


Example article title: From individual control to collective data governance

First author¹*, Second author², Third author³, Fourth author¹, Fifth author², Sixth author³ and Seventh author¹

¹Department, Institution, City, State, Country. ²Department, Institution, City, State, Country. ³Department, Institution, City, State, Country.

*Corresponding author. E-mail: sample.email@institution.edu

Received: Blank date; **Revised:** Blank date; **Accepted:** Blank date

Keywords: data governance; data stewardship; emerging data platforms; health data; person generated health data

Abbreviations: EHR, Electronic Health Records; EU, European Union; GDPR, General Data Protection Regulation; HRB, Health Record Bank; ODI, Open Data Institute; PLM, Patients Like Me

Abstract

Abstracts should be 250 words. It must be able to stand alone and so cannot contain citations to the paper's references, equations, etc. An abstract must consist of a single paragraph and be concise. Because of online formatting, abstracts must appear as plain as possible.

Policy Significance Statement

Provide a 120 word statement here that summarises the significance of the work for policymakers, written at a level understandable for a broad audience.

1. Introduction

This is a template for a full paper submitted to the Data for Policy Conference 2024. It should aim for 6,000 words in length including references. The Introduction, Methods, Results, Discussion structure is not prescriptive: we have just included these as example headings. What follows is some sample text, equations and tables to help guide you in the writing of the paper.

Note that all full paper submissions should be submitted to the Data and Policy journal at Cambridge University Press via <https://mc.manuscriptcentral.com/dataandpolicy> for simultaneous consideration for the conference and journal.

Example text follows. More data can now be analyzed faster enabling researchers to explore more and more complex interactions. Data across a wide spectrum from open data¹ that can be downloaded, through licensed data,² to sensitive personal data can now contribute to our understanding of health, social, and economic outcomes.

¹<https://data.gov.uk/>

²<https://www.ukdataservice.ac.uk/>

The United Kingdom has a long tradition of safe data use by researchers. In the 50 years that the UK Data Archive has been making data available for social and economic research, there have been no damaging disclosures of personal information by academic researchers³. While increasing use of detailed and sometimes sensitive data can contribute valuable insights for targeting policies, we cannot be complacent. In order to support our policy needs and continue to use data safely and effectively, we need a research infrastructure that protects data confidentiality while enabling researchers to undertake innovative research.

A first step toward protecting sensitive data is to keep control of it, to disseminate access, not data. Data providers internationally are increasing, moving toward systems in which researchers remotely access and analyses data that are stored within a controlled environment, with any data that are removed from the system being checked in some way to minimize chances of disclosure.

This paper will look at remote data access as one method of protecting sensitive data while enabling its effective use for research. It will look at how the Digital Economy Act can strengthen the UK data research infrastructure and make some recommendations for other improvements

2. Methods

The results of data mining endeavors are majorly driven by data quality. Throughout these deployments, serious show-stopper problems are usually unresolved, such as data collection ambiguities, data imbalance, hidden biases in data, the lack of domain information, and data incompleteness. In a traditional data science lifecycle, outliers, bias, variance, boosting, over and under sampling, and data wrangling are measures that can be tuned to mitigate output quality issues and misinformation. In our work, we performed data collection from multiple sources to mitigate issues relevant to one dataset. Additionally, normalization, descriptive analytics, hyperparameter selection, and other quality-control methods are implemented.

2.1. Data Collection

Real-world data of this study are collected from Centers for Medicare and Medicaid Services (CMS) including data on patient complications and deaths, hospital general information, cost, and value of care for each hospital.

2.1.1. Method 1

RL is a subfield of AI; it involves the process of learning of an agent by using a trial and error approach that maximizes rewards by interacting with dynamic environments

Method 2

GA is a subfield of evolutionary algorithms; they are inspired by the process of biological natural selection.

3. Results

Method 1 is deployed through a real-time R-Shiny web application. The presented web application is to be used at a hospital for real-time decision-making support. The RL algorithm is connected to the application, and the model could be retrained in real time when new data are available. A hospital manager can use the “hospitalization ratio,” a parameter in clinical severity (aggregate of patients’ situation) for determining the amount of resources needed. The mentioned RL stages are represented as

³That is not to say that there have been no breaches—researchers using data for a project they are not licensed for or passing data to unlicensed colleagues—however, this, as far as we are aware of, has never caused harm to individuals or entities in the data.

scales based on transmissibility and clinical severity. The outcomes (i.e., actions) are assigned as either “Idle” or “Share”—which represents the recommendation that the RL algorithm generates.

3.1. Validation

3.1.1. Limitatins

Another heading

In the 50 years that the UK Data Archive has been making data available for social and economic research, there have been no damaging disclosures of personal information by academic researchers. While increasing use of detailed and sometimes sensitive data can contribute valuable insights for targeting policies, we cannot be complacent. In order to support our policy needs and continue to use data safely and effectively, we need a research infrastructure that protects data confidentiality while enabling researchers to undertake innovative research.

4. Equations

Equations in \LaTeX can either be inline or on-a-line by itself. For inline equations use the $\$ \dots \$$ commands. Eg: The equation $H\psi = E\psi$ is written via the command $H\psi = E\psi$.

For on-a-line by itself equations (with auto generated equation numbers) one can use the equation or eqnarray environments D .

$$\mathcal{L} = i\psi\gamma^\mu D_\mu\psi - \frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu} - m\psi\psi \quad (1)$$

where,

$$\begin{aligned} D_\mu &= \partial_\mu - ig\frac{\lambda^a}{2}A_\mu^a \\ F_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^a \end{aligned} \quad (2)$$

Notice the use of `\nonumber` in the align environment at the end of each line, except the last, so as not to produce equation numbers on lines where no equation numbers are required. The `\label{}` command should only be used at the last line of an align environment where `\nonumber` is not used.

$$Y_\infty = \left(\frac{m}{\text{GeV}}\right)^{-3} \left[1 + \frac{3\ln(m/\text{GeV})}{15} + \frac{\ln(c_2/5)}{15}\right] \quad (3)$$

The class file also supports the use of `\mathbb{}`, `\mathscr{}` and `\mathcal{}` commands. As such `\mathbb{R}`, `\mathscr{R}` and `\mathcal{R}` produces \mathbb{R} , \mathscr{R} and \mathcal{R} respectively.

5. Tables

Tables can be inserted via the normal table and tabular environment. To put footnotes inside tables one has to use the additional “fntable” environment enclosing the tabular environment. The footnote appears just below the table itself.



Figure 1. This is a widefig. This is an example of long caption this is an example of long caption this is an example of long caption this is an example of long caption.

6. Figures

As per the L^AT_EX standards eps images in latex and pdf/jpg/png images in pdflatex should be used. This is one of the major differences between latex and pdflatex. The images should be single page documents. The command for inserting images for latex and pdflatex can be generalized. The package that should be used is the graphicx package.

7. Cross referencing

Environments such as figure, table, equation, align can have a label declared via the \label{#label} command. For figures and table environments one should use the \label{} command inside or just below the \caption{} command. One can then use the \ref{#label} command to cross-reference them. As an example, consider the label declared for Figure 1 which is \label{fig1}. To cross-reference it, use the command Figure \ref{fig1}, for which it comes up as “Figure 1”. The reference citations should used as per the "natbib" packages. Some sample citations: [Herbst-Damm and Kulik \(2005\)](#); [Gilbert et al. \(2004\)](#); [Light and Light \(2008\)](#); [Shotton \(1989\)](#); [Schiraldi \(2001\)](#); [VandenBos \(2007\)](#); [Freud \(1953\)](#).

8. Lists

List in L^AT_EX can be of three types: enumerate, itemize and description. In each environments, new entry is added via the \item command. Enumerate creates numbered lists, itemize creates bulleted lists and description creates description lists. List in L^AT_EX can be of three types: enumerate, itemize and description. In each environments, new entry is added via the \item command. Enumerate creates numbered lists, itemize creates bulleted lists and description creates description lists.

- 1. This is the 1st item
- 2. Enumerate creates numbered lists, itemize creates bulleted lists and description creates description lists.

Table 1. Tables which are too long to fit, should be written using the “table*” environment as shown here

Projectile	Element 1			Element 2 ¹		
	Energy	σ_{calc}	σ_{expt}	Energy	σ_{calc}	σ_{expt}
Element 3	990 A	1168	1547 ± 12	780 A	1166	1239 ± 100
Element 4	500 A	961	922 ± 10	900 A	1268	1092 ± 40

Note: This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote

¹This is an example of table footnote

3. Numbered lists continue.

List in \LaTeX can be of three types: enumerate, itemize and description. In each environments, new entry is added via the `\item` command.

- This is the 1st item
- Itemize creates bulleted lists and description creates description lists.
- Bullet lists continue.

9. Conclusion

Some Conclusions here.

Provenance. This article was submitted for consideration for the 2024 Data for Policy Conference to be published in Data and Policy on the strength of the Conference review process.

Acknowledgments. We are grateful for the technical assistance of A. Author.

Funding Statement. This research was supported by grants from the <funder-name><doi>(<award ID>); <funder-name><doi>(<award ID>).

Competing Interests. A statement about any financial, professional, contractual or personal relationships or situations that could be perceived to impact the presentation of the work — or ‘None’ if none exist

Data Availability Statement. A statement about how to access data, code and other materials allowing users to understand, verify and replicate findings — e.g. Replication data and code can be found in Harvard Dataverse: <https://doi.org/link>.

Author Contributions. A.A. and A.B.C. designed the study, abstracted the data wrote the first draft, and approved the final version of the manuscript. A.R.E.J., M.R.L., K.L.S., and A.D.P. revised the manuscript and approved the final version.

Additional Material. Description and links to any other relevant online materials, such as videos, presentations, papers hosted in the Data for Policy Zenodo site.

References

- Freud S** (1953) *The Method of Interpreting Dreams*, volume 4, pages 96–121. American Psychological Association.
- Gilbert DG, McClernon JF, Rabinovich NE, Sugai C, Plath LC, Asgaard G, Dickinson D and Botros N** (2004) Effects of quitting smoking and EEG activation and attention last for more than 31 days and are more severe with stress, dependence, DRD2 A1 allele, and depressive traits. *Nicotine and Tobacco Research* 6, 249–267.
- Herbst-Damm KL and Kulik JA** (2005) Volunteer support, marital status, and the survival times of terminally ill patients. *Health Psychology* 24, 225–229.
- Light MA and Light IH** (2008) The geographic expansion of Mexican immigration in the United States and its implications for local law enforcement. *Law Enforcement Executive Forum Journal* 8(1), 73–82.
- Schiraldi GR** (2001) *The Post-traumatic Stress Disorder Sourcebook*. Springer.
- Shotton MA** (1989) *Computer Addiction? A Study of Computer Dependency*. Taylor & Francis.
- VandenBos GR**, editor (2007) *APA Dictionary of Psychology*. American Psychological Association.