What should be in your Digital Toolbox? #digtoolbox

The Linnean Society of London, in collaboration with the Transcribe Bentham initiative at University College London (UCL), hosts a one-day conference on the 10th of October 2016 to showcase how innovative technology is being applied to the humanities and natural sciences.

Programme:

9:30 – 10:00	**Optional tour of Linnaean Collections**	Elaine Charwat (Linnean Society)	
10:00 – 10:30	**Registration & Tea**		
10:30 – 10:40	**Welcome**	Elaine Charwat (Linnean Society) and Dr Louise Seaward (University College London)	
10:40 – 12:30	**Session 1 (Chair: Professor Philip Schofield, University College London)**	Professor Melissa Terras (University College London) 10:40-11:30	*If you teach a computer to READ: Transcribe Bentham, Transkribus, and Handwriting Technology Recognition*
		Dr Günter Mühlberger (University of Innsbruck) 11:30-12:00	*Transkribus as a toolkit for text recognition, transcription and information extraction*
		Professor Roger Labahn (University of Rostock) 12:00-12:30	*Key concepts of Handwritten Text Recognition*
12:30 – 13:30	**Lunch**		

13:30-15:00	**Session 2 (Chair: Dr Louise Seaward, University College London)**	Dr Ulrich Tiedau (University College London) 13:30-14:00	*Asymmetrical Encounters: A digital quantitative approach to the history of mentalities in Europe, 1800–2000*
		Dr Mia Ridge (The British Library) 14:00-14:30	*The Art of Work in the Age of Mechanical Reproduction*
		Professor James Loxley (University of Edinburgh) 14:30-15:00	*Lines of Enquiry: Reordering Edinburgh's Literary History*
15:00-15:30	**Tea break**		
15:30-17:00	**Session 3 (Chair: Elaine Charwat, Linnean Society)**	Dr Elspeth Haston (Royal Botanic Garden Edinburgh) 15:30-16:00	*Automating label data capture from natural history specimens*
		Alison Harding & Lisa Cardy (Natural History Museum/Biodiversity Heritage Library) 16:00-16:30	*Unlocking biodiversity data @ the Biodiversity Heritage Library*
		Dr Victoria Van Hyning (University of Oxford/Zooniverse) 16:30-17:00	*Metadata Extraction and Full Text Transcription on the Zooniverse Platform*
17:00-17:10	**Comfort break**		
17:10-17:30	**Closing words**	Professor Melissa Terras (University College London)	
17:30-18:30	**Reception**	Linnean Society Library	

Abstracts

Professor Melissa Terras (University College London)
If you teach a computer to READ: Transcribe Bentham, Transkribus, and Handwriting Technology Recognition

For the past six years, the Transcribe Bentham project has been generating high quality crowdsourced transcripts of the writings of the philosopher and jurist Jeremy Bentham (1748-1832), held at University College London, and latterly, the British Library. Now with nearly 6 million words transcribed by volunteers, little did we know at the outset that this project would provide an ideal, quality controlled dataset to provide "ground truth" for the development of Handwriting Technology Recognition. This paper will look at the past, present and future of automated handwriting analysis for documents, showing how our research on the EU framework 7 Transcriptorium, and now H2020 READ projects, is working towards a service to improve the searching and analysis of digitised manuscript collections across Europe.

Dr Günter Mühlberger (University of Innsbruck)
Transkribus as a toolkit for text recognition, transcription and information extraction

Working with digitised historical documents is a daily task for many who work in the humanities sector. The Transkribus platform (part of the H2020 READ project) is designed to offer core services so scholars can benefit from all the advantages of the "digital turn". With Transkribus, text recognition of handwritten documents, keyword spotting and information extraction becomes available for researchers.

Professor Roger Labahn (University of Rostock)
Key concepts of Handwritten Text Recognition

Automatic text recognition is increasingly becoming an essential core component of application software in the Digital Humanities. After the "classical" OCR (Optical Character Recognition) for printed texts, we now see various successful expansions to handwritings. As this requires essentially new paradigms in algorithms and technology, we increasingly use the term HTR (Handwritten Text Recognition): Rather than processing single characters, entire sequences have to be considered, e.g. words, lines, whole text blocks or even entire pages.

We present a rather general survey on some foundations of the new approaches and explain little more details of selected basic algorithms of contemporary HTR technology. The focus is on the Recurrent Neural Network engine and to understand the fundamental decoding idea, i.e. how to come from the network's magic output to meaningful recognition results.

Finally, we show realistic examples in two main application areas: transcription of and keyword spotting in historical writings. These are results of the long-term collaboration between our research group Computational Intelligence Technology Lab (*http://www.citlab.uni-rostock.de*) with the partner enterprise PLANET intelligent systems GmbH (*http://www.planet.de*).

Dr Ulrich Tiedau (University College London)
Asymmetrical Encounters: A digital quantitative approach to the history of mentalities in Europe, 1800–2000

This short paper reports on the progress of a large, European-funded digital history research project that uses text- and sentiment-mining in long runs of large historical newspapers to investigate the question of how so-called reference cultures over time have contributed to the emergence of a common European cultural identity. In combining multilingual and transnational text-mining with literary and historical interpretative scholarship the project seeks to develop a quantitative approach to the history of mentalities in Europe.

Dr Mia Ridge (The British Library)
The Art of Work in the Age of Mechanical Reproduction

The British Library's collections are vast - an estimated 180 to 200 million collection items spanning 3000 years of global history - and hugely varied, ranging from the world's oldest printed book to websites from the 2015 UK General Election. While no more than 2% of the collection is (yet) digitised, working at this scale creates unique challenges. This talk will discuss how British Library's Digital Research team supports the creation and innovative uses of the Library's digital/digitised collections.

Professor James Loxley (University of Edinburgh)
Lines of Enquiry: Reordering Edinburgh's Literary History

Among its other attributes, Edinburgh is famed as a literary city - it hosts the largest book festival in the world, and was made the first UNESCO World City of Literature in 2004. That literary history is familiar to a wide range of readers, and has been actively cultivated for more than a century. But this cultivation often proceeds along familiar lines, repeating or rescoring an already kenspeckle narrative. The Palimpsest project set out to shake up this history through the digital exploration of large corpora of published texts. This paper will explore some of the challenges the project faced, the lessons learned, and future developments to which it might point.

Dr Elspeth Haston (Royal Botanic Garden Edinburgh)
Automating label data capture from natural history specimens

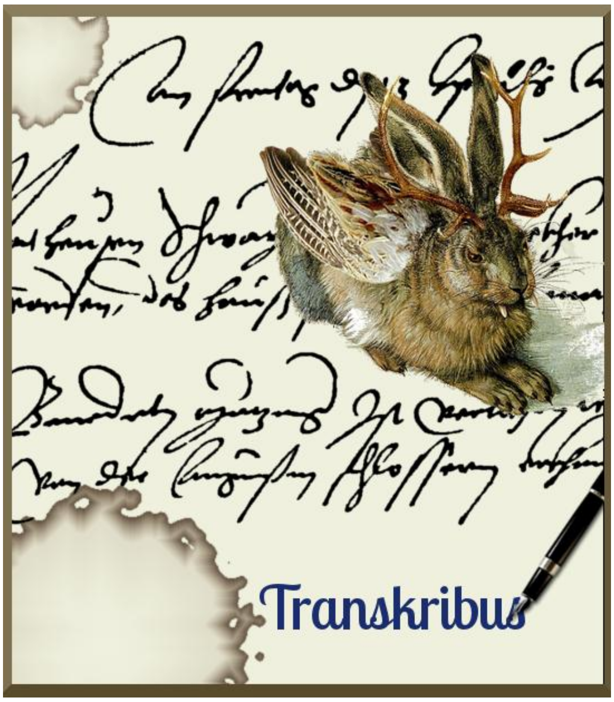
The Herbarium of the Royal Botanic Garden Edinburgh holds over 3 million preserved plant and fungi specimens, collected over a period of more than 300 years and representing nearly 2/3 of the world's flora. The need for the development and implementation of tools to speed up the process of data capture from natural history specimens is generally recognised. We have therefore been developing more automated tools and workflows, including developing the use of Optical Character Recognition (OCR) within the digitisation pipeline. More recently, we have been working with a number of other European Natural History institutes within the SYNTHESYS3 project. This project, funded by the European Union, aims to produce an accessible, integrated European resource of natural history specimens for researchers. As part of this goal, we have been identifying and testing software which utilise OCR and Handwritten Text Recognition (HTR).

Alison Harding & Lisa Cardy (Natural History Museum/Biodiversity Heritage Library)
Unlocking biodiversity data @ the Biodiversity Heritage Library

The Biodiversity Heritage Library (BHL) is formed by a consortium of natural history and botanical libraries who work collaboratively to make biodiversity literature available as part of a global community. BHL (http://www.biodiversitylibrary.org/) provides scientists, scholars and the public free and open access to over 50 million pages of digitised text and grey literature on biodiversity. BHL is mature and sustainable as a virtual organisation and digital repository. BHL supports open science and wider cultural uses by continuing to add relevant content, curating that content and enhancing access through the development and application of innovative tools. The availability and reusability of the scientific data accessible via BHL ranges from taxonomic study and training, biodiversity research, conservation and maintenance of diverse ecosystems, animal and plant disease control through to audiences beyond core science constituencies, including historical and cultural research on science, exploration and global commerce.

Dr Victoria Van Hyning (Zooniverse)
Metadata Extraction and Full Text Transcription on the Zooniverse Platform

This talk will describe the development of Science Gossip and Shakespeare's World, a metadata extraction and a full text manuscript project, respectively. It will explain how the Zooniverse method of showing the same image or task to multiple individuals, and then aggregating their responses has produced data for each project, and the successes and challenges of this method and the resulting data.



Speakers' Biographies

Session 1 (Chair: Professor Philip Schofield, University College London)

Professor Melissa Terras (University College London)

Melissa Terras is Director of UCL Centre for Digital Humanities, Professor of Digital Humanities in UCL's Department of Information Studies, and Vice Dean of Research in UCL's Faculty of Arts and Humanities. Publications include "Image to Interpretation: Intelligent Systems to Aid Historians in the Reading of the Vindolanda Texts" (2006, Oxford University Press) and "Digital Images for the Information Professional" (2008, Ashgate) and she has co-edited various volumes such as "Digital Humanities in Practice" (Facet 2012) and "Defining Digital Humanities: A Reader" (Ashgate 2013). She is currently serving on the Board of Curators of the University of Oxford Libraries, and the Board of the National Library of Scotland, and is a Fellow of the Chartered Institute of Library and Information Professionals and Fellow of the British Computer Society. Her research focuses on the use of computational techniques to enable research in the arts and humanities that would otherwise be impossible. You can generally find her on twitter @melissaterras.

Dr Günter Mühlberger (University of Innsbruck)

Günter Mühlberger works as Head of the Digitisation and Digital Preservation group at the Department for German language and Literature at the University of Innsbruck. He focuses on building up services and tools for the Digital Humanities. He has successfully coordinated several national and international projects, e.g. METADATA ENGINE (structural metadata extraction and OCR for gothic letters, 2000-2003), Digitisation on Demand/ eBooks on Demand (DoD, EOD, 2006-2012) and Europeana Newspapers (member of the executive board, OCR processing and enrichment of newspapers). He is now coordinator of the READ project.

Professor Roger Labahn

Prof. Dr. Roger Labahn received his doctoral degree in 1987 and finished his habilitation in 1994, both in Discrete Mathematics / Combinatorics. Since then he has been Senior Researcher with academic teaching duties in various fields of basic, applied and discrete Mathematics. Meanwhile, his major interest and working area changed to Machine Learning and Neural Networks with strong focus on both application oriented research and algorithm design as well as software development. Currently he leads the Computational Intelligence Technology Lab (CITlab) in the Mathematical Department of the University of Rostock. Over the last years, this group achieved international acceptance for developing algorithms and technologies for handwritten text recognition based on state-of-the-art concepts of Computational Intelligence. CITlab won several major competitions in that area and is one of the main technology partners of the Horizon-2020 READ project.

Session 2 (Chair: Dr Louise Seaward, University College London)

Dr Ulrich Tiedau (University College London)

Ulrich Tiedau is a Senior Lecturer in modern Low Countries history and society at the Department of Dutch and an Associate Director of the Centre for Digital Humanities at University College London.

Dr Mia Ridge (The British Library)

Mia's PhD in digital humanities (Department of History, Open University) was titled 'Making digital history: The impact of digitality on public participation and scholarly practices in historical research'. Formerly Lead Web Developer at the Science Museum Group, Mia has worked internationally as a business analyst, digital consultant and web programmer in the cultural heritage and commercial sectors. Mia has held international fellowships at Trinity College Dublin/CENDARI, Ireland (2014), the Polis Center Institute on 'Spatial Narrative and Deep Maps' (2012) and the Roy Rosenzweig Center for History and New Media One Week One Tool program (2013), and had short Residencies at the Powerhouse Museum (2012) and the Cooper-Hewitt Design Museum (2012). Mia has post-graduate qualifications in software development and an MSc in Human-Centred Systems. She is Chair of the Museums Computer Group (MCG) and a member of the Executive Council of the Association for Computers and the Humanities (ACH).

Professor James Loxley (University of Edinburgh)

James Loxley is Professor of Early Modern Literature at the University of Edinburgh, and the lead investigator on the Palimpsest project. He has published widely on the writing of the early modern period, on some aspects of critical theory, and led a number of collaborative interdisciplinary research projects.

Session 3 (Chair: Elaine Charwat, Linnean Society)

Dr Elspeth Haston (Royal Botanic Garden Edinburgh)

Dr Haston is Deputy Herbarium Curator at Royal Botanic Garden Edinburgh (RBGE). She is interested in the tools, processes and workflows for digitising herbarium specimens and related images and documents, and together with colleagues at RBGE and other institutes, is developing methodologies for more rapid, large-scale digitisation of herbarium specimens, including the use of Optical Character Recognition (OCR). She is interested in curation and the integration of curation and digitisation in herbaria, as well as the development of tools, processes and workflows to aid with curation in herbaria. She has carried out systematic research on the *Peltophorum* group (Leguminosae) based on molecular and morphological data. She has also undertaken research on floral development in the Gesneriaceae, with particular emphasis on the genera *Saintpaulia* and *Streptocarpus*.

Alison Harding and Lisa Cardy (Natural History Museum/Biodiversity Heritage Library)

Alison Harding is Researcher Services Librarian at the Natural History Museum. She is a qualified librarian with over 25 years' professional experience having worked at Shell, for Clive Sinclair, Wellcome Research Laboratories and Ashridge Management College. She is responsible for the Biodiversity Heritage Library workflow, administration of the institutional repository, and the day to day management of the collections held at Tring.

Lisa Cardy is a qualified librarian with over 20 years professional experience working at the British Library, the British Medical Association Library and The Guardian Newspaper Library. She is currently Head of Researcher Services and Digital Delivery in the Library and Archives at the Natural History Museum, enabling access to the library and archive collections, delivering training and support to discover and access the collections, as well as developing the systems and infrastructure to enable discovery and access. Before joining the Museum she was Access and Procurement Manager at the London School of Economics Library and volunteered with the MicroPasts project (http://micropasts.org/) and the Wellcome Library.

Dr Victoria Van Hyning (University of Oxford/Zooniverse)

Dr Victoria Van Hyning is a British Academy Postdoctoral Fellow at the University of Oxford, Pembroke College, and the Humanities PI of Zooniverse.org, a crowdsourcing research group based at Oxford, University of Minnesota and the Adler Planetarium in Chicago. She developed Science Gossip, Shakespeare's World and AnnoTate. Her monograph, *Convent Autobiography: English Nuns in Exile, 1609–1807* is forthcoming with OUP.

About the organisers:

The Recognition and Enrichment of Archival Documents (READ) project

It can be complex to work with handwritten historical documents of varying styles, languages, layouts and legibility. But technological advances are making it possible for computers to process handwritten material. The Recognition and Enrichment of Archival Documents (READ) project is an EU-funded collaboration between 14 partners drawn from the domains of computer science, archives and humanities. READ is setting up a Virtual Research Environment and is building tools capable of reading handwritten manuscripts. These technologies are made available through the Transkribus transcription platform.

To find out more about the READ project visit: http://read.transkribus.eu/

To download Transkribus visit: https://transkribus.eu/Transkribus/

Follow the READ project on Twitter: @Transkribus ; Contact us at: email@transkribus.eu

----

Transcribe Bentham

Transcribe Bentham offers volunteers the chance to explore and transcribe the vast archive of manuscripts written by the philosopher and reformer, Jeremy Bentham (1748-1832). Transcripts produced by volunteers are uploaded to an open access digital repository and used by UCL's Bentham Project in its work on the scholarly edition of Bentham's writings. Volunteers have the opportunity to make new discoveries and contribute to research – why not give it a try?

To find out more about Transcribe Bentham visit: http://blogs.ucl.ac.uk/transcribe-bentham/

To start transcribing Bentham's manuscripts, visit our Transcription Desk:

http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham

Follow Transcribe Bentham on Twitter: @TranscriBentham

Contact us at: transcribe.bentham@ucl.ac.uk

The Linnean Society of London

The Linnean Society of London is the world's oldest active biological society. Founded in 1788, the Society takes its name from the Swedish naturalist Carl Linnaeus (1707–1778) whose botanical, zoological and library collections have been in its keeping since 1829. As it moves into its third century the Society continues to play a central role in the documentation of the world's flora and fauna – as Linnaeus himself did – recognising the continuing importance of such work to biodiversity conservation.

To find out more about the Society, activities and events visit: https://www.linnean.org/

View and research the Society's Linnaean and other important collections online:

http://linnean-online.org/

Follow the Linnean Society on Twitter: @LinneanSociety ; Contact us at: info@linnean.org