Cambridge Social Data School 2023

Monday 26 – Friday 30 June

WELCOME

Welcome to the June 2023 Cambridge Social Data School!

With you here in Cambridge, we are sure that the SDS 2023 is going to be a brilliant event, four years since our last in-person edition.

The projects that you're working on are fascinating, touching on areas like social media, the environment, conflict, and others. We were impressed by receiving so many applications distinguished by their high quality, their timely approach, and their potential to have an impact in the world.

We acknowledge that we can learn from everyone, not just the teachers and keynote speakers. That's why we have created the Incubators – spaces where you can discuss your projects with peers – led by a chair who will also be able to offer some guidance and advice.

We also hope that you can make the most of our keynotes, by Sofija Stefanović, Jacopo Ottaviani, and Edgar Gomez-Cruz, their innovative research and the skills they'll share with us.

We are looking forward to you all delivering a short presentation about your projects at the end of the week. This will be an opportunity for us to see how the SDS has helped your projects develop. We hope that what you've learnt here will contribute to making them impactful interventions in society, be it in the media, academia, or at the grassroots level. But above all, we hope that the SDS 2023 can be a space where learning about data and about the social can enrich us all.

Dr Irving Huerta
Data School Convenor

IN THIS PACK

Welcome to the June 2023 Cambridge Social Data School!

We are so excited to be offering our week-long Social Data School this June. In this pack, you will find all the information you need for the five-day Data School, including:

- key contacts
- important information including software requirements
- map with key sites
- timetable
- module outlines
- teacher biographies

KEY CONTACTS

Dr Irving Huerta
Data School Course Convenor

For all <u>course enquiries</u> regarding content and curriculum, including technology, contact Irving.

Email: rih32@cam.ac.uk

Florence Harry
Data School Assistant

For all <u>admin enquiries</u> including logistics, payment, and planning, contact Florence.

Email: dataschool@cdh.cam.ac.uk

PREPARING FOR THE DATA SCHOOL

MOODLE

During the Data School, we will use Moodle, the University of Cambridge virtual learning environment. You will be invited to sign up to Moodle as a guest before the school begins. The most up to date information you need for the Data School, as well as lots of extra recourses, can be found on the Moodle page. There is also a Q&A which you can use in advance of the school.

https://www.vle.cam.ac.uk/login/index.php

SOFTWARE INSTRUCTIONS

For the Data School, you will need to access some specific software on your personal laptop. Please download the softwares in advance of the Data School. Technical instructions can be found via the QR code or via the following link.

Please note, this is a live document which may be updated before the school.

https://bit.ly/3N6PGVT

DURING THE DATA SCHOOL

The Cambridge Social Data School takes place over five days at The University of Cambridge. Each day will contain a mixture of classes, workshops, and practical sessions. A daily vegetarian/vegan lunch will be provided, as well as morning and afternoon refreshments, except for Monday where the Data School kicks off after lunch (2.00pm) with afternoon refreshments. There is also a local café and buttery on the Sidgwick Site should you wish to use it.

HOUSEKEEPING

PARKING

Parking is very limited on the Sidgwick Site, so we recommend walking if you can. Information on parking can be found at:

www.cambridge.gov.uk/parking

WIFI

During the School, please connect to the University of Cambridge WiFi as a guest. Select the 'UniOfCam-Guest' WIFI network and then connect with one of your online accounts (Amazon, Facebook, LinkedIn etc) or register using your email address.

https://help.uis.cam.ac.uk/service/wi-fi/connect-uniofcam-guest

LIBRARY ACCESS

Cambridge University Library (the UL) is one of the world's oldest university libraries – and home to one of the world's great collections of cultural treasures and research materials.

During the Data School, you may request visitor access to the library. You cannot take books out, but you are welcome to look around, visit the Tea Room, browse and use the study desks in the library. To take advantage of this opportunity, please review all information and submit a visitor request form (via the Moodle page) by 5.00pm, Monday 19 June 2023.

MAP AND LOCATIONS

The main location for the school is Criminology Building (B3&4): Sidgwick Site, 5 West Rd, CB3 9DP. A few break out sessions will be held in the Alison Richard Building (S1). You may be granted access to the University Library for the duration of the school should you wish to visit it (see above).

https://bit.ly/3MPcM1Y
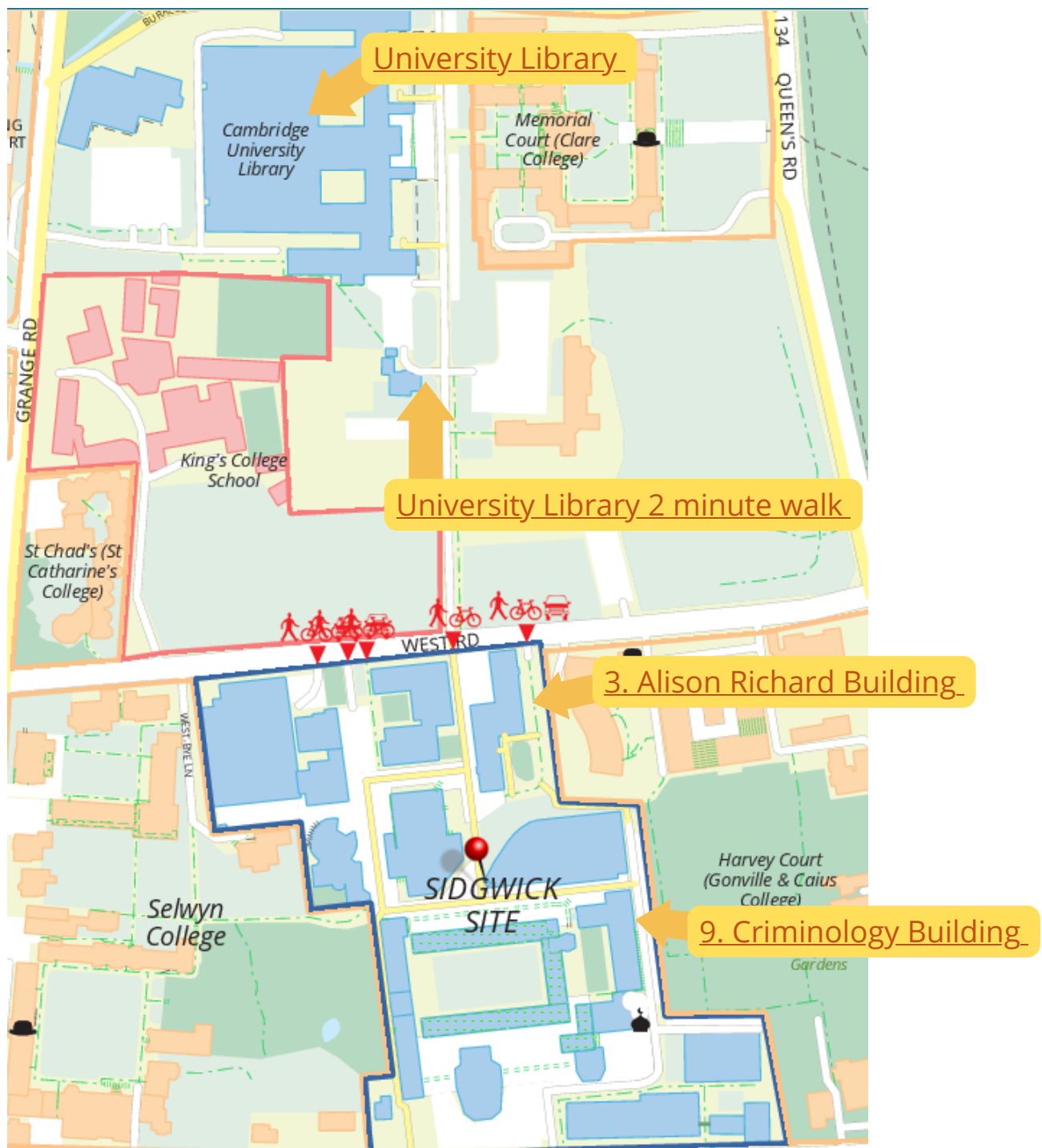
PROGRAMME

This programme is subject to change. Please check Moodle for the most up to date programme.

Day 1: Monday 26 June

Time	Session	Venue	Teacher
2.00–2.30pm	Introduction and welcome	B3&4 Criminology Building	
2.30–2.45pm	Break		
2.45–4.00pm	Presentation of Groups A-F	B3&4 Criminology Building	
4.00–5.15pm	Keynote: Decentring Knowledge	Atrium Criminology Building	Edgar Gomez Cruz

Day 2: Tuesday 27 June

Time	Session	Venue	Teacher
9.30–10.00am	Coffee	Atrium in Criminology	
10.00–11.00am	General Session 1: Methodology for Digital Investigations	B3&4, Criminology Building	Irving Huerta
11.00am–12.00pm	Group split for incubators. Groups A-C	B3&4, Criminology Building	Irving Huerta
11.00am–12.00pm	Group split for incubators. Groups D-F	S1 Alison Richard Building	Hugo Leal
12.00–1.00pm	Lunch	Atrium in Criminology	
1.00–2.00pm	Keynote: Code4Africa	B3&4, Criminology Building	Jacopo Ottaviani
2.00–3.30pm	Makako, Mapping from the Bottom Up & Digital Cartography Workshop	B3&4, Criminology Building	Jacopo Ottaviani
3.30–4.00pm	Break	Atrium in Criminology	
4.00–5.00pm	Workshop: Geolocation and Open Source Investigations I	B3&4, Criminology Building	Nik Yazikov

Day 3: Wednesday 28 June

Time	Session	Venue	Teacher
9.30–10.00am	Coffee	Atrium Criminology Building	
10:00am–12:00pm	General Session 2: Machine Learning Systems: a Critical Introduction	B3&4 Criminology Building	Anne Alexander
12:00–1:00pm	Lunch	Atrium Criminology Building	
1:00–2:00pm	Workshop: Geolocation and Open Source Investigations I	B3&4 Criminology Building	Nik Yazikov
2:00–3:00pm	Workshop: Basic Automations for Investigations	B3&4 Criminology Building	Irving Huerta
3.00–3.30pm	Break	Atrium in Criminology	
3:30–5:00pm	Workshop: Social Network Analysis with Digital Data I	B3&4 Criminology Building	Hugo Leal

Day 4: Thursday 29 June

Time	Session	Venue	Teacher
9.30–10.00am	Coffee	Atrium Criminology Building	
10:00am–12:00pm	General Session 3: Critical Approaches to Visualisation	B3&4 Criminology Building	Anne Alexander
12:00–1:00pm	Lunch	Atrium in Criminology	
1:00–2:00pm	Workshop: Social Network Analysis with Digital Data II	B3&4 Criminology Building	Hugo Leal
2:00–3.50pm	Data protection and surveillance in a networked world?	B3&4 Criminology Building	Jennifer Cobbe
3.50–4.00pm	Break		
4:00–5:00pm	Incubators (Groups A-C)	S1 Alison Richard Building	Irving Huerta
4:00–5:00pm	Incubators (Groups D-F)	B3&4 Criminology Building	Hugo Leal
6:30pm	Optional: Dinner in Cambridge (paid for individually)	Local Restaurant	

Day 5: Friday 30 June

Time	Session	Venue	Teacher
9.30–10.00am	Coffee	Atrium Criminology Building	
10:00–11:00pm	Keynote: Investigating with AI in the community	B3&4 Criminology Building	Sofija Stefanović
11:00–12:00pm	Incubators (Groups A-C)	B3&4 Criminology Building	Irving Huerta
11:00–12:00pm	Incubators (Groups D-F)	S1 Alison Richard Building	Hugo Leal
12:00–1:00pm	Lunch	Atrium Criminology Building	
1:00–3:00pm	Closing plenary, presentations and next steps	B3&4 Criminology Building	

MODULE OUTLINES

METHODOLOGY FOR DIGITAL INVESTIGATIONS

Irving Huerta, University of Cambridge

This module addresses fundamental aspects of investigative practice in digital environments and dwells on the importance of using methodology(ies) for data inquiry. Researchers doing investigations using Open Source Intelligence (OSINT) tools, data collection and analysis, as well as developing automated tools for investigations will benefit from this module. It critically reflects on the essential phases of digital investigations at large: Identification of a Problem (formulation of hypotheses), Information Gathering, Preservation, Verification, Analysis, and Dissemination.

By the end of the module, participants would have the principles to conduct investigations that effectively identify, prove, and strategically disseminate issues in the public interest, with fairness and rigour. Its scope is meant to be applied along with the rest of tools and methods from SDS 2023 modules.

GEOLOCATION AND OPEN SOURCE INVESTIGATIONS

Nik Yazikov/ Amnesty International & Cambridge University's Digital Verification Corps (DVC)

This session will cover geolocation, a crucial stage of any open-source investigation. Geolocation seeks to answer a key question: where did the events depicted happen? We will explore the basic principles of geolocation and introduce participants to a range of tools and techniques. We will cover essential resources including Google Earth Pro and Mapillary, and highlight the advantages of different data sources in a platform-agnostic manner. This workshop aims to encourage a reflexive and critical approach to open-source data, introducing practical skills while emphasising the importance of ethical and transparent research methods. Drawing from the human rights sphere, this methodology is useful for scholars and citizens using open source data such as social media content, online databases and satellite images. By the end of this session, participants will be able to identify useful clues in online content, perform reverse image searches and combine satellite information with street-view data.

MACHINE LEARNING SYSTEMS: A CRITICAL INTRODUCTION
Anne Alexander, University of Cambridge

The current generation of machine learning Artificial Intelligence systems are now widely deployed in contexts as diverse as recommender systems for online shopping and streaming music and video services, facial recognition and biometric systems used by state and private security agencies for the analysis, summarisation and generation of texts and images. This module will present the technical fundamentals of machine learning systems, exploring the challenges of structural bias, lack of transparency and the impact that the design of contemporary AI has on communities and individuals who face structural discrimination. We will demonstrate web-based platforms for creating Machine Learning models and learn about experimental techniques for exploring their potential and limitations.

BASIC AUTOMATIONS FOR INVESTIGATIONS
Irving Huerta, University of Cambridge

In this session, we are going to look into basic software and online tools to carry out a series of automations which can be helpful for investigating the social. Among the tools and methods that we are going to touch upon include automated social media content data collection, scrapping, image/video analysis, and others. This is meant to be a very practical session, with the first half being a quick demonstration of the tools, and the second half being exercises using case studies.

DATA PROTECTION AND SURVEILLANCE IN A NETWORKED WORLD?
Jennifer Cobbe, University of Cambridge

This module will explore legal and critical perspectives on data collection and use. From a legal point of view, this module will look at privacy considerations and general principles of data protection law, setting out some of the constraints and general principles applying to personal data, in particular. More critically, participants will explore questions of power arising from collection and use of data, including around surveillance business models, the data-driven economy, issues of facial recognition and other forms of surveillance, and state surveillance programmes. The module will include a lecture and group work involving relevant case studies and a presentation.

SOCIAL NETWORK ANALYSIS WITH DIGITAL DATA
Hugo Leal, University of Cambridge

"Social network" has become a catch-all term for the online spaces where we connect with other people and trade information in exchange for our personal data and attention. Considering the societal impacts of data-driven economics and politics, knowing how to reclaim and reappropriate these data to trace the form and content of online social networks is a vital skill for journalists, civil society and academics alike.

This module provides a gentle introduction to the field of social network analysis (SNA) with digital data. Social Data School participants will be given the opportunity to "learn by doing" the process of digital data collection as well as the basics of social network visualisation and analysis. After being introduced to the fundamental concepts of SNA, participants will explore all stages of a social network analysis project, including research design, data collection, data wrangling, graph visualisation, and analysis with essential network measures. The focus will be on the retrieval of electronic archival data for non-programmers, and on practical examples of network analysis with specialised software (e.g., Gephi). At the end of the two sessions, participants will be equipped with the basic tools to perform meaningful visualisations and analyses of network data. Typical use cases of SNA range from investigative journalism to NGO monitoring and academic research.

KEYNOTES

MAKOKO, MAPPING FROM THE BOTTOM UP & DIGITAL CARTOGRAPHY WORKSHOP

Jacopo Ottaviani, Code4Africa

Jacopo Ottaviani will be talking about MapMakoko, a drone mapping and data journalism project implemented in Lagos, Nigeria. Code for Africa, a non-profit organisation working on civic technology and digital transformation in African countries, worked with the local community to create a bottom-up, open source map of the area using drones, smartphones, and crowdsourcing. Jacopo will also lead a workshop on Digital Cartography. He will use Datawrapper and go through the entire map lifecycle, ranging from data research, data cleaning, geolocation and visualisation, using sample geographic datasets.

MONITORING THE ENVIRONMENT WITH GRASSROOTS ORGANISATIONS

Sofija Stefanović, University of Cambridge

Sofija will talk about undertaking data collection, as part of a pilot community water monitoring effort of around 500+ exploratory drills, during her fieldwork in the Jadar Valley, Serbia, and touch on the practicalities related to acquiring and assembling citizen monitoring stations. She'll be focusing on water quality monitoring, and drawing on previous experiences working with air quality sensors. This will include pointers to resources and initiatives that can inform similar grassroots efforts or critical data investigations.

DECENTERING KNOWLEDGE

Edgar Gomez Cruz, The University of Texas at Austin

Edgar's describes his research agenda as "a decolonising exploration of epistemology to study the digital". His talk will focus on his efforts to challenge dominant narratives of discourse, practice, and adoption of digital practices across the world. He will share part of the ongoing collaborative project 'Decentering ADM: A Review of Automated Decision-Making (ADM) in the Global South', which is a review of the current state of ADM implementation, practices, and visions in different regions in the Global South (Latin America, Anglophone Africa, South and Southeast Asia and Pacific Island Archipelagos).

DR ANNE ALEXANDER

Data School Course Convenor

Anne Alexander has been Director of Learning at CDH since its foundation and was formerly Co-ordinator of the 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 at CDH. 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).

DR HUGO LEAL

CDH Teaching Associate

Dr Hugo Leal is a Teaching Associate at CDH for the MPhil in Digital Humanities and Research Associate at the Minderoo Centre for Technology and Democracy. Hugo combines research and teaching activities at the intersection of collective action and digital technologies. He has been looking at the networked diffusion of social movements and ideas. His investigation on misinformation and conspiracy theories intends to trace the lifecycle of viral narratives, their strategic use and societal impacts in a variety of areas, from the emergence of nativism to science denial.

DR JENNIFER COBBE
University of Cambridge

Jennifer is a tech law academic working in the Compliant & Accountable Systems research group. Her research focuses on critical interdisciplinary work on law, technology, and society, with a particular interest in examining the socio-political power of tech companies and the transformative effects of their business models and ideological underpinnings on society. She investigates the structural conditions created by new and emerging technologies and explores the legal developments necessary for areas such as internet platforms, the growing use of machine learning, and automated decision-making. Additionally, she collaborates with computer scientists to develop legally compliant and accountable systems, especially in complex and interconnected environments.

NIK YAZIKOV
University of Cambridge

Nik Yazikov is the Cambridge coordinator of the Digital Verification Corps, a student-led organisation affiliated with Amnesty International's Evidence Lab and the Centre for Governance and Human Rights. Nik has recently completed a fully-funded MPhil in the department of History of Art supported by the Bailie Gifford and Selwyn Richards Edwards MPhil awards. He is also a lecturer on 'Open Source Investigations for Academics', a course run in partnership with Cambridge's Social Sciences Research Programme, and has worked as an Open Source Documentation and Verification Associate at the International Committee of the Red Cross.

JACOPO OTTAVIANI

Code4Africa

Jacopo Ottaviani is an award-winning computer scientist who manages Code for Africa's (CfA) data portfolio as Senior Strategist. Since 2011, Jacopo has designed and implemented projects, programs and strategies supported by multiple donors. He has worked in Nigeria, Kenya, South Africa, Ethiopia, Uganda and Senegal. His unique mix of technical skills, as a computer scientist and data journalist, has resulted in a series of projects published by Thomson Reuters Foundation (UK), Der Spiegel (DE), El País (ES), Al Jazeera (International) and Internazionale (IT).

DR EDGAR GOMEZ CRUZ

The University of Texas at Austin

Edgar Gómez Cruz is an Associate Professor at the School of Information at UT. He has published widely on several topics relating to digital culture in top journals, particularly in the areas of material visual practices, digital ethnography, and critical approaches to digital technologies. His recent publications include Vital Technologies: Thinking digital cultures from Latin America (2022).

SOFIJA STEFANOVIĆ

PhD student, University of Cambridge

Sofija Stefanović is a PhD student at the Centre for Doctoral Training in Artificial Intelligence For Environmental Risk. She co-develops data science tools for environmental monitoring with communities resisting mining, particularly in the context of intensified mining for "clean technologies." She is a co-founder of Zeleni talas—a civil society organisation in Serbia focusing on community-driven monitoring and increasing citizen participation in decision-making for urban environmental issues.

We look forward to welcoming you to the Social Data School this June.

If you have any questions, please do not hesitate to contact us at dataschool@cdh.cam.ac.uk.

Enjoy the Data School!