KNOWLEDGE AND ATTITUDE OF INDIVIDUALS TO PRIVACY ISSUES OF OPEN DATA: AN EXPLORATORY STUDY

MEJABI, Omenogo Veronica, AWONIYI, Abisola Azeezat, OYEKUNLE, Rafiat Ajibade & AZEEZ, Adesina Lukman
Department of Information and Communication Science
Faculty of Communication and Information
University of Ilorin,

Abstract
The increase in opening up data to the public is widely advocated but raises concerns about data privacy. Although attempts are made to anonymise data stored in databases, algorithms that enable re-aggregation of data mean that personal data can be traced back to the individual owner leading to serious violations to privacy occurring quite often. This study explores the knowledge and attitude of individuals towards open data and associated privacy issues using students as the data subjects. Students from two faculties representing individuals with legal knowledge and those with information technology knowledge are used as proxy in the study with 160 respondents. Data was collected using a questionnaire with sections on attitude to open data, knowledge of data privacy and privacy implementation techniques. It was found that 50% respondents had previous knowledge of open data. The study revealed that credit card details is the dataset mostly considered personal and of high risk to privacy if released in the open domain. However, respondents are most willing to release their pictures for research purposes and on social media. Anonymity and legal policies are the most preferred data security techniques. The study concludes that respondents' awareness and knowledge of open data and personal identifiable information (PII) is still low and recommends continuous education of data subjects to privacy compliance of open data with respect to personal identifiable information (PII) in the collection, processing, storage and opening of data.

Keywords: Open Data, Data Privacy, Data Protection, Personal Data, Personal Identifiable Information (PII)

Introduction

Publishing data on the web has become a cost-effective channel of disseminating or sharing data and information which is utilized by individuals and organizations such as government agencies, corporate entities, schools, and non-governmental organizations (NGOs). The term 'open data' has been used to describe such data on the web although more stringent definitions of open data include: data that can be freely used, re-used and re-distributed by anyone, subject only, at most, to the requirement to attribute and share-like (Open Data Handbook, 2012). Apart from information dissemination, reasons for publishing open data include the promotion of transparency and accountability, realizing social and commercial value from existing data, and encouraging participation and engagement by citizens.

Open data, is usually packaged in a manner that ensures that no personal identifiable information (PII) is published. So it seems that opening data goes hand-in-hand with closing up privacy and getting the balance right between these two conflicting priorities is an important challenge for both government and private sector enterprises (Wainewrite, 2014). This is especially important when such data are based on information about social relationships and medical history. However, Narayanan and Shmatikov (2010) noted that re-identification is possible without PII such that any information that distinguishes one person from another can be used for re-identifying data.

Ensuring the privacy of individuals when data is opened is of concern when organizations sell information pertaining to individuals to third parties as in the case of the UK National Health Service (NHS) which sold patients medical records to management consultants who uploaded it to Google servers based outside the United Kingdom. This raises the question of whether or not this is a violation of the privacy of individuals not to mention other risks such as a situation where the life insurance fee

of a particular family increases because their medical record (traced back to the family through the PII) reveals unfavourable information that would otherwise not have been available to the health insurance company. Another type of data being embraced by governments in different countries for publishing as open data is criminal information in form of crime records or case records. This can have a tremendous effect on a potential employee who has a previous conviction record. Also, neighbourhoods in which crime rate is high might discourage potential investors when cross-referenced on an open map. Thus, in general, personal data is now being claimed, processed, exchanged and analysed at a global level (Casado, 2014).

The US Department of Homeland Security (2012) identifies name, email, home address and phone number as PII. Furthermore, Social security number (SSN), Driver's license or state ID number, Passport number, Alien Registration Number, Financial account number, and Biometric identifiers are classified as sensitive PII if they stand-alone. Or if paired with another identifier, Citizenship or immigration status, Medical information, Ethnic or religious affiliation, Sexual orientation, Account passwords, last 4 digits of SSN, Date of birth, Criminal history, and Mother's maiden name, are classified as sensitive PII.

In the process of open governance, transparency and sometimes as cost-saving measures, some educational institutions make data about their students publicly available. Loss of student and family data privacy has been accelerated by the proliferation of educational programs. For example, in the United States of America (USA), the type and amount of personal and family data collected by schools are reported in the state longitudinal database and used for research by the federal government. This was made possible by the passage of the Education Science Reform Act of 2002. In Nigeria, information about students relating to admission lists, names of scholars, scholarship beneficiaries, on one side, and potentially negative impact of publication of data related to list of expelled, rusticated and arrested students, are quite common. Although informed consent is necessary before opening up individual related data, there are times when consent is not sought. For instance, Tene and Polonesky (2013) note that if Facebook asks users to opt-in if they want newsfeed to be launched, they might actually opt-out. Also, it is commonplace to publish pictures of individuals online without the consent of account holders.

Sloot (2011) tried to analyze the tension between open government policies and the protection of personal information from the legal perspective by reviewing the applicability, legitimate purpose, safeguards and also transparency and rights. He noted that access to information or reuse of information due to open government may conflict with two legal rights: intellectual property rights on the information contained in public sector documents and databases; and open government policies conflicting with privacy legislation and the protection of personal data.

Five terms are important to the discussion of open data and privacy and are defined as follows (https://ico.org.uk):

1. **Data subject:** This is a natural person that can be directly or indirectly identified in particular by reference to an identifiable number or to one or more factors specific to his/her physical, psychological, mental, economic, cultural or social identity.
2. **Data controller:** This is a natural or juristic person or public authority who acts alone or jointly with others in order to determine the purpose and means of the processing of personal data.
3. **Data processor:** This is a natural or juristic person involved in the processing of personal data.
4. **Personal data:** This is information relating to an identified or identifiable person which can relate to his or her private, professional or public life.
5. **Personal Identifiable Information (PII):** This is defined as information which if lost, compromised, or disclosed without authorization, could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual.

This study aims to explore the knowledge and attitude of individuals towards the potential privacy breaches of open data in order to gain insight into the perspective of individual's in Nigeria about this global issue. The study sought the opinion of individuals because individuals are the data originators but used students as a proxy for this exploratory study. Specifically, the study objectives were to

determine the attitude of students to open data, the types of data they consider to be personal, and to find out if students consider open data a risk to privacy. The study also attempted to identify the data security methods students prefer for protection of personal data and their preferences for obtaining consent on personal data.

Privacy Legislation and Data Protection Act in Nigeria

It is ideal that every country should have its own privacy legislation and data protection law. The constitutional law, the Freedom of Information (FOI) Act and the Data Protection Act are great instruments for privacy laws. The following sections of the 1999 Constitution of Nigeria (Nigeria Constitution, 1999, pp.15-19) provides for privacy of citizens as follows:

a) Section 37 of the 1999 constitution of the Federal Republic of Nigeria provides that: The privacy of citizens, their homes, correspondence, telephone conversations and telegraphic communications is hereby guaranteed and protected.

b) Section 39(1) of the 1999 constitution provides that: Every person shall be entitled to freedom of expression, including freedom to hold opinions and to receive and impart ideas and information without interference.

c) Section 45 deals with the restriction and derogation from fundamental human rights. Subsection A provides a get-out clause such that in the interest of defence, public safety, public order, public morality or public health, nothing in the preceding sections shall invalidate any law that is reasonably justifiable in a democratic society.

Compliance with these aspects of the constitutional requirements vary according to whether Nigeria is under military rule or democratic rule. With respect to section 37, Nigeria has kept the spirit of the requirements in that no known breaches have been publicised. Only in few instances bordering on perceived corrupt citizens or celebrities, are photographs of houses and details of the property-owner published by the news media. This is contrary to the regular breaches that occur in developed countries. For instance, in 2014, a News of the World editor was convicted for conspiring to hack people's phones for information and the scandal subsequently led to the closure of the 168-year old newspaper house (BBC News, 2014). However, Nigeria, over the course of time especially during the military era has violated section 39(1) in one way or the other although violations since 1999 (the present dispensation of democracy) are not known. Several popular newspapers (Concord, Punch and the Guardian) were shut in 1994 by the military government in power for publishing opinions thought to be detrimental to persons in government (Orr, 1995).

Apart from the Right to privacy embedded in the Nigerian Constitution (1999), the Freedom of Information (FOI) Act of 2011 makes public records and information more freely available, provides for public access to public records and information, protects public records and information to the extent consistent with the public interest and the protection of personal privacy. The relevant sections of the Freedom of Information (FOI) Act (2011, pp.5-6) are:

a) Section 14(1) states that a public institution must deny an application for information that contains personal information. This subsection provides five exemptions pertaining to information that should be denied access.
b) Section 14(2) provides the conditions in which personal information can be disclosed: if the data subject consents to the disclosure; if the information is publicly available.
c) Section 14(3) however contains a get-out clause by providing that if the public interest in the disclosure of information outweighs the protection of the privacy of the individuals to whom such information relates.

Information regarded as relevant to Section 45 of the 1999 constitution are exempt also from the FOI requests. Nigeria is yet to have a Data Protection Act although there have been several attempts to provide one. The first attempt was in 2005 when a bill for an Act to provide for Computer Security and Critical Information Infrastructure Protection Bill was proposed; the next was the Cyber Security and Data Protection Agency bill 2008; followed by the Electronic Fraud Protection Bill 2008; the Nigeria Computer Security and Protection Agency 2009; Computer Misuse Bill 2009 and the Economic and Financial Crimes Commission Act (Amendment) Bill 2010 and again the Cyber Security Information Protection Agency bill 2012, which is still being deliberated by the National Assembly (Jemilohun & Akomolede, 2015).

In 2015, the Senate of Nigeria's National Assembly sought to pass what was termed the Frivolous Petitions Bill 2015. Apart from making it more difficult to complain about public services or corrupt officers, the bill sought to gag the use of social media for freedom of expression, hence it was tagged the "Social Media Bill" by the media and civil society organisations (Association for Progressive Communications, 2015). Due to the sustained protests by Nigerians, the bill was withdrawn in May 2016 and further consideration of it was suspended (Olaniyi, 2016).

Privacy and Open Data

Solove (2006) developed taxonomy to identify privacy problems in a comprehensive and concrete manner noting that lack of clarity between people's claim that privacy should be protected and what they precisely mean creates a difficulty when making policy or resolving a case because lawmakers and judges cannot easily articulate the privacy harm. Solove, therefore, came to the conclusion that courts are not interested in privacy breach if the information is from the public domain or if intimate or embarrassing details are not revealed. There are four basic groups in Solove's taxonomy arranged around a model that begins with the data subject. The four groups are:

1. **Information collection:** This is the collection of information from data subjects which can be via surveillance or interrogation.
2. **Information processing:** This involves various ways of connecting data together and linking it to the people to whom it pertains. The various forms of information processing include: Aggregation, Identification, Secondary use and Exclusion.
3. **Information dissemination:** The data holders transfer the information to others or release the information. Information dissemination is used to identify five groups of harms: breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, and distortion.
4. **Invasion:** This deals with the impingement directly on the individual. Intrusion and decisional interference are the harms caused by invasion.

Although Bartow (2006) criticized Solove's taxonomy for not having practical application, information collection from citizens in Nigeria is not covered by data protection rights as organisations from sectors such as banks, schools, and mobile phone operators regularly collect and store data without informing their data subjects of how the data collected is used. It could be argued that the issue is more pertinent to individual data in Nigeria as organisations, especially large ones can protect themselves from breaches. This contradicts the view of Ruohomaa and Kutvonen (2010) that the taxonomy does not cover the challenges being faced in the rights of organizations to control information produced by and about them. Mobile phone operators in Nigeria have been known to release data held by them for advertising push messages without seeking the consent of the data subject thus invading the privacy of the individual.

Mackimon (2014) recommended in the article titled "Privacy in the Age of Open Data", that in order to create a balance between privacy and open data, de-identification can be achieved by: avoiding identifiers; sharing only demographic information that is important to one's research area; deleting or transformation of outliers and collapsing rare values or reporting them as missing data; transforming

dates to less identifiable formats; mixing geographical information with other demographic information; generating participant ID number randomly.

O'Hara (2010) reviewed the legal concept of privacy in the United Kingdom and he noted that Article 8 of the European Convention on Human Rights is the only tort in the United Kingdom law and it also has a number of get-out clauses which provide grounds for transparency activists to contest a privacy ruling and also that the Data Protection Act of 1998 is not specifically intended to protect privacy, but rather to balance the interests of the subjects of data with the interests of data users. O'Hara investigated the challenges of striking the right balance between the privacy of individuals and the transparency of the government and he suggested the following methods to resolve the conflict: Anonymization or pseudonymization of data sets; the aggregation of data to less specific groups and the perturbation of data; and differential privacy.

The technological aspect of data privacy was also reviewed by O'Hara (2010) and he concluded that the technical responses (that is software, protocols and tools) to protect privacy are available more quickly but are problematic in their own right noting that technologist models of behaviour are often widely inaccurate, failing to factor in mistakes, shortcuts, ingenuity, laziness, creativity and lack of engagement. Furthermore, O'Hara claims that technological fixes or patches tend to deal with more specific types of attack than to do with legal solution. He therefore recommended that there should be a debate so that the legal and technical aspects would be amalgamated into a decision making process. The issue would be considering the risk of release and the potential benefits, so the debate would be on Risk/Benefit analysis. He gave six benefits of his risk/benefit analysis with the central hub being the release of data with proper considerations to privacy under fair and just conditions. O'Hara (2010) made some recommendations concerning the United Kingdom's open government data which include the representation of privacy interest on the transparency board, using disclosure queries and access control, creation of a data asset register and sector transparency panels. Other recommendations made by O'Hara were that there should be a procedure for pre-release screening of data to ensure respect for privacy, use data.gov.uk to raise awareness of data protection responsibilities, investigate the vulnerability of anonymized databases and be transparent about the use of anonymization techniques.

Sloot (2011) reviewed personal data in light of Data Protection Directives of the European Union and noted that personal data may either be directly identifiable, such as a name, or indirectly, such as a telephone number or a combination of non-directly identifiable information, such as age and address. Sloot buttressed his point by further saying that to determine whether a person is identifiable, all means likely and reasonably to be used either by the controller, or by any other person to whom the information is disseminated, to identify a person should be taken into account and also noted that third parties that have access to the information are also able to identify individuals. Sloot (2011) also suggested anonymization so as to protect privacy of data subject and still make the data publicly available and also introduced Personal Privacy Settings where individuals would register their own privacy settings with the government. This would act as consent from the data subject and such personal privacy settings must take into consideration measures that allow the data subject: choose to whom he would like his personal data distributed; distinguish between purposes for which his personal data is re-used; distinguish between territories he wants his data to be distributed to; and select what kind of information he would like third parties to use.

Similarly, Scassa (2014) identified three broad privacy challenges raised by open government. The first privacy challenge is balancing the objectives of open government with privacy values because while some personal information is considered to be "public", its reuse is limited to purposes consistent with the goals of its original collection. Yet a tighter control over this information would limit transparency and accountability. Thus, one challenge for governments is to reconsider the nature and extent of "public" personal information that is disclosed in light of both privacy and transparency considerations which may involve both a consideration of the potential harm (direct and indirect) to individuals that may flow from disclosure as well as the extent to which all of the personal information in the records at issue is necessary to achieving the goals of transparency. A second privacy challenge noted was that data protection has largely been structured along public and private lines whereas there

is a disparity between how government and private sector considers privacy and argued that governments are held to fairly strict standards with respect to personal information collected from individuals while the regulation of privacy practices in the private sector may be relatively loose, with a focus on obtaining customer consent to increasingly complex and often large unread privacy policies. The third privacy issue raised by Scassa relates to the release of government data through open data programs within the broader big data context which might be free of personal information and may have been anonymized but when combined with other data, might pose privacy risk.

Scassa (2014) recommended that governments at all levels must consider whether the amount of personal information disclosed in the public records in the analogue environment is appropriate and necessary in a digital and networked environment; the reassessment of the degree of openness and a greater focus on both data quality and meta-data; and taking necessary actions such as the development of guidance on when data sets considered for release may raise privacy issues, and guidance as well on when those privacy issues are overridden by the need for transparency and accountability; instead of legislative actions.

Methodology

Respondents were identified from two faculties in the selected university for the study. The two faculties were chosen with the expectation that students in the faculty of Law would be more aware of the legal implications of data privacy. Likewise, the faculty of Communication and Information Sciences was chosen under the assumption that students from that faculty should be aware of open data and data privacy. A sample of 200 students based on convenience sampling, 100 from each faculty, was used. Data was collected using a questionnaire designed for the purpose. The questionnaire was structured and it included both closed and open-ended questions. The questionnaire was grouped into five sections as follows:

Section 1: Basic information about respondents
Section 2: Attitude of respondents towards open data
Section 3: Knowledge and attitude to release of personal data
Section 4: Opinions on open data and privacy issues
Section 5: Preferred technique of data privacy assurance and pre-publication consent

The sections with the relevant questions itemized in the questionnaire are shown in Table 1.

Table 1 :
Questionnaire sections and questions

Questionnaire sections	Questions
Attitude towards Open data	1. Did you know about open data before this study? 2. Have you ever explored any open data directory/portal/site? 3. Which of the following have you explored? 4. What category of data interests you most?
Knowledge of personal data and attitude to its release	1. Which of the following data items do you consider personal? 2. Which of the following personal data would you be comfortable with if made open? 3. What types of data can you willingly release for research purposes? 4. What types of data can you willingly release on social media? 5. Would you give your data out if you knew it would be seen by the public?
Open data as a risk to privacy	1. Do you consider Open data as a risk to privacy? 2. Which of the following personal data would you consider a privacy breach if made open?

	3. Do you think individuals should be in a position to determine what personal data about them can be made open? 4. Has there been any occasion where a particular dataset about you was released and you felt uncomfortable about it? 5. Would you give your consent if your personal data is to be released with risk mitigation techniques in place? 6. In the University environment, what type of information would you not be comfortable with, if made open?
Techniques of data privacy assurance	1. Which security technique do you prefer for data privacy? 2. Do you think individuals should have an opt-in and opt-out right of personal data made publicly available about them? 3. Do you think there is a need for notification if your dataset collected for one purpose is to be used for another purpose?

A pilot study was carried out in order to improve the internal validity of the questionnaire. The reliability was tested using a test-retest method with a resultant Cronbach alpha reliability coefficient of 0.79 which is in the acceptable range of adequacy for the instrument. The data was analysed using frequency distributions and cross-tabulations. Out of the 200 copies of the questionnaire distributed, 166 were duly completed and returned, out of which 160 were useable, representing a response rate of 80%.

Results

The data collected had the characteristics shown in Table 2. Female respondents were 15% more than their male counterparts while the predominant age group was 16 to 25 years – young individuals, consistent with the expected age bracket for undergraduate students. The final data set comprised 55% respondents from the faculty of Communication and Information Sciences (CIS) and 45% from the faculty of Law.

Table 2
Distribution of respondents by gender, age and faculty

Variable	Values	Frequency	%
Gender	Male	68	42.5
	Female	92	57.5
	Total	**160**	**100.0**
Age Range	16-25yrs	136	85.0
	26-35yrs	22	13.8
	36-45yrs	2	1.2
	Total	**160**	**100.0**
Faculty	CIS	88	55.0
	Law	72	45.0
	Total	**160**	**100.0**

Respondents' Attitude to Open Data

The proportion of respondents that knew about open data before the study was 55% as shown in Table 3 and the number that had explored an open data directory or portal were 40 (27% of the 148 that responded to this question). The specific sites indicated by these 40 respondents as having been

explored are presented in Table 4, and reveals Google public data explorer as the most popular, followed by the US Data.gov.

Table 3
Prior knowledge and exploration of open data

Questions	**Responses**		**Total (%)**
	No (%)	**Yes (%)**	
Did you know about open data before this study?	69 (45.1)	84 (54.9)	153 (100)
Have you ever explored any open data directory / portal / site?	108 (73.0)	40 (27.0)	148 (100)

Table 4
Open data sites explored by respondents

S/No.	**Open data sites**	**Frequency**	**%**
1.	Data.gov	12	7.5
2.	Infochimps	5	3.1
3.	Datamarket	6	3.8
4.	Google public data explorer	19	11.9
5.	Junar	2	1.2
6.	Buzzdata	1	0.6
7.	Weatherbase	6	3.8
8.	Others	1	0.6

N = 40

Respondents were presented with a list of categories of data as shown in Table 5, and the respondents were mostly interested in Crime data (22%) followed by Business data (17%), followed by Financial data (21%), and then Environmental data (12%). Only 5% of the sample showed an interest for Transport data.

Table 5
Categories of data of interest to respondents

S/No.	**Categories of data**	**Frequency**	**%**
1.	Business	27	16.9
2.	Weather	14	8.8
3.	Crime	35	21.9
4.	Financial	21	13.1
5.	Statistics	10	6.2
6.	Environmental	19	11.9
7.	Transport	8	5.0

N = 160

Respondents' knowledge of personal data and attitude to its release

This study sought to establish how much of awareness respondents had about personal identifiable information (PII). Thus, respondents were asked to identify which of Credit card details, Medical records, Account details, Pictures, Exam results; that they consider personal. Their response is presented in Table 6. Interestingly, while 53% and 52% respectively, regarded credit card details and medical records, as personal data, only 16% chose Pictures as personal data. Furthermore, fewer

respondents identified Account details (39%) and Exam results (28%) as personal data when compared to the number that chose Credit card details and Medical records. Other data sources mentioned as personal were Family information, and Home address.

Also shown in Table 6 are the choices of data sources respondents would willingly release for research purposes and pictures recorded the highest (42%) followed by medical records (27%); and on data that can be willingly released on social media, again pictures record the highest (69%) while credit card details and medical records record the least (2% respectively). The attitude towards release of pictures is consistent with its being the most selected (61%) as a data source that respondents accept can be made open.

Table 6
Responses on **p**ersonal data

S/No	Data source	Data sources considered personal		Data that can be willingly released for research purposes		Data that can be willingly released on social media		Data that respondents accept can be made open		Data sources considered a breach to privacy if opened	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
1	Credit card details	84	52.5	5	3.1	3	1.9	4	2.5	68	42.5
2	Medical records	83	51.9	43	26.9	3	1.9	10	6.2	68	42.5
3	Account details	62	38.8	13	8.1	18	11.2	14	8.8	49	30.6
4	Pictures	25	15.6	67	41.9	110	68.8	98	61.2	18	11.2
5	Exam results	44	27.5	27	16.9	11	6.9	19	11.9	38	23.8

N = 160

Respondents' perception of open data as risk to privacy

Respondents were asked if they considered open data as a risk to privacy and 76% of the sample said "Yes" (see Table 7). When asked if they thought individuals should be in a position to determine what personal data about them can be made open, 96% of respondents said "Yes" but this proportion dropped to 46% who said, "Yes", they would give their consent for the release of their personal data into the public domain, if risk mitigation techniques have been put in place. Twenty (20) respondents (14%) admitted that there had been an occasion when the release into the public domain of a dataset about them had made them uncomfortable.

Table 7
Opinions on open data and risk to privacy

Questions	Responses		Total (%)
	No (%)	Yes (%)	
Do you consider open data as being a risk to privacy?	35 (23.6)	113 (76.4)	148 (100.0)
Do you think individuals should be in a position to determine what personal data about them can be made open?	5 (3.6)	135 (96.4)	140 (100.0)
Would you give your consent if your personal data is to be released with risk mitigation techniques in place?	67 (47.9)	73 (45.6)	140 (100.0)
Has there been any occasion where a particular dataset about you was released and you felt uncomfortable about it?	126 (86.3)	20 (13.7)	146 (100.0)

When asked which of the following personal data (Credit card details, Medical records, Account details, Pictures, Exam results) they would consider a privacy breach if made open, as summarized in Table 8, the highest number of respondents (42.5%) chose Credit card details and Medical records, respectively. Again, the lowest number of respondents (11%) chose Pictures.

Table 8
Data sources considered a breach to privacy if opened

S/No.	Data source	Frequency	%
1.	Credit card details	68	42.5
2.	Medical records	68	42.5
3.	Account details	49	30.6
4.	Pictures	18	11.2
5.	Exam results	38	23.8
6.	Others: Personal poem	1	0.6

N = 160

Since the respondents for this study were students, they were asked to identify the types of data in the university environment that they would not be comfortable with, if made open. As shown in Table 9, the highest number of respondents (55%) identified Exam results, followed by Phone numbers (17.5%) which was closely followed by Names of expelled students (16%). Only 10% of the respondents identified Email address as a source of discomfort if published in the open domain.

Table 9:
University information that would cause discomfort if published in the open domain

S/No.	Data source	Frequency	%
1.	Names of expelled students	26	16.2
2.	Phone number	28	17.5
3.	Email address	16	10.0
4.	Exam results	88	55.0

N = 160

Respondent preferred techniques of data privacy assurance

Respondents were presented with the following five options of data security measures and asked to indicate what method they would prefer:

a) Anonymity: The science of taking data that contains personal identifiable information and turning it into non-identifying data. For example, a dataset containing name, sex and opinion, name can be removed from it.
b) Aggregation: This is the summarization of dataset instead of specifying data subject. For example, 45% of patients are HIV positive and 25% of them are males.
c) Perturbation: Making data publicly available by assigning fake or unique names to identifying fields. For example, instead of Noah, we can have fx***.
d) Differential Privacy: The science of using mathematical computations to add noise to the dataset so as to guarantee the privacy of data subjects. For example, it would not allow an attacker deduce that Bob or Sally did not participate in a survey.
e) Policy-based solution: There should be data subject consent, notification, opt-out rights, etc.

Their choices are summarised in Table 10 and shows a dominant preference for Anonymity (34%), nearly twice as much as those that opted for Aggregation (19%) or a Policy-based solution (16%). These three data security mechanisms were followed by Differential privacy (11.2%) and Perturbation (10.6%).

Table 10:
Preferred data security mechanism

S/No.	Data source	Frequency	%
1.	Anonymity	54	33.8
2.	Aggregation	30	18.8
3.	Perturbation	17	10.6
4.	Differential Privacy	18	11.2
5.	Policy-based Solution	26	16.2

N = 160

In Table 11, the responses to the questions:" Do you think individuals should have an opt-in and opt-out right of personal data made publicly available about them?" and "Do you think there is a need for notification if your dataset collected for one purpose is to be used for another purpose?", are presented. The results show that 122 respondents (82%) believe that individuals should have an opt-in and opt-out right to personal data that is to be made public. Also, 83% agree that there is need for notification to the data subject if data collected for one purpose is to be used for another purpose.

Table 11:
Individual rights to data privacy

Questions	Responses		Total (%)
	No (%)	Yes (%)	
Do you think individuals should have an opt-in and opt-out right of personal data made publicly available about them?	27 (18.1)	122 (81.9)	149 (100.0)
Do you think there is a need for notification if your dataset collected for one purpose is to be used for another purpose?	26 (17.4)	123 (82.6)	149 (100.0)

Discussion

Attitude towards open data

The study showed that about half of the respondents (55%) were aware of the term "open data" before the study and compares well with the 2014 Nigeria study by Mejabi et al. (2014), where they found stakeholder awareness of open data to be around 55% in most cases, except among government officials where awareness was found to be 27%. Respondents did not frequent open data sites, but the few that did, explored Google public data explorer and Data.gov. The fact that this is a sample of students may be responsible for visits to these two sites more frequently than others listed such as Infochimps, Datamarket, Weatherbase, and others.

Of note, is the high number of respondents that indicated an interest for Crime data (22%) over other types of data presented such as Weather, Financial data, Statistics, Transport which all recorded single digit frequencies. The frequency of respondents choosing Crime data was followed by Business data (17%), Financial data (13%), and Environmental data (12%). This seems to be reflective of the present socio-economic conditions in Nigeria with very high crime rates (level of crime is 74 on 100 point scale according to www.numbeo.com) and an economy that is in recession (African Economic Outlook, 2017).

Open data as a risk to individual privacy

The study also identified that majority of the respondents consider open data as a risk to their privacy because of a number of reasons such as security reasons, misuse of data, privacy concerns and re-identification of data subject. Furthermore, the study established that respondents want to give their consent before data about them is made open because they have the freedom of choice. Similar to what Solove (2006) found, that individuals want to consent to most of their activities before making it open. This implies that there is likelihood that individuals would be more inclined to opening up their dataset if their consent is sought. During awareness programs, individuals should be enlightened about the risk mitigation mechanisms in place before opening up datasets so as to encourage data sharing and to make individuals see open data as a beneficial approach to enhance knowledge and not as an ICT enhanced revolution for breaching privacy.

Personal data and its release

The study was able to determine that majority of the respondents classify credit card details as the most personally identifiable information with the most risk, followed closely by medical records, account details, exam results, pictures, and home address. Respondents do not want their credit card details to be made open at all but majority are comfortable with making pictures open. This is evident with pictures individuals post about themselves on social networks. This is why O'Hara (2010) claimed that people's attitude to privacy is cavalier because what they post about themselves online almost negates their clamour for privacy. What is not Personal Identifiable Information today might become Personal Identifiable information tomorrow. For example, Sweeney (2002) was able to identify citizens of the United States who participated in the 1990 census using their zip-code, gender and date of birth. As at that time, these attributes were not considered as personal identifiable information although information such as names and addresses were removed from the dataset.

Preferred mechanisms for ensuring data privacy and obtaining consent

Respondents mostly want their datasets to be released based on approval by the data subject and also with some form of data security mechanism in place. The study showed that anonymity is the most preferred security technique by the respondents because it prevents the data subject from being identified, it protects personal information from privacy breaches and because it puts off attackers.

According to O'Hara (2010) such data can be released when the PII have been turned into non-identifying data. It is very likely that respondents are not aware that the data security technique of anonymity and even aggregation, perturbation and differential privacy, are not full proof. For example, Netflix released a dataset of users and their movie ratings. The data consisted of a customer identification (faked), a movie, the customer's rating of the movie and the date of the rating. Narayanan and Shmatikov (2010) looked at the claim that the data was perturbed by asking acquaintances for their rankings and they found out that only a small number of the ratings were perturbed because perturbing data gets in the way of its utility. Since different movies were watched by different people due to individual taste and the dataset was sparse (because most people had not seen most movies), it was easy to pick out the individual that watched a movie if you knew the movie someone watched and the day he watched it.

The study found that respondents want to have an opt-in and opt-out right before making their dataset publicly available. Finally, respondents want to be notified if the dataset collected about them for one purpose is to be used for another purpose. Also, the opt-in, opt-out and notification of change of data use, can come under legal policies to protect the rights of the data subject. According to the United Kingdom Data Protection Act of 1998, the data owners or subjects have a right to be informed of data being kept; Right to prevent data to be used; Right of access; Right to rectify, erase or block data; Right to prevent processing of data where it might cause damage or distress; and Right of data owners concerning automatic processing of data.

There are situations where data subject consent is waived and such situations are provided as exemptions in the Data Protection Act (1998) of the UK. The findings on opt-in, opt-out rights in this study align with De Latt's (2005) discussion of an American public poll that revealed that individuals want to be in control of both initial collection of data and data sharing and that the larger percentage of the young adults (18-24) are in harmony with older Americans regarding concerns of online privacy, norms and policy suggestions.

Conclusion and Recommendations

The steady proliferation of information in this digital age has brought forth the data revolution which has made data-sharing easier as compared to the traditional means of sharing data. Open data is a means for making data freely accessible, sharable and reusable for transparency and accountability among other reasons. Privacy concerns are one of the problems in opening up data especially when it involves PII. To encourage the release of open data while ensuring privacy breaches are kept to the barest minimum, the following recommendations are made:

a. There should be increased awareness programs in closed communities such as universities and for the society at large to enhance the understanding of open data and the benefits of opening up data along with data protection measures that ensure security of published data.
b. Citizens should be educated on the need to consider pictures as very important PII, especially in the wrong hands so that careful consideration is given to releasing pictures in the open domain.
c. For effective use of individual level data, organizations can establish a trust framework similar to Mydex CIC trust framework which consists of a set of legal and technical rules by which members of a network agree to operate in order to achieve trust online. It allows individuals to connect to each other and the organization with a trusted identity. It also allows individuals to be in control of the permissioning process.
d. Finally, a holistic approach should be followed in the collection, processing and opening of datasets with emphasis on risk mitigation throughout the open data life cycle.

For further research, a wider study involving data subjects of specific open data sets should be undertaken. Also, the different data security mechanisms should be tested to determine their effectiveness and the situations in which they are best applicable.

References

African Economic Outlook (2017). Nigeria Economic Outlook. African Development Bank. Organisation for Economic Co-operation and Development, United Nations Development Programme. Retrieved on 3rd March 2017, from https://www.afdb.org/en/countries/west-africa/nigeria/nigeria-economic-outlook/

Association for Progressive Communications (2015). Open letter to the Nigerian Senate on the Frivolous Petitions Prohibition Bill (aka "Social Media" Bill). Retrieved May 14, 2017, from https://www.apc.org/en/pubs/open-letter-nigerian-senate-frivolous-petitions-pr

Bartow, A. (2006). Fair Use and the Fairer Sex: Gender, Feminism, and Copyright *L*aw.*American University Journal of Gender, Social Policy & the Law*. 14 (3), 551-584. Retrieved December 7, 2016 from http://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1267&context=jgspl

BBC News (2014). Phone hacking trial explained. Retrieved May 14, 2017, from https://www.bbc.co.uk/news/uk-24894403

Casado, C. (2014, July 31). The Future of Privacy: The New Data Protection Regulation. Retrieved on the 10th of January 2015, from http://www.mobilehealthglobal.com/in-the-news/article/28/the-future-of-privacy-the-new-data-protection-regulation

De Laat, P.B. (2005). Trusting Virtual Trust. *Ethics and Information technology,* 7, 167-180.

Jemilohun, B.O. &Akomolede, T.I. (2015). Regulations or Legislation for Data protection in Nigeria? A call for a clear legislative framework. *Global Journal of Politics and Law Research,* 3 (4), 1-16. Retrieved September 27, 2016 from http://www.eajournals.org/wp-content/uploads/Regulations-or-Legislation-for-Data-Protection-in-Nigeria1.pdf

Lexis Nexis Risk Solutions (2012, July 18). Role of Social Media in law Enforcement Significant and Growing. Retrieved from http://www.lexisnexis.com/en-us/about-us/media/press-release.page?id=1342623085481181

Mackimon, S. (2014). *Privacy in the Age of Open Data.* Retrieved September 10, 2014, from http://osc.centerforopensceience.org/2014/01/29/privacy-and-open-data

Mejabi, O.V., Azeez, A.L., Adedoyin, A., &Oloyede, M.O. (2014).*Investigation of the Use of the Online National Budget of Nigeria*. Retrieved September 10, 2014, fromhttp://goo.gl/KHgRZf

Narayanan, A., & Shmatikov, V. (2010). Privacy and Security: Myths and Fallacies of "Personally Identifiable Information". *Communications of the ACM,* 53 (6), 24-26.

O'Hara, K. (2010). *Transparent Government, NOT Transparent Citizen: A Report on Privacy and Transparency for the Cabinet Office:*Southampton, UK.

Ohm, P. (2010). Broken promises of privacy: responding to the surprising failure of anonymization. 57 UCLA Law Review 1701 (2010) 1701-1711.

Olaniyi, S. (2016). Senate throws out Frivolous Petitions Bill. *The Guardian,* 18 May 2016. Retrieved May 14, 2017, from https://guardian.ng/news/senate-throws-out-frivolous-petitions-bill/

Open Data Handbook (2012). What is open data? An Open Knowledge Foundation Project. Retrieved October 14, 2014 from http://opendatahandbook.org/en/what-is-open-data/index.html

Orr, D. (1995). Nigeria's junta keeps press on the run. Retrieved May 14, 2017, from https://www.independent.co.uk/news/world/nigerias-junta-keeps-press-on-the-run-1603549.html

Ruohomaa, S. & Kutvonen, L. (2010). Trust and distrust in adaptive inter-enterprise collaboration management. Journal of Theoretical and Applied Electronic Commerce Research, 5 (2), 118-136.

Scassa, T. (2014). Privacy and Open Government. *Future Internet*, 6, 397-413; doi:10.3390/fi6020397

Sloot, B. (2011). Virtual Identity and Virtual Privacy: Towards a Concept of Regulation by Analogy. *eGovPräsenz,* 2011-1, 41-43.

Solove, D. (2006). *A Taxonomy of Privacy.* MA: Harvard University Press.

Sweeney, L. (2002). K-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, 10, 557-570.

Tene, O., & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property,* 11(5), 240-273.

Wainewrite, P. (2014). Open data or privacy breach? Retrieved December 7, 2014, from http://diginomica.com/2014/04/22/open-data-privacy-breach/