

# #7 Analysing Secondary Data

## Practising ethics guides to built environment research

Tania Guerrero Rios & Jens Kandt

### When planning

1. What data would be ideal to answer my research question & do those data already exist?
2. Do I fully understand the purpose, methods, and context within which the candidate data sets have been collected?
3. Are the data in the public domain,
4. Would data subjects expect the type of analysis I am planning?

### When conducting

5. Have I selected the appropriate methods to answer my research questions with these data?
6. Have I documented all the steps of my analysis, including any data cleaning, filtering, and variable selection?
7. Is it possible to identify individuals?
8. How can I best safeguard the confidentiality and anonymity of data subjects?

### When producing & communicating

9. Are my illustrating outputs clearly presented and easy to interpret?
10. Is it possible to identify individual persons, households or organisations from my outputs?
11. Am I allowed to share data outputs for further use in a public repository?
12. Do I know for how long I may retain the data and how to delete them securely when appropriate?

## Practising Ethics: Guides

These guides, curated by the [Bartlett's Ethics Commission](#) in collaboration with [KNOW \(Knowledge in Action for Urban Equality\)](#), and edited by Jane Rendell, (Director of the Bartlett Ethics Commission 2015-20), offer insights by experienced researchers into how to negotiate the ethical dilemmas that can arise during a research project. The aim is to help you practise built environment research ethically. David Roberts (Bartlett Ethics Fellow 2015-20) devised the format and structure of these guides to follow the ethical issues that arise during the development of a research process – from planning, to conducting, to communicating and producing outcomes – and Ariana Markowitz wrote some of the introductory text that runs across all guides. The guides focus on the different kinds of ethical issues you might encounter as a result of using specific processes or methods, and pay attention to the particular contexts and ways in which these methods are practised. Because when practising research, methods and context inform one another, we consider this series of guides as embedded in a mode of applied ethics called **situated** or **relational ethics**. Where you see words that are highlighted, they refer back to our definitions of key **ethical** principles and to terms contained in institutional protocols as found on [Practising Ethics](#).

1. *Making Images* (David Roberts)
2. *Asking Questions* (Yael Padan)
3. *Co-producing Knowledge* (Yael Padan)
4. *Staging Research* (David Roberts)
5. *Researching, Risk, and Wellbeing* (Ariana Markowitz)
6. *Researching Internationally* (Emmanuel Osuteye)
7. ***Analysing Secondary Data* (Tania Guerrero Rios and Jens Kandt)**

### Acknowledgements

The Bartlett Ethics Commission worked alongside the Bartlett Ethics Working Group from 2015-8, and so benefitted greatly from the input of Martin Austwick, Claire Colomb, Helen Dougal, Helen Fisher, Efrosyni Konstaninou, Rowena Lamb, James O'Leary, Niamh Murtagh, Hedley Smyth, and Steve Ridge, but in particular Jens Kandt, Mike Raco, Michelle Shipworth, and Michael Walls. The project was supported educationally and economically by the Bartlett Faculty, and specifically its former Dean, Alan Penn.

## Guide #7 Analysing Secondary Data by Tania Guerrero Rios & Jens Kandt

Researchers need to consider the ethical implications of their work with respect to privacy harms, notice and consent, and the uses to which their research is being deployed. Beyond complying with relevant laws and institutional research board requirements, analysts have a duty of care to their fellow citizens not to expose them to harm through their analysis.

Rob Kitchin, "The ethics of smart cities and urban science," *Philosophical Transactions of the Royal Society A*, 374 (2016), 12.

### About this guide: why and how built environment researchers work internationally

Built environment research is as much about people as it is about places: the people who use and inhabit the places you are researching, the people who engage with those places emotionally or spiritually even if they are not physically present, the people who build them, and the people who own or manage them. In addition, you the researcher are necessarily a key actor: you devise the research approach, become a participant in the place where you gather data, and you determine how to interpret that data and what to do with it. Because people are unpredictable, research can also be unpredictable, and as a researcher you are likely to encounter unexpected **situations** that require you to think on your feet whilst navigating high expectations with limited time. Even the best-laid plans often go awry when they come into contact with reality and real people and you will need systems in place to support you throughout that process, minimising **harm** to those you are researching and participating with, as well as yourself. **Ethics** is about what kind of lives we should lead, what actions are right and wrong, what qualities of character we should develop and what **responsibilities** we have for each other and our ecosystem. To conduct research **ethically** it is important to consider the **benefits, risks** and **harms** to all connected with and affected by it.

## The ethics of analysing secondary data

Secondary data can be defined as pre-collected data that is shared with and used by a third person, such as a researcher.<sup>1</sup> By contrast, primary data are directly collected by the researchers for their own specific purpose. Secondary data analysis therefore implies a need for researchers to carefully check that the data are suitable for their intended research.

Secondary data can come in a structured format, such as quantitative records, or unstructured formats, such as qualitative findings or text. The contemporary proliferation of digital technologies has made new data sources available, such as administrative databases or high-frequency transactional data, for example from government departments, social media, mobile phones or travel smart cards (often referred to as “Big Data” due to their large volume). Because such data can be accessed at a low cost, they can provide appealing and easy-to-use empirical material for research on the built environment.

Challenges of using secondary data include uncertainty as to the suitability of using pre-collected data created for a different purpose, robustness of data collection procedures and bias.<sup>2</sup> Although researchers can often access secondary data on a bona fide basis, the data may only be processed under certain terms and conditions of use. For example, data created as part of commercial operations are often released under certain data sharing restrictions. Since the researcher has not taken part in the data collection, the researcher must ascertain whether the data have been collected legally and ethically. Even when personal data have been anonymised, re-identification may be possible when the data are combined with other sources.<sup>3</sup> If you have access to identifiable personal data, you must also ensure you report results in such a way that individuals cannot be identified.

## How to use this guide

These *Practising Ethics Guides* outline appropriate ways to engage **ethically** in research. *Analysing Secondary Data* is designed to be a point of reference at any stage of your research – from planning your project, to conducting activities in the field, to communicating what you have learned through the production of particular research outputs.

*Analysing Secondary Data* contains principles, questions, guidelines and resources. The principles in the next section inform best practice. These are not just regulatory hurdles for you to jump through at the beginning stages of your research but concepts that ground **ethical** inquiry throughout. They help you develop and refine an approach that it is **sensitive** to the physical and emotional challenges that may arise in the research process, enabling you to be a more effective researcher. The series of guiding questions act as prompts for you to **reflect** on the potential **ethical** considerations which emerge throughout a project, before, during, and after you conduct your research. The guidelines expand on the questions, illuminate the different **ethical** concerns they raise, and recommend actions which embody these principles. The resources section provides additional information.

Through the digitalisation of our society, new types of data are constantly being produced, shared, and processed. Simple daily activities, such as navigating through cities or checking into a venue, have made us both producers and consumers of data.<sup>4</sup> While the use and processing of these data sources may provide collective benefits, they are also associated with **ethical** risks in relation to **privacy** and **confidentiality**. If a particular use of data is legal, this does not by default imply that it is ethical as well. Although there are established data protection techniques to safeguard **privacy** and **confidentiality**, it is important to keep in mind key principles of research ethics when using and processing data that relate to human subjects. We discuss the most important principles in this guide.

# Principles

The people, places and research methods you use and the contexts in which they are practised will each raise their own **ethical** considerations related to a common set of principles that encourage **ethical** conduct and promote interaction based on good faith and mutual **respect**.

**Benefit not harm:** Your research should have a **benefit** to society and any **risks** involved to participants must be minimized, balanced against the potential **benefit** to the overall community, and clearly explained to participants before they give their **consent**.

**Informed consent:** You need to inform your participants about the study and what is being asked of them, including any potential **risks** or **benefits**, in order for them to make an informed and voluntary decision about whether or not to participate in the research.

**Confidentiality:** You need to inform participants of the extent to which **confidentiality** can be assured and **respect** their right to remain **anonymous** in dissemination and display.

Additionally *Analysing Secondary Data* highlights the following principles:

**Purpose limitation:** You must have a legally permitted purpose to use personal data, and you must process the data for this purpose only.<sup>5</sup>

**Transparency:** You should communicate the purpose of your research in a way that is accessible and clear, and you should document how you will process the data and for how long.<sup>6</sup>

**Data minimisation:** You should make sure the data you collect are adequate for answering your research questions, and that you store and process only those data items that you need for your research.<sup>7</sup>

**Acceptability:** The way you use the data should be expected by and acceptable to data subjects. For example, you should not try to identify individuals even if you have legal access to individual records.

**Safe storage:** All data processing present security risks and unauthorised access must be prevented as best as possible. Data security measures should be proportionate to the granularity of your data and the potential harm to data subjects.

**Safe reporting and sharing:** You should carefully select illustrations and report outputs at a safe level of aggregation to prevent unintended identification of individuals.<sup>8</sup>

**Reproducibility:** You should make sure your research can be reproduced by other researchers.<sup>9</sup> To do this, you should provide clear and transparent documentation of your databases and analytical methods.

# Questions

## ***When planning research: Assess suitability and data ownership***

1. What data would be ideal to answer my research question and do those data already exist?
2. Do I fully understand the purpose, methods, and context within which the candidate data sets have been collected?
3. Are the data in the public domain, i.e. openly accessible? If not, do I have permission to use them?
4. Would data subjects expect the type of analysis I am planning? Do I need to inform data subjects about my research?

## ***When conducting research: Minimise data to reduce risk***

5. Have I selected the appropriate methods to answer my research questions with these data?
6. Have I documented all the steps of my analysis, including any data cleaning, filtering, and variable selection?
7. Is it possible to identify individuals (directly or indirectly) from these data?
8. How can I best safeguard the confidentiality and anonymity of data subjects?

## ***When producing and communicating research outputs: Share safely and ensure reproducibility***

9. Are my illustrating outputs (charts/tables/maps) clearly presented and easy to interpret?
10. Is it possible to identify individual persons, households or organisations from my outputs?
11. Am I allowed to share data outputs for further use in a public repository?
12. Do I know for how long I may retain the data and how to delete them securely when appropriate?

## Guideline 1 When planning research: *Assess suitability and data ownership*

There are many types of secondary data sources; they can be qualitative, such as reports or interview transcripts, or quantitative, such as social surveys, administrative data sets or commercial data. Relevant data providers can be international organisations (e.g. World Bank, United Nations or OECD), government organisations (e.g. UK Office for National Statistics, Ordnance Survey, NHS or transport providers) or private organisations (e.g. social media companies, high street retailers or telecommunications companies). In recent years, an expanding range of secondary data have become available, and choosing data for your research can be an overwhelming task. You can start by thinking what the best data would be to answer your specific research questions. Look at the literature in your field and examine how similar questions have been answered by others and what methods and data have been used. In our digital age, however, suitable data may also come from unexpected sources and you should bear in mind that the data you wish to use may have been created for completely different purposes.

Many secondary data sources are public and can be accessed freely through the Internet, while some data may only be available via data sharing agreements with the data owner. In all cases, we must consider ethical issues. Consider the following contrasting examples: census statistics and social media data. Census statistics about, say, the age structure of UK population broken down by gender, are highly aggregated, i.e. they present summary data (counts of people per age group and per gender) for a large territory (the UK). Nobody would be surprised by the decision to use such data for research and its use would not plausibly pose a risk of harm to anybody. And yet, we should understand the legal context in which the data have been collected, consider the adequacy of the data for our research goal, and ensure transparency and reproducibility.

Now consider social media profiles, which may include the age, sex, and other interests and preferences of identifiable individuals. These too can be accessed freely, and one might argue that social media users have consciously placed their data in the public domain. But did users share these to provide research data to the public? And would they expect that their personal data

be used to conduct research about them? Since the answer to this last question is usually 'no', an ethical issue arises in that personal data gathered from data subjects may be used for a purpose that these subjects have not imagined or agreed to.

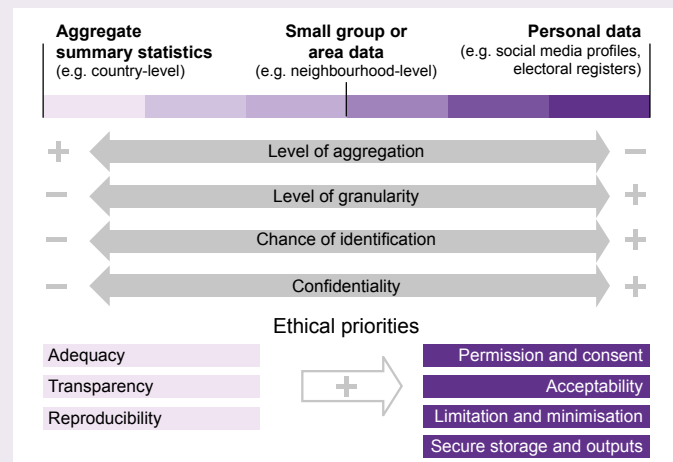


Figure 1: “The Secondary Data Spectrum: the higher the level of granularity, the more ethical concerns need to be addressed.” Source: Authors.

The two examples described occupy opposite ends of the spectrum of quantitative secondary data: at one end, highly aggregate summary statistics, and at the other end, personal data pertaining to identifiable individuals (See Figure 1). But there may also be data sources that lie in between the ends of the spectrum. For example, anonymised mobile phone records may be used to locate the activities of individuals in space. Although the records are anonymised, it may still be possible to identify individuals when linked to contextual information or accidental knowledge about a known individual's whereabouts. Higher levels of data granularity thus present a greater risk of identification, even when the data are anonymous.

The use of identifiable personal data presents an intrusion, which must be justified and balanced against the benefit your research. Ask yourself, are there less intrusive and more aggregate alternatives suitable for your analysis? If not, do you need to inform data subjects about your research? Can you obtain consent or permission from data subjects or data owners to access the data?

At this stage, it is also important to think about how you will process the data. A data management plan can help you think through how you access, process, store, share, re-use and safely dispose of the data after you complete your research.<sup>10</sup> Will you be able to minimise the data you use? How will you securely store the data? The approach you take to accessing and processing the data should be proportionate to the ethical risk it poses, i.e. the level of granularity of your data. Personal data need a higher level of management and protection than aggregate data.

In assessing the suitability of your data, make sure you understand the following.

1. The context in which the data have been collected. Have they been collected and prepared by the government, for example as part of an Open Data initiative? Are they collected as part of operational practices by a commercial organisation?
2. The purposes for which the data have been collected.<sup>11</sup> It is very likely that you will use the data for a different purpose, so you need to be mindful of its suitability to answer your own research question as well as the expectations of the data subjects about the use of these data.
3. The legal and ethical basis upon which the data have been collected. Even when participants have agreed to take part in research involving their personal data, you need to consider whether or not these participants can, and need to be, further informed about your research, and whether their consent should be sought if their data will be used for a purpose different from that originally agreed.
4. The extent to which you can access, process and share the data. Are the data publicly available? Are they only available on request under certain conditions? Do you need to enter data sharing agreements to use them? Do you need to involve your institution in this process?

If you are satisfied that the data are adequate for your research, that their use is acceptable, that you can process them safely without causing harm, then get started with your analysis.

## **Guideline 2 When conducting research: *Minimise data to reduce risk***

A scientifically sound and rigorous research design that is tailored to the data is extremely important.<sup>12</sup> Ask yourself: What do the data represent? What or who is included in the data? What or who is excluded? Could my analysis be biased, e.g. when certain types of individuals are over- or underrepresented? If these questions are not considered, inaccurate conclusions may be drawn from your analysis. This is not only a technical research problem; in the context of social science, medical or built environment research, unacknowledged bias and inappropriate methods can contribute to misguided decision-making. So, irrespective of where your data sits on the spectrum, make sure that you:

- Reflect on your conscious and unconscious assumptions about your data.<sup>13</sup>
- Ascertain what your data include or exclude.
- Document and justify all steps of your analysis, including data cleaning, filtering, variable selection, data storage arrangements and the analytical (e.g. statistical) steps you undertake.

When you use more than one database, you can keep track of them by creating a variable information log where you can include data source, version controls and licensing information.<sup>14</sup> It is also helpful to include a metadata file describing contents and variables of each data set.

Depending on where your data is located on the secondary data spectrum (See Figure 1), decide where you need to focus your ethical reflection. Aggregate data are very secure in most cases, i.e. the risk of identifying individuals and causing harm is extremely low. While using aggregate data, you should thoroughly and transparently document your analysis.<sup>15</sup> If you are dealing with personal data, you will additionally need to focus on purpose limitation, data minimisation and data security. Any data set located in the middle range of the spectrum will need similar attention, and any measures you take should be proportionate to the level of ethical risk.

If you have access to more granular data, there is significant room for discretion, and reflection on ethical practice and integrity is crucial throughout your analysis.

Good practice in data analysis includes:

- Only use data for your research purpose, no other purpose. This principle is known as “purpose limitation.”
- Minimise or filter the data set to the number of records and variables you actually need for your research. This principle is known as “data minimisation.” For example, do you need the full set of records or can you address your research question using a sample?
- Do not actively try to identify individuals; do not de-anonymise your data or perform any other actions that data subjects would not deem acceptable.

Methods to safeguard anonymity and confidentiality include removing or pseudonymising direct identifiers (e.g. name, postcode, etc); aggregating potential identifiers (e.g. workplace, salary, age into categorical bands); and anonymising or aggregating georeferenced data.<sup>16</sup> While performing your analysis you can also ensure anonymity by re-coding low-frequency occurrences or/ and excluding variables that could potentially breach anonymity.<sup>17</sup>

You should make sure that you store data securely throughout and beyond your project’s life cycle. When using personal or sensitive data, remove identifiers and store them separately from the rest of your data. Avoid storing personal data in portable devices (laptops, flash drives) and use a secured computer network through your institution instead. When you must transfer data, use secure and encrypted file transfer. Your institution may provide secure ways of sharing data outside your institution (e.g. institutional Dropbox or OneDrive). And of course, only share data with individuals who are permitted to access them. To summarise, make sure you:

- Store data safely.
- Store identifiers separately from substantive data and keep those processes that require identifiable data separate from processes that do not.
- Encrypt identifiable data.
- Understand and adhere to physical and informational security arrangements to minimise risk of unauthorised access to the data.
- Destroy research data securely after the completion of your research and any data retention period.
- Seek advice from your IT department where needed.

You should also be mindful of relevant legislation governing secure data processing. For example, the 2018 EU General Data Protection Regulation (GDPR) – implemented through the UK Data Protection Act 2018 – defines different roles, principles and rights regarding the processing of personal data.<sup>18</sup> Please also note that special security requirements apply to the processing of sensitive data (called ‘Special Categories’ data in GDPR), which include racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health, sexual life and orientation and genetic and biometric data. If in doubt, seek advice from your IT department or Data Protection officer.

Finally, remember that researchers can be held accountable for handling personal data according to data protection principles.<sup>19</sup> Therefore, if you are processing personal data, make sure that, in addition to your ethical reflections, you understand any legal requirements in relation to personal data processing. Useful guidance can be found in the “Guide to the UK GDPR.”<sup>20</sup>



## Guideline 3 When producing and communicating research outputs: *Share safely and ensure reproducibility*

When producing and communicating research outputs, remember that you are not only a data processor, you are also a data producer. Your research outputs can become data for other researchers or further research. This can take the form of qualitative findings or even data outputs for re-use, for example, when you use census data to develop an index of deprivation at the level of neighbourhoods. Therefore, the reproducibility, transparency, and communication of uncertainty are critical principles guiding this stage of your research. These principles apply to all kinds of secondary data, irrespective of where they are located on the spectrum (recall Figure 1).

When you are dealing with data on the granular end of the spectrum, additional principles apply. You must ensure that outputs are safe, specifically that no individual, household or organisation can be identified. The action required to ensure this is known as “disclosure control.”<sup>21</sup> Here is a list of actions you can take to ensure outputs are safe to be disseminated:

- Aggregate results, e.g. by geographical area, by region instead of neighbourhood or by other categories, e.g. wider income and age bands. As rule of thumb, the counts for each category should not be lower than five cases. Check your data sharing agreement if special conditions apply.
- Remove or suppress outliers, e.g. by restricting upper and lower ranges of results.
- Use safe summary statistics, e.g. the mean, instead of measures that reflect actual, raw values, e.g. the median.

If you are producing data for re-use, you are faced with a balancing act. On the one hand, you want your outputs to be sufficiently granular to be informative and valuable. On the other hand, you must minimise the risk of identification. This is challenging; think carefully about your research objectives and what you are able to share safely in the context of your particular field of research. You should always remove all personal information before storing or sharing the data for future use.<sup>22</sup> More guidance about disclosure control can be found at the Office for National Statistics.<sup>23</sup>

Reproducibility is an important research and ethical principle. This involves documenting all steps of data modification and analysis during your research and may even extend to sharing a safe analytical data set or the code you developed.<sup>24</sup> This information can be shared via public repositories on the web, such as Github, or institutional repositories or data archives such as the UK Data Service.<sup>25</sup>

But again, you must balance the level of sharing against the risk of identification. If your research involves highly granular data, sharing everything may lead to breaches of confidentiality. You can use a data management plan (as introduced above) to define what you share and at what level of aggregation, to decide where you share your outputs and whether you will share your outputs openly or subject to certain access restrictions. You can find useful guidance about sharing at the UK Data Service.<sup>26</sup>

As in any research, you should always aim for unbiased reporting of your findings, by producing clear and comprehensible outputs. Any visualisations like charts or maps should include all the elements needed to make them self-explanatory. You should cite data sources accurately and include details about the original purpose for which the data was collected.

Finally, do not forget to think about safe destruction of your raw data when no longer needed. Check your institution’s data retention policy, which assigns a maximum storage life to data, according to its type. For example, research data may have a data retention period of 10 years. Once the retention period has passed, they should be securely destroyed.<sup>27</sup> If in doubt, seek advice from your IT department.

## Resources

- Aguilera, U., O. Peña, O. Belmonte and D. López-de-Ipiña. "Citizen-centric data services for smarter cities," *Future Generation Computer Systems* 75 (November 2017): 234–47.
- British Educational Research Association (BERA). *Ethical Guidelines for Educational Research*. London: British Educational Research Association, 2018.
- Cairo, A. *The Truthful Act. Data, Charts, and Maps for Communication*. Indianapolis: New Riders, 2016.
- European Commission's Information Society Technologies (IST). *RESPECT Code of Practice for Socio-Economic Research*. Brighton: Institute for Employment Studies, 2004.
- Hasan, A. "Ethical considerations in the use of secondary data for built environment research." In *Secondary Research Methods in the Built Environment*, edited by E. Manu and J. Akotia, 26–39. Oxfordshire: Routledge, 2021.
- Kandt, J. and Batty, M. "Smart cities, big data and urban policy: Towards urban analytics for the long run." *Cities* 109 (2021) e102992.
- Kirk, A. *Data Visualisation: A Handbook for Data Driven Design*. London: Sage, 2016.
- Kitchin, R. "Big data and human geography: Opportunities, challenges and risks," *Dialogues in Human Geography* 3, no. 3 (2013): 262–267.
- Kitchin, R. "The ethics of smart cities and urban science," *Philosophical Transactions of the Royal Society A* 374 (2016): 1–15.
- Salkind, N. J. "Secondary data analysis." In *The Encyclopedia of Measurement and Statistics*, edited by Neil J. Salkind, 870–876. Thousand Oaks: Sage, 2011.
- Information Commissioner's Office (ICO), "Guide to Data Protection," accessed 24 June 2021, <https://ico.org.uk/for-organisations/guide-to-data-protection/>
- The Research Ethics Guidebook, "Secondary analysis," accessed 24 June 2021, <https://web.archive.org/web/20190208100635/http://www.ethicsguidebook.ac.uk/Secondary-analysis-106>
- UK Data Service, "Research data management," accessed 31 August 2021, <https://ukdataservice.ac.uk/learning-hub/research-data-management/>
- University College London, "UCL Data Protection Policy," (2017), accessed 24 June 2021, <https://www.ucl.ac.uk/library/sites/library/files/data-protection-policy.pdf>

## Bios

Tania Guerrero is a PhD candidate at the Department of Geography, University College London (UCL). Since 2017, she has been working as a Senior Postgraduate Teaching Assistant at the Bartlett Centre for Advanced Spatial Analysis (CASA), where she has coordinated the development of robust review procedures for research ethics (email: [tania.rios.16@ucl.ac.uk](mailto:tania.rios.16@ucl.ac.uk)).

Jens Kandt is Associate Professor in Urban Geography and Data Science at the Bartlett Centre for Advanced Spatial Analysis (CASA), University College London (UCL), where he also acts as Director of Research and Departmental Ethics Lead. His research on transport, mobility and health frequently involves the use of secondary data (email: [j.kandt@ucl.ac.uk](mailto:j.kandt@ucl.ac.uk)).

## Online references

- European Environment Agency (EEA), "Dos and don'ts of data visualization," accessed 31 August 2021, <https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts>
- GDPR, 'Complete Guide to GDPR compliance', accessed 24 June 2021, <https://gdpr.eu/>

## Endnotes

- 1 Neil Salkind, "Secondary data analysis," in *The Encyclopedia of Measurement and Statistics*, ed. Neil J. Salkind, (Thousand Oaks, California: Sage, 2011), 870–76.
- 2 Salkind, "Secondary data analysis," 870–76.
- 3 Rob Kitchin, "The ethics of smart cities and urban science," *Philosophical Transactions of the Royal Society A*, 374 (2016), 1–15.
- 4 Unai Aguilera, Oscar Peña, Oscar Belmonte and Diego López-de-Ipiña, "Citizen-centric data services for smarter cities," *Future Generation Computer Systems* 75 (November 2017): 234–47.
- 5 "Principle (b): Purpose limitation," Information Commissioner's Office (ICO), accessed 24 June 2021, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/>
- 6 "Principle (a): Lawfulness, fairness and transparency," Information Commissioner's Office (ICO), accessed 24 June 2021, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/#transparency>
- 7 "Principle (c): Data minimisation," Information Commissioner's Office (ICO), accessed 24 June 2021, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/>
- 8 "Disclosure Control," UK Office for National Statistics (ONS), accessed 29 August 2021, <https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol>
- 9 Holm Tetens, "Reproducibility, Objectivity, Invariance," in *Reproducibility: Principles, Problems, Practices, and Prospects*, eds Harald Atmanspacher and Sabine Maasen (Hoboken, NJ: Wiley, 2006), 13–20.
- 10 See UK Data Service, "Research data management," accessed 31 August 2021, <https://www.ukdataservice.ac.uk/learning-hub/research-data-management/>.
- 11 Abid Hasan, "Ethical considerations in the use of secondary data for built environment research," in *Secondary Research Methods in the Built Environment*, eds Emmanuel Manu and Julius Akotia (Oxfordshire: Routledge, 2021), 26–39.
- 12 Hasan, "Ethical considerations," 26–39.
- 13 European Commission's Information Society Technologies (IST), "RESPECT Code of Practice for Socio-Economic Research," (Brighton: Institute for Employment Studies, 2004): 1–4.
- 14 UK Data Service, "Research data management."
- 15 European Commission's Information Society Technologies (IST), "RESPECT," 1–4.
- 16 See UK Data Service, "Anonymising quantitative data," accessed 31 August 2021, <https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymising-qualitative-data/>
- 17 UK Data Service, "Anonymising quantitative data."
- 18 "Complete Guide to GDPR compliance," GDPR, accessed 24 June 2021, <https://gdpr.eu/>
- 19 "Guide to the UK General Data Protection Regulation (UK GDPR)," Information Commissioner's Office (ICO), accessed 24 June 2021, <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>
- 20 Information Commissioner's Office (ICO), "Guide to the UK GDPR."
- 21 UK Office for National Statistics (ONS), "Disclosure Control."
- 22 Hasan, "Ethical considerations," 26–39.
- 23 UK Office for National Statistics (ONS), "Disclosure Control."
- 24 Hasan, "Ethical considerations," 26–39.
- 25 See UK Data Service, "File Sharing," accessed 31 August 2021, <https://www.ukdataservice.ac.uk/learning-hub/research-data-management/collaborative-research/file-sharing/>
- 26 UK Data Service, "File Sharing."
- 27 See University College London, "UCL Data Protection Policy," (2017), accessed 24 June 2021, <https://www.ucl.ac.uk/library/sites/library/files/data-protection-policy.pdf>