

Cambridge Social Data School

16 – 29 June 2021

Online delivery

Timetable

Dates and timings (all timings in BST)

Please note that because of the fast-moving nature of Coronavirus, this timetable and content may be subject to change because of staff availability.

Date	Time	Topic
Wednesday 16 June 2021	1.30pm – 2.15pm	Introduction and welcome
Wednesday 16 June 2021	2.30pm – 3.30pm	Session 1: Digital Research Design and the Project lifecycle
Thursday 17 June 2021	1.30pm – 2.30pm	Session 2: Data protection and surveillance in a networked world Part I
Thursday 17 June 2021	3.00pm – 4.00pm	Session 3: Machine learning and computer vision Part I
Friday 18 June 2021		Self-paced study day
Monday 21 June 2021	1.30pm – 2.30pm	Session 4: Data protection and surveillance in a networked world Part II
Monday 21 June 2021	3.00pm – 4.00pm	Session 5: Machine learning and computer vision Part II
Tuesday 22 June 2021		Self-paced study day
Wednesday 23 June 2021	1.30pm – 2.30pm	Session 6: Visualising 'Fake News' Part I

Wednesday 23 June 2021	3.00pm – 4.00pm	Session 7: Social Network Analysis with Digital Data Part I
Thursday 24 June 2021	1.30pm – 2.30pm	Session 8: Visualising ‘Fake News’ Part II
Thursday 24 June 2021	3.00pm – 4.00pm	Session 9: Social Network Analysis with Digital Data Part II
Friday 25 June 2021		Self-paced study day
Monday 28 June 2021	1.30pm – 2.30pm	Session 10: Visualising ‘Fake News’ Part III
Monday 28 June 2021	3.00pm – 4.00pm	Session 11: Visualising ‘Fake News’ Part IV
Tuesday 29 June 2021	1.30pm – 3.00pm	Session 12: Closing plenary and next steps

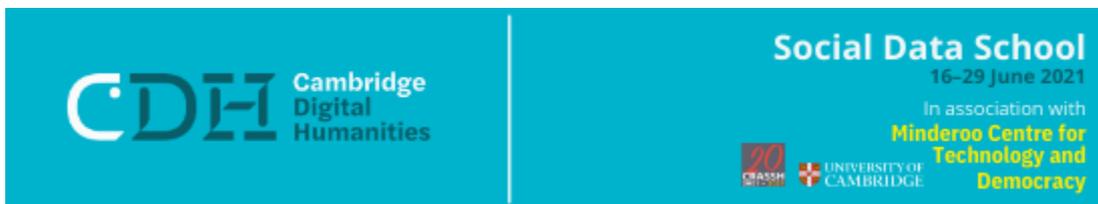
Data School live pre-session: Dr Anne Alexander

This session will include a short presentation about CDH, a Q & A session and introductions by participants.

We will organise online drop-in sessions for participants needing remote help with software installation and other technical problems, these will be advertised closer to the date of the school.

Module 1: Digital Research Design and the Project lifecycle: Dr Anne Alexander Session 1

This introductory session 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.



Module 2: Data protection and surveillance in a networked world: Dr Jennifer Cobbe Session 2 and Session 4

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.

Module 3: Machine learning and computer vision: Dr Anne Alexander Session 3 and Session 5

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 to systems 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. Participants will have the opportunity to work with web-based platforms for creating Machine Learning models and learn experimental techniques for exploring their potential and limitations.

Module 4: Visualising 'Fake News': Andrea Kocsis

Sessions 6, 8, 10, 11

The module aims to help understanding the spread of fake news with the help of a dataset collected by BuzzFeed and with the use of easily available tools. These tools include OpenRefine, which the participants are going to use for data cleaning, filtering and aggregation. During this process participants will learn the basics of regular expressions. The cleaned dataset is going to be visualised with Voyant-tools. During this step, participants will learn basic Natural Language Processing techniques without coding, such as Word Frequency and Topic Modelling, and how to visualise patterns discovered in the data. Finally, participants



gain insight into Python coding when they learn how to add a new column to the dataset from information scraped from the web in Jupyter notebooks. The purpose of this module is to provide participants with confidence in stepping in digital humanities.

Module 5: Social Network Analysis with Digital Data: Dr Hugo Leal **Session 7 and Session 9**

“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 will provide 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, the 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 (e.g., social media platforms) 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.

Biographies



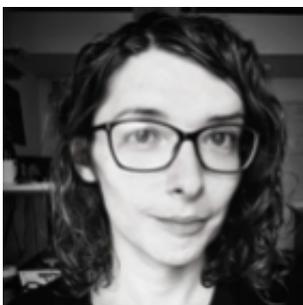
Anne Alexander
Director of CDH Learning

Anne Alexander is Director of the Learning Programme at Cambridge Digital Humanities. 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 at the Alan Turing Institute and a member of the Steering Group of the Trust and Technology Strategic Research Initiative.



Caroline Bassett
Director, Cambridge Digital Humanities

Caroline Bassett is Professor of Digital Humanities, and Director of Cambridge Digital Humanities. Her research explores digital technologies in relation to questions of knowledge production and epistemology (how does 'the digital' change scholarship, transform understanding, produce new scales or perspectives?) and in relation to cultural forms, practices, and ways of being (how can we understand the stakes of informational capitalism, what are its symptoms, how can we understand its temporalities, the forms of life it enables, and those it forecloses?).



Jennifer Cobbe
**Research Associate, Department of Computer Science and
Technology**

Jennifer Cobbe is a Research Associate and Affiliated Lecturer in the Department of Computer Science and Technology (Computer Laboratory) at the University of Cambridge, where she is part of the Compliant and Accountable Systems research group. She is also a member of the Microsoft Cloud Computing Research Centre and an affiliate of the Minderoo Centre for Technology and Democracy. She holds a PhD in Law and an LLM in Law and Governance from Queen's University, Belfast.

Jennifer is generally interested in critical interdisciplinary work on law, technology, and society. Working closely with computer scientists, her research looks at the law's relationship with and responses to new technologies and technological development; the socio-political power of tech companies, their business models and ideological underpinnings, and the structural conditions they produce; and technical and organisational mechanisms for improving legal compliance and accountability of complex systems.

**Andrea Kocsis****CDH Data Schools Methods Fellow**

Andrea Kocsis is an ESRC DTP scholar, finishing her PhD at the Cambridge Heritage Research Centre. In her work, she combines traditional archive research and digital methods (NLP, GIS) in order to compare how urban landscape has changed as a result of WWI commemorations in capital cities (Paris, London and Budapest).

She graduated in Communications major and Art Theory minor (BA and MA) as well as, Medieval Archaeology major and Archaeological Science minor (BA) from the Eötvös Loránd University in Budapest.

She finished MA in General History at the Charles University in Prague and M2 in Territory, Space and Society at the École des hautes études en sciences sociales in Paris. During her MPhil in Archaeological Heritage and Museums at Wolfson College Cambridge, she was working on the topic of the uses of medieval archaeology in Hungarian nation-building. She has recently been selected as the Friends of The National Archives Research Fellow in Advanced Digital Methods.

**Hugo Leal****Research Associate - Minderoo Centre for Technology and Democracy**

Hugo is a Research Associate at the Minderoo Centre for Technology and Democracy (MCTD), hosted by the Centre for Research in the Arts, Social Sciences and Humanities (CRASSH) at the University of Cambridge. Hugo combines research and teaching activities at the intersection of collective action and digital technologies.

His current research focuses on technological threats to democracy and public health.

Hugo holds a PhD in Political and Social Studies from the European University Institute (EUI) and came to Cambridge to work as a postdoctoral fellow in the internet branch of the project “Conspiracy and Democracy” before joining Cambridge Digital Humanities (CDH) as a methods fellow and coordinator of the Cambridge Data Schools. At MCTD, he continues his outreach efforts to provide NGOs, journalists and the general public methods and tools to better understand and interrogate the contents and effects of digital technology.

Hugo is a CDH Associate and a member of the Steering Committee of the Cambridge Infectious Diseases Research Centre, the Association of Internet Researchers and the Centre of Social Movement Studies (COSMOS).