United Nations ECE/CES/2023/24

Economic and Social Council

Distr.: General
13 June 2023

English only

Economic Commission for Europe

Conference of European Statisticians
Seventy-first plenary session
Geneva, 22–23 June 2023
Item 8 of the provisional agenda
Data ethics – a key enabler of social acceptability

An ethical approach to the development of social acceptance strategies for national statistical offices

Prepared by Canada, Ireland, United Kingdom and Eurostat

Summary

This document defines the concept of social acceptability, explores how it relates to data ethics and to the work of national statistical offices, and considers how it is a crucial component in building trust with users and data suppliers.

The document presents learnings from the experiences of other national statistical offices who have embedded this concept within their processes and procedures. It also suggests how this work could be progressed to support all NSOs to build social acceptance with their users and data suppliers.

The document is submitted to the Conference of European Statisticians' session on "Data ethics – a key enabler of social acceptability" for discussion.

I. Introduction

1. As national statistical offices (NSOs), we regularly comment that the environment in which we work is dynamic and changing at pace, with increasing requirements for more frequent, more granular and more timely outputs. One component of our work which has not changed is the requirement for us to be trusted by our data suppliers and our users. What has changed in this sphere, is the increasing importance of providing actual evidence of the nature of our trustworthiness and independence to the satisfaction of data suppliers, users and society, in a world increasingly bombarded with fake news, disinformation and biases. Our work is increasingly important as an independent, trustworthy source of data and facts and our ability to demonstrate and evidence our independence, impartiality and trustworthiness is crucial.

2. Traditionally NSOs have been physically closer to their data suppliers. In a primary collection model, consent was given to use the data by virtue of the individual or company agreeing to take part in a primary survey and the end use of the data, the creation of official statistics was transparently clear. In this model, the data supplier has a clear link to the NSO and a clear idea of what the data is being used for. The primary survey information will have explained how the data will be anonymised and how confidentiality will be maintained, providing comprehensive assurance for the data supplier directly from the organisation collecting the data.

3. Increasingly NSOs are using administrative or secondary data, i.e., data that was collected by someone else, either government departments, research offices or the private sector (privately held data) for the purpose of creating official statistics. The physical link with the original data supplier is gone. When the data supplier provides their data, it is not for the purpose of creating official statistics but rather to complete an administrative process such as for example, applying for a driving licence or completing a tax return. The engagement is not directly with an NSO and the data supplier's focus is the completion of an administrative task rather than the creation of official statistics or of a data set for use by researchers.

4. Of course, the usage of administrative data for official statistics is not new and the importance of using this data is not only recognised in statistical legislation but also in the General Data Protection Regulation Article 89 (1) which details safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes and Article 5 (1)(b) which allows that personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89 (1), not be considered to be incompatible with the initial purposes.

5. NSOs have, of course, long experience in and a record of, dealing with confidentiality (in particular when data is collected by the NSOs) and of maintaining the trustworthiness of official statistics. However, social acceptance is more linked to privacy aspects, in particular for secondary data. In this scenario, where the NSO is removed from the original data collection process, it is crucial that the NSO has built a reputation of trustworthiness. As NSOs link and match datasets from multiple diverse sources to create richer, more insightful outputs, there is an increasing responsibility on us all to communicate with and assure society that this work is being done not just in compliance with the legal right to use the data but beyond that, it is being done for the public good and in the public interest.

6. NSOs are also increasingly expanding their role beyond the creation of official statistics to one where we provide data services to others including government bodies and researchers. The expansion of the role introduces ethical challenges, as the NSO remains accountable for the responsible use of data throughout the entire data lifecycle. As such, each NSO must be able to provide and communicate the evidence of the ongoing protection and confidentiality of the data and on their trustworthiness.

7. As the world of NSOs becomes increasingly complex, with new data sources, new tools and new demands, the concept of data ethics is becoming increasingly important. The expression is relatively new, but the principles have been around for a long time. In the new

context, NSOs have to be more transparent and communicate more openly to the public what data is being collected/acquired, for what purpose and measures to be taken throughout the lifecycle to ensure ethical use.

8. The Conference of European Statisticians' (CES) recently agreed Core Values also speak to ethical considerations in recognising that NSOs must be impartial, transparent, professionally independent, respecting of confidentiality and collaborative. These values and the behaviours currently being identified to support each value, also speak to building trust with our community of users and respondents.

9. The importance and relevance of data ethics can be seen in the way the idea cuts across the current work of several expert groups and initiatives under the CES, including the High-Level Group on Modernisation of Official Statistics (HLG-MOS): (i) modernising statistical legislation, (ii) core values of official statistics, (iii) data stewardship and (iv) social licence (now social acceptability). As an additional complication, data ethics can be context dependent and dynamic. National legislation and cultural contexts can impact on the meaning of data ethics. For the purposes of this paper, data ethics can be viewed as the responsible use of data throughout the data lifecycle and as a key enabler to build trust with our users and data respondents.

10. It is not sufficient for NSOs to be confident that they are acting legally and in the public interest. We must be able to prove that data ethics is a consideration across our statistical processes. Without the trust of data suppliers and users it is difficult for NSOs to deliver for their societies and for their democracies.

II. What is social acceptance?

11. Terms such as social acceptance and social licence have begun to be used in the discourse around data ethics and have been used interchangeably and perhaps without a clear understanding of what they mean. The term social licence originated in the context of mining exploration, where a legal right to excavate was not sufficient to proceed with the mining excavation processes without consideration being given to obtaining the local community's approval to proceed with the project.

12. The term social licence also began to be used in a broader way to describe how any organisation could engage with a community to gain cooperation and acceptance. Increasingly the term "social acceptability" is being used in the context of data to describe the broader social and ethical responsibilities of official statistics beyond simply being legally compliant and having the legal authority to access data.

13. Organisations are looking to gain society's willingness to accept and support for their work. For NSOs it means gaining and retaining society's approval for our work and retaining trust. It can be seen as a standard beyond mere compliance with personal data protection laws and beyond legislation governing statistical organisations. It is about using data in a responsible way, embedding ethical considerations into statistical production and organisational processes. It is important that society accepts that not only does the NSO have the legal right to use the data, but that the NSO also has society's trust and acceptance to use the data, to provide statistics and services in the public interest as they are a public good. The alternative is that we allow non-official data and, in many cases unregulated providers to inform decision makers and the public alike which carries obvious risks. We have all seen how quickly fake news and disinformation can spread throughout societies and countries and the detrimental impact that this can have on societies.

14. At a meeting of the CES Bureau in February 2023 these two terms, social licence and social acceptance, were discussed. The Bureau agreed that when it comes to NSOs, the term social acceptance is more appropriate in the context of the trust we want to maintain and build with society and our data providers.

15. The term social acceptance might be relatively new, but NSOs have always needed the trust of data suppliers and users, whether they be Government departments, businesses, or the public, to do our work. Social acceptance is a new term for the trust that has been built

up over many years and the recent focus on the term should be seen as an opportunity for NSOs to overtly demonstrate why people can trust us with their data.

III. Why are we talking about social acceptance now?

16. Traditionally, the crucial consideration for NSOs was whether or not we had either a national or European legal right to access the data. This has been the key measurement against which we held ourselves accountable. The General Data Protection Regulation (GDPR) built additional considerations into that legal perspective. Under GDPR, NSOs must be able to demonstrate that the access to the data source is legal, that it is necessary to achieve a defined goal, that it is proportionate, in that only the data needed is accessed and that it is used for the stated purpose. More specifically, personal data protection aspects, such as data security, data traceability, and data access should be a key part of the design of any data collection or (re)use of data for statistical purposes. This framework strengthens the rights of data suppliers and the obligations of data controllers (organisations that collect and process the data).

17. Social acceptability is increasingly becoming a context against which we have to hold ourselves accountable. However, it is more nebulous. Unfortunately, NSOs may only become aware that they do not have society's acceptance to proceed when they have been perceived to have broken the trust. In a case like this, the NSO can have complied with all legal measures in accessing data, but society is uneasy or nervous of the level of access being sought regardless of the legal entitlement to the data. No NSO wants to find itself reacting to a situation where society considers that trust has been breached.

18. The idea of social acceptance is growing and gaining importance because of the changing environment we all work in. In the European Statistics System (ESS), we are increasingly using administrative data and looking to acquire legal access to privately held data via either national legislation or using the revision of Regulation (EC) no 223/2009 of the European parliament and the Council on European Statistics. The pandemic also accelerated the use of administrative data in many NSOs with the loss or reduction of primary collection processes. So, increasing access to data, increasing use of new data sources and increasing awareness within our societies of the nature and volume of the data we hold creates new data concerns.

19. NSOs operate in an increasingly open, hyper-connected world where the general public are more aware of who has access to their information and the dangers that might bring. This has resulted in greater awareness and sensitivity around issues such data privacy, data linkage, and data protection. At the same time, NSOs are combatting the growth of fake news and misinformation. This all means that NSOs are having to work harder to promote trust in official statistics and provide sound reasons and evidence for needing data for the public good.

20. This work started for many NSOs with legislation that put confidentiality at its core. The principles we adhere to and our core values also form part of that clear pathway to trust. Social acceptability is another stepping-stone to build trust in the work of NSOs. Given the lack of direct involvement of the general public at the data collection stage, the reuse of privately held data for statistical purposes requires social validation1. Social acceptability means we must be able to show the societal benefit of accessing data or linking data sets, be able to clearly explain those benefits and be able to demonstrate that we are holding ourselves accountable to a standard above the legal basis and delivering for society in an ethical and trustworthy way.

[1] See report of the expert group on facilitating the use of new data sources for official statistics "Empowering society by reusing privately held data for official statistics – a European approach", Luxembourg, 2022, ISBN 978-92-76-53600-0

IV. Development of social acceptability by building data ethics into our process and procedures – examples

21. We can all point to examples within the NSO community and beyond where perceived non-compliance with this nebulous notion of social acceptability has real consequences, when society deems that a line has been crossed and, regardless of any legal entitlement to the data, withholds its approval for its use. We need to build new structures into our procedures to ensure that we have the evidence to support our consideration of data ethics and the concept of delivering for the public good, to ensure that we have a wider societal approval for our work.

22. NSOs have been grappling with how to build trust for many years and considerable thought has been given to how NSOs can earn trust and acceptance of their work. The following examples look at how two NSOs applied an ethical lens to their organisations. They consider how they embedded ethical considerations into their processes and procedures and consider how they developed the evidence to support their claim of trustworthiness and to show why society should accept, support and tolerate their work.

23. One organisation included the ethical use of data in their key legislation and built it into their processes from the start. The other NSO responded to the withdrawal of society's acceptance to a number of projects by developing an ethical lens to inform the development and implementation of innovation and to embed the ethical considerations into their processes.

24. Key to both approaches is the development of tailored messages designed to promote broader understanding of the societal benefit of the work of NSOs.

A. Statistics Canada

25. In a modernisation effort to better respond to growing needs for more detailed and more timely information, Statistics Canada moved to an administrative data first paradigm. This paradigm shift led to new data acquisitions and new partnerships at a fast pace. However, during the development of a pilot project using personal banking data, despite having legal authority to develop such a project, it became apparent the public was not ready to go ahead with a project using such sensitive data, at least not without further consultation. This project was redesigned in collaboration with the Office of the Privacy Commissioner and international partners.

26. Having learned for this experience, Statistics Canada developed and implemented a Necessity and Proportionality Framework. The framework is an adaptation of the scientific approach to optimise privacy protection and the production of information when designing a data-gathering approach. Necessity is the principle related to data needs, who requires the information and the reasons why such information is needed. Proportionality is the extent of the effort needed to obtain the needed information in a manner that is coherent with the expected benefits of a project. Considering privacy intrusion, other ethical issues and the data quality, the proportionality principle ensures that no more information than what is needed to produce the expected benefits is going to be collected.

27. Concretely, under this framework, every data gathering proposal that involve surveys, secondary data sources (such as administrative or other alternative data sources) and record linkage initiatives must explain why it is important, what the benefits are to Canadians, who needs the information and address ethical considerations such as privacy, transparency, and fairness.

1. Data Ethics Secretariat

28. The Data Ethics Secretariat was created to implement the Necessity and Proportionality Framework into Statistics Canada's day-to-day activities. Among these activities, the Data Ethics Secretariat performs ethical reviews on new projects or data acquisitions, by holding discussions with program managers and raise ethical considerations. These ethical considerations are anchored by six guiding principles:

- Benefits for Canadians: Description and illustration of the expected benefits with concrete examples.
- Privacy and security: Description of measures taken to avoid a security breach and potential impacts should a breach happen.
- Transparency and accountability: Description of the proactive steps taken to be transparent.
- Trust and sustainability: Description of relationships with partners and data providers to determine if the relationships present a risk to undermine the public trust and if these relationships are sustainable through time.
- Data quality: Description of known sources of error and ways envisioned to address them. The expected benefits should be proportional to the quality of the data.
- Fairness and do no harm: Overall view of the initiative to ensure harmful profiles that could propagate stereotypes are not created. This is especially important for vulnerable or marginalised segments of the population.

29. The result of a review is a recommendation to the Principal Data Ethics and Scientific Integrity Officer, whose role is to ensure that Statistics Canada's scientists, researchers and managers uphold data ethics, conform to standards of scientific excellence and conduct themselves in a manner consistent with the principles of scientific integrity.

30. More sensitive cases are brought to an internal Data Ethics Committee with members from different backgrounds, with different points of view. Members come from all sectors of the organisation: social statistics, economic statistics, data collection, IT security, office of privacy management, communications, etc.

31. The Data Ethics Secretariat works on the redaction of documents on specific topics such as ethical considerations for some segments of the population, for example minors or vulnerable, marginalised populations. It also holds discussions with partners within and outside the organisation to inform or train Statistics Canada colleagues on data ethics principles, but also to refine and improve our process by learning from partners in Canada and abroad.

32. Data ethics also includes AI/ML ethics. Projects involving data science techniques are also considered from an ethical viewpoint and subjected to ethical reviews. Moreover, a Framework for Responsible Use of Machine Learning was developed to review the various uses of AI/ML techniques and ensure that they follow ethical values and principles. The four pillars of this framework are:

- Respect for people
- Respect for data
- Sound methods
- Sound Applications.

2. Advisory Council on Ethics and Modernisation of Microdata Access

33. The Advisory Council on Ethics and Modernisation of Microdata Access was also created to provide Statistics Canada with the appropriate guidance on data access, privacy and data governance to maintain and support the data needs of Canadians. The knowledge and experience that the members of the Advisory Council bring benefit the agency as Statistics Canada works to facilitate access to anonymised microdata for researchers, improve data security, and risk management protocols. Ethical issues brought before the council are not specific or transactional cases but rather concern general themes and directions for the agency. The council meets twice a year and reports are made available to the public.

3. Trust Centre

34. Statistics Canada takes proactive steps to be transparent with the "Trust Centre", available on its website. Via the Trust Centre, the public can learn about how we collect,

process, analyse and share data that benefits them. It includes sections specifically dedicated to our privacy protection practices and ethical principles.

B. Office of National Statistics of the United Kingdom

35. Data ethics is a key principle in the UK's Digital Economy Act (DEA) Research Strand 2017 that provides a legal gateway for public authorities to make deidentified data available for research for the public good. All parties involved in the disclosure, processing or use of data accessed via the DEA are required to observe ethical standards appropriate to the nature of the research.

36. To support the analytical community to collect and use data in ethically appropriate ways, the United Kingdom Statistics Authority (UKSA) has produced the following ethical principles.

- The use of data has clear benefits for users and serves the public good.
- 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.
- 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.
- 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.
- The views of the public are considered in light of the data used and the perceived benefits of the research.
- The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

37. There is little point in having data ethics principles if they are not going to be used by analysts and therefore have no impact on data collection and use practices. To ensure wide scale and consistent use of data ethics principles, analysts need high quality support to empower them to confidently use the principles. Therefore, to ensure these data ethics principles are widely embedded in statistical and research processes across the UK statistical system, the UK Statistics Authority has established a Centre for Applied Data Ethics (CADE). The CADE focuses on making data ethics accessible to analysts from across government, academia and the commercial sector, so that they can efficiently adopt an ethics by design approach to collecting and using data for analytical purposes. This has involved moving beyond just developing and communicating high-level ethical principles to developing resources to help researchers efficiently apply these principles in their work. To this end, the CADE has developed an ethics self-assessment tool that breaks down the ethical principles into easily digestible individual components and enables analysts to quickly assess their project against different aspects of each ethical principle. Through this self-assessment tool, ethically high-risk projects are quickly identified, and these projects are scrutinised at pace by the independent National Statistician's Data Ethics Advisory Committee (NSDEC) who provide the National Statistician with independent expert ethical advice on the collection and use of data for analytical purposes. To aid transparency, details of this independent scrutiny are published on the UKSA website.

38. This data ethics work is supported by training, guidance and a dedicated user support team who provide oversight of use of the tool to ensure appropriate use. This means the analysts are supported throughout the period they are considering data ethics which helps to ensure that this is a relatively robust and efficient process. Currently, researchers who submit an ethics application have an outcome within two working days. This support, training, guidance also helps to educate analysts to view data ethics as a vital part of the research process that potentially adds significant value to their research rather than a bureaucracy that must be completed to merely fulfil governance requirements. Ensuring that a culture that

recognises the importance and value of data ethics is prevalent across the analytical community is as important as establishing robust data ethics frameworks and processes.

39. To ensure that these resources are visible across stakeholders, the CADE has widely presented these services at conferences and engagement events, communicating case studies showing how these services have been successfully used on a range of high-profile research projects. Being able to communicate examples of data ethics in practice is key in showing stakeholders that the UK Statistics Authority goes beyond talking about data ethics to actually applying good data ethics practices in their data collection and use practices. This helps build social acceptability for the UK Statistics Authority's work. Widely communicating this data ethics work to other actors in the data ethics space also helps to influence wider data ethics thinking and shape the future of data ethics. Data ethics is not going away so it is important that the voice of the national statistical office is heard in shaping interpretations of data ethics to mitigate the risk of alternative interpretations being imposed upon the UK Statistics Authority by other interested actors in the data ethics space.

V. Building social acceptance

40. In order for NSOs to demonstrate our trustworthiness and to build social acceptance, society needs to understand that NSOs use data for the public good and in ways that benefit society. They have to trust that NSOs will protect their data and have reliable systems and processes in place to ensure that protection. Society must be able to see that our approach to the use of their data is open and transparent and society has to see that NSOs actively engage with them and take their views regarding our use of their data into account. In these ways we can build their trust and earn social acceptance for the work we do. These steps are central to the concept of developing social acceptance.

A. Communicating the unique role of national statistical offices

41. NSOs are working in a world where many organisations, both public and private, have access to huge volumes of data. Within this group we are unique in that we are legally mandated to collect, compile and link fragmented data sources in order to publish Official Statistics. We provide citizens and decision-makers with trustworthy, independent, high-quality information on the economy and society. The data is freely available to all and produced without interference from Government or other groups and with no agenda other than to allow our citizens to live in an informed society and to help inform public debate. The insights we publish and the work we do are a public service.

42. The NSOs must ensure the public knows why we are different from other data providers. It is crucial that we explain and communicate this difference, our lack of a political agenda, our impartiality, our independence and our commitment to high quality, at both national and European level.

B. Communicating why increasing innovation in national statistical offices is for the public good

43. NSOs are increasingly innovating to find new ways of providing important insights into societal and economics issues. This need is being fuelled by the growing demand for better and more timely insights into complex and evolving societal and economic issues and made possible by a diverse data landscape. Increasingly, NSOs are relying on linkages between administrative data sources and data held privately to enable them to develop innovative, timely insights for all. The push by NSOs to use data held by third party entities continues, and if successful, will result in ever greater volumes of primary, administrative and commercially held data being used by NSOs for official statistics.

44. The data we hold could be deemed as some of the most sensitive and confidential data in the world. NSOs need to ensure they are clearly communicating and demonstrating why they can be trusted with that data, to ease the concerns of the public and the stakeholders that hold the data.

45. However, the innovations designed to meet these demands can create new challenges for statistical organisations, particularly around legitimacy. Being able to access these new data sources and reassuring people about how we protect the confidentiality of their data not only ensures NSOs stay relevant but supports our future existence.

VI. Importance of clear communications

46. Data ethics is now an important part of the conversation. It is clear that NSOs need to be able to transparently demonstrate the practical application of an ethical perspective which transcends our legal rights to data. One approach is to build a transparent, ethical framework that informs all aspects of data collection, processing and dissemination.

47. As NSOs we need to take the time to think about how each output provides insight and be able to explain clearly why that particular insight is needed. The public good must be clearly demonstrated and communicated in a transparent manner.

48. NSOs need to be able to calm legitimate fears about over-reach and data protection and the best way to do that is to tell society what you are doing, why you are doing it, how you are doing it and where they can find out more information in a proactive transparent way.

49. NSOs must be able to provide the practical evidence to demonstrate that they are considering their work using an ethical lens.

50. When society is unprepared for something new, it can lead to uncertainty, questions, doubts, and fears. Where a society understands the benefits of the development of a new statistical product, or of accessing or merging datasets, it is more likely to be comfortable with the NSO proceeding. Communicating on the societal benefits in ways that society can understand and can access easily can help allay these concerns and pave the way for social acceptability.

VII. Recommendations

51. It is important that NSOs should socialise the ideas around data ethics and the social acceptability of the work we do both internally with our own staff and externally to our data providers, users and society using examples of how the concepts are practically applied to the work we do.

52. NSOs should clearly define what we mean by the public good and the public interest, with concrete examples to demonstrate what is meant.

53. NSOs should strengthen the engagement with citizens and business associations regarding the issue of reuse of privately held data.

54. NSOs should consider how they can demonstrate that delivering insights for the benefit of society in a way that protects the confidentiality of data is at the core of what we do in a very concrete way.

55. NSOs should consider how to communicate the message that we are interested in the totality of the data rather than in the individual, that we anonymise, apply disclosure controls and protect confidentiality to combat privacy concerns.

56. Statistical authorities should seek to improve public data literacy in connection with the reuse of private data for public purposes. The actions envisioned should target citizens as well as businesses.