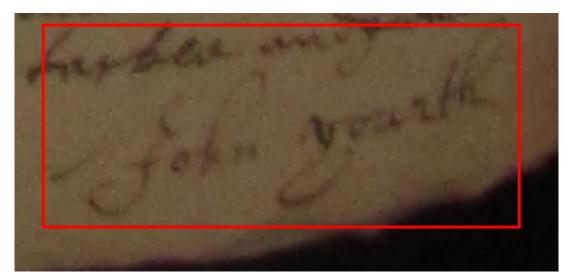
Front matter: William de Smitt

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



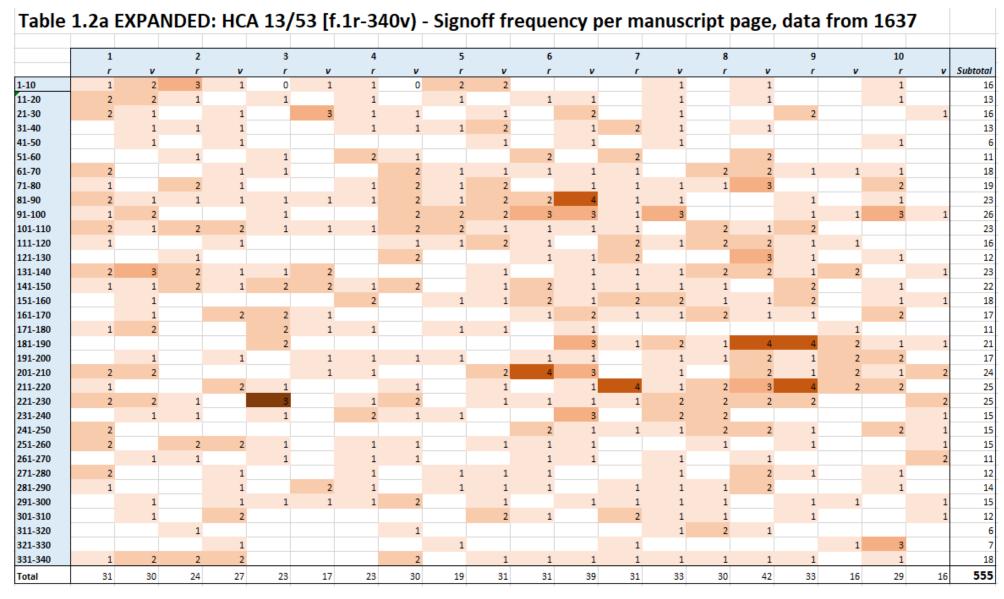
Occupations in front matter associated with "One of the	company" in body of text		
Data as of 12/09/2018			
	Additional occupations & titles in body of		
Occupations stated in main metadata	→ deposition ▼		
Carpenter [on ship] [of the Great Christopher]	One of the company [of the Great Christofer]		
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

Visualising metadata created with aid of ArchiveBots

We need visual metadata, which can be machine processed

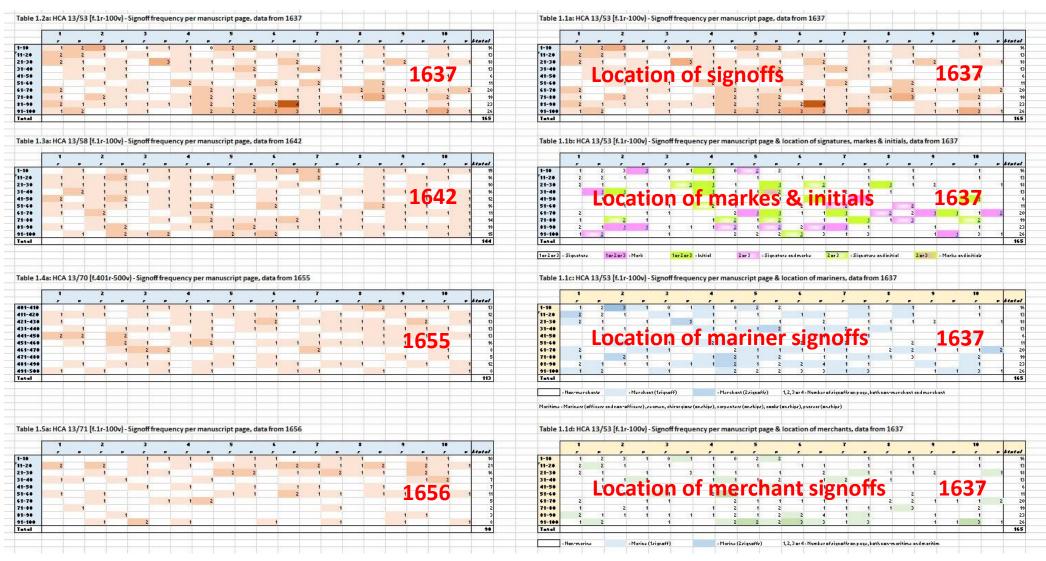




Source: SOLM-2018 database

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



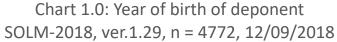


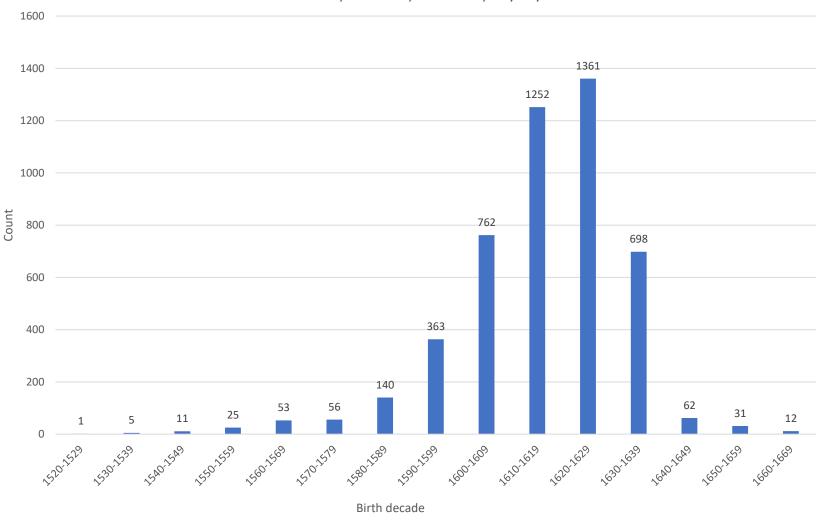
Source: SOLM-2018 database

SOLM-2018 machine learning training data set

The SOLM-2018 data set - nearly 5,000 markes, initials and signatures for individuals born between 1520 and 1669







15 SOLM-2018, as of 12/09/2018

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

17