

# Understanding Data and their Environment: Assessed Work (Essay)

---

**Essay Length:** Less than 2000 words (excluding diagrams, tables and references).

**Deadline:** 15<sup>th</sup> June 3PM

## Outline

Consider the following data situation:

*As part of the troubled families programme, the UK government wishes to understand the complex nexus of issues that appear to co-occur in family units including deprivation, poor educational and health outcomes, underemployment and crime. It has created the troubled families dataset, which combines administrative data from multiple government departments and local authorities. This will allow it to track the trajectories of the identified families and measure whether specific interventions are effective.*

*It is considering allowing researchers are to access those data to carry out deeper analyses.*

Your task is to assess the risk of this data situation and to recommend appropriate mitigations. You will need to consider ethical legal policy and technical issues inherent in the data situation.

You could approach this by applying the Anonymisation Decision-making Framework explicitly (perhaps completing the template and including it as an appendix) or you may simply want to draw on key elements of it. You do not have direct access to the data for this exercise so there are some things that you are not fully able to assess, but the report (see below) does give quite a lot of information about the scope and scale of the data.

Another approach is to describe the issues and principles that generally apply to the sharing of data, drawing on a wider literature and then apply those to the described data situation.

## Background information

You can find more information on the troubled families programme here:

<https://www.gov.uk/government/news/troubled-families-programme-annual-report-published>

And the dataset here:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/605185/Family\\_outcomes\\_-\\_national\\_and\\_local\\_datasets.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/605185/Family_outcomes_-_national_and_local_datasets.pdf)

### Distribution of the marks for the essay

Area of Essay	% share of marks
Introduction	10
Description of approach	20
Analysis of Troubled families case	50
Conclusion, implications/recommendation	10
Presentation: standard of English, overall Coherence	10

### Relevant Reading

- ARRINGTON, M. (2006). *AOL proudly releases massive amounts of user search data*, TechCrunch, available at: <http://tinyurl.com/AOL-SEARCH-BREACH> [accessed 30/5/2016].
- ATOKAR (2014). *Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset*, available at: <http://tinyurl.com/NYC-TAXI-BREACH> [accessed 30/5/2016].
- BAYARDO, R. J., & AGRAWAL, R. (2005, April). Data privacy through optimal k-anonymization. In 21st International conference on data engineering (ICDE'05) (pp. 217-228). IEEE.
- CAREY, P. (2018). *Data protection: a practical guide to UK and EU law*. Oxford University Press.
- CNN MONEY (2010). *5 data breaches: From embarrassing to deadly*, available at: <http://tinyurl.com/CNN-BREACHES/> [accessed: 30/5/2016].
- DE HERT, P., & PAPAKONSTANTINO, V. (2016). The new General Data Protection Regulation: Still a sound system for the protection of individuals?. *Computer law & security review*, 32(2), 179-19
- DIBBEN, C., GOWANS, H. ELLIOT, M. J. AND LIGHTFOOT, D. (2015). 'The Data Linkage Environment' in *Methodological Developments in Data Linkage*; Harron, K. Goldstien H. and Dibben, C. (eds) 36-62 Wiley.
- DUNCAN, G. T., ELLIOT, M., & SALAZAR-GONZÁLEZ, J. J. (2011). *Statistical Confidentiality*. Springer New York.
- ELLIOT, M. J., MACKEY, E., O'HARA, K., & TUDOR, C. (2016). *The Anonymisation Decision-Making Framework*. UKAN publications; Manchester.
- ELLIOT, M., O'HARA, K., RAAB, C., O'KEEFE, C.M., MACKEY, E., DIBBEN, C., GOWANS, H., PURDAM, K. and MCCULLAGH, K., (2018). Functional anonymisation: personal data and the data environment. *Computer Law & Security Review*, 34(2), pp.204-221.

- FLORIDI, L., & TADDEO, M. (2016). 'What is data ethics?', *Royal society* <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0360>
- HAYNES, C. L., COOK, G. A., & JONES, M. A. (2007). Legal and ethical considerations in processing patient-identifiable data without patient consent: lessons learnt from developing a disease register; *Journal of Medical Ethics*, 33(5): 302–307, DOI:10.1136/jme.2006.016907.
- LANE, J., STODDEN, V., BENDER, S. & NISSENBAUM, H. (Eds.) (2014). *Privacy, Big Data, and the Public Good*. Cambridge: Cambridge University Press.
- MACKEY, E., & ELLIOT, M. (2013). Understanding the data environment. *XRDS: Crossroads* 20(1), 36-39.
- NARAYANAN, A., & SHMATIKOV, V. (2008). "Robust de-anonymization of large sparse datasets". In *2008 IEEE symposium on security and privacy* (pp. 111-125). IEEE.
- NISSENBAUM, H. (2004). "Privacy as contextual integrity"; *Washington Law Review*, 79 (119): 101-139, available at: <http://tinyurl.com/j8xut58> [accessed 30/5/2016].
- OHM, P. (2009). "Broken promises of privacy: Responding to the surprising failure of anonymization" *UCLA L. Rev.*, 57, 1701.
- RUBINSTEIN, I. S. (2013). "Big data: the end of privacy or a new beginning?", *International Data Privacy Law*, 3(2): 74-87. Available at: <http://tinyurl.com/q3wvd53> [accessed 28/5/16].
- UK: INFORMATION COMMISSIONER'S OFFICE (2011). *Data sharing code of practice*, available at <http://tinyurl.com/ICO-SHARE> [accessed 25/5/2016].
- UK: INFORMATION COMMISSIONER'S OFFICE (2012a). *Anonymisation: managing data protection risk code of practice*, available at <http://tinyurl.com/ICO-ANON> [accessed 25/5/2016].
- UK: INFORMATION COMMISSIONER'S OFFICE (2014a). *Data controllers and data processors: what the difference is and what the governance implications are*, available at <http://tinyurl.com/ICO-CONT-PROC> [accessed 30/5/2016].
- UK: INFORMATION COMMISSIONER'S OFFICE (2014b). *Conducting privacy impact assessments code of practice*, available at <http://tinyurl.com/ICO-PIA2> [accessed 30/5/2016].
- UK: INFORMATION COMMISSIONER'S OFFICE (2016). *Guide to Data Protection version 2.4*, available at <http://tinyurl.com/ICO-DPG2-4> [accessed 30/5/2016].
- WELLCOME TRUST (2016). *Public attitudes to commercial access to health data*, available at <http://tinyurl.com/zl2es5b> [accessed 19/6/2016]

ZIMMER, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology*, 12(4), 313-325.

ZWITTER, A. (2014). "Big data ethics", *Big Data & Society*, 1(2).  
<https://doi.org/10.1177%2F2053951714559253>