Data Sharing and Data Trusts Workshop

This is a workshop for a researchers, practitioners and policy makers with an interest in data sharing. The goal of the workshop is to enable you to

a) Recognise the benefits of and obstacles to data sharing
b) Define why you may or may not want to share data
c) If you do want to share data, develop the big picture of how this should be done

This will be achieved through a series of talks, case studies and activities. The talks will provide an overview of the subject area. A case study will then exemplify the individual aspects of data sharing. The activities will guide you to collaboratively answer key questions about data sharing:

1. What is data sharing, where and how can or should data be shared
2. What role trust plays in data sharing and how it is generated and maintained
3. Which models of data access and sharing are suitable in different circumstances

As participants from different backgrounds, you will learn from the combined expertise of the researchers from the University of Southampton, the practitioners from the ODI, and also from each other. This will you to understand data sharing not only from your own perspective, but also from the perspective of others with whom you might share data.

Speakers

Johanna Walker, University of Southampton
Johanna is a Senior Research Assistant at the University of Southampton and works on the Data Pitch, Smart City Innovation Framework Implementation and European Data Portal projects. In her role in Data Pitch she led the negotiation of data sharing agreements and managed the governance arrangements. Along with Elena Simperl she authored a report on alternative models for data trusts to support the AI industry for the Office of AI. Johanna

holds a MBA with Distinction from London Business School. Her PhD thesis is on the intersection of innovation, open data and data sharing.
Johanna will talk about the benefits and pitfalls of data sharing.

Dr Kieron O’Hara, University of Southampton
Kieron is an associate professor in Electronics and Computer Science at the University of Southampton. He is a director of the Web Science Institute with responsibility for policy, and a visiting professor of Law at the University of Winchester. He is one of the leads of the UKAN network of anonymisation professionals, co-wrote the UKAN anonymisation framework, and recently helped adapt it for Australian law. He wrote the Cabinet Office’s report on transparency, open data and privacy in 2011 that anchored the UK’s open data policy, and chaired the Transparency Sector Panel for crime and criminal justice data from 2011-15.
Kieron will talk about the role of trust for data sharing.

Peter Wells, Open Data Institute
Peter is Director of Public Policy at the Open Data Institute (ODI). He works with governments and businesses around the world to build an open and trustworthy data ecosystem, where people, organisations and communities can use data to make better decisions and are protected from any harmful impacts. Peter spent over 20 years in the telecommunications industry working with both telecommunications companies and regulators to investigate new technologies, transform businesses and launch new products. Before joining the ODI, Peter worked in a voluntary role to organise an independent review of digital government policy for the Labour Party, and for Open Addresses Limited to experiment with creating a collaboratively maintained open address register.
Peter will talk about the different approaches to increasing access to data while retaining trust.

Dr Gefion Thuermer, University of Southampton
Gefion is a Research Fellow in the Web and Internet Science group at the University of Southampton. She works on the Data Pitch project, where she is responsible for producing reports about the key lessons learned from the experience of data sharing. She holds a PhD in Web Science, and has previously focused on online participation.
Gefion will be the moderator of the workshop.

Data Sharing Glossary

- **Access Controls**: Security measures applied by a data owner or provider to any data consumer with which it proposes to share its data. These include placing terms and conditions on the use or reuse of the data, or allowing the data consumer access to the data only under some specified data environment.
- **Anonymisation**: Techniques for lowering the risk of identification of data subjects from data, typically by removing or aggregating data that would (help) identify data subjects, combined with other measures, such as adding noise.
- **Confidential data**: A term from common law, to refer to data or personal information which is shared in confidence with another party, such as a lawyer, accountant or doctor, in order to allow the second party to act in their client's interest. Such information should not be shared with any third party except with the clear consent of the client; this, if it happens, is called a breach of confidence. Confidentiality agreements are often implicit, when confidentiality is a reasonable expectation of the client.
- **Data Consumer**: A person or entity that uses data for their own purposes, for example for business, academic work or in government. Data consumers might well transform data, for example by cleaning it up, merging it with other datasets, or feeding it into other systems. In order to consume data, consumers must process it.
- **Data Controller**: A legal term from the GDPR, to refer to a person, company, or other body that determines the purpose and means of personal data processing (this can be determined alone, or jointly with another person/company/body). The Data Controller is responsible for what happens with the data, and held accountable for any breaches of data protection.
- **Data Environment**: The context in which data is held. Data environments are characterised by agents with access to the data, other datasets with which the data may come into contact, governance arrangements for the data, and infrastructure used to store it. Typically, a dataset will be stored in a range of different data environments. Data sharing involves moving the shared data from one environment into another.
- **Data Owner** or **Provider**: An entity that owns a dataset; this could, for example, be a company with sales data, or a GP with a patient database. For data sharing to take place, a data owner or provider must facilitate access to the data for a data consumer. Note that if the data owner facilitates such access to personal data, then this will be regulated by GDPR; if it is not personal data, then it won't be.
- **Data Processing:** A legal term from the GDPR, to refer to any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction

- **Data Processor**: A legal term from the GDPR, meaning a natural or legal person, public authority, agency or other body which processes personal data on behalf of the Data Controller. A Data Processor is not necessarily also a Data Controller: It could be a third party that a Data Controller delegates the processing to, such as an external analyst.
- **Data Sharing**: The sharing of data between entities, typically for a specific purpose. This can happen between companies, or departments within an organisation. The data owner or provider provides a data consumer with access to some of its data. If the data is personal data, then the sharing will be regulated by GDPR; if not, not.
- **Data Subject**: A legal term from GDPR, referring to a living person that is or can be identified through data. The term does not extend to institutions, organisations, or deceased individuals.
- **Data Trust**: There is no one definition of what a Data Trust is (yet). Data Trusts were recommended by the independent review 'Growing the artificial intelligence industry in the UK' as a way to make Data Sharing easier. As a working definition, O'Hara suggests that A data trust works within the law to provide ethical, architectural and governance support for trustworthy data processing. Different types of Data Trusts are currently a topic of debate; Hardinges' article provides a good overview.
- **Functional Anonymisation**: A risk management approach to anonymisation that accepts that whether data is anonymous or not is a function of the relationship between those data and their environment, and not a property of the data itself. Hence functional anonymisation goes beyond manipulation of the data, and encompasses manipulation of the data environment.
- **GDPR**: The General Data Protection Regulation; an EU regulation that came into force in May 2018. GDPR provides new definitions of terms such as data processing or anonymisation, and defines different bases on which data processing is allowed. It goes further than previous legislation in protecting data subjects, and as an EU regulation, unifies data protection across the EU, and thereby allows the flow of data across the single market. The UK has implemented GDPR through the Data Protection Act 2018.
- **Metadata**: Data that describes the properties of data. Metadata can be attached to a dataset, and can therefore be used to understand whether that dataset is of interest to potential consumers, without giving consumers access to the data itself. Of particular importance is metadata describing the provenance of data.
- **Open Data**: Data that is freely available on the internet, without access controls.
- **Provenance**: Metadata that gives a record of the inputs, entities, systems, and processes that were involved in the creation of data, providing a record of its origins.
- **Pseudonymisation**: Techniques involving the substitution of identifiers that are easily attributed to individuals with, eg, an ID number, which is stored separately. Re-identification of the data is possible by reference to the original key.
- **Synthetic Data**: Created algorithmically rather than generated by real-world events. It is generally used to explore datasets before sharing, as a stand-in for test datasets of production or operational data, to validate models, and to train machine learning models.

Pre-Workshop Reading

Jack Hardinges, ODI: *What is a Data Trust?*

https://theodi.org/article/what-is-a-data-trust/

Good overview of the current discussion around the nature and purpose of Data Trusts, and different models for how they could be implemented.

GovLab: *An introduction to Data Collaboratives*

http://datacollaboratives.org/introduction.html

Comprehensive overview of Data Collaboratives, which are data sharing institutions for public good spearheaded by the US-based GovLab.

Aida Mehonic, Alan Turing Institute: *Can Data Trusts be the backbone of our future AI ecosystem?*

https://www.turing.ac.uk/blog/can-data-trusts-be-backbone-our-future-ai-ecosystem

Key ethical challenges for data sharing, particularly of health and other (ultra) personal data.

Sean McDonald, Centre for International Governance Innovation: *Reclaiming Data Trusts*

https://www.cigionline.org/articles/reclaiming-data-trusts

Why data trusts are currently of interest, their intellectual foundations and five key policy priority areas.

Post-Workshop Reading

Summaries & Opinions

Andrew Collinge (2018: *A 'New Deal' for City Data?*

https://data.london.gov.uk/blog/a-new-deal-for-city-data/

How London is planning to use data trusts to ensure effective stewardship and productive use of data, and increase citizen-government trust.

Nesta (2017): *The New Ecosystem of Trust*

https://www.nesta.org.uk/blog/new-ecosystem-trust/

A comprehensive overview of a possible taxonomy of data trusts. Among other things, it defines different types of data trusts along dimensions of public value and individual control.

Peter Wells & Jack Hardinges, ODI, (2018): *Defining a data trust*

https://theodi.org/article/defining-a-data-trust/

A brief blog outlining the ODI's approach to defining a data trust.

Teresa Scassa (2018): *Digital governance and Sidewalk Toronto: Some thoughts on the latest proposal*

http://www.teresascassa.ca/index.php?option=com_k2&view=item&id=290:digital-governance-and-sidewalk-toronto-some-thoughts-on-the-latest-proposal&Itemid=80

Insight into the assumptions that are embedded in a specific example of a (controversial) proposed civic data trust, and the importance of mapping what data will be collected, by whom, in what form and for what purpose. Challenging questions about the public-private relationship are raised.

Practical Guidance

Sophie Stalla-Bourdillon, Laura Carmichael (2018). ***Legal and privacy toolkit v2***

Data Pitch H2020 Project Deliverable

https://datapitch.eu/datapitch-d3-5-legal-and-privacy-toolkit-v2/

Practical guidance to navigate the Data Protection aspects in Data Sharing processes.

Reports & White Papers

Element AI: ***Data trusts: reinforced data governance that empowers the public***

https://hello.elementai.com/data-trusts.html

Element AI creates AI products for decision making. This white paper addresses 5 key questions of control, power, privacy and innovation with data and considers whether data trusts can assist in providing solutions.

Professor Dame Wendy Hall & Jérôme Pesenti (2017): ***Growing the Artificial Intelligence Industry in the UK***

London: Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy

https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

An independent review for the UK Government. It makes 17 recommendations for supporting the AI industry in the UK, amongst them, the development of data trusts to make more data available.

Kieron O'Hara (2019): ***Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship***

https://www.southampton.ac.uk/wsi/enterprise-and-impact/white-papers.page

The purpose of this White Paper from the Web Science Institute is to explore the questions of (a) what existing structures can data trusts exploit, and (b) what relationship do data trusts have to trusts as they are understood in law?

Royal Society (2017): ***Data management and use: governance in the 21st century***

https://royalsociety.org/-/media/policy/Publications/2017/Data_management_and_use_governance_in_the_21st_century_2017_seminar_report.pdf

A report on a Royal Society seminar with representatives of 14 organisations, calling for stewardship of the entire data governance landscape.

Academic Papers

Sylvie Delacroix & Neil Lawrence (2018): ***Disturbing the 'One Size Fits All' Approach to Data Governance: Bottom-Up Data Trusts***

http://dx.doi.org/10.2139/ssrn.3265315

Data trusts as legal mechanisms that allow data subjects to authorise (or otherwise) use of their data via the terms of the Trust.

Cindy Gallois, Peta Ashworth, Joan Leach & Kieren Moffat (2017): ***The language of science and social licence to operate***

Journal of Language and Social Psychology, 36(1), 45-60

https://doi.org/10.1177/0261927X16663254

Social License to Operate (SLO) is the ongoing acceptance or approval of an operation by those community stakeholders who are affected by it and who can affect its profitability. This paper addresses this form of informal agreement and its implications.

Elaine Mackey & Mark Elliot (2013): ***Understanding the data environment***

XRDS, 20(1), 36-39

https://dl.acm.org/citation.cfm?doid=2517249.2508973

Understanding how to protect data privacy and anonymity via a better understanding of what might threaten them, and how that risk might be increased by sharing across different data environments.

Iryna Susha, Marjin Janssen & Stefaan Verhulst, (2017): ***Data Collaboratives as a New Frontier of Cross Sector Partnerships in the Age of Open Data: Taxonomy Development***

Proceedings of the 50th Hawaii International Conference on System Sciences, 2691–2700

https://pdfs.semanticscholar.org/4682/85434d7eb14f0610ffaf6a6f0d591286e9ac.pdf?_ga=2.172306922.857938525.1557485306-1271185416.1551178080

Focusing on data collaboratives rather than trusts, this is nonetheless an insightful overview of key considerations of 14 different aspects of data sharing, 6 affecting supply and 8 affecting demand.

Meg Young et al. (2019): ***Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing***

Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). ACM, New York, NY, USA, 191-200

https://faculty.washington.edu/billhowe/publications/pdfs/young_open_v_closed_semi_synthetic_data.pdf

Legal and technical approaches for maintaining privacy, fairness, accountability and transparency while sharing data, based on transport cases.