

Archivists, computer
scientists and users
working on historical
literacy

SOLM-2018 data set

Archives & AI

Tuesday, September 4th, 2018
The National Archives
United Kingdom

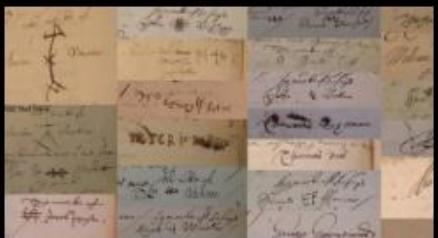
Project portfolio

<http://www.chronoscopic.org>

MarineLives



Signs of Literacy



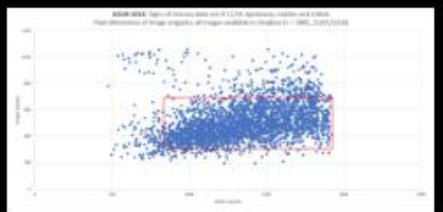
EM Textiles, Garments & Dyestuffs Glossary



Maphackathon



SOLM-2018



EM Maritime & Mercantile Gazetteer



Some perspective

Labeled Faces in the Wild

UNIVERSITY OF MASSACHUSETTS AMHERST, MASS.

Labeled Faces in the Wild Home

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 The aligned images include "funneled images" (ICCV 2007), LFW-a, which uses an unpublished method of alignment, and "deep funneled" images (NIPS 2012).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.

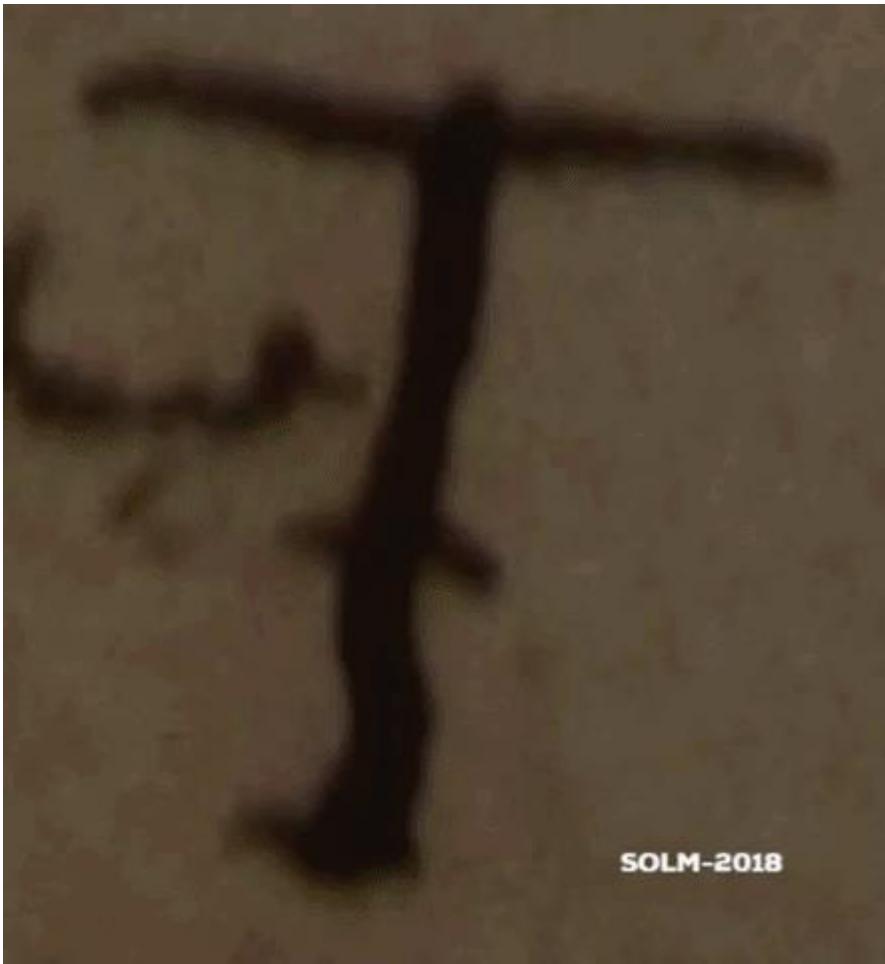
Labeled Faces in the Wild: A Survey

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

In *Advances in Face Detection and Facial Image Analysis*, edited by Michal Kawulok, M. Emre Celebi, and Bogdan Smolka, Springer, pages 189-248, 2016.

Pattern seeking

How many different letters can you recognise?



31 images of initials in the SOLM-2018 database
Animated GIF viewable only in Powerpoint

Initials – Ts and Js, and a few others thrown in



We need to work together, if we are going to make sense of our digitised manuscript archives



'From punishing to pleasurable, how cursive writing is looping back into our hearts', [Washington Post online edition, 03/09/2018](#) - Emilia Schweitzer, left, and Danielle Falls write with homemade quills during a summer day camp devoted to cursive instruction in Danbury, Conn. (Monica Jorge/For The Washington Post)

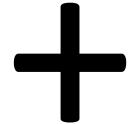


Image processing

Machine learning

Natural Language Programming

Key word spotting

Controlled vocabularies

Programmable decision rules

SOLM-2018 database

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

We will be previewing the database at the TNA Archives & AI symposium on Tuesday, September 4th and at the Sheffield Digital Humanities Congress on Thursday, September 6th, 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 markes, initials and signatures by individuals engaged in marine and shore trade occupations from the following English towns and areas for the C16th, C17th and C18th.

Aldeburgh [Suffolk]
Barnstaple [Devon]
Bermondsey
Bristol
Colchester [Essex]
Dartmouth [Devon]
Deptford

Dover [Kent]
Falmouth [Devon]
Faversham [Kent]
Foy [Cornwall]
Greenwich
Harwich [Essex]
Hull

Ipswich [Suffolk]
Newcastle
Plymouth [Devon]
Rochester [Kent]
Rotherhithe
Southampton
Southwark

Stepney
Weymouth [Dorset]
Woodbridge [Suffolk]
Yarmouth [Norfolk]



SOLM-2018 data set

Analysis of signatures by groups of 10 signoffs organised by occupational type to assess consistency of data and effect of sample size in SOLM-2018 data set

Date as of 22/06/2018 - Signoffs = 1,680



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

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 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!!!!

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 4th and at the Sheffield Digital Humanities Congress on Thursday, September 6th, 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

Stanford | Text Technologies

[HOME](#) [ABOUT](#) [NEWS](#) [PUBLICATIONS](#) [RESEARCH](#) [PEOPLE](#) [COLLEGIUM](#)

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
ARCHIVE EXPLORER

Essex Record Office
@essexarchive

The storehouse of Essex's history
© Chelmsford, Essex
@essexrecordoffice

Southampton City Council

Dorset History Centre
@DorsetArchives Follows you

Dorset History Centre collects & preserves archives & local studies for Bournemouth, Dorset & Poole for all to enjoy. Use policy d4u.org.uk/jkVKV

Kent County Council
kent.gov.uk

Kent Archives
@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.

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

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



TYNE & WEAR archives&museums

TWAM Museums [@TWAMmuseums](#)

Regional museum, art gallery and archives service managing Tyneside and the Wear. Tweets by [@TyneandWear](#)

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.

Norfolk Record Office
@NorfolkRO

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.

Bristol Archives
@bristolarchives Follows you

Home to ten centuries of history: we collect + preserve Bristol's archives (and the British Empire & Commonwealth Collection) for current + future generations.

THE NATIONAL ARCHIVES

THE HUNTINGTON

SOUTH WEST HERITAGE TRUST

Devon Archives
@DevonHeritage

Tweets from the team at the South West Heritage Trust about Devon's fascinating archive collection. #SWHTdiscovery

Kresen Kernow - Cornwall's new archive centre

Last updated: 07/06/2018

Select Language Powered by Google Translate

Print page Email page Subscr...

kresen kernow

heritage lottery fund
LOTTERY FUNDED

Signs of Literacy Kaggle Research Competition

The collage includes:

- Kaggle logo and interface elements.
- A diagram showing a group of people with icons representing data analysis and machine learning.
- A section titled "kaggle is a competition platform for interesting data science" with icons for a question mark, a document, a bar chart, a tree, and a neural network.
- Logos for Google Cloud and the NCAA.
- A diagram illustrating a data pipeline from "Data" through "Feature extraction" and "Model" to "Prediction".
- A section titled "SOLVING THE TITANIC KAGGLE COMPETITION IN AZURE ML" featuring a ship icon and several small robot icons.
- A section titled "kaggle" with a pie chart and a screenshot of a competition page.
- A screenshot of a competition page for "SOLVING THE TITANIC KAGGLE COMPETITION IN AZURE ML".
- A section titled "Creating a Titanic Model in R (Part 1)" with a screenshot of a code editor.

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.

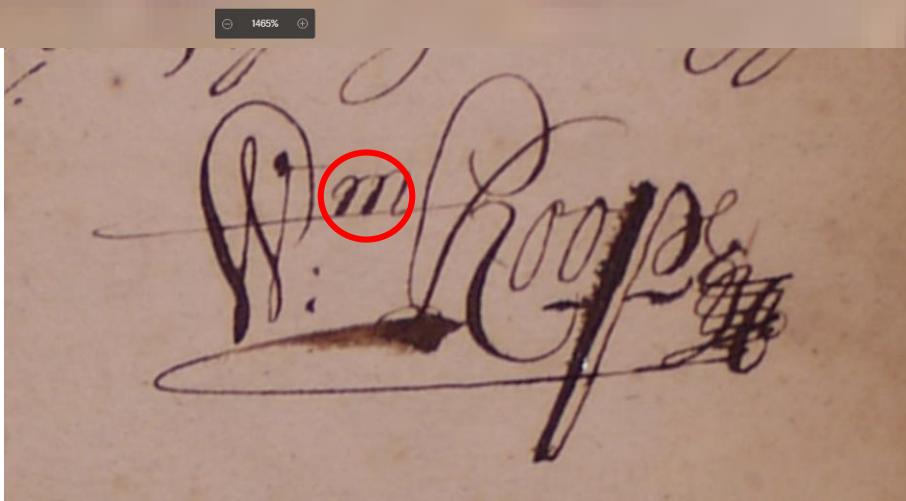
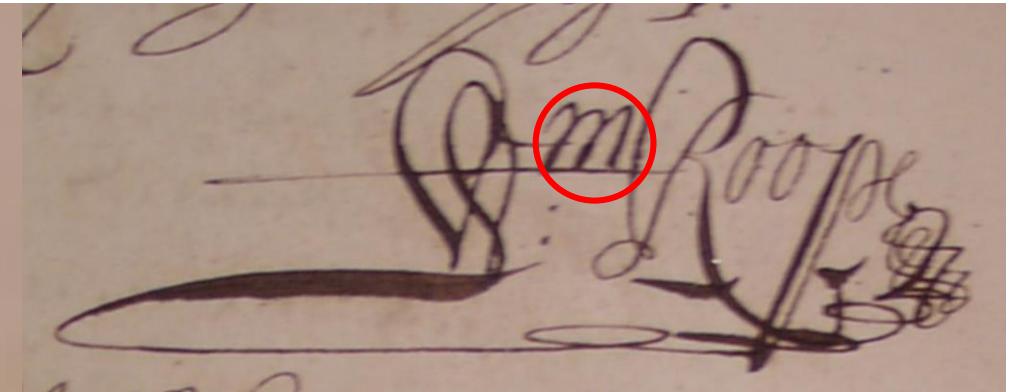
Kaggle will cover the running costs of the competition. We will provide the prize pool, and are now seeking to raise US \$30,000 from potential sponsors and partners.

The Proof of Concept will contain two parts:

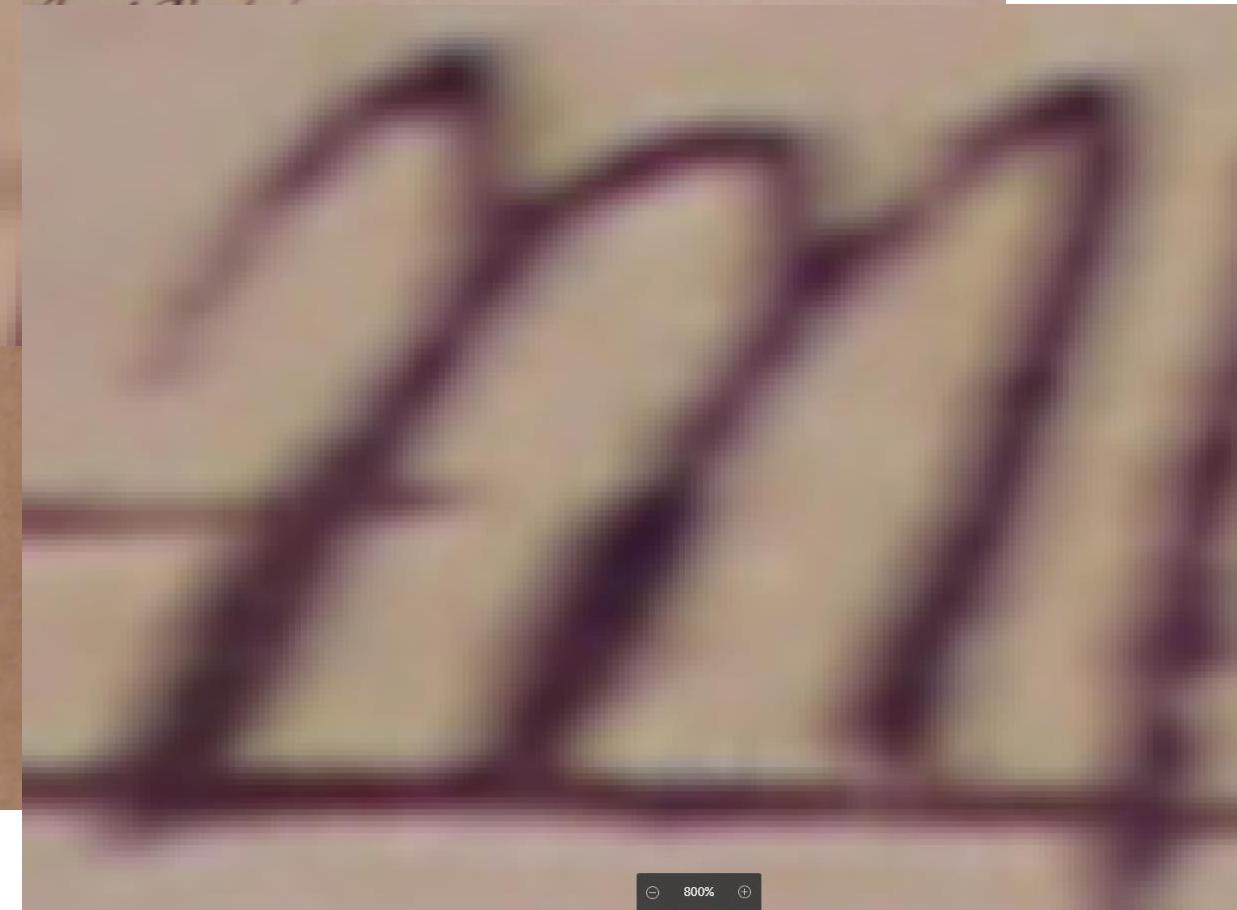
- (1) Algorithmic identification of marks, initials and signatures.
- (2) Algorithmic discrimination between degrees of "sophistication" within the three categories of "mark"; "initial(s)", and "signature".

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.

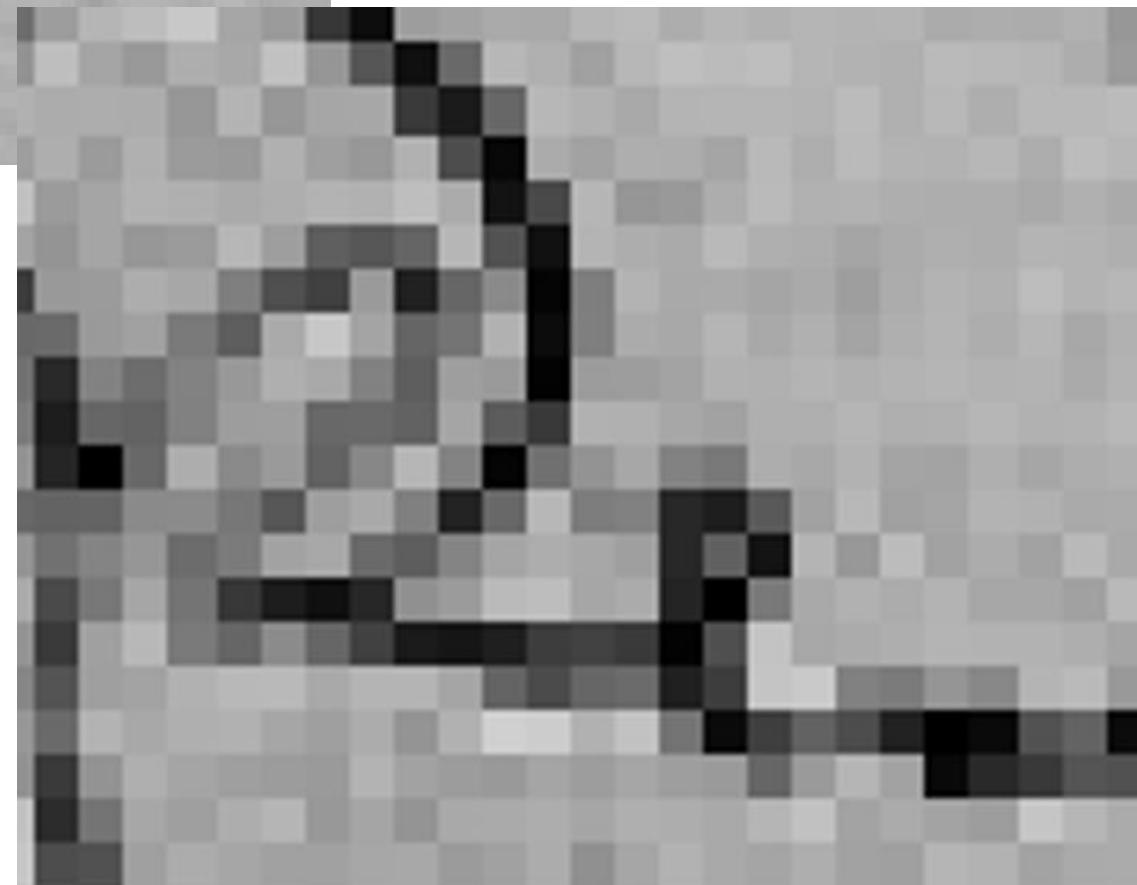
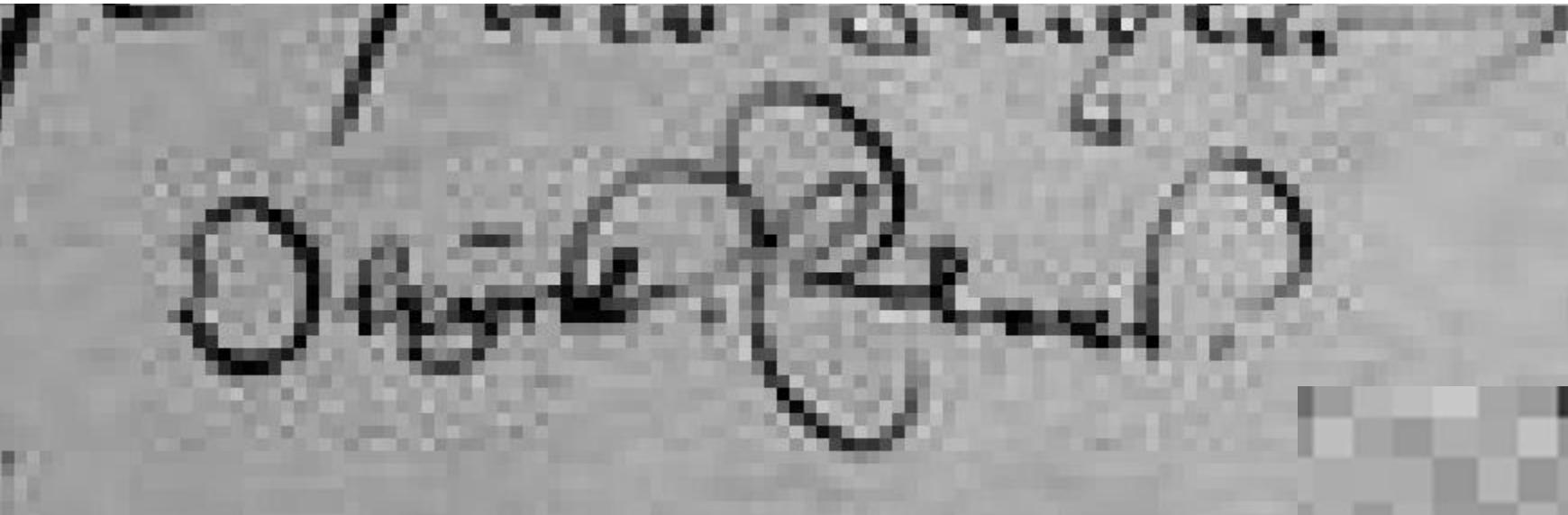
High pixel definition



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



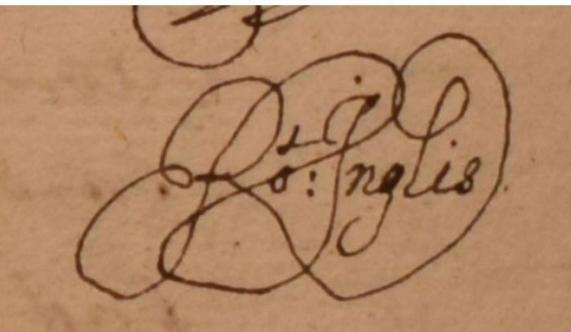
Color	Color Code	Percentage
#e0a080	0.855975	
#c08060	0.084403	
#806040	0.039371	
#604020	0.013208	
#402020	0.007044	

File to upload:
 No file chosen

Number of colors:

 delta:

(1-255)
 Reduce Brightness: Yes No
 Reduce gradient: Yes No



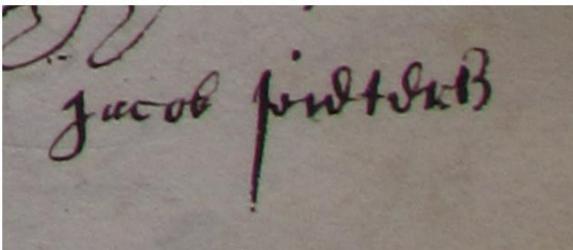
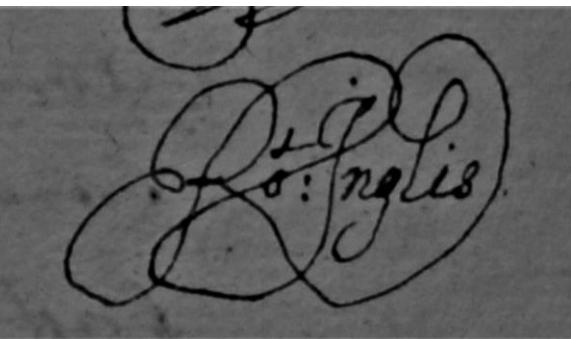
Color	Color Code	Percentage
#c08060	0.883721	
#806040	0.063798	
#604020	0.048605	
#402000	0.003876	

File to upload:
 No file chosen

Number of colors:

 delta:

(1-255)
 Reduce Brightness: Yes No
 Reduce gradient: Yes No



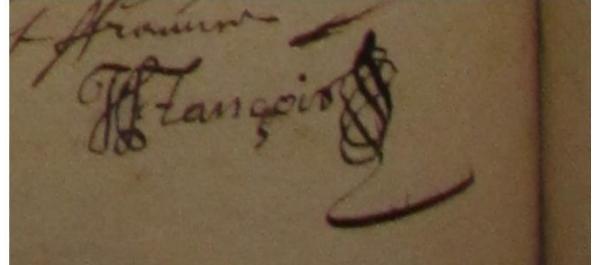
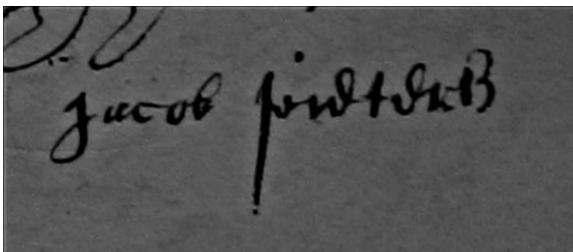
Color	Color Code	Percentage
#808080	0.969271	
#604040	0.030729	

File to upload:
 No file chosen

Number of colors:

 delta:

(1-255)
 Reduce Brightness: Yes No
 Reduce gradient: Yes No



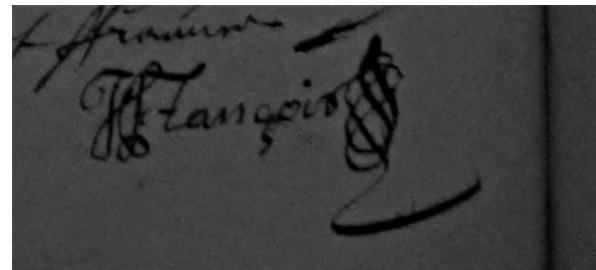
Color	Color Code	Percentage
#806040	0.806762	
#604020	0.125143	
#402020	0.062000	
#202000	0.006095	

File to upload:
 No file chosen

Number of colors:

 delta:

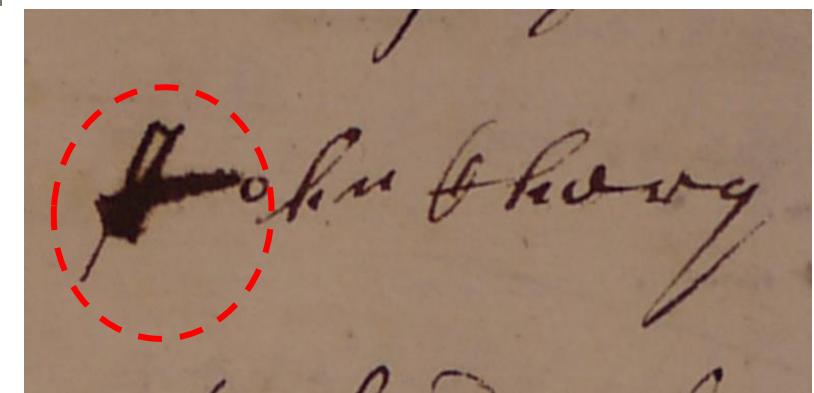
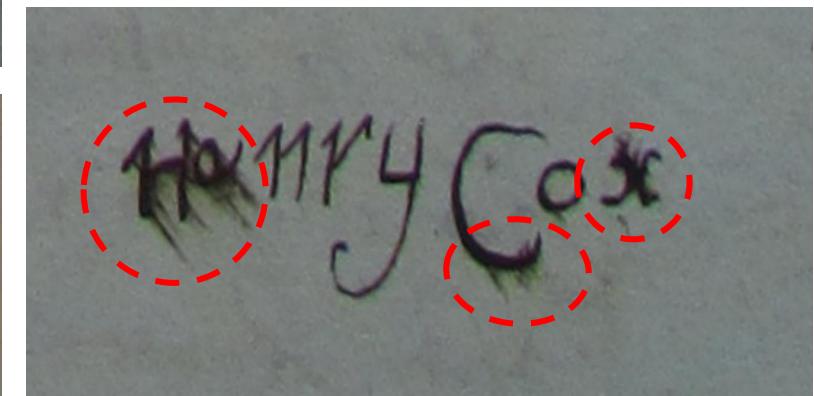
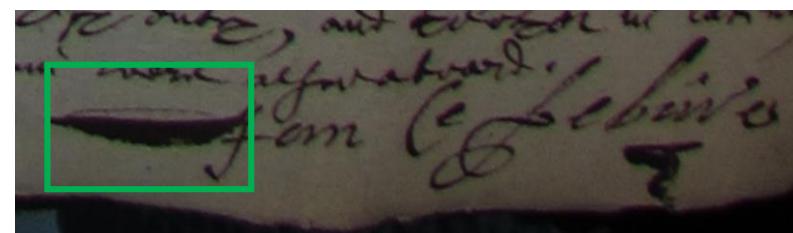
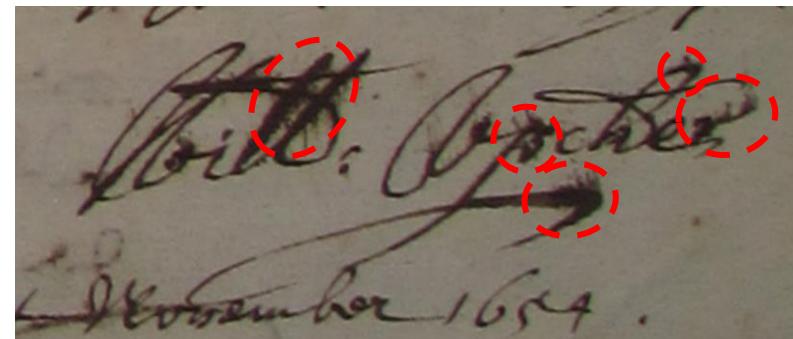
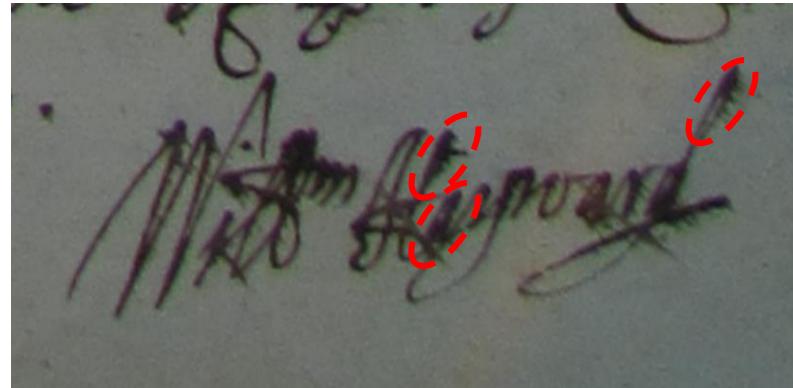
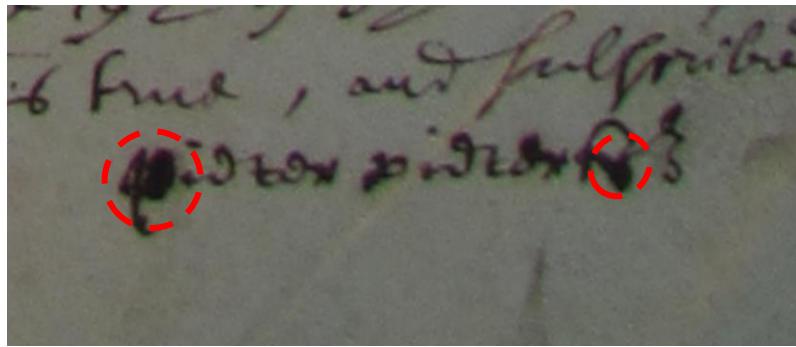
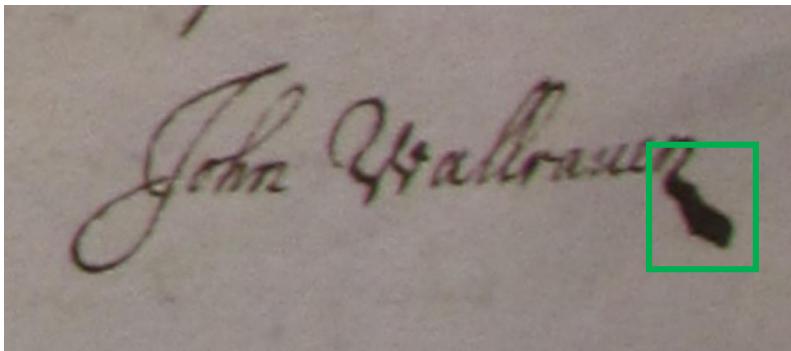
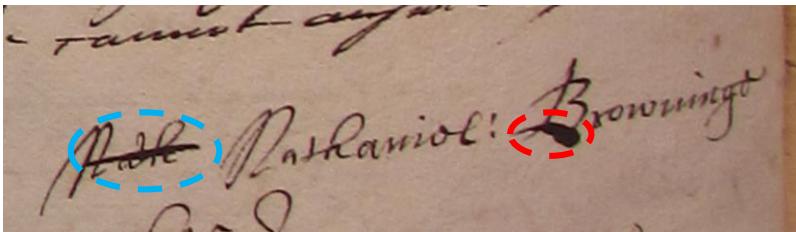
(1-255)
 Reduce Brightness: Yes No
 Reduce gradient: Yes No



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; 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



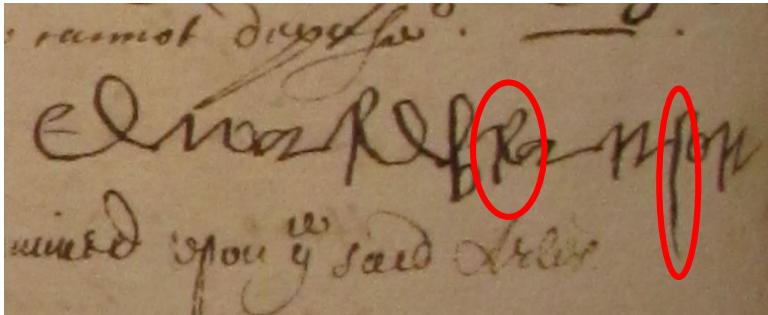
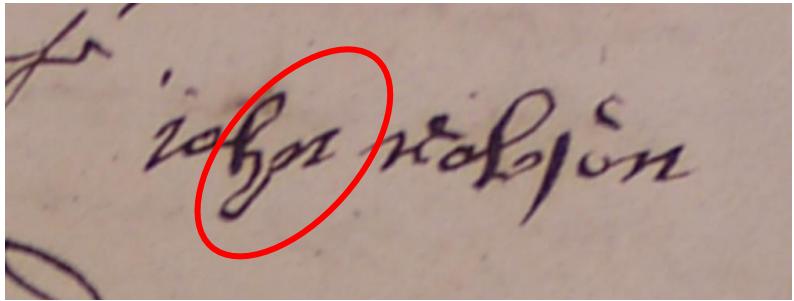
Stylistic feature or smudge?



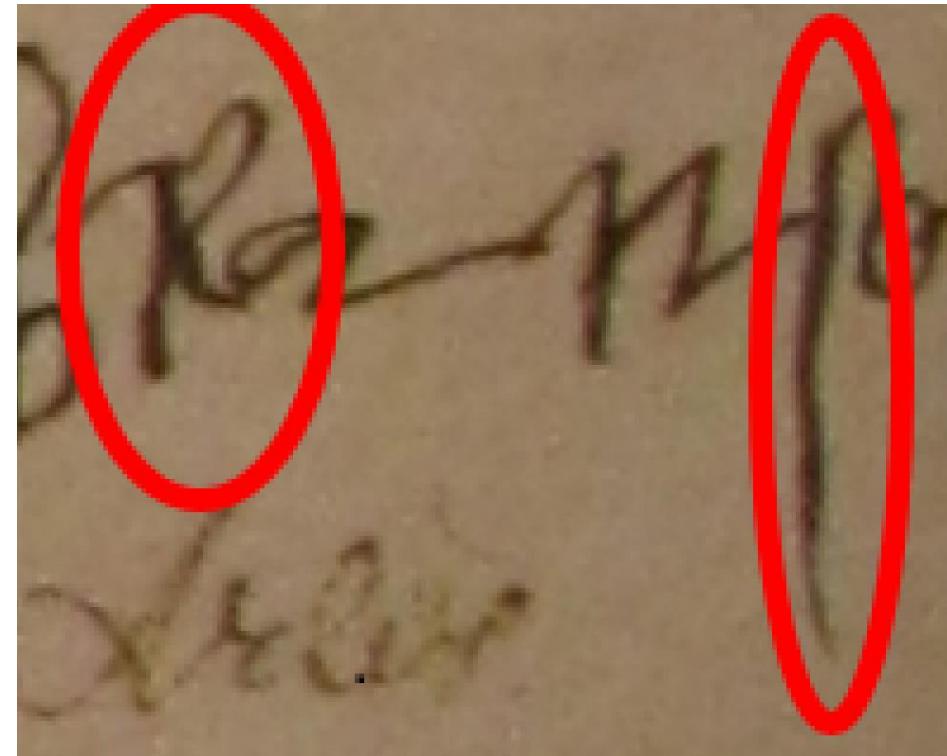
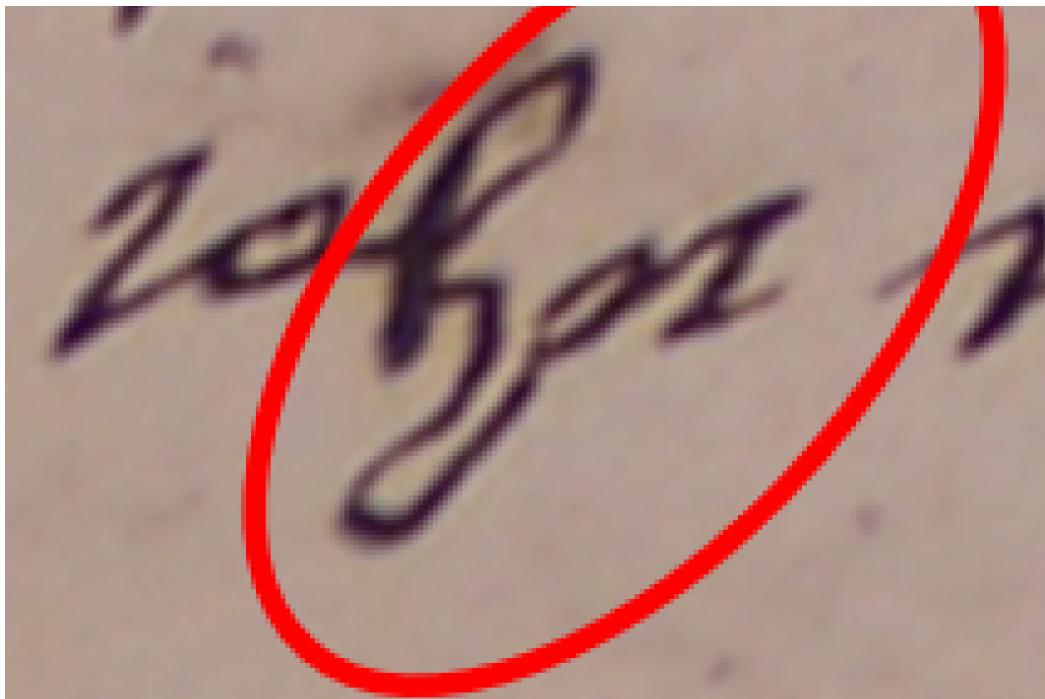
Deletion

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

Detection of “shake” in straight and curved lines



HYPOTHESIS: Shaky lines may be a sign of poor signature execution (and by inference, poor handwriting execution) suggesting either lower level of literacy than smooth executed lines, or the effect of illness or age

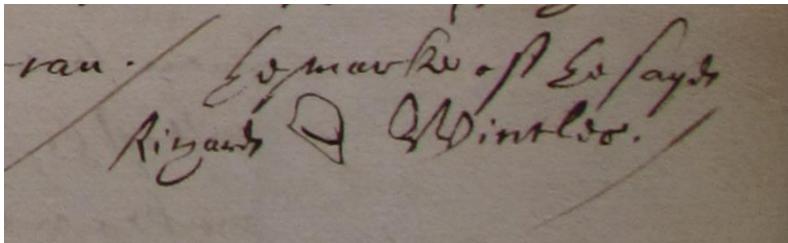


Shaky straight lines and/or loops

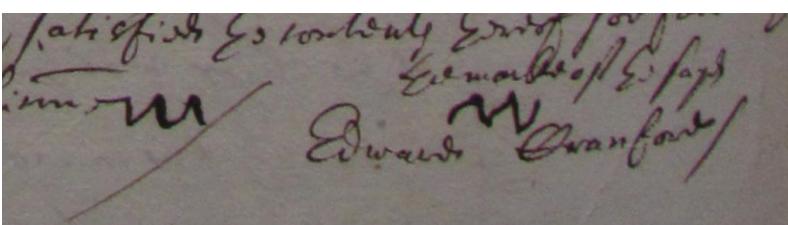
Clockwise, from top LH: KaggleTestSnippet_HCA_1371_f.435v.PNG,
KaggleTestSnippet_HCA_1368_f.483v.PNG,
KaggleTestSnippet_HCA_1368_f.483v_PIXELS.PNG,
KaggleTestSnippet_HCA_1371_f.435v.PNG_PIXELS.PNG

Porters handling coals, whale oil, ginger & corn

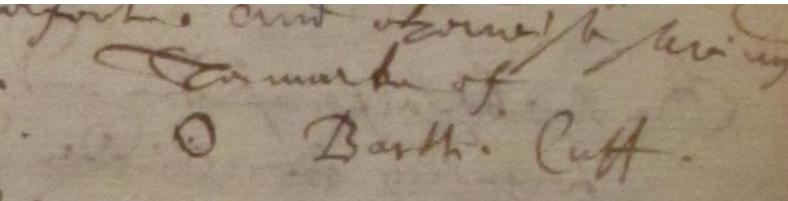
Richard Wincles, thirty-three year old porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Imployment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 s per man ([HCA 13/70 f.554r](#))



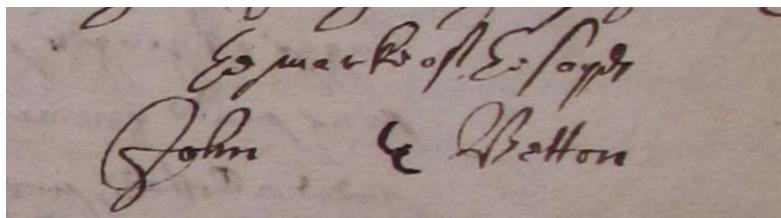
Edward Cranford, forty-four year old coale heaver or porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Imployment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 s per man ([HCA 13/70 f.555v](#))



Bartholomew Cuff, sixty year old porter of the Stillyard, of the parish of Allhallows the Greate, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key and loading them away into a warehouse ([HCA 13/70 f.555v](#))



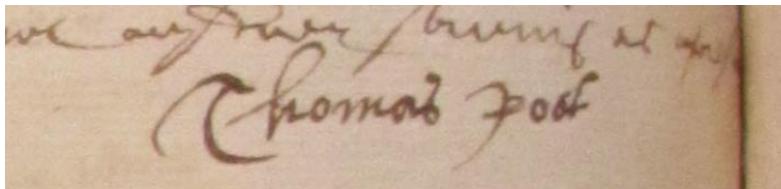
John Betton, fifty-four year old citizen and white baker of London, of the parish of Saint Buttolph Algate, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.449r](#))



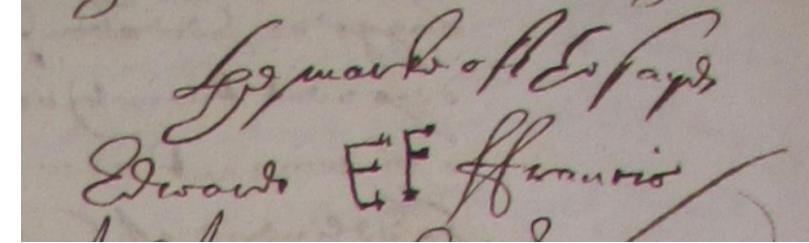
Edward Sherwin, fifty-six year old cittizen and leatherseller, of the parish of Little Allhallowes, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.449v](#))



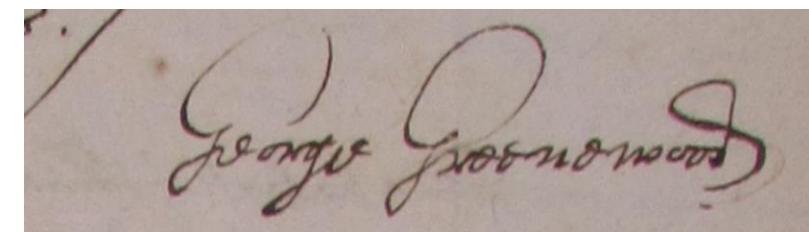
Thomas Roots, twenty-nine year old porter, of the parish of Greate Allhallowes, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key, as one of the Stillyard porters, and loading them away into a warehouse ([HCA 13/72 f.330v](#))



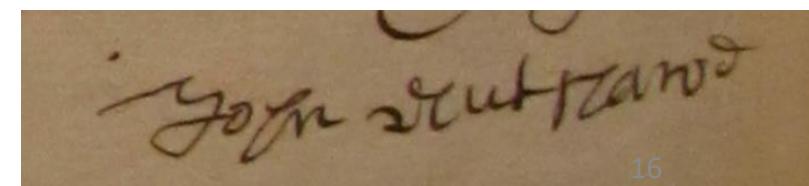
Edward ffrancis, citizen and merchant taylor of London, of the parish of Saint Olave in Southwarke, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.450v](#))



George Greenwood, thirty year old citizen and vintner of London, of the parish of Saint Buttolph Bishopsgate, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.454r](#))

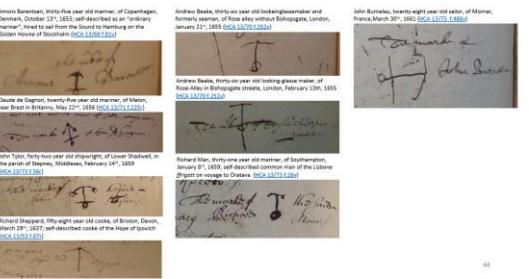


John Nutshall, fifty-five year old corne porter, of the parish of Saint Saviours Southwarke, Nov. 19, 1653; employed with a barber chyrurgeon/corne meter, an additional corne-meter, and other labourers to unlade a cargo of what in the *ffortune* of Stettin, moored against Limehouse; eight years of experience as a corne porter ([HCA 13/70 f.352v](#))



SOLM-2018 IIIF anchors manifest in Mirador viewer

Anchors



44

HCA Depositions: Anchors

Change Layout Full Screen

Richard Shepperd

Andrew Beake (2)

Andrew Beake (1)

John Tylor

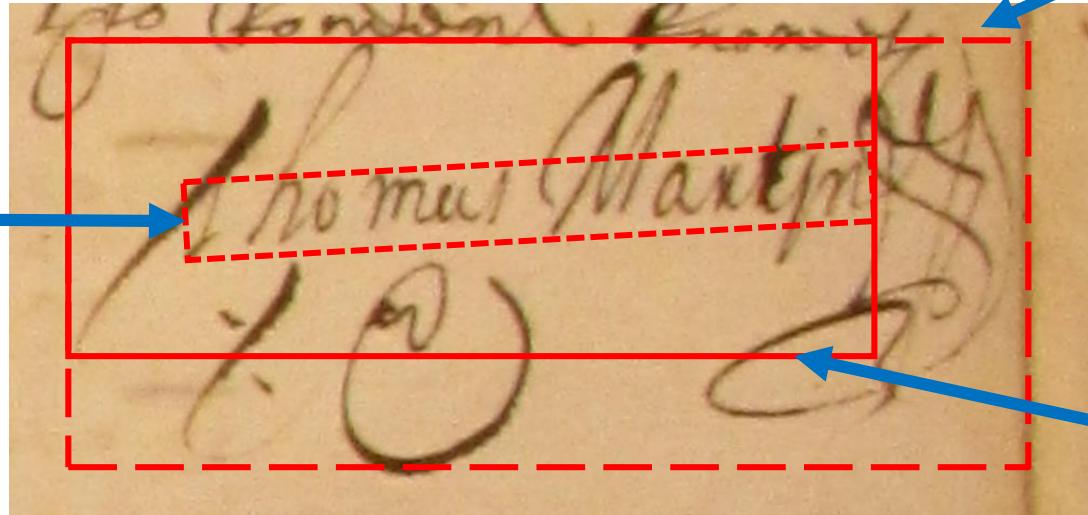
John Burnelau

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

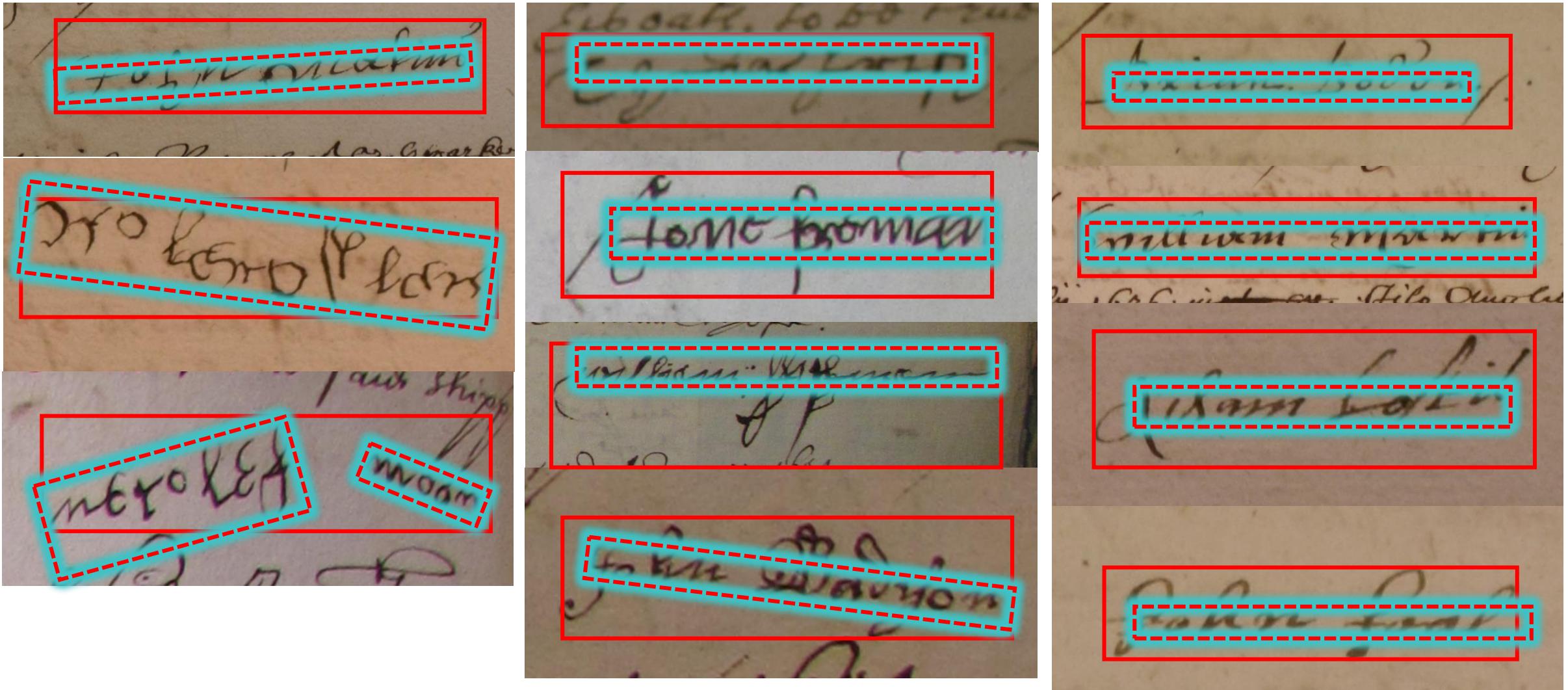
Inside boundary box: 9.0 x 1.1

Middle boundary box: 9.75 x 4.25

Outside boundary box: 12.75 x 5.75

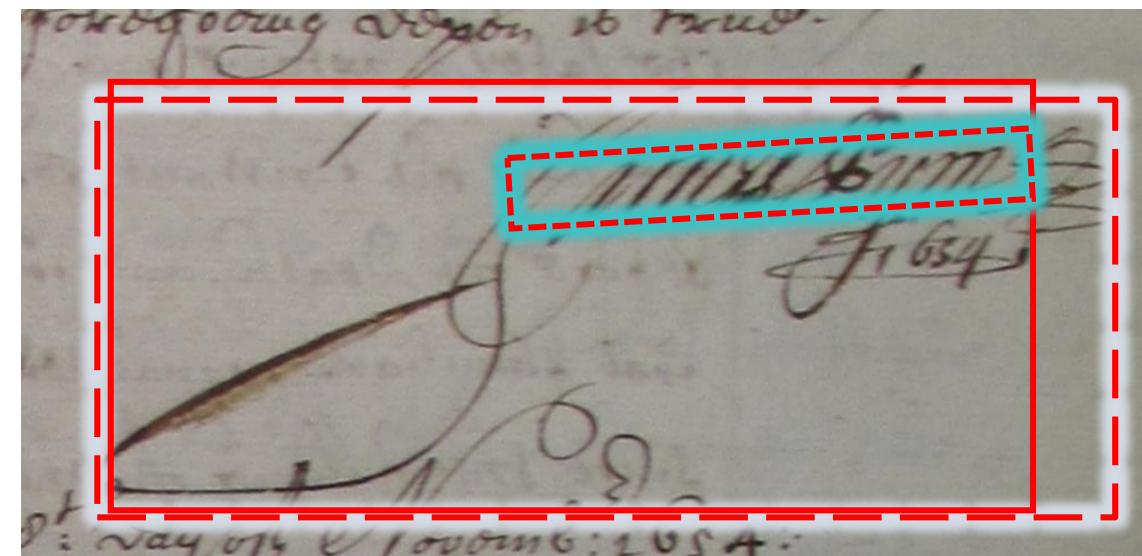
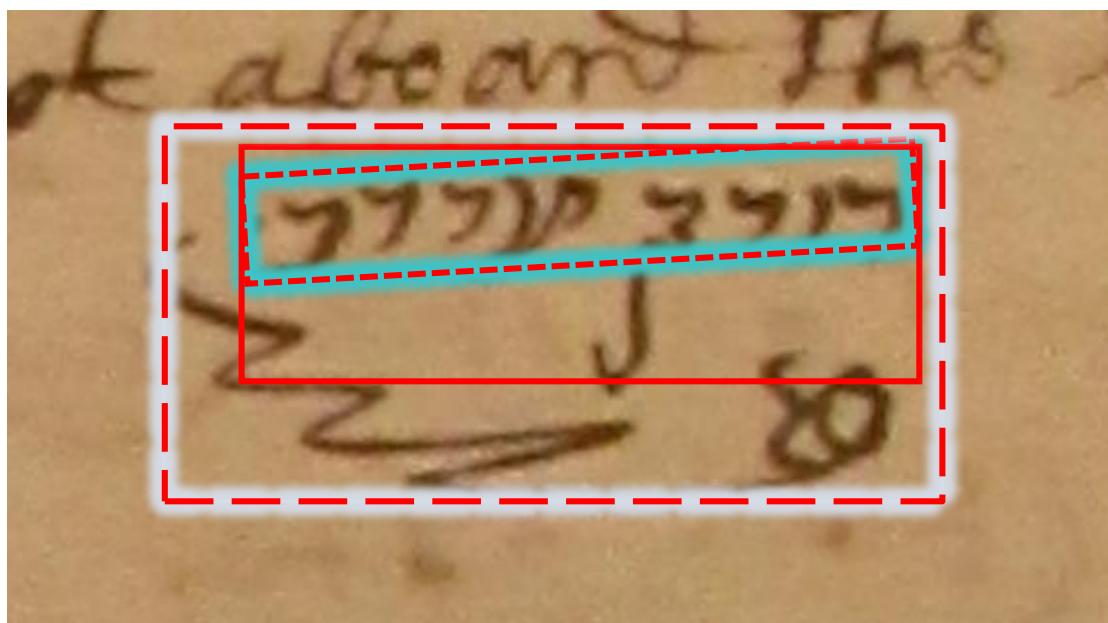
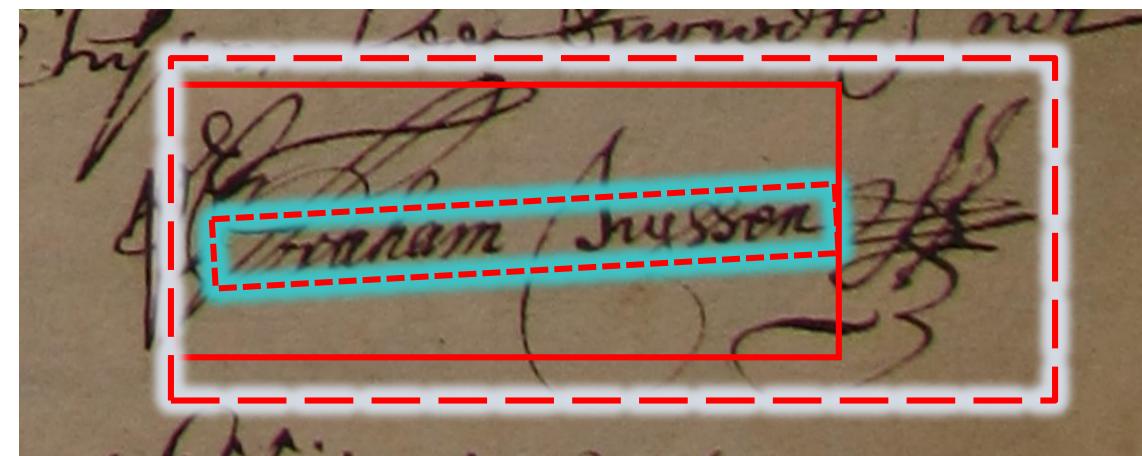
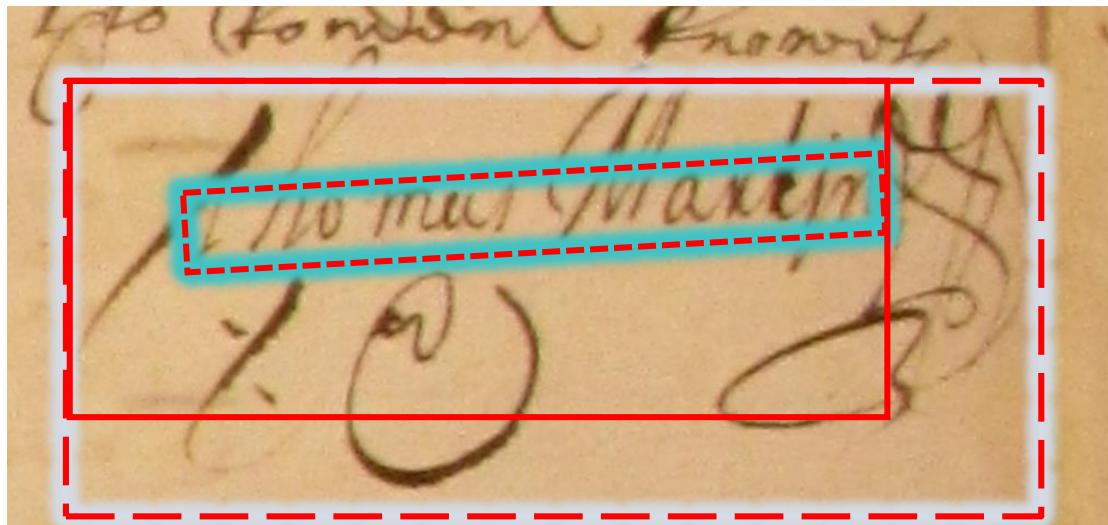
Rotation from horizontal: ca. 340 degrees

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



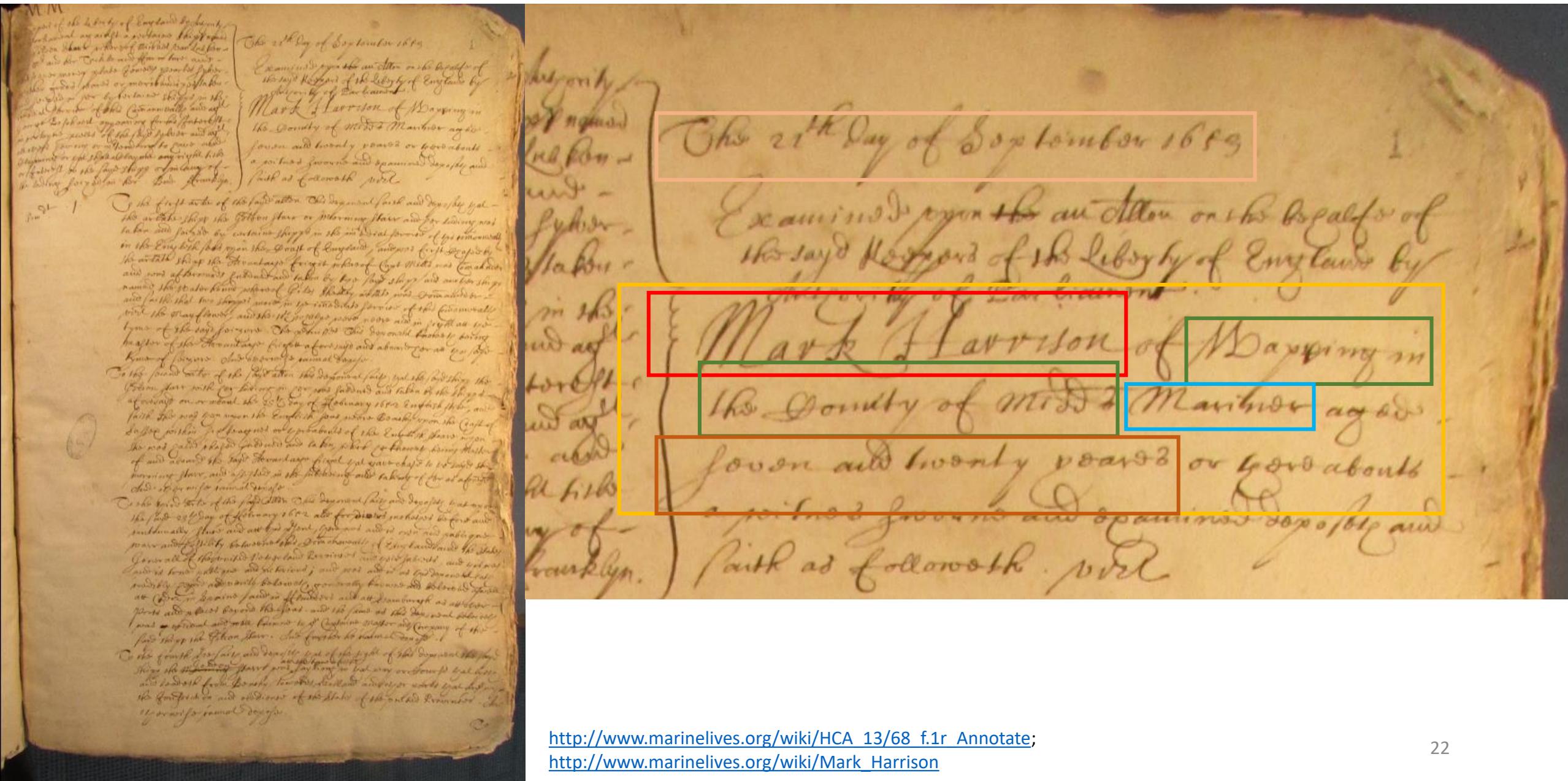
Source: Clockwise from top LH side: KaggleTestSnippet_HCA_1368_f.34v.PNG,
KaggleTestSnippet_HCA_1370_f.366r.PNG, KaggleTestSnippet_HCA_1370_f.134r.PNG,
KaggleTestSnippet_HCA_1368_f.58r.PNG

Legal deposition

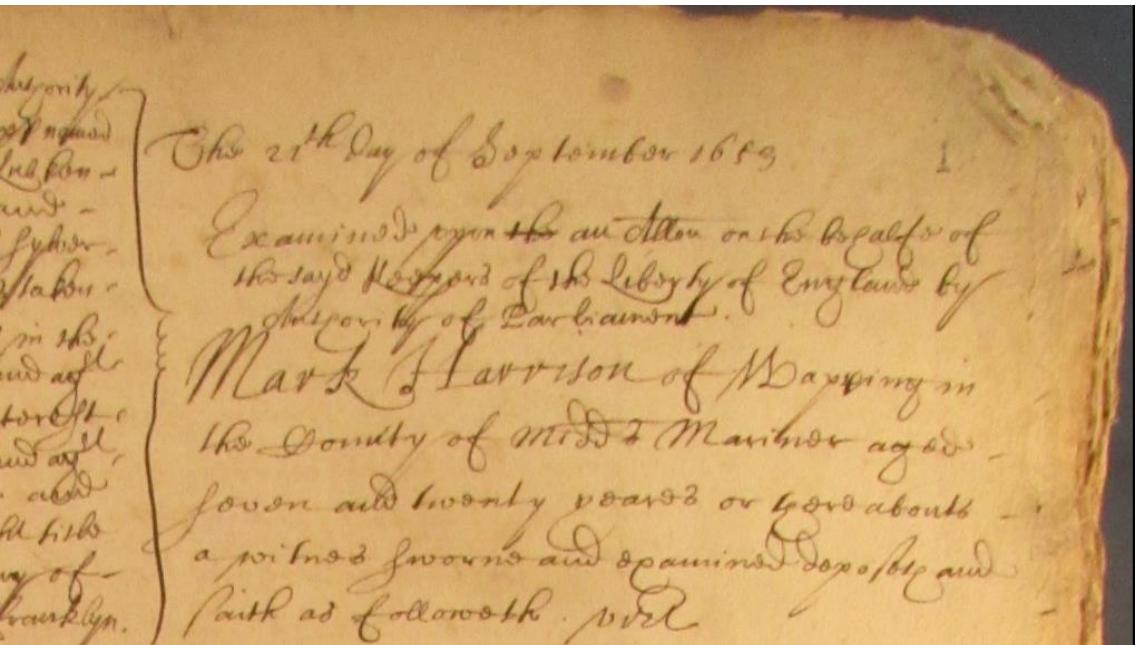
Deposition of Mark Harrison; mariner and master; resident in Wapping, Middlesex; age 27;
Dated September 21st 1659 (TNA, HCA 13/68, ff. 1r-3r)



Machine based recognition of metadata



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 Harris² and of⁴ what happened⁷ in² the county of Middlesex mariner⁸ aged seven and twenty years

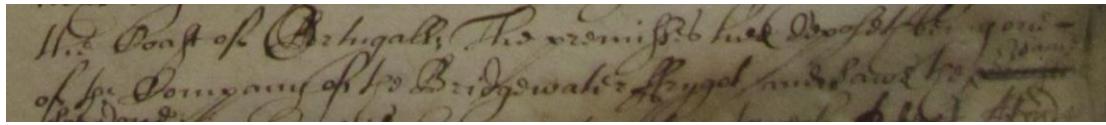
mark Harrison³ of² walking⁸ in² the county of Middlesex mariner¹⁵ aged seven and twenty years or² there⁴ about

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

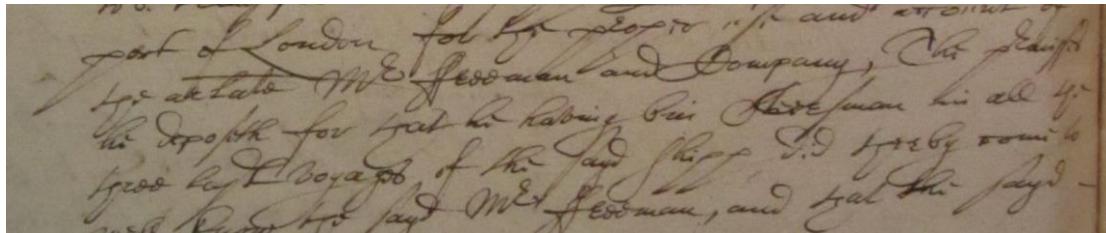
Can we use key word spotting to excavate raw metadata?

LANGUAGE DENOTING OCCUPATION



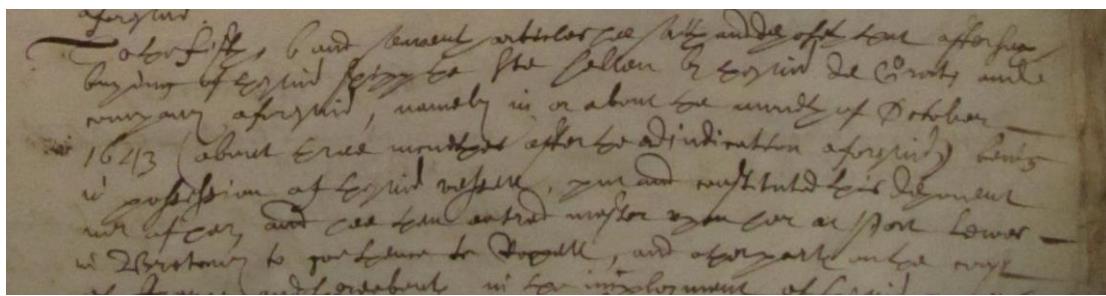
W^t Roast of Orthigall, The premissee hee deposeth
of the Companye of the Bridgewater ffrygott, and sawe her
in the same shipp in the said voyage.

"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]



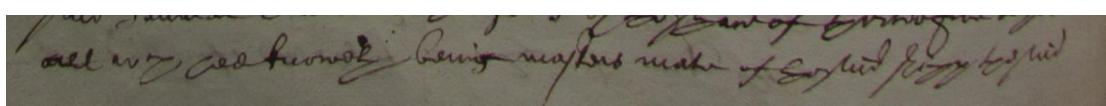
part of London for the sayd shipp and comon
for late Mr. Godman and Company, the shipp
the deponeth for that he had beene Steersman in all ye
free last voyages of the sayd shipp S. I. syngt downe
and to the sayd Mr. Godman, and that the sayd

"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]



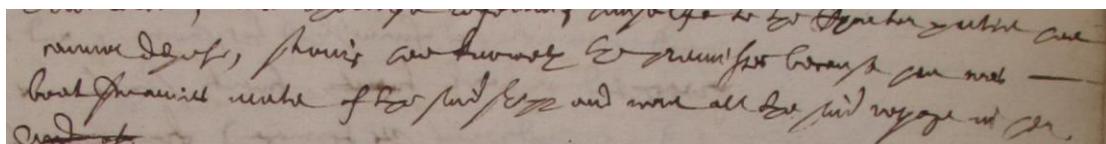
Yours.
To oblyf, and present witnessse by ande off late after
buying of Esq^r J^r to the Bellin or Esq^r de Grotte and
many other apysit, namely in or about the moneth of October
1643 (about three moneths after the adiudication aforesaid) being
in possession at Esq^r de Grotte, and was constytuted his Deponent
out of Esq^r and Esq^r master of the shipp at Port Lewes
in Bretany to go thence to Foggall, and afterwards in the same
shipp to Lorient in Bretany.

"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 knoweth being masters mate of sayd shipp by his
name deput, having now knoweth Esq^r de Grotte to be the
boat swaines mate of sayd shipp and made all sayd shipp voyage w^t her.

"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 knoweth Esq^r de Grotte to be the
boat swaines mate of sayd shipp and made all sayd shipp voyage w^t 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 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 which hee knoweth being masters mate of the said shipp the said voyage" [HCA 13/70 f.669v] [CONCLUSION: Master's mate]

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

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	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	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_687r.PNG - CREATE HCA 13/68	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
KaggleTestSnippet_HCA_1371_f_99r.PNG	HCA 13/71	Signature and	Mariner; Boatswain [of man of war]
KaggleTestSnippet_HCA_1370_f_484r.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

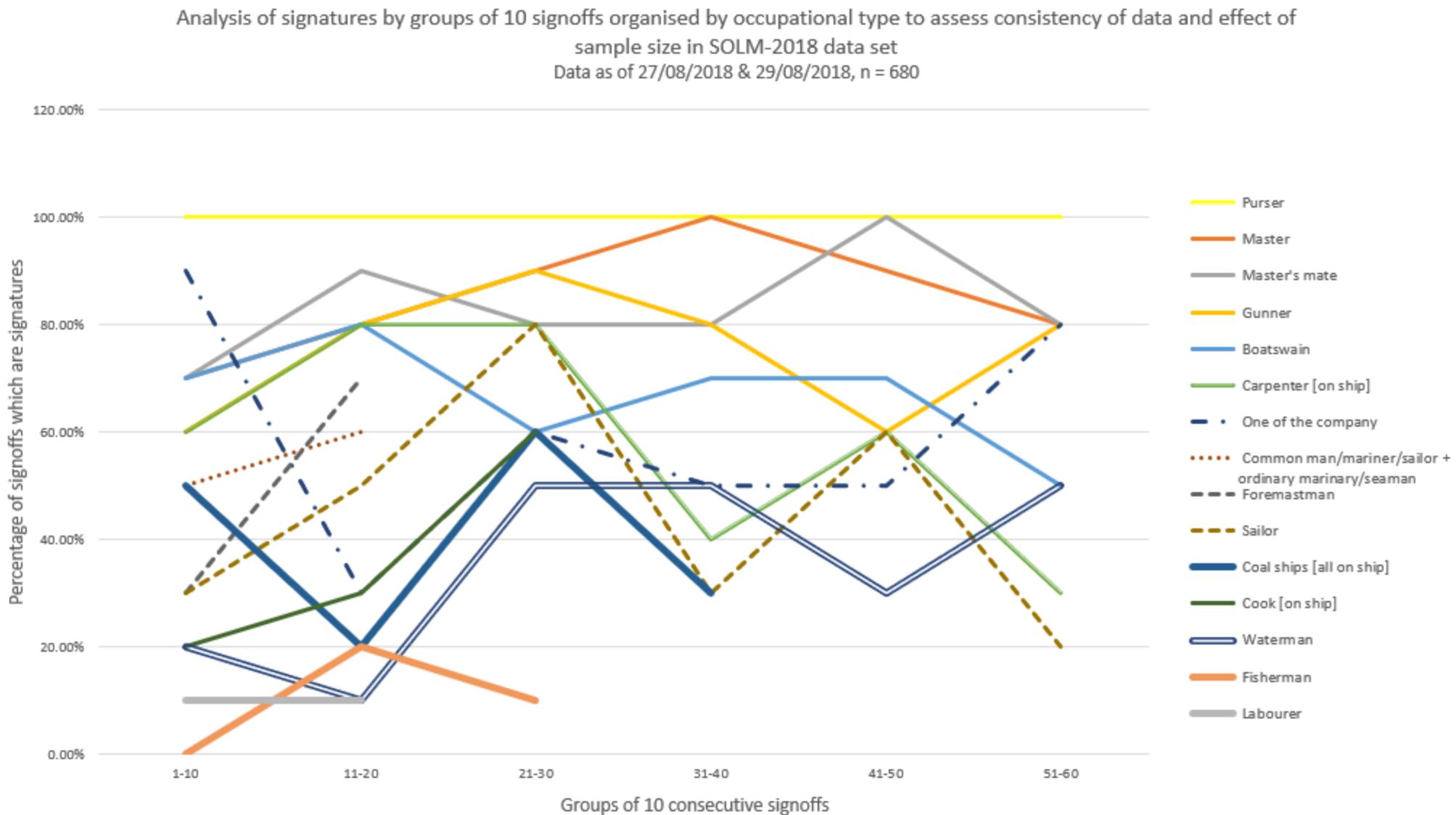
File name&c	Volume	Type	Occupation	Month	Year	Estimated	names	name	last name
				Wage	paid	age	Age	Age	Age
KaggleTestSnippet_HCA_1370_f_546r.PNG	HCA 13/70	Signature	Mariner; Boatswain	47	1654	1607 John	Baker	Baker	Cookson
KaggleTestSnippet_HCA_1370_f_571v.PNG	HCA 13/70	Signature	Mariner; Boatswain	40	1654	1625 Peter	Simonson	Simonson	Lee
KaggleTestSnippet_HCA_1370_f_596v_One.PNG	HCA 13/70	Signature	Mariner; Boatswain	30	1654	1625 John	Lee	Lee	Lee
KaggleTestSnippet_HCA_1370_f_636r.PNG	HCA 13/70	Signature	Mariner; Boatswain	26	1654	1629 John	Lee	Lee	Lee
KaggleTestSnippet_HCA_1370_f_671v.PNG	HCA 13/70	Marke	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1368_f_631v.PNG	HCA 13/68	Signature	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1371_f_27r.PNG	HCA 13/71	Initials	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1371_f_27v_One.PNG	HCA 13/71	Initials	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1371_f_27v_Two.PNG	HCA 13/71	Initials	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1368_f_640r.PNG	HCA 13/68	Signature	Mariner; Boatswain	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1368_f_687r.PNG - CREATE HCA 13/68	HCA 13/68	Signature	Mariner; Boatswain [of the Civill Society]	26	1655	1629 JOHN	Lee	Lee	Lee
KaggleTestSnippet_HCA_1371_f_77v.PNG	HCA 13/71	Signature	Mariner; Boatswain	22	1655	1635 Joseph	Cornwell	Cornwell	Cornwell
KaggleTestSnippet_HCA_1370_f_378r.PNG	HCA 13/70	Signature	Mariner; Boatswain	22	1655	1635 ROBERT	Shorring	Shorring	Eare
KaggleTestSnippet_HCA_1371_f_99r.PNG	HCA 13/71	Signature and	Mariner; Boatswain [of man of war]	22	1655	1635 ROBERT	Shorring	Shorring	Eare
KaggleTestSnippet_HCA_1370_f_484r.PNG	HCA 13/70	Signature	Mariner; Quartermaster; Boatswain	22	1655	1635 WILLIAM	Conquett	Conquett	Jackson
KaggleTestSnippet_HCA_1371_f_139v.PNG	HCA 13/71	Signature	Mariner; Boatswain	22	1655	1635 WILLIAM	Conquett	Conquett	Jackson
KaggleTestSnippet_HCA_1371_f_167r.PNG	HCA 13/71	Signature	Mariner; Boatswain [of the John and Mary]	22	1655	1635 WILLIAM	Conquett	Conquett	Jackson
KaggleTestSnippet_HCA_1371_f_279r.PNG	HCA 13/71	Signature	Mariner; Boatswain	41	1655	1634 CORNELLIUS	Peterson	Peterson	Mathyson
KaggleTestSnippet_HCA_1370_f_546r.PNG	HCA 13/70	Signature	Mariner; Boatswain	34	1655	1621 HENDRICK	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_571v.PNG	HCA 13/70	Signature	Mariner; Boatswain	34	1655	1621 HENDRICK	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_596v_One.PNG	HCA 13/70	Signature	Mariner; Boatswain	30	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_636r.PNG	HCA 13/70	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_671v.PNG	HCA 13/70	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1368_f_631v.PNG	HCA 13/68	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_687r.PNG	HCA 13/70	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1371_f_77v.PNG	HCA 13/71	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_378r.PNG	HCA 13/70	Signature	Mariner; Boatswain	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1371_f_99r.PNG	HCA 13/71	Signature and	Mariner; Boatswain [of man of war]	29	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_484r.PNG	HCA 13/70	Signature	Mariner; Boatswain	30	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1371_f_139v.PNG	HCA 13/71	Signature	Mariner; Boatswain	30	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1371_f_167r.PNG	HCA 13/71	Signature	Mariner; Boatswain	30	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1371_f_279r.PNG	HCA 13/71	Signature	Mariner; Boatswain	30	1654	1624 THOMAS	Stralund	Stralund	Akers
KaggleTestSnippet_HCA_1370_f_546r.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_571v.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_596v_One.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_636r.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_671v.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1368_f_631v.PNG	HCA 13/68	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1371_f_77v.PNG	HCA 13/71	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_378r.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1371_f_99r.PNG	HCA 13/71	Signature and	Mariner; Boatswain [of the John and Mary]	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1370_f_484r.PNG	HCA 13/70	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1371_f_139v.PNG	HCA 13/71	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1371_f_167r.PNG	HCA 13/71	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch
KaggleTestSnippet_HCA_1371_f_279r.PNG	HCA 13/71	Signature	Mariner; Boatswain	28	1655	1630 WILLIAM	Elton	Elton	Crouch

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

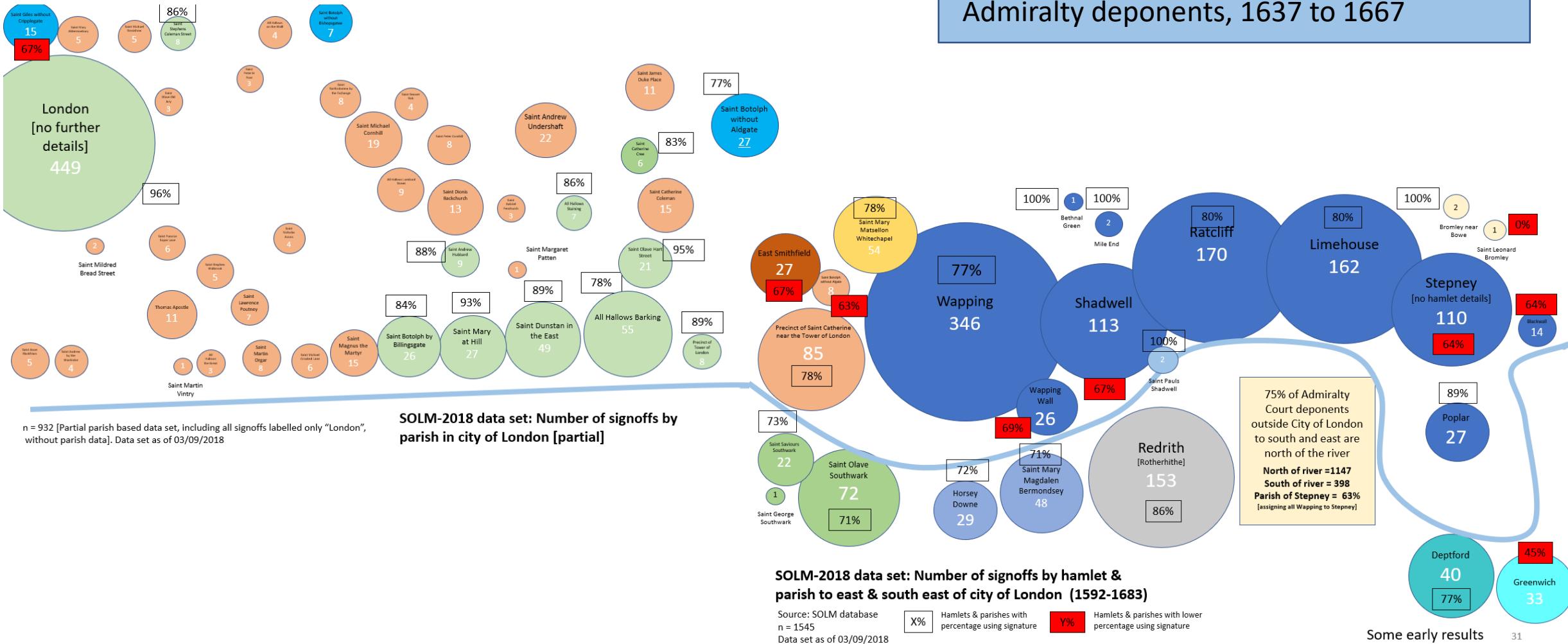
	1 r	2 v	3 r	4 v	5 r	6 v	7 r	8 v	9 r	10 v	Subtotal									
1-10	1	2	3	1	0	1	0	2	2		16									
11-20	2	2	1		1	1	1	1	1	1	13									
21-30	2	1		1	3	1	1	1	2	1	16									
31-40		1	1	1		1	1	2	1	1	13									
41-50		1		1			1	1	1		6									
51-60			1		1	2	1		2	2	11									
61-70	2			1	1		2	1	1	1	18									
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	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

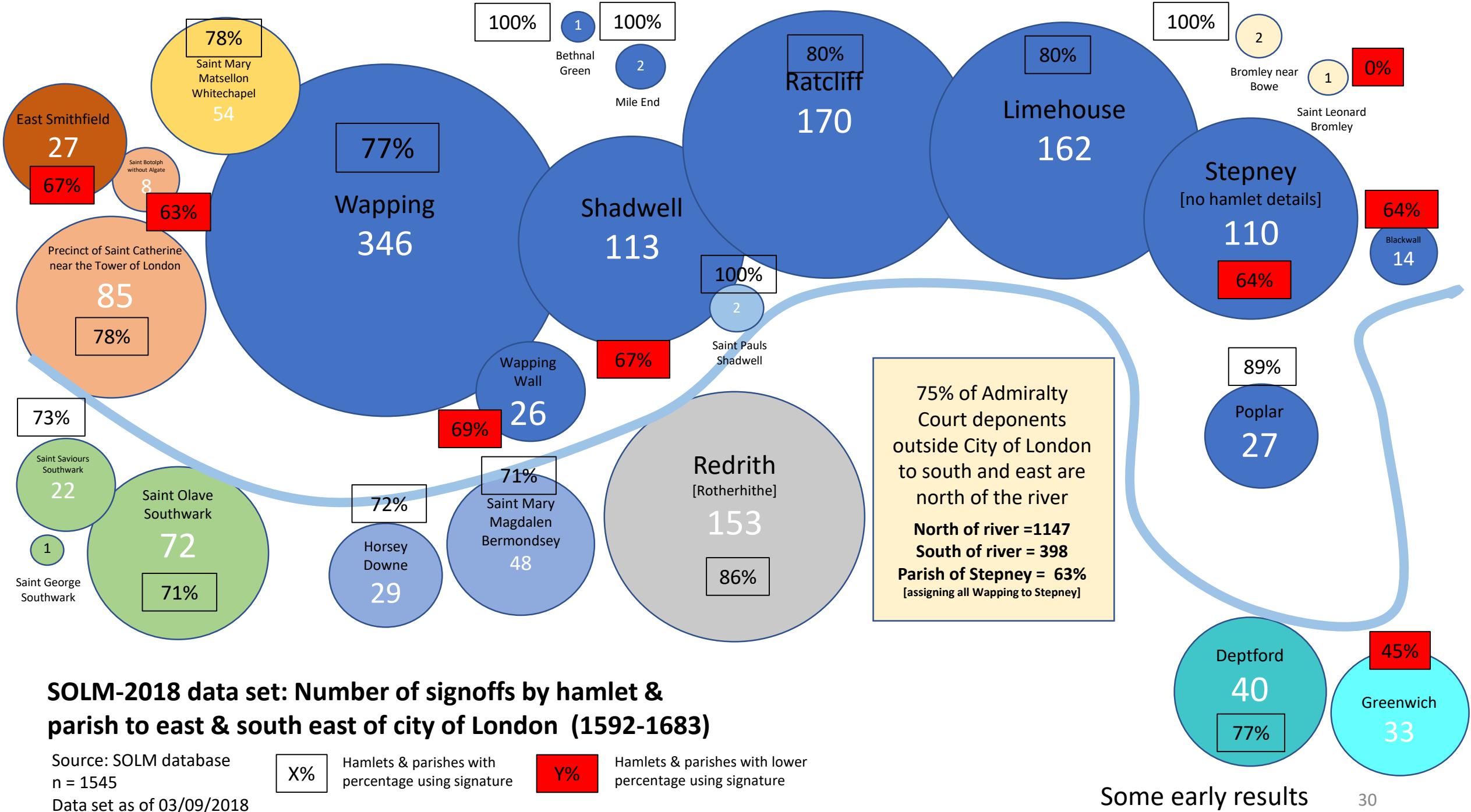
Some early results



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
m.hailwood@bristol.ac.uk

Weblinks:

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

GitHub:

<https://github/Signsofliteracy/Signoff>

Twitter:

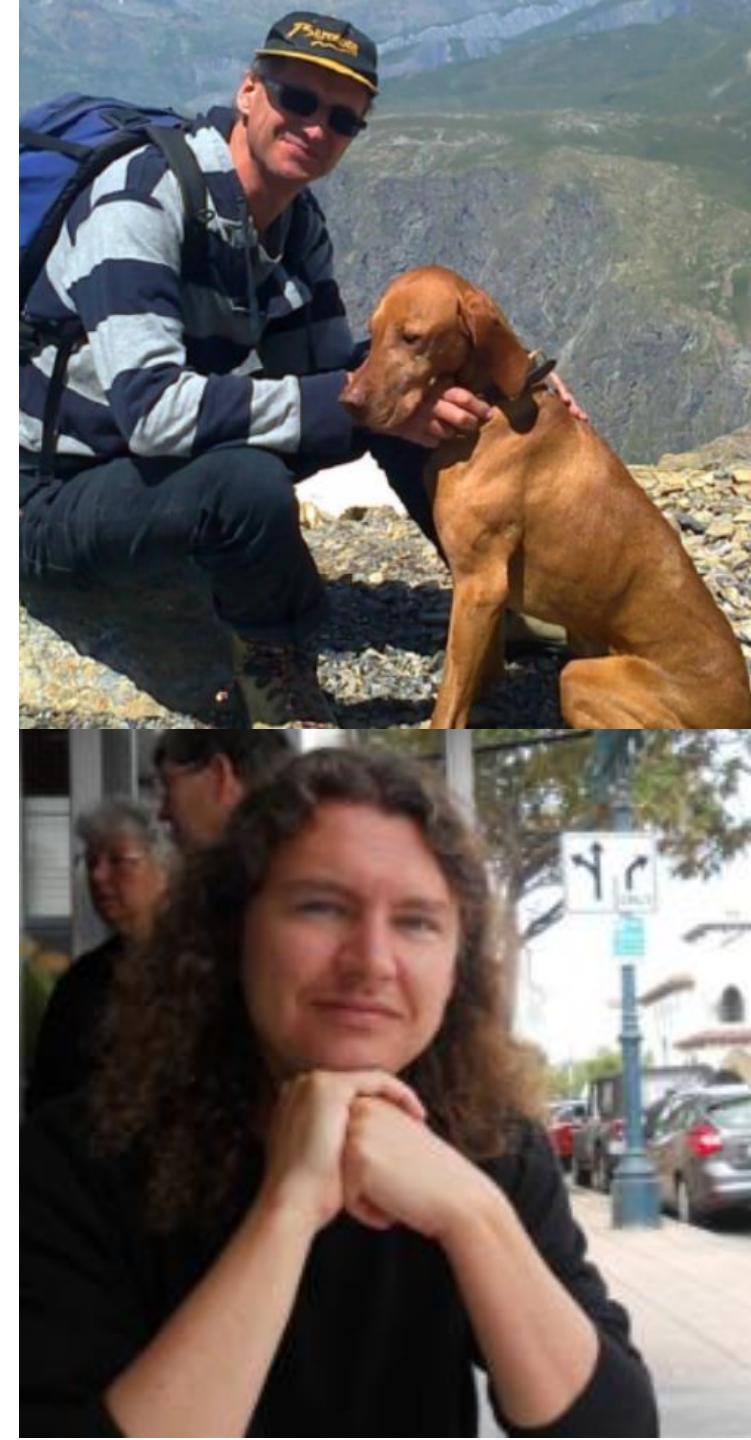
[Marinelivesorg](https://twitter.com/Marinelivesorg)

Working
with:



The National Archives

kaggle Competitions



Thanks to our one hundred and ninety-six contributors since 2012

We would like to recognise and thank our friends and supporters who have contributed to the MarineLives project, to the Maphackathon event, to the Early Modern Textiles, Garments and Dyestuffs glossary, and to the Signs of Literacy initiative, whether as volunteer transcribers, annotators, commentators, glossary contributors, software developers, advisors, interviewees, workshop participants, workshop and conference speakers, or as PhD Forum participants (in alphabetical order)

Dr Aquiles Alencar-Brayner	Thomas Davies	Viveka Hansen	M. L. Logue	Benjamin Redding	Dr Samantha Thompson
Dr Kimberly Alexander	Jonathan Dent	Elaine Harrington	Jelle van Lottum	Irmila Regulova	Roger Towner
Dr Aaron Allen	Melvyn Dresner	Dr Liam Haydon	Ismail Malik	Ethan Reynolds	Dr Alexis Truax
Elizabeth Ames	Dr Stuart Dunn	Phillipa Hellawell	Grace Mallon	aniel Richards	Dr William Tullett
Dr Roberta Anderson	Professor Kai Eckert	Laura Herbert	Paula Marmor	ariet Richardson	Martin Turner
Vicky Annand	Bob Egan	Dr Helmer Helmers	Simon Marsh	ndrew Richens	Oliver Turner
Dr Carolyn Arena	Eglantine	Dave Henderson	Dr Alan Marshall	n Mia Ridge	David Underdown
Deborah Ashby	Dr Charlene Eska	Dr Stefan Hessbrüggen-Walter	Dr Barbara McGillivray	n Dominique Ritzé	Rebecca Unsworth
Dr Gary Baker	Louise Falcini	Professor Tim Hitchcock	Dr Angela McShane	n Gavin Robinson	Dr Brodie Waddell
Gabor Bakos	Emilie-Jane Farrimond	Dr Philip Hnatkovich	Angela Middleton	en Robson	Rebecca Want
Iona Bartzi	Dr Karwah Fatah-Black	Rachel E. Holmes	John Miller	n Helen Rogers	Samuel Watson
Rachel Bates	Dr Amelia Fay	Jeffrey Hopper	Anne Mills	n Jill Rogers	Dominic Webb
Rowan Beentje	Dr Janet Few	Dr Jenni Hyde	Kate Morant	cel Hee Rømer	Ken Whittaker
Nga Bellis-Phan	Christian Folgar	Dr Tom Irvine	Matthias Müller-Prove	eline Romano	Jill Wilcox
Michael Bennett	Sara Fox	Steve Ives	Professor Steve Murdoch	n Margaret Schotte	Peter Williams
Amber Berkeley	Dr Anthony Firth	Alex Jackson	Dr Shavana Musa	n even Schrum	Royline Williams-Fontenelle
Dr Richard Blakemore	Erin French	Stefan Jäggi	Maurice Nicolson	nđrei Seleznov	Ad van der Zee
Lior Blum	Dr Ian Friel	Elin Jones	Stephanie Ostrich	ura Seymour	Dr Kathrin Zickermann
Ffion Boyd	Professor Cheryl Fury	Sue Jones	Frances Owen	Inn Sheridan	Dr Suze Zijlstra
Katie Broke	Steve Garnett	Menno Jonker	Gordon O'Sullivan	n Deborah Sherlock	Dr Cäcilia Zirn
Dr James Brown	Dr Perry Gauci	Dr Sam Kaislaniemi	Dr Katherine Parker	n Rainer Simon	and the ever helpful but
Dr Andy Burn	Marja Geesink	Ross Keel	Tim Parry-Williams	a Sjoberg	anonymous @_mapnut
Elio Calcagno	Jaap Geraerts	Dr Patricia Keller	David Pashley	ntr Sloup	
Joshua Calloway	Helen Good	William Kellett	Dr Michael Pearce	n Edmond Smith	
Michelle Maria Early Capistrán	Dr Jamie LH Goodall	Sara J Kerr	Dr Cathryn Pearce	Illy Smith	
Rachel Carter	Guy Grannum	Heather Knight	Elizabeth Pimblett	n Stephen Snelders	
Cynthia Chin	Colin Greenstreet	Iris Kramer	Dr Sophie Pittman	ofessor Humphrey Southall	
Nicola Clarke	Francesca Greenstreet	Dr Marcin Krygier	Professor Simone Paolo	n Craig Spence	
Freyalyn Close-Hainsworth	Justin Grierson	John Kuhn	Ponzetto	aniel Stewart-Roberts	
Dalton Coker	Tracey E Griffiths	Dr Craig Lambert	Mark Ponte	n Hannah Steyne	
Dr Giovanni Colavizza	Dr Adam Grimshaw	John Layt	Dr Peter Pridal	n Stoll	
Dr Justin Colson	Karen Gunnell	Deborah Leem	Dr Jo Pugh	ichael Stoner	
Sam Cottrell	Yerevag Hagopian	Dr Sjoerd Levelt	Seif Al Rashidi	n Chad Stolper	
Kevin Daniels	Dr Mark Hailwood	Amalia Skarlatou Levi	Dr Peter Rauxloh	eter Taylor	
Thierry Daunois	Finn Halligan	John Levin	Patrizia Rebullia	n Carl Thompson	
Dr John Davies		M. L. Logue			

Discussion

For discussion at Archives & AI symposium, Tuesday, September 4th, 2018

	2018	2019	2020	2021	2022	2023
Q1		Historical literacy network meeting (1)				
Q2	IIIF Conference Staatsarchiv Amsterdam workshop DH Benelux	Historical literacy network meeting (2)				+ <div style="border: 1px solid black; padding: 10px;"><p>Machine learning enabled manuscript archivists</p><p>Image processing</p><p>Natural Language Programming</p><p>Key word spotting</p><p>Controlled vocabularies</p><p>Programmable decision rules</p><p>IIIF visibility</p></div>
Q3	Archives & AI Sheffield DH Congress	Kaggle competition Huntington library visit				
Q4	Kaggle visit Stanford University visit Transkribus Vienna users conference					
	SOLM-2018	SOLM-2019	SOLM-2020	SOLM-2021	SOLM-2022	SOLM-2023