



Social Data Science needs Law

Andrew McHugh, Urban Big Data Centre, University of Glasgow

Legal Approaches to Data: Scraping, Mining & Learning

May 27, 2021



JOINTLY FUNDED BY



University
of Glasgow

About the Urban Big Data Centre

- UK Government (Economic and Social Research Council) & University of Glasgow funded
- Priorities:
 - Data infrastructure and collections
 - Priority research strands: transport & mobility; neighbourhood, housing & environment; education, skills & productivity; big data & urban governance
 - Combining social science research with data analytics and computing science
- Overall aims:
 - Achieve public policy impact
 - Critically evaluate role and value of big data and urban analytics
 - Enhance data and methods

“Promoting innovative research methods and the use of big data to improve social, economic and environmental well-being in cities”

Data science has implicit legal questions

- Getting hold of data to support our research and data service users
 - Content licensing / contracts / sublicensing
 - Web scraping
- AI and machine learning
 - Trained models
 - Sensitivity of training data
- Software release
 - Licensing and commercialisation

Case study – AirBnB (1)



- AirBnB is an online marketplace for arranging or offering lodging, homestays or tourism experiences
- What are the impacts of the rapidly growing sharing economy for private rented sector property market?
 - AirBnB is not enthusiastic about facilitating such research
 - As an unregulated sector there is little reliable data available
 - Some third party providers make these data available (under US copyright law) but problems for robust academic work:
 - Sampling limitations
 - Data quality issues
 - Black box models

Case study – AirBnB (2)



- We wanted to automate the **scraping** of AirBnB on a systematic basis to capture and store publicly accessible web content and facilitate research
- Aiming to understand legality of our proposed data collection
- Methods designed to enable us to demonstrate empirically and reliably:
 - Scale and growth of AirBnB
 - Spatial change and focus of short-term lets
 - Occupancy
 - Price
 - Availability

Case study – AirBnB (3)



- Exploring questions such as:
 - What impact does AirBnB have on the availability of private renting stock?
 - Where is that impact greatest?
 - What is the relationship with areas of social deprivation? Does the rise of the sharing economy increase inner-city gentrification and the suburbanisation of poverty?
 - Are AirBnB properties falling below the standards for the PRS e..g in terms of occupancy levels?
 - What impact does AirBnB have on the existing hospitality industry?

Case study – AirBnB (4)



- Scraping means automating the transfer of data from a web site
 - Text pattern matching
 - HTTP programming
 - HTML / DOM parsing
 - Computer vision
- Unsupported by the web content owner, prone to break if the website changes
- Negotiate website controls intended to block nuisance hosts

Relevant Forthcoming Events from UBDC



- Practical approaches for web scraping for research – using Airbnb as an example data provider
 - Tuesday 22 June, 10:00 - 12:00 (BST)
Registration coming soon
- Using daily Airbnb web scraped data to provide spatial and temporal understanding of short-term lets activity
 - Thursday 24 June, 10:00 - 11:00 (BST)
Registration coming soon

Thanks – Questions?

To stay up to date with our research and other activities you can sign up to our newsletter on our website and follow us on social media:

www.ubdc.ac.uk

