

DRAFT FOR REVIEW

SOLM-2018 data set

# Archives & AI

Tuesday, September 4<sup>th</sup>, 2018  
The National Archives  
United Kingdom

# Project portfolio

<http://www.chronoscopic.org>

## MarineLives



## Signs of Literacy



## EM Textiles, Garments & Dyestuffs Glossary



## Maphackathon



## SOLM-2018



## EM Maritime & Mercantile Gazetteer



# Some perspective

Labeled Faces in the Wild

Labeled Faces in the Wild Home

UNIVERSITY OF MASSACHUSETTS AMHERST, MASS.

Menu

- LFW Home
  - Mailing
  - Explore
  - Download
  - Train/Test
  - Results
  - Information
  - Errata
  - Reference
  - Resources
  - Contact
  - Support
  - Changes
- Part Labels
- UMass Vision

**NEW SURVEY PAPER:**  
Erik Learned-Miller, Gary B. Huang, Aruni RoyChowdhury, Haoxiang Li, and Gang Hua.  
**Labeled Faces in the Wild: A Survey.**  
In *Advances in Face Detection and Facial Image Analysis*, edited by Michal Kawulok, M. Emre Celebi, and Bogdan Smolka, Springer, pages 189–248, 2016.  
[Springer Page] [Draft pdf]

**NEW RESULTS PAGE:**  
WE HAVE RECENTLY UPDATED AND CHANGED THE FORMAT AND CONTENT OF OUR [RESULTS PAGE](#). PLEASE REFER TO THE [NEW TECHNICAL REPORT](#) FOR DETAILS OF THE CHANGES.

Welcome to Labeled Faces in the Wild, a database of face photographs designed for studying the problem of unconstrained face recognition. The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set. The only constraint on these faces is that they were detected by the Viola-Jones face detector. More details can be found in the technical report below.

There are now four different sets of LFW images including the original and three different types of "aligned" images. The aligned images include "funneled images" (ICCV 2007), LFW-a, which uses an unpublished method of alignment, and "deep funneled" images (NIPS 2012). Among these, LFW-a and the deep funneled images produce superior results for most face verification algorithms over the original images and over the funneled images (ICCV 2007).

**Related:**

[new] Collected resources related to LFW - updated 2017/05/09.  
LFW Deep Funneled Images.  
LFW attributes file (see Attribute and Simile Classifiers for Face Verification, Kumar et al.).  
Face Detection Data set and Benchmark (FDDB), our new database for face detection research.  
Faces in Real-Life Images workshop at the European Conference on Computer Vision 2008, run by Erik Learned-Miller, Andras Ferencz, and Frederic Jurie.

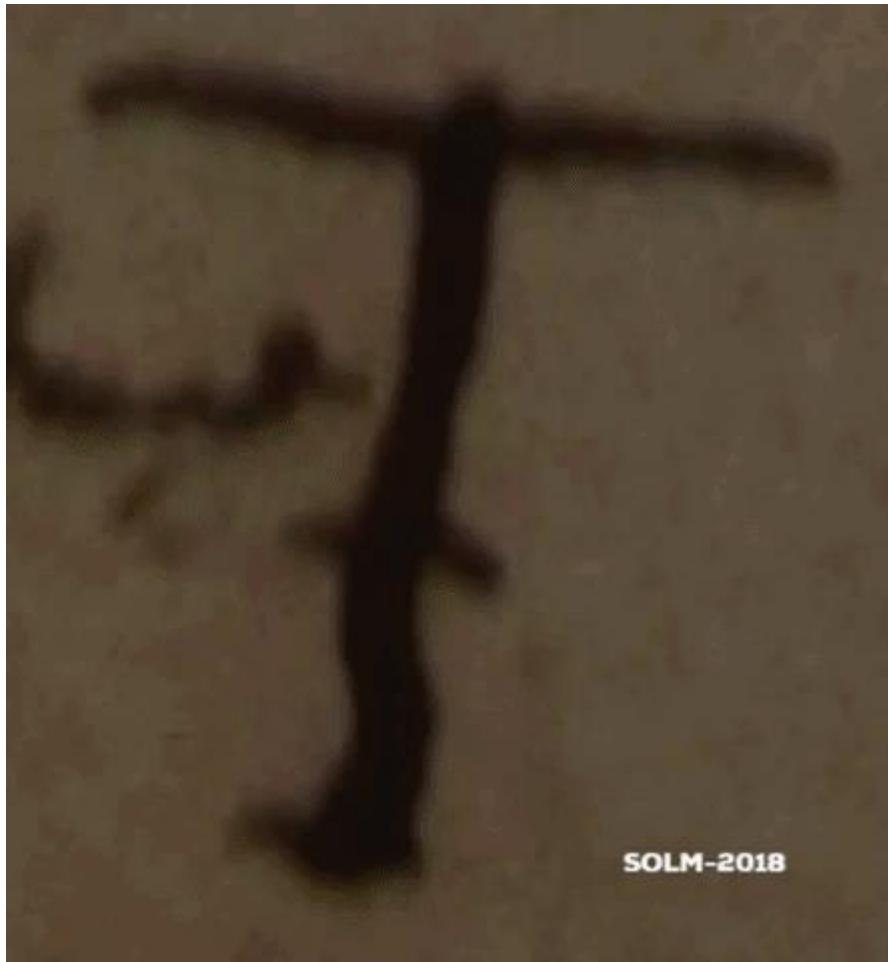
**Abstract** In 2007, Labeled Faces in the Wild was released in an effort to spur research in face recognition, specifically for the problem of face verification with unconstrained images. Since that time, more than 50 papers have been published that improve upon this benchmark in some respect. A remarkably wide variety of innovative methods have been developed to overcome the challenges presented in this database. As performance on some aspects of the benchmark approaches 100% accuracy, it seems appropriate to review this progress, derive what general principles we can from these works, and identify key future challenges in face recognition. In this survey, we review the contributions to LFW for which the authors have provided results to the curators (results found on the LFW results web page). We also review the cross cutting topic of alignment and how it is used in various methods. We end with a brief discussion of recent databases designed to challenge the next generation of face recognition algorithms.

## Labeled Faces in the Wild: A Survey

Erik Learned-Miller, Gary Huang, Aruni RoyChowdhury, Haoxiang Li, Gang Hua

# Pattern seeking

How many different letters can you recognise?



Click for animated GIF

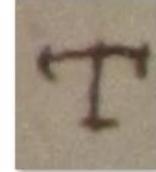
## Initials – Ts and Js



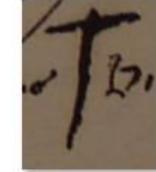
T6.png



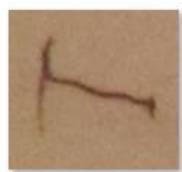
T7.png



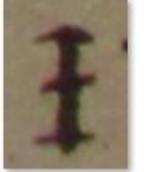
T8.png



T9.png



T10.png



J5.PNG



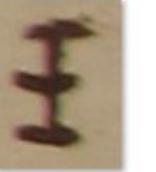
J6.PNG



J7.PNG



J8.PNG



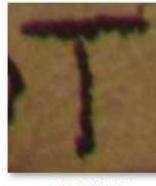
J9.PNG



T11.PNG



T12.PNG



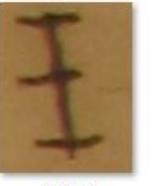
T13.PNG



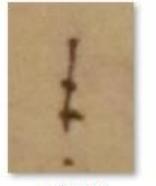
T14.PNG



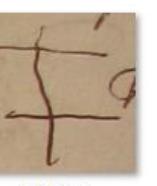
T15.PNG



J10.PNG



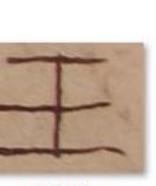
J11.PNG



J12.PNG



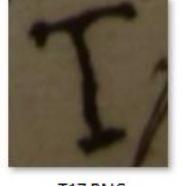
J13.PNG



J14.PNG



T16.PNG



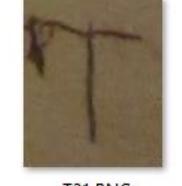
T17.PNG



T18.PNG



T20.PNG



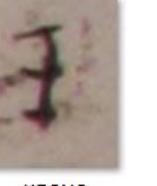
T21.PNG



J15.PNG



J16.PNG



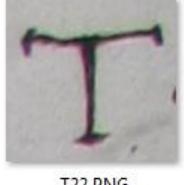
J17.PNG



J18.PNG



J19.PNG



T22.PNG



T23.PNG



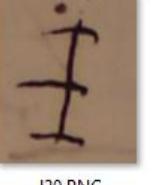
T24.PNG



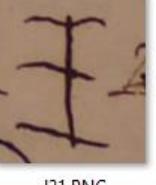
T25.PNG



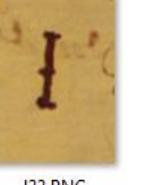
T26.PNG



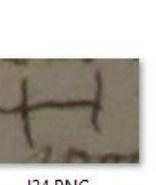
J20.PNG



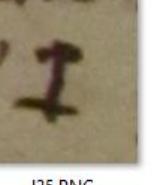
J21.PNG



J22.PNG



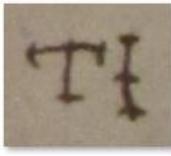
J24.PNG



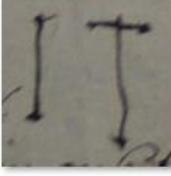
J25.PNG



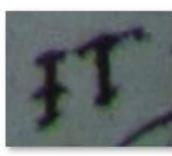
T28.PNG



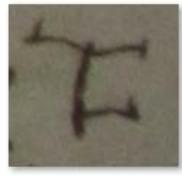
TJ28.PNG



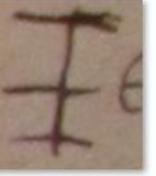
TJ30.PNG



TJ31.PNG



TL29.PNG



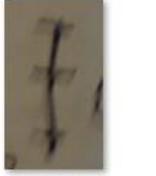
J26.PNG



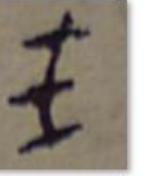
J27.PNG



J28.PNG



J30.PNG



J31.PNG

# SOLM-2018 database

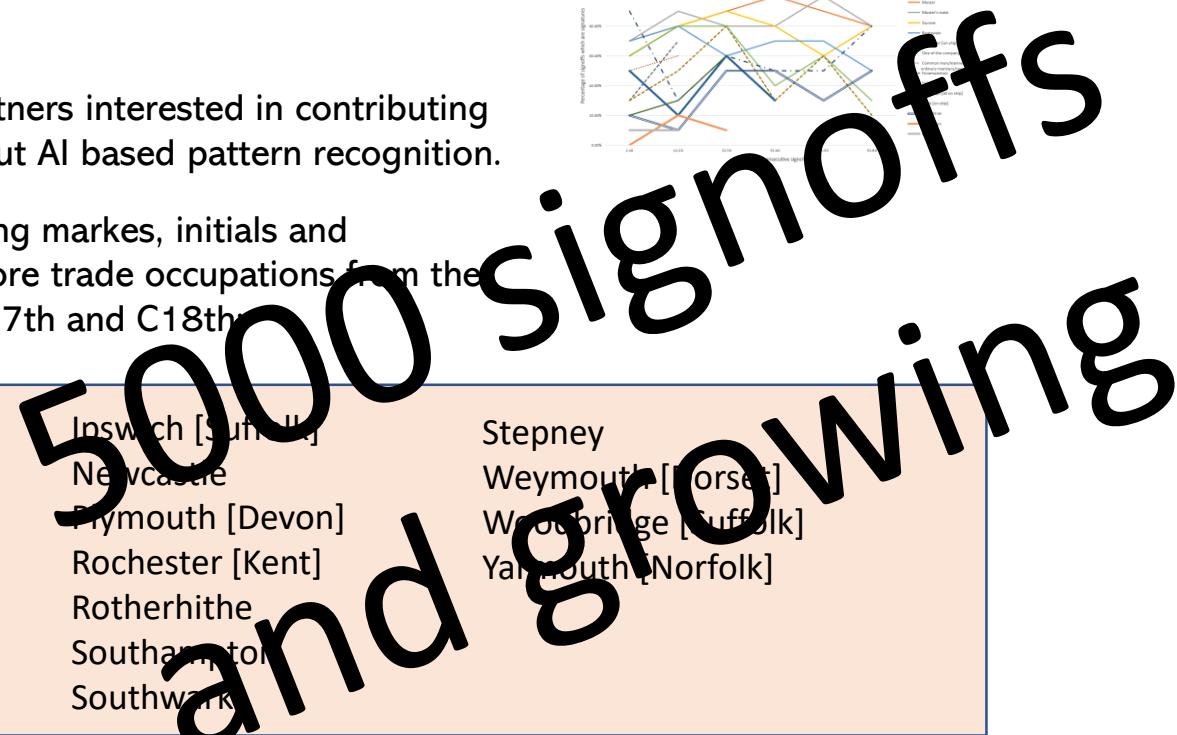
The **SOLM-2018 database** is a tool for historians and computer scientists to work with marks, initials and signatures. It has been designed to support the exploration of historical literacy and the development of tools for automatic metadata creation.

We will be previewing the database at the TNA Archives & AI symposium on Tuesday, September 4<sup>th</sup> and at the Sheffield Digital Humanities Congress on Thursday, September 6<sup>th</sup>, 2018.

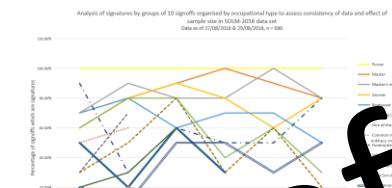
We are looking for UK and international archival partners interested in contributing content to the SOLM-2018 tool and in learning about AI based pattern recognition.

We are especially interested in manuscripts containing marks, 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]
Barnstaple [Devon]	Falmouth [Devon]
Bermondsey	Faversham [Kent]
Bristol	Foy [Cornwall]
Colchester [Essex]	Greenwich
Dartmouth [Devon]	Harwich [Essex]
Deptford	Hull



SOLM-2018 data set



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

# Our vision is a SOLM-2023 database with 1 million marks, 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 15 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!!!!

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

We will be previewing the database at the TNA Archives & AI symposium on Tuesday, September 4<sup>th</sup> and at the Sheffield Digital Humanities Congress on Thursday, September 6<sup>th</sup>, 2018.

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 marks, 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]	Scarborough [North Yorks]
Barnstaple [Devon]	Falmouth [Devon]	Newcastle	Weymouth [Dorset]
Bermonsey	Faversham [Kent]	Plymouth [Devon]	Woodbridge [Suffolk]
Bristol	Foy [Cornwall]	Rochester [Kent]	Yarmouth [North Yorks]
Colchester [Essex]	Greenwich	Rotherhithe	
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>

5000 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 partners in the United Kingdom and internationally

**Kaggle Competitions**

**Bristol Archives**  
 Home to ten centuries of history: we collect + preserve Bristol's archives (and the British Empire & Commonwealth Collection) for current + future generations.  
 Bristol, United Kingdom  
[bristolarchives.org.uk/bris](http://bristolarchives.org.uk)  
 Joined April 2011

**The National Archives**

**THE HUNTINGTON**

**DORSET HISTORY CENTRE**

**Devon Archives**  
 Tweets from the team at the South West Heritage Trust about Devon's fascinating archive collection. #SWHTdiscovery  
 Exeter  
[swheritage.org.uk/#archives/cmcl](http://swheritage.org.uk/#archives/cmcl)  
 Joined January 2012

**TYNE & WEAR ARCHIVES & MUSEUMS**

**TWAM Museums**  
 Regional museum, art gallery and archives service managing Tyneside and the Wear. Tweets by [@TWAMmuseums](#)  
 Tyne and Wear  
**Essex Record Office**  
 ARCHIVE EXPLORER

**Essex Record Office**  
 @essexarchive  
 The storehouse of Essex history.  
 Chelmsford, Essex  
[essexrecordoffice](#)

**SOUTHAMPTON CITY COUNCIL**

**Southampton Council**  
 News and updates from Southampton City Council. Need to do something? Most tasks can be completed on our website: [southampton.gov.uk](http://southampton.gov.uk)  
 Southampton, UK  
[southampton.gov.uk](http://southampton.gov.uk)  
 Joined February 2009

**Dorset History Centre**  
 Dorset History Centre collects & preserves archives & local studies for Bournemouth, Dorset & Poole for all to enjoy. Use policy [d4u.org.uk/jjkVK](http://d4u.org.uk/jjkVK)  
 Dorchester, England  
[dorsetforyou.gov.uk/libraries-hist...](http://dorsetforyou.gov.uk/libraries-hist...)  
 Joined July 2015

**Kent Archives**  
 Get involved and keep up-to-date with Kent's Archive and Local History service. This account is monitored from 8:30am - 5pm Monday to Friday.  
 Maidstone, South East  
[kent.gov.uk/archives](http://kent.gov.uk/archives)  
 Joined March 2017

**Hull History Centre**  
 News, tips and events from Hull's home page featuring the Hull archives, Hull City Local Studies Library and [hullhistorynews.org.uk](http://hullhistorynews.org.uk)  
 Hull, UK  
 Hull Local Studies Library  
[hullhistorycentre.org.uk](http://hullhistorycentre.org.uk)  
 What's On 2009

**Norfolk Record Office**  
 Norfolk Record Office  
 @NorfolkRO  
 Situated At the Archive Centre, Martineau Lane, with staff also based at Norfolk Heritage Centre (@NorfolkHC) and King's Lynn Borough Archive.  
 Norwich, Norfolk

**PICTURAE**  
 LET'S UNLOCK OUR HERITAGE TO THE WORLD.  
 Gemeente Amsterdam  
 Stadsarchief

**Alle Amsterdamse Akten**  
 DE BESTE VERHALEN UIT HET ARCHIEF VAN DE AMSTERDAMSE NOTARISSEN

**Stanford Text Technologies**  
 STANFORD TEXT TECHNOLOGIES IS AN INTERDISCIPLINARY ENTERPRISE COMBINING BOOK HISTORICAL APPROACHES WITH DIGITAL METHODS AND TOOLS TO INVESTIGATE THE LONG HISTORY OF HUMAN COMMUNICATION FROM THE EARLIEST TIMES TO THE PRESENT DAY. [FIND OUT MORE](#)

**Westminster Abbey**

**Essex Record Office**

**Suffolk Archives**  
 @KeyToThePast  
 Suffolk County Council's Record Office. Follow for archival news, events, updates on collections and more! Discover the treasures of 900 years of Suffolk's past  
 Suffolk, UK  
[suffolkarchives.co.uk](http://suffolkarchives.co.uk)  
 Joined February 2015

**TYNE & WEAR ARCHIVES & MUSEUMS**

**TYNE & WEAR ARCHIVES & MUSEUMS**

**TYNE & WEAR ARCHIVES & MUSEUMS**

# Signs of Literacy Kaggle Research Competition, Nov 2018 – Jan 2019

The collage includes:

- Kaggle logo and a large 'K' icon.
- A section titled "kaggle is a competition platform for interesting data science" featuring icons for a question mark, a document, a bar chart, a person, and a neural network.
- Logos for "DataCamp", "Google Cloud", and "NCAA".
- Icons representing data science concepts like "Data Cleaning", "Feature Engineering", "Model Selection", and "Cross Validation".
- A section titled "SOLVING THE TITANIC KAGGLE COMPETITION IN AZURE ML" featuring a ship icon.
- Icons of robots and a person at a desk.
- A screenshot of a Microsoft Word document titled "Signs of Literacy Kaggle Competition - Final Report.docx".
- A screenshot of a Microsoft Power BI dashboard titled "Signs of Literacy Kaggle Competition - Final Report".
- A pie chart showing distribution across categories.
- A screenshot of a Kaggle competition page for "Creating a Titanic Model in R (Part 1)".
- A small image of three people working together.

**Signs of Literacy Kaggle Research Competition, 2018**  
Colin Greenstreet on LinkedIn  
April 30, 2018

Google owned Kaggle has selected us as one of a small number of pro bono competitions they support each year on the merits of our proposal, and the potential impact on the research field and community of the competition.

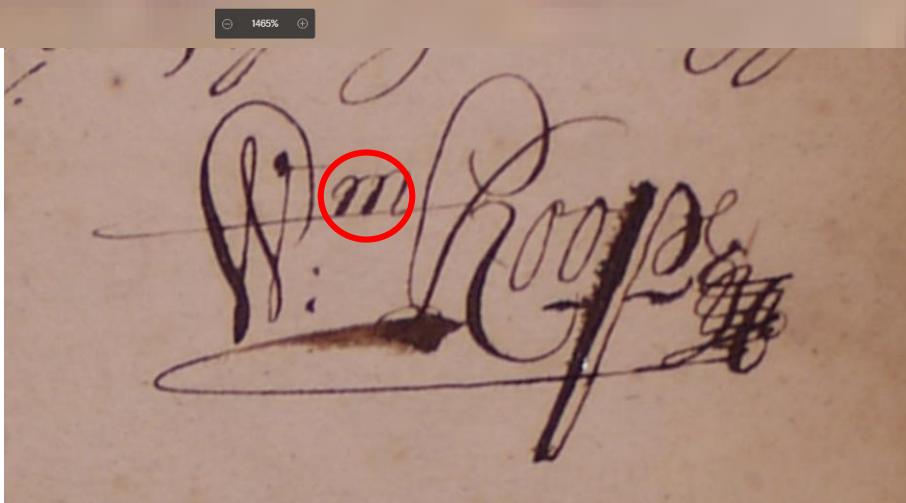
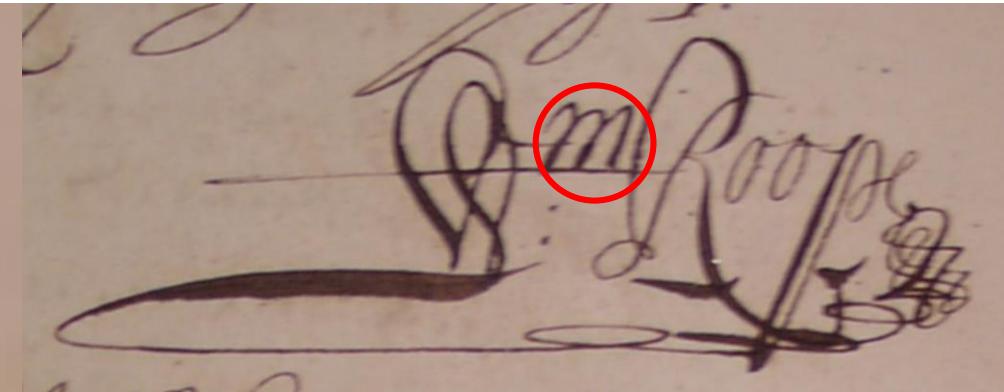
Kaggle will cover the running costs of the competition. We will provide the prize pool, and are now seeking to raise US \$30,000 from potential sponsors and partners.

The Proof of Concept will contain two parts:

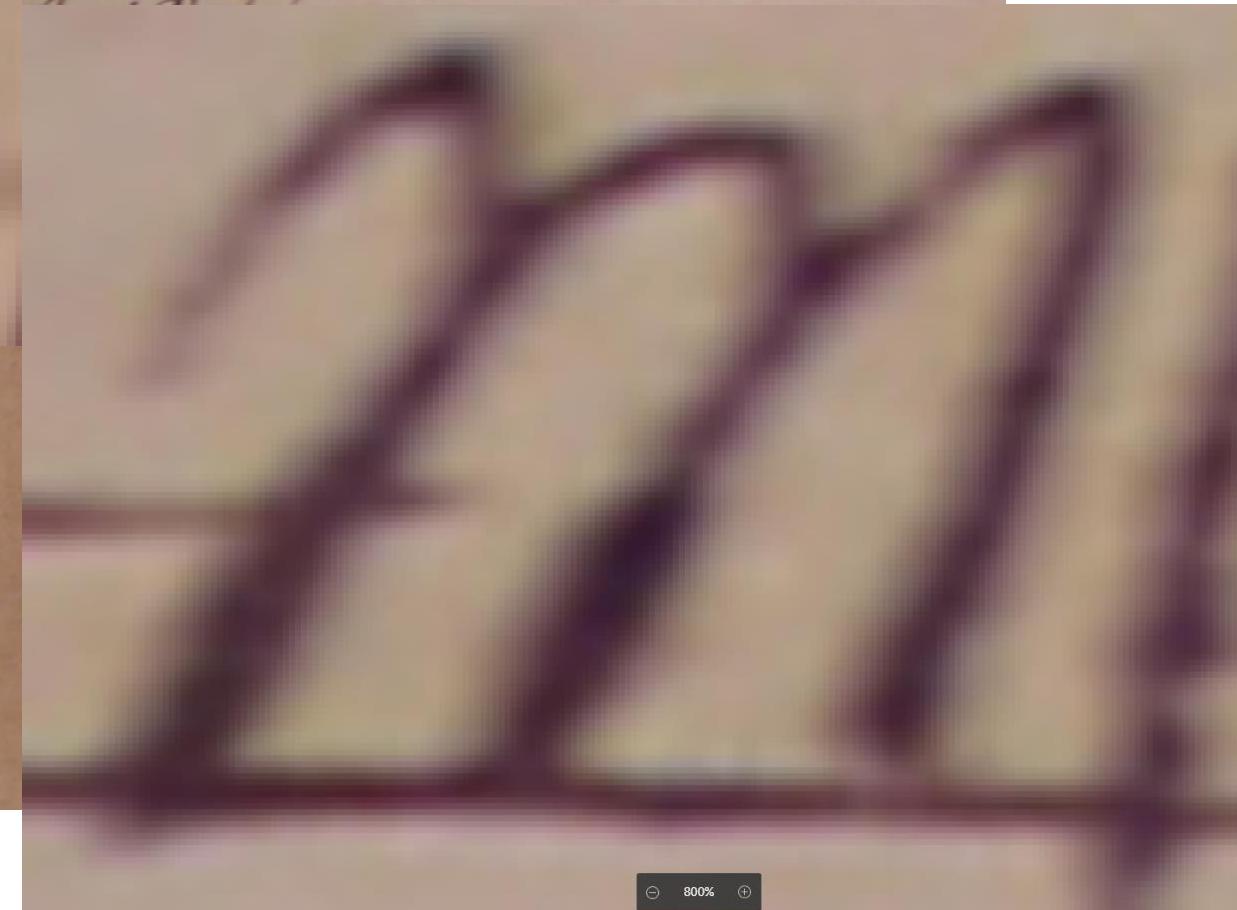
- (1) Algorithmic identification of marks, initials and signatures.
- (2) Algorithmic discrimination between degrees of "sophistication" within the three categories of "mark"; "initial(s)", and "signature".

Having proven the concept, we will seek out an image or vision oriented computational laboratory with which to develop a grant funded collaboration to take the work further in 2019 and beyond.

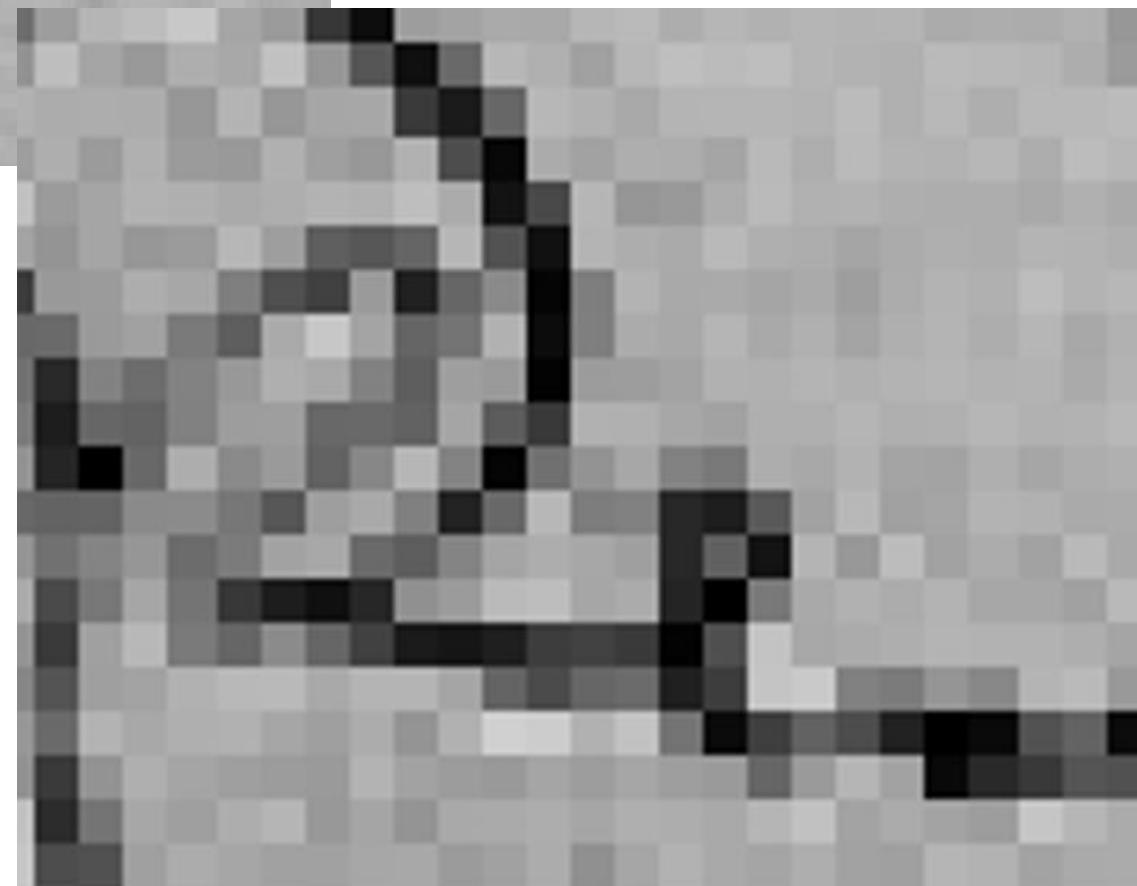
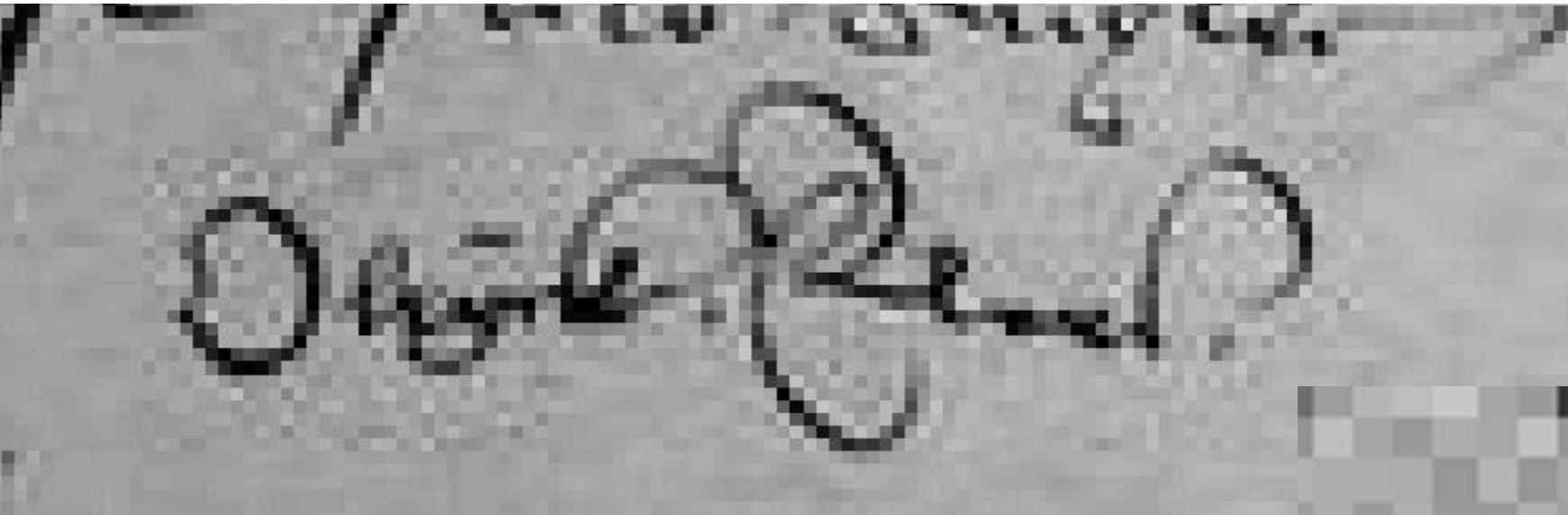
# High pixel definition



Source: KaggleTestSnippet\_HCA\_1373\_f.199r.PNG,  
KaggleTestSnippet\_HCA\_1373\_f.199v\_One.PNG



## Low pixel definition



# Colour analysis – image colour extract PHP, hexadecimal colours

The figure consists of four panels, each showing a snippet of handwritten text and its corresponding color analysis results.

- Panel 1:** Shows the text "Nicholas Harrison". The color analysis table is as follows:

Color	Color Code	Percentage
#e0a080	0.855975	
#c08060	0.084403	
#806040	0.039371	
#604020	0.013208	
#402020	0.007044	

- Panel 2:** Shows the text "Bo:nglis". The color analysis table is as follows:

Color	Color Code	Percentage
#c08060	0.883721	
#806040	0.063798	
#604020	0.048605	
#402000	0.003876	

- Panel 3:** Shows the text "Jacob pintorB". The color analysis table is as follows:

Color	Color Code	Percentage
#808080	0.969271	
#604040	0.030729	

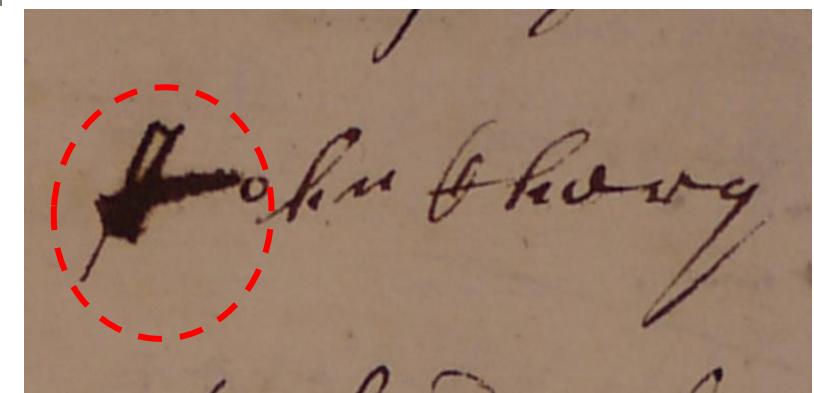
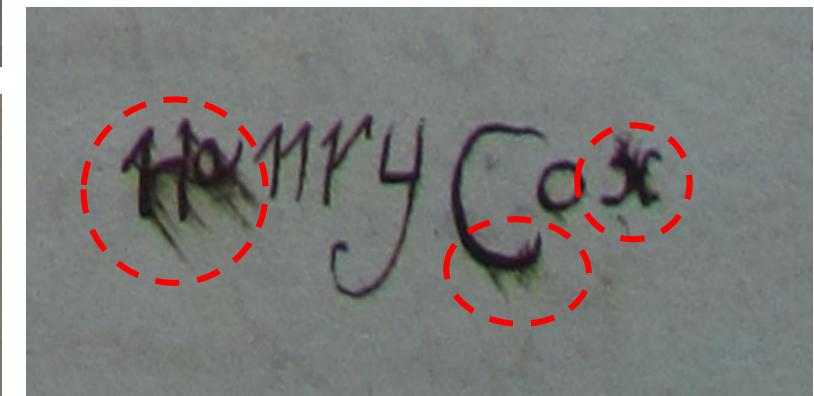
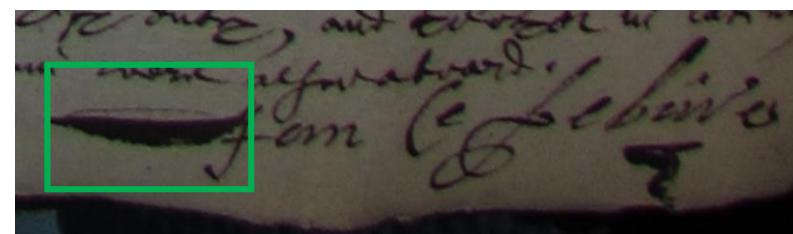
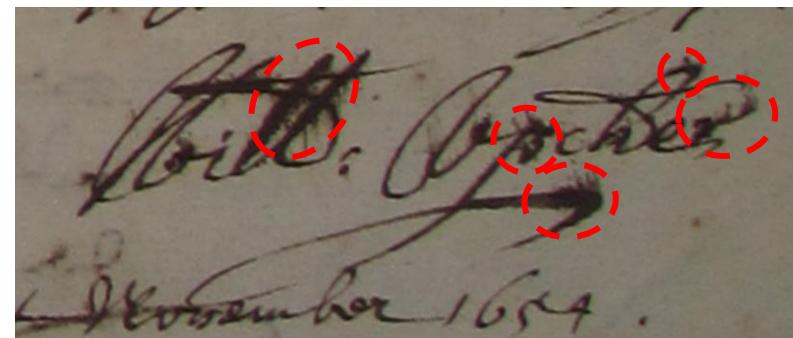
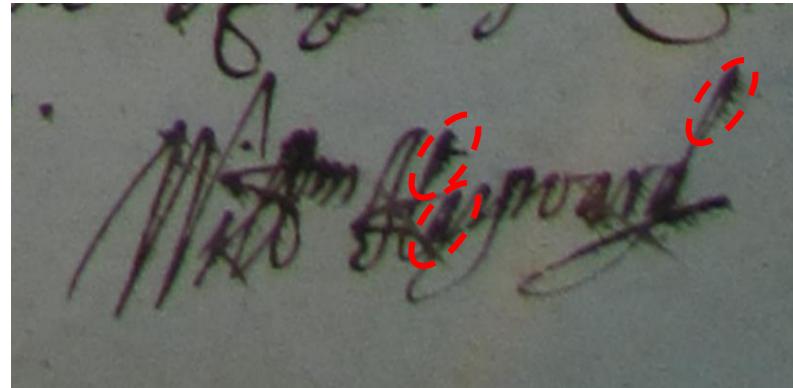
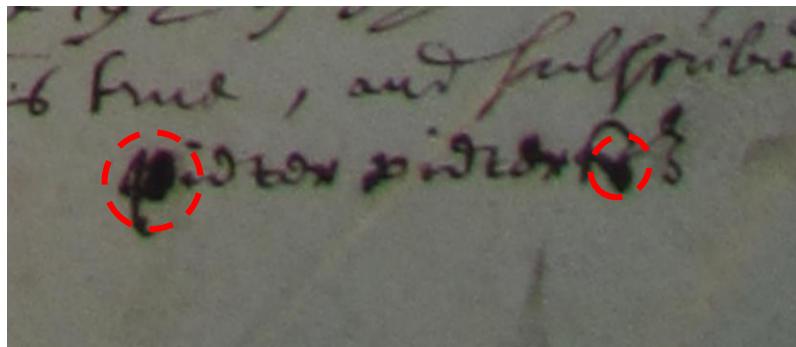
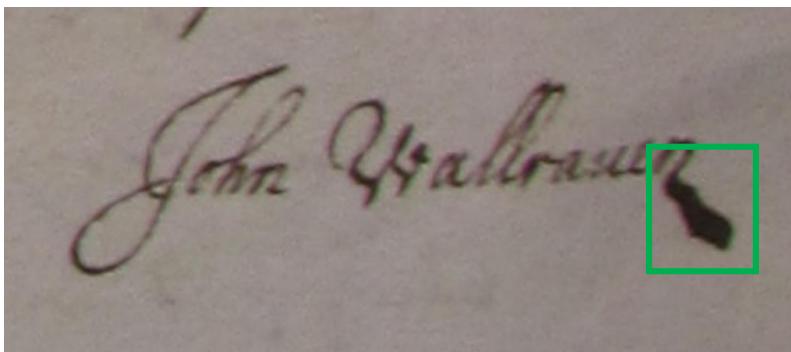
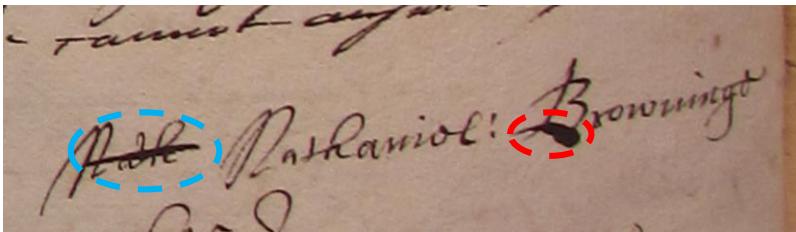
- Panel 4:** Shows the text "H. Langius". The color analysis table is as follows:

Color	Color Code	Percentage
#806040	0.806762	
#604020	0.125143	
#402020	0.062000	
#202000	0.006095	

Below each panel is a grayscale version of the same image, representing the processed input for the color extraction tool.

Source: Sample images from SOLM-2018 (KaggleTestSnippet\_HCA\_1353\_f.275v.PNG, KaggleTestSnippet\_HCA\_1353\_f.270v\_Two.PNG, KaggleTestSnippet\_HCA\_1370\_f.463r\_One.PNG, KaggleTestSnippet\_HCA\_1368\_f.497v.PNG) processed in [http://www.coolphptools.com/color\\_extract#demo](http://www.coolphptools.com/color_extract#demo); same images reprocessed in Photos SW package, with adjustments set to 0% light, 0% colour, 100% clarity

# Detection and analysis of blots, smudges, stylistic features, & deletions



Ink blots or smudges



Stylistic feature or smudge?

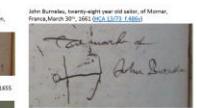
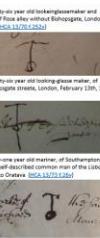


Deletion

Source: Clockwise from top LH side:  
KaggleTestSnippet\_HCA\_1370\_f.387v.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.13r.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.167r.PNG,  
KaggleTestSnippet\_HCA\_1371\_f.456r.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.15r.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.19r.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.41v.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.17v.PNG

# SOLM-2018 IIIF anchors manifest in Mirador viewer

## Anchors



44

HCA Depositions: Anchors

The Mirador viewer interface displays a large central image of a historical manuscript page featuring several anchor markings. Below this main image, five smaller thumbnail images are shown, each corresponding to one of the anchor drawings. The thumbnails are labeled with the names of the individuals whose anchors they represent: Richard Shepperd, Andrew Beake (2), Andrew Beake (1), John Tylor, and John Burnelau. The thumbnail for Andrew Beake (1) is highlighted with a blue border, indicating it is the active or selected item. The interface includes standard Mirador navigation controls such as arrows for page navigation, zoom controls, and a search bar.

Richard Shepperd

Andrew Beake (2)

Andrew Beake (1)

John Tylor

John Burnelau

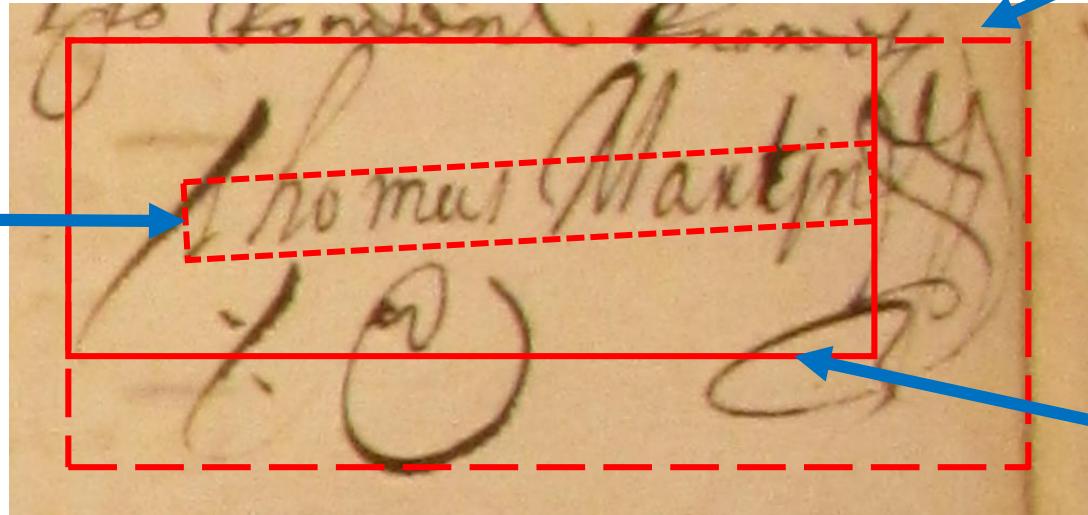
Mockup of a IIIF manifest in Mirador viewer, using <http://projectmirador.org/demo/> ;  
[http://www.marinelives.org/wiki/HCA\\_13/70\\_f.252v\\_Annotate](http://www.marinelives.org/wiki/HCA_13/70_f.252v_Annotate)

# Boundary boxes marking the visual geometry of a signature

Inside boundary box,  
excluding uppers and  
downers

Outside boundary  
box, including  
flourish

Middle boundary  
box, including all  
letters, but excluding  
flourish



## Statistics

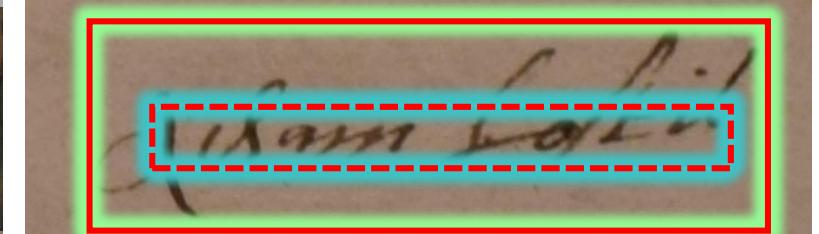
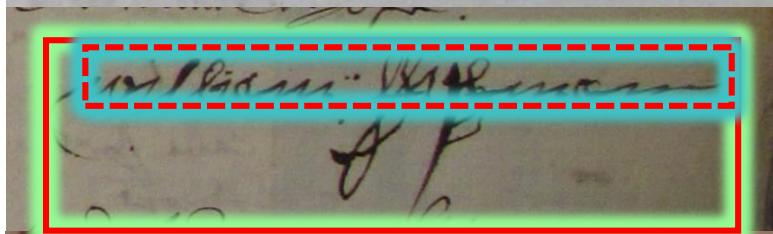
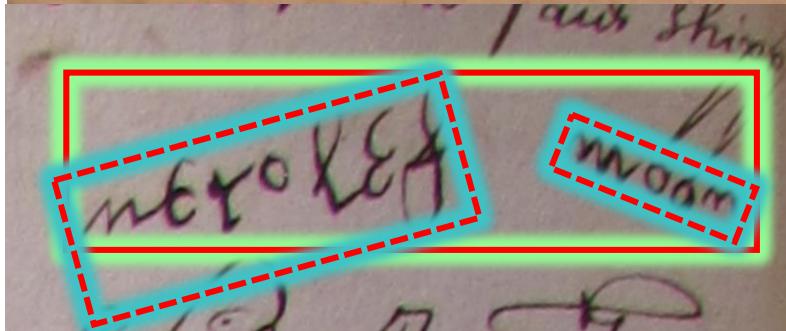
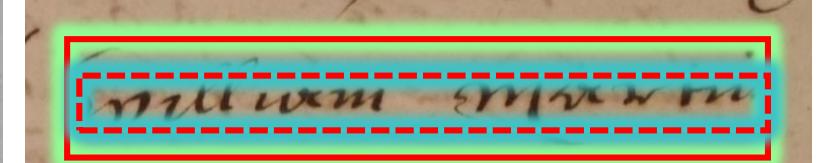
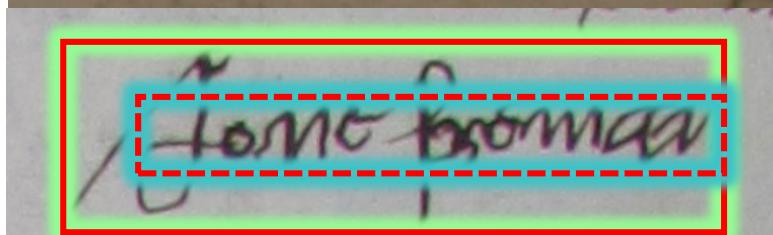
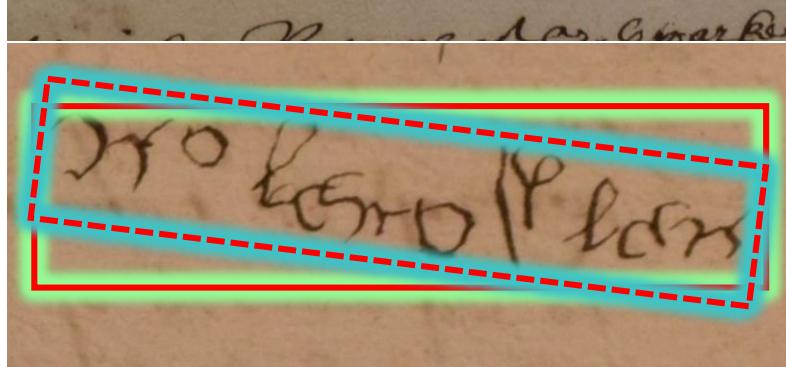
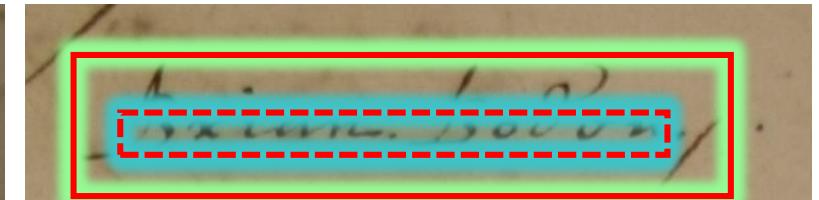
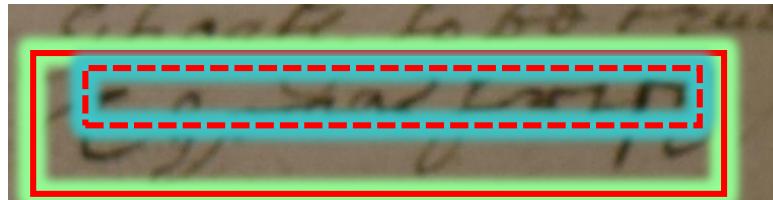
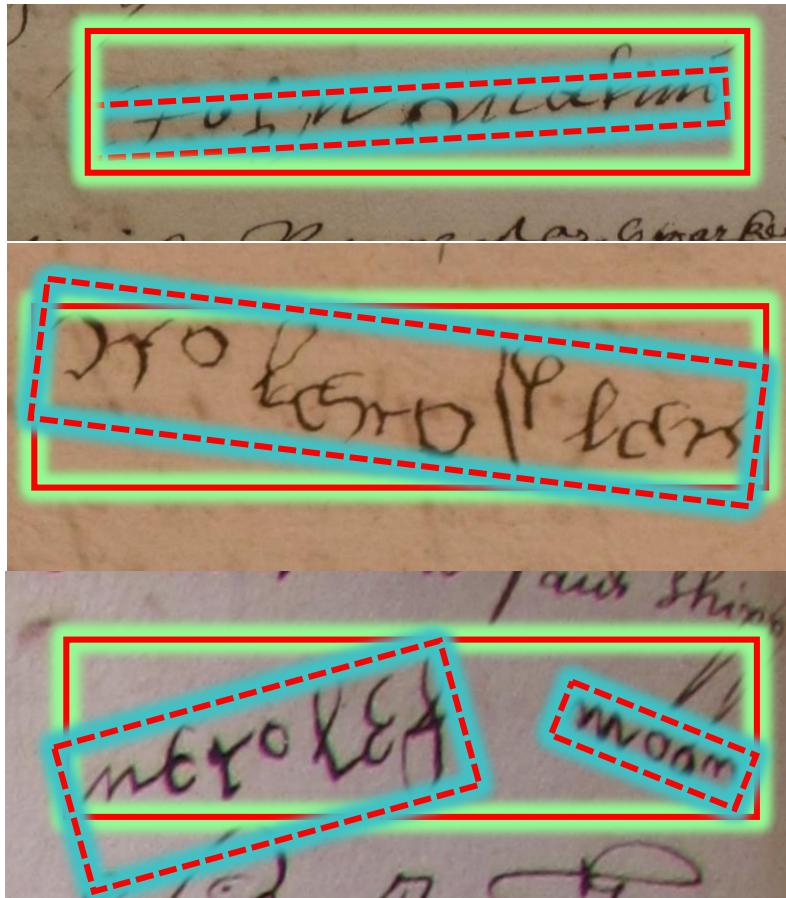
Inside boundary box: 9.0 x 1.1

Middle boundary box: 9.75 x 4.25

Outside boundary box: 12.75 x 5.75

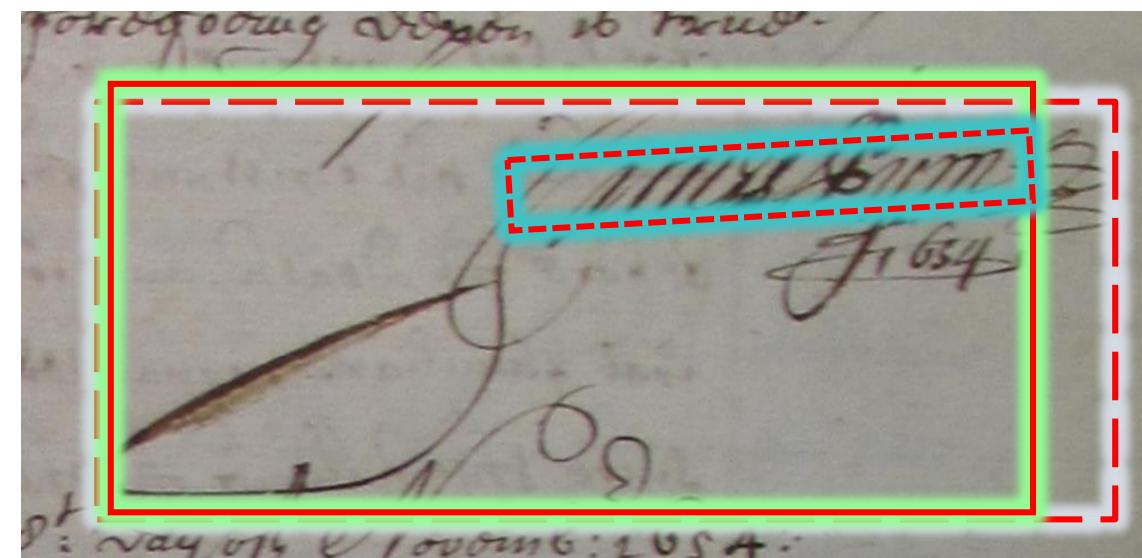
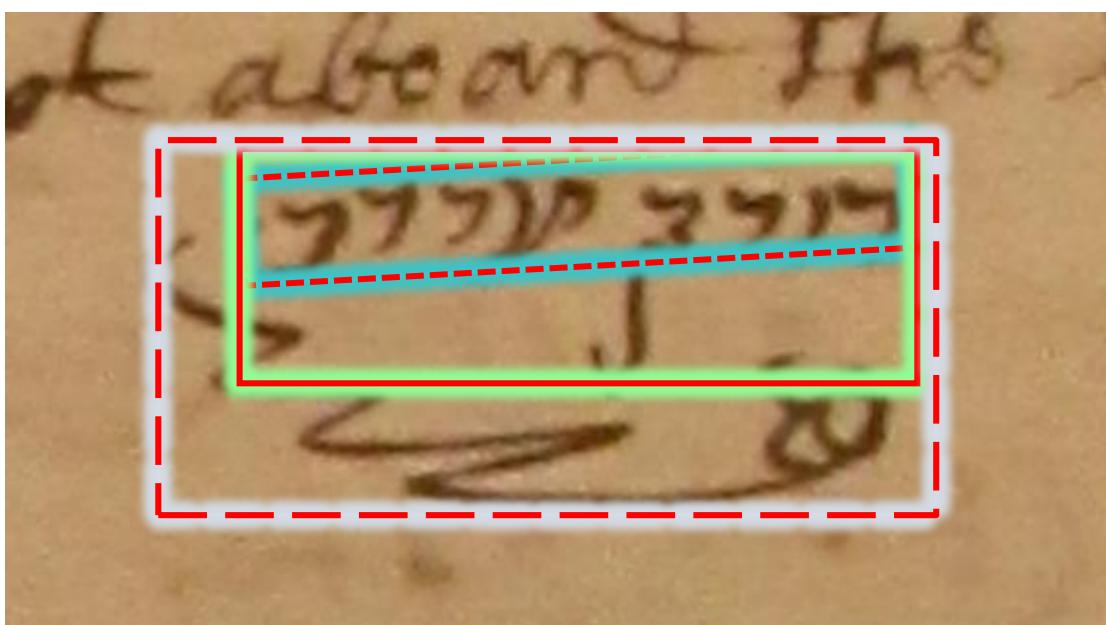
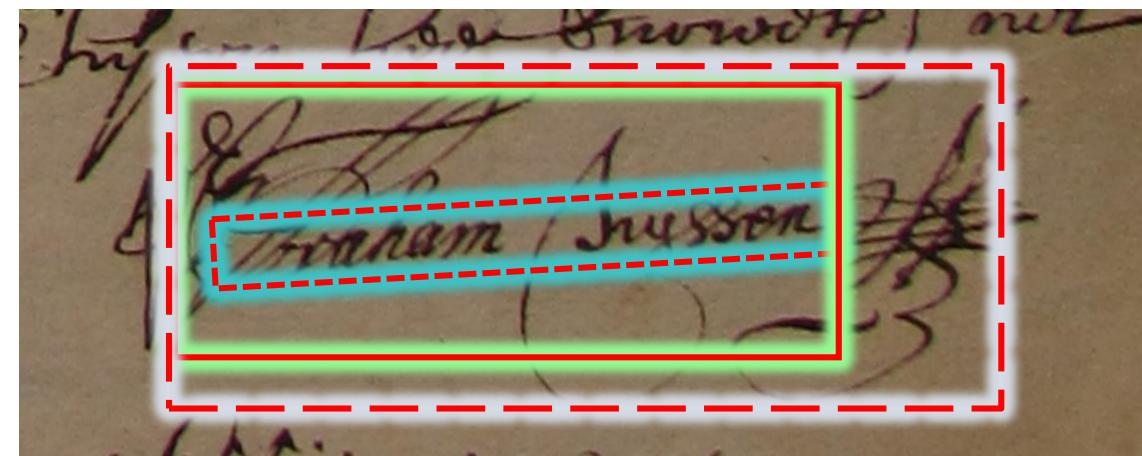
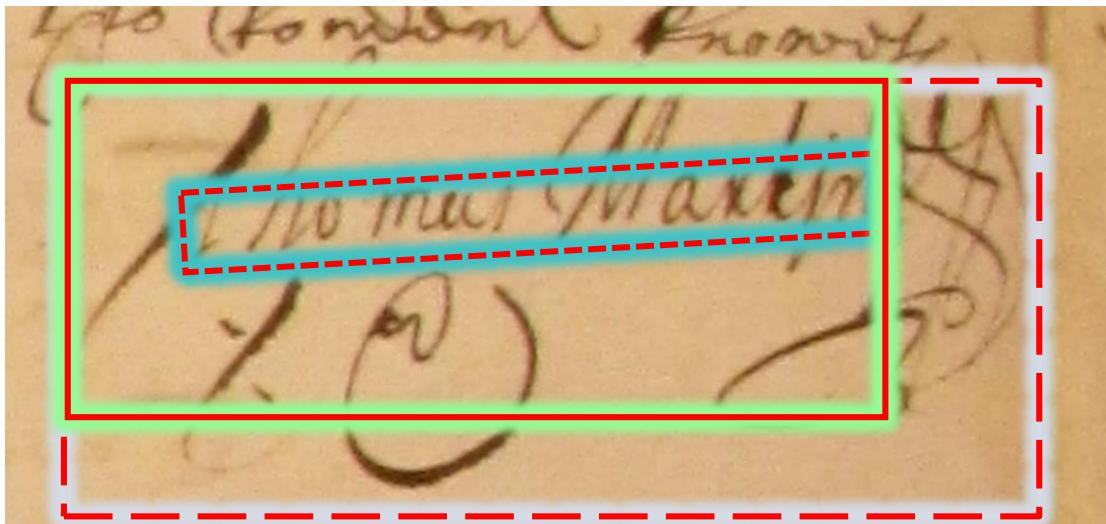
Rotation from horizontal: ca. 340 degrees

# Simple signatures, no flourishes



Source: Down from top LH side: KaggleTestSnippet\_HCA\_1353\_f.24v.PNG, KaggleTestSnippet\_HCA\_1353\_f.188r.PNG;  
Down from top Middle: KaggleTestSnippet\_HCA\_1353\_f.66r.PNG; KaggleTestSnippet\_HCA\_1370\_f.193r\_One.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.203r.PNG, KaggleTestSnippet\_HCA\_1370\_f.218r.PNG  
Down from top RH SIDE: KaggleTestSnippet\_HCA\_1353\_f.28v.PNG, KaggleTestSnippet\_HCA\_1353\_f.29v\_One.PNG,  
KaggleTestSnippet\_HCA\_1353\_f.35r.PNG, KaggleTestSnippet\_HCA\_1353\_f.36v.PNG

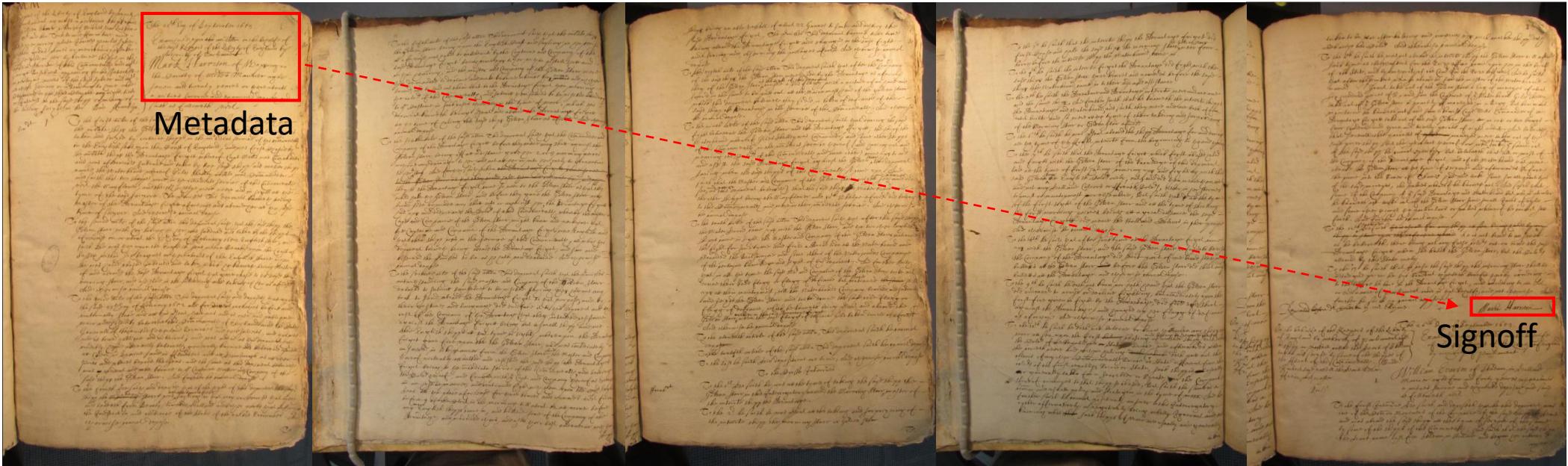
# Visual geometries of flourishes – C17th Irish, Dutch, English & Moroccan merchants



Source: Clockwise from top LH side: KaggleTestSnippet\_HCA\_1368\_f.34v.PNG,  
KaggleTestSnippet\_HCA\_1370\_f.366r.PNG, KaggleTestSnippet\_HCA\_1370\_f.134r.PNG,  
KaggleTestSnippet\_HCA\_1368\_f.58r.PNG

# Legal deposition

Deposition of Mark Harrison; mariner and master; resident in Wapping, Middlesex; age 27;  
Dated September 21<sup>st</sup> 1659 (TNA, HCA 13/68, ff. 1r-3r)

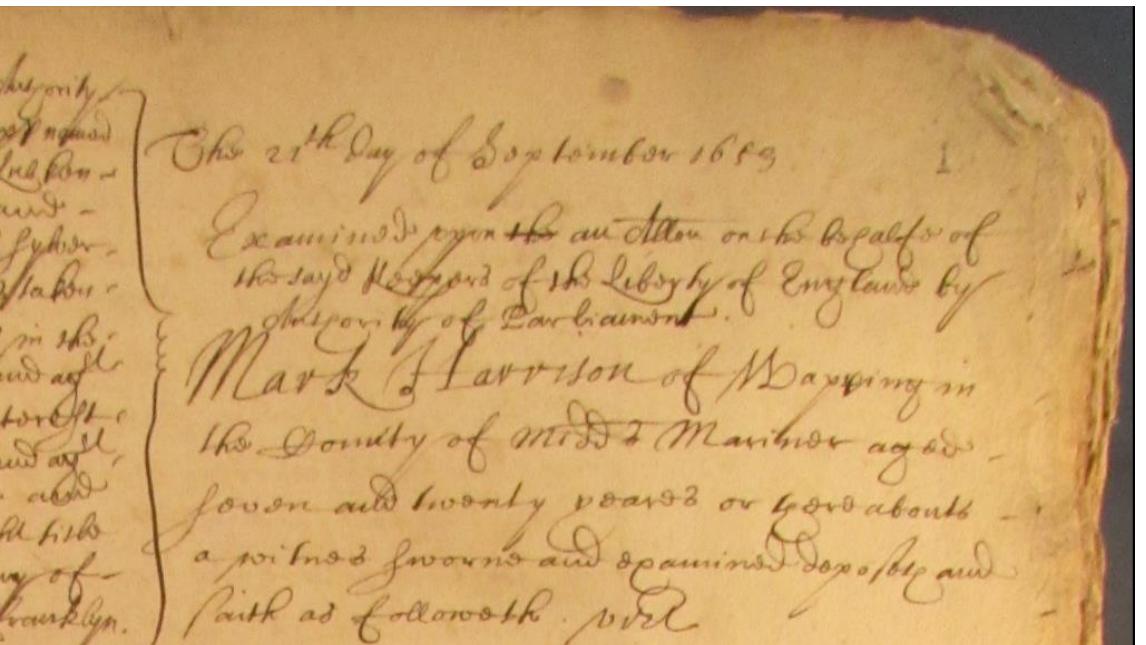


# Machine based recognition of metadata

The 21<sup>st</sup> Day of September 1689

Examined upon the affaile on the behalfs of  
the sayd Negroes of the Liberty of En gland by  
Mark T. Garrison of Newbury in  
the County of Middlesex aged  
seven and twenty years or there abouts  
sworn from and examined before me and  
signed as followeth. ver

# 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<sup>2</sup> and of<sup>4</sup> what happened<sup>7</sup> in<sup>2</sup> the county of Middlesex mariner<sup>8</sup> aged seven and twenty years

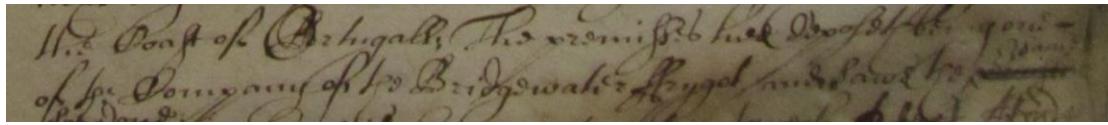
mark Harrison<sup>3</sup> of<sup>2</sup> walking<sup>8</sup> in<sup>2</sup> the county of Middlesex mariner<sup>15</sup> aged seven and twenty years or<sup>2</sup> there<sup>4</sup> about

mark<sup>2</sup> Harrison<sup>3</sup> of<sup>2</sup> what<sup>4</sup> happened in the county of Middlesex mariner<sup>10</sup> aged seven and twenty years or<sup>4</sup> there<sup>4</sup>

mark<sup>2</sup> Harrison<sup>3</sup> or<sup>3</sup> walking in the county of Middlesex mariner<sup>8</sup> aged<sup>2</sup> seven and twenty years<sup>3</sup> or<sup>2</sup> the<sup>3</sup> about<sup>5</sup>

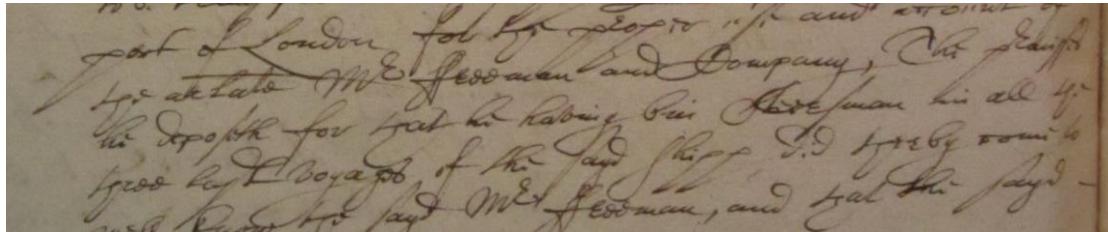
# Can we use key word spotting to excavate raw metadata?

## LANGUAGE DENOTING OCCUPATION



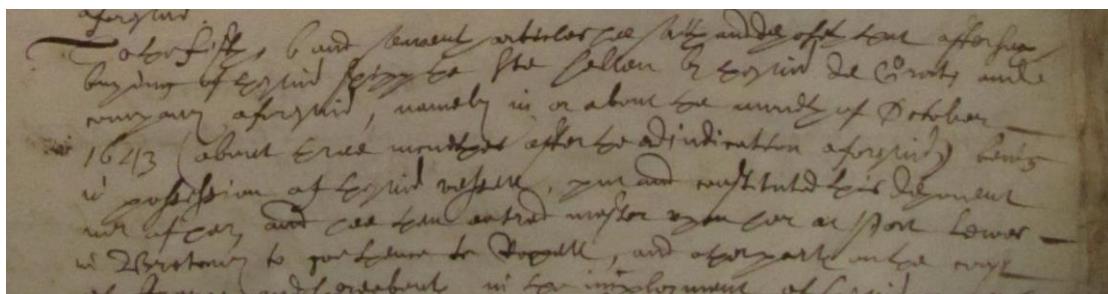
W<sup>t</sup> Roast of Orthigall, The premissee hee deposeth  
of the Companye of the Bridgewater ffrygott, and sawe her  
in the same shipp in the said voyage.

"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]



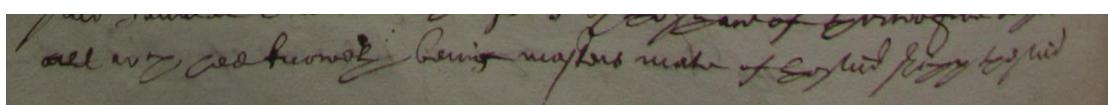
part of London for the said shipp and comon  
she ar late Mr. Godman and Company, the shipp  
the deponeth for that he had his Steersman in all ye  
free last boyage of the said shipp S. I. by his owne  
hande paid Mr. Godman, and that the said

"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]



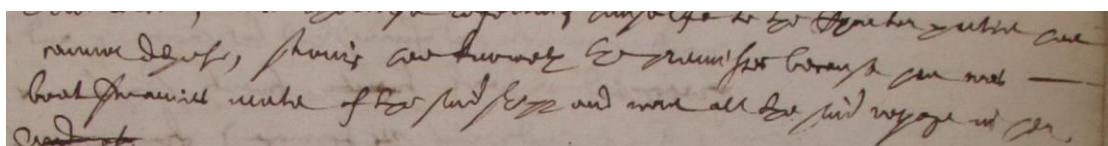
Yours.  
To oblyf, and present witnessse by ande of late after  
buying of said shipp the Santa Hellen or S. H. de Grotte, and  
having aposid, namely in or about the moneth of October  
1643 (about three moneths after the adiudication aforesaid) being  
in possession at said shipp, and wrought bid him self  
out of say and came into entred master of her at Port Lewes  
in Bretany to go thence to Foggall, and afterward in the same  
shipp to Lorient in France.

"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 vessel, 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 my ded knowell being masters mate of sayd shipp by mid

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



name deput, having now knowell to sayd shipp and hee  
had command of her last year and made all sayd voyage wth her.

"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 adjudication aforesaid) being in possession of the said vessel, 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.67r] [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 boatswain
KaggleTestSnippet_HCA_1370_f.671v.PNG	HCA 13/70	Marker	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	Initials	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.657r.PNG - CREATE HCA 13/68	HCA 13/68	Signature	Mariner; Boatswain [of the Civil 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
KaggleTestSnippet_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

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

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

Table 1.2a: HCA 13/53 [f.1r-100v] - Signoff frequency per manuscript page, data from 1637

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10	1	2	3	1	0	1	1	0	2	2	16
11-20	2	2	1	1	1	1	1	1	1	1	13
21-30	2	1	1	3	1	1	1	1	1	1	18
31-40	1	1	1	1	1	1	2	1	1	1	13
41-50	1	1	1	1	1	1	1	1	1	1	6
51-60	1	1	1	1	1	1	2	1	1	1	11
61-70	2	1	1	1	2	1	1	1	2	1	20
71-80	1	2	1	1	1	2	1	1	1	1	19
81-90	2	1	1	1	1	2	1	2	4	1	23
91-100	1	2	1	1	1	2	2	3	3	1	26
Total	1	2	1	1	2	2	3	3	1	3	165

1637

Table 1.3a: HCA 13/58 [f.1r-100v] - Signoff frequency per manuscript page, data from 1642

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10	1	1	1	1	1	1	1	1	1	1	15
11-20	1	1	1	2	1	1	1	1	1	1	16
21-30	1	1	1	1	1	1	1	1	1	1	10
31-40	1	1	1	1	1	1	1	1	1	1	16
41-50	2	1	1	1	1	1	1	1	1	1	12
51-60	1	1	1	2	1	1	1	1	1	1	16
61-70	1	1	2	1	1	1	1	1	1	1	11
71-80	1	2	1	1	1	1	2	1	1	1	14
81-90	1	1	2	1	1	1	2	1	1	1	19
91-100	1	2	1	2	1	2	1	2	1	1	15
Total	1	2	1	2	1	2	1	2	1	1	144

1642

Table 1.4a: HCA 13/70 [f.401r-500v] - Signoff frequency per manuscript page, data from 1655

	1	2	3	4	5	6	7	8	9	10	Total	
	r	p	r	r	r	r	r	r	r	r		
401-410			1	1	1	1	1	1	2	1	1	13
411-420	1	1	1	1	1	1	1	1	1	1	1	12
421-430	1	1	1	1	1	1	2	1	1	1	1	13
431-440	1	1	1	1	1	1	1	1	1	1	1	13
441-450	2	2	2	1	1	1	1	1	1	1	1	13
451-460	1	2	1	2	1	1	1	1	1	1	1	16
461-470		1	2	2			1	2				8
471-480		1			1	1	1		1			5
481-490	1	1	1	1	1	1	1	1	1	1	12	
491-500	1	1	1	1	1	1	1	1	1	1	8	
Total			1	1	1	1	1	1	1	1	113	

1655

Table 1.5a: HCA 13/71 [f.1r-100v] - Signoff frequency per manuscript page, data from 1656

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10			1	1	1	1	1	1	1	1	10
11-20	2	2	2	1	1	1	1	1	2	2	21
21-30	1	1	1	1	1	2	2	1	1	1	16
31-40	1	1	1	1	1	1	1	1	1	1	7
41-50	1	1	1	1	1	1	1	1	1	1	7
51-60	1	1	1	1	1	2	1	1	1	1	11
61-70	1	1	1	1	2	1	1	1	1	1	5
71-80	1	1	1	1	1	1	1	1	1	1	2
81-90	1	1	1	1	1	1	1	1	1	1	3
91-100	1	2	1	1	1	1	1	1	1	1	8
Total			1	2	1	1	1	1	1	1	98

1656

Table 1.1a: HCA 13/53 [f.1r-100v] - Signoff frequency per manuscript page, data from 1637

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10	1	2	3	1	0	1	1	0	2	2	16
11-20	2	2	1	1	1	1	1	1	1	1	13
21-30	2	1	1	3	1	1	1	1	1	1	18
31-40	1	1	1	1	1	1	2	1	1	1	13
41-50	1	1	1	1	1	1	1	1	1	1	6
51-60	1	1	1	1	1	1	2	1	1	1	11
61-70	2	1	1	1	1	1	1	1	1	1	20
71-80	1	2	1	1	1	1	2	1	1	1	19
81-90	2	1	1	1	1	2	1	2	4	1	23
91-100	1	2	1	1	1	2	2	3	3	1	26
Total	1	2	1	1	1	2	2	3	3	1	165

Location of signoffs

1637

Table 1.1b: HCA 13/53 [f.1r-100v] - Signoff frequency per manuscript page & location of signatures, marks & initials, data from 1637

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10	1	2	3	1	0	1	1	0	2	2	16
11-20	2	2	1	1	1	1	1	1	1	1	13
21-30	2	1	1	3	1	1	1	1	1	1	18
31-40	1	1	1	1	1	1	2	1	1	1	13
41-50	1	1	1	1	1	1	1	1	1	1	6
51-60	1	1	1	1	1	1	2	1	1	1	11
61-70	2	1	1	1	1	1	2	1	1	1	20
71-80	1	2	1	1	1	2	1	1	1	1	19
81-90	2	1	1	1	1	2	2	4	1	1	23
91-100	1	2	1	1	1	2	2	3	3	1	26
Total	1	2	1	1	1	2	2	3	3	1	165

Location of marks & initials

1637

Table 1.1c: HCA 13/53 [f.1r-100v] - Signoff frequency per manuscript page & location of mariners, data from 1637

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	
1-10	1	2	3	1	0	1	1	0	2	2	16
11-20	2	2	1	1	1	1	1	1	1	1	13
21-30	2	1	1	3	1	1	1	1	1	1	18
31-40	1	1	1	1	1	1	2	1	1	1	13
41-50	1	1	1	1	1	1	1	1	1	1	6
51-60	1	1	1	1	1	1	2	1	1	1	11
61-70	2	1	1	1	1	1	2	1	1	1	20
71-80	1	2	1	1	1	2	1	1	1	1	19
81-90	2	1	1	1	1	2	2	4	1	1	23
91-100	1	2	1	1	1	2	2	3	3	1	26
Total	1	2	1	1	1	2	2	3	3	1	165

Location of mariner signoffs

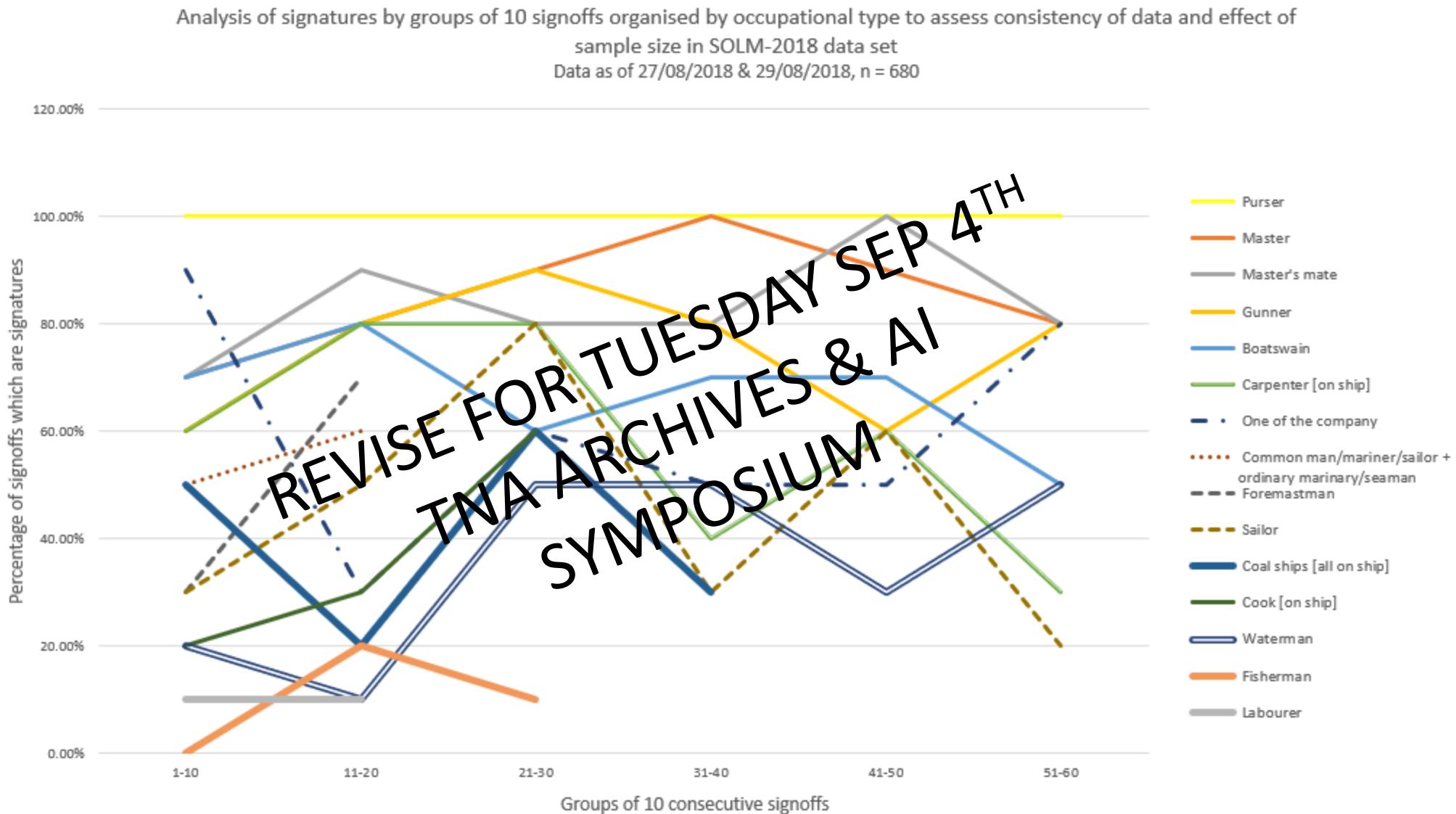
1637

Table 1.1d: HCA 13/53 [f.1r-100v] - Signoff frequency per manuscript page & location of merchants, data from 1637

	1	2	3	4	5	6	7	8	9	10	Total
	r	p	r	r	r	r	r	r	r	r	

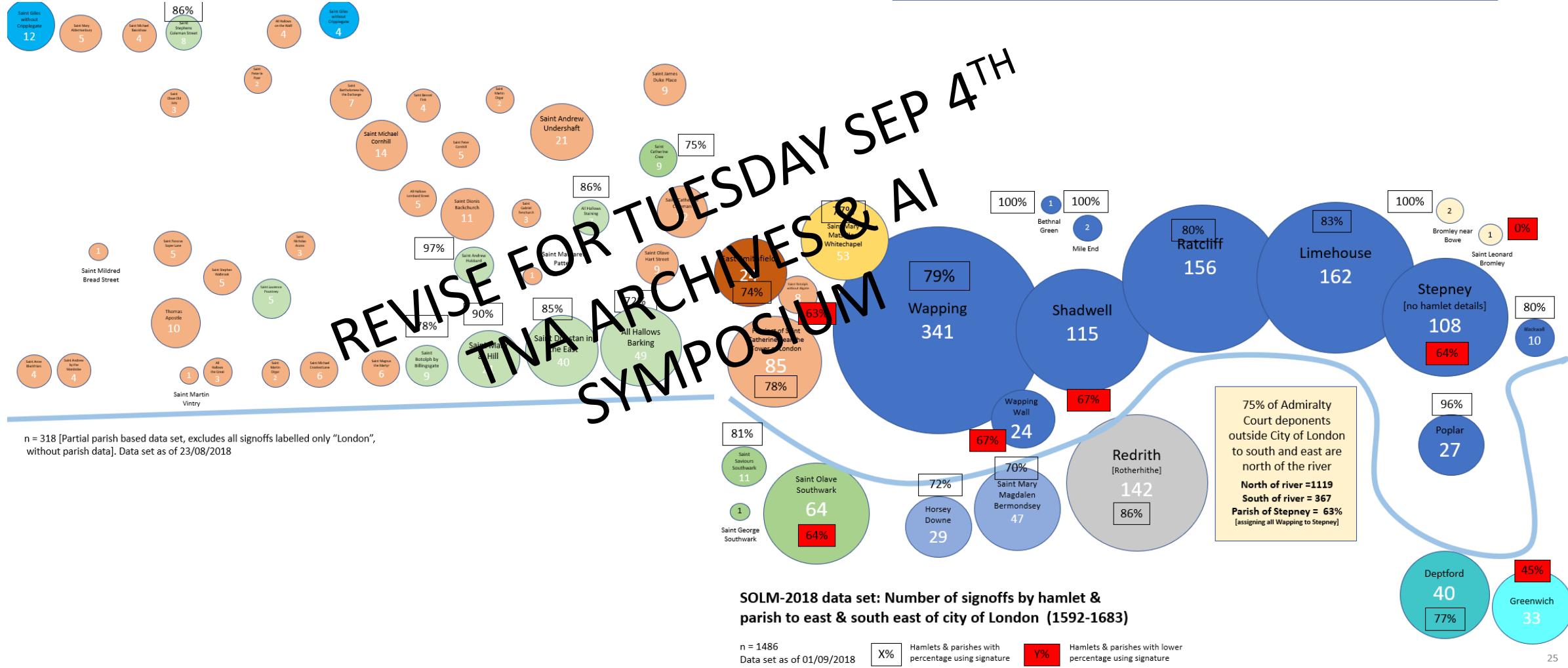

<tbl\_r cells="12" ix="2" max

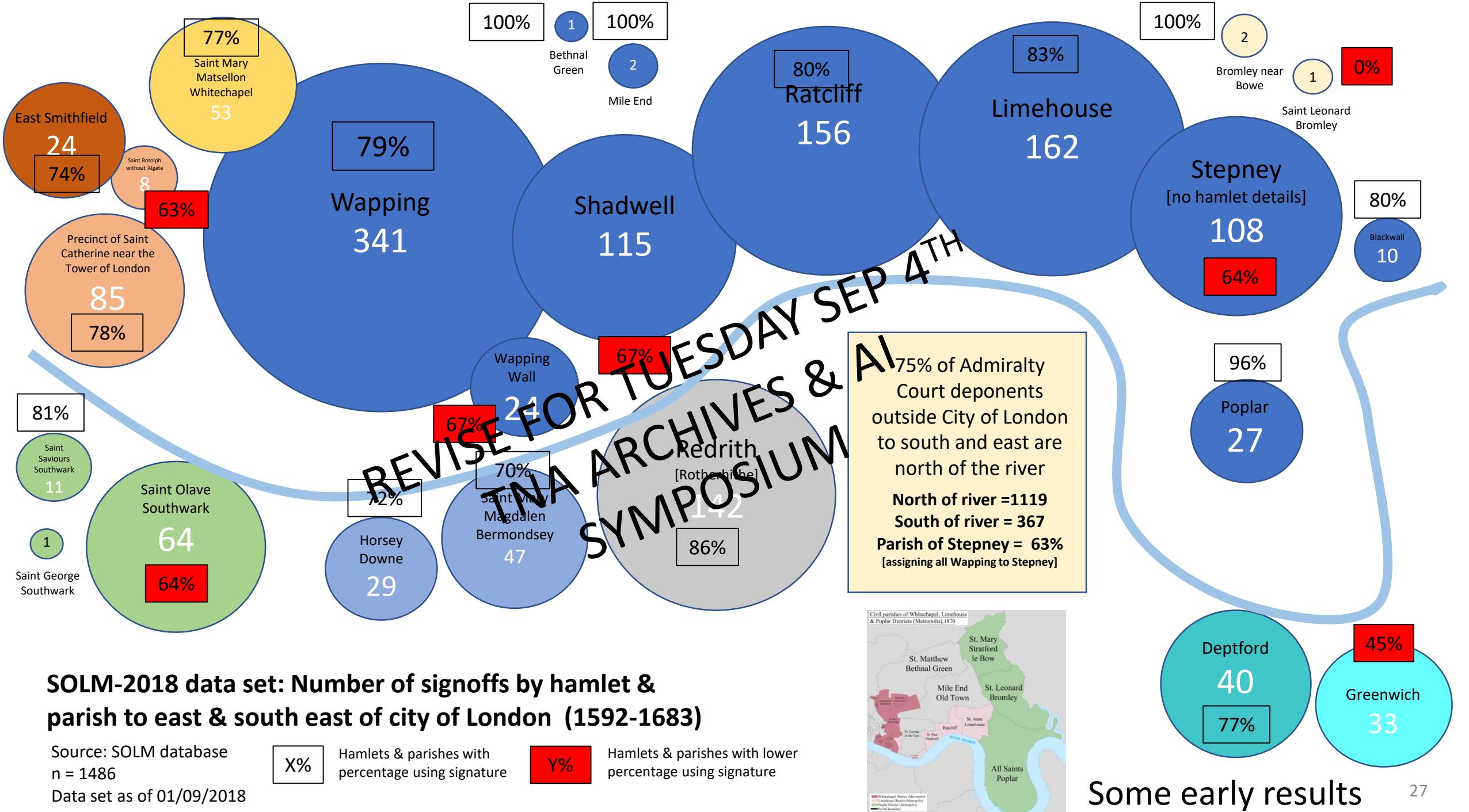
# Some early results



## Some early results

Early/mid-C17th London – a linear maritime city, as seen in the location of High Court of Admiralty deponents, 1637 to 1667





# Contact details

**Colin Greenstreet  
Community organiser,  
Signs of Literacy**

**Dr Mark Hailwood  
Lecturer in History, 1400-1700,  
University of Bristol**

Email:

[colin.greenstreet@gmail.com](mailto:colin.greenstreet@gmail.com)  
[m.hailwood@bristol.ac.uk](mailto:m.hailwood@bristol.ac.uk)

Weblinks:

<http://signsofliteracy.org>  
<http://marinelives.org>  
<http://chronoscopic.org>

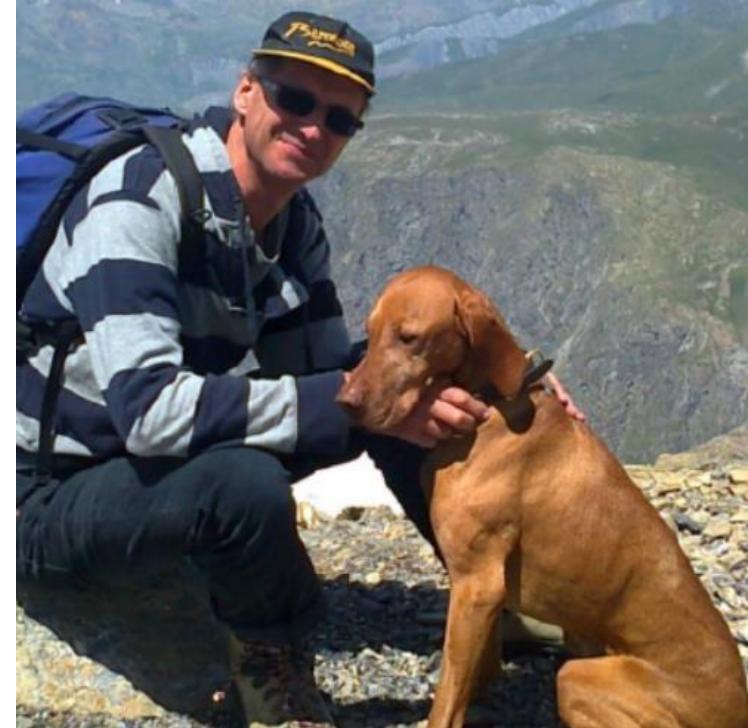
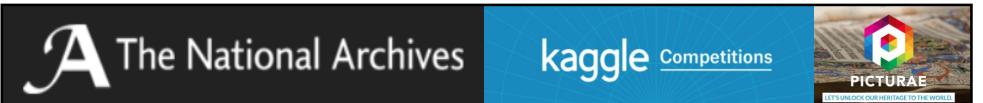
GitHub:

<https://github/Signsofliteracy/Signoff>

Twitter:

[Marinelivesorg](https://twitter.com/Marinelivesorg)

Working  
with:



# **Discussion**

**Archives and AI symposium  
SOLM-2018  
Supplementary material**