

POSITION PAPER

A manifesto for increasing access to data in engineering

Leigh Dodds , Pauline L'Hénaff, James Maddison and Deborah Yates

Open Data Institute, London, United Kingdom

*Corresponding author. E-mail: leigh.dodds@theodi.org

(Received 06 March 2020; accepted 02 April 2020)

Keywords: Data infrastructure; data sharing; data stewardship; innovation; open data

Abstract

This paper introduces a set of principles that articulate a shared vision for increasing access to data in the engineering and related sectors. The principles are intended to help guide progress toward a data ecosystem that provides sustainable access to data, in ways that will help a variety of stakeholders in maximizing its value while mitigating potential harms. In addition to being a manifesto for change, the principles can also be viewed as a means for understanding the alignment, overlaps and gaps between a range of existing research programs, policy initiatives, and related work on data governance and sharing. After providing background on the growing data economy and relevant recent policy initiatives in the United Kingdom and European Union, we then introduce the nine key principles of the manifesto. For each principle, we provide some additional rationale and links to related work. We invite feedback on the manifesto and endorsements from a range of stakeholders.

Impact Statement

Researchers and professionals working in engineering and related sectors need access to data to help drive innovation required to address a variety of social, economic, and environmental challenges. If we are unable to access necessary data, then we risk limiting our ability to benefit from new technologies and approaches. When increasing access to data, we must also consider potential negative impacts, for example, on privacy, and ensure we are building capabilities necessary to guide and support its use. This paper introduces a manifesto designed to help drive the behavior change necessary to help unlock the value of data. The nine principles cover a range of factors that will build capability, opportunity, and motivation for sharing and using data. The goal is to help a variety of stakeholders across industry, government, and academia to understand their role in driving change and highlight how a range of existing activities align with these principles.

1. Introduction

The ability to easily collect, store, and use large volumes of data is driving change across our economy. Data is helping to increase productivity, support the application of new technologies, enabling innovative research, and the development of new products and services.

Many studies have attempted to assess the value of the data economy and from increasing access to data. For example, European Commission (2017) estimates that the European Union (EU) data economy was worth €300 billion in 2016, and estimates that this will have increased to €739 billion in 2020. McKinsey (2018) projected that data-enabled applications of artificial intelligence (AI) will generate \$13 trillion in new global economic activity by 2030.

European Data Portal (2020) estimates that the value of open data for the EU28+ was €184 billion in 2019, and forecast it to reach between €199.51 and €334.21 billion by 2025. The report also looked at employment figures, with 1.09 million open data employees in 2019 and 1.12–1.97 million open data employees forecast by for 2025.

However, as a recent report by the Bennett Institute (Coyle et al., 2020) has highlighted, putting a precise figure to the value of data is extremely difficult. This is in part due to the variety of types of data, and the unanticipated ways in which they can be reused. The report argues that the specific economic characteristics of data, and the data economy, mean that market forces alone will not be able to realize its full potential. One contributing factor (London Economics, 2019a) is the mutual uncertainty across data producers and consumers of the potential uses, users, volume, variety, or quality of data that is available.

Based on this research, it is clear that to increase the social, economic, and environmental value from data, will require the activities of a range of stakeholders across government, businesses, and academia. Specifically, there is a need to change how data are being accessed, used, and shared.

2. Realizing the Value of Data in Engineering and Related Sectors

The open data, open government, and open research communities are examples of movements whose goals are to increase access to data by changes to norms, practices, and approaches across a broadly defined set of organizations, disciplines, and communities.

Driving behavior change requires organizations that are collecting and stewarding data need to have the capability, motivation, and opportunity to share and open up data (see Michie et al., 2011). While broad movements can help to build motivation, these need to be supplemented by more focused programs that can help to build stronger alignment across stakeholders.

There are a variety of examples of more targeted sectoral approaches to driving change in ways that are intended to help address a range of social, environmental, and economic challenges. These include Open Banking in finance (Open Data Institute, 2019a), OpenActive in the physical activity sector (McKenna, 2019) and GODAN (n.d.) in agriculture.

The sharing and use of data are governed by a variety of nested rules. Legislation, to protect privacy, or regulation to mandate access to data to create more equitable markets provide the broader legal context within which organizations and institutions operate. Within that context, specific sector programs, contracting standards, and principles will also shape the governance of data. Professional and organizational norms and policies provide a more specific set of rules. As the open research movement has shown, individual incentives and capabilities are also a factor in driving behavior change (Fane et al., 2019).

Changing these rules and developing a culture of trusted sharing of data requires a range of stakeholders to take co-ordinated action.

3. Exploring a Manifesto for Change

In our recent report (Lloyd's Register Foundation, 2019b), we explored a range of potential benefits for increasing access to data in engineering and a range of related sectors, including construction, transport, and health and safety. Our report includes some short case studies and identifies a number of existing barriers to change. Many of these are common to those we have encountered in other sectors.

Recognizing the need for co-ordinated action, we presented a manifesto (Open Data Institute, 2019e) for sharing engineering data. We believe that the principles outlined in that manifesto can help to fulfil two goals.

First, we believe it presents a shared vision of a world where we are able to realize increased value from data, to help to improve safety, engineer a more resilient built environment and adapt to a changing climate.

Second, the principles provide a useful framework for understanding how a variety of existing initiatives are contributing toward delivering that vision for the future. By understanding overlaps and complementary approaches across programs and policy initiatives we hope to help those leading these existing activities identify opportunities for closer collaboration. If gaps are identified, then we hope to

incentivize individual stakeholders in taking a leading role within their community or profession. With this in mind, the manifesto includes recommendations for a range of stakeholders including government, professional bodies, funders, the privacy sector, and academia.

In the following sections, we provide some additional background on each of the principles in the manifesto with pointers to relevant work.

3.1 Data is infrastructure

The “Data for the Public Good” report (National Infrastructure Commission, 2017) highlights that

Data is now as much a critical component of national infrastructure as steel, bricks and mortar. Data is part of infrastructure and needs maintenance in the same way that physical infrastructure needs maintenance.

Data infrastructure consists of data assets, technologies, standards, and guidance that inform their collection and use, and the organizations and communities that manage, use, and benefit from it (Dodds and Wells, 2019).

The concept of spatial data infrastructure dates back to 1993 (National Research Council, 1993). Since then, governments and the geospatial standards and data community have organized around developing, using and maintaining national data infrastructure to support use of spatial data. This has delivered a variety of economic benefits (ESRI, 2010). Assessment of the state of national spatial data infrastructure is now routinely used to review the marketplace for geospatial data and services (GeoBuiz, 2019).

While there are likely to be challenges in adoption of common standards and approaches (Lyubka and Temenoujka, 2017) a stronger focus on shared data infrastructure across the engineering and related sectors will help to ensure access to data, drive standards adoption, maintain quality, and ensure security.

3.2. Data must be stewarded

Stewardship involves the responsible management of a resource. When applied to data within an organization, stewardship involves a focus on managing data as an asset, to maximize its value while mitigating potential harms, within a well-defined data governance process that will include, for example, ethical and responsible approaches to data collection and use, maintaining quality, managing access to data, and ensuring legal compliance.

At an organizational level, stewardship of data will take a broader perspective with a goal of delivering responsible, trustworthy, and sustainable access to data across a business ecosystem, to an industry and society as a whole.

Good stewardship helps to make data discoverable and accessible. Stewardship requires a range of activities across the lifecycle of collecting, using, managing, and archiving of data.

The Gemini Principles, developed by the Center for Digital Built Britain emphasize the importance of value creation, curation, and public good benefits of data (CDBB, 2019). These are all important elements of good stewardship.

While there have been efforts to document and understand what good stewardship and management of engineering data looks like in a research context (Howard et al., 2010), these approaches are not widely adopted, especially outside of academia.

Lack of clear rights and responsibilities around the stewardship of data and digital resources leads to poor stewardship. Initiatives like Project 13 (Infrastructure Client Group, 2018) aim to address these issues by encouraging strong collaboration, better sharing of digital resources and data, and supporting the wider digital transformation required to support delivery of construction projects.

4. Opening and Sharing Data Unlocks Value

Data exists on a spectrum from closed, to shared, to open (Open Data Institute, 2017). As access to data increases, then more people can access data, allowing us to unlock more value from that same dataset. But,

further value can be unlocked by linking and combining this data with other data which is already more accessible.

The Royal Academy of Engineering data-sharing project (RAENG, 2018) illustrated a variety of benefits of increasing access to data, explored some practical challenges and provided guidance on approaching trustworthy sharing of data.

If, due to the economic characteristics of data, the market will be unable to realize its full value, then an alternative approach is required. This requires governments, businesses, and communities to be more intentional about deciding where on the data spectrum, different types of data should reside. Making data as open as possible, while protecting people's privacy, commercial confidentiality, and national security.

This is particularly important for "foundational" datasets, like geospatial data, which might typically be combined with a large number of other types of data assets.

Wherever data exists on the data spectrum, making it Findable, Accessible, Interoperable, and Reusable (FAIR, see Wilkinson et al., 2016) are important activities.

4.1. Explore new data sharing models

There are a wide variety of different technical, legal, practical, and institutional approaches that support increasing access to data (Open Data Institute, 2019b). These approaches include research data portals, data review boards, and a growing variety of institutional data stewardship models (Manohar et al., 2020), including data trusts (Open Data Institute, 2019c).

New technical approaches to sharing data are also emerging. These use a combination of cloud computing platforms, synthetic data (Thereaux, 2019), virtualization, and similar techniques to provide scalable, secure sharing of sensitive data. Practical applications include the variety of data commons approaches explored by Sage Bionetworks to support health research (Kellen, 2019) and the "data safe havens" framework (Alan Turing Institute, 2019).

Building appropriate incentives and support into funding models to ensure that researchers are able to make data FAIR and as open as possible, to comply with open access and open data mandates is also important.

4.2. Use challenges to drive innovation that solves problems

The open data movement has progressed from a standpoint of "open by default" to one of "publish with purpose" (Calderon, 2018). Greater impact has been achieved where opening up and sharing data has been driven by the needs of addressing a specific social, environmental, or economic challenge.

Clarity around which categories of data have higher value, for example, to address specific problems in international development (Open Data Institute, 2015) or as a means to drive economic growth (European Commission, 2020), is helping to prioritize further release of data. But, continued investment and engagement in challenge prizes and similar models can help to inspire innovative approaches (NESTA, 2019).

For example, the EU funded DataPitch program supported innovators in startups and academia to work on a variety of sector-wide and organization specific challenges. The program has helped to encourage collaboration across a wide range of different types of organization, overcoming a variety of data sharing challenges (London Economics, 2019b). The individual projects have led to cost savings from participating organizations and supported the development of new products and services (Open Data Institute, 2020).

In engineering, programs like the Construction Innovation Hub (Construction Innovation Hub, 2019), the Lloyd's Register Safety Accelerator (Lloyd's Register Foundation, 2019a), and the Data-centric engineering program (Alan Turing Institute, 2020) are demonstrating the value of a challenge lead approach.

4.3. Regulation must adapt to new technologies and uses of data

The pace of change around new approaches to data collection and applications of machine-learning and AI is creating a variety of regulator challenges, for example, in addressing privacy concerns, or in assuring the safety of digital, robotic, and autonomous systems.

Existing regulation of the engineering and related sectors will need to be extended or adapted to meet this changing environment. Drawing on recommendations from the Council for Science and Technology (Council for Science and Technology, 2018), a recent UK government white paper on modernizing regulation to support the industrial strategy, included the creation of a “Regulatory Horizons Council” as a means of supporting responses to this changing environment (BEIS, 2019).

The regulators themselves will also need to develop their own capacity to use data, their understanding of the increased role data plays in engineering and infrastructure projects, and an understanding of when, where and how to intervene in the sector. Approaches like the regulatory sandboxes being explored in the financial sector (FCA, 2015) illustrate one way in which regulators might work more closely with innovators to develop this understanding.

4.4. Building data literacy and skills

There is ongoing demand for data scientists and data engineering skills (RAENG, 2019). Meeting this need will be necessary for organizations to build the capacity to make use of data, and an understanding of how to safely share it.

But, the growing importance of data and digital skills in a variety of engineering professions (RAENG, 2020) and disciplines means that, while not everyone needs to be a data scientist, there is a need to increase data literacy and skills across a much wider audience.

This will require not just embedding the development of data skills into updated curricula in universities and apprenticeship schemes, but a review of existing professional development, certification and training programs to build necessary skills across the sector.

4.5. Ensure data is used legally and ethically

The engineering profession has always had strong ethical principles and codes of practice (NSPE, 2018). Concerns over applications of machine-learning, algorithmic decision making and uses of data has led to the publication of “ethical AI” principles from a range of organizations around the world. A recent review compared 36 different sets of principles to help identify areas of potential convergence (Fjeld et al., 2020).

Moving forward requires guidance that will help turn these principles into practices (Morley et al., 2019) that will inform the design and delivery of data projects. Tools like the Data Ethics Canvas (Open Data Institute, 2019d) are being successfully applied by a range of public and private sector organizations.

Researchers and data practitioners will need to stay abreast of an evolving legal and regulatory environment, to adapt their working practices, and deploy appropriate privacy preserving technologies to help comply with data protection regulations.

4.6. Share knowledge and insight

To deliver on our wider vision for unlocking value of data for public good requires more than just opening data. Sharing of data, code, models, guidance, and insights is all necessary to maximizing value of data for society. Open access, open source, and open data all have a role to play.

Using the full range of open approaches will be necessary to create value for the public good, and a world where data works for everyone.

5. Conclusion

In this paper, we have discussed how increasing access to data across the engineering and related sectors can help to address a variety of social, economic and environmental challenges. We have identified the need for action by a range of stakeholders to support behavior changes by creating the necessary capabilities, opportunities, and motivation to increase access to and use of data.

Our manifesto for change is intended to articulate a shared vision and help to build alignment across a range of existing initiatives and programs. We have discussed some of the rationale for the individual elements of the manifesto with reference to recent research and policy debates.

We are building on this work by working with a range of organizations across the engineering, built environment, transport and related sectors on practical projects. We welcome feedback on, and further endorsements of the manifesto.

We look forward to continuing to work with stakeholders across the engineering sector to support the creation of an open, trustworthy data ecosystem.

Funding Statement. This research was supported by a grant from the Lloyd's Register Foundation, Grant Number GA100182.

Competing Interests. The authors declare no competing interests.

Data Availability Statement. Data availability is not applicable to this article as no new data were created or analyzed in this study.

Authorship Contributions. Writing-original draft, L.D.; Writing-review & editing, L.D., P.L., J.M., and D.Y.

References

- Alan Turing Institute** (2019) Data Safe Havens in the Cloud. Available at <https://www.turing.ac.uk/research/research-projects/data-safe-havens-cloud> (accessed 4 May 2020).
- Alan Turing Institute** (2020) Data-Centric Engineering Challenges. Available at <https://www.turing.ac.uk/research/research-programmes/data-centric-engineering/data-centric-engineering-challenges> (accessed 4 May 2020).
- BEIS** (2019) Regulation for the Fourth Industrial Revolution. Available at <https://www.gov.uk/government/publications/regulation-for-the-fourth-industrial-revolution> (accessed 4 May 2020).
- Calderon A** (2018) Publishing with Purpose. Available at <https://medium.com/@opendatacharter/publishing-with-purpose-introducing-our-2018-strategy-ddbf7ab46098> (accessed 4 May 2020).
- CDBB** (2019) Gemini Principles. Available at <https://www.cdbb.cam.ac.uk/DFTG/GeminiPrinciples> (accessed 4 May 2020).
- Construction Innovation Hub** (2019) About the Construction Innovation Hub. Available at <https://constructioninnovationhub.org.uk/about/> (accessed 4 May 2020).
- Council for Science and Technology** (2018) Reforming the Governance of Technological Innovation. Available at <https://www.gov.uk/government/publications/reforming-the-governance-of-technological-innovation> (accessed 4 May 2020).
- Coyle D, Tennison J and Kay L** (2020) The Value of Data. Available at <https://www.bennettinstitute.cam.ac.uk/publications/value-data-summary-report/> (accessed 4 May 2020).
- Dodds L and Wells P** (2019) Data Infrastructure. *State of Open Data*. Available at <https://stateofopendata.od4d.net/chapters/issues/data-infrastructure.html> (accessed 4 May 2020).
- ESRI** (2010) Spatial Data Infrastructure. Available at <https://www.esri.com/library/bestpractices/spatial-data-infrastructure.pdf> (accessed 4 May 2020).
- European Commission** (2017) Final Results of the European Data Market Study Measuring the Size and Trends of the EU Data Economy. Available at <https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy> (accessed 4 May 2020).
- European Commission** (2020) European Legislation on Open Data and the Re-Use of Public Sector Information. Available at <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information> (accessed 4 May 2020).
- European Data Portal** (2020) The Economic Impact of Open Data: Opportunities for Value Creation in Europe. Available at <https://www.europeandataportal.eu/sites/default/files/the-economic-impact-of-open-data.pdf> (accessed 4 May 2020).
- FCA** (2015) Regulatory Sandbox. Available at <https://www.fca.org.uk/firms/innovation/regulatory-sandbox> (accessed 4 May 2020).
- Fane B, Ayris P, Hahnel M, Hrynaszkiewicz I, Baynes G and Farrell E** (2019) The State of Open Data Report 2019. Available at <https://doi.org/10.6084/m9.figshare.9980783.v2> (accessed 4 May 2020).
- Fjeld J, Achten N, Hilliglos H, Nagy A, Srikumar M** (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482 (accessed 4 May 2020).
- GeoBuiz** (2019) GeoBuiz 2019 Report. Available at <https://geobuiz.com/geobuiz-report-2019/> (accessed 4 May 2020).
- GODAN** n.d. Available at <https://www.godan.info/aboutgodan> (accessed 4 May 2020).
- Howard T, Darlington M, Ball A, Culley S and McMahon C** (2010) Understanding and Characterizing Engineering Research Data for its Better Management. Available at <https://researchportal.bath.ac.uk/en/publications/understanding-and-characterizing-engineering-research-data-for-it> (accessed 4 May 2020).
- Infrastructure Client Group** (2018) Project 13: From Transactions to Enterprises. Available at <http://www.p13.org.uk/> (accessed 4 May 2020).
- Kellen M** (2019) From Open Systems to Trusted Systems: New Approaches to Data Commons. Available at <https://sagebionetworks.org/in-the-news/from-open-systems-to-trusted-systems-new-approaches-to-data-commons/> (accessed 4 May 2020).
- London Economics** (2019a). Independent Assessment of the Open Data Institute's Work on Data Trusts and on the Concept of Data Trusts. Available at <https://theodi.org/wpcontent/uploads/2019/04/DataTrusts-economicfunction.pdf> (accessed 4 May 2020).

- London Economics** (2019b) Data Pitch Evaluation. Available at <https://datapitch.eu/news/data-pitch-impact-assessment-our-learnings/> (accessed 4 May 2020).
- Lloyd's Register Foundation** (2019a) Safety Accelerator Wins Award for Excellence in Technology Innovation. Available at <https://www.lr.org/en-gb/latest-news/safety-accelerator-wins-lloyds-list-award-for-excellence-in-data-and-technology-innovation/> (accessed 4 May 2020).
- Lloyd's Register Foundation** (2019b) Insight Report on Sharing Engineering Data. Available at <https://www.lrfoundation.org.uk/en/publications/insight-report-on-sharing-data/> (accessed 4 May 2020).
- Lyubka P, Temenoujka B** (2017) A brief overview of current status of European spatial data infrastructures – Relevant developments and perspectives for Bulgaria. *Geo-Spatial Information Science* 20(2), 97–108. <https://doi.org/10.1080/10095020.2017.1323524>
- Manohar S, Kapoor A and Ramesh A** (2020) Understanding Data Stewardship: Taxonomy and Use Cases. Available at <https://www.aapti.in/blog/data-stewardship-a-taxonomy> (accessed 4 May 2020).
- McKenna B** (2019) Open Data Programme OpenActive Ramps up Efforts to Get People Off the Sofa. Available at <https://www.computerweekly.com/news/252462423/Open-data-programme-OpenActive-ramps-up-efforts-to-get-people-off-the-sofa> (accessed 4 May 2020).
- McKinsey** (2018) Notes from the AI frontier: Modeling the Impact of AI on the World Economy. Available at <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy#part1> (accessed 4 May 2020).
- Michie S, van Stralen MM and West R** (2011) The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science* 6, 42. <https://doi.org/10.1186/1748-5908-6-42>.
- Morley J, Floridi L, Kinsey L and Elhalal A** (2019) From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-019-00165-5>
- National Infrastructure Commission** (2017) Data for the Public Good. Available at <https://www.nic.org.uk/publications/data-public-good/> (accessed 4 May 2020).
- National Research Council** (1993) *Toward a Coordinated Spatial Data Infrastructure for the Nation*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/2105>.
- NESTA** (2019) Challenge Prizes: A Practice Guide. Available at <https://www.nesta.org.uk/toolkit/challenge-prizes-a-practice-guide/> (accessed 4 May 2020).
- NSPE** (2018) NSPE Code of Ethics for Engineers. Available at <https://www.nspe.org/resources/ethics/code-ethics> (accessed 4 May 2020).
- Open Data Institute** (2015) How to Prioritise Open Data to Drive Global Development. Available at <https://theodi.org/article/how-to-prioritise-open-data-to-drive-global-development/> (accessed 4 May 2020).
- Open Data Institute** (2017) The Data Spectrum. Available at <https://theodi.org/about-the-odi/the-data-spectrum/> (accessed 4 May 2020).
- Open Data Institute** (2019a) Open Banking, Preparing for Lift Off. Available at <https://www.openbanking.org.uk/wp-content/uploads/open-banking-report-150719.pdf> (accessed 4 May 2020).
- Open Data Institute** (2019b) The Data Access Map. Available at <https://theodi.org/project/the-data-access-map/> (accessed 4 May 2020).
- Open Data Institute** (2019c) Data Trusts: Lessons from Three Pilots (Report). Available at <https://theodi.org/article/odi-data-trusts-report/> (accessed 4 May 2020).
- Open Data Institute** (2019d) The Data Ethics Canvas. Available at <https://theodi.org/article/data-ethics-canvas/> (accessed 4 May 2020).
- Open Data Institute** (2019e) A Manifesto for Sharing Engineering Data. Available at <https://theodi.org/article/engineering-data-for-the-public-good-a-manifesto/> (accessed 6 May 2020).
- Open Data Institute** (2020) Case Study: The Value of Sharing Data to Drive Open Innovation. Available at <https://theodi.org/article/case-study-the-value-of-sharing-data-to-drive-open-innovation/> (accessed 4 May 2020).
- RAENG** (2018) Towards Trusted Data Sharing: Guidance and Case Studies. Available at <https://www.raeng.org.uk/policy/publications/interactives/data-sharing> (accessed 4 May 2020).
- RAENG** (2019). Growing Demand for Data Science Leaves Britain Vulnerable to Skills Shortages. Available at <https://royalsociety.org/news/2019/05/data-science-skills-shortages/> (accessed 4 May 2020).
- RAENG** (2020). Global Engineering Capability Review. Available at <http://reports.raeng.org.uk/global-engineering-capability-review/cover/> (accessed 4 May 2020).
- Thereaux O** (2019) Anonymisation and Synthetic Data: Towards Trustworthy Data. Available at <https://theodi.org/article/anonymisation-and-synthetic-data-towards-trustworthy-data/> (accessed 4 May 2020).
- Wilkinson M, Dumontier M and Aalbersberg I** (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Cite this article: Dodds, L. L'Hénaff, P. Maddison, J. and Yates, D. 2020. A manifesto for increasing access to data in engineering. *Data-Centric Engineering*, 1: e5. doi:<https://doi.org/10.1017/dce.2020.3>