Key Points for an Ethical Evaluation of Healthcare Big Data

Authors:

Pilar Leon-Sanz

Date Submitted: 2019-09-30

Keywords: health administration and big data, consent forms, confidentiality, data protection, big data and life sciences, biomedical ethics

Abstract:

Background: The article studies specific ethical issues arising from the use of big data in Life Sciences and Healthcare. Methods: Main consensus documents, other studies, and particular cases are analyzed. Results: New concepts that emerged in five key areas for the bioethical debate on big data and health are identified—the accuracy and validity of data and algorithms, questions related to transparency and confidentiality in the use of data; aspects that raise the coding or pseudonymization and the anonymization of data, and also problems derived from the possible individual or group identification; the new ways of obtaining consent for the transfer of personal data; the relationship between big data and the responsibility of professional decision; and the commitment of the Institutions and Public Administrations. Conclusions: Good practices in the management of big data related to Life Sciences and Healthcare depend on respect for the rights of individuals, the improvement that these practices can introduce in assistance to individual patients, the promotion of society's health in general and the advancement of scientific knowledge.

Record Type: Published Article

Submitted To: LAPSE (Living Archive for Process Systems Engineering)

Citation (overall record, always the latest version):	LAPSE:2019.1072
Citation (this specific file, latest version):	LAPSE:2019.1072-1
Citation (this specific file, this version):	LAPSE:2019.1072-1v1

DOI of Published Version: https://doi.org/10.3390/pr7080493

processes

MDPI

Article

Key Points for an Ethical Evaluation of Healthcare Big Data

Pilar Leon-Sanz

Department of Biomedical Humanities, School of Medicine, University of Navarra, 31008 Pamplona, Navarra, Spain; mpleon@unav.es; Tel.: +34-948-425-600

Received: 19 June 2019; Accepted: 28 July 2019; Published: 1 August 2019

Abstract: Background: The article studies specific ethical issues arising from the use of big data in Life Sciences and Healthcare. Methods: Main consensus documents, other studies, and particular cases are analyzed. Results: New concepts that emerged in five key areas for the bioethical debate on big data and health are identified—the accuracy and validity of data and algorithms, questions related to transparency and confidentiality in the use of data; aspects that raise the coding or pseudonymization and the anonymization of data, and also problems derived from the possible individual or group identification; the new ways of obtaining consent for the transfer of personal data; the relationship between big data and the responsibility of professional decision; and the commitment of the Institutions and Public Administrations. Conclusions: Good practices in the management of big data related to Life Sciences and Healthcare depend on respect for the rights of individuals, the improvement that these practices can introduce in assistance to individual patients, the promotion of society's health in general and the advancement of scientific knowledge.

Keywords: biomedical ethics; big data and life sciences; data protection; confidentiality; consent forms; health administration and big data

1. Introduction

"Big data" refers to the processing of large quantities of data, with the aim of discerning patterns and thus gaining novel insights. The volume and variety of data, as well as the velocity with which it is captured, analyzed and interlinked, requires the use of innovative and continuously evolving technological approaches [1] (p. 5). Big data mining applied to the field of health has received critical interest due to two factors: On the one hand, the huge potential it possesses to advance research, biomedical practices, and the promotion of public health; on the other hand, increasing awareness of the issues of vulnerability this implies [2] (p. 304).

Some researchers consider that the analysis of big data would be, in itself, as ethically neutral as any other statistical methodology [3]. However, the use of biological and health big data also introduces heuristic or interpretative algorithms that can be rather uncertain [4] and lead to both beneficial and detrimental effects. For example, in machine learning algorithms procedures, it is difficult to control error or uncertainty, biases may be introduced in the programming process, which affects the ethical consideration they receive. Hence, there is interest in identifying the specific bioethical issues that have arisen from the use of data mining. This multidisciplinary issue affects very different types of professionals. As we will observe, ethics relates specially to the use of the results and the privacy of the data.

Big data has changed the control or precautionary measures established so far. It is necessary to identify the key factors in order to obtain a balance of the benefits and risks that this technology can entail for professionals, patients, and society [5,6].

2. Materials and Methods

Through examples (research or published cases) and based on consensus documents and recent bibliography, key biomedical ethical points are defined and the consequences derived from the use of large databases are analyzed. We observe that ethical valuation arises from the nature of big data.

This is a relatively new issue in the field of biomedical ethics, but there are already important consensus documents approved by agencies and public institutions, such as: *Big Data: Seizing Opportunities, Preserving Values* published by the Executive Office of the President of USA (2014) [7]; *Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues* by the Nuffield Council on Bioethics (2015) [8], *Big Data and Health—Data Sovereignty as the Shaping of Informational Freedom* by the German Ethics Council (2017) [1], also the documents published by the International Medical Informatics Association (IMIA) [9] and the American Medical Informatics Association (AMIA) [10]. In addition, the World Medical Association published in 2016 the *Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks* [11] and we also have the *Report of the International Bioethics Committee of UNESCO on Big Data and Health* passed by the International Bioethics Committee of UNESCO (IBC) (2017) [12].

Moreover, there has been a development on the regulation of Data Protection, especially in Europe, where the General Data Protection Regulation [13] was adopted in May 2016 and fully enforced in the European countries in 25 May 2018. As Haug underlines, "the regulation enshrines in law the principles of protection of privacy and personal data that have been internationally agreed on since 1980" [14]. In the article we will refer to some of the changes introduced by this regulation which involves big data development such as the pseudonymization of data, although a deeper analysis of this European Regulation exceeds the objectives set in this research.

3. Results

Firstly, we examine issues related to the accuracy and validity of the algorithms; secondly, we review the peculiar characteristics of the data that condition the confidentiality aspects of big data and the consent procedures for the use of the data. The new right to "not be profiled"; the debate on objectivity versus subjectivity regarding clinical decisions, and the guarantee function of the State and the institutions will also be discussed.

3.1. Precision and Validity of Algorithms

In 2016, a research project that included 1.4 billion de-identified tax records (between 1999 and 2014) on people between 40 to 76 years of age that evaluated the factors associated with differences in life expectancy of the population was published [15]. The conclusions of the study warned against the non-nuanced use of information obtained from the analysis of big data. In effect, the research coincides with other previous studies in which life expectancy increases with the highest level of wealth. But the new analysis showed that the correlations between average life expectancy, wealth, and lifestyles were more complex than previously thought. That is, the simple analysis of millions of data does not come closer to reality if the appropriate algorithms are not used.

From the example [15], we can see that the fundamental ethical requirements of big data include the technical precision of data analysis, accuracy, and statistical performance [16]. Otherwise, the information obtained may be subject to biases and errors, which would not allow for adequate standards for their application in the health area.

In the referred study [15], we observed that the definition of the objectives of the research influences the ethical consideration of the analysis of large databases. It is also important that possible benefits should justify the deployment of the data. We perceive that this is a requisite for the validity of the analysis, the quality and the adequacy of the group of data included. Additionally, it is necessary to avoid the bias of the person who sets up or adjusts the performance of algorithms applied to the data set or to the objectives of the research.

In addition, care should be taken not to extrapolate the results beyond the scope of the study, considering the limitations derived from the uncertainty, which always exists, in relation to the accuracy of the data and the statistical power of the data analyzed. This aspect acquires more relevance if we consider that the findings are frequently used for new analyses. In any case, disagreements over the information obtained should be addressed before its application to the care of patients or in society.

Transparency and Confidentiality in the Use of Data and Operations

In the field of big data, transparency refers both to the origin of the information as well as to the context and operations carried out because of the complexity of the algorithms applied. Professionals, researchers, citizens, policy makers, and stakeholders must be able to understand the relevance of the data and the implications of operations. Lack of clarity could lead to a lack of confidence. It is essential that the authorities and society in general can rely on the good practices of these analyses for the subsequent use of the results, for example in the area of public health.

The study on life expectancy described in [15] shows that transparency is necessary so that professionals can understand and interpret the meaning of the data, or verify and subsequently apply available data.

3.2. Big Data and Confidentiality

Privacy, the right to confidentiality, and the way of preserving information have been important issues in the implementation of information technology in medicine [17]. Special attention has been paid to the protection of genetic data [18].

The right to confidentiality implies that the personal information provided by a person will not be disclosed later without their authorization, except in the cases established ethically and legally.

In the mining of health data, the right to confidentiality is conditioned by two characteristics. Firstly, by the multiple origins of the data—some come from medical assistance, others from research, from the area of public health, from administrative scope, or as a consequence of the registration of social activities (Table 1).

Table 1. Characteristics about the origin of the data in the big data of Health Care [1].

Diversity of Origin	Type of Records	Ethical Requirements
Private sources	Clinics	Quality of the data
Public sources	Laboratory	Confidentiality and privacy
	Research	Informed consent (transfer of data)
	Public health	Custody of data
	Administrative	Communication of results
	Social/consumer activities	Transparency (origin/contexts/operations)

[1] Own elaboration.

The second characteristic is the reuse or the secondary use of databases (Table 2). The reuse of data lead to new opportunities: It avoids the costs and inconveniences of gathering similar information for different purposes. This would be the case of clinical data that can also be used for the planning of health services, for medical research, or, in the case of insurance companies, for actuarial purposes. But the change of context or purpose can also make the data acquire a very different meaning [19,20]. For example, if security forces had access to clinical databases, the health or disease indicators could become "guilt indicators" [8] (p. 18). Incidental findings might arise that could condition people or their lives. In addition, people often show different sensitivities regarding the use of their data.

Table 2. Secondary uses of data [1].

Benefits	Consequences	Ethical Requirements
New opportunities	Changing the meaning of the data	Adapting the objectives
Reduction of costs	Possible identification of sources	Accuracy of algorithms
Avoiding difficulties for obtaining data	Difficulties for obtaining consent	Social justification and transparency
Allowing the combination of databases	Identification of groups	Confidentiality and privacy
Provides access to data from various areas	Creation of profiles	Consent to the transfer and flow data
	Incidental findings	Avoiding discrimination and stigmatization
		Protection of vulnerable populations
		Guarantees in commercial uses

[1] Source: Own elaboration.

Coding or Pseudonymization and Anonymization of Data

We generally agree that data and information have a personal character if they are linked to a name; that is, if they are identified. Standard practice considers that if the data are in a public domain, are anonymous, or are anonymized, they do not need the approval of the interested parties for their use since the data cannot be related to an individual.

The EU Regulation on data protection has introduced a new concept: "Pseudonymization" that means "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" [13] (article 4, point 5). This new concept is controversial: On the one hand, it is a security technique or a measure to add security and confidentiality to the processing of personal data, but on the other, the pseudonymized data has to be considered as identifiable information [13] point 26. In addition, the European regulation does not define concrete procedures of pseudonymisation, and this regulation does not contemplate previous methods as dissociating or anonymizing.

In any case, coding or pseudonymization facilitates the study of the data. In medical research, it is very common to analyze records of health data, diseases, or treatments from various health centers using this practice. The coding and pseudonymization also allows the communication of the findings obtained through big data mining in case they are useful for particular individuals.

However, in the case of big data, it is difficult to guarantee the confidentiality of a person or a group of people when databases from different sources are combined, even when the data were entered in an encoded or anonymized form [1] (dissenting vote, p. 51).

To illustrate this point: In 2012, an anonymous database provided by the Boston Group Insurance Commission (without names, addresses, social security numbers, or any other type of identifying information) was combined with the database of the voters of the State (which included name, zip code, address, sex, date of birth, because that was in the public domain). After the combination, it was possible to identify specific politicians and citizens [8] (p. 67).

Consequently, to establish the criteria that would safeguard the confidentiality of people, we should assume that the combination of databases can lead to the identification of persons or groups of people, depending on which tools are used and what other information is available [21,22].

Experts say that, in general, the more people involved in a study, the more anonymity is preserved as a combination of the variables will be repeated in several people. But the potential uses and applications of big data may be unpredictable, and several lines of research have been opened in order to avoid possible identifications [22–24].

3.3. The Right "Not to Be Profiled"

The combination of big data can lead to more or less temporary and even permanent associations (Table 2). An example could be the linking of disease record data with the location of environmental contaminants to examine or monitor possible links between them. Thus, as a result of the big data, differentiated personal profiles or organized specific groups may be established. For this reason, it becomes necessary to protect the so-called "group identity" within the scope of the big data. In fact, the Council of Europe has proposed the recognition of the right of "not being profiled" as a new right of persons [12] (p. 64). This right could be important to avoid the discrimination of individuals or groups of people.

We can observe two situations in which we frequently establish profiles or groups of people.

3.3.1. Public Health

One of the most important applications of the big data in medicine is the public health field. Uses of big data allow us to find patterns and correlations between environmental conditions, lifestyles, or social behaviours, and morbi-mortality. These applications facilitate the development of public health policies both in collective and individual settings, since the analyses can determine the predisposition or the risk that a person has to developing a disease.

But those procedures have to respect the freedom of individual action. There is an ethical risk if the measures proposed to achieve a healthier population, could limit the freedom of action and the lives of people by imposing, as mandatory, life styles that, in themselves, are only optional.

In addition, although in most cases big data related to public health use anonymous or anonymized data, multicenter epidemiological studies that include large cohorts of patients or healthy subjects (national and international) are not. Therefore, precautions regarding confidentiality of data or consent to participate in the studies are important.

These precautions acquire special relevance when the results may lead to a possible discrimination or stigmatization of any population group, either by socio-economic level, by risk of developing diseases or by the estimation of a reduced effect of a therapeutic strategy.

From an ethical point of view, the analyses that correlate information on disability, mental illnesses, genetic diseases, sexual orientation, drug addictions, juvenile delinquency, political or religious issues, are especially sensitive [8] (pp. 107–108).

3.3.2. Big Data and Health Management

The use of big data is very convenient for patient safety. Statistical tests allow professionals to establish programs that prevent iatrogenesis and increase the quality of care. The AMIA [10], for example, has repeatedly pointed out that big data reinforces the safety in drug treatments and avoids adverse effects.

In addition, the databases on health activity allow the identification and characterization of care practices and prescription profiles. This information can be used by those who organize and pay for medical assistance, whether public or private systems, regional, state, or international regulatory agencies. Big data analysis can be used to implement effective systems for cost containment, risk management, and safety and quality assurance programs [25].

However, governance criteria should also include ethical paradigms because there are many voices who claim that public and private hospitals and health care organizations such as insurance companies use that information almost exclusively to control costs and to evaluate the practice of the professionals instead of to guarantee the quality of care [26,27].

The findings of big mining can lead to health centers or departments of hospitals encouraging a certain profile of prescription according to standards derived from particular policies or health economic interests [28,29].

3.4. New Ways to Obtain Consent for the Use of the Personal Data

Informed consent has been one of the main guarantees established to protect the autonomy and responsibility of people in medical care and in research [30] (p. 6). However, in the field of big data, informed consent refers not only to the authorization of the use of data, but also to the possibility of deciding on the flow of personal information [5]. We have already reviewed the importance of reusing or combining various data sets.

Below, we list new modes of consent—some explicit, others implicit—for obtaining personal data, from the most explicit to the least.

3.4.1. "Dynamic Consent"

"Dynamic consent" implies a temporary update of the consent for the use of the data so that, as the studies progress, the included persons may re-consent to different uses from the initial one, or allow the use of the data for other purposes. For example, it would be a type of consent applicable when new objectives are proposed in longitudinal studies over time in which measurements of the variables are taken from the same people at different times (cohort study). Dynamic consent allows individuals some control of their own data. Frequently, this consent is made on a virtual or online basis through the platforms that maintain the research projects. As the Nuffield Council Report explains [8] (p. 74), through "continuous" consent, people can collaborate effectively with studies while respecting their autonomy.

3.4.2. "Broad Consent"

A person gives "broad consent" when the information is going to be used in more than one study [12] (p. 12), as long as the projects are related to a specific area or line of research.

In the field of big data, international documents maintain the need to obtain specific consent for data processing, also in the case of secondary uses (WMA, CIOMS) [11]. The secondary analysis of the data is admitted without a new informed consent provided the requirements indicated in Table 3 are met.

Table 3. Ethical requirements for secondary analysis of data, without a new informed consent [1].

Ethical Requirements for "Broad Consent"
Appropriate legal foundation
Evaluation by the Research Ethics Committee
Adequate technical procedures in order to prevent researchers and third parties from accessing personal data, such as pseudonymization
Overriding public interest in this health research
Infeasible to obtain a new consent
Data must have been collected according to ethical and legal requirements

[1] Source: Report of the International Bioethics Committee of UNESCO on Big Data and Health. Paris, 2017, n. 59 [12].

As we observe in the table, "broad consent" is similar to the consent provided in the research with biological samples included in collections. "Broad consent" requires that an IRB or a specific committee guarantees that the rights and interests of the people are adequately protected and that neither the research nor the data used can harm the individual or be used in conditions or for purposes not included in the consent granted.

"Broad consent" is accepted because it is not a general consent. In some way, the need to obtain a specific consent for the processing of the data itself is maintained.

Some publications affirm that many people would prefer to re-consent if personal information were to be used for a second purpose [31]. We have observed that the conditions included in the new regulations such as the EU Regulation on data protection say that "it may be necessary for reasons of public interest to collect data without the consent of the interested party" [32] (p. 54). Of course,

these cases will be subject to appropriate measures that allow the protection of the rights and freedoms of natural persons, as we have observed in Table 3.

3.4.3. The Automated Transfer of Data or the "Opting Out" Approach

This modality implies that if a person does not declare against the use of their data, it is assumed that the data that is recorded as a consequence of computer activity can be used. The case occurs when someone publishes a page on the internet, provides information on social networks, uses a mobile phone application or connects with other users through email [32].

Although the automated data transfer option ("opting out") has become widespread, it does not mean that it is the most suitable way to protect people. This complex debate has not yet been resolved because it requires people to have sufficient training to know the implications of data transfer. Additionally, it can be difficult or almost impossible to unsubscribe from some databases that also have commercial purposes (such as mobile and Internet telephone operators, commercial and consumer behavior databases) [8] (p. 6).

3.4.4. The Voluntary Communication of Data and the Crowdsourcing Phenomenon

Since 2005, the term "crowdsourcing" has been applied to the process by which, voluntarily, people answer requests (mostly online) for information. This process makes personal data available to third parties, including those related to health. There are numerous pages on the internet, as well as networks and applications, developed for this purpose. Crowdsourcing has been used, above all, in the area of public health. For example, in 2013, public health professionals from the Colorado State University, in collaboration with other institutions, created an initiative (in wiki format) to gather information on food production practices and the distribution systems of groceries. This demonstrates a shift in thinking and practice with respect to privacy and public utility of the personal data [33].

3.5. Professional Decision and Data Mining

Best Care at Lower Cost: The Path to Continuously Learning Health Care in America, a report from the US Institute of Medicine [34], highlights the interest of analyzing large sets of health data for medical practice. One of those benefits is the support the data provide for clinical decisions.

In effect, big data provides evidence that facilitates the development of medical guidelines. For example, the Guide redefined arterial hypertension and the therapeutic approach published by the American College of Cardiology and other associations in 2017 [35]. However much experience a professional has, it will always be based on a limited number of cases, while big data provides some "universalization or objectification" of clinical facts, so, paradoxically, the analysis of big data facilitates personalized medicine based on evidence [12] (p. 22). The medical ethical recommendations suggest that this is a question that connects with the debate on subjectivity versus objectivity on which the clinical decision is based.

In any case, statistical methods do not eliminate the responsibility of the professional (in the case cited, to diagnose and indicate, or not, a treatment for arterial hypertension). Nor do they completely eliminate the uncertainty inherent in a clinical decision. Medical reasoning must take into account the values, needs, and priorities of individual patients, which is not a computable ability.

Therefore, the search for computerized algorithms for clinical decision making is an important strategy, but it cannot determine in advance a professional decision. What it does offer are guidelines based on statistical tests which, at times, may not fit the specific cases.

3.6. Responsibility of Institutions and Public Administrations

The International Bioethics Committee of UNESCO (IBC) "considers [that there are] four measures to be crucial for protecting individual rights and fostering public good while recognizing the unavoidable loss of control by individuals about the use of their data in times of Big Data: governance, education, capacity building, and benefit sharing" [12] (pp. 22, 114). Here we will refer to the first two.

3.6.1. Guarantee Function of Institutions and Public Administrations

Governance must promote the adequate control of big data, and also establish the priorities of health policies. Governmental actions should be especially warranted when there are mercantile or economic interests in the uses of the data.

Policy makers must pay special attention to the impact of big data on the most vulnerable populations (minors, modified capacity, and so on), and promote a solidarity that allows the sharing of the advances, infrastructures, and results that favor the common good. In this sense, various government projects and those of public and private scientific companies, through Open Data and Open Science platforms, facilitate the carrying out of new research. As a result, there will be more active participants in big data or in data mining and fewer professionals or populations that are mere "data collectors in international collaborative projects". The correct distribution of the benefits obtained in this way is, in the opinion of the IBC, "the only way to unite development and respect for all" [12] (p. 43).

- Is it an international regulation required?

There are numerous voices that demand more control of big data in the form of laws, regulations, or guidelines that provide social confidence and security, that takes into account both the fast technological development and the diversity of professionals and the actors involved. An international framework for the cross-border flow of data has also been proposed. This proposal seeks to share and exchange information in a safe and ethical manner [36,37]. This would reduce possible harm to people and protect basic human rights (Table 4).

Table 4. Purposes of an International Regulation [1].

Ethical Aims
Regulate data transfer and flow (local, national, international)
Avoid the application of diverse regulations to the same data
Promote a good use of results
Link individual versus collective and transnational interests
Adapt the distribution of results
Avoid inequalities
Protect vulnerable populations
Establish Health policies
Promote ethical criteria of governance
Guarantee good commercial uses

[1] Own elaboration.

3.6.2. Training in Big Data

Big data in Life sciences and Health Care is relatively new and in continuous development, from which derives the ethical institutional imperative to procure training in ethical aspects. It is also necessary to develop guides for good medical practices on the use of big data related to health.

The training of health professionals must be continuous. For example, the NHI has established a Program (BD2K) [38] that aims to train and help biomedical researchers in the proper use of big data. This is the only way to evaluate the designs or the results, and also the "incidental or unexpected findings" of data evaluation.

In the case of patients and subjects included in the corresponding big data projects, the training facilitates the understanding of the meaning and impact of the possible uses of their data.

In the design of educational programs, the intergenerational gap that exists between professionals [12] (p. 10) and the general population must be taken into account.

4. Discussion

This section focuses on three issues that have emerged: The dubious objectivity of the outcomes of big data; the debate between respect for privacy and the public interest of the findings; and the practical difficulties in obtaining consent in the transfer and data flow.

4.1. Ethical Importance of the Technical Design of Big Data

In research, big data and data mining are a powerful generator of hypotheses, particularly for clinical trials, which facilitates the achievement of objectives, provided that they assume the ethical standards established for scientific research [5] (p. 33).

As there is not a single method of data classification, so, as noted in the results (point 3.1 of this article), the algorithm that best suits the aims of the analysis must be identified. Ethically, it is critical that the applied algorithm adapts to the objectives of the analysis that is to be carried out [19,39]. Sue Halpern [40] has drawn attention to this point with a harsh criticism of the results provided by some algorithms applied to data from Facebook and other social networks, so it can be concluded that different types of algorithms involve diverse kinds of ethical challenges.

At this point, the training, experience and intentionality of the analysts are also key, because, as they carry out the study, they may (consciously or not) favor one algorithm over another [8] (p. 91).

4.2. Respect for Privacy versus Public Interest

Big data has modified the consideration of the intersection between public and private interests.

Until now, health professionals know that, in certain cases, public interest conflicts with the right of particular individuals. But, as stated in the UNESCO declaration [12] (p. 66), the relationship between privacy and public interest is not simply one of opposition. Both rights are mutually implied—there are private interests in the achievement of common objectives and a public interest in the protection of privacy that fosters cooperation between these two spheres. This complex relationship leads to the need to reconcile, in the field of big data, the articulation of the private within the public and the public within the private.

In the case of the analysis of the large data bases, the collecting, storing, processing and analyzing the data over and over again (the so-called cybernetic loop), can lead to intrusions in the private sphere, for example, through the sending of information, also with commercial or advertising purposes. Experts point out that privacy decreases with the number of variables analyzed, but increases with a bigger number of cases.

However, surveys and field studies show that the opinion of some sectors of society about the use of personal data has changed if there is a public interest for them. The new tendency of public opinion is to prioritize collective interest over the individual right to privacy. Thus, for example, when citizens of Western Australia were consulted about the use of, for purposes of health policy and management, and other researches, a database of more than three decades which included all types of personal records related to the morbi-mortality of the population, the majority response was to support the project, and people questioned why the data had not yet been used [41]. The report of the Nuffield Council, for its part, describes similar examples in Europe and the United Kingdom [8] (pp. 132–133).

4.3. Difficulties of Electronic Procedures for Obtaining Consent

Consent must be an explicit act, even if it is requested by electronic means, and requires sufficient information about what it is being consented to. As the European Regulation states, "consent should not be considered as granted freely if the interested party has no genuine or free choice or cannot reject or withdraw the consent without detriment" [13].

However, in the analysis of large databases it may be difficult for a non-expert person to understand the technical possibilities of the information, which might condition the consent. As mentioned above, the data can come from very different sources and there are countless technical possibilities that

have opened the secondary use of medical data, which can make it difficult or impossible to grant the corresponding informed consent or guarantee rights of access, rectification, cancellation, tracing, or opposition to the use of personal data. Nor can the way in which data can be removed from big data projects be easily established.

Surveys through computer media and social networks for gleaning health-related data can be very efficient, but in these cases, the consent can involve important challenges since they require measures for respecting the identity and autonomy of the participants.

The informed consent that is given electronically also raises doubts about the difficulty, or in some cases the impossibility, of identifying the user who is completing it and therefore the capacity, competence, and voluntariness of the person granting the consent. It is difficult to verify if the person completing the data is a person with modified capacity or a minor. And, in these cases, the consent would not be valid.

Further, consent documents are often long and, sometimes, difficult for most people to understand. The IBC considers that it is doubtful that the "one click, accept all" procedure can be really a free and informed consent [12] (p. 56).

5. Conclusions

Big data is more than a large number of data sources. Society recognizes the enormous potential of big data and of data mining to develop more efficient health services, to improve research, and to guide local, national and global health policies.

The bioethical perspective on big data underlines the impact of the objectives, the context, and the purposes of the information analysis. The article has shown the importance of guaranteeing:

1. The quality, accuracy, adequacy, and validity of the data and algorithms, questions related to transparency and confidence in the use of the data;
2. The need for adequate information for professionals, researchers, citizens, and policy makers to understand the implications of big data;
3. The requirement to respect the privacy of individuals and of groups of people with similar profiles, because the possibility of discrimination derived from the information obtained through big data should be avoided;
4. The consent of the persons to the cession of data and to the flow of the information, as well as good practices regarding the ways to obtain them;
5. The responsibility of health professionals in general, of researchers, of the managers, and computer specialists in their professional performance has been emphasized;
6. That Institutions and Public Administrations have to support the development of big data, taking into account equity and solidarity to avoid inequalities or discrimination, especially of vulnerable persons or groups.

Analysis of large databases is a complex social practice, where ethical tensions and possible conflicts of interest exist. Spaces for ethical reflection should be built to account for their use, to resolve possible contradictions, and promote procedures based on the search for good—not merely acceptable—solutions.

Ultimately, good uses and good practices will be a function of the improvements that the analysis of large masses of data can introduce to the care of the individual patient and the health of society in general.

Funding: This research received no external funding.

Acknowledgments: Earlier versions of this text have been discussed with Javier Carnicero Giménez de Azcárate. I am very grateful for his comments and suggestions as well as for the comments and suggestions of the anonymous reviewers.

Conflicts of Interest: The author declares no conflict of interest.

References

1. German Ethics Council. Big Data and Health–Data Sovereignty as the Shaping of Informational Freedom 2017. Available online: http://www.ethikrat.org (accessed on 21 July 2019).
2. Mittelstadt, B.D.; Floridi, L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Law Gov. Technol. Ser.* **2016**, *29*, 445–480.
3. Seltzer, W. The promise and pitfalls of data mining: Ethical issues. In Proceedings of the American Statistical Association, Section on Government Statistics, American Statistical Association, Alexandria, VA, USA, 2005; pp. 1441–1445.
4. Galeano, P.; Peña, D. Data science, big data and statistics. *TEST* **2019**, *28*, 289–329. [CrossRef]
5. León, P. Bioética y Explotación de Grandes Conjuntos de Datos. In *La Explotación de Datos de Salud: Retos, Oportunidades y límites*; Sociedad Española de Informática de la Salud: Pamplona, Spain, 2016; pp. 25–41.
6. Marckmann, G.; Goodman, K.W. Introduction: Ethics of Information Technology in Health Care. *Int. Rev. Inf. Ethics* **2006**, *5*, 2–5.
7. Executive Office of the President. *Big Data: Seizing Opportunities*; Preserving Values: Washington, DC, USA, 2014.
8. Nuffield Council on Bioethics. Linking and Use of Data in Biomedical Research and Health care: Ethical Issues. 2015. Available online: http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf (accessed on 21 May 2019).
9. International Medical Informatics Association. Code of Ethics for Health lnformation Professionals. 31 January 2011. Available online: http://www.imia-medinfo.org/new2/node/39 (accessed on 21 May 2019).
10. American Medical Informatics Association. Code of Professional and Ethical Conduct; Principles of professional and ethical conduct for AMIA members. November, 2011. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 141–143.
11. World Medical Association. Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. 2016. Available online: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/ (accessed on 21 May 2019).
12. UNESCO. *Report of the International Bioethics Committee of UNESCO on Big Data and Health*; UNESCO: Paris, France, 2017; Available online: https://unesdoc.unesco.org/ark:/48223/pf0000248724 (accessed on 21 July 2019).
13. EU. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. *Off. J. Eur. Union* **2016**, *119*, 1–88.
14. Haug, C.J. Turning the Tables—The New European General Data Protection Regulation. *N. Engl. J. Med.* **2018**, *379*, 207–209. [CrossRef] [PubMed]
15. Chetty, R.; Stepner, M.; Abraham, S.; MPhil, S.L.; Scuderi, B.; Turner, N.; Bergeron, A.; Cutler, D. The Association between Income and Life Expectancy in the United States, 2001–2014. *JAMA* **2016**, *315*, 1750–1766. [CrossRef] [PubMed]
16. Tavani, H.T. *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*, 4th ed.; John Wiley: Hoboken, NJ, USA, 2013.
17. León Sanz, P. Aspectos éticos de la seguridad de la información en los entornos sanitarios. In *Seguridad de la Información en Entornos Sanitarios*; Carnicero Giménez de Azcárate, J., Ed.; Sociedad Española de Informática Sanitaria y Navarra de Gestión para la Administración: Pamplona, Spain, 2008; pp. 25–42.
18. EU. Regulation 2012/0011 of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data. Available online: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT%20TA%20P7-TA-2014-0212%200%20DOC%20XML%20V0//EN (accessed on 21 May 2019).
19. Goodman, K.W. *Ethics, Medicine, and Information Technology: Intelligent Machines and the Transformation of Health Care*; Cambridge University Press: Cambridge, UK, 2015.
20. A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data. 2014. Available online: http://www.research.ed.ac.uk/portal/en/publications/a-review-of-evidence-relating-to-harm-resulting-from-uses-of-health-and-biomedical-data(d11d0bb4-7003-4558-8391-0567007e9258).html (accessed on 21 May 2019).

21. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **2002**, *10*, 557–570. [CrossRef]
22. Wel, L.; van Royakkers, L. Ethical issues in web data mining. *Ethics Inf. Technol.* **2004**, *6*, 129–140.
23. Aggarwal; Charu, C.; Yu, P.S. *Privacy-Preserving Data Mining. Models and Algorithms*; Springer: Boston, MA, USA, 2008.
24. Ohm, P. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Rev.* **2009**, *57*, 1701–1777.
25. Al-Saggaf, Y. The Use of Data Mining by Private Health Insurance Companies and Customers' Privacy. *Camb. Q. Health Ethic* **2015**, *24*, 281–292. [CrossRef] [PubMed]
26. Anderson, J.G.; Goodman, K. *Ethics and Information Technology: A Case-Based Approach to a Health Care System in Transition*; Springer: New York, NY, USA, 2002.
27. Lin, K.C.; Yeh, C.L.; Huang, S.Y. Use of Data Mining Techniques to Detect Medical Fraud in Health Insurance. *Int. J. Eng. Technol. Innov.* **2012**, *2*, 126–137. [CrossRef]
28. Kaplan, B. Selling health data: De-identification, privacy, and speech. *Camb. Q. Healthc. Ethics* **2015**, *24*, 256–271. [CrossRef] [PubMed]
29. Orentlicher, D. Prescription Data Mining and the Protection of Patients' Interests. *J. Law Med. Ethic* **2010**, *38*, 74–84. [CrossRef] [PubMed]
30. The Universal Declaration on Bioethics and Human Rights. *Int. Soc. Sci. J.* **2005**, *57*, 745–753. [CrossRef]
31. Willison, D.J.; Keshavjee, K.; Nair, K.; Goldsmith, C.; Holbrook, A.M. Patients' consent preferences for research uses of information in electronic medical records: Interview and survey data. *BMJ* **2003**, *326*, 373. [CrossRef] [PubMed]
32. Al-Saggaf, Y.; Islam, M.Z. Data Mining and Privacy of Social Network Sites' Users: Implications of the Data Mining Problem. *Sci. Eng. Ethics* **2015**, *21*, 941–966. [CrossRef] [PubMed]
33. Wicks, P.; Pickard, T.; Francke, U.; Swan, M. Crowdsourced Health Research Studies: An Important Emerging Complement to Clinical Trials in the Public Health Research Ecosystem. *J. Med. Internet Res.* **2012**, *14*, e46.
34. Committee on the Learning Health Care System in America; Institute of Medicine; Smith, M.; Saunders, R.; Stuckhardt, L.; McGinnis, J.M. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*; National Academies Press: Washington, DC, USA, 2013.
35. Whelton, P.K.; Carey, R.M.; Aronow, W.S.; Collins, K.J.; Himmelfarb, C.D.; DePalma, S.M.; Gidding, S.; Jamerson, K.A.; Jones, D.W.; MacLaughlin, E.J.; et al. Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary. *J. Am. Coll. Cardiol.* **2017**, *24429*. [CrossRef]
36. Organisation for Economic Co-operation and Development (OECD), Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. 2013. Available online: http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm (accessed on 21 May 2019).
37. Verschuuren, M.; Badeyan, G.; Carnicero, J.; Gissler, M.; Pace Asciak, R.; Sakkeus, L.; Stenbeck, M.; Deville, W. The European data protection legislation and its consequences for public health monitoring: A plea for action. *Eur. J. Public Health* **2008**, *18*, 550–551. [CrossRef]
38. National Institutes of Health, Data Science. About BD2K. Available online: https://datascience.nih.gov/bd2k/ (accessed on 21 May 2019).
39. Anderson, J.G.; Aydin, C.E. Evaluating the Impact of Health Care Information Systems. *Int. J. Technol. Assess. Health Care* **1997**, *13*, 380–393. [CrossRef]
40. Halpern, S. *They Have, Right Now, Another You*; The New York Review of Books; Hederman: New York, NY, USA, 2016.
41. Goodman, K.W.; Meslin, E.M. Ethics, Information Technology and Public Health: Duties and Challenges in Computational Epidemiology. In *Public Health Informatics and Information Systems*; Magnuson, J.A., Fu, P.C., Eds.; Springer: London, UK, 2014; pp. 191–209.