



Data Ethics Framework

Workbook

v1-0

What is data ethics? Data ethics is an emerging branch of applied ethics which describes the value judgements and approaches we make when generating, analysing and disseminating data. This includes a sound knowledge of data protection law and other relevant legislation, and the appropriate use of new technologies. It requires a holistic approach incorporating good practice in computing techniques, ethics and information assurance.

How to use the Data Ethics Framework: The Data Ethics Framework guides the design of appropriate data use in policing and the wider public sector. This guidance is aimed at anyone working directly or indirectly with data within the City of London Police, including data practitioners (statisticians, analysts and data scientists), policymakers, operational staff and those helping produce data-informed insight. The Data Ethics Framework builds on the core values of the Police Code of Ethics - honesty, integrity, confidentiality, equality and diversity - to encourage ethical data use to build better services and inform policy.

Teams should work through the framework together before starting the design or discovery phase of a new project, policy or service. Use the workbook to consider legal and ethical questions to inform the best approach for your use of data.

Each part of the framework is designed to be regularly revisited throughout your project, especially when any changes are made to your data collection, storage, analysis or sharing processes.

For advice support and guidance, please contact Information Management Service - IMSupport@

Data sources

Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.

Is any personal data involved, or data that is otherwise sensitive?

Limitations in data sources

Are there limitations that could influence your project's outcomes?

Consider:

- > bias in data collection, inclusion/exclusion, analysis, algorithms
- > gaps or omissions in data
- > provenance and data quality
- > other issues affecting decisions, such as team composition

Sharing data with others

Are you going to be sharing data with other organisations? If so, who?

Are you planning to publish any of the data? Under what conditions?

Ethical and legislative context

What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?

Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).

Rights around data sources

Where did you get the data from? Is it produced by an organisation or collected directly from individuals?

Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?

Your reason for using data

What is your primary purpose for collecting and using data in this project?

What are your main use cases? What is your business model?

Are you making things better for society? How and for whom?

Are you replacing another product or service as a result of this project?

Communicating your purpose

Do people understand your purpose – especially people who the data is about or who are impacted by its use?

How have you been communicating your purpose? Has this communication been clear?

How are you ensuring more vulnerable individuals or groups understand?

Positive effects on people

Which individuals, groups, demographics or organisations will be positively affected by this project? How?

How are you measuring and communicating positive impact? How could you increase it?

Negative effects on people

Who could be negatively affected by this project?

Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?

How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organisations using the data.

Minimising negative impact

What steps can you take to minimise harm?

How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure?

How are you measuring, reporting and acting on potential negative impacts of your project?

What benefits will these actions bring to your project?

Engaging with people

How can people engage with you about the project?

How can people correct information, appeal or request changes to the product/service? To what extent?

Are appeal mechanisms reasonable and well understood?

Openness and transparency

How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements?

Can you ask peers for feedback on the project? How will you communicate it internally?

Will you publish your actions and answers to this canvas openly?

Ongoing implementation

Are you routinely building in thoughts, ideas and considerations of people affected in your project? How?

What information or training might be needed to help people understand data issues?

Are systems, processes and resources available for responding to data issues that arise in the long-term?

Reviews and iterations

How will ongoing data ethics issues be measured, monitored, discussed and actioned?

How often will your responses to this canvas be reviewed or updated? When?

Your actions

What actions will you take before moving forward with this project? Which should take priority?

Who will be responsible for these actions, and who must be involved?

Will you openly publish your actions and answers to this canvas?

Data Ethics Framework

	0	1	2	3	4	5
1. Start with clear user need and public benefit	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 that 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 transparent 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 transparent 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 1: Start with a clear user need and public benefit

To consider:

Describe the user need:	
Does everyone in the team understand the user need?	
How does this benefit the public?	
What would be the harm in not using data science - what needs might not be met?	
Do you have supporting evidence for the approach being likely to meet a user need or provide public benefit?	

Principle 2: Be aware of relevant legislation and codes of practice

To consider:

List the pieces of legislation, codes of practice and guidance that apply to your project:	
Do all team members understand how relevant laws apply to the project?	
If necessary, have you consulted with relevant experts?	
Have you spoken to your information assurance team?	
If using personal data, do you understand obligations under data protection legislation?	

Principle 3: Use data that is proportionate to the user need

To consider:

If using personal data, have you answered the questions for determining proportionality? You must include evidence to support any decision.	
If using personal data, what measures are in place to control access? How widely are you searching personal data?	
How can you meet the project aim using the minimum personal data possible?	
Is there a way to achieve the same aim with less identifiable data?	
Can you use synthetic data?	
Has the data being used been provided for your analysis?	
By using data that the public have freely volunteered, would your project jeopardise people providing this again in the future?	
Could you clearly explain why you need to use that data to members of the public?	
Is there a fair balance between the rights of individuals and the interests of the community?	

Principle 4: Understand the limitations of the data

To consider:

Identify the potential limitations of the data source(s) and how they are being mitigated:	
What data source(s) is being used?	
Are all metadata and field names clearly understood?	
What processes do you have in place to ensure and maintain data integrity?	
Is there a plan in place to identify errors and biases?	
What are the caveats?	

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?	

Principle 7: Embed data use responsibly

To consider:

Describe the steps taken to ensure any insight is managed responsibly:	
How many people will be affected by the new model, insight or service?	
Who are the users of the insight, model, or new service?	
Do users have the appropriate support and training to maintain the new technology?	
Have future events been planned for?	
Is your implementation plan correlated with the impact of a particular model?	
How often will you report on these plans to senior reporting officers?	