

# Social Data School (January 2023)

Modules (to be scheduled between 11am-6pm GMT)

This programme may be subject to changes.

---

## **Methodology for Digital Investigations**

**(Irving Huerta)**

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 2022 modules.

## **Video Data Analysis**

**(Tom Kissock)**

This module aims to highlight how Video Data Analysis can be used for explorative and investigative research. It covers how the method has evolved since the 1990's, to examine how it can be used and critically understood today. The two sessions focus on how to approach videos for research, data gathering, and storytelling/content creation purposes, and will provide analytical techniques for extrapolating insights from video content. Some of these techniques will include how to code and sample video content, sequence videos, and construct concepts from situational and spatial dynamics. It will also present a short case study of how it's possible, in some scenarios, to triangulate videos with other videos via a 'ready-made multi-camera' method.

Importantly, the course will also address future trends in Video Data Analysis and proposes an experimental methodology for capturing, documenting, and analysing livestreamed videos that are synchronic, ephemeral, and therefore difficult to analyse.

## **Introduction to Critical AI**

**(Anne Alexander)**

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.

## **Social Network Analysis with Digital Data**

**(Hugo Leal)**

“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.

## **Geolocation and Open Source Investigations**

**(Amnesty International and Cambridge University’s Digital Verification Corps)**

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.

## **Data Spatialisation using Python and Blender (TBC)**

### **Session 1 (Nicholas Masterton)**

In the first session we will look at the interface of Blender, and talk about the various workspaces and 3D tools available and how they relate to data visualisation and spatialisation. From here we will import a geographical shapefile using an addon called Blender GIS and manipulate it to create a 3D height field. We will use a node-based shader to apply a gradient to it. Through this process we will develop an understanding of how to manipulate objects in 3D space and how to use colour and shading to communicate gradation within the data.

In the second session we will look into the Blender text editor, the interactive console, and the system console to understand ways of working with python. We will look at a script which is able to read a csv (comma-separated values) file. We will look at the process of iterating through columns and rows of data, using python to output a result into 3D space. This will allow us to develop a methodology for spatialising datasets which are bespoke, hand-crafted, or obscure.