

Ethical governance for repurposing geospatial data: Abstract for LESSON 2019 – Zurich, October 2019

Thomas King, RSS Data Ethics SIG (tk.socstats@gmail.com).

ABSTRACT

There is great potential to reuse existing data to geographically plan and target delivery of services in public and private sectors, but it requires suitable governance. However, most of the recent concern around data ethics has focused on technology, and advanced applications, described as artificial intelligence (AI). Statistical issues, apart from privacy protection procedures, are largely overlooked despite their much wider application (see rss.org.uk/data-ethics). Existing data can be reused, through reanalysis in its original intended form, through repurposing to new applications e.g. service evaluation, and recombination to create new data resources (Steinmann et al., 2016). While the ethical governance of reanalysis has various precedents for research projects, repurposing and recombining data is much more problematic. Geospatial data, in the form of a standard location, is often used in indexing transactions, whether service delivery or digital communications, but it also serves to identify the individual or their community. As the location is intrinsic to the transaction, there is no opportunity to opt out of its inclusion in the administrative system database. Thus any standard terms and conditions may imply the holder of the data is free to use the data as they determine. While boundaries are often imposed e.g. restricting individual data for legal reasons, an ethical analysis would suggest internal uses need governance more generally. Specifically, use ought to respect reasonable expectations of those originating it, and be respectful of the dignity of social groups, particularly those considered vulnerable.

Appropriate governance of data should not be designed around regulation and restrictions, and geospatial data is a particularly salient example. The social world is geographically structured, and services are designed to reflect different needs, priorities and demands of local communities. Some of these obviously benefit from having access to individual spatio-temporal transactions, the most established being public transport systems. While the attendant risks of the open publication of such transaction data identifying individuals are clear, the marginal impact of doing so is not: a person travelling on public transport is also visible to their fellow passengers; stalking is already understood as a personal privacy problem, and is the most obvious threat from identification. Geodemographic data is already used to provide intelligence to companies wishing to plan their business strategy, although the terms used to characterise consumer groups are not always respectful. However, relying solely on

administrative sources of data risks introducing bias into decisions, whether from poor coverage of certain groups, or varying validity of attributes. While national surveys have established that people have little realisation of how their data is being used, they are often concerned but their political representatives have lacked capacity for oversight. An alternative increasingly employed in projects using health data is to involve a range of non-experts in governance, finding they readily pick up the issues and shift priorities. While media report considerable concern about the ethical basis for the use of new types of data, the informed citizen response is often overlooked, even as it emphasises the benefits of using data.

REFERENCE

Steinmann, Michael, Matei, Soren A. and Collmann, Jeff. "A theoretical framework for ethical reflection in big data research." In *Ethical Reasoning for Big Data: an exploratory analysis* edited by Jeff Collmann and Soren Matei, 11-27. Basel, CH: Springer, 2016.