ESOMAR

GLOBAL RESEARCH BUSINESS NETWORK
APRC • EFAMRO • ARIA • AMRA

DISCUSSION PAPER:

USE OF SECONDARY DATA IN MARKET, OPINION, AND SOCIAL RESEARCH AND DATA ANALYTICS

Reg Baker, Market Research Institute International and
Norbert Wirth, SUPERCRUNCH by GfK

25 February 2018

ESOMAR is the global voice of the data, research and insights community, speaking on behalf of over 6000 individual professionals and 550 companies who provide or commission data analytics and research in more than 130 countries, all of whom agree to uphold the ICC/ESOMAR International Code**.**

GRBN, the Global Research Business Network, connects 45 research associations and over 3500 research businesses on five contents. www.grbn.org

1 INTRODUCTION

We are awash in data. The volume of data being produced, the velocity with which it is generated, and the variety of forms and structures in which it comes are all unprecedented. By one estimate, 2.5 Exabytes (almost 2.7 billion Gigabytes) of data is produced every day (Khoso 2016). More data was created in the past two years than in the entire previous history of the human race. And, it continues to increase, not incrementally, but exponentially (Marr 2015).

All this data presents an equally unprecedented opportunity for market, opinion and social research and data analytics (hereafter referred to simply as "research"). In theory at least, it is now possible to deliver deeper insights about people's behaviour, needs and attitudes, making possible timelier and better-informed decisions by providers of goods and services, governments, individuals and society at large.

There also are significant challenges. Chief among them is the need to safeguard the privacy of those individuals whose data is collected and processed in research. Regulators across the globe are united in their resolve to establish a new set of legal protections that balance the rights of individuals with the economic and social benefits inherent in data reuse. How this will play out in actual legislation and rule-making is still unclear. In the same vein, the traditional ethical guidelines developed by researchers at a time when most data was collected directly from research participants must be rethought. There also are technical challenges posed by the sheer volume of data now available and the variety of formats and structures in which it comes, orders of magnitude more complex than those of the world of surveys from which so many researchers come.

This paper reviews the key ethical, legal, technical and data quality challenges researchers face when working with these new data sources. Its goal is to start a conversation among researchers aimed at clarifying their responsibilities to those whose data we use in research, the clients we serve and the general public. It uses the term *secondary data* to mean data collected for another purpose and subsequently used in research. It expands on the traditional definition of secondary data to account for new types and sources of data made possible by new technologies and the Internet. It is used here in place of the popular but often vague term, "big data," and is meant to include data from various sources, such as transactions generated when people interact with a business or government agency; postings to social media networks and the Internet of Things (IOT). It is distinct from *primary data*, meaning data collected by a researcher from or about an individual for the purpose of research.

Research is a global business, requiring that we take a broad view rather than tying guidance to the specific legal requirements of a single region or country. This is not to say that researchers no longer have a primary responsibility to comply with local regulations in the markets where they collect, store and process data. Rather, we have chosen to simplify the discussion by focusing on the key principles that form the basis for privacy and data protection regulations worldwide. Three frameworks stand out: the Asia-Pacific Economic Cooperation (APEC) Privacy Framework (2005), the Organisation for Economic Co-operation and Development (OECD) Privacy Principles (2013), and the US Federal Trade Commission's (FTC) Fair Information Practice Principles (2007). To one degree or another, these three frameworks share four broad principles relevant to our discussion: notice and consent; use limitation; information security safeguards; and data integrity.

In all of this, researchers should remain mindful of the self-regulatory character of research and apply ethical standards that may exceed local legal requirements. Researchers have their own version of the Hippocratic Oath mandating that data subjects not be harmed as a direct consequence of their data having been collected and/or processed for research. Harm, in this context, has a broad meaning that includes "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" (ESOMAR 2016). Market, opinion, and social research, and data analytics relies on the public's confidence (and that of their elected representatives) in the integrity of the research process, and maintaining that confidence requires that data subjects not suffer adverse consequence, regardless of applicable legal requirements.

2 NOTICE AND CONSENT

Throughout its history, research has relied mostly on data collected directly from research participants. These collections required (and continue to be required in the codes of conduct of research industry and professional associations) the consent of the participant, defined in the ICC/ESOMAR Code (2016) as "the freely given and informed indication of agreement by a person to the collection and processing of his/her personal data."[1] Article 4 of the Code specifically requires that:

(a) When collecting personal data directly from a data subject for the purpose of research:

i. Researchers must identify themselves promptly and data subjects must be able to verify the identity and bona fides of the researcher without difficulty.

ii. Researchers must clearly state the general purpose of the research as soon as methodologically possible.

iii. Researchers must ensure that participation is voluntary and based on information about the general purpose and nature of the research that is adequate and not misleading.

iv. Researchers must inform data subjects if there is any activity that will involve re-contact and data subjects must agree to be re-contacted. The only exception to this is re-contact for quality control purposes.

v. Researchers must respect the right of data subjects to refuse requests to participate in research.

Privacy protection frameworks typically specify similar requirements for any data collection, regardless of the collector or purpose. They are expressed in various ways such as individual control, individual participation, use limitation, choice, notice, etc.

With the proliferation of data collections all kinds and attendant public concerns about privacy, issues of notice and consent have taken on new meaning. As researchers rely less on data collection and more on reuse of data collected by someone else for a purpose other than research we require a new mechanism to ensure that we are consistent with our industry's traditional ethical standards, safeguard the privacy of those individuals whose data we use in research, and conform to local law. Fortunately, we are not alone.

In 2012, the OECD solicited input from experts worldwide as it prepared a revision of its privacy principles, first developed in 1980. In one response, Cate and Mayer-Schönberger (2013) argued that the traditional reliance on notice and consent is now problematic. They pointed to the obvious reality that many people simply do not read privacy notices as they have become more prevalent, complex, and difficult to understand. Instead, they simply click "Agree." They also noted the tendency for notices either to specify a very narrow purpose, limiting the potential to use for other purposes unknown at the time of collection, or have become increasingly broad and permissive. They concluded that:

> … the notice and consent system, on which data collectors and data users have come to rely, was designed to empower individuals to make decisions about their

[1] Defined as "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.

> personal data, but the evolution of data collection and data use has severely weakened that power while imposing increasing burdens on data subjects and on society. While notice and consent may provide meaningful privacy protection in appropriate contexts, this approach is increasingly ineffective as the primary mechanism for ensuring information privacy. In addition, the advent of big data and new analytical tools has shown us that many valuable and innovative uses of data are not known at the time of collection. Data collected for one purpose can often be repurposed in ways that greatly benefit society.

Having come to this conclusion, they recommended a new approach, one that shifts the responsibility for privacy protection away from individuals (and a robust notice and consent process) to data collectors and data users "who should be held accountable for how they manage data rather than whether they obtain individual consent."

Around that same time the World Economic Forum convened its own symposium as part of their ongoing project, Rethinking Personal Data. They came to much the same conclusion as Cate and Mayer-Schönberger, acknowledging that while the basic data protection principles are still valid, they do not work in today's world. "In particular, notice and consent was highlighted as not delivering real effective choice to individuals to ensure permissioned, trusted flow of data." (World Economic Forum 2012)

As regulatory bodies develop new approaches for protecting privacy in this changed environment they seem to be of two minds. For example, the US Federal Trade Commission (FTC) recently stipulated that companies do not need to get consent "before collecting and using consumer data for practices that are consistent with the context of the transaction or the company's relationship with the consumer or are required or specifically authorized by law." (Federal Trade Commision 2012) Consent is still required when using the data for a purpose that is different from what was claimed at the time of collection or when the data being collected is considered sensitive.

The EU's General Data Protection Regulation (GDPR) set to go into effect in the spring of 2018 takes a more stringent view. It reaffirms and strengthens the consent requirement by requiring a "freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."[2]

The codes of conduct of research associations worldwide continue to insist on a robust notice and consent process when researchers collect data directly from research participants. However, there also is a need to provide more detailed guidance for the use of data collected by someone else for some purpose other than research and under conditions that may be much different from those used in primary research.

3 USE LIMITATION

With the exception of research syndicated data, use of secondary data typically involves a change of purpose from what was presented at the time of collection. In general, a change in purpose requires the consent of those whose data is was collected - typically referred to as *data subjects* in most legislation. This can be an onerous task, although in some cases it may be sufficient to notify data subjects rather than contact them to gain consent for the new purpose. This can be as simple as posting a notice on the data controller's website, providing data subjects with the opportunity to withdraw consent.

The OECD and APEC describe two exceptions. One is straightforward: "by the authority of law." (APEC, p. 17; OECD, p. 14) and such exceptions are sometimes specified in regulations. The second is more complex and requires that a judgment be made as to whether the new purpose is *compatible* with the original purpose. This judgement requires

[2] Article 4 (11).

that we strike a balance among all stakeholders, with particular focus on the legitimate interests of the data controller/collector on the one hand and the data subjects on the other. The legitimate interest of the data controller may be most easily understood by considering a use case.

An online retailer routinely collects data on its customers - what they buy, what promotions they take advantage of, how often and what they return, where they live, etc. - for the overt purpose of sales and fulfilment. While not explicitly disclosed to customers, the company may also choose to analyse that data to improve its understanding of its customer base, their product preferences, susceptibility to promotions, etc. While that may represent a change in purpose, it passes the legitimate interest test for the data controller and therefore is a compatible use.

Other use cases may not be as clear cut, and consideration of the legitimate interests of the data subjects may tip the scales one way or the other. The key criterion is protecting data subjects from harm as result of their data being used for a different purpose. As noted at the outset, harm, in this context, has a very broad meaning that includes financial loss, invasion of privacy (including targeted marketing messages), physical injury, denial of opportunity and damage to reputation. One useful tool for understanding the potential harm to data subjects is a *privacy impact assessment*, discussed below.

This issue is now front and center with regulators due to the growth of *automated decision-making systems*, that is, rules-based systems that use profiles of individual data subjects to make management decisions. Those decisions can cover a wide range of activities with varying levels of impact on individual data subjects, some of which may be clearly discriminatory. Examples include: granting of credit; employment offers; criminal prosecution; treatment of illnesses; product pricing; and tax audits. (See, for example, Future of Privacy Forum 2017). As a result, research has been drawn into a broader discussion about the collection, use, and processing of personal data, especially when using online methods to collect behavioural data.

3.1 Privacy Impact Assessments

A privacy impact assessment (PIA)[3] provides a road map for systematically identifying and mitigating the risk to data subjects' privacy when planning primary data collections or use of secondary data. Such risks can arise from many sources including but not limited to:

- collection of excessive or irrelevant information (including sensitive information);
- overly long data retention practices;
- use of data in ways unanticipated by data subjects;
- sharing of data in a form that makes re-identification possible;
- merging/fusion of separate data sets;
- disclosure to third parties without consent; and
- ineffective information security practices.

The design of a PIA will vary depending on the organisation's business, its internal processes, and those of any subcontractors used. It typically involves these four generic steps:

1. Chart the flows of information through the organisation, project or data life cycle.
2. Identify the potential privacy risks and harms.

[3] For a more thorough discussion of PIAs see the UK's Information Commissioner's Office (ICO) publication, Conducting Privacy Impact Assessments: Code of Practice.

3. Develop and evaluate privacy solutions.
4. Integrate the most effective solutions into organisational processes or project plans.

PIAs typically are conducted at the individual project level or some other specific planned use of data. Ideally, they are refreshed throughout a project life cycle as data inputs evolve and new or unanticipated risks emerge. Organisations generally retain a record of each assessment for future reference.

This type of evaluation process meets the requirement of Article 5 of the ICC/ESOMAR Code (2016), describing the broad outline of the responsibilities researchers face when using or at least planning to use secondary data. They must ensure that:

- The intended use is compatible with the purpose for which the data was originally collected.
- The data was not collected in violation of restrictions imposed by law, through deception, or in ways that were not apparent to or reasonably discernible and anticipated by the data subject.
- The intended use was not specifically excluded in the privacy notice provided at the time of original collection.
- Any requests from individual data subjects that their data not be used for other purposes are honoured.
- Use of the data will not result in harm to data subjects and there are measures in place to guard against such harm.

In theory, at least, researchers have an advantage over those working in other commercial enterprises because they have no interest in the identity of data subjects in order to take direct action toward them or to change their opinions, attitudes, or behaviours. The data is used only in statistical and scientific analyses to uncover patterns and provide new insights. The emphasis in research is on unlocking the potential social and economic value made possible by additional use of existing data and not, for example, to construct profiles of individual data subjects for targeting purposes. In practice, researchers are under increasing pressure to deliver profiles of individual data subjects for targeting. This is not a legitimate research output and projects designed for this purpose are not research and must not be claimed as such.

Researchers have a powerful argument for using secondary data in their research in the absence of specific prohibitions against such use at the time of collection. However, such a right generally is not clearly stated in existing law. For example, the GDPR recognises the need to balance the economic and social value in facilitating the reuse of data for analytic and statistic purposes. It allows for further processing of personal data as long as the purpose is compatible with the original purpose at the time of collection. It further states that “for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes.” (Article 5(1)(b)) Nonetheless, it is still unclear whether even a research purpose will require some level of notice to data subjects when the purpose is changed.

4 INFORMATION SECURITY SAFEGUARDS

There is no such ambiguity when it comes to protecting the personal data of those whose data is collected and/or used in research. Emerging legal requirements are not unique to secondary data. Rather, they reflect widespread public concerns about threats to privacy and the potential harm inherent in the irresponsible use of personal data, regardless of its origins.

Research organisations are finding it increasingly necessary to implement information security management systems that strengthen the administrative, physical and technical

controls needed to manage and protect the confidential and sensitive data they hold, including personal data. These controls include but are not limited to:

- improved privacy policies describing the ways an organisation gathers, uses, discloses and manages personal data;
- enhanced staff training that includes an overview of applicable laws, industry codes of conduct, information security practices and the internal procedures that ensure successful implementation of the organisation's privacy policy;
- implementation of physical and technical barriers to access including alarm systems, security cameras, passwords, firewalls and encryption;
- requirements for timely notification of relevant authorities and affected data subjects in the event of a data breach;
- shortening of data retention periods;
- broader use of de-identification methods (e.g., pseudonymisation, aggregation, masking, encryption, etc.);
- improved data transfer protocols that secure information during transfer and require a level of security protections by the recipient that are comparable to those of the sender; and
- clearly defined rules for sharing data that ensure no personal data is shared without the consent of individual data subjects.

One useful resource for research organisations when developing and implementing their information security practices ISO 2700:2013—Information technology security techniques. While not mandating specific actions, the standard includes suggestions for documentation, internal audits, continual improvement and corrective and preventive action.

5 DATA QUALITY AND TRANSPARENCY

Data quality in the context of privacy frameworks typically means ensuring that data collected on individual data subjects is accurate. Researchers' concerns about data quality are different. We tend to evaluate data in the aggregate rather than at the individual level and with a focus on the traditional concerns of representativity and measurement.

In an industry that has evolved over many decades, researchers have accumulated vast experience in producing, analysing, and interpreting primary data. The entire value chain from translating an empirical research question into a research design, developing questionnaires, collecting data, processing and analysing data to interpreting the findings is well established. Recommended practices are incorporated into guidelines developed by research professional and industry associations and even expressed in international standards such as ISO 20252:2012. The properties of the gathered primary data are well known and there is a rich toolbox of robust and proven approaches to developing insights from it.

With the rapidly emerging use of secondary data collected by other means, often in the form of multiple data sources being merged or integrated, we face an entirely new set of challenges. Secondary data is by definition not generated for a specific research purpose, and therefore needs to be approached with a new set of precautions. Concepts of data quality and transparency have new relevance.

Article 7 of the ICC/ESOMAR Code (ESOMAR 2016) states, in part, that:

> (b) Researchers must ensure that findings and any interpretation of them are clearly and adequately supported by data.

(c) Researchers must on request allow clients to arrange for independent checks on the quality of data collection and data preparation.

(d) Researchers must provide clients with sufficient technical information about the research to enable them to assess the validity of the results and any conclusions drawn.

When research is based on a blend of primary and secondary data, or exclusively on secondary data, researchers must rethink how data integrity and transparency can be ensured. The data generating mechanism of all data sources used must be properly understood, otherwise it is simply impossible to demonstrate how the "findings and interpretation of them are adequately supported by data." This will often require substantial technical skills and expertise as well as domain knowledge.

Arguably the biggest risk when analysing secondary data is the misperception that volume - i.e., the amount of data used to address a specific research question - automatically produces quality results.

> Nearly all of the discussion about the future of marketing research either sweeps data integrity under the rug or ignores it in the mistaken belief that more data or observational data or sensor data or whatever automatically means quality data. With data more important than ever, the importance of data quality has never been greater. (Kantar Futures 2017)

The adverse effect of misjudging volume for quality is significantly amplified when multiple data sources are used. Data can be heavily biased and this problem does not go away with large datasets.

Let us illustrate this challenge with a hypothetical example. A client is interested in an improved understanding of the customer shopping journey. Geolocation data is available from a single mobile service provider for customers who have explicitly opted in, allowing for their geolocation data to be used for a market research program. This might result in millions or even billions of data records over time. Yet despite its impressive volume, this data still only represents a subpopulation, not the entire population. Generalisations based on this data may be misleading.

Secondary data can only contribute to the answer of a research question if it contains relevant signals or, in more general terms: information. The choice of data sources is determined by the use case and the research question. Researchers design primary data collections specifically to provide information about the needs, wants and behaviours of a clearly defined target population. Secondary data generally is not designed around specific research questions and therefore requires a thorough exploratory analysis and assessment in order to understand if it is useful or not. The signal to noise ratio of secondary data sources is highly variable and volume does not necessarily change that.

In short, each secondary data source used needs to be fit for purpose in two ways: (1) it must be a reasonable representation of the population of interest and (2) it must accurately measure the underlying constructs deemed relevant to the research.

5.1 Assessing Data Quality

If we are to have confidence in the integrity of secondary data, we need a clear view of the quality of the original data sources - the collection processes, the metadata, the population covered, etc. The rules under which the data was collected, cleaned and processed, while adequate for the original purpose, may be inadequate for an analytic purpose or may not be complete enough to develop the rules needed to merge with other data. Transaction data may have missing data elements or not fully cover the population of interest. Unstructured data such as text, images and video posted on social media networks may have been converted to a standard structure to support analysis, opening the door to a wide range of

potential errors and lost or mischaracterised meaning. Internet of Things data may be corrupted or have gaps due to sensor malfunction, outage or gaps in coverage.

Assessing the quality of a secondary data requires an understanding of three key concepts. The first is *data governance*, meaning the rules by which the organisation that collected the data manages the integrity, usability, security and availability of the data it collects. Ideally, these rules should be at least equivalent to the standard processes researchers have developed over decades for collecting and processing primary data. Too often, this is not the case. Concerns about the effectiveness of data governance across industries are widespread, leading many to openly worry about the quality of secondary data that otherwise might be of use to researchers. See, for example, Redman (2016).

The second is *data curation*, meaning “the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science and education."(Cragin et al. 2007). Data curation is much like data archiving in that it documents data and preserves it over time, making reuse and analysis possible. It is especially challenging when dealing with secondary data because by definition the data was not collected for a research purpose and therefore may not be fully documented or preserved in a format that facilitates reuse.

Finally, there is *data provenance*, meaning “the process of tracing and recording the origins of data and its movement between databases.“ (Buneman et al. 2000) Data provenance essentially involves the reconstruction of the history of each data element in as much detail as possible. It looms especially large when using a dataset constructed from multiple sources, where a number of merging, linking, transforming or aggregating steps already have been done. For further discussion of data quality issues with secondary data see Baker (2016.)

The technical and quality challenges posed by secondary data, particularly in large and dynamic multi-source data environments, call for a well-defined methodology of quality assessment. This does not mean that all data has to be of highest quality, but it is nonetheless vital that the quality be assessed and documented.

5.2 Transparency

Meeting the transparency requirements of the ICC/ESOMAR Code starts with a full and formal evaluation of data as just described. It is a key step toward increasing a client’s understanding and confidence in research results. But, it is not the only step.

Earlier in this paper we described the evolution over time of a set of standards for primary data collection designed to promote high quality data. These are generally understood by both researchers and clients alike. A similar set of standard approaches has evolved for analysis of and reporting on primary data, some of which are now being challenged by the availability of very large secondary datasets and the technology to manage and process them. Clients may not be fluent in neural networks, sentiment analysis, genetic algorithms, logistic regression, data visualisation and other techniques that form the basic toolkit of the data scientist. Transparency in this environment takes on new meaning. Providing clients with “sufficient technical information about the research to enable them to assess the validity of the results and any conclusions drawn” is more difficult. The temptation to obscure rather than enlighten is real and to be avoided.

6 SUMMARY AND CONCLUSIONS

This paper has attempted to identify the challenges researchers face as they broaden the range of data sources they use to support timelier and better-informed decisions by providers of goods and services, governments, individuals, and society at large. While the ethical foundations of the research profession remain unchanged, the legal environment is in flux as regulatory bodies grapple with the twin challenges of a dramatic increase in data (much of which is personal data) and related public concerns about privacy.

We have reviewed several principles common to the global privacy frameworks that form the basis for data protection regulations in the major markets. Two things stand out as especially challenging for researchers.

The first is the interplay between notice and consent on the one hand and purpose limitation on the other. Researchers are accustomed to requirements of notice and consent when designing primary data collections. They agree that individuals have the right to control how their data is used. A key question is how best to operationalize that belief in the context of secondary data use in ways that reflect the ethical and legal requirements of a self-regulated industry. How do we ensure that the original data collection was legal and legitimate? How do we balance the legitimate interest of all stakeholders? How do we ensure that the risk of harm to data subjects is thoroughly vetted and the right protections put in place? Where there is a clearly-stated legal exception for research, how do we temper that with our longstanding ethical principles? Is there or should there be a higher bar?

The second issue is largely a methodological one, but nonetheless central to ethical research practices. It focuses on transparency about data quality and research results in our interactions with clients. For the better part of a century researchers have been developing and fine-tuning practices for the collection, analysis and delivery of results that are widely understood and generally accepted by researchers and data users alike. Secondary data changes all that. How do we establish a new set of practices that validate the quality of the data we use? How do we describe our analyses in ways that are understandable and open to validation? How do we create a sense of comfort with clients that encourages them to act on the insights we deliver?

Finally, while we have not described in any detail the changing responsibilities researchers face in implementing more robust information security practices. These clearly are front and centre. They are not directly related to the increasing use of secondary data, but the need to elevate our game in this regard is crystal clear. A positive view of market, opinion, and social research and data analytics on the part of both the public and regulators is essential to our long-term success. Regardless of how we acquire data we need to be more resolute in our protection of it than at any point in our long history.

7 REFERENCES AND SUGGESTIONS FOR FURTHER READING

Asia-Pacific Economic Cooperation (2005). APEC Privacy Framework. Retrieved on July 19, 2017 from https://www.apec.org/-/media/APEC/Publications/2005/12/APEC-Privacy-Framework/05_ecsg_privacyframewk.pdf

Baker, Reg (2017). Big Data: A Survey Research Perspective. In Paul Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West (eds) *Total Survey Error in Practice*. New York: John Wiley and Sons, Inc.

Buneman, P., Khanna, S., and Tan, W.C. (2000). Provenance: Some Basic Issues. Lecture Notes in Computer Science, Vol. 1974, *Foundations of Software Technology and Theoretical Computer Science*, (FST TCS 2000), 87-93.

Cate, Fred and Mayer-Schönberger, Viktor (2013). Notice and Consent in a World of Big Data. *International Data Privacy Law* 3 (2): 67-73.

Cragin, Melissa H., Heidorn, P. Bryan, Palmer, Carole L., and Smith, Linda C. (2007) An Educational Program in Data Curation. ALA Science & Technology Section Conference, Retrieved on July 21, 2015 from https://www.ideals.illinois.edu/handle/2142/3493

ESOMAR (2016). The ICC/ESOMAR International Code on Market, Opinion, and Social Research and Data Analytics. Retrieved on July 17, 2017 from https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_English_.pdf

Federal Trade Commission (2012). FTC (2012). Protecting Consumer Privacy in an Era of Rapid Change. Retrieved on July 1, 2017 from https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

___(2007). Fair Information Practices Principles. Retrieved on June 30, 2017 from https://web.archive.org/web/20090331134113/http:/www.ftc.gov/reports/privacy3/fairinfo.shtm

Future of Privacy Forum (2017), "Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making." Retrieved on February 21, 2018 from https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/

Information Commissioner's Office (2014). Conducting Privacy Impact Assessments: Code of Practice. Retrieved on July 18, 2017 from https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf

International Organization for Standardization (2013). ISO 27001—Information technology security techniques. Geneva: International Standards Organisation.

___(2012). ISO 20252—Market, opinion and social research -- Vocabulary and service requirements. Geneva: International Organisation for Standardization.

Kantar Futures (2017).The New Roles of Marketers and Researchers. Retrieved on July 26, 2017 from http://thefuturescompany.com/the-new-roles-of-marketers-and-researchers/.

Khoso, Mikal (2016). How much data is produced every day? Retrieved on July 4, 2017 from http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/

Marr, Bernard (2015). Big Data: 20 Mind-Boggling Facts Everyone Must Read. Retrieved on July 1, 2017 from https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/ - 70fcb15717b1

Organisation for Economic Co-operation and Development (2013). The OECD Privacy Framework. Retrieved on July 9, 2017 from http://www.oecd.org/internet/ieconomy/oecd_privacy_framework.pdf.

Redman, Thomas C. (2016). Bad Data Costs the U.S. $3 Trillion Per Year. *Harvard Business Review*. Retrieved on July 24, 2017 from https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year

TechTarget (2012). Data Curation. Retrieved on July 22, 2017 from http://whatis.techtarget.com/definition/data-curation

Tiao, Shirley (2018). GDPR and Big Data: 4 Steps to Compliance. Retrieved on February 15, 2018 from https://blogs.oracle.com/bigdata/gdpr-big-data-steps

Woodie, Alex (2017). GDPR: Say Goodbye to Big Data's Wild West. Retrieved on December 13, 2017 from https://www.datanami.com/2017/07/17/gdpr-say-goodbye-big-datas-wild-west/

World Economic Forum (2012). Unlocking the Economic Value of Personal Data Balancing Growth and Protection. Retrieved on July 5, 2017 from http://www3.weforum.org/docs/WEF_IT_UnlockingValueData_BalancingGrowthProtection_SessionSummary.pdf