The Public Perception of Privacy Research and Analysis for open Data

Chi-Cheng Chung
Department of Information Management,
National Taichung University of Science and Technology,
129, Sec. 3, Sanmin Rd.,Taichung City 404,
Taiwan, R.O.C.

Shu-Fen Chiou
Department of Information Management,
National Taichung University of Science and Technology,
129, Sec. 3, Sanmin Rd.,Taichung City 404,
Taiwan, R.O.C.

Abstract—With the development of big data in various fields, Open data is more and more important. Open data can help us to analyze and get knowledge for our life, social economy, and etc. However, there are some problems when using open data. One of these is how to keep users' privacy. In our country, there is less to study in the privacy section. Thus, we focus on the privacy issue. In this paper, we explore open for privacy and cognitive information. The results are showed that there are currently 7 percent to open data person, agrees anonymous open data has 5 percent, feel anonymous privacy violations are still 4 percent, while in use under the premise inform open data, there are 40% the man agreed to a non-anonymous open. But regardless of the anonymous or non-anonymous, open data supporting measures will need to complete, especially in our country stage of development, the need for more attention to this aspect, otherwise it will cause open data hinder the future development.

Keywords— Anonymous, open data, privacy, questionnaire

I. INTRODUCTION

The revolution in information technology, took place around us, but so far most have focused on science and technology (T), and now it is time to look to information (I) time [1]. Big data is used to build models from the Google search through fifty million words, to find out 45 key search words to predict the results of real data almost in line with the official announcement in the first. [2] While big data "parent = sample" is not necessarily a huge data, in other words, big data refers to the complete set of data, so the big data set of data does not have to Terabyte (TB) or very huge data, and even some of the information set than a digital photo is still small. But before big data, random sampling quite remarkable feat, but also a good shortcut, so that the development of digital technology is not complete, it is possible to analyze large amounts of data. In the enterprise, random sampling is also used as a tool for quality control, not only easier, but also more cost-saving [3]. The main reason is because the data can't be collected for analysis, will choose a random sampling. However, despite the success of random sampling, but also hidden a lack of random sampling just like compressed digital video or audio files as the sample is bound to have missing data, and in the example at random sampling estimate the overall situation, there will be a 3% error [1]. Although some research or investigation, there is no damage, but the lack of details, the less the details, we can't explore all possible potential problems [1].

Using big data analysis of personal privacy information, to analyze the first mayor of New York City as an example Flowers, Flowers took office the first thing is to solve the problem of illegal buildings New York residence [1]. While New York City each year about complaints of illegal buildings twenty-five million. It has no good method to determine whether is a really or false complaint. Therefore Flowers team to think, it seems that as long as a big data will be able to solve this situation. First, they get residential data, and then collect information from other organizations, such as utility bills exception, whether foreclosures, and some of the old members of police experience and so on. Then put data into the model to determine which buildings may be fire, then after the relevant information to the inspectors to deal with, and the results have surprised everyone in the big data analysis has not been used before, and only 13% buildings need to be given orders to evacuate, but the use of a huge amount of data, even 70% of the buildings to be evacuated [1].

Application examples Flowers of New York City Mayor would tell, big data is certainly possible to improve human life or is used in predicting the future of things. But before making the big data analysis, acquisition data is the most difficult part, so now most of the big data analysis, almost using open data of the government, firm or variety agencies to analyze and prediction. But regardless of the use to predict the future or in the improvement of living things, databases are privacy issues involved, although open data in the release data, the information relating to personal privacy will be anonymized, but in order to open data when not undermine the relevance of data itself, each seemingly anonymous data, in fact, has a certain significance. Like the University of Texas at Austin's research staff, the Netflix announced half a million users in the race rental video data [3], and other public information to do comparison, soon found an anonymous user ratings, studies have shown that as long as there Comments over six popular films have 84% of the opportunity to identify the identity of the client, but also to know if someone in day rates, there is 99% of the opportunity to identify it [1].

In the implementation of open data must also be carried out under sound legal, because the open data on behalf of the public to view privacy to others, and in order to allow open data safer, most of them anonymously to increase personal privacy security[4,5]. In addition to the basic law, the idea of the most important contributors of data is very important, because there involves personal privacy information is provided by each contributor of information, in other words, "data contributor" is Construction of a database source, so in addition to data laws and regulations, analyzing,

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

storing, etc., and the source of data is also worthwhile for us to discussed.

The above content derived, open data is able to create one of the main part of the big data applications, but at the same time the use of data, most of us do not care data contributors feel. Despite the opening of these data in order to improve human life, but still need to respect the idea that every data contributors. Therefore, in this study through questionnaires to understand the general public for the big data and open data view, as well as in firm or government public database, although you can query data to individual behavior, but the relevant information under such as telephone number, name, identity card, etc. are all anonymous, the people are still worried about whether or not even accept corporate or open Government data [6].

II. OPEN DATA

A. Definitions

Open Data refers to the government, individuals or private entities by picking out the inside of the released data, and is not limited to intellectual property rights or related to management constraints, the use of data and the concept of authorization, and can be used in different areas, like a business, analyze trends, entrepreneurship, problem solving and so on [7,8]. According to "Open Data Handbook" definition of open data can give anyone is free to use, and can be reused with scattered, but on its limitations, can only ask the user to indicate the source of data with the provider [9]. In addition, the open data need to have the following characteristics:

1) We can obtain all of the data, it must be based on convenience, we can modify the form out, and it is better to be able to directly download via the Internet.The cost must be reasonable when it remake.

2) Open data must be allowed to re-use and distribution, as well as mixing with other data used in the statement.

3) Anyone can use, as well as re-use and dissemination of data, and can't limit the scope of data [10].

B. Open Government Data

Open government data means the Government will data release and comply can download, use, meaning everyone can use, and does not have legal or related issues [11]. OECD by Governments will now released from the following types data:

1) *Business information.* (Including the Chamber of Commerce information, public information.)

2) *Registration*

3) *Geographic information.* (Including address information, aerial photographs, buildings, geographic information, etc.)

4) *Legal information.* (Including national judgments, foreign and international court decisions, legislation and treaties.)

5) *Weather information.* (Including meteorological information and models and weather forecasts.)

6) *Social Information.* (Including economic, employment, health, population, public administration statistics.)

7) *Traffic Information.* (Including traffic congestion information, road construction, public transport and vehicle registration.) [12, 13]

C. The Concept of Open Government Data(United States Sebastopol meeting: Open Government Data Eight Principles)

In 2007 December 7 to 8, Sebastopol meeting California by 30 people concerned about open government data is involved. And at Carl Malamud, Tim O'Reilly, Lawrence Lessig leading end of the meeting, sorting out eight principles on open government data, the government in compliance with the following principles will be deemed open:

1) *Complete:* All open data can be obtain. Open data refers not privacy, security, or other restrictions.

2) *Primary:* Data stored in the most complete form, without the use of process after the Data.

3) *Timely:* Data should be open as soon as possible, in order to ensure that the data value.

4) *Accessible:* User instructions should be sources of data, and to allow widespread use.

5) *Machine processable:* Open Data for lawful case, a variety of machine-readable formats provide, does not exclude or reject any member of the public or organizations to use Data need to have a non-exclusive, publicly available and without any use restrictions in order to achieve the widest range of applications [14].

6) *Non-discriminatory:* Data must be disclosed to any person to use, and does not require authorization.

7) *Non-proprietary:* Data formats without limitation, and disseminating information format not use special way.

8) *License-free:* Data from copyrights, patents, trademarks and trade secrets of control, but in the privacy, security, or other special rights of use restrictions may be considered.[12,15]

D. Data Format

In the open data format, Tim Berners-Lee [16] suggested that it be divided into five stars architecture: One Star publisher issue using relatively simple format. Although the user can read, but want to get Data, still need other software or use artificial means (Software such as: ORC, Excel, Google forms); Three stars publisher will be saved as CSV file format does not require dedicated software to open (for example: Notepad, but Notepad file visualization of open lower), and the user will need to download before been used; Four stars publishers when publishing a given fixed address, Data can be read directly through the Internet, also can be converted into the format desired by the user; Five stars publishers giving a fixed position and can be read directly in the Internet, and can be linked to other information, as summarized in Table I.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

TABLE I.   THE FIVE-STAR RATING ANALYSIS OUTLINED IN TABLE

Star Rating	Brief Introduction
★	Download read through the network, but requires dedicated software can read data from the archive.
★★	Download via the Internet, but requires dedicated software is available to open the file to read the data
★★★	Download via the Internet, does not require dedicated software to open the file to read the data.
★★★★	Through the network can be directly read or download data
★★★★★	It has a four-star advantage, and provide additional data from the original table links to other data tables.

E. Data Format

"Open Data Now" author Joel Gurin, in the open data international forums cited McKinsey study, indicating that the current global open data can create business opportunities can be up to 3 trillion to $ 5 trillion. While the US in education, health, transportation, energy and finance these five areas, at least 500 companies have been open data, and hope that other governments are free open data [17]. When Obama took office the same day the United States, he immediately signed a "memorandum of transparency and open government," and "transparent government", "citizen participation" and "collaboration" to "open government" as a basic principle, and in the subsequent proposed Open Government Data plans to implement the memorandum[18, 19, 20]. In a memorandum of content organized into the following three and be interpreted:

- Transparent government: that the government release the relevant information through the Internet provides people read, and provide the public with feedback conduit to reach optimization [18, 19, 21, 22].

- Citizen Participation: to provide opportunities for public participation in the decision-making process, to improve the quality of government decision-making [18, 19, 21, 22].

- Collaboration: requiring agencies through innovative tools, methods and systems that come with non-profit organizations, businesses, and individuals to seek cooperation and feedback to assess the situation through cooperation of government and find new opportunities for cooperation [18, 19, 21, 22].

And, after reference to federal chief technology officer and public comments, the US administration and the Budget Office released "Open Government Directive" memorandum requiring agencies to publish information within the time limit government, strengthen government information quality, and strengthen the "open government" culture and the establishment of "open Government" architecture, and release of the organs have data on the Internet. For example, to identify the most valuable inside information within a month, we need to deal with the situation published on the Internet within two months, and within three months, together with the Federal Office review whether there is implementation of

open government. The current results can be found in www.data.gov [18, 19].

But also because the White House's attention, as well as a substantial use of technology professionals from Silicon Valley into the government system, and therefore such as NASA, the Ministry of Health, the Environmental Protection Agency, etc. departments, are going to open big data and calculation software services to reach the benefits of open government . In addition, the White House released worry may cause privacy violations, so the signing of "Consumer Privacy Protection Act" in February 23, 2012, the bill provides the umbrella for the personal information and open data [23].

In 2013, our government created the DATA.GOV.TW open platform which provides users to download data or to give the feedback. In 2014, the platform had got 3,054 records for open data. The United States announced data format compared to the Taiwan data set is still mostly three-star or less open structured or unstructured data, there appears less compliant with open standard four-star, or can provide inquiry data across data sets the five-star standards [24], so the quality of the Control Manager also needs to be improved, it is desirable to make the follow-up open data have quality specifications, even like the United States to regulate the quality of each department data. In addition, the open data is also due to currently issue decrees and advocacy has led many departments as you want to open but they fear negative publicity, so the choice is not open. Taiwan amending the law in terms of speed and was unable to keep up with the changing times, but also caused a lot of open data because of legal issues can't be completely open, this part of our government in the future is worth learning correction and effort, in addition to the policy guidance is also open data there is a real impact [25], and open data is not simply open the outside to use, but the implications to the future development of big data applications, as well as data that can be created by the open interest, which are not to be underestimated [26].

F. Open Data Risk

In understanding the risks, we need to know, not everyone has an open data containing personal information. Data such as factory machinery, airports and weather information, etc., for these companies, they do not need personal data can be analyzed to obtain the value of which is not related to personal privacy of sensitive data [27]. And with these data, it would be completely without causing any risk to privacy. But the main problem is not open Data that is likely to endanger privacy, but a variety of possibilities through algorithms to predict various things, such as personal medical process is stored into the data analysis to examine the possibility for disease prediction or improve current drugs abuse [28] or retail store customer data through to analyze consumer preferences [29] and crime prevention and pre-emptive [6] etc. This will trigger a series of theoretical issues: free will and dictatorial Data, if the use of large data used in the crime, the results derived from the statistical determination made, but with the individual will differ, how to decide?

The main risk of open Data is still a major issue of privacy, because if there is an open Data involves personal data, or behavioral data, there may be negative propaganda and malicious use [27], as long as the current through the big data computing, we can predict a person behavior, and this

part needs to be improved by amending the law. In the past few decades, the basic principles of personal data privacy protection, mainly the first control to the individual, and then by the individual to decide whether to allow others to use, and allows the method used [30]. This approach is now very respect everyone's privacy, but the era of big data, this method has not as useful. In the current popularity of open data, we should change the previous practice, would only be responsible for the traditional principles of their own personal data, change to the principle usage data on its use are required to take responsibility. And for high-risk category, due to the current regulations are not applicable[31], the basic principles of legislation must be used, the data user must use risk assessment methods prior to use, but also ways of reducing the risk of harm caused by the method. In this way, you can achieve the creative re-use of data, while the relevant supporting measures to avoid the final result in personal injury, as well as in the future for the development of open data down barriers[1]. PhD and Master Thesis currently open Data in Taiwan, mostly in the Discussion on the application or the currently open Data regulations related to this field to open this Data is helpful, but also represents the open Data has begun to be taken seriously, but not enough the place is rarely explore this part of the personal data privacy, and the privacy of personal data related to explore is different ignored. In the past, when no open Data, each Data only represents a movement, but the advent of open data, each Data that is representative of a person, can be achieved even predict behavior through big data analysis. Therefore, we must respect each after Data providers, care about their feelings, about their idea, because in the era of open data, you and I are open Data provided by a member, we need to be understood, respected and used feeling.

III. OPEN DATA FLOW PLAN

We explore the Data for the above documents open to suggestions, open data and the establishment of the future process on the basis, the implementation of open Data will be relatively smooth. So in the first part, we introduce process planning open data, and then we explore each process to support the proposed model and easy to follow open Data reference.
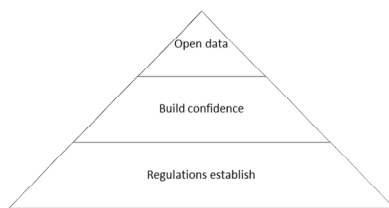

Fig.1 Open Data flow model

A. Process Planning

According to the literature, we sort the three basic processes of open Data in Fig. 1. The lowest level is the regulations establish, confidence is in the intermediate layer, and the uppermost layer is open Data.

B. Flow Description

In the era of accumulating data, since 1980, the accumulated data is doubling every three years, and now the popularity of wired or wireless networks, as well as a multitude of devices capable of generating data, such as mobile phones integrated GPS, PDA, thereby creating spreadsheets and media files, etc. [32], will increase the speed of rapid data accumulation, analysis and storage in the future so made use of the data is likely to be a major problem. Therefore, data analysis and storage in the future is likely to be made use is a major problem. In the current analysis of the data stored there have been quantum computing, quantum computers composed by quantum bits, or qubits, for example: 100 qubits store size is 2^{100}. Currently the development of quantum computers have been a number of products, for example: in in 2011 D-Ware Release D-Wave one of 128 qubits and in 2013 launched the D-Wave two of 512 qubits [33]. In addition the acquisition and use of data in the previous business and the agencies use their own data to analyze, and not much of a question. But now there is open Data issues, open Data to government transparency, citizen participation, collaboration of these three as the goal [18, 19], and now most of the countries belonging to the referendum democracy and the sovereignty of the people of the country, but also need to focus on; But at the same time open Data, we must define whether this Data is sensitive data, such as health Data, employment, and financial income [10, 34] and so on. So how to protect Data contributors and to ensure correct use of Data, it is important to establish in advance the Regulations, because if Regulations is not perfect in the case of open Data will cause a certain degree of risk or slow the pace of development of open Data, such as the EU data protection laws [35, 36], except when making specification for data users, but also for the data owner to use the specification, such as the police department has a huge crime information, but are related to personal privacy, so if there is no the case is approved, police behavior is a data access is a crime [6]. Therefore, the establishment of laws and regulations, in addition to understand open Data is what is sensitive data, and sensitive data, how norms in the open data that more commonly used method is to use a string of code to replace sensitive data, and except for the open data compliance data range, data owners and users in the use of must be standardized, so that data owners and users responsible for the data, rather than contributors to each data. There is also need to comply with the principles of data protection and privacy, we must prohibit the transfer of data, like the European legislation prohibits the transfer of data to the outside of the European Economic Area [37], rather than in 2010 Octopus Enterprise in the interest of money, will user data sold to third parties, so it is necessary to strictly prohibited! [38]

After the Regulations established, we have to tell the general public that our Regulations is perfect, but in addition to improving regulations are some warning and mandatory, so brief propaganda legislation is necessary. In addition to propaganda Regulations, the positive modifications are necessary, like Hong Kong or Malaysia, although there are data protection laws, but integrity protection law was inadequate [39]. So when the Regulations of loopholes, if not immediately repair or carry out emergency measures, with the growth of data will cause an increased risk and more severe cases can cause people to lose confidence in the Regulations, not to believe that Regulations to protect their Data. Although the open Data process model, the figure build confidence is above Regulations build, but if the general public to lose confidence in the Regulations, even if there is in perfect Regulations, still making it difficult for more

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

people to support open Data . But in addition to the basic regulations publicity, and the potential benefits of open data he had hidden also need to tell the general public, such as you can reach through the open large amounts of data to analyze achieve flood damage assessment, drug risk assessment, customer evaluation, market forecasting, and etc. [40].

The final part, as well as Regulations build and build confidence is completed, thus, we can reduce the open Data security concerns of people, and the open data resistance will be reduced. Let hold data institutions or enterprises in the implementation of open data will follow the law, so that people have the data more secure when making open data, and also guarantee contributor data. And open data standards except in accordance with regulations other than quality also need to pay attention, because the open data is mainly transparent government, citizen participation, collaboration, so when open data [18, 19], we should be able to perform data using relatively simple software, like Tim Berners-Lee suggested five-star architecture [16] and eight principles of the American open data. We can also refer to other developing better country, such as New York's open data format, or other better examples are worthy of our consideration. Also in the open data before we have to pay attention to issues related to open data and data sources described manner, such data is centralized or decentralized, whether usage data of various government departments, data sources as well as data have privacy and copyright and other issues [41].

IV. RESEARCH METHODS

Due to the rise in the current open data, while most research major at the application level and discuss the legal aspects, and less for the general public's views are discussed, such as whether to accept anonymous open and so on. Research aimed at the general public conducted a questionnaire, and in order to avoid affecting the result, not the ease of use when questionnaire sample, sample surveys are all conducted through the Internet platform, and after the screening to eliminate invalid samples to achieve the authenticity of the sample.

A. Questionnaire Design and issuance

This research investigated the views of the general public to understand the general public for the open data understanding and perception.

■ Questionnaire Design

Questionnaire is to the general public for the open data support and understanding of the research questionnaire Likert 5-point scale (1) represents strongly disagree (5) represents strongly agree. Survey and sample collection takes about one month, the first question we first research to measure every facet, after design survey questions, give two

scholars confirm the contents of the question and make changes, this action will be repeated until the scholars complete all the recommendations. Then we also pre-test questionnaire, a questionnaire before the object is 14 graduate students, the first post-test questionnaire completed during the adjustment to enhance validity. Because this questionnaire belong to a single question, so on the recommendation of the previous review questionnaire scholars and graduate students, increase the negative question to prevent casual answer questionnaire to enhance reliability, after the above-mentioned process, a total of 9 positive question to the last question, and question 6 negative question, a total of 15 topics for open data to understand the question.

■ Sample collection

In order to determine the participants were randomly surveyed and strange objects are unknown to us, we first after our questionnaire electronically, mainly large network platform for the current PPT questionnaires, sample collection time from November 2014 starting on the 7th, a total investigation time after two weeks. Finally, the sample collected a total of 260 parts, 15 parts of the sample is invalid, the remaining 245 after deleting invalid samples.

B. The questionnaire and sample overview

The basic information like gender, age and education of questionnaire people is shown in Tables II-IV.

TABLE II.   GENDER DISTRIBUTION TABLE

	Males	**women**	**Sum**
No. of people	89	156	245
%	36.3%	63.7%	100%

TABLE III.   QUESTIONNAIRE AGE DISTRIBUTION TABLE

Age	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**Sum**
No. of people	85	105	24	19	4	3	5	245
%	34.8	42.8	9.8	7.8	1.6	1.2	2	100

[a.] A=16~20, B=21~25, C=26~30, D=31~35, E=36~40, F=41~45, G=46 Over the age.

TABLE IV.   EDUCATION DISTRIBUTION TABLE

	Junior high school	**Senior high school**	**University**	**Master**	**PH.D**	**Sum**
No. of people	4	35	174	30	2	245
%	1.6%	14.3%	71%	12.3%	0.8%	100%

TABLE V.   The results of Q1-Q3

	A	**B**	**C**	**D**	**E**	**Sum**	**Average**	**Standard deviation**	**Variance**
Q1	2.9%	22.9%	46.9%	20.4%	6.9%	100%	2.94	.908	.825
Q2	20.2%	36%	21.4%	15.7%	6.7%	100%	3.47	1.175	1.381
Q3	12.4%	35.4%	28.6%	17.4%	6.2%	100%	3.30	1.088	1.184

[b.] A= Strongly agree, B= Agree, C= ordinary, D= Disagree, E=Strongly Disagree

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

V. ANALYSIS AND DISCUSSION ON THE OUTCOME

According to our questionnaire, we the general public to open data view and privacy anonymous basis sequentially view the general public (1) for an open data understanding, (2) whether to support the open data, (3) The open data anonymous and privacy of view, and open data of the current situation, the challenges to give advice.

A. Q1:For open data to understand

In Table V, research results from questionnaire showed that 2.9 percent of people strongly agreed and 22.9 percent agreed to open data understanding, while 20.4 percent disagreed and 6.9% of people strongly disagree open data for understanding, about 3 percent of the people do not understand what information is open to do, but accounted for 46.9% of the ordinary high follow-up and we have to send E-mail to ask for respondents to fill in "ordinary" reasons, most of the answer is to see literally, or from books and others chat content etc., so only limited to open data surface meaning to understand.

Research are currently open data unknown accounted for 20.4%, while completely unaware of 6.9%, but still may affect future open data, due to the current for the open data for the ordinary 46.9 %, while the open data for ordinary attitude of people, there may be external factors change their minds, for example, a message said, Open Government data leading to a capital leakage people, seriously affecting people's private life, understanding "ordinary" most of the people will not support the biased, if changed to Open Government data for IT industry generates great leap forward in changing the life of great help, to understand the "ordinary" people will tend to support. Basic causes of these problems are the open data lack of knowledge, will have such a result. In the current understanding of the advantages of open data fill in the "strongly agree" respondents found mostly belonged to the IT-related industries, we now understand the importance and benefits of open data, is still in the information-related industries people, if you want to improve the current situation, or prevention of negative information influence the future development of open data is for open data prior publicity, told the open data benefit non-IT industry people, and the problems may be caused and how to solve, to achieve the correct receiving messages.
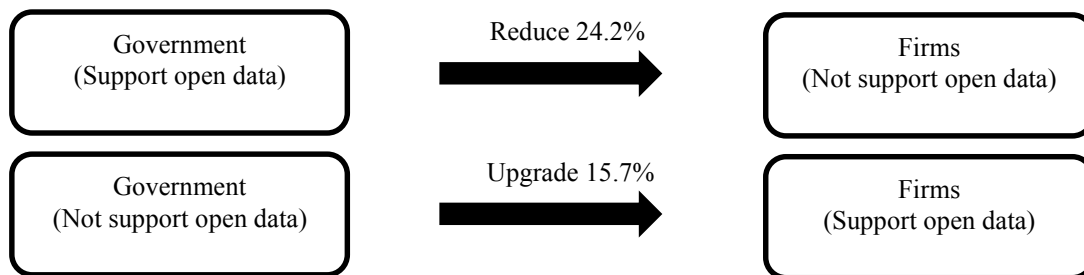

Fig.2 Support for open data (Firms and government) attitude change Fig.

TABLE VI. OPEN DATA (FIRMS AND GOVERNMENT) ATTITUDE CHANGE ANALYSIS TABLE

	Open firms data willingness to lower	open government data willingness to lower	Unchanging	sum
Number of people	43	28	107	178
Percentage	24.2%	15.7%	60.1%	100%

TABLE VII. Whether to support the open government data (open to fill ordinary person to understand data

	Strongly agree	Agree	ordinary	Disagree	Strongly Disagree	Sum
Number of people	16	39	28	23	9	115
Percentage	13.9%	33.9%	24.4%	20%	7.8%	100%

B. Whether to support open data

■ Q2-Q3:Support open government data or Open firms Data

In Taiwan, most of the open data is still Open Government data, because legal restrictions or because of privacy issues, hardly any Open firms Data, but information is also owned by the enterprise closer and personal privacy. Such as the purchase of information, inquiries and other information. Therefore, this study was to investigate whether to support the open data part, we will separately investigate

whether to support Open Government data (Q2) and support Open firms Data (Q3). In the survey, if the for open data understanding question, fill in the "disagree" or "strongly disagree", it can not answer the question of follow-up, follow-up questions and be able to fill in the number of total 178 , so do not pay more to explore in the back of repeat reasons.

In Table V, Q2 results show that support the Open Government data attitude to strongly agree accounted for 20.2 percent and 36 percent agreed, while 15.7% of people attitude to disagree and 6.7 percent of people attitude to strongly disagree Open Government data , while 21.4% of

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

ordinary people currently. And support Open firms Data attitude to strongly agree accounted for 12.4% and 35.4% agree, disagree the attitude of 17.4% and strongly disagree there is 6.2 percent ordinary 28.6%. From the survey, the more people support the open data, so in the current open data by the government to be open data strategy is correct.

■ Q4: For the open government data or Open firms Data view of different people

In the analysis of the information at the same time, we found a very interesting thing. Previously on whether to support open government data attitude to agree, strongly agree to support Open firms Data becomes ordinary, disagree, strongly disagree and the support open Government data statement to disagree, strongly disagree to support Open firms Data becomes ordinary or agree, strongly agree. In the results of the survey, the former than the support (strongly agree and agree) people open government data, the encounter will reduce open firms data accounted for 24.2%, while the original is less support (disagree, strongly disagree) open government data people encounter Open firms Data wishes but increased 15.7%.

This part of the issues can be discussed in the future, although higher support open government data visitors, but it is worth noting that the original does not support the open government data people, why it would be more support open firms data, but this part of the right must be interpreted by future investigation to know the results, so the results of the study will not be guesswork. We show the concept in Figure II and results in Table VI.

■ Q5-Q6: For open data understanding as ordinary people

In the study survey for the open data attitude to ordinary people accounted for 46.9%, while they were in the follow-up whether to support the open government data and supports the open firms data to answer the two options, support open government data, there are 13.9% of people attitude to strongly agree and 33.9% of people attitude to agree, while 20% of people attitude to disagree and 7.8 percent of people attitude to strongly disagree, but there are still 24.4% ordinary people still attitude (see Table VII). Whether in support of open firms data, 8.7% of people attitude to strongly agree and 31.3% of people attitude to agree, while 22.6% of people attitude to disagree and 7.8 percent of people attitude to strongly disagree, 29.6 % of people are still in common (see Table VIII). As for the current attitude as ordinary people, they are for open data future development can not be ignored, because they understand the open data may still be scanty, so the advantages of open data for publicity and future impact is more important .

TABLE VIII.   Whether to support the firms open data (open to fill ordinary person to understand data analysis) (10)

	Strongly agree	Agree	ordinary	Disagree	Strongly Disagree	Sum
Number of people	16	39	28	23	9	115
Percentage	13.9%	33.9%	24.4%	20%	7.8%	100%

TABLE IX.   The results of Q7-Q9

	A	B	C	D	E	Sum	Average	Standard deviation	Variance
Q7	10.1%	29.2%	34.8%	20.8	5.1%	100%	3.19	1.038	1.078
Q8	11.8%	36.5%	24.1%	22.5%	5.1%	100%	3.28	1.093	1.195
Q9	14.1%	51.4%	19.2%	8.5%	6.8%	100%	3.58	1.051	1.104

d. A= Strongly agree, B= Agree, C= ordinary, D= Disagree, E=Strongly Disagree

TABLE X.   Q10: I know uses data in the open condition of anonymity, would change their views (whether ordinary mind choose, disagree, strongly disagree)

	Strongly agree	Agree	ordinary	Disagree	Strongly Disagree	Sum
Number of people	2	39	28	12	11	92
Percentage	2.2%	42.4%	30.4%	13%	12%	100%

TABLE XI.   The results of Non-anonymous open data

	A	B	C	D	E	Sum	Average	Standard deviation	Variance
Q11	1.1%	5.6%	8.4%	40.5%	44.4%	100%	1.79	.902	.813
Q12	5.6%	30.9%	24.2%	31.5	7.8%	100%	2.95	1.080	1.167

f. A= Strongly agree, B= Agree, C= ordinary, D= Disagree, E=Strongly Disagree

C. For the open data anonymous and Privacy

Open data in the current problems encountered, privacy is still more important issues, but also because of privacy led to the government or firms is not willing to open data, rather than open the data can be used immediately mostly belonging or valuable data, so we conducted an anonymous survey of

this part of the open data privacy, in order to facilitate future government or firms in the open data reference.

- Q7-Q10: Open data Anonymous

Principal investigator for the case where the general public is "open data if anonymous" support, but also to survey about how many people do not support an open data have included personal privacy. In addition, for the general public in the know in advance if you have purposes data, whether in support of open data situation changed. In this section there are two parts, the first part of the "The data anonymous after opening whether it will rejection", the second part, "I know the purpose data, I am willing to give up the anonymity open my personal data".

In Table IX, first, for "even in the case of anonymous open data, I still feel the violation of privacy (Q7)" issues were analyzed and found from the survey, 10.1 percent of people strongly agreed and 29.2 percent agreed that their privacy has been violated, while 20.8 percent disagreed and 5.1% of people strongly disagree that their privacy has been violated, 34.8 percent of people think there is no difference. But still noted that the nearly 40 percent of visitors: even in the case of anonymous open data, I still feel the violation of privacy, so this part of the constraint data users can go through the regulations, but not to the general public for responsible for their own data, but data should be responsible for data users to increase the general public confidence in the anonymous open data.

In the "condition of anonymity, I do not mind my personal data is open (Q8)," the survey, we can learn strongly agree the current 11.8% and 36.5% agreed have people to support open data anonymous. Disagree 22.5% and "strongly disagree" 5.1% of people do not know was open data. Therefore, in the case of open data in an anonymous, almost half of the visitors to accept open data, only 5.1% of people in extreme opposition, while the remaining 94.9 percent of the people still able to communicate changes to the open data view.

We ask the Q9 question that "I know the purpose data, I am willing to open my anonymous profile". If the general public knowing purposes open data, open data support anonymous (strongly agree, agree) to enhance the 17.2% previously not supported (disagree, strongly disagree) Anonymous "open data" people reduced 12.3 percent. The study also indicates for the original general, disagree, strongly disagree, "whether anonymous data open mind (Q10)" of people were analyzed. Originally expressed ordinary, disagree, strongly disagree accounted for 51.7 percent of the people, but "knows the anonymous use of open data" premise, 2.2% of people changed strongly agree and 42.4% of people changed agree to continue holding 13% disagree, strongly disagree with 12%, although there are still 30.4% of people hold the same attitude (see Table XII).

- Q11-Q12:Non-anonymous open data

There are two main issues of non-anonymous, the first is "no condition of anonymity, I do not mind if my data is open (Q11)," the second is "if the open data can improve the real life, I am willing to give up anonymity (Q12)." The first issue is mainly to bring out a second issue, because in the subsequent analysis of large data, may be part of the study the need for research in non-anonymity, so I chose non-anonymous chain problem under investigation.

In the non-anonymous open data survey, the results are not unexpected, only 1.1% of people strongly agree and 5.6 percent agreed to support non-anonymous open data, while 40.5 percent and 44.4 percent of people do Disagree people strongly disagreed support non-anonymous open data (as shown in Table XI). But in the face of open data is to improve the life of the option, the "strongly agree," the people there are 5.6% and 30.9% of people agreed non-anonymous open data, Disagree of the people there are 31.5% and strongly disagreed 7.8% of people still choose not to support, but the gap has significantly changed, the remaining 24.2% holding ordinary attitude. From the survey shows that in the general public in the know "open data" using the premise, from the original are not supported (strongly disagreed, disagree) to support (strongly agree, agreed) non-anonymous "open data", increased 29.8% While non-anonymous open from this study have increased 29.8%, but regardless of anonymous or non-anonymous, before the "open data", you must first improve the "open data" regulations to improve the safety of the general public personal information.

VI. CONCLUSIONS

The development of open data, analysis of the rise of big data, all have a great impact on our future. Current advanced countries around the world are hoping to promote open data, to achieve the White House Office of Management and Budget's "transparent", "public participation", "collaboration" The main goal and assisted the development of big data. The big data analysis is the future mainstream, we can live through big data analysis to achieve various projections, in the future can not be underestimated. Big data and open data is the most important pre-work, because of open data, for data analysis personnel have more sources of information for analysis, and for lack of resources than the analysts are also of great help.

From the research survey, currently the open data for about 30% of people and not understanding also accounted for 30%, which is open data may be some obstacles in the current development, the people do not understand very susceptible to the negative effects against the open data, therefore, the open data should be other ways to advocacy, so that people can understand the benefits of open data. In the study of open data as well as the ordinary understanding accounted for 70%, but 70% of this, whether it is open or enterprise open opposition to the government is still 30%, which is worth exploring. Because open data for national development assistance, there is no doubt, so will need to understand why there is any objection, is that the lack of current laws, or because being influenced by others, or there are other problems are unknown to us need to be aware of. We can not deny that Taiwan open data is still in its infancy, need to rely on a lot of advice and help to successfully promote open data in the future, for the development of open data the initial stage, any wrong decision, are likely to cause open data significant obstacles in the future.

And in the future regardless of the supporting measures "open data", such as legal or data use agreements, still we need to prepare well in advance in case of anonymous or non-anonymous. "Open data" by amending the law to use data or agreements, to the general public for the "open data" reassuring feeling, so that the general public "open data" is bona fide use, can change our lives, but also wary of each

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIDB - 2015 Conference Proceedings

data use for those who use the information should be cautious. In the current "open data" should also get rid of the past era of personal data only for the individual responsible for the concept into every data users are required to be responsible for each data to achieve "open data" security and ensure "open data" is only used on the future of useful research.

REFERENCES

[1] Viktor Mayer-Schonberger、Kenneth Cukier (2014)。Big Data:A Revolution That Will Transform How We Live, Work, and Think。Amazon Digital Services, Inc.。

[2] A.F. Dugas et al. (2012). Google Flu Trends: Correlation With Emergency Department Influenza Rates and Crowding Metrics. *Clinical infectious diseases*, 54(4), 463-469.

[3] W.J.J. Roberts. (2014). Factor analysis parameter estimation from incomplete data. *Computational Statistics & Data Analysis*, 70, 61-66.

[4] Ramzi A. Haraty, Bassam Zantout. (2014). A collaborative-based approach for avoiding traffic analysis and assuring data integrity in anonymous systems. *Computers in Human Behavior*. Retrieved from http://www.sciencedirect.com/science/article/pii/S074756321400483X.

[5] Xavier Tracol. (2014). Legislative genesis and judicial death of a directive: The European Court of Justice invalidated the data retention directive (2006/24/EC) thereby creating a sustained period of legal uncertainty about the validity of national laws which enacted it. *Computer Law & Security Review*, 30(6), 736-746.

[6] Bert-Jaap Koops. (2013). Police investigations in Internet open sources: Procedural-law issues. *Computer Law & Security Review*, 29(6), 654-665.

[7] Maxat Kassen. (2013). A promising phenomenon of open data: A case study of the Chicago open data project.Government Information Quarterly. *Government Information Quarterly*, 30(4), 508-513.

[8] L. Fernando Ramos Simón, Rosario Arquero Avilés, Iuliana Botezan, Félix del Valle Gastaminza, Silvia Cobo Serrano. (2014). Open Data as Universal Service. *New perspectives in the Information Profession. Procedia - Social and Behavioral Sciences*, 147(25), 126-132.

[9] Daniel Dietrich, Tim McNamara, Tim McNamara, Antti Poikola, Rufus Pollock, Julian Tait, Ton Zijlstra. Open Data Handbook。Online retrieval time：February 1,2015。from：http://opendatahandbook.org/zh_TW/introduction/index.html

[10] TH Schee(2012)。Open Campus & Open Data。July 26, 2012。From：http://fertta.com/2012/07/opencampus-opendata/

[11] Fons Wijnhoven, Michel Ehrenhard, Johannes Kuhn. (2015). Open government objectives and participation motivations. Government Information Quarterly, 32(1), 30-42.

[12] Ubaldi, B. (2013). "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives". *OECD Working Papers on Public Governance, No. 22, OECD Publishing*. Retrieved from http://www.oecd-ilibrary.org/docserver/download/5k46bj4f03s7.pdf?expires=1422016437&id=id&accname=guest&checksum=C2E887D062CE31466EF53850239DAF76

[13] Jo Bates. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. Government Information Quarterly, 31(3), 388-395.

[14] Institute for Information Industry(2013)。Open Government Data。From：http://opendata.tca.org.tw/spaw2/uploads/images/Open%20Government%20Data%202014/2014Open%20Government%20Data-U.S..pdf

[15] 「Open Government Data Principles」(2007). Retrieved from:https://public.resource.org/8_principles.html

[16] Tim Berners-Lee (2010).*Is your linked open data 5 star?* Retrieved May 16, 2014, Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

[17] "Open Government Information: current status, vision, strategy" forum。Online retrieval time: February 4, 2015. From:http://creativecommons.tw/blog/20111129#sthash.y6skQbbR.dpuf

[18] Transparency and Open Government Memorandum, THE WHITE HOUSE, from: http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment (last visited Jan.08, 2013).

[19] 「Data.gov. Retrieved Retrieved from:http://www.data.gov/open-gov/

[20] Gwanhoo Lee, Young Hoon Kwak. (2012). An Open Government Maturity Model for social media-based public engagement. Government Information Quarterly, 29(4), 492-503.

[21] Teresa M. Harrison, Djoko Sigit Sayogo. (2014). Transparency, participation, and accountability practices in open government: A comparative study. Government Information Quarterly, 31(4), 513-525.

[22] Patrice McDermott. (2010). Building open government. Government Information Quarterly, 27(4), 401-413.

[23] Open Data go forward. From : http://www.bnext.com.tw/article/view/id/34175

[24] 「Open government survey data, which people love to use?」Retrieved from:http://www.ithome.com.tw/news/89376

[25] Patrick Andreoli-Versbach, Frank Mueller-Langera. (2014). Open access to data: An ideal professed but not practised. Research Policy, 43(9), 1621-1633.

[26] [Open Data Topic] Part2 open government data five big problem. Online retrieval time: February 4, 2015. Retrieved from: http://www.bnext.com.tw/article/view/id/32352

[27] B. van der Sloot. (2015). Do privacy and data protection rules apply to legal persons and should they? A proposal for a two-tiered system. *Computer Law & Security Review*, 31(1), 26-45

[28] Ji-Jiang Yang,Jian-Qiang Lic,Yu Niub. (2015). A hybrid solution for privacy preserving medical data sharing in the cloud environment. *Future Generation Computer Systems*, 43-44,74-86.

[29] Tun-Min (Catherine) Jai,Nancy J. King. (2015). Privacy versus reward: Do loyalty programs increase consumers' willingness to share personal information with third-party advertisers and data brokers? *Journal of Retailing and Consumer Services*, Available online 31.

[30] Graham Greenleaf, Whon-il Park. (2014). South Korea's innovations in data privacy principles: Asian comparisons. *Computer Law & Security Review*, 30(5),492-505)

[31] Wolfgang Kuchinke, Christian Ohmann, Robert A. Verheij, Evert-Ben van Veen, Theodoros N. Arvanitis, Adel Taweee, Brendan C. Delaney. (2014). A standardised graphic method for describing data privacy frameworks in primary care research using a flexible zone model. *International Journal of Medical Informatics*, 83(12),941-957.

[32] Nazanin Aminzadeh, Zohreh Sanaei, Siti Hafizah Ab Hamid. Mobile storage augmentation in mobile cloud computing: Taxonomy, approaches, and open issues. Simulation Modelling Practice and Theory, 50, 96-108.

[33] C.L. Philip Chen,Chun-Yang Zhang. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.

[34] Euan MacLennan. (2012). Data protection in private practice. Journal of Herbal Medicine, 2(3), 97-100.

[35] Paul de Hert, Vagelis Papakonstantinou. (2014). The Council of Europe Data Protection Convention reform: Analysis of the new text and critical comment on its global ambition. Computer Law & Security Review, 30(6), 633-642.

[36] Sylvia Kierkegaard, Nigel Waters, Graham Greenleaf, Lee A. Bygrave, Ian Lloyd, Steve Saxby. (2011). 30 years on – The review of the Council of Europe Data Protection Convention 108. Computer Law & Security Review, 27(3), 223-231.

[37] Hazel Grant. (2009). Data protection 1998–2008. Computer Law & Security Review, 25(1), 44-50.

[38] Warren B. Chik. (2013).The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. Computer Law & Security Review, 29(5), 554-575.

[39] Rebecca Ong. (2012). Data protection in Malaysia and Hong Kong: One step forward, two steps back?. Computer Law & Security Review, 28(4), 429-437.

[40] Angela Lausch, Andreas Schmidt, Lutz Tischendorf. (2015). Data mining and linked open data – New perspectives for data analysis in environmental research. *Ecological Modelling*, 295, 5-17.

[41] Peter Conradie, Sunil Choenni. (2014). On the barriers for local government releasing open data. Government Information Quarterly, 31(Supplement 1), S10-S17.