



Department for
Digital, Culture,
Media & Sport

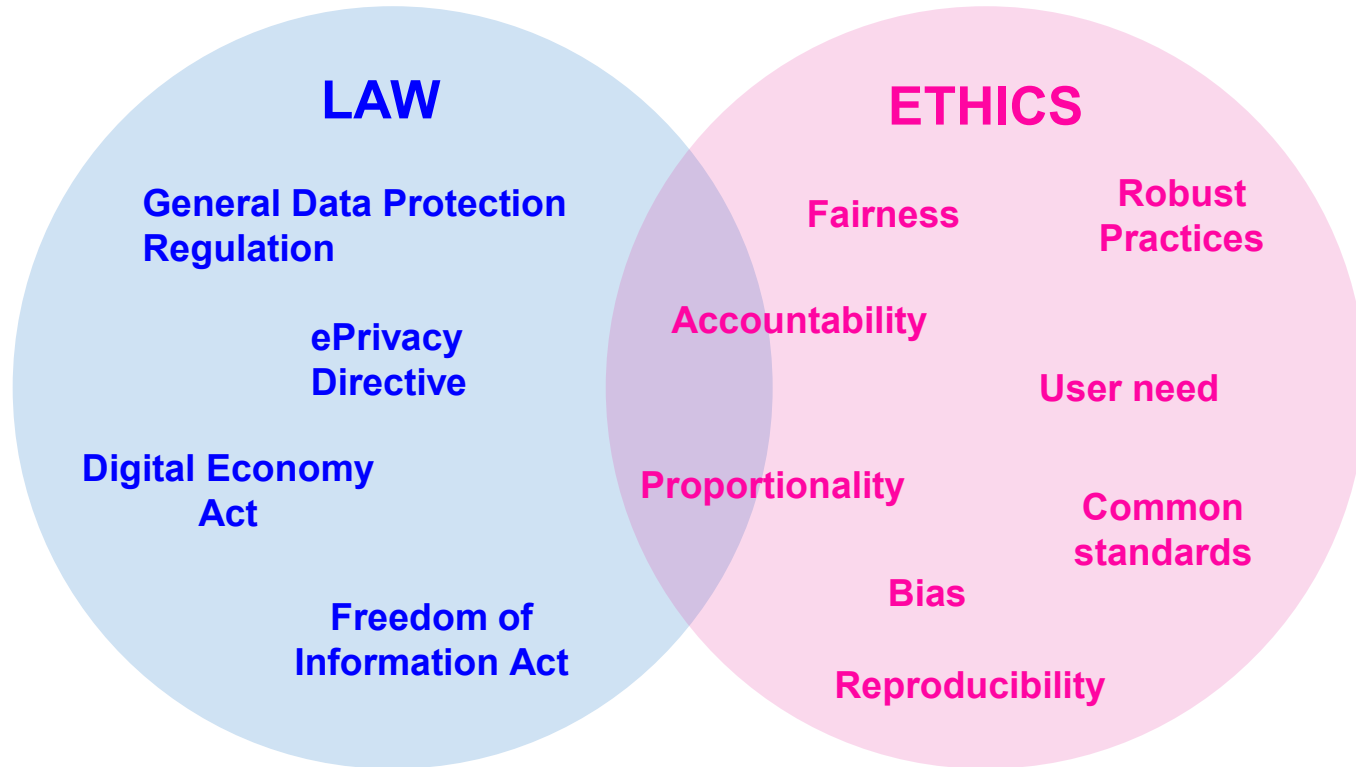
Data Ethics in the Public Sector - Data Ethics Framework

sarah.gates@culture.gov.uk

@gates_sj

DCMS

Data ethics in the public sector



Less than a tenth (8%) of the UK public say they have a good understanding of how their information is made available to others.



May 2016

Public dialogue on the ethics of data science in government

Although there was limited awareness of data science, there was broad support for the need for government to find new ways of using data.

People are open to you trying to tackle most policy challenges but they care about how you do it



Cabinet Office

Data Science Ethical Framework

Data science carries both huge opportunities and a duty of care. Technology is changing so rapidly; as are the public's views. In this new and changing landscape, this document is not about creating additional hurdles, but rather about making innovation easier. It does this by bringing together the relevant law in the context of new technology, and prompting consideration of public reaction so that government data scientists and policymakers can be confident to innovate appropriately with data.

Developing the ethics around data science can't be done by government alone. This framework is a first iteration - a beta, if you like - of a set of principles wider than the legal framework, to help stimulate innovative and responsible action.

I look forward to listening to, and participating in that debate.

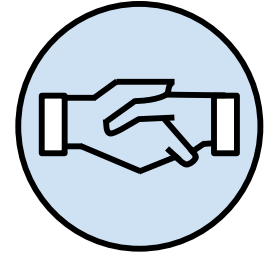
The Rt Hon Matt Hancock MP
Minister for the Cabinet Office and
Paymaster General

- to advise data scientists on the appropriate use of data beyond simply complying with data protection legislation
- to clearly communicate the standards we expect data projects to follow and demonstrate this to the public to maintain trust.



**We have taken an
open policy making
approach**

Using feedback to update the framework: Essex Partners



The Application

- Essex Partners are investigating whether they can develop an approach to predicting risk which might enable early intervention on social problems.
- As the information is personal and relates to vulnerable groups, Essex Partners recognised the importance of using data ethically.

The Outcome

- The framework has helped Essex Partners analyse the ethical issues around their domestic abuse prototype and Gangs, Violence and Vulnerability prototype.
- It led to the decision to use combined datasets to ensure more accurate analytical output.

Data Ethics Principles

- 1 Start with clear user need
- 2 Be aware of relevant legislation and codes of practice
- 3 Use data that is proportionate to the user need
- 4 Understand the limitations of the data
- 5 Use robust practices and work within your skillset
- 6 Make your transparent and be accountable
- 7 Embed data use responsibly

Guidance
Data Ethics Framework
 Updated 30 August 2018

Contents

- [Ministerial Foreword](#)
- [Introduction](#)
- [How to use the Data Ethics Framework](#)
- [The Data Ethics Framework principles](#)
- 1. [Start with clear user need and public benefit](#)
- 2. [Be aware of relevant legislation and codes of practice](#)
- 3. [Use data that is proportionate to the user need](#)
- 4. [Understand the limitations of the data](#)
- 5. [Ensure robust practices and work within your skillset](#)
- 6. [Make your work transparent and be accountable](#)
- 7. [Embed data use responsibly](#)
- [The Data Ethics Workbook](#)

Ministerial Foreword



Making better use of data offers huge benefits, in helping us provide the best possible services to the people we serve.

However, all new opportunities present new challenges. The pace of technology is changing so fast that we need to make sure we are constantly adapting our codes and standards. Those of us in the public sector need to lead the way.

Guidance

5. Use robust practices and work within your skillset

How to implement principle 5 of the Data Ethics Framework for the public sector.

Published 13 June 2018
 From: [Department for Digital, Culture, Media & Sport](#)

Contents

- [Multidisciplinary teams](#)
- [Getting help from experts](#)
- [Accountability of algorithms](#)
- [Social bias in algorithms](#)
- [Reproducibility](#)
- [Test the model](#)

Related content

- Collection
- [Data ethics guidance](#)

Accountability of algorithms

You should always use the most simple model to achieve the desired outcome.

As machine learning algorithms are designed to learn and develop semi-independently, the processes used can become more difficult to interpret and understand. Teams need to have a reasonable understanding of how the machine learning or pipeline of machine learning models has worked to meet the user need.


You must be able to explain this to non-specialists.

You can design your machine learning process to improve accountability.

Staged process

This involves tackling prediction tasks as a pipeline of machine learning models which should facilitate the potential for overall interpretability. The staged process must be clearly documented with all necessary assumptions and caveats articulated.

Often we do not need to know exactly how a machine learning model has arrived at a particular result if we can show logical steps to achieving the particular outcome. This includes exactly what training data was used and which variables have contributed most to a particular result.

A solid pink vertical bar is located on the left side of the slide, extending from the top to the bottom.

The framework is aimed at anyone working directly or indirectly with data in government, including data practitioners (statisticians, analysts and data scientists), policymakers, operational staff and those helping produce data-informed insight.

Data Ethics Workbook

Data Ethics Framework						
	0	1	2	3	4	5
1. Start with clear user need	User need is not well defined					User need is clearly defined
Description of the user need with supporting evidence						
2. Be aware of relevant legislation and codes of practice	Needs clarification or expert input					Relevant laws are well understood
List the pieces of legislation, codes of practice and guidance that apply to your project.						
3. Use data which is proportionate to the user need	Reuse not proportionate					Reuse of data is clearly proportionate to achieve user need
Describe how the data being used is proportional to the user need						
4. Understand the limitations of the data	Unreliable, unsuitable data					Data is representative and accurate
Identify the potential limitations of the data source(s) and how they are being mitigated						
5. Use robust practices and work within your skillset	Needs further expert input					Methodologies clearly designed and understood
Explain the relevant expertise and approaches that are being employed to maximise the efficacy of the project						
6. Make your work open and be accountable	No scrutiny or peer review available					Oversight built in through life cycle of project
Describe how you have considered making your work open and accountable						
7. Embed data use responsibly	No ongoing plan determined					Evaluation plan developed and resource in place to deliver it
Describe the steps taken to ensure any new model, policy or service is managed responsibly						

Principle 5: Use robust practices and work within your skillset

To consider:

Explain the relevant expertise and approaches that are being employed to maximise the efficacy of the project

Describe the disciplines involved and why.

Are there expertise that the project requires that you don't currently have?

Have you designed the approach with a policy team or subject matter expert(s)?

Has all subject matter context, from policy experts or otherwise, been taken into account when determining the appropriate loss function for the model?

If necessary, how can you (or with external scrutiny) check that the algorithm is achieving the right output decision when new data is added?

How has reproducibility been ensured? Could another analyst repeat your procedure based on your documentation?

How confident are you that the algorithm is robust, and that any assumptions are met?

What is the quality of the model outputs, and how does this stack up against the project objectives?

If using data about people, is it possible that a data science technique is basing analysis on proxies for protected variables which could lead to a discriminatory policy decision?

Principle 6: Make your work transparent and be accountable

To consider:

Describe how you have considered making your work transparent and accountable

Have you spoken to your organisation to find out if you can speak about your project openly?

Have you considered how both internal and external engagement could benefit your project?

How interpretable are the outputs of your work?

How are you explaining how approaches were designed in plain English to other practitioners, policy makers and if appropriate, the public?

Can you openly publish your methodology, metadata about your model, and/or the model itself e.g. on Github?

Can you get peers to review your Pull Requests?

Procurement

To consider:

Describe how the technology you are procuring is conducive to the principles of this framework:

Can you answer the questions in the Data Ethics Workbook?

What was the original intended use of the software?

How has the algorithm been trained?

What data was used to train the algorithm?

How interpretable is the algorithm?

How has the algorithm been tested for different failure modes relevant to the intended task?

What are the plans for re-assessing the algorithm's performance at set times?

Are assessment methods, e.g. for bias, compatible with the software being procured?

Do assessment methods require the support of the provider, or can they be carried out independently?

Can you scrutinise the performance of the software free from the constraints of intellectual property restrictions?



Department for
Digital, Culture,
Media & Sport

Thank you for listening!

sarah.gates@culture.gov.uk

@gates_sj

DCMS