



# Guideline on Processing Secondary Data for Research

# Contents

1. INTRODUCTION .....	2
2. PURPOSE AND SCOPE .....	2
3. DEFINITIONS .....	3
4. KEY PRINCIPLES .....	6
RESPONSIBILITIES TO DATA SUBJECTS.....	6
5. STUDY DESIGN - PRIVACY BY DESIGN.....	6
5.1 Privacy Impact Assessments .....	7
5.2 Research design and safeguards.....	7
5.3 Additional Guidance.....	8
6. ESTABLISHING GROUNDS FOR PROCESSING PERSONAL DATA .....	8
6.1 Determining data provenance .....	8
6.2 Choosing specific grounds .....	9
6.2.1 Notification and consent.....	9
6.2.2 Legitimate interest.....	9
6.2.3 Compatible purpose .....	10
6.2.4 Contract.....	11
6.2.5 Public task and interest.....	11
7. DATA SECURITY .....	11
7.1 Privacy protection .....	11
7.2 Documentation.....	12
RESPONSIBILITIES TO CLIENTS AND OTHER DATA USERS.....	12
8. TRANSPARENCY .....	12
8.1 Project design .....	12
8.2 Subcontracting .....	12
8.3 Documentation.....	12
8.4 Machine Learning .....	13
RESPONSIBILITIES TO DATA OWNERS.....	14
9. RIGHTS OF DATA OWNERS .....	14
RESPONSIBILITIES TO THE GENERAL PUBLIC .....	14
10. PUBLISHING RESULTS .....	14
11. REFERENCES.....	14
12. PROJECT TEAM.....	14

## 1. Introduction

Throughout its history market, opinion and social research and data analytics (hereafter “research”) has delivered information and insights about people’s behaviour, needs and attitudes to inform decision making by providers of goods and services, governments, individuals, and society at large. In doing so it has mainly relied on primary data collected through direct interaction with and/or observation of participating individuals.

Over the last two decades, we have seen a digital revolution with dramatic increases in the ability to create, collect, store, process, and analyse information through the Internet, social media and mobile technology. This is radically changing the way people live and work. As a result, research is being transformed by an increased reliance on secondary data already available in digital form that is subsequently used in research. The role of the researcher is evolving from interviewer/data collector to data curator, focusing more on organising and integrating data. The research and insight function is extending beyond primary data collection and analysis to managing, synthesising and analysing secondary data available from a diverse range of sources, often evolving the use of new analytic concepts and techniques. The result is an entirely new approach to research whereby researchers assemble and analyse large databases, often containing personal data, to uncover patterns and deliver powerful new insights.

At the same time, there is increasing public concern about the ability of individuals (hereafter “data subjects”) to determine when their personal data is collected, how it is used, and for what purposes. This creates a pressing need for clear ethical and professional guidance on how to handle that data responsibly.

Furthermore, researchers have an ethical responsibility to decision-makers and other data users to be open and fully transparent about the specifics of the data processing and analysis. Such transparency is essential for users of the research to judge its quality and determine whether it is fit for purpose.

## 2. Purpose and Scope

This guideline describes the ethical responsibilities of researchers, regardless of the type of organisation in which they work, when processing secondary data, meaning data that already exists. Secondary data can include publicly available data that can be freely used, reused and redistributed by anyone with no existing local, national or international legal restrictions on access or usage. On the other hand, a lot of secondary data may come from a wide variety of sources including transactional databases created when data subjects interact with companies or government agencies; social media networks; syndicated data; sensors and scanners that comprise the Internet of Things; data aggregations constructed from a variety of sources; and many other similar types of data including from primary research conducted for other purposes.<sup>1</sup> Working with secondary data from these sources will often entail legal and ethical responsibilities.

---

<sup>1</sup> Not included are forms of passive data collection in which a researcher interacts with data subject, for example, to gain their consent to observe and record behavior. See the *ESOMAR/GRBN Guideline for Researchers and Clients Involved in Primary Data Collection*.

Although this guideline is primarily directed at researchers, the audience also includes their clients and other data users and providers, to highlight their responsibilities and to set expectations about what is and is not possible given established ethical and legal requirements.

The requirements and best practices described herein are not meant to reflect the legal requirements of any specific country or region. Rather, they are designed to complement the [\*ICC/ESOMAR International Code on Market, Opinion, and Social Research and Data Analytics\*](#), existing ESOMAR/GRBN guidance documents (including the [\*ESOMAR/GRBN Guideline on Duty of Care: Protecting Research Data Subjects from Harm\*](#), as well as the [\*ESOMAR Data Protection Checklist\*](#), and the codes and guidelines of national associations worldwide. As such, they should not be consulted in isolation.

This ESOMAR/GRBN guidance does not take precedence over national law. Researchers responsible for international projects should take this guideline's provisions as a minimum requirement and fulfil any other responsibilities set down in law or by nationally agreed standards. It is not legal advice and must not be relied upon as such. It remains the responsibility of researchers to keep abreast of any legislation which might affect their research and to ensure that all those involved are aware of and agree to abide by its requirements.

Throughout this document the word "must" is used to identify mandatory requirements. We use the word "must" when describing a principle or practice that researchers are obliged to follow. The word "should" is used when describing implementation. This usage is meant to recognise that researchers may choose to implement a principle or practice in different ways depending on the design of their research.

### 3. Definitions

For the purpose of this document these terms have the following specific meanings:

**Automated decision-making systems** means rules-based systems that make repetitive management decisions without human intervention.

**Client** means any individual or organisation that requests, commissions, or subscribes to all or any part of a research project.

**Consent** means freely given and informed indication of agreement by a person to the collection and processing of their personal data.

**Data analytics** means the process of examining data sets to uncover hidden patterns, unknown correlations, trends, preferences, and other useful information for research purposes.

**Data broker** means any individual or organisation that aggregates data from first party data, second party data, and public data sources and processes it to create profiles of individual data subjects that are offered for sale or lease.

**Data owner** means any individual or organisation that collects and/or curates data, and retains or obtains copyright and license to the data used in secondary analysis.

**Data provenance** means the origin of a piece of data and the tracking of its movement across databases.

**Data subject** means any individual whose personal data is used in research.

**Deductive disclosure** means the inference of a data subject's identity via cross-analysis, small samples or through combination with other data (such as a client's records or secondary data in the public domain).

**First party data** means data collected by an individual, business or other type of organisation from data subjects with whom it has a direct relationship. Examples include customers, members, website visitors and research participants.

**Harm** means tangible and material harm (such as physical injury or financial loss), intangible or moral harm (such as damage to reputation or goodwill), or excessive intrusion into private life (including unsolicited personally targeted marketing messages).

**Passive data** collection means the collection of data by observing, measuring or recording an individual's actions or behaviour.

**Personal data (sometimes referred to as personally identifiable information or PII)** means any information relating to a natural living person that can be used to identify an individual, for example by reference to direct identifiers (such as a name, specific geographic location, telephone number, picture, sound, or video recording) or indirectly by reference to an individual's physical, physiological, mental, economic, cultural or social characteristics.

**Primary data** means data collected by a researcher from or about a data subject for the purpose of research.

**Privacy** means the right of an individual to be free from intrusion or interference and assumes that the individual has the ability to a certain degree to control, edit, manage and delete information about themselves and to decide how and to what extent such information is communicated to others.

**Privacy impact assessment (sometimes referred to as PIA or DPIA)** means a process to identify and mitigate data subjects' privacy risks.

**Privacy notice (sometimes referred to as a privacy policy)** means a published summary of an organisation's privacy practices describing the ways an organisation gathers, uses, discloses and manages a data subject's personal data.

**Profiling** means the collection and processing of personal data with the intent to analyse or predict a data subject's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements in order to take direct action toward data subjects for a non-research purpose.

**Research**, which includes all forms of market, opinion and social research and data analytics means the systematic gathering and interpretation of information about individuals and organisations. It uses the statistical and analytical methods and techniques of the applied social, behavioural and data sciences to generate insights and support decision-making by corporations, governments, non-profit organisations and the general public.

**Researcher** means any individual or organisation carrying out or acting as a consultant on research, including those working in client organisations and any subcontractors used.

**Second party data** means data provided by a first party to other organisations for sale or use. Examples include syndicated data, social media postings, customer data, subscription lists and so on.

**Secondary data** means data that has already been collected and is available from another source.

**Segmentation** means an analytic technique aimed at dividing a broad target population into subsets or groups of individuals or organisations who have, or are perceived to have, common needs, interests, and priorities, and then designing and implementing strategies to interact with them. Segmentation differs from profiling in that its focus is on well-defined groups of people with shared characteristics rather than individual data subjects.

**Sensitive data (Special Category Data in some jurisdictions)** means specific types of personal data that local laws require be protected at the highest possible level from unauthorised access in order to safeguard the privacy or security of an individual or organisation, and which may require additional explicit permission from the data subject before processing. The designation of sensitive data varies by jurisdiction and can include, but is not limited to, a data subject's racial or ethnic origin, health records, biometric and genetic data, sexual orientation or sexual habits, criminal records, political opinions, trade union membership, religious or philosophical beliefs. It can also include other types of data (not necessarily legally defined) such as location, financial information, and illegal behaviors such as the use of regulated drugs.

**Third party data** means data aggregated from a broad range of first and second party data sources. Unlike first party and second party data, third party data organisations typically do not have a direct relationship with data subjects.

**Terms of Use (sometimes called Terms of Service)** means the policy, for example for a website or online service, that requires its users and third parties to accept as a condition of using the service.

## 4. Key Principles

Throughout the long history of market, opinion, and social research and data analytics researchers have recognised that individual data subjects have an inherent right to determine when and how their personal data is collected and used. To this end our work has been governed by three overriding principles:

- When collecting personal data from data subjects for the purpose of research, researchers must be transparent about the information they plan to collect, the purpose for which it will be collected, with whom it might be shared, and in what form.
- Researchers must ensure that personal data collected and used in research is thoroughly protected from unauthorised access and/or use and not disclosed without the consent of data subjects.
- Researchers must always behave ethically, comply with all applicable laws and regulations, and not do anything that might harm data subjects or damage the reputation of market, opinion and social research and data analytics.

These principles<sup>2</sup> form a foundation of trust on the part of the general public, whose data researchers rely on, and the clients who commission research to help them make better decisions. They remain as important today as at any time in our long history.

Secondary data challenges researchers to adapt to a changing environment in which they have less control over the terms of collection and any consent mechanism that may have been required, may not have been sufficiently robust or not present at all. At the same time, they also must ensure that any personal data contained in secondary data and used in research is not disclosed without a legal basis and that the use of personal data does not result in harm.

## Responsibilities to Data Subjects

### 5. Study Design - Privacy by Design

Where the secondary data includes personal data, researchers have ethical responsibilities to those data subjects whose personal data they rely on and fulfilling those obligations to the data subjects begins at the design stage. Some guidance may be provided by the applicable regulatory and data protection requirements. However, these vary considerably from country to country, with some being more restrictive than others and some having no data protection laws at all. While researchers must be aware of and adhere to the laws in those countries where they collect or process data, meeting their ethical responsibilities requires more than simply complying with applicable laws.

---

<sup>2</sup> The Organisation for Economic Cooperation and Development (OECD) espouses a similar set of privacy principles that comprise a privacy framework reflected in many existing and emerging privacy and data protection laws worldwide. See [OECD Privacy Framework](#) for details.

One effective way of doing so is through a practice often described as “[privacy by design](#)”.

The essence of privacy by design is implementation of a process that emphasises an upfront, proactive, end to end project design process in which privacy is taken into account from the beginning and constitutes the default setting. As applied here it has three main components: (a) a foundation of clearly-articulated privacy principles; (b) a process (e.g., a privacy impact assessment) for assessing the privacy risks in a specific project design; and (c) an infrastructure of information security practices and privacy protection approaches, policies and procedures that mitigate those risks.

One of the foundational principles of privacy by design and global privacy frameworks worldwide is data minimisation, loosely defined as the practice of limiting the collection of personal data to that which is directly relevant and necessary to accomplish a specific purpose. Practical considerations of time and money encourage data minimisation in primary data collection. When working with secondary data, the often vast amounts of data available and the computing power to process it, can lead researchers to focus on amassing “all the data”, leaving judgments about what data is relevant to the analysis stage. As a consequence, the need for well-thought out and robust data protection practices increases substantially when working with large amounts of secondary data. A rigorous privacy impact assessment is an essential tool for doing so.

## 5.1 Privacy Impact Assessments

A carefully conducted privacy impact assessment or PIA (also referred to as Data Protection Impact Assessment) ensures that any risks and potential harms that data subjects might be exposed to as a result of their personal data being used for research are identified. This then allows the researcher to adopt a specific study design that includes protections of data subjects’ personal data and privacy so that they do not experience any harm as a result of their personal data being used for research. Simply stated, a PIA is a process to systematically identify and thus enable the mitigation of risks to data subjects’ personal data and privacy over a project’s life cycle. It typically involves these steps:

1. Chart the planned flow of information through the project and all organisations and countries involved.
2. Identify risks and assess their severity, and impact likelihood .
3. Develop and evaluate solutions that mitigate any identified risks.

## 5.2 Research design and safeguards

Based on the PIA, researchers must endeavour to design their research so that the further data processing does not risk the privacy of data subjects either directly or indirectly, for example through deductive disclosure or through combining with other data. Organisations must put safeguards in place to mitigate these risks, using measures like reducing the granularity of the data to lower the probability of a data subject being identified without prior consent. There should also be an agreement with the client to ensure that no non-research activity, taking direct action towards an individual with the intent to



change their attitudes, opinions, or behaviour, will be directed at data subjects as a direct consequence of their data having been used for research.

### 5.3 Additional Guidance

For a more detailed treatment of PIAs consult the ESOMAR/GRBN guideline, [\*Duty of Care: Protecting Research Data Subjects from Harm\*](#). In addition, the [\*ESOMAR Data Protection Checklist\*](#) provides a step-by-step evaluation process to identify gaps and develop solutions in an organisation's information security infrastructure and practices. Researchers should consult it as part of the risk mitigation phase of a PIA.

## 6. Establishing Grounds For Processing Personal Data

Data protection frameworks worldwide increasingly require that individuals and organisations establish clear and compelling grounds before collecting and/or processing personal data. These requirements apply to researchers. Even in jurisdictions where no legal requirements exist, researchers have a responsibility to respect the privacy and rights of data subjects. This requires that they establish some basis for collecting and/or processing any personal data.

Implicit in this requirement is the need to identify the owner of the data to be processed and to secure permission to process the data. Researchers must not access or scrape personal data from websites or other online sources without the consent of the data owner (see section 9, Responsibilities to data owners).

### 6.1 Determining data provenance

Before accessing any secondary data source containing personal data, researchers must first determine the provenance of individual data sets, i.e., the origins of data and its subsequent processing, in as much detail as possible (see section 8). This can be difficult when using a database constructed from multiple sources, where a number of merging, linking up, transforming or aggregating steps may already have been performed. The difficulty will vary depending on whether the data is first, second or third party. For example, when dealing with first and second party data, the data owner is often easy to identify and the circumstances of collection determined. When working with third party data, which generally is multi-sourced, even establishing provenance can be a major undertaking. Data brokers, for example, typically build profiles of individual consumers from dozens of sources, making it difficult to verify what data subjects were told at the time of collection and what limitations may have been placed on its use.

One relatively straightforward method for doing so is to acquire and review for each data source the Terms of Use (ToU), privacy notice, or other similar document provided to data subjects at the time of collection. Researchers must only use secondary data sources containing or constituting personal data that are adequately supported by information that specifies how the data was collected, under what terms, and for what purpose. Above all else, researchers must verify that the personal data was collected legally and transparently, which is essential when determining whether the data can be processed for a research purpose.

## 6.2 Choosing specific grounds

While the requirement to establish ethical or legal grounds for processing personal data is increasingly common worldwide, there are often significant differences across jurisdictions in terms of available grounds, how to qualify, and what specific data collection and/or processing activities are permitted. Therefore, researchers must fully understand the requirements within all jurisdictions applicable to the personal data being processed and ensure that they comply with the relevant law.

### 6.2.1 Notification and consent

When engaged in primary data collection, researchers have generally relied on consent from data subjects before collecting and processing any form of personal data. This includes being transparent about the information they plan to collect, the purpose for which it will be collected, how it will be protected, with whom it might be shared and in what form.

When dealing with secondary data, the rigour of consent practices as expressed in the terms of Use (ToU) varies widely. Some may have many of the same elements as the classic research consent process, while others may miss important elements. In addition, data subjects may have agreed to ToU without careful reading prior to indicating their agreement, making any consent based thereon questionable.

If a researcher intends to rely on consent as the ground for processing there must be sufficient information to determine that:

- The data was legally and transparently collected without deception or in ways not obvious or reasonably discernible and anticipated by the data subject.
- The data subject was required to opt-in to sharing personal data.
- The purpose or purposes to which the data would be put was clearly specified.
- Use of the data for research was not specifically excluded in either the ToU or privacy notice provided at the time of collection.
- Any requests from individual data subjects that their data not be used for purposes other than those described at the time of collection are honoured.

Failing to meet any one of these five conditions requires that researchers consider other grounds.

### 6.2.2 Legitimate interest

In certain jurisdictions, legitimate interest provides an alternative ground to consent that can be used for processing personal data where the data is being used in a manner that data subjects would reasonably expect, and the processing is unlikely to have a significant impact on their privacy. An example would be to ensure quality controls in research such as avoiding duplications or excluding data subjects that do not match the research requirements.

Legitimate interest explicitly considers a balancing of the interests of all stakeholders – the data subjects, the data holder, the client or other end user and even society at large. Individual stakeholders may have differing interests in processing the data to discover new and potentially useful insights, and those interests may conflict. As a consequence,

researchers must balance these competing interests with the greatest weight being given to the interests of data subjects. This translates to a requirement for an especially rigorous PIA and strong privacy and data protection measures to guard against any potential harm to data subjects.

When determining whether legitimate interest can be used as a basis for processing, researchers must ensure that their interests or those of their client are not prioritised over the fundamental rights and freedoms of data subjects. In this determination, researchers must follow and document a three-stage approach addressing these criteria:

- Purpose – is a legitimate interest being pursued?
- Necessity – is the processing necessary to fulfill the purpose?
- Balancing – do the data subjects’ rights and interests override the stakeholders’ interest?

This process of considering and weighing the interests of data subjects must be documented in some way, for example, as a Legitimate Interest Assessment. Legitimate interest must not be used for processing sensitive data or when automated decision making is used (see [ESOMAR/GRBN Guideline on Duty of Care](#)).

### 6.2.3 Compatible purpose

Compatible purpose may also provide an alternative ground. The use of secondary data often involves a change of purpose from what was presented to data subjects at the time of collection. In some jurisdictions, a change in purpose may require the consent of the data subject. This can be challenging and may require that the researcher contact data subjects with the details of the new purpose. In others, it may only be necessary to post a notice on the data collector’s website, providing data subjects with the opportunity to withdraw consent.

There are instances in which a change of purpose may not require consent for the new purpose. One is straightforward: “by the authority of law.” Compatible purpose, where the new purpose is similar, is potentially another.

Establishing compatible purpose requires careful consideration of the relationship between the original purpose at the time of collection and the new purpose, balanced by the reasonable expectations of data subjects about potential future uses of the data and the existence of a compatible purpose in the applicable data protection legislation. It also assumes that appropriate mitigating measures are in place to ensure fair processing and limiting any impacts on the privacy of data subjects.

As an example, online retailers typically collect information about customer purchase behaviour, payment methods, response to promotions, and other personal data needed for product delivery and support. It is reasonable to assume that the retailer will use that data to improve its understanding of what products to offer, at what price, how best to promote them, etc. In this instance, use of the data may be compatible with the original purpose of collection. This may include the delivery of targeted marketing messages to individual customers who have opted in to receive them.

In some jurisdictions, statistical research is considered a compatible purpose. Even in such cases, researchers must observe the privacy protection safeguards described in this guideline.

### **6.2.4 Contract**

Contract might also be used in limited situations as a basis for processing personal data. Researchers might be able to use this basis if they need to process a data subject's personal data in order to fulfil contractual obligations they have towards the data subject, for instance in the administration and management of access panels.

### **6.2.5 Public task and interest**

Processing of personal data that is necessary for the performance of a task in the public interest or for official functions is another basis that researchers may be able to consider. The conditions for using this basis tend to be tightly defined and vary between countries. As a result, it is primarily used for public sector research and/or to a lesser extent private sector research which clearly demonstrates public interest.

## **7. Data Security**

Researchers must ensure that during processing (a) the privacy of data subjects is fully protected and (b) no errors are introduced during processing and analysis. In both cases researchers must have in place a set of procedures and standards designed to accomplish these goals.

### **7.1 Privacy protection**

The [\*ESOMAR Data Protection Checklist\*](#) provides a road map to an infrastructure of technologies, standards and processes designed to prevent the inadvertent disclosure or loss of personal data. Researchers should use it as an initial assessment tool of their privacy protection program to identify gaps and develop solutions.

A key concern is that personal data is not disclosed to clients.

Unless applicable privacy laws and/or regulations stipulate a higher requirement, researchers must only communicate a data subject's personal data to a client under the following conditions:

- the data subject has given explicit consent and
- the purpose is for research only.

Further, it is essential that researchers obtain from clients a written guarantee that the client will not attempt to re-identify data subjects including reverse engineering de-identification protections unless the above conditions are met. For further details consult the [\*ESOMAR/GRBN Guideline on Duty of Care: Protecting Research Data Subjects from Harm\*](#).

Researchers must ensure that any personal data shared with a subcontractor be limited to what is required to perform the subcontracted task(s) and that the subcontractor has

the necessary information security procedures in place to protect the data. The subcontractor's responsibility for data protection must be clearly documented and agreed.

## 7.2 Documentation

Researchers must fully document the specific processing steps performed including any cleaning, merging with other data sources, weighting, imputation (if used) and specific analyses undertaken. The documentation should be specific enough for a data user to understand how the data may have been altered in the course of processing. For further details see section 8.3 below.

# Responsibilities to Clients and Other Data Users

## 8. Transparency

### 8.1 Project design

Researchers must design their research to meet the objectives, specifications and quality proposed and contractually agreed. They must be transparent about the way in which research is to be executed from beginning to end. This information typically is communicated to clients at the proposal stage, and then modified as the work progresses.

The ISO standard, [\*ISO 20252:2019 – Market, opinion and social research, including Insights and Data Analytics – Vocabulary and service requirements\*](#), provides a detailed list of project design features that should be disclosed to clients and other data users at the proposal stage and updated as the research unfolds to ensure full transparency of the specific data used in the research and analyses to be performed.

### 8.2 Subcontracting

Researchers must inform clients, prior to work commencing, when any part of the research work is to be subcontracted outside the researcher's own organisation. On request, clients must be told the identity of any such subcontractor.

### 8.3 Documentation

For research to have value it must be used and acted upon, and it only will be acted upon when a data user accepts the validity of the data, the analyses performed, and the accuracy of its outcomes. As noted in 7.2, researchers must document specific processing and analysis steps performed and specific analyses undertaken.

When working with primary data collections such as surveys and focus groups, there is a broad set of familiar metrics that can be relied upon, coupled with a disciplined approach to measurement that researchers and clients alike mostly understand. The requirements laid out in ISO 20252 ensure a level of transparency that enables clients and other data users to make informed judgements about validity and fitness to purpose in the traditional primary research setting. Those same requirements apply

when traditional analytics are used with secondary data. See paragraph 11.3 of [the ESOMAR/GRBN Guideline on Primary Data Collection](#) for minimum details on what researchers must share with clients at the conclusion of the research noting that the more complex and varied range of secondary data sources will require additional information to provide transparency.

The increasing use of newer algorithmic analytics such as machine learning poses a new set of challenges. While these techniques are often described as “opaque” or “black box,” many of the traditional metrics still apply.

At a minimum, researcher must document the following (with the level of granularity to be agreed with the client):

- the name of the organisation that funded the research, the organisation that conducted it, and any subcontractors used;
- the research objectives;
- the definition of the target population;
- the universe of data sources used and why they were selected;
- documentation to demonstrate provenance and permission to use the data;
- a description of the data items included and their source where possible, or, if imputed, the method used;
- methods of statistical analysis, where applicable;
- where data was combined from multiple sources, a summary of the techniques used and how their accuracy was assessed;
- where appropriate, methods used to edit or clean the data;
- an assessment of the level of omissions in the data; and
- the frequency and process for evaluating the reliability, accuracy, and validity of the analysis.

Researchers also should consider the following:

- identifying, assessing, and addressing the risks associated with any complexities and the nature of the research undertaken;
- including a statement of substantive limitations affecting the validity of findings;
- identifying the needs and expectations of interested parties, and assuring them that their requirements are considered (e.g., clients, communities, regulators);
- providing clarity, transparency, identification, and traceability to enable audit and replication; and
- documenting any methods that are known or suspected to produce bias.

## **8.4 Machine Learning**

There is an additional set of reporting requirements when machine learning is used. The typical goal of a machine learning exercise is to construct a model capable of classifying new incoming data. The generally accepted method for assessing the accuracy of these classifications or predictions is to define a series of well-designed test samples from

which accuracy metrics can be computed and evaluated. These metrics should include, but are not limited to, classification accuracy, logarithmic loss, confusion matrix, area under curve, F1 score, and mean absolute error.

## Responsibilities to Data Owners

### 9. Rights of Data Owners

When researchers work with second and third party data, they must recognise and respect the rights of data owners and any restrictions that they might place on the use of the data.

## Responsibilities to the General Public

### 10. Publishing Results

When the results of a research project will be published either by the researcher or their client, the researcher has a responsibility to ensure that the published results are not misleading. To that end, clients are strongly encouraged to consult with the researcher on the form and content of publication of the findings.

Researchers must also make available on request technical information sufficient to assess the validity of published findings.

Researchers must inform clients that their name must not be associated with the dissemination of conclusions from a market research project unless those conclusions are adequately supported by the data.

### 11. References

[The ICC/ESOMAR International Code on Market, Opinion, and Social Research and Data Analytics](#)

[The ESOMAR/GRBN Guideline Duty of Care: Protecting Research Data Subjects from Harm](#)

[The ESOMAR/GRBN Guideline for Researchers and Clients Involved in Primary Data Collection](#)

[The ESOMAR Data Protection Checklist](#)

[ISO 20252:2019 – Market, opinion and social research, including Insights and Data Analytics – Vocabulary and service requirements](#)

### 12. Project Team

Reg Baker, ESOMAR

Debrah Harding, Market Research Society/GRBN

Anna-Lee Bridgstock, LADbible  
Mario Callegaro, Google  
John Crockett, First West Credit Union  
Philippe Guilbert, SYNTEC Conseil  
Benoit Hubert, IPSOS  
Kathy Joe, ESOMAR  
Peter Milla, Peter Milla Consulting  
George Terhanian, Electric Insights  
Raimund Wildner, Nuremberg Institute for Market Decisions

People who provided feedback:

Pepe Aldudo, ANEIMO  
Catherine Ayland, British Healthcare Business Intelligence Association (BHBI)  
Philip Derham, Derham Insights Research  
Ken Field, IPSOS  
Stefan Klug, DemoSCOPE  
Annie Pettit, E2E Research, Pvt. Ltd.  
Adam Philips, Real Research  
Rejeev Ramachandran, Research, Data & Strategy consultant  
Kamlesh Shukla, Fieldnet Global