UK Statistics Authority Centre for Applied Data Ethics

Ethics checklist for use of geospatial data for analysis and statistics

When designing a statistical or geospatial analysis project, consider the steps in the checklist below. As with all ethical considerations, context and the bigger picture is everything – the first point is the most important.

☐ 1. Think

Step back from the project and think hard about the work you are about to undertake and where ethical issues might arise.

In particular, what might you do that:

- breaks confidentiality;
- discriminates or disadvantages;
- is biased or unfair or excludes groups or individuals;
- misrepresents the truth.

Try and think about those things that might go wrong and take steps to avoid them – using the advice provided in our guidance and by seeking further support if necessary.

☐ 2. Do no harm

Think about whether you should undertake this piece of work at all. Just because you can – does not mean that you should. Just because the data fits together, does not mean that it tells a truth. Will this analysis do more good than harm? Will it be a positive influence in the world?

Consider in particular the UK Statistics Authority's general ethical principles below:

☐ 3. Public Good

Have the benefits of using geospatial data been thought through and clearly documented?

☐ 4. Methods and Quality

Have the limitations of the data, methods and technologies been considered? How will you document these?

☐ 5. Transparency

Has transparency in the collection, use, retention and sharing of the data been considered? How will you publish your results and methods?

☐ 6. Legal Compliance

Has relevant regulation been considered in relation to the dataset used, both in the UK and if necessary, internationally?

☐ 7. Public Views and Engagement

Have potential public views regarding particular uses of geospatial data across different contexts been considered? How will you engage with any communities impacted? Staying up to date with current research and initiatives on public views regarding geospatial data use may be helpful.

☐ 8. Confidentiality and Data Security

Have appropriate mechanisms to maintain confidentiality of datasets been considered? How will the security of the data be maintained?

Consider these points more specific to geospatial data in a statistical context:

☐ 9. The choice of geography:

Is the geography that you have chosen for your analysis (and particularly for the release of any results) the right one to avoid disclosure of confidential information? Have you considered and mitigated the risk of disclosure by differencing?

☐ 10. Disclosure by location:

Might any of the data that you will release provide specific information about individuals or small groups (say the location, route or habits of individuals or groups that might be put at risk)? Has the potential for the disclosure of locations to result in a negative impact on groups or individuals been fully considered?

☐ 11. Ensuring inclusion:

Has the potential for individuals or groups to be excluded from datasets due to reduced engagement with digital technologies, services or infrastructure been considered? Have you considered using additional data sources to improve inclusion? If the data cannot be complete or representative, you must take account of this in your analysis and document it clearly when reporting the results.

☐ 12. Avoiding bias:

Have you considered the potential for bias in your data or even in the choice of study? What are your assumptions about this area or topic? Are you sure that you are not mirroring or reinforcing an unfair bias?

☐ 13. Unintended consequences:

Might the data or your results be re-used outside of their original context and purpose in the future to the disadvantage of individuals or communities? What can you do to try and protect against this possibility?

☐ 14. Double-check any strange results:

Unexpected geographical patterns in study results should always be challenged before publication to ensure that there are not any overlooked outliers or missing themes. Unexpected results can often be a clue to an unidentified or misunderstood relationship – and the more influential the study the greater the potential for harm (or good) as a result! The Code of Practice for Statistics and HM Treasury Aqua Book provide further information on verification and validation of analysis.

☐ 15. Mapping and geovisualisation:

Do the choice of ranges, colours and symbolisation you have chosen reflect the real relationship rather than a particular story you have chosen to tell? Is the choice of map type appropriate to the topic? The data, not the medium or desired policy should define the story.

☐ 16. What will the user think?:

It is a good test to look at your final results or mapping and consider: What would a reader coming to this completely fresh read from these results? How would they read this map? Is that reading a fair reflection of the data? If not, you should think again. A second opinion may be useful here.