

**The Alan Turing Institute Data Study Group | Spring 2018**

**Monday 16<sup>th</sup> April – Friday 20<sup>th</sup> April 2018**

**Accenture Exclusive**

Challenge form

- 1, A-D: Defining the science challenge
- 2, A-D: Dataset and data access
- 3, A-D: Impact

**1: Defining the science challenge**

A) What is the business/operational/societal background to this science challenge?  
[description should be suitable for a well-educated lay-person]

**MarineLives: “Pattern recognition of signatures and marks in historical manuscripts as a basis for sub-population identification”**

**Goals of the challenge**

The proposed MarineLives science challenge for the Spring 2018 Alan Turing Institute Data Study Group serves two purposes.

Firstly, to enrich understanding of our medium sized occupational, socio-demographic and image data using pattern recognition techniques drawn from data science.

Secondly, to develop a use case for both historians and data scientists to show the power of working together.

**Social and educational background**

[MarineLives](#) is a collaborative volunteer driven project, founded as a spinoff from a National Archives hackathon in early 2012. It is dedicated to the collaborative transcription, linkage and enrichment of primary manuscript documents from the [English High Court of Admiralty](#), 1627 to 1677, together with thematically related manuscripts from international manuscript and printed document collections. These data are a powerful source of commercial, social, material, legal and linguistic insight, and are of relevance to historians, historical sociologists, historical and corpus linguists, and archaeologists. We publish as we go, under a CC BY 3.0 licence, on our [semantic media wiki](#). Currently we have 12,606 text pages and 11,843 images available and nearly six million words of full text transcriptions.

MarineLives is unique amongst expert sourcing and crowdsourcing projects in its application of project management and facilitation techniques drawn from management consultancy and pharmaceutical R&D. We provide a high degree of support and training to our volunteers through facilitators, modelled on engagement managers, underpinned by [well developed wiki based training material](#), and enlivened with regular Skype facilitated team meetings for volunteers drawn from England, Scotland, Wales, Ireland, France, the Netherlands, the United States and Canada.

We are actively looking for new archival and library partners, in California, London and the Netherlands, as we scale up our activities. In the next five years we would like to grow our full text corpus to twenty-five million words and our image collection to 50,000 images.

The originality, drive and impact of MarineLives was recognised by our volunteers, our supporters, and by a number of leading academics, in testimonials submitted to the [Huntington Library](#), San Marino, California, in November 2017. We are also in discussions with the British Library about potential collaboration in the field of hand writing recognition, metadata extraction and enrichment, and collaborative transcription. Our collaboration with the [University of Warwick History Department](#) for the teaching of palaeography, and our past collaboration with the Bath Spa University History Department, are models for the expansion of our university reach.

### **Charitable aims**

We are registering a charitable incorporated organisation under the name of Chronoscopic Education, to be the legal, funding and technology home for MarineLives and our sister project [Maphackathon](#).

The aims of Chronoscopic Education are threefold:

- (1) To further the teaching of palaeographic, digital research and project management skills, at universities and schools.
- (2) To apply insights from small teams in management consultancy and R&D project management, together with tools from data science, to the subject matter of history.
- (3) To build a virtual manuscript-based archive and associated research community, which will foster a culture of collaborative scholarship.

B) What is the science challenge that you want to solve? [description should be 1-3 questions that can, in principle, be answered quite simply (Yes, No, a number, etc)]

(1) Can AI be used to discriminate within sets of signatures and signature marks of Admiralty Court witnesses by reference only to their visual and other physical characteristics?

(2) Can AI be used to extract metadata concerning name, occupation, place of residence, age and date of witness statement from digital images, without reference to a manual transcription?

(3) Can AI based visual and other physical discrimination within sets of signatures and marks of Admiralty Court witnesses be used alone, or in combination with other characteristics of Admiralty Court witnesses and witness statements (name, occupation, place of residence, age, date of witness statement), to identify non-intuitive sub-populations?

Ideally all three sections of the challenge would be addressed at the DSG. However, section two, whilst a powerful question, is optional. We have sufficient pre-transcribed data, extracted manually from images, concerning name, occupation, place of residence, age and date of witness statement, to enable (1) and (3) to be addressed, without (2).

C) Why is the solution to the science challenge not obvious? [description should identify the main difficulties, or approaches that have been considered/tried before.]

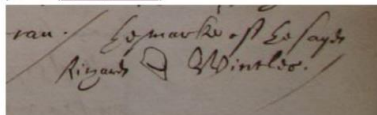
(1) Manual techniques lack rigour and lack agreed scoring criteria.

- Manual inspection of images reveals differences in technical control of a pen (solidity; regularity of line, presence or absence of ink smudges); in presence or absence of flourishes; in style or formal school of handwriting; in supplementary information; and other aspects.
- However, there is an absence of agreement as how best to distinguish and classify signatures and marks, and no agreed quantified scoring criteria.

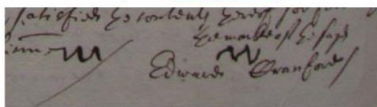
## Examples

### Dockyard & warehouse porters, handling coals, whale oil, ginger & corn

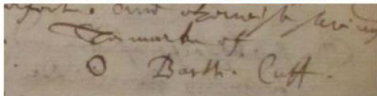
Richard Wincles, thirty-three year old porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Employment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 s per man ([HCA 13/70 f.554r](#))



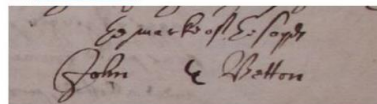
Edward Cranford, forty-four year old coale heaver or porter, of the parish of Stepney, Middlesex, Dec. 15, 1656; employed as a labourer with fifteen other men to unload coals from the *Employment* moored near Execution Dock, Wapping, into lighters for fixed rate of 12 s per man ([HCA 13/70 f.555v](#))



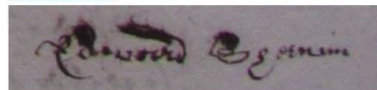
Bartholomew Cuff, sixty year old porter of the Stillyard, of the parish of Allhallows the Greate, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key and loading them away into a warehouse ([HCA 13/70 f.555v](#))



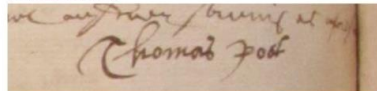
John Betton, fifty-four year old citizen and white baker of London, of the parish of Saint Buttolph Algate, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.449r](#))



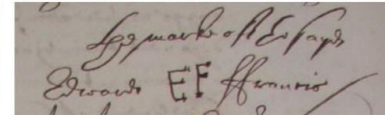
Edward Sherwin, fifty-six year old citizen and leatherseller, of the parish of Little Allhallows, London, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.449v](#))



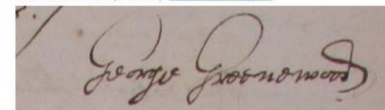
Thomas Roots, twenty-nine year old porter, of the parish of Greate Allhallows, London, May 15, 1658; assisted in the landing of whale oil from lighters at the Stillyard Key, as one of the Stillyard porters, and loading them away into a warehouse ([HCA 13/72 f.330v](#))



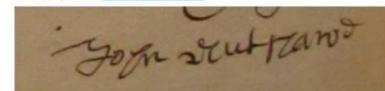
Edward ffrancis, citizen and merchantaylor of London, of the parish of Saint Olave in Southwarke, Jul. 31, 1655; self-described as a porter employed by the Commissioners for Prize Goods to deliver ginger from a warehouse at Ralphes Key ([HCA 13/70 f.450v](#))



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](#))

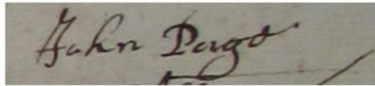


John Nutshall, fifty-five year old corne porter, of the parish of Saint Saviours Southwarke, Nov. 19, 1653; employed with a barber chyrurgeon/corne meter, an additional corne-meter, and other labourers to unlade a cargo of what in the *ffortune* of Stettin, moored against Limehouse; eight years of experience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

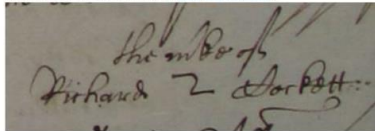


## Common men, foremast men, gunners, carpenters, boatswains, master's mates on board coal ships

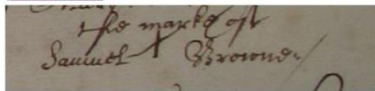
John Page, twenty-four year old mariner, of Alborough, Suffolk; Nov. 24, 1654; self-described as gunner, on the deck of the *Mary* of Alborough, when she collided with the *Freeman*, whilst coming up the River Thames in Long-Reach, known also as Gallions Reach ([HCA 13/70 f.28r](#))



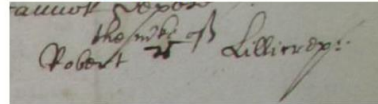
Richard Cockett, fifty year old mariner, of Alborough, Suffolk; one of the company of the *Mary* and self-described as master's mate; Nov. 24, 1654; in the *Mary* of Alborough, when she collided with the *Freeman*, whilst coming up the River Thames in Long-Reach, known also as Gallions Reach ([HCA 13/70 f.29r](#))



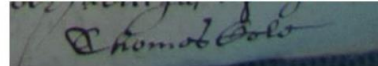
Samuel Browne, fifty year old mariner & boatswaine of the *Mary*, of Alborough, Suffolk; Nov. 24, 1654; in the *Mary* of Alborough ["the ship been a constant trader for coals from Newcastle to London"], when she collided with the *Freeman*, whilst coming up the River Thames in Long-Reach, known also as Gallions Reach ([HCA 13/70 f.30r](#))



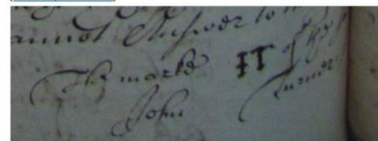
Robert Lillicrep, twenty-eight year old mariner & one of the company of the *Mary*, of Wapping, in parish of Stepney, Middlesex; Nov. 25 1654; self-described as "an ordinary mariner", in the *Mary* of Alborough, when she collided with the *Freeman*, whilst coming up the River Thames in Long-Reach, known also as Gallions Reach ([HCA 13/70 f.31r](#))



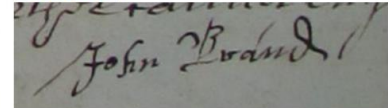
Thomas Cole, mariner & gunner of the *Mary Rose*, of Acton near Ipswich; Nov. 30 1659; in the *Mary Rose*, when she collided with the *John* of Ipswich, both laden with coals, whilst coming out of Tinnmouth with many other coal ships ([HCA 13/73 f.418r](#))



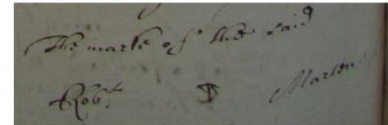
John Turner, forty-one year old shipwright & carpenter of the *Mary Rose*, of Ipswich; Nov. 30 1659; in the *Mary Rose*, when she collided with the *John* of Ipswich, both laden with coals, whilst coming out of Tinnmouth with many other coal ships ([HCA 13/73 f.419v](#))



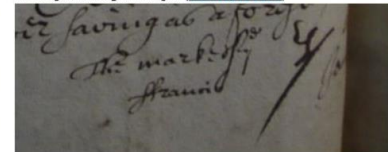
John Brand, forty-four year old, of Acton [near Ipswich]; Dec. 1 1659; self-described "foremast man" in the *Mary Rose*, when she collided with the *John* of Ipswich, both laden with coals, whilst coming out of Tinnmouth with many other coal ships ([HCA 13/73 f.421r](#))



Robert Marten, fifty-one year old mariner, of Acton near Ipswich; Dec. 1 1659; self-described "foremast man" in the *Mary Rose*, when she collided with the *John* of Ipswich, both laden with coals, whilst coming out of Tinnmouth with many other coal ships ([HCA 13/73 f.422v](#))



Francis Jennings, nineteen year old mariner, of Shadwell in parish of Stepney, Middlesex; Jan. 27, 1659; self-described "comon man" in the *Agreement*, when she came from Newcastle in Oct. 1658 in a great fleet of English colliers, bound for London, & collided with the *Abigail*, sinking the *Abigail* ([HCA 13/73 f.33v](#))



[See also Colin Greenstreet, 'Using pattern recognition of signatures & marks to explore occupational fluidity', pub. Feb. 10, 2018. <https://www.linkedin.com/pulse/using-pattern-recognition-signatures-markes-explore-colin-greenstreet/>, accessed 12/02/2018]

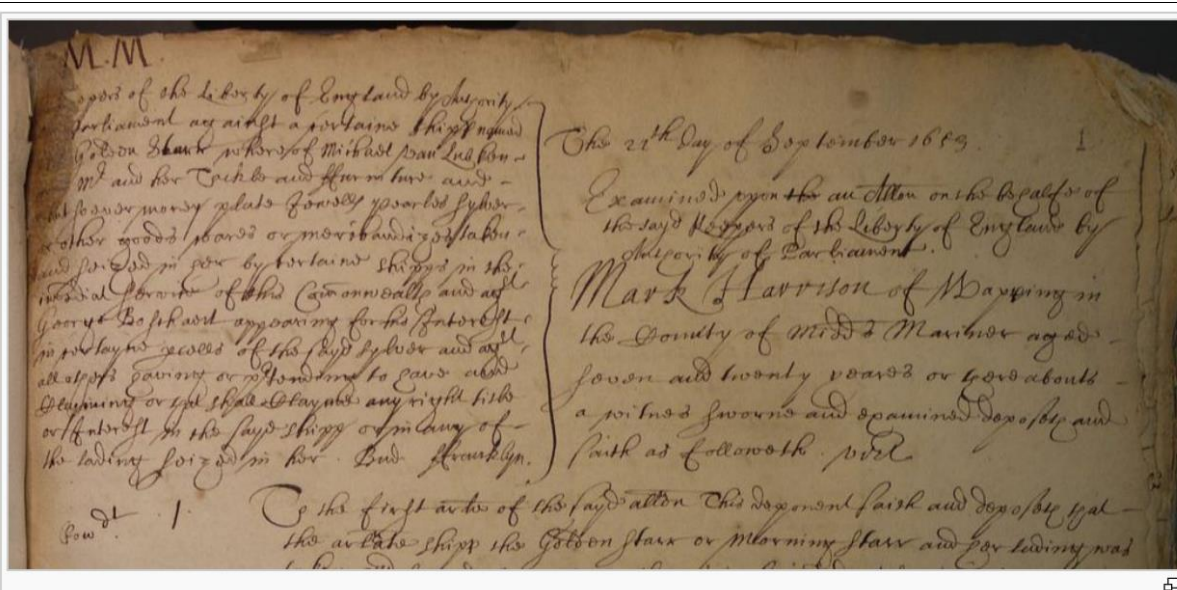
(2) Pattern recognition has not been applied to historical signatures and marks as a discrete field of study (as opposed to being part of manuscripts processed by HTR engines).

- Put simply, this is an unsolved challenge to an interesting question, with an interested audience of non-technical historians, who would like to see what data science can bring to the party.
- Specifically, we are interested in largely unsupervised clustering techniques which might be used and crossed with metadata associated with individuals.

(3) Progress is being made on automatic and semi-automatic document layout and baseline recognition, by scientists associated with the READ initiative and at a number of academic institutions.

Example [[HCA 13/68 f.1r](#)]: First witness statement in a series, with date, name, place of residence, occupation & age in discrete areas on page.





- However, there is a lack of use cases demonstrating the successful extraction of metadata from digital manuscript images.
- Several rival methods exist by which metadata might be recognised and extracted, including different approaches to HTR, and key word spotting. It is not clear which is superior for the application we propose.

D) What is your scientific basis for believing the science challenge can be solved?  
[description might include the available data and possible methods, and why they seem relevant.]

#### **Validation of science challenge by historians, palaeographers, archival based technical staff and by HTR specialists**

In the course of preparing this proposal, we have consulted with historians and palaeographers about the usefulness of the proposed research challenge to them as individuals and to their fields, and with archival based technical staff and HTR specialists as to the level of their interest in AI based pattern recognition applied to signatures and marks, and to the attempted machine-based extraction of metadata from images.

Dr Brodie Waddell (Birkbeck), a C17th social historian, and Dr Perry Gauci (Lincoln College, Oxford), a C17th and early C18th social and economic historian, have confirmed their interest in the questions and the approach. Social historian Dr Mark Hailwood (Bristol) highlighted analytical issues in a 2014 article. He has confirmed to us his strong interest in the proposed use of AI, as he extends his own research interests in the area of seventeenth century literacy.

Mark Hailwood, 'The Rabble that Cannot Read'? Ordinary People's Literacy in Seventeenth-Century England', Oct. 13, 2014, The Many-Headed Monster, web resource. <https://manyheadedmonster.wordpress.com/2014/10/13/the-rabble-that-cannot-read-ordinary-peoples-literacy-in-seventeenth-century-england/>, accessed 12/02/2018.

Dr Jo Pugh, Digital Development Manager, National Archives, has identified significant interest at the TNA in pattern recognition, natural language processing and metadata

extraction as tools to enrich digital manuscript collections and metadata. He notes the relevance of the proposed MarineLives research challenges to work planned on the PROB 11 collection of over one million wills, and the relevance of occupational identification and metadata extraction to other collections such as SC, WARD and REQ. He has expressed personal interest in the MarineLives research challenge and has offered to put in some limited time to assist MarineLives. Jo is a past volunteer on one of our summer transcription programmes, and was the organiser of the TNA hackathon from which MarineLives was spun out.

Dr Günter Mühlberger, senior project manager of the Digitisation and Digital Preservation group, Department of German Language and Literature, University of Innsbruck & coordinator of the READ/Transkribus project, has also expressed interest in our proposed use of pattern recognition on signatures and marks as a route into segmentation of medium scale populations. We have confirmed with his team the operational details of using Transkribus tools to mark-up ca. 4,000 images both for signatures and marks and for metadata around names, occupations, places of residence, ages and dates of witness statements.

### **Well developed field of scientific enquiry dedicated to modern signature verification**

Luiz G. Hafemann et al., 'Offline Handwritten Signature Verification – Literature Review', arXiv:1507.07909v4 [cs.CV] 16 Oct 2017.

<https://arxiv.org/pdf/1507.07909.pdf>, accessed 12/02/2018

### **Considerable scientific interest and technical progress in development of manuscript document layout recognition, handwriting text recognition and key word spotting**

#### Journal articles

K. Zagoris, I. Pratikakis and B. Gatos, "Segmentation-Based Historical Hand-written Word Spotting Using Document-Specific Local Features," 2014 14th International Conference on Frontiers in Handwriting Recognition, Heraklion, 2014, pp. 9-14.

K. Zagoris, I. Pratikakis, and B. Gatos, "A framework for efficient transcription of historical documents using keyword spotting," in Historical Document Imaging and Processing (HIP'15), 3rd International Workshop on, August 2015, pp. 9-14.

K. Zagoris, I. Pratikakis, B. Gatos. 2017 Unsupervised Word Spotting in Historical Handwritten Document Images using Document-oriented Local Features. Transactions on Image Processing. Under Review.

#### Transkribus

Günter Mühlberger, 01 Presentation: Transkribus - the status quo and future plans, Transkribus User Conference, Nov. 2-3, 2017, Youtube video, pub. Jan 23, 2018.

[https://www.youtube.com/watch?v=7V8iXFzx\\_E&feature=youtu.be](https://www.youtube.com/watch?v=7V8iXFzx_E&feature=youtu.be), accessed 12/02/2018

02 Presentation: Transkribus in practice - Reports from Users, Transkribus User Conference, Nov. 2-3, 2017, Youtube video.

<https://www.youtube.com/watch?v=qnY3i6d4lHw&feature=youtu.be>, accessed 12/02/2018

03 Presentation: Transkribus keyword searching and indexing, Transkribus User Conference, Nov. 2-3, 2017, Youtube video.

<https://www.youtube.com/watch?v=APpugG4Yzso&feature=youtu.be>, accessed 12/02/2018

04 Presentation: More reports from users: first part, Transkribus User Conference, Nov. 2-3, 2017, Youtube video. <https://www.youtube.com/watch?v=Ih4mGsXT7NU&feature=youtu.be>, accessed 12/02/2018

04 Presentation: More reports from users: second part, Transkribus User Conference, Nov. 2-3, 2017, Youtube video. <https://www.youtube.com/watch?v=QE9jeupnK4Y&feature=youtu.be>, accessed 12/02/2018

05 Presentation: Panel discussion, Transkribus User Conference, Nov. 2-3, 2017, Youtube video. <https://www.youtube.com/watch?v=KJ-cp3u2qr4&feature=youtu.be>, accessed 12/02/2018

### Github

Transkribus/VCG-DUTH-Word\_Spotting\_By\_Example, Github, 2018. <https://github.com/Transkribus/VCG-DUTH-Word-Spotting-by-Example>, accessed 12/02/2018

### **Considerable archival interest and technical progress in development of manuscript document layout recognition and handwriting text recognition**

Alex Hailey (Curator, Modern Archives and Manuscripts), 'Using Transkribus for handwritten text recognition with the India Office Records', British Library, Digital scholarship blog, Jan. 23rd, 2018. <http://blogs.bl.uk/digital-scholarship/2018/01/using-transkribus-with-the-india-office-records.html>, accessed 12/02/2018

Nora McGregor (Digital Curator, British Library), '8th Century Arabic science meets today's computer science, Or, Announcing a Competition for the Automatic Transcription of Historical Arabic Scientific Manuscripts', British Library Digital Scholarship blog, Feb. 8th, 2018. <http://blogs.bl.uk/digital-scholarship/2018/02/8th-century-arabic-scientists-meet-todays-computer-scientists.html>, accessed 12/02/2018

FTNA Postdoctoral Research Fellowship (Datafication); Six month National Archives fellowship to explore crowdsourcing, user experience design, the use of automated transcription technology, entity extraction, natural language processing and data visualisation, applied to PROB 11, a collection of over 1 million digitised Prerogative Court of Canterbury wills, 1384-1858, held at TNA. <https://www.civilservicejobs.service.gov.uk/csr/jobs.cgi?jcode=1572699>, accessed 12/02/2018

Digital Methods Workshop series 2017-18: Beyond words: challenges in reading historical document collections at scale. Workshop 2: Technical challenges. Speakers: John Sheridan, Digital Director, TNA; Daniel Bruder, Cambridge Computer Laboratory; Prof. Jane Winters, School of Advanced Study, University of London. Feb 6, 2018. <https://www.eventbrite.co.uk/e/beyond-words-workshop-2-technical-challenges-tickets-41451741274>, accessed 12/02/2018

## **2: Dataset and data access**

A) What data can you make available to challenge participants? [detail per dataset: data inventory, size, variable descriptions, description of data collection mechanism, level of data sensitivity/confidentiality, data owner etc.]

### Data sets

- Digital images [JPEG]
  - 4,000-6,000
  - Containing ca. 4,000 signatures and/or marks & 4,000 groups of metadata in the image (name, occupation, place of residence, age, date of witness statement)
- Marked-up images [JPEG/Transkribus]
  - 4,000-6,000
  - Text Regions marked up with Transkribus tools for both signature/marks and metadata
  - Mark-up in XML
- Full text transcriptions [In Semantic Media Wiki forms]
  - 6,000 pages [ca. 4 million words]
- Marked-up transcriptions identifying pages with signatures and/or marks [in Semantic Media Wiki forms]

### Confidentiality

- [CC BY 3.0](#) licence
- Permission by MarineLives and Transkribus to further process Transkribus modified and marked-up digital images

### Dataowner

- MarineLives (with data transformed with the assistance of Transkribus tools)

B) What are the most frequent data quality issues in the data, based on your previous experience? [if you have never looked at data like this before, please state that.]

### Image clarity

- Images are [JPEG](#); size range = 2 MB to 7 MB.
- Uploaded into [MediaWiki](#) and made available within the MarineLives [Semantic Media Wiki](#) in customised [semantic page forms](#) to allow detailed inspections of regions within image, such as signatures and marks at end of witness statements, and name, occupation, place of residence, age and date of witness statement at beginning of such statements.
- 98% of images are high quality and sufficient for reliable HTR (as tested with Transkribus).

### Transcription accuracy

- Audit history exists for date of creation of page holding image and transcription, and also editorial history.
- Manuscript text is transcribed using a semi-diplomatic editorial policy – this expands contractions and abbreviations, but leaves capitalisation, punctuation and spelling as seen in the manuscript original, without modernisation.
- 85% of transcriptions are high quality, with perhaps two to four character errors in a 2500 character page, but with a somewhat higher rate of capitalisation errors



- 95% of transcriptions of metadata at the start of a witness statement (name, occupation, place of residence, age, date of witness statement) are of high quality, with very few character or capitalisation errors.
- Transcriptions of signatures are somewhat lower quality, due to the range of handwriting styles used in signatures. We will redo all transcriptions of signatures prior to the proposed April DSG.

C) If there are any issues listed above, in (B), that are not present in the data described in (A), please flag this here and explain why.

N/A

D) What is the proposed mechanism for data access for DSG researchers? [please describe the complete process in concrete steps.]

#### **Data set definitions and documentation**

- We will provide detailed documentation of all data sets offered to DSG researchers, including definition and contextualisation of terms.

#### **Data set support**

- We will provide expert support from MarineLives project team members to answer queries about data sets, data terms and their contextualisation.

#### **Palaeographic queries and manual classification of signatures and marks**

- Our [active use of social media](#), including a large [Twitter followership](#) and [wiki user base](#), amongst academic historians and palaeographers (and the general, but sophisticated, public) enables us to mobilise experts at short notice to answer specific questions.
- Should DSG researchers wish to explore a feedback loop between machine based pattern recognition and classification of these patterns, and human based manual pattern recognition and classification, we could set something up to run in parallel with the April DSG. For example, online conjoint analysis by volunteers between pairs of signatures and/or marks, selected at random, or selected for close similarity of (say) technical quality of execution of a signature. This could be used as further input to a machine learning based approach to an analysis of the technical execution of signatures and marks.

#### **SMW API**

- We will give all DSG researchers usernames and passwords to access the MarineLives SMW.
- We will give one or more DSG researchers systems op status for the MarineLives wiki, if this is required.
- We will give DSG researchers remote access to the designer and coder of our semantic media wiki, should there be questions about design aspects of the wiki,

operational aspects of the SMW API, or about the technical characteristics of our data sets.

#### **Transkribus access**

- We will arrange for all DSG researchers to have access to Transkribus and MarineLives/Transkribus data, either through our existing Transkribus username and password, or through supplementary usernames and passwords, which will be issued by Transkribus.

### **3: Impact**

A) Assuming a solid, state-of-art solution to the science challenge, what is the least possible business/operational/societal impact? [ie what would be the minimal achievement of a successful challenge?]

The minimum achievement would be to address section 1 and 3 of the challenge, ideally in the affirmative.

We would propose using the minimum achievement results in the following ways:

- Publicise in conference presentation and potential journal article (with the Turing Institute as lead author).
- Repackaging the data as a reusable data set, to enable other AI and data science groups to attempt to replicate or improve the results.
- Strengthening our technical credibility (through successful work with the Turing) as we continue to develop our relationships with the Huntington Library, the British Library, and the National Archives.
- Strengthening our broader credibility with potential financial donors and technical supporters for our planned educational charity, Chronoscopic Education.

B) Assuming a solid, state-of-art solution to the science challenge, what is the best possible business/operational/societal impact? [ie what would be the maximal achievement of a successful challenge?]

The maximal achievement would be to address section 1, 2 and 3 in the affirmative

We would propose using the maximal achievement in the following ways:

- Same as in the minimal achievement, but definitely a journal article.
- Greater impact on our credibility with potential archival partners, technical partners and financial donors to support our planned educational charity.
- An increased ability to attract technically proficient and interested volunteers to work with MarineLives and with Chronoscopic Education.

C) What is the anticipated scope for internal follow-up on this work within your organisation?

### **Internal follow-up**

(1) Expand signature/marke data set

- in collaboration with TNA
- in collaboration with Huntington Library
- in collaboration with British Library

(2) Use as input to definition of key word spotting project later in 2018

-in collaboration with Transkribus

(3) Test Turing generated algorithms against expert historian input and against general public input

D) What is the anticipated scope for collaborative follow-up, together with the Turing?

### **Potential collaborative follow-up with the Turing**

Collaboration is based upon defining interests of mutual benefit and on trust. It takes time to do both, and is built largely at an individual level, though within an institutional framework. Our volunteers, supporters and advisors see the impact and opportunity of technological innovation on the academic and private study of the humanities. History is no exception. Technological drivers include digital imaging, both of manuscripts and books; semantic markup and linkage technologies, such as TEI and linked data; and the availability of low cost large scale open databases containing images, metadata and full text, hosted by libraries, archives, and university research teams.

We have developed a vision for MarineLives and Chronoscopic Education, which will help historians and archivists explore the impact of this technological change on their personal and institutional work processes. We need sophisticated technical partners at the individual and institutional level to help us, and we are open to exploring how that might take place, whether around specific technologies, or methodologies, or specific projects.

We would be interested in potential collaboration with the Data Science and Digital Humanities group at the Turing Institute.