UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting of the 2022/2023 Bureau
Cork (Ireland), 12-13 October 2022

ECE/CES/BUR/2022/OCT/2
27 September 2022

For discussion and
recommendations

Item 2 (a) of the Provisional
Agenda

IN-DEPTH REVIEW OF DATA ETHICS

Prepared by Canada and the United Kingdom with contribution by Eurostat

This in-depth review was mandated by the CES Bureau and it explores the state of development of data ethics, what it encompasses and what should be the way forward. The paper provides the definition of data ethics, summarises the state of practice and development of data ethics in a number of countries and statistical professional organisations, explores the related issues and challenges, and presents conclusions and recommendations.

Additional short notes from the CES Bureau members are welcome in advance of the Bureau meeting.

The Bureau is invited to discuss the issues, challenges, conclusions and recommendations identified in this paper, and consider the need for follow-up work.

I.   BACKGROUND

1.   As societies keep on changing, they become increasingly complex. As a result, the world of official statistics needs to constantly adapt. With refined, more diverse, timely and larger quantities of data needs, National Statistical Offices (NSO) are faced with having to expand much beyond the traditional ways of organizing statistical programs. To this end, NSOs are now going outside the more traditional survey-taking approach and are using new alternative data sources and are expanding their methods and toolkits to data science and modern approaches to gathering and integrating data. Such broad and rapid advancements are bringing to the fore the importance of ensuring that all this be conducted in ethical ways, hence the need for data ethics.

2.   While the term data ethics has only been in use for a few years, the concept of data ethics and the underlying ideas are not completely new. In their processes and culture, NSOs have been guided by values and principles such as the importance of producing professionally guided statistics free from interested influences; avoiding or limiting bias; the importance of protecting confidentiality of data; and the importance of grounding methods on science. Traditionally however, when thinking about ethics, NSOs have focused on business ethics as it relates to the behavior of staff and managers, and when referring to data the focus has mostly been on privacy and confidentiality aspects. For example, NSOs did not question so much the issues related to data acquisition (beyond legal) but rather were concerned with ensuring security of data once within the NSO and limiting access to microdata.

3.   With society being more involved with and aware of data through the increased presence of the digital and virtual world, people have come to both appreciate the importance of the data and fear the potential negative effects of their use. In this context, NSOs cannot take for granted the social acceptance of their statistical activities, whether they be new and innovative or even traditional. Social acceptance needs to be earned, nurtured and regained when reduced for all kinds of reasons directly related to the actions of the NSO or not related to that at all. To earn the trust and social acceptance of society, data ethics has an important role in guiding NSOs in the realm of what can be done versus what should be done.

4.   In a sense, many aspects of data ethics have been in place in NSOs for a long time. For instance, there has always been a great deal of attention devoted to protecting data and ensuring that only those that need to have access to detailed records do have such an access. This mainly materialized in terms of IT security and protection of confidentiality. However, with the new possibilities to gather data, combined with a number of greatly publicized breach events that happened in large international organizations, the public is seeing an increased responsibility on the side of NSOs to ensure that data are more comprehensively ethically gathered and guarded. Public acceptance now requires enhanced data ethics processes and increased transparency throughout the data lifecycle[1].

5.   Data ethics is related to all aspects of the data in all steps of statistical programs. For instance, there can be ethical decisions needed in all steps of the GSBPM[2]. Right from the start of the design of a program, questions will arise on which data to collect and/or acquire. Even if certain types of data or data sets would be perfectly adequate to respond to the data needs, society may not understand the importance, may not be willing that such data be used or may be sensitive to exactly which part will be used and how. When collecting data via surveys, the process appeared to be more transparent as respondents knew which information was gathered about them. Now with administrative or alternative types of data, transparency needs to take on new forms and additional considerations must be given to other ethical values.

6.   To illustrate data ethics, if an NSO is to consider the acquisition of new types of data, particularly in the case of micro-data, it would be considered unethical to simply lean on the legal provisions (if they exist) and not pro-actively provide the information to citizens. Further, the importance of the need and the public good must be clearly present and demonstrated. On the other hand, a fully transparent approach that would clearly lay out what is to be acquired and why it is necessary is an ethical approach.

7.   When considering the international community, the development and implementation of data ethics approaches and methods has been unequal across countries. As it is still a new topic, data ethics needs to be well defined and developed further. This review is about exploring the state of development of data ethics, what it encompasses and what should be the way forward.

8.   This review is organized into the following sections. Section II introduces the purpose and scope of the review. Section III provides the definition and scope of data ethics and how it is related to other aspects of ethics. Sections IV presents the state of practice and development of data ethics in a number of countries as well as statistical professional organisations. Then Section V explores the issues and challenges related to data ethics and Section VI presents the conclusions and recommendations.

---

[1] Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical journal of the IAOS* 35, 549-558. DOI 10.3233/SJI-190581. IOS Press.
[2] UNECE (2022). Generic Statistical Business Process Model. Generic Statistical Business Process Model - Generic Statistical Business Process Model - UNECE Statswiki.

II.   INTRODUCTION

9.   The Bureau of the Conference of European Statisticians (CES) regularly reviews selected statistical areas in depth. The aim of the reviews is to improve coordination of statistical activities in the UNECE region, identify gaps or duplication of work, and address emerging issues. The review focuses on strategic issues and highlights concerns of statistical offices of both a conceptual and a coordinating nature. The current paper provides the basis for the review by summarizing the international statistical activities in the selected area, identifying issues and problems, and making recommendations on possible follow-up actions.

10.   The CES Bureau selected "Data Ethics" for an in-depth review for its October 2022 meeting. UK and Canada, with guidance from the UNECE Ethical Leadership Task Team were requested to prepare the paper providing the main basis for the review.

11.   A number of countries (UK, Canada, Albania, Hungary) as well as Eurostat and the International Statistical Institute were consulted. Further, the context of many NSOs was considered through a survey on ethics that had been conducted in 2021 by the Ethical Leadership Task Force of the UNECE headed by Italy and Canada.

III.   SCOPE, DEFINITIONS AND LITERATURE REVIEW

A.   What is data ethics?

12.   Definitions of data ethics build on the conceptual shift from computer to information ethics, where there is now a focus on the information that is created by computers and other technologies, rather than the technology itself (Floridi and Taddeo 2016). Given the wide range of data types and methods, and that data use and applications vary dependent on culture and contexts, it is difficult to broadly define data ethics. General definitions relate to moral problems and providing moral solutions in relation to data, algorithms and practices (Floridi and Taddeo 2016).

13.   Common issues in data ethics include de- and re-identification, trust and transparency, responsibility, consent, and the secondary use of data. It is argued that data itself are not an ethical concern, but concern comes from how data are used, particularly as ongoing technological advancement means the future uses of data are unknown (Hand 2018). For example, matching detailed information on spouse or on children could lead to issues of fairness and potentially some forms of stigmatization such as when asking questions on child abuse or violence. In this case, the ethical approach could be to limit what is asked, mitigate the process or simply re-think the approach or in some cases even cancel the project. Conversely, it would be unethical to simply proceed and not consider ethical issues. A balance between the risks and benefits is required to maximise the opportunities that come from new technologies and the use of data. Floridi and Taddeo (2016) argue that ethical regulation can inhibit the value of data science in society, but that this should be balanced by appropriately addressing ethical concerns to reduce potential harms.

B.   Artificial Intelligence and data ethics

14.   Other applications of data ethics include artificial intelligence (AI) ethics and, as a method for processing and analysing big data, there are specific ethical considerations. AI ethics has developed from the increasing availability of big data, capacity for storage and advancement of

machine learning algorithms. The rapid evolution of machine learning and algorithmic capability means that new and unexpected data ethics concerns may develop and lead to harm.

15.   Ethical guidelines on the use of AI aim for the technology to be harnessed in a safe way. The Alan Turing Institute[3] guide for responsible design and implementation of AI systems states that, to avoid unanticipated harms, human responsibility for the design and use of AI is important to safeguard and protect individuals in society who may be affected by its impacts. However, it is argued that general guidelines and principles can be vague given the differences and complexities of AI use across contexts, and that rigid frameworks, particularly within law, may limit the use of AI (Hagendorff 2020), and to what extent could the impacts of this be unethical.

C.   Business and data ethics

16.   Business ethics relates to the values and moral obligations of organisations in terms of corporate responsibility, organisational values and employee conduct. Here the context is related to how employees and/or managers would behave in front of specific situations. So, it has subject matter dimension and a human resources dimension. This intersects with data ethics when the data practices of organisations are considered, such as the collection, storage and use of data. Much of this relates to data held by companies on their customers. The Institute of Business Ethics[4] points out that companies 'have a responsibility to promote transparency and prevent misuse of personal data' as misconduct can affect 'reputation, customer relationships and ultimately revenues'.

17.   There are debates around the reuse of corporate data, often personal data from customers, especially for purposes different to when it was originally collected. This is a general issue faced by NSOs, in terms of wider data repurposing. A further argument in business data ethics is around the ownership of data collected and held by private organisations. Access to corporate data may be controlled by organisations when it could be used to help measure what is missed in national statistics, ultimately being useful to government decision-making which could benefit wider society (Taylor 2016).

D.   Privacy and data ethics

18.   Data ethics considerations related to privacy include confidentiality, informed consent, trust, transparency and the appropriate use of personal data. It is argued that privacy is an ethical issue and that privacy concerns with data in research are about the ethical consequences of its use and the rights of individuals (Kisselburgh 2020). While data use is regulated by privacy laws and personal data is protected through anonymisation and aggregation, it is argued that aggregate data can still be sensitive given the ability to predict personal characteristics, like age and gender, to a high degree (Taylor 2016).

19.   Data ethics policy initiatives can complement data privacy legislation, but the meanings of data ethics in a privacy context are complex and dependent on culture, meaning it can be difficult to define data ethics in a privacy context (Hasselbalch 2019). Data privacy legislation can also be difficult to apply to emerging technologies and new uses of data (Mulligan, Koopman and Doty 2016). While laws address some elements of data ethics, other considerations around moral problems are harder to define, particularly as understandings and public attitudes of privacy, data use and ethics shift (Hand 2018).

---

[3] https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf
[4] https://www.ibe.org.uk/resource/business-ethics-and-big-data.html

E.   **Fundamental principles, core values and codes of practice**

20.   Data ethics is codified in the principles, values and codes of practice of international and national statistics organisations. Adopting and adhering to such principles ensure that statistics are produced and used in an ethical way for benefit, and that reduces harm. The United Nations Fundamental Principles of Official Statistics highlight common considerations related to data ethics like confidentiality, equality, impartiality, accountability, transparency and preventing of data misuse. There is an emphasis on the responsibility of statistical agencies to ensure their data practices are ethical by following methods and procedures for collecting, processing, storing and presenting data and that 'official statistics - reliable and objective information - is crucial for decision making. These principles were developed and adopted in the UNECE region 30 years ago and subsequently by the United Nations Statistical Commission in 1994, and in January 2014 by the United Nations General Assembly.

21.   In 2022, core values of official statistics were agreed and endorsed by the Conference of European Statisticians[5]. These core values upheld by the official statistics community include being relevant (e.g. "though an engaged, responsive, user-centric approach, based on clear and regular communication with all stakeholders"), impartial (e.g. "s objective, in all aspects of our work, acting with fairness and integrity to serve the public good") and transparent ("official statistics, methods, processes, products and quality reports are communicated to the public through appropriate channels, and are open to scrutiny") as well as respecting confidentiality ("protects privacy by ensuring that data collection is limited to what is necessary and proportionate; upholds the confidentiality of data on individual people, households and businesses; acts in an ethical way, following professional ethics standards"). As a national example, the Office for Statistical Regulation Code of Practice for Statistics in the United Kingdom sets a standard for the producers of official statistics. The Code of Practice is based on three pillars – trustworthiness, quality and value. An element of the 'trustworthiness' pillar is that data ethics standards are met in order to be transparent about how data is managed and used, recognising the rights of data subjects.

IV.   **COUNTRY PRACTICES**

A.   **Data Ethics at Statistics Canada**

22.   Statistics Canada has always adhered to the highest ethical standards. With all the changes in society and the increasing digital world approaches have to be modernized to remain relevant in terms of information provided to citizens, and also to ensure the level of ethics is maintained. In its response to better serve society with data insights, Statistics Canada shifted its focus from a survey-first approach to an approach that prioritizes the use of non-traditional or alternative data sources. The use of these types of data has been at the core of the modernization initiatives at Statistics Canada. Along with the use of non-traditional data sources, technological advances have also taken place. For example, Machine Learning and web-scraping capacity are all techniques that can be used to obtain and understand data. But with this innovation has come new ethical questions. Just because it is possible to do something does not mean it should be done, even when equipped with a legal mandate.

23.   It is in this context that Statistics Canada has adopted a Necessity and Proportionality Framework and established a Data Ethics Secretariat as well as governance committees to ensure an appropriate use of data throughout its life-cycle.

---

[5] https://unece.org/sites/default/files/2022-07/ECE_CES_2022_2-2211176E.pdf

Necessity & Proportionality Framework

24.   In 2019, Statistics Canada developed and implemented a Necessity and Proportionality Framework. The framework is an adaptation of the scientific approach to optimize 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. Taking into account 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. Under this framework, every proposal for a new project or data acquisition 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. At the beginning, the framework was applied only to data gathering initiatives that involved administrative and alternative data, but was quickly extended to surveys. Further, the framework is also applied to record linkage initiatives and now to any project that involves new data or new uses of data.

Data Ethics Secretariat

25.   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
- Privacy and security
- Transparency and accountability
- Trust and sustainability
- Data quality
- Fairness and do no harm.

26.   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.

27.   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 organization: social statistics, economic statistics, data collection, IT security, office of privacy management, etc.

28.   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, marginalized populations. It also holds discussions with partners within and outside the organization to inform Statistics Canada colleagues but also refine and improve our process by learning from partners in Canada and abroad.

29.   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 responsible ML framework, which includes both respect for people and respect for data, was developed to review the various uses of AI/ML techniques and ensure that they follow ethical values and principles.

Advisory Council on Ethics and Modernization of Microdata Access

30.   The Advisory Council on Ethics and Modernization 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 anonymized 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.

B.   Data Ethics at the Hungarian Central Statistical Office

31.   The Hungarian Central Statistical Office (HCSO) handles ethical issues on a case-by-case basis and is looking to integrate these procedures into a wider ethical framework covering both statistical production and institutional operations. As a member of the UNECE HLG-MOS Ethical Leadership Task Team, the HCSO is following with interest the best practices and experiences of other countries to inform the development of its own ethical framework.

32.   The current ethical practices of the HCSO pertaining the data life-cycle can be parsed in terms of the input, throughput and output phases of the statistical business process.

33.   The HCSO only requests data that is strictly necessary **input** for the development, production and dissemination of official statistics[6]. In rare cases where data providers hand over a surplus of information to which the Office is not privy, the data is deleted or returned to the providers with a request to only send the data to which the statistical office has justified access. While acknowledging the burden on data providers, the course of action is determined by ethical considerations taken by the head of the relevant unit and backed by the legal mandate of the Office.

34.   As regards the **throughput** phase, the staff of the HCSO is entrusted with the processing of the data solely for the production of official statistics[7]. In addition to signing confidentiality agreements for the protection of data, HCSO employees are prohibited from using the data for personal gain and for other purposes than production of given statistics that falls within the responsibility of the given person and unit. There are IT security and access controls in place to prevent unauthorized access. In our experience, these safeguards are secondary to a workplace culture that is defined by integrity and the example set by ethical leadership.

35.   The **output** of the statistical production process arrives in the hands of users in the form of official statistics. Discussions held by the public and within the statistical community[8] in the last years have highlighted ethical issues and the challenges of handling mis- and dis-information,

---

[6] See the "Hungarian Generic Statistical Business Process Model, version 2.3" at https://www.ksh.hu/docs/bemutatkozas/eng/estfm_eng.pdf
[7] See the "Confidentiality policy of the Hungarian Central Statistical Office" at https://www.ksh.hu/docs/bemutatkozas/eng/avpol_web_eng.pdf
[8] See Eurostat's "Rules of engagement for handling disinformation, 2022 edition" at https://ec.europa.eu/eurostat/documents/3859598/14279253/KS-GQ-22-003-EN-N.pdf/d0017a54-e394-3887-d4ce-875aea0606fe?t=1643963795410

including the extent of the roles of NSOs in this matter. This is an area that needs further conceptual development in order to define non-ethical use of official statistics and formulate guidelines for NSOs. The HCSO does not ascertain the ethical use of its statistics by users. However, on occasions when the Office has received ambiguous data requests for research purposes, users have been asked for further clarification on the nature and purpose of the research with the intention of better understanding and fulfilling their needs.

C.   Data Ethics at INSTAT, Albania

36.   "The ethics as an element related not only to methods and procedures chosen as well as behaviours, is an integral part towards retaining trust in official statistics."

37.   Ethics is being considered an important principle to be followed not only in terms of organisational behaviour but as well in the production process and access of data. The Institute of Statistics in Albania follows the European Code of Practice and the Fundamental Principles of Official Statistics, important in ensuring appropriate and reliable statistics where ethics is an integral part.

38.   The **Statistical Council** is an advisory body composed of experts in the field of statistics, which support INSTAT in ensuring that official statistics reflects the user needs for statistical information and promote the implementation of statistical principles and professional standards. The council also discuss with management of statistical structures at line Ministries or other statistical agencies on matters that require their intervention to ensure that official statistic meets the professional standards.

39.   While the Statistical Council is the voice of users the **Committee of Producers of Official Statistics**, created by inter-institutional agreement, brings together the highest representatives of INSTAT and Other National Agencies (Bank of Albania and the Ministry of Finance) in coordinating the work for drafting, implementing and monitoring the Official Statistics Programme. The committee discuss issues related to the expansion of official statistics in the country, but also cost-effectiveness as a result of inter-institutional cooperation.

40.   To safeguard confidentiality the **Confidentiality Committee and the Technical Group on Confidentiality** have been established. The purpose of establishing the Committee and the Technical Group is to guarantee the principle of statistical confidentiality by taking preventive measures to ensure statistical confidentiality during the processes of collecting, processing and publishing official statistics, as well as prohibiting the use of data for non-statistical purposes. A dedicated regulation on Microdata access is publicly available and aims to define organizational and technical measures for security in the administration and access to micro data for scientific research purposes in support of the scientific community.

41.   Moreover, the **Quality Committee** deals with the evaluation of standards established through the quality guidelines, quality management and the proposal for their further improvement. Improvements in the statistical production process they can come through recommendations given during the statistical audit and self-assessment processes. This Committee is responsible for proposing standards for quality and the structure of quality reports for all statistical products.

42.   The **Code of Ethics** built based on national laws, as well as international standards on ethics and **integrity of employee**s, regulates ethical and professional conduct of employees of the Institute of Statistics. It serves as a guide for all INSTAT staff, related to day-to-day or consistent

decision-making actions with the mission, the principles that guide us and our core values. It does not claim to contain all the answers regarding any ethical problems that may arise. The Code is not a substitute for common sense and does not eliminate various regulations implemented by INSTAT. It serves as a bridge between daily needs and actions, and expresses a sense of commitment to the mission of INSTAT. The Code is also a tool to encourage discussions about ethics and improve attitudes or reactions at work (those reactions related to ethical dilemmas and ambiguities). Periodical training of new staff are conducted related to the Code of Ethics.

43.   The **treatment of ethics in all typology used institutional,** data exchange and access, human resources, is an action to be undertaken by the statistical offices. For the first time on 2021, a separate module dedicated to ethical behaviour has been included in the staff satisfaction survey. Based on the results, we have calculated an index to measure the Ethical Environment based on the following dimensions:

- Ethical management, (HR, business ethics and data ethics)
- Ethical practices (e.g transparent behaviours, equal treatment of people and ethical leadership style)
- Data ethics from INSTAT staff point of view
- Perceptions of witnesses of a certain event (illegal or unethical).
- Ethical standards of the organization.

44.   Staff have given their opinion on the importance of ethics through the production cycle of official statistics, challenges and behaviours and what need to be improved in ethical behaviour.

45.   In production process INSTAT uses GSBPM model for standardization since 2017. In 2018 INSTAT prepared a Performance Assessment Framework (PAF), and identified a performance indicator to measure performance of institution. PAF is similar to GAMSO. In the next year INSTAT plans to replace PAF with GAMSO.

46.   It is time to identify the proper instruments to recognize data ethics in all the statistical production chain, communicate it to producers and users of official statistics, move forward and enhance building trust in official statistics.

D.   Data Ethics at the UK Statistics Authority (UKSA)

How the UKSA defines data ethics?

47.   The UKSA's data ethics work focusses on ensuring that the collection, use and dissemination of data across the analytical community is ethically appropriate. Data ethics is defined using 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.

48.   These ethical principles were created in collaboration with the National Statistician's Data Ethics Advisory Committee (NSDEC) who provides the UK's National Statistician with independent and transparent expert ethical advice on the collection and use of data for analytical purposes. The majority of NSDEC members are independent from government and have expertise in areas such as: data law, data ethics, social research, health research and data science. There is also a lay member for a valuable non-expert perspective.

How the UKSA implements data ethics?

49.   The UKSA five-year strategy, Statistics for the Public Good[9], published in 2020, sets out an ambition for the organisation to be a recognised world leader in **applied** data ethics. In February 2021, the UKSA established a Centre for Applied Data Ethics (CADE), with five staff members, to deliver this ambition. To date, CADE has focused on providing resources (e.g. tools, guidance and training) and user support to help researchers efficiently apply good data ethics practices in their collection and use of data. To ensure that these resources are visible across the analytical community, the CADE has presented these services at 32 analytical conferences and engagement events, communicating case studies showing how these services have been successfully used on a range of high-profile research projects. The CADE ensure that these services have impact outside of government by making them accessible to academic and commercial researchers looking to access government data safely, through trusted research environments across the UK.

50.   The CADE has mainly focused their energies on making data ethics accessible to the research community so that they can efficiently adopt an ethics by design approach to collecting and using data for analytical purposes. This has involved moving beyond 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 researchers to quickly assess their project against different aspects of each ethical principle. This is supported by training, guidance and dedicated user support who provide oversight of use of the tool to ensure appropriate use.

51.   Use of the self-assessment tool has been embedded into the UKSA's safe data access processes so that most researchers accessing data for research purposes need to use the tool to get approval to collect and/or access data. Through this self-assessment tool, ethically high-risk projects are quickly identified, and these projects are scrutinised at pace by the independent NSDEC. Details of this independent scrutiny are published on the UKSA website.

52.   In line with the UKSA strategy, a primary focus for this work is to enable efficient use of these services to make timely access to data for analytical purposes possible. The user support, guidance and training provided by the CADE means the researcher is supported throughout the

---

[9] Statistics for the public good – UK Statistics Authority

period they are considering data ethics which helps to ensure that this is a relatively robust and efficient process. Currently, researchers who submit an application have an outcome within two working days.

53.   The UKSA measure the impacts of this work across different stakeholder groups through regular user reviews with users of our data ethics services and collecting key service delivery metrics. The results from these are regularly reported to an independent Advisory Committee made up of stakeholders from across government departments, research funders and the analytical community who provide transparent advice and guidance on the strategic direction, outputs and impact of the Centre. These user reviews and metrics play an important role in helping to inform the development of the services the Centre provides moving forwards.

54.   In the last year, 324 research projects from across the analytical community used the UKSA's ethical principles to efficiently consider the ethics of their collection and/or use of data. Consequently, the UKSA's ethical framework has become one of the most used ethical frameworks in the UK.

55.   To communicate this work to the public the UKSA has published a data ethics policy and a dedicated web presence on the UK Statistics Authority website, which transparently communicates the steps the organisation takes to ensure that data is used ethically, and freely makes their services available to the analytical community. The UKSA also commit to using data ethically in survey participant information across our range of surveys and work with communications teams to make data ethics a central theme of the organisation's public engagement work on the use of data for analytical purposes.

56.   Data ethics is also a key principle in the UK's data sharing framework within the Digital Economy Act 2017. The Research Accreditation Panel is an independent oversight committee which advises the UKSA in its capacity as the statutory accrediting body for this data sharing framework which allows for the sharing of public authority data with the research community for analysis that will benefit UK society and the economy. Under this framework, all parties involved in the disclosure, processing or use of data accessed via the framework are legally required to observe ethical standards appropriate to the nature of the research. As well as ensuring that appropriate ethical standards are consistently observed in all uses of the data sharing framework, the Research Accreditation Panel provides independent advice to the UKSA on how the use of the data sharing framework can be maximised across the research community to increase the public benefit provided by the data sharing powers to UK citizens. The Panel ensures that the UKSA continues to operationalise the data sharing framework in an ethically appropriate and secure way that upholds data privacy and confidentiality.

How the UKSA influences wider thinking on data ethics?

57.   To position the UKSA as a thought leader in the data ethics space, the CADE has completed the following activities:

- Convened a round table to explore questions surrounding how we can best address emerging ethical challenges in the use of data for research and statistics, and how we can support researchers to overcome these issues. This event was attended by approximately 30 participants from a range of organisations from across the UK with an interest in data ethics. The outputs from this have informed our work plan.
- Collaborated with a range of other actors to produce practical user guidance on a range of applied data ethics topics that are listed below. These topics have been selected following

user engagement which suggested a demand amongst analysts for ethics guidance on these topics. Guidance pieces are made available on the UKSA website and are promoted through blogs from senior leaders which are publicised on the Authority's social media channels:

- o Ethical considerations in the use of geospatial data;
- o Considering public good in research and statistics;
- o Considering public views and engagement regarding the use of data for research and statistics;
- o Using data from third parties for research and statistics; and
- o Considering the ethics of working with vulnerable groups.

- Regularly take part in meetings with other government departments working on data ethics to ensure the voice of researchers and statisticians contributes to wider government thinking on data ethics. These departments are also represented on the CADE's independent advisory committee to help ensure the Centre works in a coordinated way with others.
- Provide data ethics training to a range of government departments and professional groups across government.

58. Internationally, the UKSA has played a lead role in the UK's contribution to the data ethics agenda on the international stage. The CADE's work in this area has focused on leading a workstream dedicated to exploring data ethics in the use of Machine Learning to produce official statistics as part of an UNECE programme of work on Machine Learning. As part of this work, new data ethics guidance focusing on the use of machine learning for research and official statistics was published, and the CADE delivered data ethics training at UNECE and UNESCAP events to statisticians working in National Statistical Organisations across Europe and the Asia-Pacific region. The Centre also organised and chaired a successful data ethics session at a UNECE conference on modernising statistical legislation and contributes to international initiatives, such as UNECE Task Team on Ethical Leadership and UNESCAP Asia-Pacific events on Big Data and Governance.

E. Data ethics at Eurostat

59. The mission statement of Eurostat sets out the objective "to provide high quality statistics and data on Europe" and specifies the core values of Eurostat[10]. Eurostat produces European statistics in partnership with the national statistical institutes and other national authorities of the European Union. The development, production and dissemination of European statistics are governed by the statistical principles: *professional independence, impartiality, objectivity, reliability, statistical confidentiality and cost effectiveness*, which are defined in the regulation on European statistics. The statistical principles are further elaborated in the European Statistics Code of Practice (CoP)[11], which is also part of the European Statistics regulation[12]. The CoP defines additional principles related to quality of outputs, statistical processes and institutional environment. The adherence of the European Statistical System (ESS) towards these principles is assessed against a set of indicators for each of the principles by an independent board, the European

---

[10] See https://ec.europa.eu/eurostat/about/who-we-are
[11] See https://ec.europa.eu/eurostat/web/products-catalogues/-/ks-02-18-142
[12] Regulation (EC) No 223/2009 of the European Parliament and of the Council of March 2009, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R0223-20150608&from=EN

Statistical Governance Advisory Board (ESGAB). The ESGAB regularly analyses the implementation of the CoP and advises measures for improvement.

60.   The CoP builds the cornerstone of the European quality framework, which, in addition, is composed of the quality assurance framework and general quality management principles.

Figure 1
The European Statistics Code of Practice
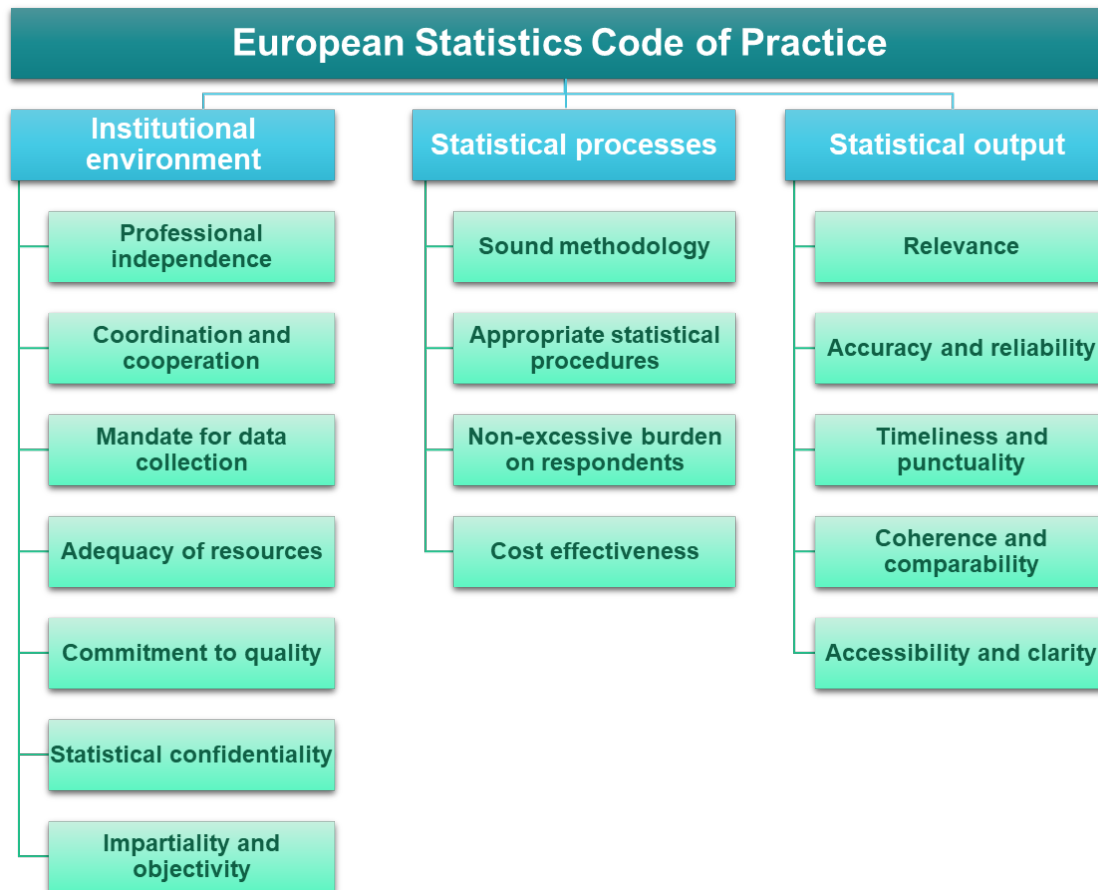
61.   Being a non-legislative self regulatory instrument the CoP and the quality framework for European statistics constitutes an ethical framework for the development, production and dissemination of European statistics. Especially, the principles on *commitment to quality, statistical confidentiality, impartiality and objectivity, sound methodology* as well as the principles related to statistical output are closely related to data ethics as they ensure the ethical use and the quality of data through the statistical production process.

62.   The indicators were originally developed having surveys and censuses as data collection mode in mind. They were extended considering use of administrative sources, which shares some characteristics with the so called big data sources. As part of the activities to include new data sources into the production of European statistics, Eurostat analysed the main issues of integrating them into the production of official statistics and their impact on the principles of the CoP and their indicators[13]. The analysis was done following the main steps of the statistical business production process (data acquisition, data processing, and dissemination of statistical output)

---

[13] See: https://ec.europa.eu/eurostat/cros/content/ethical-review-big-data-0_en

taking into account the characteristics of selected big data sources. The main identified issues are related to the fact that these data sources are not designed for the purpose of producing official statistics, the selection of data and data providers, the access to the data, need for metadata, application of adequate algorithms for data analysis, transparency of processing and analytics, and collaboration with academia and data providers (link to the reports). The analysis was based on desk research as big data sources were hardly used in the production of official statistics at that time. Hence, it will be necessary to review the issues and define indicators to complement the CoP. The work on potential ethical issues resulted in draft guidelines in developing, producing and disseminating statistics based on big data sources[14]. The guidelines distinguish between different types of data sources as potential ethical issues vary very much depending on the type of data source. The draft guidelines related to web scraping served as a starting point for developing guidelines in the context of the web intelligence hub of Eurostat.

63.   The web intelligence hub of the ESS[15], which is currently being developed to provide web content retrieval and web scraping capabilities for the European statistical system is one domain of implementation. The ESS recently agreed on as set of guidelines that govern the retrieval and use of data collected from the world wide web[16]. These guidelines will be implemented by defining a set of applied rules, which are currently being elaborated and discussed within the ESS.

64.   In addition to being the leading organisation of the European Statistical System, Eurostat is a directorate general of the European Commission. The European Commission developed a comprehensive strategy on artificial intelligence between 2018 and 2021[17] resulting in proposing a regulatory framework on AI to foster the uptake of AI while at the same time addressing risks and protecting fundamental rights of EU citizens.

65.   Based on these developments, the European Commission committed to design an internal set of guidelines, a dedicated internal governance and a code of conduct for the development and use of AI systems. The guidelines will pay particular attention to the formation of a "data and AI mature" workforce.

F.   Data ethics statistical professional organisations

66.   In 2020, the Royal Statistical Society in the UK established a Data Ethics and Governance section after it had supported the work of a data ethics interest group for three years. Interest is high and rising for areas such as persons in data, groups, practices, impacts and trustworthy collective structures.

67.   The International Statistical Institute's Advisory Board on Ethics is investing in broadening the coverage of its principles to includes new disciplines related to data science and AI and will thus be involved in data ethics.

68.   On its side, the International Association for Official Statistics (IAOS) created the Krakow group in 2022. The group "aims to contribute to the burning challenges Official Statistics meet in today's datafied societies"[18].

---

[14] See: https://ec.europa.eu/eurostat/cros/content/draft-ethical-guidelines-big-data-and-european-statistics_en
[15] See: https://ec.europa.eu/eurostat/cros/content/project-overview_en
[16] See: https://ec.europa.eu/eurostat/cros/content/web-content-retrieval-guidelines-0_en
[17] See: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
[18] https://www.iaos-isi.org/index.php/latestnews/339-krakow-group

69.   Finally, statistical conferences and journals have seen an marked increase in the number of sessions and articles being produced on data ethics and more generally ethical issues that relate to the context of data in NSOs.

V.   ISSUES AND CHALLENGES

70.   This deep dive has identified several issues and challenges for NSOs which will be discussed in this section. These include:

a) Successfully embedding data ethics policies and principles into data gathering and analytical activities;
b) Getting support for data ethics policies and principles at all levels across the statistical community;
c) Communicating to different stakeholders that data ethics policies and principles have teeth; and
d) Influencing wider thinking on data ethics.

A.   Successfully embedding data ethics policies and principles into data gathering and analytical activities

71.   One challenge that has been identified is around ensuring data ethics policies and principles are successfully implemented across the statistical community. It is not enough for NSOs just to say they have data ethics policies and principles. To make these meaningful, they need to be embedded and implemented throughout the data collection and statistical production processes in a consistent fashion. Only in this way can data ethics policies and principles have a significant impact on all activities across the statistical system.

72.   Key to successfully embedding and implementing data ethics principles and policies across a statistical system is ensuring high-level data ethics policies and principles are accessible to analysts. Analysts need help to apply data ethics principles and policies to their work. This help can take the form of tools enabling analysts to apply high level data ethics policies and principles to their work, user support to support them to do this and data ethics training to communicate the importance of good data ethics amongst the analytical community. It is important that these activities demystify high level data ethics principles and policies for analysts and make them applicable to real life analytical scenarios that analysts are experiencing. This helps move data ethics from the world of high-level ethical theory to useable applied practices that can become successfully embedded in data collection and statistical processes.

73.   Given increasing pressures on NSOs to produce fast paced analysis and timely statistics, ensuring analysts can use these resources efficiently is also key to successfully embedding and implementing data ethics policies and principles. In short, if these processes are not easy to use and are inefficient then they are less likely to be used by analysts. There is a need for specialist data ethics support teams to ensure these processes are easy to use for analysts. Good data ethics within NSOs does not just happen by itself, it needs resources to develop good data ethics practices within a statistical context and support the application of these practices in efficient ways across the statistical community. Further, data ethics tools and practices should be developed and be such that data ethics activities are not seen as a road block by managers who feel they have to get through these to pursue their program.

B. Getting support for data ethics policies and principles at all levels across the statistical community

74. Another challenge that has been identified during this deep dive is the need to get buy in for data ethics policies and principles at all levels across the statistical community. Achieving a common understanding of the importance of data ethics across the statistical system is crucial to enabling any data ethics work within an NSO to have any significant lasting impact.

75. To achieve this buy in, targeted engagement with senior and working level stakeholders is required to champion the importance of data ethics. This engagement should include communicating the significant potential consequences of collecting and using data in unethical ways and explaining why data ethics differs from other established processes in place to safeguard the use of data, such as security policies and data protection legislation. It also involves communicating why data ethics is important now. NSOs currently have access to more data than ever before. This opens up more analytical possibilities to do public good and cause harm than were previously available. With these greater possibilities comes an even greater responsibility to behave in an ethically appropriate manner. This new context that NSOs find themselves in makes it even more important that there is not just consideration of what the NSO **can** do but also robust consideration of what the NSO **should** do, and **how** they should do it from an ethical perspective.

76. It is also important that stakeholders recognise the value that widespread adherence to data ethics principles and policies can provide to ensuring robust data collection and data analysis within an organisation. Adherence to good data ethics policies and principles, such as considering who may be excluded from a sample or considering the participant information that a sample is provided with, often significantly improves the quality of NSO's statistical activities. These activities are not just the right thing to do from an ethical perspective, they are often also the best thing to do from a methodological perspective thus enriching the quality of the statistical outputs produced.

C. Communicating to different stakeholders that data ethics policies and principles have teeth

77. It is sometimes a challenge for NSOs to overcome public scepticism and show that these data ethics activities have teeth. It is therefore important that NSOs are transparent about their data ethics activities. This transparency needs to go beyond communicating data policies, principles and processes and articulate the extent to which these activities are implemented across the statistical community and how they are implemented in a consistent fashion. In this way, NSOs can meet the challenge of proving they do what they say they do and maintain or increase the public's trust in their activities.

78. Stakeholders may question decisions that are made about the ethics of the collection and use of data by people from within the NSO whose statistical portfolio of outputs will benefit from these activities. It is therefore important that NSOs are open to independent scrutiny of their decisions about the ethics of the collection and use of data and that these decisions are informed by appropriate independent expertise. This independent scrutiny should be communicated to stakeholders. Further, the NSOs should make sure to document their data ethics activities and decisions to be in a position to highlight and demonstrate the soundness of their approach. These activities help NSOs to meet the challenge of persuading stakeholders that the ethical decision-making process is credible and independent of any vested interests.

D.   Influencing wider thinking on data ethics.

79.   A large range of academic, government, commercial and third sector organisations are influencing debates on data ethics in countries across the world. Given these different voices, it can sometimes be challenging for NSOs to get their voice head and shape wider thinking on data ethics. It is important that NSOs rise to this challenge as they are prolific and important data users and therefore it is important that the statistical community significantly influences wider work on data ethics both within countries and internationally. Confidently communicating their data ethics work to these other sectors and working with them on data ethics related topics enables NSOs to do this. This reduces the risk that data ethics becomes something that happens to NSOs without them being actively involved in framing thinking on data ethics.

V.   CONCLUSIONS

80.   Data ethics goes beyond traditional topics of security and privacy and encompasses ethical considerations throughout the whole data life cycle from data gathering to data dissemination.

81.   Ethical considerations and principles are not new in official statistics. However, an evolving data landscape and social attitudes towards collection and use of data, bring them centerstage as a necessary component of trust and social acceptability which need to be gained and nurtured. To achieve that, NSOs are starting to approach data ethics in a much more transparent, practical and systemic way. These practices have an important role to play in solidifying their reputation as trusted data gatherers, users and producers.

82.   Despite the growing importance of data ethics, in the context of official statistics, there is no common international definition of data ethics yet and no agreement on recommended practices. The understanding of data ethics and how to apply data ethics within NSOs can differ depending on the context.

83.   The level of development and implementation of data ethics practices is clearly increasing but varies among NSOs. It is clear from the paper that some NSOs have made considerable progress and have shown that one can use data ambitiously and efficiently for statistical purposes whilst ensuring that the collection and use of these data is ethical. The evidence from these countries suggests that robust and timely consideration of data ethics is an important enabler for facilitating innovative data practices that are publicly acceptable.

84.   Data ethics processes are not a blocker to innovative collection and use of data for statistical purposes. Transparent data ethics processes that are well embedded across an NSO help build trusted relationships with the stakeholders (e.g. data subjects, the wider public and data providers) and secure social acceptability. But even before this, they instill a stronger culture of the sense of public benefits related to data needs for society. Robust data ethics also considers the ethics of how research is done and how it is communicated. Therefore, it also helps to ensure methodologically robust data practices and appropriate levels of transparency within an NSO.

85.   For good data ethics practices to be successfully embedded across an organisation, a specialist data ethics support resource needs to be in place to develop these practices and help analysts apply them in their work. It is also important that a culture of good data ethics is embraced by all employees and managers within an organisation to ensure these resources are used and taken seriously.

86.   NSOs are not the only actors in the data ethics space. There is a risk that NSOs are playing only a limited role in shaping wider data ethics thinking. This needs to change if NSOs want to shape the future of data ethics rather than data ethics being something that happens to them.

VI.   RECOMMENDATIONS

87.   There is a need to **better define and communicate what is meant by data ethics** in different contexts and why it is an important enabler for the use of data for statistical purposes. This should help to **develop a common understanding across different international contexts**. To deliver this **a session for senior NSO leaders should be organized** to discuss best practices and produce high level international data ethics standards for NSOs. This session would build on the success of this deep dive and enable the collection of more examples of data ethics practices across the international community and promote more collaboration across the international statistical system. This would make a valuable contribution to informing NSOs data ethics work, drive up data ethics standards internationally and help the statistical community to speak with a more united voice in the wider data ethics space.

88.   To **promote a culture across NSOs where data ethics is taken seriously**, senior management in NSOs should indicate to their employees their firm intention to adhere to data ethics principles and the concept, benefits, values and issues related to data ethics should be widely communicated within NSOs. Data ethics should also be **a topic that is under the responsibility of some evaluation committee**; either a data ethics committee or a more general ethics committee or added to the mandate of some other management/scientific committee. NSOs should involve appropriate independent members in these committees so that decisions about data ethics can be informed by independent experts and civil society who are free from any vested interests.

89.   In order to **ensure that data ethics principles are successfully embedded into NSOs activities during the whole data life-cycle**, from data gathering to data dissemination, NSOs should **identify a person/group responsible to communicate and provide support related to data ethics**. This person/group should have some means to exercise pressure onto program managers to help them abide by ethical principles. It should be the responsibility of this person/group to provide easy to use tools, user support and training to empower analysts to conduct analysis that is consistent with these principles. These tools and activities should be simple and as burden-free as possible for the analysts using them.

90.   **NSOs should be transparent about their data ethics processes and communicate their data ethics work widely** to show stakeholders that they take data ethics seriously. This communication should move beyond just saying what they do to ensure their collection and use of data is ethical to providing evidence that they do what they say they do and presenting the impacts of their data ethics work.

91.   As an important producer and user of data, **NSOs should** confidently **communicate their data ethics work to other actors in the data ethics space** so that NSOs can be leading actors in developing thinking on data ethics and data ethics practices both nationally and internationally.

Bibliography

Floridi and Taddeo 2016, https://doi.org/10.1098/rsta.2016.0360
Hagendorff 2020, https://doi.org/10.1007/s11023-020-09517-8
Hasselbalch 2019, https://doi.org/10.14763/2019.2.1401
Hand 2018, https://doi.org/10.1089/big.2018.0083
Kisselburgh 2020, https://doi.org/10.1007/978-3-030-82786-1_17
Mulligan, Koopman and Doty 2016, https://doi.org/10.1098/rsta.2016.0118
Taylor 2016, https://doi.org/10.1098/rsta.2016.0116

* * * * *