UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
Workshop on Ethics in Modern Statistical Organisations
26-28 March 2024, Geneva, Switzerland
15 March 2024

Structure of ethical issues in new data ecosystems (pre-workshop version)

Marianne Johnson, Timo Koskimäki, Markus Sovala (Statistics Finland)
markus.sovala@stat.fi

Abstract

The current ethical frameworks for statistics, and for social research, rely conceptually on the idea of an organisation – or researcher – collecting data directly from the units they aim to study. In the era of data ecosystems and frequent sharing of data, the assumption of direct data collection has become obsolete.

The statistical community has recognised that the new data-era generates new ethical issues. Trust-enhancing measures like ethical commissions, ethics-related web segments and ethics communication strategies has been set up to overcome the everyday distrust situations.

The paper analyses the new ethical issues using the concept of professional ethics as a tool for analysis. A standard professional ethics code consists of a value statement, a set of claims to subjects and stakeholders, and a set of promises (code of conduct). The most prominent example is the medical profession, but this kind of deconstruction can be applied to any other professions, enabling a systematic study of ethical issues related to professional practices.

We analyse examples of recent ethical debates, related to data and statistics, from the point of view of professional ethics, to better understand the reasons for the emergence of negative public debates and other indications of distrust.

The key result is that current ethical codes for statistics do not recognize the new issues related to use of data. A proposal is made to amend the codes by adding descriptions and guidance related to (at least) the following situations:

- Secondary and tertiary use of data
  Example on secondary use is NSO using privately held or administrative data for statistics, example on tertiary use of data is researchers using micro-data provided by NSO
- Combining data from different sources
  In 2024 UN Statistical Commission meeting very many participating countries indicated that they are going to base their 2030 census on administrative data. We urgently need ethical guidelines on combining administrative data.
- Use of micro-data (or very granular statistics) in knowledge-based decision making.
  Data-based decisions typically impact specific groups of individuals or even individuals. There are already now many borderline cases on ISI principle 12, protection of the subjects. One example is to determine taxable value of a dwelling using micro-data on dwelling prices and statistical modelling; many decisions relating to social benefits impact (smallish) groups of people, and statistics, even micro data, are commonly used to allocate benefits.

UNECE

Structure of ethical issues in new data ecosystems (pre-workshop version)

Marianne Johnson, Timo Koskimäki, Marianne Johnson (Statistics Finland)

Markus.sovala@stat.fi

Paper

Introduction

1. The statistical community has recognised that the new data-era generates new ethical issues. Trust-enhancing measures like ethical commissions, ethics-related web segments and ethics communication strategies has been set up to overcome the everyday distrust situations[1]. However, it has been difficult to determine what was the cause that triggered the social distrust on statistics.

2. In this paper, we analyse the causes using the concept of professional ethical standard as a tool to understand the distrust situation. The study of professions, including professional ethics, is a well-established field of study[2]. As the ethical codes are intended to demonstrate core values and preferred code of conduct to stakeholders, including citizens and the media, they are a natural reflection surface to emerging distrust-situations. As a result of emerging distrust, there has also been a vivid discussion on ethical standards among professional (official) statisticians[3].

3. First, we reflect the basic ideas of the study of professions and relate that to ethical codes and principles related to official statistics. Then we provide two examples of recent public debates on ethics and statistics. We conclude by reflecting our examples with the professional codes of statisticians.

Features of a professional ethics

4. A standard professional ethics structure consists of the following elements:

- a value statement, specifying the professional promise; what is the specific common good that the profession claims to provide to the society and individual people, e.g. health professionals are promoting public health and curing diseases.
- a set of claims to subjects and stakeholders, e.g. permission to violate intimacy, cause pain when diagnosing, professional integrity and education system, resources
- a set of promises (code of conduct), e.g. respect for the patient's right to self-determination, the duty to 'do good', – the duty to 'not do bad', to treat all people equally and equitably.

5. In the formal ethical codes this structure of values, promises and claims is supported by practical guidance highlighting how professionals should behave (e.g. Statistics Finland 1993). The current professional ethics for statistics – the ISI Declaration - also recognises this structure: "[…] the Principles inherently reflect the obligations and responsibilities of – as well as the resulting conflicts faced by – statisticians to forces and pressures outside of their own performance, namely to and from:

• Society

• Employers, Clients, and Funders

• Colleagues

• Subjects

[1] ECE/CES/2023/24

[2] For an overview see Suddaby and Muzio 2015, Bateman 2012

[3] e.g. IAOS 2022; ECE/CES/2022/2

The ISI Declaration on professional ethics then elaborates these dimensions to four shared values - respect, professionalism, truthfulness and integrity - and then continues with a list of 12 ethical principles[4]. One can find similar ethical characterisations from many statistical organisations (ECE/CES/2023/24)

6. For this paper, we will use the structure of the ISI declaration of professional ethics as tool for analysis. The examples we use here come from Finland and Norway. The Finnish case is about the use of statistical micro data for knowledge-based decision making: Should the state know, behind your back, how much you will potentially cost for the health-care service provider. The Norwegian case is about the National Statistical Office access to new, privately held data sources[5]. How much does the National Statistical Office need to know about the individuals in the country.

The Incidents

Micro-data for knowledge-based decision-making

7. In Finland it has taken decades of planning to find a solution to restructure public social- and health care services. One aim has been to take the responsibility to provide these services from the municipalities and instead have new welfare areas, consisting of ten ore more local governments (municipalities). This change was realised in 2023. At the same time, there was also a plan on having private health care providers take a bigger role in health care system. The welfare areas would buy the basic health –care services from private enterprises. The price to be paid to private enterprises would be based on the health-related characteristics of the client. The clients would be free to choose between different service providers, and the service-provider would be compensated by the welfare area depending on the health-related characteristics of the clientele. This latter part of the plan has not realised and is currently not on the political agenda. However, it provides an interesting case of an ambitious plan to apply knowledge-based decision making in the health-care system.

8. Establishing a system of private production of publicly provided services would be implemented by channelling central government funding for services in advance. This would be achieved by the state compensating for the provided services in advance. A model was to be developed for calculating how much funding should be distributed to the different health care actors. The task was given to the Finnish Institution for Health and Welfare (THL), from where an application was sent out to Statistics Finland to get use of individual level data gathered by Statistics Finland from different registers for statistical production.

9. The plan was to use background information from Statistics Finland (such as age, education, language, occupation, place of residence, socioeconomic status etc) and link this data to information on the persons health conditions from healthcare and medication registers from THL and the Social Insurance Institution KELA. With this data on individual level THL would be able to assess for each citizen a risk value, that would give an estimate of the persons upcoming costs as a user of the health care system. The plan was that the calculated risk value would only be known by THL and KELA and that not even individuals themselves would be told the results of the calculations.

10. There were many legal problems with the plan. The EU general data protection regulation states that the use of data should be transparent, and everybody should have the right to check their data. The Statistics Act states that data gathered for statistical purposes can only be given to another statistical agency (which the THL is) for statistical purposes. Statistics Finland can give access to pseudonymized data to be used in scientific research and statistical analyses. The law explicitly rules out that data obtained from Statistics Finland could be used for decisions relating to the individual. The case as such did not involve such decisions, but the sole existence of the coefficient was perceived as a risk that it could be used on individual decision making.

11. There was wide media coverage of the plan that came to be known as the Capitation reimbursement model. First the articles were neutral and brought forth the ideas behind the model with informative interviews with the

[4] ISI 2023
[5] The description of the Norwegian case is made by the authors of the paper and not Statistics Norway

experts at THL. The benefits of the model were put forth and it was implied that the government had much better possibilities to calculate the individual risk values than insurance companies, as the government agencies already have so extensive data on individuals available.

12. Quite fast the articles started to question if giving 'health points' to persons was the right way to go ahead. The Data Protection Ombud took part in the discussions, as well as other legal scholars, and the plan was condemned as among other things citizens' rights had not been taken into account. One of Statistics Finland directors also pointed out the ethical issues on a semi-official blog-platform of Statistics Finland[6]

13. Statistics Finland sent out a letter of inquiry to THL asking for a better description on how the data from Statistics Finland was planned to be used and on what grounds they would have the right to handle the data.

THL came around and submitted an answer to Statistics Finland as well as a new data application that differed in many ways from the previous application. THL still requested much of the same data from Statistics Finland to be linked to health data from THL and medication data from KELA, but now the data was to be used for a research project to come up with a model for distributing funding to each Welfare area. The data would not include identifiers and would be used over Statistics Finland's remote access system by researchers stated in the data permit.

14. As the public discussion was critical towards purely individual coefficients, the legislators ended up to a solution where no individual data would be permitted for the calculation of coefficients. Instead, a rather limited set of variables to be used was defined in the draft legal act. The legal act was never passed, due to resignation of the Government.

Access to new, privately held data sources

15. Norway introduced a new Statistics Act in 2019. The Act granted Statistics Norway rights to access to privately held data for statistical purposes. One of the first attempts to get access to privately held data was to gather data on private consumption using cash-register records combined with payment card data. However, one key actor in the process, the supermarket chains contested the Norwegian statistical office's authority to request regular submission of purchases data collected by these stores. The data was intended to be collected directly from point-of-sale systems and would encompass nearly all grocery purchases made by the entire Norwegian population. Although the purchase data does not include personal identification numbers, Statistics Norway would be able to link more than 70 % of all grocery purchases (receipts) to persons and households through debit card transactions from the banking systems. After this it is possible to link the purchase data to other data on individuals and households already held by Statistics Norway, and thus to e.g. generate information on categorized product purchases by household size, income, education level and geographical region, as well as possibly produce new statistics on dietary habits.

16. One of the supermarket chains filed a complaint with the Norwegian Data Protection Agency (DPA). After investigation, the DPA concluded that the public authority (the State) was intruding upon citizens' privacy by collecting such data. It emphasized that there are limits on what data that official agencies should handle when it comes to personal data, even though the intentions are good. Laws state that everybody has the right for respect of their and their families' private life. The DPA found Statistics Norway lacked sufficient legal grounds to process transactional personal data as proposed and, under Article 58(2)(f) of the GDPR, banned the data processing.

17. Primarily, the issue revolves around concerns about government overreach in data collection rather than mistrust in Statistics Norway's utilization of individual data. Even political parties took part in the debate saying that they do not want Norway to become a surveillance society. Statistics Norway contends that it is wrong to identify the agency as "the state" and asserts its' authority to gather necessary data under the Norwegian

[6] Koskimäki 2018

Statistical Act. They are not interested in individuals but in statistics. Developing new methods is an integrated part of statistics production. There are also high standards set for producing high quality statistics which Statistics Norway aims to achieve by using the best data available. However, there is a recognition within Statistics Norway of the critical importance of maintaining high levels of public trust.

18. The challenge of accessing purchase data is a setback to Statistics Norway's efforts to explore using supermarket receipts for the Household Budget Survey (HBS). Future requests would imply continuing to engage in open dialogue about data usage and conducting necessity assessments for each new planned data collection, as well as continued consultations with the Data Protection Agency.

Reflection

19. Our data – reflection of the case studies with respect to ISI ethical principles and professional values – is presented in Annex 1. According to our judgement, 6 out of the 12 ethical ISI principles were relevant from the point of view of our cases. Considering the ISI professional values, three instances related to value "truthfulness", two to "professionalism" and one to "respect"

The following principles were classified as **truthfulness** issues:

2. Clarifying Obligations and Roles

The respective obligations of employer, client, or funder and statistician regarding their roles and responsibility that might raise ethical issues should be spelled out and fully understood.

8. Maintaining Confidence in Statistics

In order to promote and preserve the confidence of the public, statisticians should ensure that they accurately and correctly describe their results, including the explanatory power of their data.

9. Exposing and Reviewing Methods and Findings

Adequate information, including open-source software, should be provided to the public to permit the methods, procedures, techniques, and findings to be assessed independently.

The following principles were classified as **professionalism issues**:

3. Assessing Alternatives Impartially

Available methods and procedures should be considered, and an impartial assessment provided to the employer, client, or funder of the respective merits and limitations of alternatives, along with the proposed method.

10. Communicating Ethical Principles

In collaborating with colleagues and others in the same or other disciplines, it is necessary and important to ensure that the statisticians' ethical principles are clearly understood by all participants, and properly reflected in the inquiry.

One principle was classified under the value **Respect**:

12. Protecting the Interests of Subjects

Statisticians are obligated to protect subjects, individually and collectively, insofar as possible, against potentially harmful effects of participating.

20. Our first observation is, that all the issues studied relate to the conflict between expectations of the audience (people and media) and the actions of data institutions. In both cases the audience was taken by surprise. In Finland this was probably because throughout the entire process, it was, even for professional audience, unclear whether the research methodologies would be publicly available and whether it would be possible to evaluate the quality of the results. Also, it was unclear who exactly would have access to these sensitive coefficients. Both THL and especially KELA, the two institutions that would have access to coefficients, have also administrative and surveillance functions. They were, in a way, not perceived as research institutes but more as parts of state apparatus. The Norwegian case was more transparent, the approach was quite straightforward to produce official statistics; perhaps the description of the new statistics to be produced was a bit loose. Despite

this, the Statistical Office was perceived as part of – potentially repressive - state apparatus, not as research institute or independent statistics producer.

21. All the statisticians and researchers acted in good faith and carefully followed their ethical codes of practice. Despite this, these data- actions were considered by the audience suspect, intrusive and threatening. Even the cases that we have here classified as ethical issues, are in essence cases where the statistician or researcher has carefully followed the code, but in the eyes of the public, they were not acting ethically. We think the root cause for the distrust we have analyzed here are not the behavior of the institutions or their staff, the issue is that the ethical codes are not fit for the new data situations we face.

22. Statistical institutes and the statistical community have done a lot of work to grasp the new ethical challenges. Trust centra has been established, strategic communication on data confidentiality and data security has been enhanced and explanatory memoranda on the new data ecosystems are being produced. These actions may fail, as they only try to remedy the symptoms, not the root cause which is the fact that statistical institutions have entered to the new areas – privately held data and data services – that are not familiar to our audiences. These new types of tasks are not reflected in our ethical code, not even in the most recent one, ISI code that was revised 2023. Thus, there is no authoritative norm to support data decisions in these new settings. To update the norms would not only benefit the statistical community but all institutions and experts dealing with data and analysis.

23. The ethical codes should (at least) cover the following new situations and provide guidance on how act:
- Secondary and tertiary use of data – example on secondary use is NSO using privately held or administrative data for statistics, example on tertiary use of data is researchers using micro-data provided by NSO
- Combining data from different sources – in 2024 UN Statistical Commission meeting very many participating countries indicated that they are going to base their 2030 census on administrative data. We urgently need ethical guidelines on combining administrative data
- Use of micro-data (or very granular statistics) in knowledge-based decision making. Data-based decisions typically impact specific groups of individuals or even individuals. There are already now many borderline cases on ISI principle 12, protection of the subjects. One example is to determine taxable value of a dwelling using micro-data on dwelling prices and statistical modelling; many decisions relating to social benefits impact (smallish) groups of people, and statistics, even micro data, are commonly used to allocate benefits.

---

24. It should be noted, however, that the Finnish case is an excellent example on evidence-based decision making. As such, it would not have violated the integrity of the subjects. It would probably also fall in to the "do good" category as the application of the coefficients would result in better allocation of resources and thus better services. The statistics Norway case would also fall in the "do good" -category. To do as planned would have saved a lot of taxpayers' money and, at the same time, improved technical quality and relevance of official statistics.

25. We must be prepared for the ethical discussion about the right way to use data gathered for statistical purposes. As the statistical offices are becoming data repositories for a vast array of different governmental registers that can be linked to each other, it is inevitable that society has other use for the data than only the production of statistics. The statistical offices could be a one stop shop to provide much needed micro data for evidence-based decision making. The data needed within government, and the laws concerning the use of data collected for statistics as well as the data protection legislation, do not always meet. There are e.g questions , concerning the need to protect statistical units in cases when the data has not been obtained by the statistical office directly from the statistical unit, but from other sources. Statistics Finland has e.g. turned down an application by the Ministry of Education for school-wise information on the student's parents' socioeconomic status (income, national origin, education). The ministry would have wanted datato be able to give more funding to schools by applying positive discrimination. According to current national legislation, Statistics

Finland cannot disclose the names of our statical units (in this case schools) so the schools needing more funding could not be identified.

Annex: Key elements of ISI code of ethics, Case studies, and Violated values

	Case Finland	**Case Norway**	**Violated values**
1. Pursuing Objectivity Statisticians should pursue objectivity without fear or favour, only selecting and using methods designed to produce the best possible results.	Not an issue	Not an issue	None
2. Clarifying Obligations and Roles The respective obligations of employer, client, or funder and statistician regarding their roles and responsibility that might raise ethical issues should be spelled out and fully understood.	Role of NSI not Clear, role of the research organization not clear	Use of the requested information not specified in detail.	Truthfulness - processes were not transparent.
3. Assessing Alternatives Impartially Available methods and procedures should be considered, and an impartial assessment provided to the employer, client, or funder of the respective merits and limitations of alternatives, along with the proposed method.	No alternatives were considered to the basic idea	No alternatives to micro-linking were considered.	Professionalism – social acceptability was not considered carefully enough.
4. Conflicting Interests Statisticians avoid assignments where they have a financial or personal conflict of interest in the outcome of the work.	Not an issue	Not an issue	None
5. Avoiding Preempted Outcomes Any attempt to establish a predetermined outcome from a proposed statistical inquiry should be rejected, as should contractual conditions contingent upon such a requirement.	Not an issue	Not an issue	None
6. Guarding Privileged Information Privileged information is to be kept confidential. This prohibition is not to be extended to statistical methods and procedures utilized to conduct the inquiry or produce published data	Perceived as risk by subjects?	Perceived as risk by subjects?	None
7. Exhibiting Professional Statisticians shall seek to upgrade their professional knowledge and skills, and shall maintain awareness of technological developments, procedures, and standards which are relevant to their field, and shall encourage others to do the same.	Not an issue	Not an issue	None

8. Maintaining Confidence in Statistics In order to promote and preserve the confidence of the public, statisticians should ensure that they accurately and correctly describe their results, including the explanatory power of their data	The calculated coefficients would not be made available to subjects, details of the models used unclear	The usage of the requested data unclear	Truthfulness – lack of transparency
9. Exposing and Reviewing Methods and Findings Adequate information, including open source software, should be provided to the public to permit the methods, procedures, techniques, and findings to be assessed independently.	Was perceived inadequate by subjects	Was perceived inadequate by subjects	Truthfulness – lack of transparency
10. Communicating Ethical Principles In collaborating with colleagues and others in the same or other disciplines, it is necessary and important to ensure that the statisticians' ethical principles are clearly understood by all participants, and properly reflected in the inquiry.	Failure to communicate to subjects and stakeholders; differing professional codes between statistician and economists	Failure to communicate to subjects and stakeholders	Professionalism – social acceptability not considered carefully enough
11. Bearing Responsibility for the Integrity of the Discipline	Not an issue	Not an issue	None
12. Protecting the Interests of Subjects Statisticians are obligated to protect subjects, individually and collectively, insofar as possible, against potentially harmful effects of participating	Was perceived as threat, unclear whether there is impact to the subjects	Was perceived as intrusive	Respect – the promise to "not do bad "not convincing, the promise to "do good" not convincing from the subjects' point of view.

References

Bateman, Connie (2012):
Professional Ethical Standards: The Journey Toward Effective Codes of Ethics
Work and Quality of Life, 2012, pp 21 – 34

ECE/CES/2022/2 Core values of official statistics (downloaded 11.3.2024)

ECE/CES/2023/24 An ethical approach to the development of social acceptance strategies for national statistical offices by Canada, Ireland, UK and Eurostat (2023). (downloaded 11.3.2024)

IAOS (2022): IPS02.Core values of official statistics – What are they and how do we demonstrate them?

ISI (2023): International Statistical Institute Declaration on Professional Ethics (downloaded 11.3.2024)

Koskimäki, Timo (2018): Sinunkin elämällesi oikea hinta. Blogi Tieto & Trendit 11.4.2018.

United Nations (2014): UN Fundamental Principles of Official Statistics (downloaded 11.3.2024)

Suddaby, Roy and Daniel Muzio (2015): Theoretical Perspective on the Professions. In: The Oxford Handbook of Professional Service Firms. Oxford University Press 2015

Statistics Finland (1993): Toimi oikein tilastoalalla. Tilastokeskuksen ammattieettinen opas. Käsikirjoja 30, Tilastokeskus, Helsinki 1993. (Guide on professional ethics in Official Statistics, in Finnish only)