UK Statistics Authority

Version 2.3

National Statistician’s Data Ethics Advisory Committee

Guidelines on using the ethics self-assessment process

Introduction

To help maintain public trust and confidence in the use of data for research and statistics it is important that researchers (in ONS, the wider Government Statistical Service, and beyond) whatever their particular discipline (statistics, economics, social research, operational research, other) who use data for statistical, analytical and wider research purposes do not just consider what can be done with the data, methods, expertise and technology available to them. It is equally important that researchers consider what *should* be done and how it *should* be done. This ethical self-assessment has been developed to provide a framework to help all researchers to think about the ethics of their research.

This guidance is designed to support researchers and statisticians to complete the National Statistician's Data Ethics Advisory Committee's (NSDEC) ethical self-assessment form. NSDEC's ethical self-assessment enables researchers to self-assess the ethics of their research by scoring their research against NSDEC's ethical principles.

The NSDEC's ethical principles are:

1. The use of data has clear benefits for users and serves the public good
2. The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately
3. The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality
4. Data used and methods employed are consistent with legal requirements such as Data Protection Legislation[1], the Human Rights Act 1998, the Statistics and Registration Service Act 2007 the common law duty of confidence, and the Equality act 2010.
5. The views of the public are considered in light of the data used and the perceived benefits of the research
6. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

---

[1] “Data Protection Legislation” means the full, applicable data protection framework as set out in the Data Protection Act 2018. This encompasses general processing (including the General Data Protection Regulation and the applied GDPR).

This self-assessment process is designed to provide researchers with an easy-to-use framework to consistently review the ethics of their projects. Although the self-assessment is not designed to automatically resolve the ethical issues in individual projects, it is designed to help identify an accurate and consistent estimation of the “ethical risks” of research proposals. Identification of any potential ethical risks should be used to shape discussions that will drive improvements in research proposals and activities. Ensuring that researchers and statisticians continuously consider research in light of the NSDEC’s ethical principles will ensure that the use of data for research and statistical purposes continues to be ethical and for the public good.

How to use the self-assessment

We recommend that self-assessments are conducted as early as possible in the project timeline, as this will help to determine and ensure the most ethically sound route for research. We also advise that you revisit the self-assessment throughout the project lifecycle to ensure that any changes to the proposed project are considered in light of the ethical principles.

Although this framework is presented as a self-assessment, it need not be a process that you complete on your own. It is also important to remember that the self-assessment process is designed to consider the ethics of your particular project – therefore, it is still the researcher’s responsibility to ensure that the project satisfies all of the relevant legal requirements relating to their project. We therefore recommend that you discuss your research projects and/or self-assessment form with the following (where relevant and appropriate):

1. Senior director/manager of your branch/business area responsible for the research project
2. The relevant Information Asset Owner(s)
3. Legal Services team and (where different) the Data Protection team
4. Communications and Media relations team

All completed self-assessments should be sent to the UK Statistics Authority’s Data Ethics team, at Data.Ethics@statistics.gov.uk.

To help you navigate through the process we have included a user checklist at the end of this document.

The self-assessment form

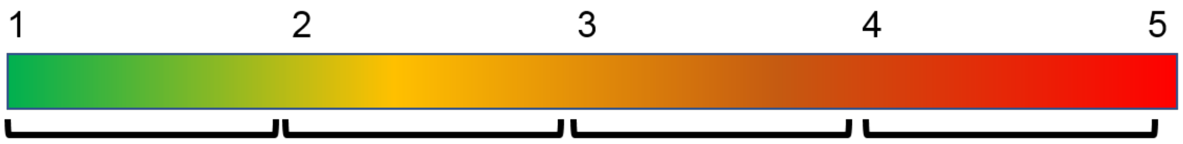
The self-assessment form consists of 3 main sections:

1. **Basic Information**
2. **Weightings and sensitive research areas**
3. **Item scoring scales.**

Information and guidance for completing each of these sections is provided in the next three sections.

The self-assessment tool calculates the average ethical risk score for the research project and provides you, as the researcher, with a suggested course of action based on this score. Whatever the course of action, you should share your completed self-assessment form with the Data Ethics team, at Data.Ethics@statistics.gov.uk before proceeding with the project.

The suggested courses of action are as follows:



Low risk	Average risk	High risk	Very high risk
Project may proceed after confirmation from the Data Ethics team	Consult with the Data Ethics team to discuss actions to mitigate any highlighted risks before proceeding with the project	Consult with the Data Ethics team. If risks cannot be mitigated then this project should be presented to NSDEC for a full independent ethical review before proceeding	Consult with the Data Ethics team. If risks cannot be mitigated then this project should be presented to NSDEC for a full independent ethical review before proceeding

1. Basic information

This section consists of four parts: 1) Project title; 2) Project timeline; 3) Project summary; 4) Public benefit.

Project title
Please provide a short title to describe the project. Please make sure that the title is indicative of the project.

Project timeline
Please provide some details about your project timeline. This should include key dates such as the start and end date of your project, as well as any key dates for dissemination activities (such as project reports and outputs).

Project summary
Please provide a short summary of the project. This should include the following information (where relevant):

1. Project partners and/or sponsors
2. Research aims
3. Methods proposed (for data collection and analysis)
4. A list of data sources and variables (with a justification for each)
5. Plans for dissemination of research findings
6. Any useful and relevant background information

Public benefit

The Statistics and Registration Service Act 2007 sets out the statutory objective of the UK Statistics Authority to promote and safeguard the production and publication of official statistics that serve the public good. Public good is defined as:

- informing the public about social and economic matters;
- assisting in the development and evaluation of public policy; and
- regulating quality and publicly challenging the misuse of statistics.

The Research Code of Practice and Accreditation Criteria states that for a project to secure accreditation, the primary purpose of the project must be to serve the public interest in one or more of the following ways, to:

- provide an evidence base for public policy decision-making;
- provide an evidence base for public service delivery;
- provide an evidence base for decisions which are likely to significantly benefit the economy, society or quality of life of people in the UK, UK nationals or people born in the UK now living abroad;
- replicate, validate, challenge or review existing research and proposed research publications, including official statistics;
- significantly extend understanding of social or economic trends or events by improving knowledge or challenging widely accepted analyses; and/or,
- improve the quality, coverage or presentation of existing research, including official or National Statistics.

Please therefore use this section to describe how undertaking this research will benefit the public. Please go beyond the statistical outputs that will be produced to explain the public benefits of producing these statistics. What impact is your research going to have on people's lives? See our guidance on considering and articulating public good in research projects for further information and help in completing this section.

2. Weightings and sensitive research areas

There are five characteristics (listed below) which determine the ethical risk of a project, with 'ethical risk' being defined as the 'negative consequences of unethical actions'. To measure the differential complexities of various ethical decisions these have been included in the self-assessment as weighted measures. The weights have been reviewed by NSDEC based on its consideration of a large number of projects since 2015.

Data linkage projects

Linking data can lead to useful insights and offers new opportunities for existing datasets. As different information about a data subject is pulled together, linking data increases the risk of re-identification of individual data subjects. Data linkage may be also perceived as profiling, and hence might not be publicly acceptable.

Sensitive personal data and processing

Personal data means any information relating to an identifiable person who can be directly or indirectly identified, in particular by reference to an identifier. This definition provides for a wide range of personal identifiers to constitute personal data, including name, identification number, location data or online identifiers. Sensitive personal data are special categories of personal data as defined in law. These special categories include personal data on racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data and biometric data (where processed to uniquely identify an individual), data concerning health, and data concerning a person's sex life or sexual orientation. Due to the risk of disclosing the identity of data subjects along with other personal information, it is important that researchers put in place additional safeguards. This is mandated by law (Data Protection Act 2018, and the EU General Data Protection Regulation).

Patient level health data

There are particular sensitivities when using patient level clinical (health) data for research and statistics. Consideration needs to be given to the public acceptability of using such data and respecting patient confidentiality.

Research including children and vulnerable adults

A 'child' means any person who is under the age of 18, and a 'vulnerable adult' means any person aged 18 or over, who for any reason, is unable to take care of themselves, or unable to protect themselves from harm or exploitation. Research which includes data on children and vulnerable adults can lead to a substantial societal benefit. However, such research requires researchers to be more vigilant of ethical risks especially around public acceptability. It is important that the project considers any risks to the organisational reputation as well as some public engagement or public acceptability testing. See our guidance on considering public views and engagement for research and statistics projects for further information on public acceptability.

Data sources

Legal and ethical frameworks are not equally established for all data sources and technologies, and issues around the data quality, privacy and whether (or not) a data subject's consent is required, might arise by the collection, use and sharing of specific categories of data. These sources currently include social media, rich media sources, metadata or paradata, web-scraped data, and big data (including sensor and mobile data). For further information on ethical considerations when using different types of data, see our high-level ethics checklist for third party data, our guidance on location data and our guidance on the use of machine learning techniques.

Weights have been developed to take account of these complexities in the self-assessment process and are applied to the overall self-assessment score. As legislation, regulation, and methodology around these areas evolve we will review these weights. Some weights may be adjusted, and new weight categories may be introduced.

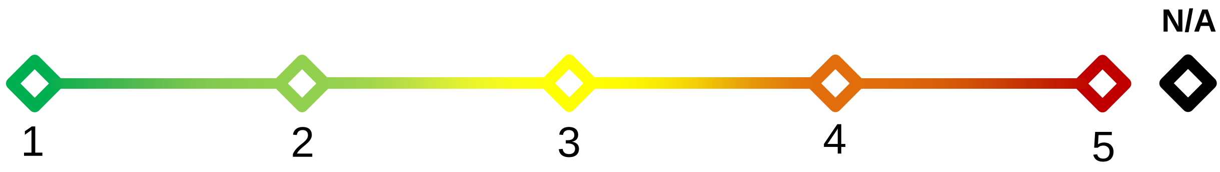
On the self-assessment form:
If any of these characteristics are relevant to your project, please indicate this on the self-assessment form by placing a “1” in the corresponding cell on the form.

If you would like more information about how these weights have been calculated, and how they impact the self-assessment score, then please contact the Data Ethics team, at Data.Ethics@statistics.gov.uk.

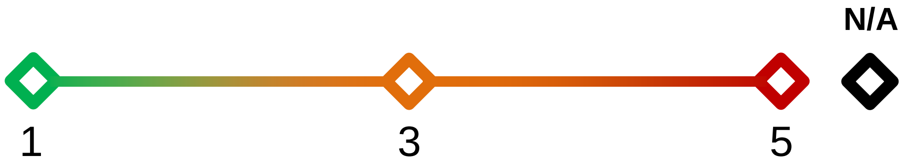
3. Item scoring scales

The scoring scales

In this section, you are asked to assess your project against 22 items grouped against the six NSDEC ethical principles. For the majority of items, we ask you to respond to each of these items using a five-point Likert type scale.



For some items that require a less granular response, we ask you to respond using a three-point scale. For all items, we also ask you to add a short **justification** of your selected rating for each item. Where appropriate and justified, some items can be omitted when completing the self-assessment by selecting N/A, but again a short summary of justification is required as to why a response is not applicable.



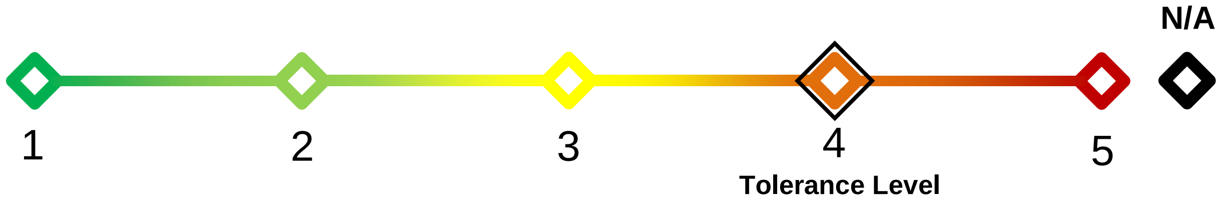
‘The items’ section below provides guidance for how to consider your response to each of the items.

The average of the rating for all items is a good indicator of the overall ethical risk of the project. However, this could lead to a high ethical risk score being averaged out by the results of the rest of the items. To avoid this, we have introduced tolerances.

Tolerances

As mentioned above, to avoid responses that might indicate ethical issues being averaged out of the overall score, we have introduced tolerances against each item that is scored. You will therefore notice that the colour gradients on the self-assessment form differ for each item, with the red indicating the upper tolerance limit. For example, you will notice that the tolerance level for the Public Good item is set low (at 3), as the public good should always be an integral part of the research aims. Without these tolerances, a project could therefore achieve an average score that is “Low Risk”, despite there being no clear public good. When these tolerance limits are reached or exceeded, researchers should consider appropriate actions to mitigate the ethical risk, and if mitigations are not possible we will recommend that the project is sent to NSDEC for a full independent ethical review. Researchers should also set out a justification as to why mitigations are not possible.

In this guidance, tolerance levels against each item are indicated by a diamond shape. In the example below, the tolerance limit is set at 4.



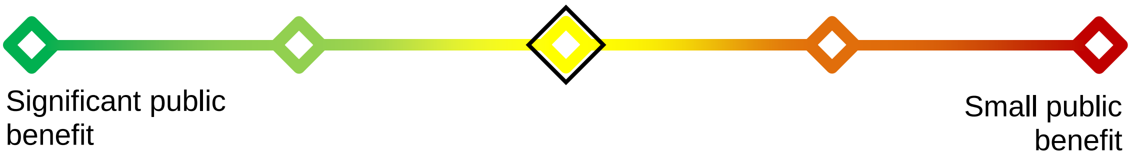
The items

In this section, we provide guidance on how to consider your responses to each of the 22 items grouped against the six NSDEC ethical principles. We also describe which items have the potential to be omitted where such a response can be clearly justified.

Principle 1: The use of data has clear benefits for users and serves the public good

See our guidance on considering and articulating public good in research projects for further information to help you in completing this section.

1. Public good

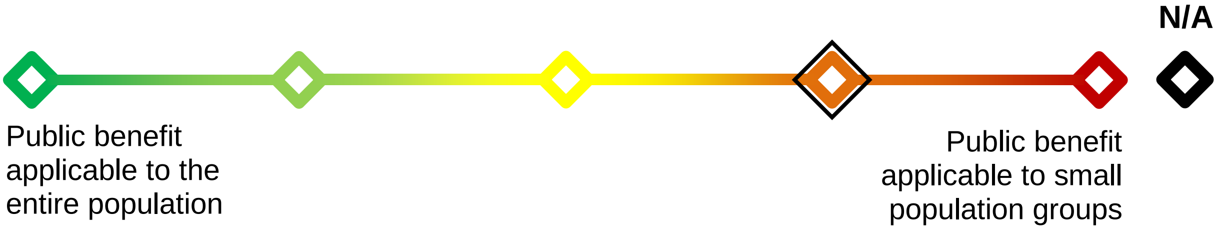


Assessing the public good is by default highly subjective. However, when assessing the public good of your research, you should consider the definitions of public good and public interest set out in the Statistics and Registration Service Act 2007 and the Research Code of Practice and Accreditation Criteria respectively (see the Public Benefits section on Page 4 of this guidance).

It might also help you to consider:

i. how beneficial would your research be to the society as a whole; and
ii. whether it is necessary to conduct this research to realise these benefits.

2. Population coverage



When considering the public benefit of the project you should assess how many people would be affected. If the study is focused a small proportion of the population, or a particular group, then:

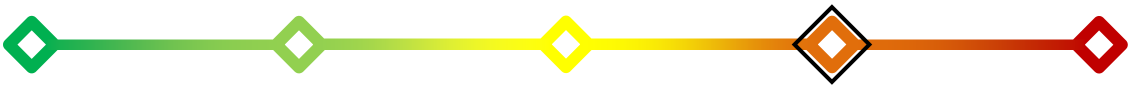
i. the research might disproportionally benefit or disadvantage a group;
ii. the societal impacts of the research might be limited; and
iii. the risk of breaching confidentiality via re-identification increases.

N/A: Omit this item if the scope of the research is specific to a particular group. However, you should justify why the research is focused on that group, and whether this, or other groups, might be adversely affected by this research.

3. Potential harm

You should consider whether the project could cause any potential negative consequences to the public, and whether these are proportionate to the proposed public benefits of the project. Where appropriate, you should also consider whether the activities involved with conducting the research project could cause any potential harm or distress to *any* of the individuals involved, including the research participants, the research team, or the research facilitators.

4. Biases



Identifying and managing bias is essential in research, and to ensure its integrity it is important that you consider:

- the data sources used and most importantly how these are produced;
- the methods and algorithms employed, their assumptions and constraints; and
- the outcomes of your research and how your research is presented.

Principle 2: The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately

5. Direct identification

Direct identification means using the published research outcomes to derive the identity of data subjects without the use of additional data sources. Statistical research may require access to datasets with a higher level of granularity, and to produce and publish statistics researchers might risk breaching the confidentiality of data subjects. You should make sure that adequate statistical disclosure controls are strictly applied to prevent research outcomes being used to directly identify data subjects or attributes identifying population groups.

6. Indirect identification

Data or research outcomes cannot be used indirectly to identify data subjects or specific population groups

Don’t know, or unsure if data or research outcomes could be used to indirectly identify data subjects

Data or research outcomes could indirectly identify data subjects or specific population groups

Indirect identification involves using additional data sources along with research outcomes to derive the identity of data subjects or a set of proxy attributes that can identify individuals or population groups. Although you cannot prepare datasets for every eventuality, you should consider whether the current level of de-identification is proportionate to the datasets being used, and (as much as reasonably possible) if there are any other datasets available which could be used to indirectly identify individuals.

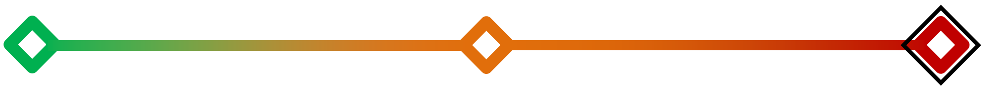
7. Data Security

Strict data security arrangements and confidentiality/ privacy safeguards

Some data security arrangements and confidentiality/ privacy safeguards

Data security is an essential requirement for any research environment. The level of security required should be proportionate to the data collected, used, processed and curated. Depending on the granularity and sensitivity of data, we must ensure that public data is handled in a secure and responsible manner.

8. Consent



Fully informed consent is obtained from all individuals for all stages of the project

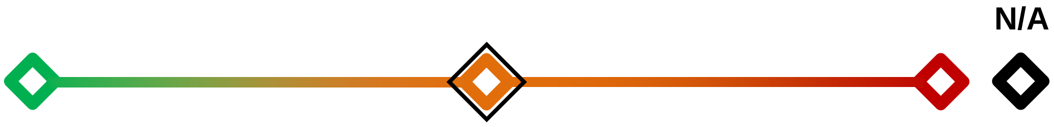
Consent has not been obtained, but is not required as a lawful basis

Fully informed consent has not been obtained, and is required as a lawful basis

In cases where consent is being used as the lawful basis for processing personal data, individuals should be given a say in how their data are used. Unless there are good reasons to the contrary, their consent should be sought in advance of the project taking place. Consent must be well informed and ‘opt-in’ rather than ‘opt-out’.

As indicated by the middle response for this item, there are some instances when consent is not required as the legal basis for processing data (e.g. mandatory surveys, or linkage of UK-wide datasets to produce aggregate UK statistics). However, it is important to note that ethical consent is not necessarily the same as consent required under the Data Protection Legislation. Regardless of whether or not you use consent as a lawful basis for processing personal data, you should still consider whether or not it is ethical for individuals to be given the choice of taking part.

9. Permitted use of data



Permission has been given specifically for this research project, or the proposed use of data is within the same context for which permission was previously given

Don’t know, or unsure if the proposed use of data is beyond the initial context for which permission was originally given

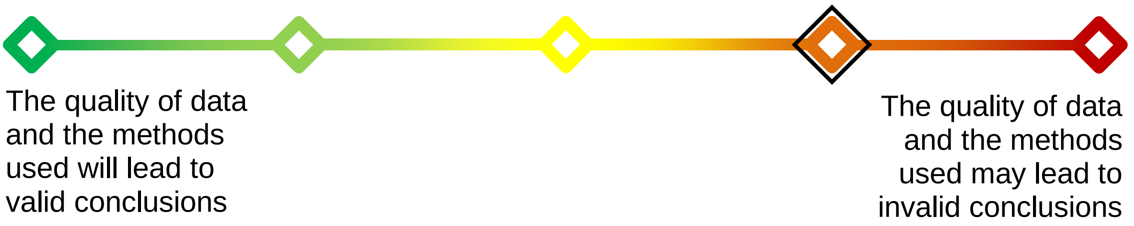
The proposed use of data is beyond the initial context for which permission was originally given

If you have secured approval from a data owner to acquire or use a dataset then you will need to ensure that any further research based on that dataset falls within the context of the original agreement to use this dataset. For further information on ethical considerations when using third-party data, see our high-level ethics checklist focused on this data type

N/A: There may be situations where permission to access certain data is not required. In such instances, you must still provide a solid justification, along with necessary evidence, to explain why permission is not required.

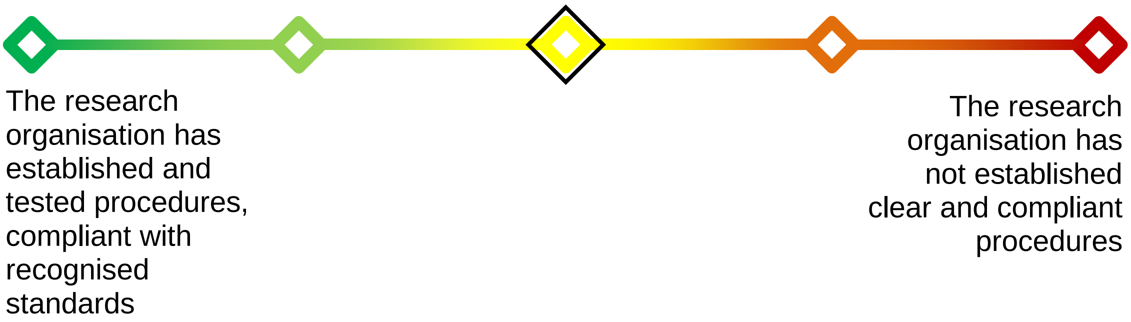
Principle 3: The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality

10. Validity



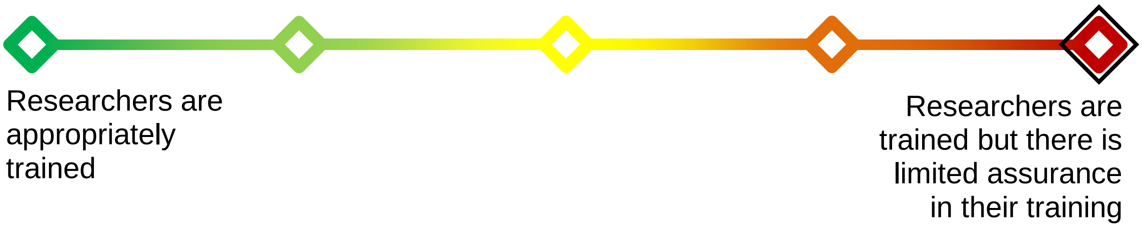
In many cases you might use a dataset without knowing the quality of the data, the methods used to collect, process and visualise the data, and any assumptions made during those processes. All these factors may compromise the validity of the research. You should therefore strive to meet recognised standards of data quality and clearly state any hypotheses and assumptions.

11. Standards



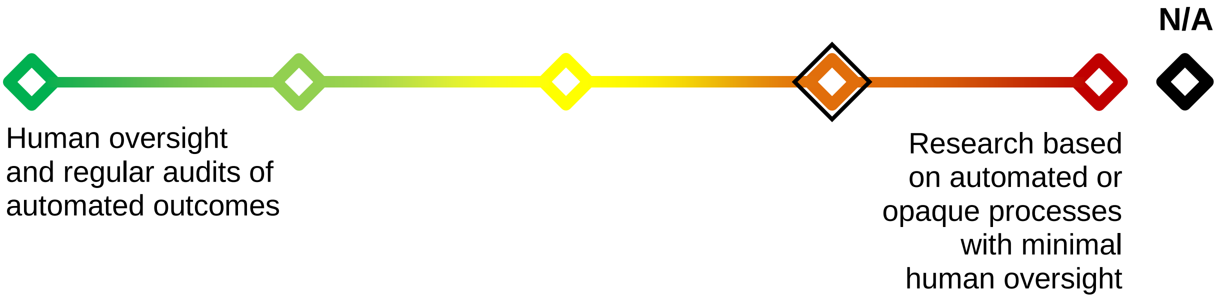
Compliance with recognised standards does not only ensure the validity of the research but also the reproducibility of results. Organisationally it improves the resilience of the organisation to public scrutiny and is a vital part of building public trust and confidence. Apart for auditable research procedures, researchers should have policies in place to assure the security of the research environment, for example to manage data breaches.

12. Training



It is essential that researchers have an updated training portfolio over a broad spectrum of research skills and experience. Documenting these skillsets within the research team enables for more flexible working and ensures continuity and knowledge transfer. Organisationally this provides assurance that, apart from the technical systems, staff have the required expertise to undertake the research specified.

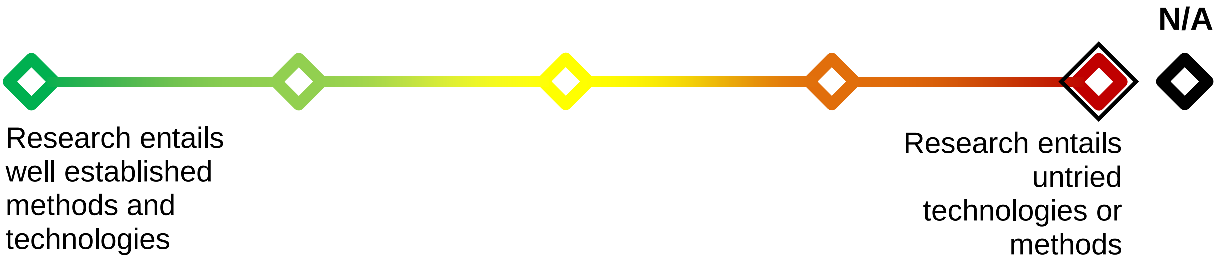
13. Human Oversight



The extended use of ‘off-the-shelf’ software solutions, and the use of code sharing platforms, requires you to be vigilant of assumptions and constraints which may not always be documented. Human oversight is a critical safeguard of any research governance process, requiring an emphasis on the quality of methods used, especially as automated processes become more opaque.

N/A: Omit this item in case of fully **transparent** automated or manual processes with well documented assumptions.

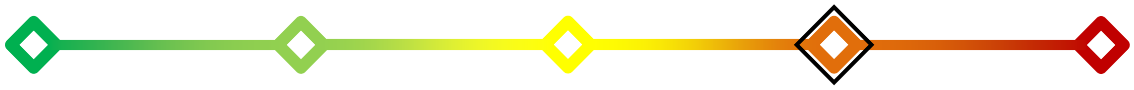
14. New technologies



Established methods and technologies have been tested extensively over long periods, are well documented, and have been subjected to scientific scrutiny. This offers assurance to the public that personal data are handled safely and provides confidence in the quality of research/statistical outputs. New technologies may entail a wider variety of unforeseen risks, from security to methodology, which may not have been discovered. Of course, the research community draws on innovation and should not miss the opportunity to transition to new technologies. Researchers should remain vigilant of the data sources and methods used in their projects and make sure that adequate security arrangements are in place.
If you are using location data or Machine Learning, see our guidance on ethical considerations in the use of geospatial data, and Machine Learning.

N/A: Omit this item for small-scale exploratory projects and feasibility studies which are not used to produce any research/statistical outputs.

15. Potential to realise benefits



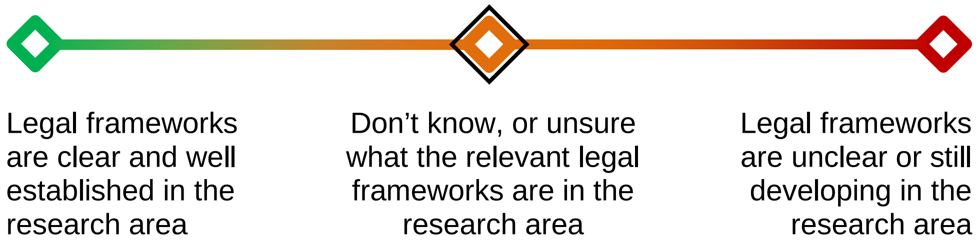
It may not be enough to state the public benefit of your research project; you also need to make sure that the methods used, and the outcomes derived, can be used to realise the public benefit. Complex statistical outputs, increased number of assumptions, or the level of granularity and geography might not properly inform the public or decision-makers.

Principle 4: Data used and methods employed are consistent with legal requirements such as Data Protection Legislation, the Human Rights Act 1998, the Statistics and Registration Service Act 2007 and the common law duty of confidence

16. Established legal gateways and agreements

It is your responsibility to ensure that you acquire data required for your research using appropriate legal gateways and agreements, and for the purpose these gateways and agreements were intended for. For instance, data may be acquired from a government department for the production of a specific statistical output based on a Memorandum of Understanding (MoU) between the two organisations. Using the dataset for any other reason but the production of the specific statistical output would not fall within in the data sharing agreement.
For further information on ethical considerations when using third-party data, see our high-level ethics checklist focused on this data type.

17. Established legal frameworks

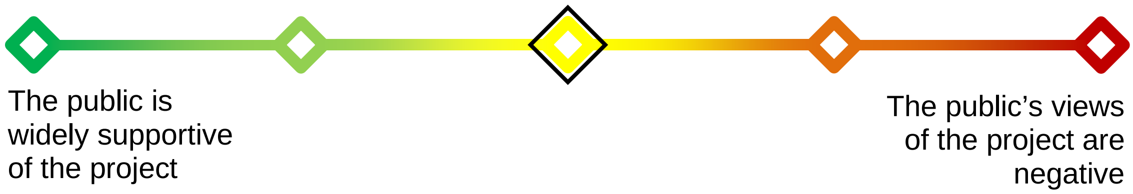


Working within a developing legal framework entails risks, especially for long-term projects, and as such requires researchers to remain alert to assess the impact of new laws relevant to their project. Due to the trans-national nature of some data, researchers should also consider the need to comply with international legislation when appropriate. Researchers should consult with legal professionals to ensure the legal compliance of their approach.

Principle 5: The views of the public are considered in light of the data used and the perceived benefits of the research

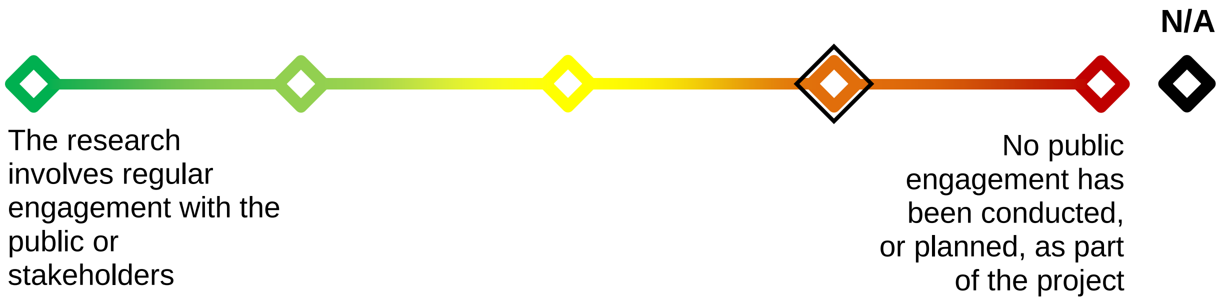
See our guidance on considering public views and engagement for research and statistics projects when completing this section

18. Public views



Research does not happen in isolation, so the wider environment in which researchers operate should always be taken into account. This does not mean that the public's views must be sought for every project, as this would be disproportionately time and resource consuming, but an overall awareness of public acceptability must be considered. Information from engagement events for similar projects, government initiatives, public polls and literature reviews are reasonable alternatives to large public consultations, focus and expert groups.

19. Public engagement



Securing public engagement with research projects is one strategy for facilitating research projects. This could include engagement with the public or with specialist/ interest groups. Although there are several approaches to public engagement, it is most effective to maintain regular engagement throughout the life cycle of the project. A project might be acceptable at the design phase but may warrant further engagement at a later stage e.g. when producing outputs.

N/A: Omit this item when no public engagement is required and can be clearly justified (e.g. for the production of statistics as part of statutory responsibilities; or the same, or a very similar, research project has already completed public consultation or public acceptability testing)

Principle 6: The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public

20. Public access to outcomes

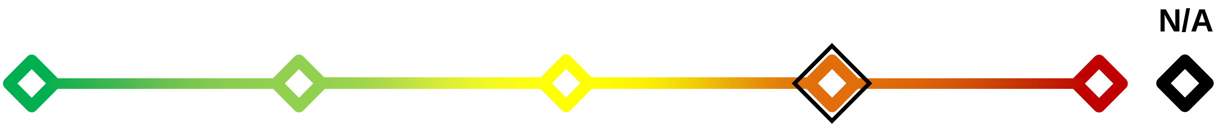
Research outcomes are (or will be) openly available to the public

Don't know, or unsure if research outcomes will be openly available to the public

Research outcomes are not (and/or will not be) openly available to the public

The use of data produced by the public offers an exciting opportunity to the statistical community but comes with a responsibility to be transparent to the public in the way we use their data. It is imperative that we share the research outcomes with the public and ensure that they remain openly accessible. This transparency principle is enshrined in the Codes of Practice for Statistics and Research under the Digital Economy Act 2017, and is also set out in the UK Research and Innovation's Open Access Policy.

21. Sharing of methods and tools



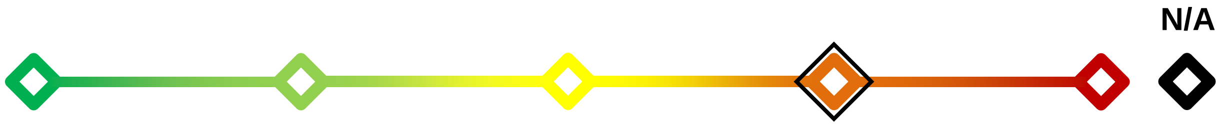
Both methods and tools are widely available to the public

Methods or tools are not widely available to the public, or will only be shared internally

In parallel with research outcomes, researchers often develop new methods and tools to enable future research to be more effective. Where appropriate, it is good practice for researchers to make these new methods and tools available for others to use, as this enables wider research impact and innovation throughout the research community.

N/A: There are some cases where researchers may not be able to share these tools and methods, and in those instances, this item can be omitted. Firstly, when reverse engineering the tools or method could compromise the confidentiality of the statistical outputs produced. Secondly, when there is a legal agreement in place that prevents us from doing so, for example tools and methods are produced in partnership with a third party which retains intellectual property rights.

22. Data curation and re-use



Data will be curated based on data retention policies and will be available for re-use by the public

Data will not be available for re-use, or data retention policies are not in place

You should select an appropriate retention period for the data to ensure that your research can be reproduced and validated. Due to the significant costs in re-acquiring and preparing data we encourage you to re-use raw and linked datasets when possible. You should remain vigilant of the sensitivity of identifiable datasets to be retained when selecting retention periods and data re-use.

N/A: Omit this item when data sharing agreements or original consent does not allow re-use of the dataset.

Training and Support

The Data Ethics team are your main point of contact for ethical queries, and can also help to facilitate further discussions and offer general assistance to researchers by providing advice on the self-assessments.

To support colleagues across the Government Statistical Service (GSS) and wider research community, we can also provide ethics training on request.

For more information, please contact us at Data.Ethics@statistics.gov.uk.

User Checklist

This checklist is to help highlight the key steps you should take when using the self-assessment form to evaluate the ethics of your project. It also highlights some of the key people or business areas that you may wish to discuss your self-assessment/ project with in order to ensure the legality of your project.

☐ Made the appropriate checks that the use of data is legal
☐ Completed the Basic Information section

☐ Completed the Weightings and Sensitive Research Areas section
☐ Completed the Item Scoring Scales section
☐ Shared self-assessment with appropriate colleagues to receive feedback
☐ **Shared self-assessment with the Data Ethics team for feedback** (this is an essential step, regardless of your self-assessment score).

Please send completed self-assessment forms to Data.Ethics@statistics.gov.uk for feedback before commencing your project.