

# International Journal of Population Data Science

Journal Website: [www.ijpds.org](http://www.ijpds.org)



## The Value and Challenges of Making Survey and Digital Trace Datasets Available for Open Access

Riza Battista-Navarro<sup>1</sup>, Marta Cantijoch<sup>1</sup>, Alex Cernat<sup>1</sup>, Conor Gaughan<sup>1</sup>, and Rachel Gibson<sup>1</sup>

<sup>1</sup>University of Manchester, Manchester, United Kingdom

### Introduction & Background

Over the last two decades, the digital revolution has led to an explosion of new data sources commonly referred to as digital footprint or trace data (DTD). This rapid expansion in digital data sources has pushed survey research into a new era of development that now centres on its linkage with various participant DTD. This culture shift has unlocked a range of novel opportunities for social scientists to access rich new sources of insight into human behaviour which can be used to augment, validate or even replace conventional self-reported survey data. However, when it comes to making such data open access, there remains a critical gap about maintaining respondent anonymity when it comes to openly releasing DTD.

### Objectives & Approach

This paper will focus on demonstrating the conceptual and methodological value and challenges in producing anonymised and standardised variables from survey respondents' digital trace data (DTD). We will do this using existing YouGov datasets collected over two time periods in the US 2020 and 2024, and a third collected in the UK 2022. The US datasets link individual survey responses to their Twitter/X feeds and the UK to their browsing history. All three datasets were designed to address research questions about the effects of digital media consumption and exposure on citizen attitudes and behaviours. This paper aims to establish a standardised and automated process for variable generation which is replicable and can produce anonymised variables from the DTD which can be safely linked to respondent survey data and openly shared with the wider research community.

### Relevance to Digital Footprints

The aim of this work is to encourage other researchers working with digital footprint data to consider the ethical and legal implications they face when looking to make their DTD open access. Our work aims to resolve the conflict between open access and data protection, bridging the gap by establishing a process for deriving anonymous unit-level variables which can be released in lieu of the raw DTD. While not designed to

be an entirely prescriptive method, this paper strives to inform strategies for making DTD open access and to start the process of creating better standardised practices within the discipline.

### Conclusions & Implications

While this paper is still a work in progress, work is underway for variable generation and will result in the creation and release of a standardised procedure for the anonymisation of DTD. These variables will be created for two specific types of DTD: social media and web-browsing data. However, these variables will be translatable to various other types of DTD and this paper will be accompanied by step-by-step code and codebook which can be used by other researchers. This paper will have significant ethical and methodological implications for how researchers working with DTD make their data open access and will hopefully improve transparency and collaboration within the discipline.

### Keywords

Digital trace data, Open Research, Survey methods, Digital footprint, Data linkage

