

ETHICAL ISSUES IN MANAGEMENT OF BIG DATA IN LIBRARIES AND OTHER INFORMATION CENTERS

Abubakar, Fati (CLN)¹, Atehiran, Folashade (CLN), Musa Hussaini PhD² & Musa Imam Shakarau³
fati.abu@futminna.edu.ng¹; osinafadeke@gmail.com²; hussaini.musa@futminna.edu.ng³
musa.imam@futminna.edu.ng⁴

Abstract

This paper addresses some ethical issues that arise from the management of big data in libraries. The concept of big data was explained. The benefits to society have brought significant improvement in the processing of information for research and policy use. Some challenges faced by researchers and policymakers when big data is utilised in libraries and information centers include user privacy and information security issues. Some of the roles of librarians and information professionals in the big data era include: data ethicists, data advisors/consultants –and use at reference desks in points of need. Components of ethical competence for library and information practitioners are ethical sensitivity, ethical knowledge, and ethical behaviour. Some of the competencies for data ethics identified are: aware of big data trends and developments, Knowledge of key ethical principles, codes and policies, Moral sensitivity, judgment, and character. Guided reading, listening, and viewing are some of the education expected of data ethicists. Recognising an ethical issue, and getting the facts makes the framework for ethical decision-making. The paper concludes that the establishment of scientific and effective network ethics for network, ethical construction plays a very important role in the era of big data.

INTRODUCTION

Big Data has become a popular parlance in the contemporary world. Its popularity is attributed to the increasing advancement in technology and the establishment of various databases meant for various research and policy actions. While its contribution to the growth and development of society is unquestionable, it must be noted that it is not without its ethical challenges which may impact the quality and reliability of the data itself if such ethical issues are not carefully addressed. Thus, this presentation aims to outline and explain the ethical issues that arise from the management of big data in libraries and information centres to ensure that big data are managed professionally to achieve the desired result.

The Concept of Big Data

While its origin can be traced to the 1960s and 70s, big data have become the popular parlance and has, in recent times, received remarkable scholarly attention. Such attention is borne out of the need to address the increasing usage of big data and its attendant ethical issues which may affect the quality of the data being utilized. Bonide (2022) explains that big data is largely differentiated from other data from its volume, variety, purpose or intent to use, veracity, and usage. It implies that big data is essentially different from the common forms of data that exist within the realms of non-library settings. Thus, as the name implies, big data is any data whose volume is large and is sourced from various origins for multiple uses. According to Nair (2020), big data essentially is made up of four Vs. That is the volume, variety, velocity, and veracity.

Floridi (2016) Ying (2015), Blumenthal and Kenton (2018) identified them as follows:

User Privacy and Information Security Issues

Privacy refers to individuals' ability to restrict and control how organizations use and disclose their personal information. Three aspects of privacy that underlie ethical issues in this context include controlling how organizations can access, modify, and use personal data. First, even when individuals give consent for organizations to collect and share their data, individuals need to be able to control what data organizations collect and aggregate about them and who will have access to their aggregated data. While the separate databases that contain an individual's data might be anonymous, the aggregation process might re-identify the individual and make the data available to other parties without the individual's knowledge. Second, individuals need to be able to modify data about themselves, which includes updating or deleting data to remedy incorrect, incomplete, or out-of-date data. In particular, individuals require the ability to modify aggregated or shared data that might misrepresent them. Third, individuals need to influence how organizations use data about themselves. Although individuals may consent for the primary organization to use their data, they need to audit how and for what purposes other parties who have access to it will further exploit it. Moreover, big data analytics lead to organizations' creating and sharing new knowledge about individuals. The new knowledge from aggregated data might reveal sensitive and unwanted information about individuals, create discomfort for them, and possibly have unintended consequences such as discrimination.

The volume describes the quantum of data involved. Typically, this may range from tens of terabytes to hundreds of petabytes. The bottom line is that the quantity of data in question is large. The variety implies that the data is heterogeneous in its source and form. Data comes from various sources such as interviews, opinion polls, product usage, service utilization, etc. Again, these data come in various forms and media such as texts, audio files, videos, etc. They can also be structured, semi-structured, or unstructured. This points to the different varieties in which big data may be presented. The velocity connotes the speed with which the data in question are gathered and stored for retrieval and usage. For veracity, it emphasizes the validity and reliability of the data at hand. With the increase in technological advancement, large volumes of data are accessed, processed and stored within seconds.

The benefits of big data to society are numerous and have brought a significant improvement in the processing of information for research and policy uses, however, it is worthy of mention that the digitalization of data has its risks and harbors certain challenges which can affect the quality of the data itself. These challenges raise ethical concerns and such concerns need to be addressed if the big data is to be utilized by researchers and policy makers. Thus, these ethical issues are discussed below.

Ethical Issues in Management of the Big Data in Library and Information Centres

There are some major ethical concerns raised in the management of big data in the library and information centres Mitteltadt and

Privacy right is the basic right of citizens, which is protected by the Constitution and other relevant laws. No one can interfere with the personal life of citizens. The development of network information services, privacy and security of personal information are faced with some new threats: firstly, in the network environment, the user's personal information, book lending, literature search, consulting, website access, etc., can be stored in the corresponding network database system. Technically speaking, there is no information on the Internet one hundred percent technical support. Any information, as long as the Internet, is possible to get to other people from the network. Secondly, the library in the development of network services may not be conscious of personal information rights violations. To regulate the management of books, the library may neglect the protection of the privacy of the readers, for example, they put the readers' information on the school's electronic bulletin board for those who haven't returned the book in time, or pass the student's information to the school board because they visited unhealthy websites. The influence on the students of this approach could be easily imagined. If the school library took a student's view and dealt with these cases flexibly, satisfactory results would be believed to be obtained.

Equal and Free Access to Information

Traditionally based on the extension of the books and periodical resources limited or the status hierarchy system, library for conditions such as the nature of the institution, the user level, the title of the user site on Library copies, borrowing period made some different rules, which is actually an expression of the inequality. In the network information service, electronic

books, electronic journals, etc., can be used for a lot of people online at the same time, but the use of network information services relying on computer and network facilities. As of now, even in the developed countries, the population of computers and network facilities is still a minority group, the distribution of computers and networks with income, education, gender, race and other factors show great differences. Therefore, the network information service in the solution of some inequalities, but also triggered a new industry and regional inequality. This will further expand the existing digital divide between industries, between regions, between people and people. IFLA 2002 issued by the IFLA Internet declaration "pointed out:" Library and information services to the public to provide information on the Internet accessibility, support groups and individuals are free to knowledge acquisition, and the development of the cause of success. Barriers to the flow of information must be cleared to provide access to the Internet's basic approach to help people overcome barriers to resources and technology (Yu & Breivold 2008). From the overall perspective, the network information service has brought the user freedom and liberation, but there are some obstacles in the specific use. For example, some network resource data format is not uniform. In the user access to network information services, some of the user interface design is not easy to use, and access to services is also more complicated. Some libraries restrict the user to access to certain information levels by setting up an account, and some libraries are charged with the use of network information service facilities, which affects the user's free use of information.

Abuse of Network Technology

In the era of big data, data mining, sorting, collection and analysis are mainly based on the Internet. The development and application of Internet technology at the same time also have some potential safety problems. Some lack of ethical and moral people in the use of the Internet, cannot correctly and effectively use it, in the course of the application of some similarity to the phenomenon of network crime. These network crimes are mainly manifested in the network fraud, manufacturing computer viruses and other issues. This makes a number of network experts use the Internet technology to dig and collect other people's information online for fraud, violations of people's personal privacy, and do some other immoral things. These problems are derived from the current Internet technology development. At the same time, with the application of Internet technology, it is also a certain degree of indirect impact on the traditional ethics, making people's understanding of the Internet lack of reason, their thinking modes on the emergence of some changes, and making people desire to get more psychological earnings which may trigger more network crimes, and harm the society.

Ethical Governance

While formal governance concerns formal policies, standards, and accountabilities about data, informal governance concerns culture and is determined by what the library and information centres actors believe and do based on their values, norms, and shared beliefs. Aspects of ethical governance that can create ethical issues include building ethical norms, establishing rules and procedures, and internalizing the costs. Unethical practices can become accepted and

legitimized in an organization's culture (even despite formal policies). To prevent this problem, organizations need to be vigilant in data governance and governments must impose sanctions against unethical practices. Education and training is a means to build an appropriate set of shared norms, values, and beliefs and, ultimately, influence the actions of organizational actors with regard to data practices. Library and information centres need to establish new rules and procedures to regulate and reinforce appropriate employee behaviour. These rules can make data flows more transparent for customers and other organizations. Where organizations share data, inappropriate use of the data by one organization in the chain can have negative effects for all the other organizations in the chain. These rules can also cover policies for data security where data is subject to conflicting laws and regulations in different locations. Organizations should balance the costs and benefits of big data analytics between individuals and organizations. Big data analytics can generate huge financial and market benefits for organizations, which might mean they extensively monitor and measure individuals' behaviours and manipulate their choices and behaviours. Currently, organizations mostly focus on generating value for themselves rather than internalizing some of the costs associated with big data analytics.

Awareness

Awareness concerns what individuals know and understand about big data analytics practices, such as how Library and information centres analyze their data to offer products and services. Ethical issues arise when individuals lack awareness about why Library and information centres use and the processes involved in big data analytics.

Aspects of awareness that underlie ethical issues in big data analytics include understanding what big data analytics is, understanding rights regarding big data analytics, and understanding who holds the data and for what purpose. First, individuals need to learn about big data analytics, how it operates, and how it influences their choices and behaviour. Individuals need to engage in public data literacy programs and recognize appropriate uses and consequences of big data analytics. By doing so, they can learn to better balance the personal costs and benefits of big data analytics. Second, individuals need to recognize policies, regulations, and laws that exist to protect them from the potential negative consequences of big data analytics (e.g., the European General Data Protection Regulation (GDPR)³). Subsequently, they can engage with governments and influence how regulations about big data analytics. Third, individuals need to know what data organizations collect about them, who owns and controls this data, and which third parties have access to it. Library and information centres often collect big data implicitly without clear informed consent, and they frequently hide secondary uses of the data from individuals. Terms and conditions are obscure, and opting out from them can be difficult. Individuals need to recognize these practices since the analytics conducted on their data will ultimately influence their lives.

Choice

Big data analytics can restrict individuals' choices. As a result, it can discriminate against individuals and unfairly manipulate their behaviour. Aspects of choice that underlie ethical issues in big data analytics include the limiting of individuals' choices, incorrect analytics, and gamification. First,

big data analytics may limit individuals' choices based on their past behaviour, location, age, gender, and so on. Library and information centres profile and categorize individuals according to their personal data and then send them services and products based on the resulting profile. As a result, individuals may lose freedom of choice and face a less-than-free market. Second, the digital profiles that Library and information centres create using aggregated data may not correctly represent individuals. This may arise due to low-quality aggregated data or inappropriate algorithms, which can yield an incorrect profile and predictions. Further, Library and information centres may still target and unfairly discriminate against individuals based on such data/algorithms. Third, organizations may gamify individuals to further analyze and manipulate their behaviour, such as by personalizing rewards until a certain customer shows certain behaviour.

Data Quality

Big data quality underlies the correctness of decisions that organizations make using big data analytics. Although data quality is well defined for conventional data systems, it does not apply to the context of ethical big data analytics. Specifically, aspects of data quality that can cause ethical issues in big data analytics include data quality criteria for big data, quality of aggregated data, and creation and maintenance of metadata. First, big data often comprises complex social data that organizations source in multiple formats and usually has an unstructured form. Although structured data has quality dimensions including accuracy, timeliness, and completeness, it does not apply in the context of big data, so that adequate managerial guidelines are not available.

Second, organizations combine and aggregate data about individuals from multiple sources. Aggregated data might reveal information about individuals who otherwise had anonymity at the initial data-collection point (called the mosaic effect). We know little as to whether or not the aggregated data about individuals accurately represents them. Third, when organizations source big data, they may not establish data definitions and not capture and maintain metadata information. Individuals might have contributed their data in different contexts for different reasons, particularly in the case of social media. The data definition problem is exacerbated by the fact that the data's meaning might change over time through sharing and aggregation processes, which can create even greater potential for unethical data use.

Algorithmic Decision Making

Decision making concerns the processes and outcomes of decisions that Library and information centres make using big data analytics. Decisions made using big data analytics typically rely on complex statistical and computational methods. Aspects of decision making that underlie ethical issues in big data analytics include reliability of algorithms, lack of human involvement, and accountability of decisions. First, as the data increases in size, speed, and complexity, algorithms become more important in making sense of data, generating insights, and predicting the future. Library and information centres use algorithms to predict the future based on historical and subjective data, and, in most cases, they predict based on correlations only (as opposed to establishing a causal effect). Library and information centres have no means to ensure that they have made an ethically appropriate

decision about an individual. For example, an algorithm may inadvertently lead to racial profiling or some other act of discrimination because no theory exists to explain the relationships in the data since big data analytics relies mainly on inducing insights. Second, decisions made using big data analytics are either automated, have no human involvement, or are visualized for the human decision maker. Visualizations typically convey a particular message or story and hide underlying assumptions, limitations, biases, and data-quality issues. The human decision maker has no means to understand how, or against what criteria, the decision has been made, which can limit the decision maker's ability to properly interpret the results. Third, when individuals make decisions using complex algorithms that humans find difficult or impossible to understand, the responsibility for decision outcomes becomes blurred. Such a situation could be problematic if, for example, poor-quality data or an unsuitable algorithm led to an unethical and discriminatory decision.

Roles for Librarians and Information Professionals as Data Ethicists

- **Data advisor/consultant** – dealing with queries about data access, privacy/protection, sharing, and use at reference desks or point of need
- **Data literacy educators** – extending information literacy instruction to data use in diverse contexts: learning, the workplace, and everyday life
- **Interdisciplinary facilitators** – helping multidisciplinary research teams negotiate unfamiliar information resources and intellectual norms

• **Policy developers** – contributing to or leading on formulation of data management policy and production of guidelines on implementation

• **Research ethics partners** – supporting ethics and RCR/RDM training, developing online resources, and serving on institutional review boards

• **Cross-cultural mediators** – anticipating cultural differences and alerting stakeholders to issues arising from plural and conflicting values – an ethical responsibility to engage!

components of ethical competence for library and information practitioners

Ethical Sensitivity

Ethical sensitivity is an “ability to recognize an ethical problem or ethical aspects of situations.” ethical sensitivity is a key element facilitating decision-making in library and information practices. Ethical sensitivity is therefore to be used to designate the ability to recognize areas of ethical tension that can emerge in different situations, especially in relation to others. It is a source of reflection, decision making, and responsible action by librarian and information professionals.

Ethical Knowledge

Ethical knowledge can be defined as knowledge based on a combination of philosophical knowledge, theoretical knowledge (normative, deontological, scientific), and practical knowledge while taking into account the contexts and people involved. These different types of knowledge will be mobilized in the ethical competence process. Librarians and information professionals need to extend their ethical

knowledge beyond normative and deontological aspects.

Ethical Reflection

Ethical reflection appears as an iterative reflective process requiring that various alternatives to ethical problems be considered. Undeniably, this process will be affected by the beliefs and values put forward by stakeholders, including those of the nurse involved, and it is important to keep them updated in light of their influence. Ethical reflection is “to reflect critically about what the librarians and information professionals know, are, and do (ethical reflection).”

Ethical Decision-Making

Ethical decision-making can be defined as a decision-making process aimed at making a reasonable and responsible choice from among a number of alternatives. Ethical choice then takes shape through ethical action

Ethical Action

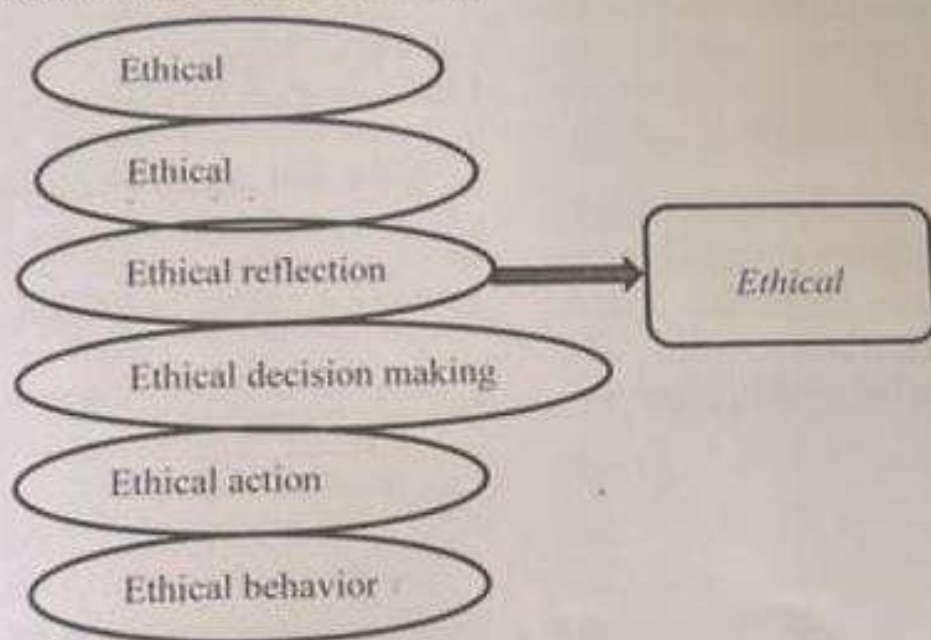
Ethical action may be defined as action that is the product of critical reflection and analysis motivated by the desire to assist and accompany a person along his or her path. It is thus a specific form of action, adapted to a particular situation and context, which differentiates it from ethical behaviour. Ethical action is the result of reflection, analysis, and judgment. Such action is based on an in-depth understanding of individuals and their context, as well as a consideration of their vulnerability. Thus, ethical action involves effecting an action that was chosen following identification of ethical tension, a reflection based on acquired knowledge, and consideration of a variety of alternatives and their implications with a view to providing care and promoting individual well-being

Ethical Behaviour

Ethical behaviour is the reflected in library or information professionals know-how and personal skills. "ethical comportment is used to refer to the embodied, skilled know-how of relating to others in ways that are respectful, responsive, and supportive of their concerns."

Ethical behaviour is characterized by an attitude of moderation and respect for others, which reflects the ethical reflection on which it is based.

In the diagram, the distinction is clear between the acknowledgement of a problem,



Components of ethical competence

(Lechasseur et al., 2018)

Competencies for Data Ethicists

- Awareness of big data trends and developments

- including local initiatives, programs and projects

- Knowledge of key ethical principles, codes and policies

the mobilization of various internal and external resources, and taking action. Furthermore, critical reflection underlies a contextualized mobilization of knowledge.

Ethical sensitivity, ethical knowledge, ethical reflection, ethical decision-making, ethical behaviour, and ethical action are therefore all components of ethical competence and resources to be mobilized. Their identification and linkage allow for an understanding of how ethical competence is developed by library and information professional's.

- both professional (library and information) ethics and research ethics (including Internet/online/virtual research ethics)

- Moral sensitivity, judgment, motivation, and character

- the ability to recognize ethical tensions, to reflect/think critically and make informed decisions, to understand professional obligations and responsibilities, and to act and behave with respect for others

- Skills in critical thinking, reflection, analysis, and advocacy – already part of the information specialist's professional skillset

Generic and contextually situated know-

Education for Data Ethicists

- Guided reading, listening, and viewing
 - Blog posts, Books, Journal articles, Magazine features, News stories, Videos...

- Case studies/reports/materials

- Original or published sets of paradigmatic cases for analysis and discussion, with questions and/or commentaries to encourage critical reflection

- Formal debates (with voting)

- Divide students into groups and teams, provide a debate motion, assign to particular sides (proposing or opposing), supply instructions/resources, and ask team members to develop arguments supporting the views assigned

- Onsite/online (synchronous web conference or asynchronous discussion board)

- Role plays (and observation)

- Ask students to play a character in a case study portraying a realistic, difficult ethical situation that illustrates divergent perspectives on the same problem

- Provide descriptions of the situation and script/leading lines to get them started

- Immersive experiences

- Students spend a day observing researchers working with data in a laboratory

Framework for Ethical Decision Making

Recognize an Ethical Issue

1. Could this decision or situation be damaging to someone or to some group?

2. Does this decision involve a choice between a good and bad alternative, or perhaps between two "goods" or between two "bads"?

Get the Facts

3. What are the relevant facts of the case? What facts are not known?

Can I learn more about the situation? Do I know enough to make a decision?

4. What individuals and groups have an important stake in the outcome?

Are some concerns more important? Why?

5. What are the options for acting? Have all the relevant persons and groups been consulted? Have I identified creative options?

Evaluate Alternative Actions

6. Evaluate the options by asking the following questions:

- Which option will produce the most good and do the least harm?

(The Utilitarian Approach)

- Which option best respects the rights of all who have a stake? (The Rights Approach)

- Which option treats people equally or proportionately? (The Justice Approach)