4 Informed Consent in Digital Data Management

Elisabeth Hildt and Kelly Laas

Abstract This article discusses the role of informed consent, a well-known concept and standard established in the field of medicine, in ethics codes relating to digital data management. It analyzes the significance allotted to informed consent and informed consent-related principles in ethics codes, policies, and guidelines by presenting the results of a study focused on 31 ethics codes, policies, and guidelines held as part of the Ethics Codes Collection. The analysis reveals that up to now, there is a limited number of codes of ethics, policies, and guidelines on digital data management. Informed consent often is a central component in these codes and guidelines. While there undoubtedly are significant similarities between informed consent in medicine and digital data management, in ethics codes and guidelines, informed consent-related standards in some fields such as marketing are weaker and less strict. The article concludes that informed consent is an essential standard in digital data management that can help effectively shape future practices in the field. However, a more detailed reflection on the specific content and role of informed consent and informed consent-related standards in the various areas of digital data management is needed to avoid the weakening and dilution of standards in contexts where there are no clear legal regulations.

Keywords: informed consent, code of ethics, guidelines, personally identifiable information, big data, surveillance, privacy

4.1 Introduction

Digital data, discrete information signals produced by machine language systems that represent other kinds of data, can be copied indefinitely and spread easily. Digital technologies allow many ways to create, store, and replicate data, extract information from data sets, and transform it for future use. Digital data allows new ways for individuals and organizations to interact with one another (National Academy of Sciences 2009; Clark et al. 2015).

Digital data arises from a variety of contexts and is becoming an increasingly valuable commodity to be collected, stored, shared, and sold. In many instances, users of apps and social media provide personal information to companies and receive services in return (van Dijik 2014). In research, researchers are encouraged to collect and deposit data in digital archives for secondary use or are obligated by funding agencies to allow for greater transparency in research (for example, National Institutes of Health 2004; National Science Foundation 2017). In business, the collection of data through sensor technologies, and the widespread use of the internet in daily life, have increased the amount of information available to companies and the different ways this information can be collected and used (Institute for Business Ethics 2016).

Elisabeth Hildt
Illinois Institute of Technology, Chicago, Illinois, USA,
email:ehildt@iit.edu

Kelly Laas
Illinois Institute of Technology, Chicago Illinois, USA
email: laas@iit.edu

In this contribution, we are primarily interested in digital data that provide information about individuals. Examples include medical and health-related data, data resulting from online research, data generated through social media, smartphones or fitness devices, geospatial data, or data created from online purchases.

Of particular concern in this context is personally identifiable information, i.e., information that either alone or in conjunction with other information can be used to identify, trace, or contact an individual person. Examples include name; personal identification number such as passport number, social security number, financial account number or credit card number; address information (street address, email address); asset information such as Internet Protocol (IP) or Media Access Control (MAC) address or other host-specific identifiers; personal characteristics including biometric data; information about an individual that is linked or can be linked to one of the above, such as date of birth, activities, geographical indicators, employment information, medical information, education information and financial information (National Institute of Standards and Technology 2010).

Whenever personally identifiable information is collected, stored, used, or deleted, issues concerning privacy, confidentiality, ownership, informed consent and data security may arise (Moor 1997; Clark et al. 2015). Though ethical issues related to the use of personally identifiable data are nothing new, this data's digital nature does raise these questions in new ways.
A recent prominent case is the political data firm Cambridge Analytica. It improperly collected the private information of more than 87 million Facebook users without their knowledge and sold psychological profiles of American voters to a political consulting firm connected to Donald Trump during the 2016 election (Cellan-Jones 2018; Rosenberg & Frenkel 2018; Kang & Frenkel 2018). Other recent cases include WhatsApp sharing user account information with Facebook (Denham 2016), or Google scanning the content of Gmail users' email messages for marketing purposes (Statt 2017).

In the research community, a much-discussed case arose in 2008 when a group of researchers officially released the de-identified profile data collected from the Facebook accounts of a cohort of 1,700 college students from a U.S. University (Lewis et al. 2008). However, it proved easy to identify the university, and the inclusion of data elements such as students' majors, nationalities, and extracurricular activities made it likely that individual students could be re-identified (Zimmer 2010). There are other more recent examples. A study published in June 2014 manipulated the News Feeds of almost 700,000 Facebook users without informing them of their being involved in a research study (Kramer et al. 2014; Kleinsman & Buckley 2015). Moreover, a controversial face recognition study using facial images uploaded on a dating site spurred discussion as to whether the researchers were entitled to use the images without the consent of the dating site users (Leetaru 2017).

Cases like these have helped raise awareness of ethical issues in digital data management in many different fields, including research involving online data collection and Big Data. Institutions engaged in digital data management have become aware of the need to address these issues and to set priorities and specify rules in this area of practice. Becoming aware of this need, some have developed codes of ethics, policies, or guidelines shaping data management practices. While this may in part be a reaction to a specific problem that occurred in the past, many of these

standards may also serve as proactive goals and help to shape the future of digital data management (Metcalf 2014). The development of policies, guidelines, and ethics codes relating to digital data management can be seen in various contexts. It includes ongoing revisions of the collection of major research ethics regulations known as the Common Rule (Metcalf & Crawford 2016; Vitak et al. 2016). The regulations can draw from well-established standards in related fields, as concern over the proper handling of data is widespread through the ethical guidelines of medicine, life sciences, and social sciences, among others.

4.2 Role of Ethics Codes and Guidelines in Process

Codes of ethics and ethical guidelines reflect morally permissible standards of conduct that members of a group make binding upon themselves and ideally should change as the group faces new ethical issues or questions. Codes of ethics also call upon members of that group to go beyond the standard dictates of the law and ordinary morality (Davis 2015). At their best, codes of ethics help lay the foundation for how members of a profession should act in a given situation, and help build trust between members of that profession and the public (Davis 1991).[1]
Since their inception, professional codes of ethics have often sought to direct how practitioners gather, use, store, share, and ultimately dispose of their data. For instance, the American Anthropological Association, in their 1971 "Principles of Professional Responsibility."[2] discusses the paramount responsibility anthropologists have to the individuals being studied and goes on in 1(c) to outline,

> "Informants have a right to remain anonymous. This right should be respected both where it has been promised explicitly and where no clear understanding to the contrary has been reached. These strictures apply to the collection of data by means of cameras, tape recorders, and other data-gathering devices, as well as to data collected in face-to-face interviews or in participant observation. Those being studied should understand the capacities of such devices; they should be free to reject them if they wish, and if they accept them, the results obtained should be consonant with the informant's right to welfare, dignity, and privacy."

The updated 2012 Statement on Professional Responsibility has greatly enlarged this, including a new focus on digital data.

> "The use of digitalization and of digital media for data storage and preservation is of particular concern given the relative ease of duplication and circulation. Ethical decisions regarding the preservation of research materials must balance obligations to maintain data

[1] See the Ethics Codes Collection of Illinois Institute of Technology's Center for the Study of Ethics in the Professions (http://ethicscodescollection.org), a digital repository of around 3,000 professional codes that seeks to trace the development and use of ethics codes across many professions.

[2] http://ethics.iit.edu/ecodes/node/3162

integrity with responsibilities to protect research participants and their communities against future harmful impacts."[3]

Besides the growing relevance of digital data management, this example of the comparison between the American Anthropological Association's 1971 and 2012 statements on professional responsibility exemplifies how codes of ethics are works in progress and how these documents develop and expand over time. They respond to social and technological developments and are initiated or modified following disruptions of everyday professional practice (Metcalf 2014).

The implementation of the General Data Protection Regulation (GDPR) in 2018 has also had a profound impact on guidelines and policies for businesses and industry associations that suddenly had to meet the expanded requirements for informed consent in handling the personal data of their users. In 2016, the European Union passed the GDPR, a legal regulation that seeks to protect individuals in contexts involving personal data collection and analysis. This legal framework went into effect in May 2018 and has had a profound impact on the use of digital data in sectors worldwide and on ethics codes relating to digital data management.

The GDPR, "…applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the European Union, regardless of whether the processing takes place in the European Union or not…" and to the processing of personal data of data subjects who are in the European Union, where the processing activities are related to the offering of goods and services or the monitoring of behavior (Article 3, GDPR). The primary goal of the GDPR is to protect the rights of data subjects by giving them insight into and control over the collection and processing of their personal data (Abiteboul 2019). In Chapter II of the GDPR, the regulation lays out fundamental principles relating to the processing of personal data: lawfulness, fairness, and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, and confidentiality, and accountability. In Chapter III, it lays out the rights of the data subject: the right to be informed, the right to access, the right to rectification, the right to erasure, the right to restrict processing, the right to data portability, the right to object, and rights in relation to automated decision-making and profiling (EU 2016/679).

Unlike the GDPR, ethics codes and ethical guidelines are not legally binding. Whereas the GDPR lays out a general legal framework, ethics codes and ethical guidelines are much more context-specific. Often, they reflect in more detail about the meaning or significance of a particular principle or concept in their respective context. Insofar, even though legal regulations such as the GDPR clearly trump whatever may be written in a code of ethics, the guidance found in ethics codes clearly is reflective of the ethical aspects involved in the respective field.

4.3 Ethics Codes and Guidelines in Digital Data Management

In guidelines, policies, and ethics codes developed in the many fields of digital data management, a wide variety of ethical principles and concepts are addressed, including (Table

[3] Section 6. http://ethics.iit.edu/ecodes/node/6005

1): dignity, respect for persons and communities, informed decision-making and informed consent, transparency, beneficence, justice, risk minimization and fair distribution of benefits and risks, accountability, procedural fairness, non-discrimination, accessibility, dissemination, reciprocity, engagement, recognition and attribution, respect for law and public interest, authorship, ownership, and custodianship (see for example Averweg & O'Donnell, 2007; Centre for Social Justice and Community Action, Durham University 2012; Dittrich & Kenneally, 2012; Global Alliance for Genomics and Health 2014; Clark et al., 2015; Oxfam, 2015). Among these, questions related to privacy and informed consent are frequently considered of vital importance. The various guidelines, ethics codes, and policies stress different concepts and principles, use different definitions for the various concepts and standards, and frame the concepts they use in different ways. They also discuss these elements in a variety of different contexts. Because of these factors, it is necessary to analyze the various documents in more detail for a more comprehensive discussion.

In what follows, we shall focus our analysis on informed consent and informed-consent related standards. There are two reasons for this: First, informed consent is one of the most prominent standards in digital data management. It is also of central relevance in the GDPR. Second, for decades, informed consent has been playing a crucial role in a broad spectrum of online and offline management of personal data. Especially in the context of clinical practice and research involving humans, there has been a particularly high awareness of data management-related ethical issues both in non-digital and digital contexts.

In medicine, a long tradition of policies and guidelines relating to data management exists, including, most prominently, the Declaration of Helsinki and the Belmont Report (World Medical Association, 1964/2013; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978; UNESCO, 2003; Global Alliance for Genomics and Health, 2014). Documents devised in this field may prove helpful for developing policies and guidelines relating to other fields of digital data management. A significant example of this strategy is the Menlo Report – Ethical Principles Guiding Information and Communication Technology Research (Dittrich & Kenneally, 2012). Developed for the Department of Homeland Security to provide a framework for ethical guidelines for computer and information security research, it relies on the Belmont Report issued in 1979, which identifies three basic ethical principles underlying research with human subjects: respect for persons, beneficence, and justice (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). In the following section, we will explore informed consent in the medical context to better understand the longstanding tradition of this concept and how informed consent-related standards can be and have been applied to digital data management.

4.4 Models of Practice: Informed Consent

Informed consent in the medical context is the requirement of a formal agreement by a patient to permit a healthcare intervention after having been provided adequate information on the context, risks, and benefits. The concept of informed consent has a long tradition in medicine and research involving human subjects (Nuremberg Code, 1949; World Medical Association, 1964/2013; National Commission for the Protection of Human Subjects of Biomedical and

Behavioral Research, 1979; Faden & Beauchamp, 1986; Mason & O'Neill, 2017). In medicine and research involving human subjects, several aspects of the concept of informed consent are of central relevance: transparent information has to be provided on the relevant aspects, benefits and risks; the informed consent has a gatekeeper function, i.e., it is to be given before anything else happens; a waiver is possible; participants can quit at any time without negative implications; on request of the participant, the data collected has to be destroyed; the data collected is to be used only for the purpose or purposes specified; if the data is to be used in additional contexts (data sharing), informed consent is needed; the data collected is stored only for a limited duration of time that is clearly specified; and special protection for non-competent individuals (children, etc.) has to be in place.

Informed consent has also been considered of central relevance in the context of information and communication technology. Notably, "The Menlo Report – Ethical Principles Guiding Information and Communication Technology Research" (Dittrich & Kenneally, 2012), using the Belmont Report as a basis, discusses respect for persons and informed consent as one of the central standards governing information and communication technology research.
The Menlo Report proposes a framework for ethical guidelines to be used in research about or involving information and communication technology (ICT), and discusses four core ethical principles, and reflects on the role of these principles in the context of ICT: Respect for Persons; Beneficence; Justice; and Respect for Law and Public Interest.

The Menlo Report restates the principle of Respect for Persons in the context of Information and Communication Technology Research (ICTR) as follows:

> Respect for persons: "Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection." (p.5)

Thus, the Menlo Report considers informed consent and informed consent-related aspects as central standards governing information and communication technology research based on the principle of respect for persons. It stresses the overall relevance of informed consent in research involving digital data by drawing direct connections to the medical context.
Furthermore, the Menlo Report states that (Dittrich & Kenneally, 2012, p. 7): "In the ICTR context, the principle of Respect for Persons includes consideration of the computer systems and data that directly interface, integrate with, or otherwise impact persons who are typically not research subjects themselves."

In the study described in the following, we shall analyze and discuss the relevance and role of informed consent and informed consent-related standards in the context of ethics codes and guidelines referring to ICT and digital data management.

4.5 Study Methodology

This study examines 31 different codes of ethics and guidelines (see appendix 2) from the Ethics Codes Collection held by the Center for the Study of Ethics in the Professions at the Illinois Institute of Technology (http://ethicscodescollection.org). This publicly available collection includes around 3,000 normative documents from approximately 1,750 different institutions. While not fully representative of all the ethics codes and guidelines from the various fields of digital data management, this cross-section of ethics codes from government (European, U.S., and international), business and industry associations (marketing, management, and social media), and non-governmental and professional associations (social aid societies, as well as professional associations from the areas of computer science, health, and information sciences) provides a set of representative ethics codes that guide how to handle digital data, and what ethical principles they draw upon in providing this guidance.

The database was searched using a keyword search looking for guidelines that included at least one or more of the terms "informed consent," "data," "digital data," “privacy,” or “confidentiality.” Documents needed to mention the handling of data from human subjects explicitly, and at least in part discuss principles, mechanisms, and strategies for handling potentially confidential data from users, patients, or data subjects. In cases where there were multiple versions of the document in the collection, we opted to use the most recent version. From our initial set of 43 documents, we narrowed the set to 31 individual documents based on the above set of criteria. Our final set of documents were developed from 2002 to 2020 and represent a broad swath of institutions, sectors, and fields.

Figure 4.1: Number of Documents in Sample by Sector

Sector	Number of Documents
Business and Industry Associations	9
Government and Intergovernmental Organizations	9
Non-Governmental Organizations	6
Professional Associations	7

The remaining 31 documents were divided into four different categories: business and industry associations that include organizations such as Accenture, Facebook, and the Mobile Marketing Association; government organizations, like the U.S. Federal Trade Commission and UNESCO; non-governmental/educational organizations such as Oxfam and the University of Melbourne's Carlton Connect Initiative (which does health research); and professional associations such as the Association for Computing Machinery.

In the analysis of documents, we used the methodology of qualitative content analysis. Both of the authors began in March of 2020 by reading four example documents, and, based on the various aspects of the standard of informed consent in medical contexts, principles governing the use of data in human subjects research, and data ethics, in particular, the principles and rights from the GDPR, developed an initial list of preliminary informed consent-related ethics topics (see supplementary table 1). We then engaged in reliability testing, with both coders individually coding the four documents and comparing results. Through discussion, we improved our final set of codes to 18 that represent the various aspects of informed consent.

The coding process began in April of 2020. All codes entered by the authors were collected in a spreadsheet individually. Each document received a score between 0 and 2 for each of the 18 codes, with 0 referring to an absent code, 1 referring to a minor reference to the code, and 2 to a substantial or developed reference to the code. A substantial reference includes a paragraph or heading relating to the code, and a minor reference might consist of only the mention of the term, with little or no development of the topic. Our coding strategy took into account the length of the document. Therefore, in a one-page set of guidelines, a single bullet point or sentence would be counted as a 2, whereas a 2 would warrant a paragraph or more in a ten-page document.

To ensure the reliability of the coding, each document was coded by the two authors separately, followed by a reconciliation process during which the coders discussed differences and attempted to reconcile differences.

Limitations of this study need to be kept in mind when interpreting the results. Our data set included ethics codes, policies, and guidelines available to the Ethics Codes Collection (ECC). While the ECC is the most extensive collection of professional, business, and governmental codes of ethics and guidelines in the world, it by no means represents the entirety of the digital ethics landscape. Codes are only added to the ECC when copyright permission can be obtained. Otherwise, a link is added to the collection leading the user to a version posted on the authoring institution's website. This approach limited our analysis to only published and publicly available documents. Internal documents developed by businesses and other institutions were not included in the study. Our collection of documents also only included those available in the English language, which ensures that our dataset does not adequately represent ethics codes from non-western countries.

4.6 Overview – Appearance of Informed Consent-Related Standards in Ethics Codes and Guidelines

The following divides our sample of documents that discuss informed consent in digital data management by sector. You can see the mean score of all the documents in that sector for each principle/concept we coded for in each table.[4]

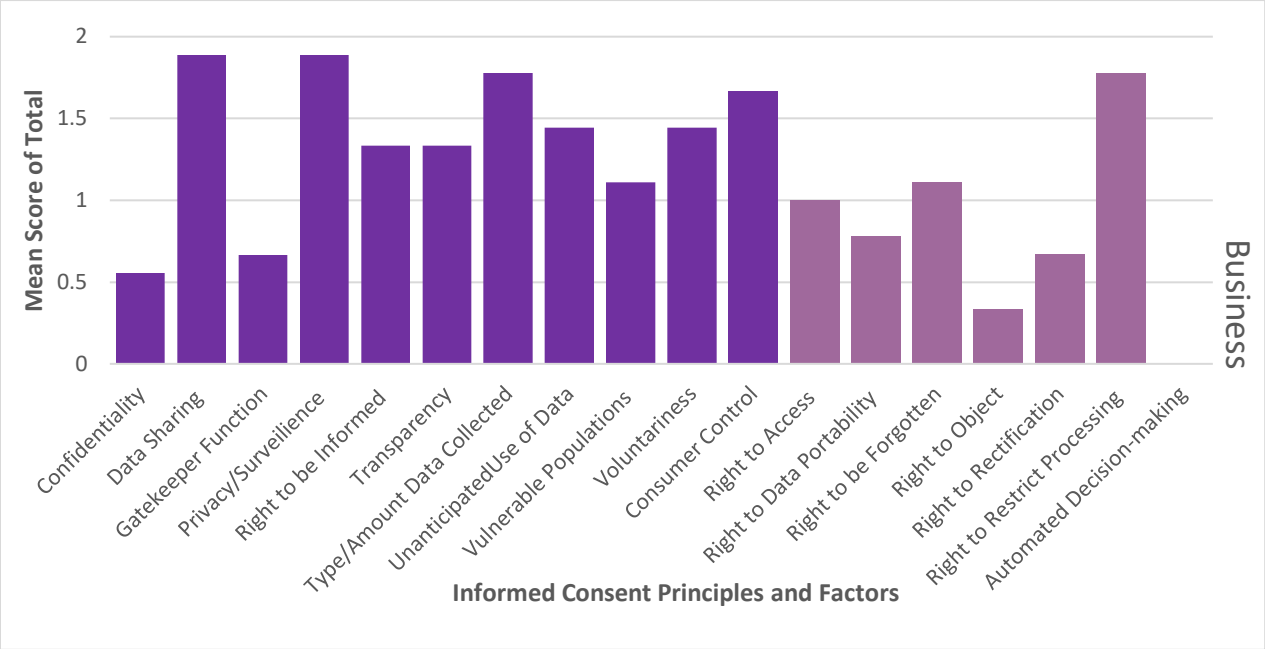
[4] See Supplementary Table 1 for a list and definitions for the 19 codes being used here.

Figure 4.2: Mean score by sector

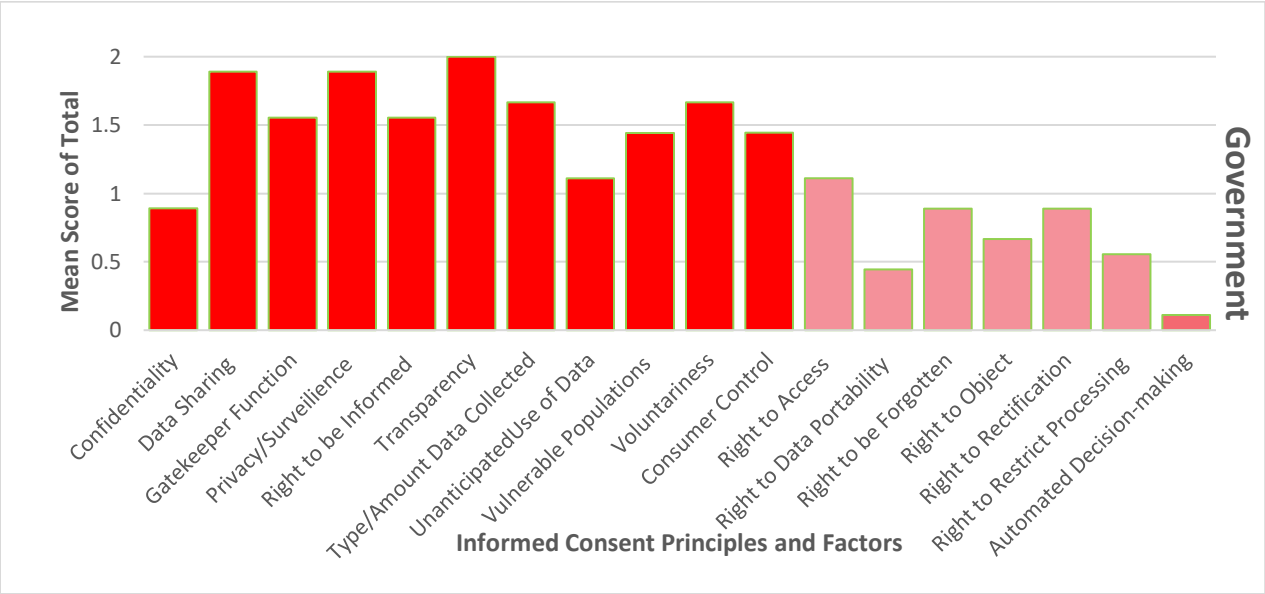
Codes in yellow represent different aspects of consumer control, as outlined by the GDPR.

Sector	Confidentiality	Data Sharing	Gatekeeper Function	Privacy/Surveillance	Right to be Informed	Transparency	Type/Amount of Data Collected	Unanticipated Use of Data	Vulnerable Populations	Voluntariness	Consumer Control	Right to Access	Right to Data Portability	Right to be forgotten	Right to Object	Right to Rectification	Right to Restrict Processing	Automated Decision-making
Business	0.6	1.9	0.7	1.9	1.3	1.3	1.8	1.4	1.1	1.4	1.7	1	0.8	1.1	0.3	0.6	1.7	0
Government	0.9	1.9	1.6	1.9	1.6	2	1.7	1.1	1.4	1.7	1.4	1.1	0.4	0.9	0.7	0.9	0.6	0.1
NGO	2	2	2	2	1.8	1.8	2	1.5	1.3	1.7	1.3	1.3	0	0.8	0.8	0.5	0.7	0.3
Professional	1.3	1.9	1.6	1	2	1.7	1.7	0.6	0.9	1.1	1.6	0.4	0	0	0.3	0.3	1	0.4

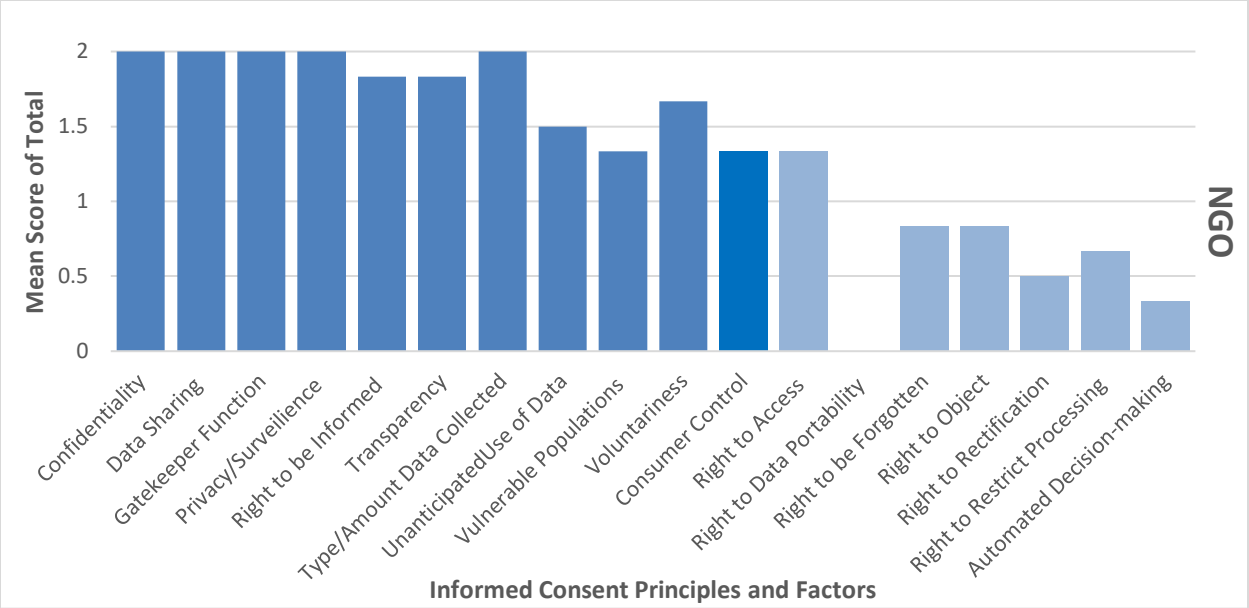
Figure 4.3: Categorization of Informed Consent-Related Codes by Sector
Codes in lighter colors represent different aspects of consumer control, as outlined by the GDPR.



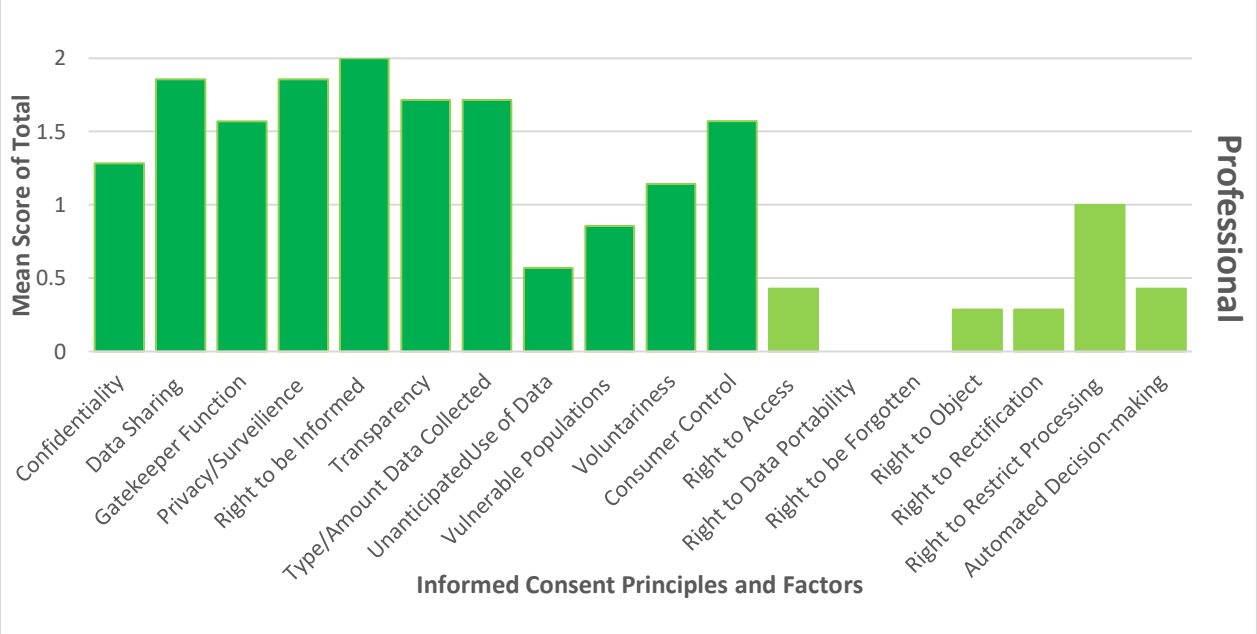
Business = 9 documents



Government = 9 documents



NGO= 6 documents



Professional =7 documents

All four sectors scored the highest in areas of data sharing and the type/amount of data collected, and lowest in the areas of automated decision-making, the right to data portability and the right to object. Traditionally, issues of data sharing and the type/amount of data being collected have been a critical component of traditional concepts of informed consent in the biomedical fields, regardless of how the data is being collected. Institutional review board protocols include questions about these research issues that fall under these regulations under the Common Rule. In areas of business, the Federal Trade Commission uses several regulations (such as the

Children's Online Privacy Protection Act, the Fair Credit Reporting Act, and the Can-SPAM Act, to name a few) (FTC 2019). Issues of data portability and the right to opt-out of automated decision-making are seen at a much lower rate due to the attention being drawn to these issues relatively recently through the GDPR implemented in 2018 and the relatively recent growth of digital data in business, governmental, and health decision-making. As professional associations and governments continue to update these guidelines and professional codes, further attention will hopefully be paid to these critical issues. Data portability is also likely a less relevant topic in some professional fields where the data being gathered probably belongs to research subjects rather than consumers who have legitimate reasons for wanting to move their data from one platform to another.

When looking at business and professional association guidelines and policies, these score high in many codes drawn from the GDPR and on several elements of consumer control but score lower in the areas of confidentiality, gatekeeper functions, and transparency. This may reflect the driving need to meet the demands of national and international regulations rather than the other sectors who often come from a life or social sciences research background, and thereby include more traditional biomedical and research-based principles of informed consent like confidentiality, the duties of the researcher as a gatekeeper, and issues of transparency.

Government guidelines for handling digital data scored high in data sharing, privacy/surveillance, and transparency but scored in the middle on consumer control and many of its related aspects. These guidelines were published in the years 2002-2012 and, therefore, do not reflect more expanded notions of consumer control in the ethical handling of digital data and informed consent. However, issues of privacy and transparency have been on the radar of U.S. governmental institutions for close than 50 years with the passage of the Privacy Act of 1974 (U.S.C. § 552a) that established a code of fair information practices that govern the collection, maintenance, use and dissemination of information about individuals that federal agencies maintain.

NGO documents – many of which deal with either health, human research, or social advocacy – score highest in the areas of confidentiality, data sharing, gatekeeper function, privacy/surveillance, and the type/amount of data collected. Again, this likely stems from field-specific principles arising from biomedical research and other traditions arising from human subject research.

In general, the professional codes of ethics scored relatively high in terms of principles that appear in traditional models of practice but score lower on codes drawn from new regulation and guidelines – such as issues surrounding consumer control.

4.7 A Closer View on Informed Consent-Related Standards in Ethics Codes and Guidelines

In what follows, we'll discuss in more detail the role granted to some of the informed consent-related standards in information and communication technology and digital data management. Given the broad spectrum of informed consent-related standards, it is not possible to elaborate on

all the aspects represented in the coding approach described above in this chapter. Instead, we'll focus on some prominent examples. These relate to informed decision-making and informed consent's gatekeeper function, transparency, consumer control, type and amount of data collected, and data sharing.

4.7.1 Gatekeeper Function

As in medicine, informed consent in digital data management has a gatekeeper function, i.e., informed consent must be given before any kind of activity or data collection commences. However, there are several differences between these fields: In medicine, informed consent presupposes a doctor-patient relationship and typically requires a medical doctor to convey relevant information in a conversation to a patient. This interchange allows the doctor to ask questions and to verify whether the patient has understood the information. However, none of this is possible in digital data management. Here, individuals give their consent by clicking a button, often with the details of how their personal data will be gathered, utilized, and possibly sold buried in a "terms and conditions" agreement. There is usually no face-to-face interaction in online environments, and it is not even clear whether the users have read and understood the information provided (Clark et al., 2015).

In this context, the *European Data Protection Supervisor Opinion 4/2015: Towards a new digital ethics* stresses that, as human beings are not entirely rational, the fact that individuals have given informed consent for the processing of their personal information does not entitle others to unlimited use:

> "Under EU law, consent is not the only legitimate basis for most processing. Even where consent plays an important role, it does not absolve controllers from their accountability for what they do with the data, especially where a generalized consent to processing for a broad range of purposes has been obtained."

In Western medicine, there is a general agreement that informed consent, usually written consent, is required, except in emergency situations. In emergencies in which the individuals receiving medical treatment are not able to give informed consent, proxy consent is considered an alternative. Furthermore, in special situations, a waiver is possible.

In contrast, in digital data management, there is a broader spectrum of positions. The Code of Ethics for Community Informatics Researchers (Averweg & O'Donnell, 2007, p. 2-3) expresses a traditional view, similar to medical conventions. It requires that research should commence only after free and informed consent has been given, ordinarily in writing, by prospective participants.

However, there are also more liberal views concerning the need to obtain informed consent. For example, The OECD Privacy Framework (2013) states in the Collection Limitation Principle that "there should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject." The question, of course, is when it is appropriate or when it may not be appropriate or not necessary to obtain consent.

Of relevance here is the context of data collection, which may make it difficult to obtain informed consent, such as passive methods of collecting data where no interaction can take place during which a user can give or deny consent. The Menlo Report discusses the possibility of researchers seeking waivers of informed consent in those cases in which obtaining informed consent would make it impossible to achieve research objectives. Accordingly, this requires that (Dittrich & Kenneally, 2012, p. 8):

> "(1) The research involves no more than minimal risk to the subjects; (2) The waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) The research could not practicably be carried out without the waiver or alteration; and (4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation."

The Menlo Report mentions situations in which it would be too difficult to identify all individuals from whom consent should be sought or to practicably obtain consent as situations in which a waiver of informed consent or a waiver of documentation of informed consent may be the only option. For example, in a communication traffic modeling study, it may not be feasible to obtain consent from millions of users.

However, there are also more ambiguous situations or contexts in which data may be collected or analyzed in ways individuals are not aware of. This may be the case when digital data is used in unanticipated ways without asking those who contributed for their consent (Clark et al., 2015; Zimmer 2010). For example, when research is done based on material posted on social media. There is no consent yet on how to deal with situations like this.

On the one hand, some have claimed that there is no need for informed consent because the material posted on social media can be accessed freely online (Zimmer, 2010). This position can be seen as being backed up by the American Anthropological Association's Ethics Statement (2012). It states, "…the observation of activities and events in fully public spaces is not subject to prior consent", and therefore it may be concluded by analogy that prior consent may also not be needed for the observation of public internet spaces such as openly accessible forums or social media. It is questionable whether this analogy is valid; however, especially as social media research involves data resulting from systematic data collection, which would be much more difficult to obtain in public spaces. Furthermore, re-identification issues may arise, as mentioned above, in the 2008 study involving Facebook accounts (Zimmer, 2010).

On the other hand, the fact that data pertaining to individuals that may include personally identifiable information is collected without their knowledge clearly is a problem in itself. In view of this, some have argued that various possibilities for obtaining consent for research involving social media posts may be available, such as contacting those who wrote the posts and asking for permission or gaining consent from respective groups beforehand (Clark et al., 2015).

4.7.2 Transparency

Transparency refers to the requirement that information on the relevant aspects of data collection and data management must be provided in a clear, comprehensible, and accessible way. This also

includes potential risks. Another transparency requirement is that users are aware of what actions are performed, for example, that users know about data collection taking place. Transparency is considered central relevance by various ethics codes and guidelines, especially in digital data analytics and health informatics (see, for example: Global Alliance for Genomics and Health, 2014; Digital Analytics Association, 2011; European Commission, 2016).

The Web Analyst's Code of Ethics by the Digital Analytics Association (2011) strongly advocates transparency for practitioners who adhere to their ethics code, stating, "I agree to educate my clients/employer about the types of data collected, and the potential risks to consumers associated with those data." The Code of Ethics requires practitioners to encourage their clients and employers to fully disclose consumer data practices in clear language and educate these parties in how technologies could be perceived as invasive.

The Global Alliance for Genomics and Health stress in their Framework for Responsible Sharing of Genomic and Health-related Data (2014) that information has to be developed and provided on the purposes, processes, procedures, and governance frameworks involved, and that the information provided "should be presented in a way that is understandable and accessible in both digital and non-digital formats." (p.4).

Concerning the requirement to present transparent information, the European Commission's Draft Code of Conduct on Privacy for Mobile Health Applications (2016, p. 7) says: "Note that consent requires that users have been provided with clear and comprehensible information first. Key information shall not be embedded in lengthy legal text."

These paragraphs relate to the well-known challenges to transparency that exist in the presentation of information in ways that are difficult to understand, especially the provision of long and complex "terms and conditions" texts with a lot of details. This may lead to the majority of users failing to read the information and people just clicking on "accept" to get rid of it.

In the context of online behavioral advertising, the Interactive Advertising Bureau's IAB Code of Conduct uses a narrow and rather indirect approach to transparency. Whereas the code of conduct says in the section on transparency that, "Third Party and Service Providers should give a clear, meaningful, and prominent notice on their own websites that describe in detail their data collection and use practices," on the page that contains the advertisement and where the data is collected, only a clear, meaningful, and prominent link to the above disclosure has to be provided. Thus, this is an indirect procedure that requires the consumer to find the link and follow it to the related homepage of the third party that offers disclosure on data collection. Transparency and explicit information transfer require that individuals, first and foremost, know that their data is being collected. Without this knowledge, informed consent simply is not possible.

4.7.3 Consumer Control

Important, informed consent-related standards relate to the requirement that consumers/users are able to control whether or not to participate in research activities or to allow data collection and

to exert control over the ways their data is being used. The GDPR specifies the right to access, the right to rectification, the right to erase (or the right to be forgotten), the right to restrict processing, the right to data portability (the right to shift your data from one service provider to another by moving personal data), the right to object, and rights in relation to automated decision-making and profiling.

Particular challenges arise when consumers are not aware of their choices or when parties assume tacit consent. Opt-in and opt-out mechanisms are ways that attempt to deal with this problem. In order to avoid a situation in which users are not aware of their data being collected, the National Information Standards Organization (NISO) write in their NISO Privacy Principles (2015) in the section on informed consent: "The default approach/setting should be that users are opted out of library services until they explicitly choose to opt-in."

Whereas the NISO Privacy Principles assume that the standard approach is that users are opted out and take steps to actively opt-in if they want so, the *Web Analyst's Code of Ethics* by the Digital Analytics Association focuses on user's ability to actively opt-out of data collection practices – implying that the default setting is opt-in. It says in a paragraph on consumer control:

> "I agree to inform and empower consumers to opt-out of my clients/employer data collection practices and to document ways to do this. To this end, I will work to ensure that consumers have a means to opt-out and to ensure that they are removed from tracking when requested."

While in general, the availability of an opt-out option is considered central for consumer control, depending on the context, consumer control may be more or less challenging to achieve. In particular, issues may arise when the default option is an opt-in option, or when opt-out options are offered of which users may not be aware of.

In online behavioral advertising, the default option is that a consumer's data are collected. The Interactive Advertising Bureau's IAB Code of Conduct states that "A Third Party should provide consumers with the ability to exercise choice with respect to the collection and use of data for Online Behavioral Advertising purposes or the transfer of such data to a non-Affiliate for such purpose." In reality, however, consumers may not be aware of this opt-out option, which may be challenging to find.

The same is true for consumer device-based data collection in and around retail shops. The consumers may be unaware of data collection, even if, as suggested by the *PrivacySIG Code of Conduct*, the retail shops use stickers "to signal to the shopper that Retail Intelligence is being practiced around this location." PrivacySIG is a Special-Interest Group consisting of companies active in retail intelligence. For organizations subscribing to this opt-out approach, store customers must find the notices about the tracking system being used, navigate to the opt-out page offered by PrivacySIG, and then enter their unique MAC address to opt-out of any future tracking by organizations in the PrivachSIG. Similar suggestions are made by the Future of Privacy Forum (2013). Clearly, this assumes a high level of diligence, action, and comprehension on the part of the individual customer.

4.7.4 Type and Amount of Data Collected

In general, most guidelines we surveyed specified that only data that is relevant for a particular purpose should be collected. For example, the OECD Privacy Framework (2013) say in their "Data Quality Principle," that, "Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date."

The NISO Privacy Principles (2015) distinguish between different types of personal data, saying that certain types of personal data (for example on gender, race, socioeconomic status, ability) are considered more sensitive, and therefore the decision to collect and use them should require higher levels of scrutiny and justification, and, once collected, the data should receive extra protection.

Concerning the amount of data to be collected, the authors of the *Accenture: Universal Principles of Data Ethics* (p. 8) write that collecting data just for the sake of more data may complicate analysis, and goes along with risks and unpredictable harmful future consequences. Several ethics codes and guidelines stress that personal data collected is to be used only for the purpose specified. The individuals gave their consent (OECD Privacy Framework, 2013; NISO Privacy Principles, 2015; Accenture Universal Principles of Data Ethics, undated).

4.7.5 Data Sharing

While digital data can be easily shared, there is general agreement that (with the exception of cases of law enforcement) personal data should not be shared without the informed consent of those to whom the data pertain. Whenever personal data is used in additional contexts, informed consent is needed (American Anthropological Association 2012, Dittrich & Kenneally, 2012; OECD Privacy Framework, 2013)

For example, the Menlo Report (p. 7) states: "[…] informed consent for one research purpose or use should not be considered valid for different research purposes."

The authors of the Accenture Principles for Data Ethics also direct attention to the reuse of data sets. In principle 2, they say: "Correlative use of repurposed data in research and industry represents both the greatest promise and greatest risk posed by data analytics."
However, the European Commission's Draft Code of Conduct on Privacy for Mobile Health Applications (2016) outlines that secondary processing of data for historical, statistical, or scientific purposes, even when these purposes were not originally communicated, may still be possible with anonymized or pseudonymized data:

> "Any processing of personal data must be compatible with the purposes for which you originally collected the personal data, as communicated to the users of your app. Secondary processing of the data for historical, statistical or scientific purposes (assuming that these purposes were not originally communicated) is, however still considered as compatible with original purposes if it is done in accordance with any national rules adopted for such secondary processing. This means that, in order to process data for such secondary purposes, you will need to determine which national laws apply, and respect any restrictions."

4.8 Discussion

This chapter reflects on the role of informed consent and informed consent-related standards in codes of ethics and guidelines pertaining to digital data management. The ethics codes and guidelines reveal the informed consent-related aspects considered of relevance in the respective contexts and provide details on the roles informed consent-related standards have in the respective areas. They are not legally binding, however, and are always subordinated to the respective legal framework. Furthermore, in some of the ethics codes and guidelines, modifications that adjust to recent legal changes may be expected in the not-too-distant future.

Overall, in our analysis, we found that in most of the ethics codes, policies, and guidelines examined, informed consent and informed consent-related standards are considered of relevance, and a transfer of informed consent-related standards from medicine to digital data management is taking place. Central relevance is allotted, especially in the context of digital data management in (health-related) research.

However, in other contexts, such as marketing or mobile applications, we found the standards modified, weakened, or broadly reshaped. Examples include parties assuming tacit consent or offering only opt-out options of which users may not be aware. There is also a limited understanding of what should be considered personally identifiable information that seems to either exclude Mac or IP addresses or to be sure to pseudonymize the "unique data" (PrivacySIG, Undated) or to de-identify or de-personalize personal information or unique device information as soon as technically possible (Future of Privacy Forum 2013).

This observation is in line with the results of a 2005 study that examined the privacy policies of 22 online retailers and online travel agencies. The author Irene Pollach (2005) found a high level of complexity in the language used and states that companies "… benefit from obfuscating, mitigating, and enhancing data handling practices in that this helps them to obtain data they would not have access to if users were fully informed about data handling practices." (Pollach, 2005, p. 232).

Even when informed consent documents exist, misconceptions can still arise, as shown in a 2012 article by Erika Check Hayden in the journal *Nature* entitled "Informed Consent: A broken contract" (2012). The article discusses a case of the gene-testing company 23andMe, which asked participants to sign an informed consent document allowing their data to be used in research, and that this research might lead to the company patenting and commercializing products or services. Despite this, confusion occurred, illustrating the divide between researchers and companies and the public in how they understand their data is likely to be used. The article goes on to outline some options to improve transparency, including researchers sending participants regular emails documenting how their data is being used, relying on individuals uploading their own data, and how future technologies might allow participants to track their use of data over time (Hayden 2012).

Overall, in digital data management, a central issue centers around the question of how difficult it is for users to make free and well-informed decisions concerning their personal data and to

exert effective user control. Lack of transparency and conditions impeding effective user control contribute considerably to this problem.

However, behavioral and cognitive factors also play a role. Alessandro Acquisti and Jens Grossklags (2005) found that many parameters affect an individual's privacy decision-making, including inconsistencies in discounting (preferring to opt for a reward received sooner than avoiding later negative consequences later) that lead to under-protection and over-release of personal information. The authors stress that individuals may lack information to make privacy-related decisions, and even when they have sufficient information, they likely trade long-term privacy for short-term benefits (Acquisti & Grossklags 2005).

Daniel J. Solove discusses several cognitive problems that impede privacy self-management (Solove, 2013, p. 1888): "(1) people do not read privacy policies; (2) if people read them, they do not understand them; (3) if people read and understand them, they often lack enough background knowledge to make an informed choice; and (4) if people read them, understand them, and can make an informed choice, their choice might be skewed by various decision-making difficulties."

As the above reflections show, it is necessary to raise users' awareness of the relevance of privacy and the possibilities of data protection and user control. On the other hand, there is a clear need for institutions involved in digital data management to increase transparency and develop ethics codes, policies, and guidelines that include effective informed consent-related standards.

A transfer of the model of informed consent to digital data management comes with chances and limitations. The transfer is most evident in those digital data management fields involving human subjects research; it comes with some strains in other fields such as online marketing or commercial data uses. Whereas individual autonomy has been considered central in medicine and in research involving humans for decades, in commercial contexts, the focus is less on individual autonomy but more on a company's financial interests.

Overall, however, it is a more than plausible assumption that similar activities around personal data management involve similar ethical issues and require similar strategies, independently of whether they rely on digital or non-digital data management.
While it may not be possible to transfer without adjustments the informed consent model rooted in medicine to digital data management, the informed consent model certainly serves as an essential reference point. The model delineates a high standard that helps to protect the users' privacy and autonomy. The model of informed consent can provide guidance for digital data management, such as:

- attempt to obtain consent in as many contexts as possible, even if this may not always be feasible;
- prefer opt-in options over opt-out options;
- provide comprehensive and transparent information so that the users are enabled to make an informed decision;
- seek to collect as little personally identifiable information as possible;
- only keep data as long as necessary;

- de-identify data if this does not render the data unusable for its intended purpose.

One of the issues to be further discussed is which kinds of data collection require informed consent. Whereas it is generally agreed that personally identifiable information that allows us to identify, trace, or contact a person requires the person's informed consent, the same may hold for non-identifiable data that tracks individuals' behavior. The latter applies to data collection and analysis for purposes like economic benefit or political influence. To the maximum extent possible, individuals should know what information pertaining to them is planned to be used for, and they should be allowed to agree or disagree and opt out of these potential uses.

4.9 Conclusion

Even though our analysis is not exhaustive, it can be said that up to now, only a limited number of guidelines and ethics codes on digital data management are available. For sure, there is space for more documents covering ethical issues in digital data management. As technological change is going on and new ways to collect, use, share, and dispose of digital data are evolving, there is a need to reflect on past and current practices, rethink existing priorities, and reflect on future ethical guidance. Existing ethics codes and guidelines in digital data management help to raise awareness of ethical issues in the field and may serve as a starting point for further developments. Overall, ethics codes, policies, and guidelines may help to develop a framework that organizations and bodies can use to guide their collection, use, sharing, and disposal of digital data.

Acknowledgments: This research was funded through a generous grant from the John D. and Catherine T. MacArthur Foundation.

References

Abiteboul, Serge, and Julia Stoyanovich. 2019.. "Transparency, Fairness, Data Protection, Neutrality: Data Management Challenges in the Face of New Regulation." *Journal of Data and Information Quality* 11(3): Article 15. https://doi.org/ 10.1145/3310231.

Accenture. Undated. "Universal Principles of Data Ethics." https://www.accenture.com/t20160629T012639Z__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf. Accessed 29 March 2019.

Acquisti, Alessandro, and Jens Grossklags. 2005. "Privacy and Rationality in Individual Decision Making." *IEEE Security & Privacy* 3(1): 26-33.

American Anthropological Association. 1971. “Principles of Professional Responsibility.” http://www.americananthro.org/ParticipateAndAdvocate/Content.aspx?ItemNumber=1656. Accessed 1 April 2019.

American Anthropological Association. 2012. *Principles of Professional Responsibility*. Available at: http://ethics.americananthro.org/category/statement/. Accessed 8 March 2019.

Average, Udo, and Susan O’Donnell. 2007. “Code of Ethics for Community Informatics Researchers.” *The Journal of Community Informatics* 3 (1). http://ci-journal.net/index.php/ciej/article/view/441/307. Accessed 26 March 2018.

Cellan-Jones R. (2018). Facebook data – as scandalous as MPs’ expenses? *BBC News.* 19 March. http://www.bbc.com/news/technology-43458110 Accessed 4 April 2018.

Center for the Study of Ethics in the Professions, Illinois Institute of Technology. 2018. *Ethics Codes Collection.* http://ethicscodescollection.org. Accessed 25 February 2020.

Centre for Social Justice and Community Action, Durham University. 2012. *Community-based participatory research: A guide to ethical principles and practice*. Available at:http://www.livingknowledge.org/fileadmin/Dateien-Living-Knowledge/Dokumente_Dateien/Toolbox/LK_A_CBPR_Guide_ethical_principles.pdf Accessed 17 December 2017).

Clark, Karin, Matt Duckham, Marilys Guillemin, Assunta Hunter, Jodie McVernon, Christine O’Keefe, Cathy Pitkin, Steven Prawer, Richard Sinnott, Deborah Warr, and Jenny Waycott. 2015. *Guidelines for the Ethical use of Digital Data in Human Research*. Melbourne: The University of Melbourne, Melbourne School of Population and Global Health. https://www.carltonconnect.com.au/wp-content/uploads/2015/06/Ethical-Use-of-Digital-Data.pdf Accessed 12 February 2018.

Davis, Michael. 1991. Thinking Like an Engineer: The place of a code of ethics in the practice of a profession. *Philosophy and Public Affairs* 20(2): 150-167.

Davis Michael. 2015. “Codes of Ethics.” In Holbrook JB and Mitcham C (eds) *Ethics, Science, Technology and Engineering, 2nd Edition.* Farmington Hills, MI: Gale, Cengage Learning, pp. 380-383.

Denham, E. 2016. *Information Commissioner updates on WhatsApp / Facebook investigation.* In: *ICO Information Commissioner’s Office Blog*. 7 November. https://iconewsblog.org.uk/2016/11/07/information-commissioner-updates-on-whatsapp-facebook-investigation/ Accessed 17 December 2019.

Digital Analytics Association. 2011. *The Web Analyst’s Code of Ethics*. https://www.digitalanalyticsassociation.org/codeofethics Accessed 18 November 2017.

Dittrich David, and Erin Kenneally. 2012. *The Menlo Report – Ethical Principles Guiding Information and Communication Technology Research*. United States, Department of

Homeland Security. http://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf. Accessed 12 October 2019.

European Commission. 2016. *Draft Code of Conduct on Privacy for Mobile Health Applications*. https://ec.europa.eu/digital-single-market/en/news/code-conduct-privacy-mhealth-apps-has-been-finalised. Accessed 21 November 2019.

European Data Protection Supervisor. 2015. *Opinion 4/2015: Towards a new digital ethics.* Available at: https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf. Accessed 07 January 2018.

Faden Ruth R., and Tom L. Beauchamp. 1986. *A History and Theory of Informed Consent*. Oxford: Oxford University Press.

Future of Privacy Forum. 2013. *Mobile Location Analytics Code of Conduct.*https://fpf.org/wp-content/uploads/10.22.13-FINAL-MLA-Code.pdf. Accessed 07 January 2018.

Global Alliance for Genomics and Health. 2014. *Framework for Responsible Sharing of Genomic and Health-related Data.* https://www.ga4gh.org/ga4ghtoolkit/regulatoryandethics/framework-for-responsible-sharing-genomic-and-health-related-data/. Accessed 18 November 2019.

Hayden, Erika C. 2012. "Informed Consent: A broken contract." N*ature* 486: 312-314. https://doi.org/10.1038/486312a.

Information Commissioner's Office, United Kingdom. 2017. *Guide to the General Data Protection Regulation (GDPR)*. https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf. Accessed 23 March 2018.

Institute for Business Ethics. 2016. Business Ethics and Big Data. *Business Ethics Briefing*, 52. https://www.ibe.org.uk/userassets/briefings/b52_bigdata.pdf. Accessed 19 March 2018.

Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information. 2009. "The HIPAA Privacy Rule." Nass SJ, Levit LA, Gostin LO, editors. 2009. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington (DC): National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK9579/

Interactive Advertising Bureau. Undated. *IAB Code of Conduct*. https://www.iab.com/wp-content/uploads/2015/06/IAB_Code_of_Conduct_10282-2.pdf. Accessed 5 December 2019.

Kang, Cecilia, and Sheera Frenkel. 2018. "Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users." *New York Times.* 4 April. https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html Accessed 8 April 2018.

Kleinman John, and Sue Buckley. 2015. "Facebook Study: A Little Bit Unethical But Worth It?" *Bioethical Inquiry* 12:179-182. https://doi.org/ 10.1007/s11673-015-9621-0

Kramer ADI., Guillory JE. and Hancock JT. 2014. "Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 1111(24): 8788-8790. DOI: 10.1073/pnas.1320040111.

Leetaru Kalev. 2017. "AI 'Gaydar' And How The Future Of AI Will Be Exempt From Ethical Review." *Forbes*. 16 September. https://www.forbes.com/sites/kalevleetaru/2017/09/16/ai-gaydar-and-how-the-future-of-ai-will-be-exempt-from-ethical-review/#704e7602c09a. Accessed 17 January 2018.

Lewis Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. 2008. "Tastes, Ties and Time: A new social network dataset using Facebook.com." *Social Networks.* 30 (4): 330-342. https://doi.org/10.1016/j.socnet.2008.07.002.

Mason, Neil C. and Onora O'Neill. 2017. *Rethinking Informed Consent in Bioethics*. New York: Cambridge University Press.

Metcalf, Jacob. 2014. *Ethics Codes: History, Context, and Challenges*. Council for Big Data, Ethics, and Society. http://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/. Accessed 13 November 2018.

Metcalf, Jacob, and Kate Crawford. 2016. "Where are human subjects in big data research? The emerging ethics divide." *Big Data & Society*, 3 (1): 1-14. https://doi.org/10.1177/2053951716650211.

Moor, James H. 1997. "Towards a theory of privacy in the information age." *ACM SIGCAS Computers and Society*, 27 (3): 27-32. https://doi.org/10.1145/270858.270866.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age.* Washington, D.C.: National Academy Press. http://www.onlineethics.org/?id=34249&preview=true. Accessed 7 February 2018.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, United States, Department of Health, Education and Welfare. (1979). *Belmont Report.* https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/. Accessed 12 January 2018.

National Information Standards Organization. 2015. *NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software- Provider Systems (NISO Privacy Principles.* :https://groups.niso.org/apps/group_public/download.php/16064/NISO%20Privacy%20Principles.pdf. Accessed 09 December 2019.

National Institute of Standards and Technology. (2010). *Guide to Protecting the Confidentiality of Personally Identifiable Information.* (PII), ES-1/ES-2)

https://www.nist.gov/publications/guide-protecting-confidentiality-personally-identifiable-information-pii. Accessed 12 December 2019.

Nuremberg Code. 1949. *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10"*, Vol. 2, pp. 181-182. Washington, D.C.: U.S. Government Printing Office. https://history.nih.gov/research/downloads/nuremberg.pdf Accessed 7 January 2018.

Organization for Economic Co-Operation and Development. (2012). *The Protection of Children Online*. https://www.oecd.org/sti/ieconomy/childrenonline_with_cover.pdf Accessed 12 March 2018.

Organization for Economic Co-Operation and Development. (2013). *The OECD Privacy Framework.* http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf Accessed 13 January 2018.

Oxfam. 2015. *Oxfam Responsible Program Data Policy*. https://policy-practice.oxfam.org.uk/publications/oxfam-responsible-program-data-policy-575950 Accessed 20 January 2018.

Pollach, Irene. 2005. "A Typology of Communicative Strategies in Online Privacy Policies: Ethics, Power and Informed Consent." *Journal of Business Ethics*, 62:221-235. https://doi.org/10.1007/s10551-005-7898-3.

PrivacySIG. Undated. *Code of Conduct*. http://www.privacysig.org/code-of-conduct.html Accessed 16 February 2018.

Rosenberg, Matthew and Sheera Frenkel. 2018. "Facebook's Role in Data Misuse Sets Off Storms on Two Continents." *The New York Times* 18 March. https://www.nytimes.com/2018/03/18/us/cambridge-analytica-facebook-privacy-data.html?smid=tw-share. Accessed 10 April 2018.

Ruof, Mary C. 2004. "Vulnerability, Vulnerable Populations, and Policy." *Kennedy Institute of Ethics Journal.* 14(4): 411-425. https://doi.org/10.1353/ken.2004.0044.

Solove, Daniel J. 2013. "Introduction: Privacy self-management and the consent dilemma." *Harvard Law Review,* 126: 1880-1903.

Statt, Nick. 2017. "Google Will Stop Scanning Your Gmail Messages to Sell Targeted Ads. *The Verge*. https://www.theverge.com/2017/6/23/15862492/google-gmail-advertising-targeting-privacy-cloud-business Accessed 8 January 2018.

Turilli, Matteo, and Luciano Floridi. 2009. "The Ethics of Information Transparency." *Ethics and Information Technology.* 11(2): 105-112. https://doi.org/10.1007/s10676-009-9187-9.

United Nations Educational, Scientific and Cultural Organization. (2003). *International Declaration on Human Genetic Data*. http://ethics.iit.edu/ecodes/node/5863. Accessed 10 January 2018.

United States, Federal Trade Commission. 2019. *Privacy and Data Security Update: 2019.* https://www.ftc.gov/system/files/documents/reports/privacy-data-security-update-2019/2019-privacy-data-security-report-508.pdf

United States, National Institutes of Health (NIH). 2004. *National Institutes of Health (NIH)2003NOT-OD-03-032: Final NIH Statement on Sharing Research Data. NOT-OD-03-032: Final NIH Statement on Sharing Research Data.* https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html Accessed 12 January 2018.

United States, National Science Foundation (NSF). 2017. *Nation Science Foundation (NSF) 2018 Grant Proposal Guide*, Chapter II.C.2.j. https://www.nsf.gov/pubs/policydocs/pappg18_1/index.jsp Accessed 6 March 2018.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12 (2): 197.

Vitak, J, Shilton, K. and Ashktorab, Z. (2016). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. San Francisco, CA: Association of Computing Machinery. DOI: 10.1145/2818048.2820078

World Medical Association. (2013). *WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ Accessed 12 January 2018.

Zimmer, M. 2010. "But the Data is Already Public": on the ethics of research in Facebook." *Ethics and Information Technology*, 12, 313-325.

Supplementary Table 1

This is a list of the 19 central concepts and principles used when coding the 32 ethical documents in this study.

Concept/ Principle	**Definition**
Confidentiality	The treatment of information that an individual has disclosed to an organization/researcher in a relationship of trust, and the expectation (and duty of the organization to ensure) that the information will not be shared with others without permission in ways that are inconsistent with the understanding of the original disclosure. Confidentiality specifically pertains to data (Institute of Medicine 2009).
Consumer Control	Overall control of data by consumer/user
Data Sharing	Data that shared with organizations/individuals outside of the original collector of data.
Gatekeeper Function	Responsibilities of an individual/organization to the apply criteria of informed consent in a way that protects the interests of individuals allowing the organization.
Privacy	The right of an individual to control the extent, timing, and circumstances of sharing one's data, and to keep this data private. (GDPR, Art. 20).
Right to Access	From the GDPR, Art. 15: The right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, access to the personal data.
Right to be Informed	The right to be informed before any kind of activity or data collected commences that involves his or her personal data.
Right to Data Portability	From GDPR, Article 20: The right to receive personal data from a vender, and transfer it to another vender - this helps keep the data subject informed about what data a vender has, as well and to prevent vender lock-in, thereby enabling a data subject to move to a

	new vendor without having to reconstruct her entire history
Right to be Forgotten	From GDPR, Article 17: The right to obtain from the controller the erasure of personal data concerning him or her without undue delay.
Right to Object	From GDPR, Article 21: the right to object, on the groups relating to his or her particular situation, at any time to processing of personal data in certain situations.
Right to Rectification	From GDPR, Article 16: The right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her.
Right to Restriction of Processing	From GDPR Article 18: The right to limit ways in which an organization uses their data/ right to withdraw
Rights Related to Automated Decision-Making	From GDPR, Article 22: The right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
Transparency	Information on the relevant aspects of data collection and data management has to be accessible and provided in a clear, comprehensible, and accessible way (Turilli and Floridi 2009).
Type/Amount of Data Collected	The various kinds and scope of data being collected about a user. This may include health information, economic information, and various kinds of personal information over a period of time.
Unanticipated Use of Data	The use of data in a way that is not outlined by the ethics document or policy – for instance, by an outside organization or for a different research study or use not explicitly stated.
Voluntariness	The user has provided access to his or her data freely and without coercion or undue influence.
Vulnerable Populations	Groups or communities at a higher risk of harm as a result of limitations due to age, mental ability, social, economic, or political status. This often includes children,

	individuals with mental disabilities, prisoners, and other groups facing socio-economic disadvantages (Ruof 2004)

Supplementary Table 2: Informed Consent Guidelines and Policies Examined in Study

These documents represent materials contained in the Ethics Codes Collection by the Center for the Study of Ethics in the Professions at the Illinois Institute of Technology (http://ethicscodescollection.org) that directly deal with digital data management. They do not represent the entirety of guidelines and policies that deal with informed consent and digital data but do provide an illuminating overview of the kinds of documents that exist.

1	Accenture	*Universal Principles of Data Ethics*		Business - Management
2	American Anthropological Association	Principles of Professional Responsibility	2012	Professional Association – Social Sciences
3	Association for Computing Machinery	2018 ACM Code of Ethics and Professional Conduct: Draft 3	2018	Professional Association – Computer Science
4	Association of Internet Researchers	Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee	2019	Professional Association – Information Sciences
5	Association of National Advertisers	Guidelines for Ethical Business Practice	2019	Industry Association - Marketing
6	Canada, Office of the Privacy Commissioner of Canada	Guidelines for Obtaining Meaningful Consent	2002	Government Agency
7	Canadian Marketing Association	Code of Ethics (2004)	2019	Industry Association – Marketing
8	Carlton Connect Initiative, University of Melbourne	*Guidelines for the Ethical Use of Digital Data in Human Research*	2015	NGO/Higher Education- Research
9	Digital Advertising Alliance	Self-Regulatory Principles for Online Behavioral Advertising	2009	Industry Association – Marketing
10	Digital Analytics Association	Web Analyst's Code of Ethics	Undated	Industry Association – Marketing
11	European Commission, European Data	European Data Protection Supervisor Opinion 4/2015: Towards a new digital ethics	2015	Government Agency

	Protection Supervisor			
12	European Commission, Industry Partners	Draft Code of Conduct on Privacy for Mobile Health Applications	2016	Government/Business Partnership
13	Facebook	Data Policy	2018	Business – Social Media
14	Global Alliance for Genomics and Health	Framework for Responsible Sharing of Genomic and Health-related Data	2012	Standards Setting Agency – Health
15	Google	Google Privacy Policy	2020	Business – Social Media
16	Global Privacy Assembly	International Standards on the Protection of Personal Data and Privacy	2009	Professional Association – Information Sciences
17	Human Genome Organization	Statement on Human Genomic Databases	2002	Professional Association – Health
18	Interactive Advertising Bureau	IAB Code of Conduct	2018	Industry Association – Marketing
19	International Committee of the Red Cross	Rules on Personal Data Collection	2020	NGO – Health
20	Mobile Marketing Association	Mobil Marketing Association Global Code of Conduct	2008	Industry Association – Marketing
21	National Research Council Canada	Draft Code of Ethics for Community Informatics Researchers	2007	Government Agency
22	National Information Standards Organization	NISO Privacy Principles	2015	Standard Setting Agency – Information Sciences
23	Organization for Economic Cooperation and Development	OECD Privacy Framework	2013	Government – International
24	Organization for Economic Cooperation and Development	*The Protection of Children Online.*	2012	Government

25	Organization for Economic Cooperation and Development	Guidelines for Human Biobanks and Genetic Research Databases	2009	Government – Intergovernmental Agency
26	Oxfam	*Oxfam Responsible Program Data Policy*	2015	NGO – Social Advocacy
27	PrivacySig	Code of Conduct	Undated	Industry Association - Marketing
28	Twitter	Privacy Policy	2020	Business – Social Media
29	United Nations, Educational, Scientific, and Cultural Organization	International Declaration on Human Genetic Data	2003	Government – Intergovernmental Agency
30	United States, Department of Homeland Security	Menlo Report	2012	Government Agency
31	United States, Federal Trade Commission	Final FTC Privacy Framework and Implementation Recommendations	2012	Government Agency