

# CULTURAL HERITAGE DATA SCHOOL



**MONDAY 17TH - TUESDAY 18TH APRIL 2023**

Cambridge Digital Humanities  
The University of Cambridge



**UNIVERSITY OF  
CAMBRIDGE**

# CAMBRIDGE DATA SCHOOL CULTURAL HERITAGE APRIL '23



## WELCOME!

### Welcome to the April 2023 Cambridge Cultural Heritage Data School!

We are so excited to be offering our prestigious Data School in this new, compact, two-day format, and to be welcoming participants back to Cambridge for our first in-person school since the Coronavirus Pandemic.

In this pack, you will find all the information you need for the two-day Data School, including:

- Important information
- Key contacts
- Map with key sites
- Timetable
- Module outline
- Teachers' bios

## KEY CONTACTS



### DR IRVING HUERTA

*Data School Course Convenor*

Email: [rih32@cam.ac.uk](mailto:rih32@cam.ac.uk)

**For all course enquires regarding content and curriculum, including technology, contact Irving.**



### FLORENCE HARRY

*Data School Course Assistant*

Email: [dataschool@cdh.cam.ac.uk](mailto:dataschool@cdh.cam.ac.uk)

**For all admin enquiries including logistics, payment, and planning, contact Florence.**

# CAMBRIDGE DATA SCHOOL CULTURAL HERITAGE APRIL '23



## BEFORE YOU ARRIVE

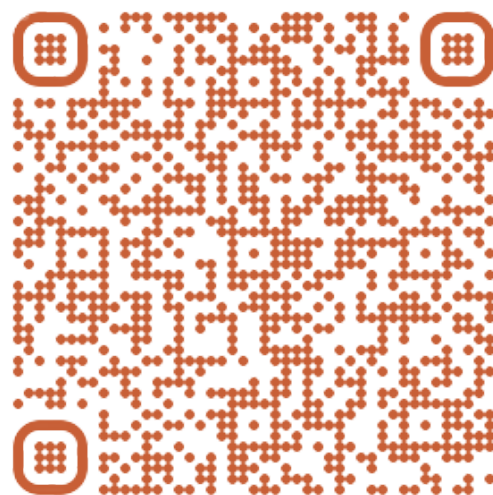
**Before you arrive, please make sure:**

- You have **booked transport** to and from Cambridge, **and accommodation** for the duration of the event.
- You have **read and understood the T&Cs**, including our cancellation policy
- You have taken out **travel insurance** and, if from abroad, **health insurance** (both highly recommended).
- You are bringing a laptop and have **downloaded appropriate software** (see below).

***This will ensure you have a great experience and the Data School runs smoothly!***

## PREPARING FOR THE DATA SCHOOL

For the Data School, you will need to access some specific software on your personal laptop. Please follow the QR code for technical instructions and requirements in advance of the Data School. You can email us if you have any issues. Thanks!



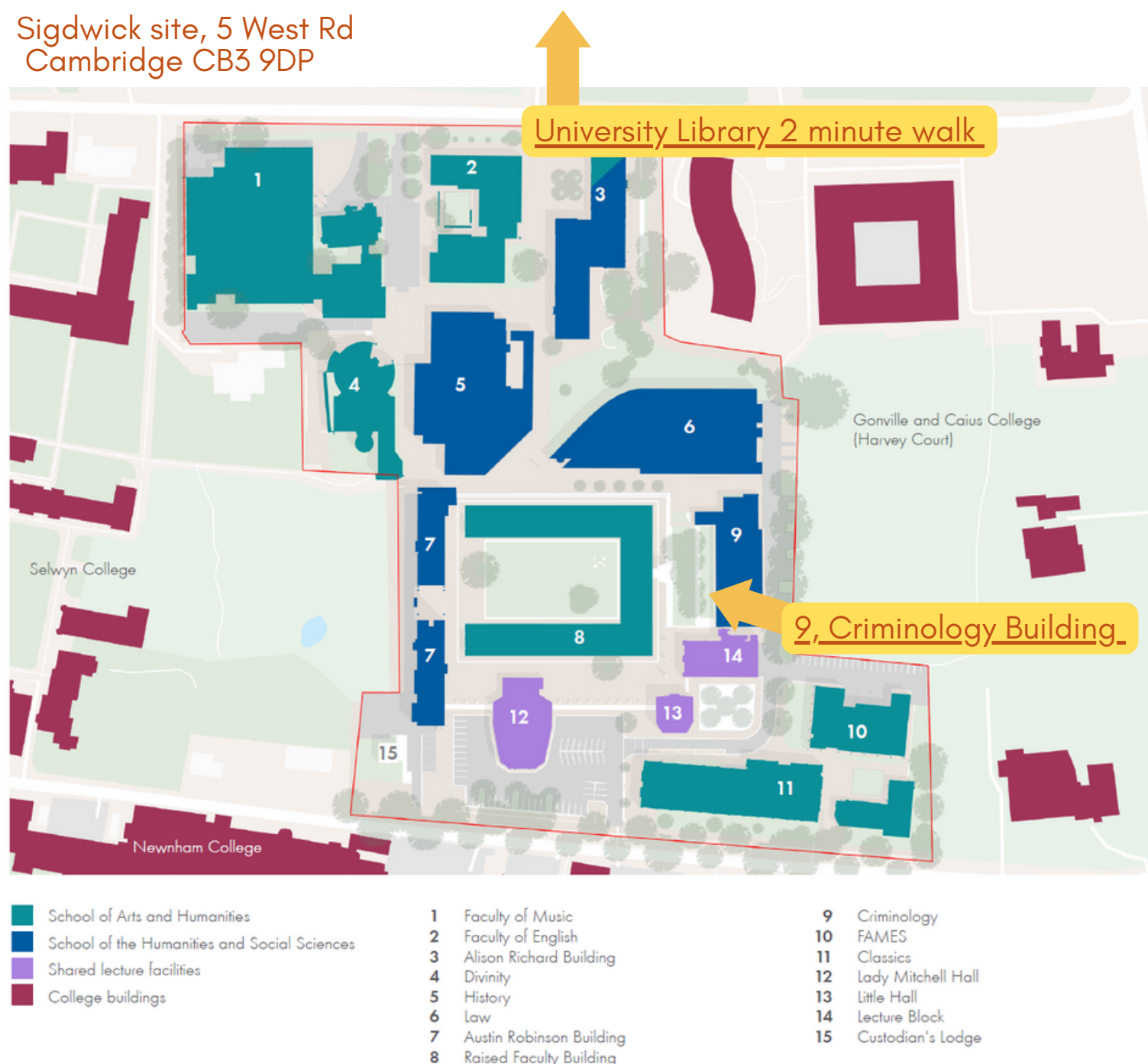
**SOFTWARE**

# CULTURAL HERITAGE DATA SCHOOL, APRIL '23

C·D·H

## MAPS AND LOCATIONS

Sigdwick site, 5 West Rd  
Cambridge CB3 9DP



Main locations for the Data School: Criminology Building & University Library

If you have any concerns about accessibility or any other questions, please get in touch so we can assist!

# TIMETABLE



DAY 1 | MONDAY 17 APRIL

| TIME          | SESSION  | LOCATION             | TEACHER                        |
|---------------|--|----------------------|--------------------------------|
| 9:30–10:00am  | Introduction and welcome                           | Criminology Building |                                |
| 10:15–10:45am | Meeting the Manuscripts                            | Library              | Suzanne Paul                   |
| 10:45–12:30am | Principles of GLAM Imaging (DCU)                   | Library              | Maciej Pawlikowski             |
| 12:30–1:30pm  | Lunch (provided)                                   |                      |                                |
| 1:30–2:30pm   | Workshop: Working with digital images at scale     | Criminology Building | Anne Alexander                 |
| 2:30–2:45pm   | Break  |                      |                                |
| 2:45–3:45pm   | Workshop: Machine vision – a critical introduction | Criminology Building | Anne Alexander                 |
| 3:45–4pm      | Lightning Talk: The Constable Project              | Criminology Building | Caroline Bassett/Andy Corrigan |
| 4–5pm         | Digital Spaces: Interoperability and Interaction   | Criminology Building | Andy Corrigan                  |
| 6:30pm        | Dinner (paid individually)                         | Restaurant in Town   | Irving, Anne, Florence         |

# TIMETABLE



DAY 2 | TUESDAY 18 APRIL

| TIME          | SESSION  | LOCATION             | TEACHER                                 |
|---------------|--|----------------------|---|
| 9:30-9:45am   | Lightning Talk: Image and Text Data in Research Projects at the Fitzwilliam Museum | Criminology Building | Alexandra Gushurst-Moore                |
| 9:45-10:30am  | Round Table with colleagues from the Fitz Research Team                            | Criminology Building | Alexandra GM and the Fitz Research Team |
| 10:30-11:30am | Digital Research Design and the Project Lifecycle                                  | Criminology Building | Anne Alexander                          |
| 11:30-11:45am | Break  |                      |   |
| 11:45-12:45pm | Collecting and Wrangling Textual Data  | Criminology Building | Huw Jones                               |
| 12:45-1:30pm  | Lunch  |                      |   |
| 1:30-2:30pm   | Text Analysis and Visualisation I  | Criminology Building | Andrea Kocsis                           |
| 2:30-2:45pm   | Break  |                      |   |
| 2:45-3:45pm   | Text Analysis and Visualisation II   | Criminology Building | Andrea Kocsis                           |
| 3:45-4:15pm   | Closing Plenary  | Criminology Building |   |

# CULTURAL HERITAGE DATA SCHOOL, APRIL '23



## MODULE OUTLINES

### **PRINCIPLES OF GLAM IMAGING** *With **Maciej Pawlikowski** (Head of the Digital Content Unit at the Cambridge University Library)*

This session will cover learning about principles of archival imaging standards and practical approach to taking images fit for project purpose. It will provide the participants with the basic vocabulary and understanding of methodological approach to digitisation applicable to any project. It may also address any more advanced imaging topics such as image stitching, Optical Character Recognition, Multispectral Imaging or Photogrammetry if these are in the interest of the participants. Students will visit the Digital Content Unit at Cambridge University Library.

### **WORKING WITH DIGITAL IMAGES AT SCALE (WORKSHOP)** *With **Dr Anne Alexander** (Cambridge Digital Humanities)*

This session introduces the IIIF image data framework, which has been developed by a consortium of the world's leading research libraries and image repositories and demonstrates a range of different machine learning- based methods for exploring digital image collections. Students will have the opportunity to try ImageGraph, a simple visual search engine developed by Leo Impett, convenor of CDH's MPhil.

### **MACHINE VISION: A CRITICAL INTRODUCTION (WORKSHOP)**

*With Dr Anne Alexander*

Machine vision systems can potentially help humanities researchers see historical and cultural image collections differently, and could provide tools to answer new research questions. This session provides an introductory overview of basic tasks in machine vision, such as Image Classification, Object Detection and Image Captioning within a critical framework highlighting the challenges of algorithmic bias and the limits of automation as a method for humanistic enquiry.

# CULTURAL HERITAGE DATA SCHOOL, APRIL '23



## MODULE OUTLINES CONTINUED

### DIGITAL RESEARCH DESIGN AND THE PROJECT LIFECYCLE

*With Dr Anne Alexander*

This introductory module explores the lifecycle of a digital research project across the stages of design – data capture, transformation, analysis, presentation and preservation. It also introduces tactics for embedding ethical research principles and practices at each stage of the research process. We will discuss the importance of documentation of data provenance, look at the practical and ethical challenge of common methods used for bulk data capture including use of APIs and working with data collected by others. The second session in the module will introduce the data-cleaning tool OpenRefine and a set of exercises for participants to work through in their own time.

**COLLECTING AND WRANGLING TEXTUAL DATA** *With Huw Jones (Cambridge Digital Library and CDH).*

**MORE INFO COMING SOON.**

**TEXT ANALYSIS AND VISUALISATION I & II** *with Dr Andrea Kocsis (Assistant Professor in History and Data Science at Northeastern University London & CDH Methods Fellow)*

**MORE INFO COMING SOON.**

### DIGITAL SPACES: INTEROPERABILITY AND INTERACTION

*With Andy Corrigan*

**MORE INFO COMING SOON.**

*Programme subject to ammendment.*

**FOR LATEST UPDATES REGARDING  
MODULES, FOLLOW QR CODE.**



# TEACHER BIOS

---



## DR ANNE ALEXANDER

*Data School Course Convenor*

Dr Anne Alexander has been Director of Learning at CDH since its foundation. She was previously Co-ordinator of the Cambridge Digital Humanities Network. Her research interests include ethics of big data, activist media in the Middle East and the political economy of the Internet. She is a member of the Data Ethics Group and the Humanities and Data Science Special Interest Group at the Alan Turing Institute



## DR IRVING HUERTA

*Data School Course Convenor*

Irving Huerta is a Research Associate and Data School Convenor. His background is in journalism, collaborating with organisations like Forensic Architecture, the International Consortium for Investigative Journalism and others. He is interested in the intersection between politics, media, and accountability. His research revolves around the politics of uncovering wrongdoing (i.e. the struggles of its practitioners, the production of truth, and the transformation of reality). Irving holds a PhD in Politics from Goldsmiths University and has taught Political Communication and Investigative Journalism in universities and civil organisations such as the Centre for Investigative Journalism in the UK and Fundación Gabo in Colombia.



## HUW JONES, HEAD OF DIGITAL LIBRARY UNIT

*Session: Collecting and Wrangling Textual Data*

Huw is the Head of the Digital Library Unit, Digital Humanities Coordinator and the first point of contact with many DH research projects through CDH Labs. He has a broad range of experience in content-driven DH work, including collaborations with the Newton Project, the Darwin Correspondence Project, the Casebooks Project and many others. He has a particular focus on text encoding, and co-convenes the TEI strand of the Oxford Digital Humanities Summer School.



# TEACHERS' BIOS

---



## **ANDY CORRIGAN, CAMBRIDGE DIGITAL LIBRARY CO-ORDINATOR**

*Session: Digital Spaces: Interoperability and Interaction*

Andy Corrigan has a background in archaeology and specialises in graphics, illustration, historic building surveys and the application of photography. Andy began working at Cambridge University Library on both the Darwin Correspondence Project and Cambridge Digital Library in 2011 and has developed his experience in public and academic engagement, complex image preparation, digital asset management, and the workflow of digital content to both digital and print publication.

In 2016, he took on his current role as the Cambridge Digital Library Co-ordinator, focusing on engagement with digital content, such as using the Library's digital collections in teaching and research and its analysis. Andy has a particular interest in the matrix of technology, resources, research and pedagogy and filling the gaps between them.



## **DR ANDREA KOCSIS, GUEST LECTURER**

*Session: Text Analysis and Visualisation I & II*

Andrea comes from an interdisciplinary and international background. Before has degrees in Communications, Archaeology, History and Geography and received her Mphil and PhD in Heritage Studies from the University of Cambridge. Her doctoral research focused on the impact the national WWI commemorations had on the urban landscape of capital cities, London, Paris and Budapest. She used digital humanities methods, such as NLP and GIS, during this research. As a Cambridge - ESRC intern at BT, she took part in distant reading and machine learning research on misinformation, while at the National Archives, she was a Research Fellow in Advance Digital Methods working with the crowdsourced dataset of the Operation War Diary. Currently, she works as an Assistant Professor in History and Data Science at Northeastern University London.



# TEACHERS' BIOS

---



## MACIEJ PAWLIKOWSKI

*Session: Principles of GLAM Imaging (DCU)*



Maciej Pawlikowski is the Head of the Digital Content Unit at Cambridge University Library. The Unit provides visual content to Cambridge Digital Library and academic researchers fulfilling over 2,000 imaging and licencing requests creating average of 150,000 digital assets annually. He has led number of digitisation projects, which include, Parker on the Web, Darwin Manuscripts, and the Friedberg Genizah digitisation project. As a member ISO TC/42 JW26 workgroup, he contributes to standards for archival imaging, photography of transparent materials and multispectral imaging. Over the years, he has developed and designed equipment, protocols, and techniques to accommodate specialised imaging of cultural heritage artefacts.

# SPEAKERS' BIOS

---

**MORE INFO COMING SOON.**

# CAMBRIDGE DATA SCHOOL CULTURAL HERITAGE APRIL '23



THANK YOU!

**We can't wait to welcome you to our first, in-person Cultural Heritage Data School in this new, two-day format.**

If you have any questions, please do not hesitate to contact us at [dataschool@cdh.cam.ac.uk](mailto:dataschool@cdh.cam.ac.uk).

We look forward to welcoming you soon.

**The CDH Team**

SEE YOU IN CAMBRIDGE SOON



Cambridge  
Digital  
Humanities