



Cabinet Office

Draft Data Science Ethical Framework

December 2015

Note: this is an early draft document and not formal government policy. It is designed to be iterated as it is used, and is shared in the expectation that it will encourage feedback and further improvement

data.blog.gov.uk

Data science and ethics introduction

Who is this guidance for?

This guidance supports those working with data and those making policy or operational decisions with data to feel confident to innovate. It balances the use of new data and techniques with respect for privacy and making sure no-one is *unfairly* adversely affected.

Why is guidance needed for data science?

Data science is a new practice for government which provides new opportunities to create insight and improve public services. Digital advances are both producing huge amounts of new forms of data, and allowing computers to more quickly process these and make decisions without human oversight. New opportunities, and many new challenges we have not had to consider before.

The Data Protection Act sets out some important principles about how you can use data. There are also standards for the quality and integrity of data processes in the analytical, health and other professions. Those working with data should be aware of how these and how they support the use of data science. This guidance ensures that these principles are met in the new rapidly evolving data landscape, so that people are confident to use the law to innovate and use data science.

There are huge opportunities to use data to improve insight for better policymaking and to make government operations more efficient. However, the public will often not be aware of what these techniques are and how they could affect them. These perceptions and attitudes may change over time. Working with data in a way which makes the public feel uneasy, without adequate transparency or engagement, could put your project at risk and also jeopardise other projects across government.

This guidance helps you navigate the legal aspects applicable to data science, and also helps you think through some of the ethical issues which sit outside the law.

How to use the guidance

Data science projects have a number of stages; discovery work to explore what it is possible to do with the data, the actual delivery work, refining the accuracy of the insight, and the ongoing use of that insight by policymakers or operational staff. This guidance will help you think through the methodology of a project and ask appropriate questions about how the project is conducted at each stage.

The guidance is based around six principles. **Fundamentally, the public benefit of doing the project needs to be balanced against the risks of doing so.**

- 1 Start with clear user need and public benefit
- 2 Use data and tools which have the minimum intrusion necessary
- 3 Create robust data science models
- 4 Be alert to public perceptions
- 5 Be as open as possible
- 6 Keep data secure

The guidance we are sharing today is very high level and we are working with departments to develop a more detailed checklist and case studies so people can clearly see how ethical issues play out.

The guidance is iterative, and is being developed with feedback from departments and external stakeholders. It also complements other ethical frameworks for analysis such as those relating to health data and from the [National Statistician](#).

Data science guidance: summary

1 Start with clear user need and public benefit

Data science offers huge opportunities to create evidence for policymaking, and make quicker and more accurate operational decisions. Being clear about the public benefit will help you justify the sensitivity of the data (principle 2) and the method (principle 3) that you want to use.



2 Use data and tools which have the minimum intrusion necessary

You should always use the minimum data necessary to achieve the public benefit. Sometimes you will need to use sensitive personal data. There are steps that you can take to safeguard people's privacy e.g. de-identifying or aggregating data to higher levels, querying against datasets or using synthetic data.



3 Create robust data science models

Good machine learning models can analyse far larger amounts of data far quicker and accurately than traditional methods. Think through the quality and representativeness of the data, flag if algorithms are using protected characteristics (e.g. ethnicity) to make decisions and think through unintended consequences. Complex decisions may well need the wider knowledge of policy or operational experts.



4 Be alert to public perceptions

The Data Protection Act requires you to have an understanding of how people would reasonably expect their personal data to be used. You need to be aware of shifting public perceptions. Social media, commercial data and data scraped from the web allow us to understand more about the world, but come with different terms and conditions and levels of consent.



5 Be as open as possible

Being open allows us to talk about the public benefit of data science. Be as open as you can about the tools, data and algorithms (unless doing so would jeopardise the aim, e.g. fraud). Provide explanations in plain English and give people recourse to decisions which they think are incorrectly made. Make sure your project has oversight and accountability built in throughout.



6 Keep data secure

We know that the public are justifiably concerned about their data being lost or stolen. Government has a statutory duty to protect the public's data and as such it is vital that appropriate security measures are in place.



DRAFT - NOT A STATEMENT
OF GOVERNMENT POLICY