Assessment on China's Open Government Data Platforms: Framework, Status and Problems

Lei Zheng
Fudan University
Room 619, Wenke Building
No.220, Handan Road, Shanghai, China
+86-21-55665676, 200433
zhengl@fudan.edu.cn

Feng Gao
Open Data China
Room 619, Wenke Building
No.220, Handan Road, Shanghai, China
+86-21-55665676, 200433
fenggaosh@163.com

ABSTRACT
The research establishes an assessment framework on government open data initiatives based on China's circumstance and then applies the framework to assess eight open government data initiatives in China. The study finds a number of problems in current open data practices in China. Based on the findings, the study provides relevant policy recommendations for fostering open government data practices in China.

CCS Concepts
Applied Computing • Law, social and behavioral sciences Data → Economics

Keywords
Open data; Government; Assessment; China

1. INTRODUCTION
In the process of carrying out administrative functions and managing public affairs, government agencies have collected and stored large amounts of data. Opening those data to the society could benefit the whole society while government data are extensively and effectively used and shared. Opening government data refers to making government data accessible to anyone and to be used and shared freely. Open government data will not only improve government transparency and governance capacity, but also promote economic development and social innovation.

Since the US and the UK launched their national open data platforms in 2009 followed by other countries such as Canada, France, Norway, Kenya, South Korea, Singapore and, a wave of open data movement was set off globally. In China, Beijing and Shanghai governments are the first two local governments who carries out open government data practices in 2012. Since Shanghai's pilot project named "Shanghai Municipal Government Data Service Platform" launched in June 2012, other cities such as Beijing, Nanhai, and Wuhan have also launched their open data platforms successively. On the national level, the development of a centralized national open data platform has also been put on the agenda.

The progress of open government data can hardly be done without policy support, management improvement and resources investment by government. Thus, continuous tracking and assessment of the progress of China's open government data practices could help governments and various interested parties to identify gaps, find out problems and ultimately foster open government data practices in China.

Currently, the "Open Data Barometer" [1] assessed by the World Wide Web Foundation and the "Open Data Index" [2] organized by the Open Knowledge are the two open data assessment projects widely recognized by open data communities. In the Open Data Barometer 2014, China was ranked 46 out of 86 countries in total, while in Open Data Index 2014, China was ranked 57 out of 97 in total. However, these two international assessments are focused on the national level, while the current practices of open government data in China, however, occur mainly at the local government level. Therefore, the two international assessments cannot be directly applied to assess the current status of China's open government data practices in local governments.

Based upon abovementioned assessment frameworks and practices, on account of China's national context, this study establishes a systematic assessment framework and then conducts an assessment on China's open government data practices in various provinces and cities. The goal is not to simply rank the practices of those local governments, instead, it intends to systematically and objectively evaluate the current status of and identify gaps in open government data practices in China, so as to provide policy suggestions for future development.

2. ASSEEEMENT FRAMEWORK AND METHODS
2.1 Assessment Framework
With three dimensions, namely readiness, implementation and impact, the "Open Data Barometer" ranks and assesses projects of open government data in different countries and regions. The "Open Data Index", however, is aimed to access whether various

dg.o '16, June 08-10, 2016, Shanghai, China
 ISBN 978-1-4503-4339-8/16/06..$15.00
DOI: http://dx.doi.org/10.1145/2912160.2912213

national governments open their ten key datasets. Based on existing assessment projects such as "Open Data Barometer", "Open Data Index" and "Open Data 500" [3], the World Wide Web Foundation and the Governance Lab at New York University summarized and abstracted a common assessment framework [4] for open data. The framework includes the following four dimensions:

1) Context: the context and environment within which open data is being provided, such as the legal and regulatory environment; organizational context; political will & leadership; technical capacity; the wider social environment; and the commercial environment and capacity of firms to engage with open data.
2) Data: the nature and qualities of open datasets. Including the legal, technical, practical and social openness of data, and issues of data relevance and quality.
3) Use: the context of use of the open dataset including the category of users accessing the dataset, the purposes for which the data will be used, and the activities being undertaken.
4) Impact: the benefits to be gained from using the open dataset. Potential benefits can be studied according to social, environmental, political/governance, and economic, commercial dimensions.

Based on the Common Assessment Framework, and by taking into account that the development of China's open government data practices is still in their initial stage and the usage and impact of open data are too early to be observed in China, the research focuses its assessment on government -- the supply side of open data first. Thus, the assessment framework in this study will focus on the three dimensions of "Context" and "Data" and "Platform" including thirteen indicators, so as to comprehensively and systematically assess China's open government data practices in various regions (see Figure 1).

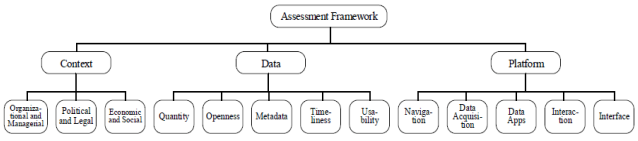


Figure 1. The Assessment Framework for the Study in China

2.2 Data Collection

Based on the aforesaid assessment framework, the research mainly uses three methods for data acquisition, namely, document analysis, data scraping and manual observation. The cut-off date of data collection is May 20, 2015, and all data are collected from open channels.

1. Document Analysis

Taking the assessment on the "Context" dimension as an example, the organizational context, management structure, the environment of the IT industry and innovation industry, document analysis is adopted. Relevant government policy documents and statistical yearbooks are searched and studied.

2. Data scraping

Regarding to indicators with the "Data" dimension, the research team scrap data from each open data platforms to automatically collect all relevant information about datasets on these platforms. The information collected includes titles of dataset, publishing agencies, dates of update, the frequency of update, downloading volume, view volume, etc.

3. Manual Observation

In view of data that cannot be automatically collected (for example, whether a data platform provides the data rating function), the research adopts the method of manual observation. Researchers visit the platform websites and manually observe and record various functions against the indicators in the assessment framework.

2.3 Sample Selection

The study first selects the open government data practice in six regions, i.e. Beijing, Shanghai, Wuhan, Wuxi, Foshan, Nanhai through the means of media and official reports. The study then selects open government data practices in other regions through the Baidu search engine by using the key words of "数据+gov.cn" and "data+gov.cn". In consideration of the maturity of local open data practices, accessibility of relevant information, diversity of samples in terms of administrative levels and regions, the assessment decides to selects eight regions (i.e. Beijing, Shanghai, Guizhou, Wuhan, Wuxi, Zhanjiang, Haishu District of Ningbo City, Nanhai District of Foshan City) from 13 regions as assessment samples. (See Table 1).

3. FINDINGS

3.1 Data

As of May 20, 2015, the open government data samples selected in this research involves opening of 1963 datasets in total. Among them, Wuhan is the city that makes the most opening in number (635), while Guizhou is the least (17). As for the openness of data, open data should be provided in machine-readable formats (e.g. XLS and non-PDF), accessible and downloadable. Based on such requirements, Shanghai is the city who released the most machine-readable data in number (398) among other regions. On average, 81.1% of open data conform to the requirements of open data formats (i.e. machine-readable). At present, however, only Beijing and Haishu District of Ningbo City provide 100% machine-readable data, while other regions still have room for improvement (see Figure 2).

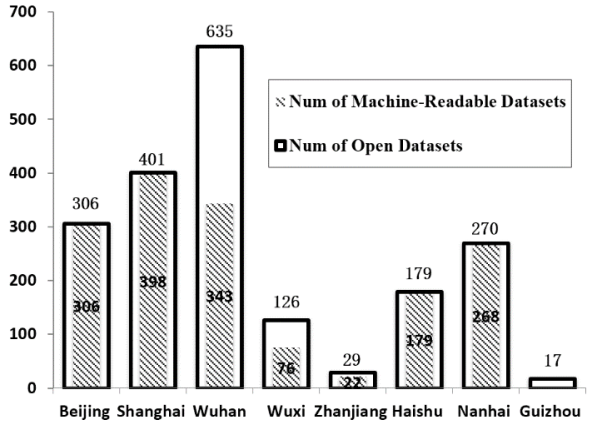


Figure 2. Number of total open datasets and number of total machine-readable open datasets

An open file format is a published specification for storing digital data, which can be used and implemented by anyone. An open

format can be implemented by both proprietary and free and open source software [5]. The purpose is to ensure that a data file is accessible without a specified, especially a paid, software. Based on the standard, the CSV format is considered a common open format, while XLS is not.

At present, only Beijing city provides open formats for all data on its platform, and the Naihai District of Foshan City provides open formats for 98.15% of its data, while other samples don't provide open formats at all. The study also notes that currently only Nanhai District provides a variety of formats (CSV, XLS, TXT, XML and JSON) in its data opening. Such way of providing data can better facilitate data users with different needs to make use of the datasets more conveniently and efficiently.

Open license is another measurement for assessing data openness. It is found that, except Guizhou, all regions and cities selected in the study have provided a license on their platforms. An open data license usually contains exemption clause or user agreement with regard to datasets provided. Only two regions, Wuxi and Haishu District of Ningbo City, ensure that the datasets provided on their platforms will be permanently free, while other regions only set a relatively vague statement about the duration of free usage and inexplicit indication on how datasets will be authorized after expiration. Furthermore, open licenses in studied samples do not clearly guarantee users' right of free use, especially the right of commercial use. Similar problems exist with regard to users' rights of free dissemination. In some regions and cities, a license even contains clauses such as "distribution of the data resources obtained from the website is not allowed, whether paid or unpaid", which does not conform to the requirement of open license either.

The frequently of datasets update datasets is also an indicator for the timeliness of datasets. The research conducts a statistical study on the update frequency of datasets committed by local platforms, except for Nanhai District and Guizhou province, which do not commit an update frequency. The study shows that overall 86.25% of datasets seem to be static data which are updated annually as needed only. Only 13.75% data are dynamic data, which are updated monthly, weekly, daily, or in real time. (See Figure 3).

The research team compares the actual update frequency of datasets against their committed frequency of update, and finds that most regions mostly provide static data, and few regions update datasets in compliance with their commitment. The result shows that among all samples, Wuxi is the city with the highest proportion (62.5%) in compliance with their commitment of datasets update, while the average proportion of all regions is only 17.21% (See Figure 4).

Table 1. Samples selected for assessment

Research objects	Administrative level	Located area	Name of open data platforms	Website of open data platform
Beijing Municipality	Municipality directly reports to central government	Northern China	Beijing Government Data Resources Portal	www.bjdata.gov.cn
Shanghai Municipality	Municipality directly reports to the central government	Eastern China	Shanghai Government Data Service Portal	www.datashanghai.gov.cn
Wuhan City	Sub-provincial city	Central China	Wuhan Government Open Data Service Portal	www.wuhandata.gov.cn
Wuxi City	Associate sub-provincial city	Eastern China	Wuxi Government Data Service Portal	opendata.wuxi.gov.cn
Zhanjiang City	Prefecture-level city	Southern China	Zhanjiang Data Service Portal	data.zhanjiang.gov.cn
Haishu District of Ningbo City	District under the jurisdiction of sub-provincial city	Eastern China	Haishu District Open Data Platform	data.haishu.gov.cn
Nanhai District of Foshan City	District under the jurisdiction of prefecture-level city	Southern China	Nahai Data	data.nanhai.gov.cn
Guizhou Province	Province	Southwestern China	Guizhou on the Cloud	www.gzdata.gov.cn

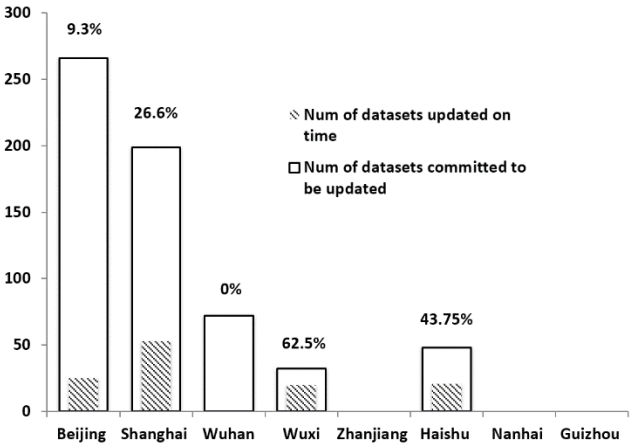
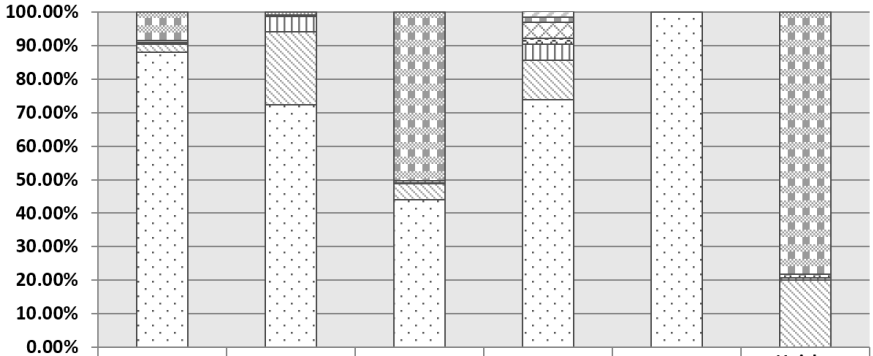


Figure 4. Number and proportion of dataset updated as committed

The practices of sample regions in providing metadata of datasets also vary. At present, all sample regions are able to provide basic metadata information, including the data name, release institution, data classification and data description, etc. This research also investigates whether sample regions provide certain key metadata information. The first type of key information is the release date of datasets. For such information, Nanhai District and Haishu District of Ningbo City do not clearly indicate, and it can only be calculated through historical data records provided. The second type is the update time of datasets. For such information, currently only Haishu District and Nanhai District indicate clearly in the data page, while Beijing, Shanghai, Wuxi, and Zhanjiang only provide them in the data directory instead of data page, which is inconvenient for users to view those information. Until the study period, Wuhan City and Guizhou Province has not provide metadata yet.

Municipality and Nanhai District do not require real-name registrations, while the rest all require users to provide real ID in order to register.

Other than opening original government datasets, open data platforms are also required to display applications that utilize open government data. At present, all sample regions have set up application channels to showcase data applications on their platforms. Among them, Beijing and Haishu District not only showcase data applications, but also listed the datasets used by those application respectively. As of May 20, 2015, totally 158 applications are listed on the eight platforms selected in the study with an average of 20 applications per platform (See figure 5). Among them, Shanghai accounts for the largest (73 applications), while Haishu District accounts the smallest (0 application). Except Shanghai, all sample regions allow developers to upload applications, so as to motivate users to develop and utilize



	Beijing Municipality	Shanghai Municipality	Wuhan City	Wuxi City	Zhanjiang City	Haishu District of Ningbo City
No update	0%	0%	0%	1.60%	0%	0%
Update as needed	8.50%	0.70%	50.40%	1.60%	0%	78.20%
Update promptly	0.70%	0%	0%	0%	0%	0%
Update in real-time	0.30%	0%	0.60%	4.80%	0%	0%
Update daily	0%	0.50%	0%	1.60%	0%	1.10%
Update weekly	0%	4.50%	0.20%	4.80%	0%	0.60%
Update monthly	2.60%	21.90%	4.70%	11.90%	0%	20.10%
Update yearly	87.90%	72.30%	44.10%	73.80%	100%	0%

Note: Nanhai District of Foshan City and Guizhou Province are not included since they do not state the data update frequency.

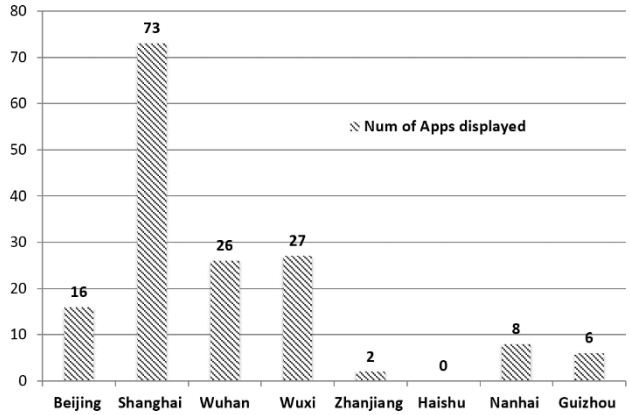
Figure 3 Update frequency distribution of open government data

3.2 Platform

In the platform dimension, the research also assesses whether open data platforms provide data navigation to help users explore data resources. It is found that all sample regions provide search functions. Except Guizhou Province, all regions provide navigation by category, through which users can navigate by release institutions and/or data categories. Shanghai, Wuhan, Zhanjiang also provide sort and search functions according to specific attributes such as download volume and update dates.

Currently, only datasets on open data platforms of Shanghai, Wuhan and Nanhai District are able to be previewed to have a glance of sample datasets before fully downloading them. As for the convenience of downloading datasets, currently the downloading processes of Wuxi and Zhanjiang are the most convenient ones, which enable the users to download datasets freely without going through the registration and login process. Among the five regions which require registration, only Beijing

government data actively.



Note: Haishu District of Ningbo City is only available for test data, so the number is 0.

Figure 5. Number of applications in application channels of sample regions

At present, all data platforms in sample regions except Wuxi City provide interaction function with users, including the datasets review rating function and datasets request function. In view of datasets review, the platforms of Shanghai and Zhanjiang allow users to submit evaluation without registration and login, whiles platforms in other sample regions require users to log in first before they can submit corresponding reviews.

In terms of review methods, Wuhan and Zhanjiang adopt star rating, Beijing, Haishu District and Nanhai District allow text-based comments, and Shanghai adopt both of them. For data request function, all sample regions except Wuxi allow users to submit data needed. However, according to the results of testing conducted in the study, all regions fail to respond to test requests within one week. Moreover, among all regions, only Haishu District made users' datasets requests public so that they can also be viewed by other users.

As for interaction, the research also focuses on whether it is the competent organizations that create social media accounts related to "big data" "open data" to publicize and promote the projects of open government data and provide communication channels for the public. According to the assessment results, now only Nanhai District of Foshan City created Weibo and WeChat accounts for its open government data projects. Through such social media accounts, it pushed the projects related news information (including the management system, policy, platform update, etc.). However, on the "Nanhai Data" platform of Nanhai District of Foshan City there is no link to the social media accounts currently.

3.3 Context

In respects of contextual and environmental support, the development of open government data practices need support from high-level administrative executives. Shanghai and Nanhai District are at the forefront in this respect among other regions in the study. The study observed whether local open government data practices receive open support from top local administrators. The deputy mayor of Shanghai and the former District Committee Secretary of Nanhai District once voiced in public to support opening data, which was also reported by the news media. For other regions, though there are also evidences of supports from administrative leadership in internal government meetings, public support from local administrators has not been observed yet.

Shanghai is the only region that publishes its open data work plan publicly. "The Annual Work Plan of Sharing and Opening of Shanghai Government Data Resources" released by Shanghai not only makes clear the overall strategy and annual target, but also clearly lists key areas for data opening. The work plan also explicates the related supporting projects such as enactment of government data sharing policies and platforms for opening government data. In other regions, however, only generic policy documents related to Smart City and Big Data initiatives are made in public. Neither corresponding policy is formulated specially for the sake of open data, nor relevant work plans is made.

Lastly, from the point of view of management structure, the current open government data initiatives are generally managed by the information-related department in government, for example, the economic and information commission. However, the administrative level of those departments is in same with other functional departments. Only Nanhai District and Shanghai municipality made progress to some extent in terms of their management structure. Nanhai District has established a Data Coordination Bureau to plan and manage all data-related business in government. Meanwhile, Shanghai also converted the previous Division of Information Technology Application (a division of Shanhai Municipal Commission of Economy and Information) into the Division of Big Data Development in early 2015, to which the task of open data is clearly designated.

4. CONCLUSIONS

Based on the comparative analysis of assessment, the research summarizes the main problems existing in China's open government data practice from six aspects.

1. Low quantity, low value, and low machine readability.

Overall, the size and value of datasets opened by local governments in China are still far from meeting social needs, and are unable to meet the requirements of economic development and social innovation. A large number of data which are high in value and does not involve national security, commercial secrets and personal privacy are not opened to public. In addition, the data format adopted determines the technical openness of datasets. The open government data practices in sample regions have not achieved machine-readable formats for all data released. This would make it difficult for data users, especially non-technical users to fully utilize datasets to create value out of them. The research found that all datasets provided by "Guizhou on the Cloud" are not available for downloading, and the 17 datasets available in the portal are all linked to external websites rather available locally in the same website.

According to the study, currently only two sample regions provides open formats for datasets. In the respect of data openness, China's local practices of open data still has much room for improvement.

2. Lack of frequently updated data.

Open government datasets are supposed to be dynamic and available in real time. Dynamic data of high value are important to inspire enterprises and individual developers to create values by making full use of data. However, it is found in the research that, among the data currently released by the sample regions, 86.25% of them are static data, and only 13.75% of them are dynamic and update monthly, weekly, daily, or in real-timely, which are far from satisfying and inspiring users' demands and interests.

Data updating and maintenance are an important task in open government data projects. It is found in the research that, among dynamic datasets that are supposed to be update regularly by open government data projects in the eight regions, only 17.21% of datasets were updated in compliance with their commitments. This makes those open datasets fail to be considered and used as stable and reliable data resources.

3. Inexplicit open data license.

The open license of data is an important elements for open data ecosystem. It not only can be used to ensure free use and spread of open data, but also is good for the public to understand open culture, so as to develop a business model based on open license. It is found in research that all sample regions adopted a website disclaimer/user agreement that provides one-time overall

authorization for all open datasets released. Relevant clauses do not conform to the principles of legal openness for open data. Specifically, problems exist in two aspects. First, some licenses in sample regions contain clauses that obviously violate the principle of openness of dataset. Second, the licenses are often vague in terms of user's rights, and failed to definitely and clearly empower and guarantee users' rights.

4. Lack of convenient channels for data accessibility.

In respect of data accessibility, current open data portals in China lack convenient channels for data acquisition. Most platforms require a prior registration and login processes, which are quite complicated and are not conducive to users' quick and easy access to the data, and thus imposes a negative impact on the motivation and satisfaction of data usage. Moreover, most open data platforms in sample regions do not provide the function of data preview, or only provide basic picture or external links, which are not sufficient for convenient data acquisition for users.

5. Lack of high quality data applications.

In general, applications provided by the open data platforms in sample regions are small in volume. Although some platforms have opened data application channels, those applications are not available for downloading and using. Although some platforms provide downloadable and useable data applications, most of them are developed with datasets provided on the platforms, or do not indicate in detail which datasets are used in developing a specific application. Besides, for some platforms, even though they provide the function of application submission, they do not require those applications submitted to be developed based on open data available on the platform, which cannot effectively promote sufficient usage of open data by the society. For some platforms, the submission of an application is limited to a few approved corporate users. General individual users who are permitted to submit applications are small in volume. which may greatly frustrates users' enthusiasm and creativity of utilizing open data.

6. Lack of convenient, timely, effective and open interaction with users.

In respect of interaction, open data platforms in most sample regions provide evaluating functions for certain datasets and data request functions, and establish dialogue mechanisms between users and governments. However, most interaction functions are convenient for use. In most cases, the users must first register and log in to be able to use. Secondly, the interaction functions on those platforms lack timely and effective responses. Although users' evaluations can be submitted, they have never been make public on the websites. Whether they are actually submitted, or are still under review, or didn't pass review remains unknown. It is also found in the test that most data request functions on theses platforms fails to make timely and effective responses. According to the study, only the platform of Haishu District lists feedback and comments received from the public, while other sample regions close these information in closet, which is not helpful for encouraging public's participation.

5. RECCOMENDATIONS

According to the status and main problems in the open government data initiatives in the sample regions, seven suggestions are proposed for the development of opening government data in China.

1. Enhance management structure and enact policies and work plans.

Currently, many regions are short of effective management architectures, policies and plans targeted at open government data projects. First of all, it is recommended to establish or designate competent departments to take charge of open government data work, and empower those departments with full authorities to integrate data provided by other business departments. Meanwhile, division of work and responsibilities of business departments should also be clarified. Second, it is recommended to formulate practical policies, specifications and work plans, so as to define principles, requirements, objects, forms and boundaries of open government data to ensure the normalization and systematism of data opening. Finally, it is suggested to prepare annual work plans and make them public, as an effort to enhance social public's awareness of and participation in open government data.

2. Strength leadership support.

It is far from being enough to leave departments in charge only to promote open government data projects. Explicit support from local administrators is no doubt a critical factor in driving open data initiatives. Prime Minister and Vice Prime Ministers of the State Council have shown their support for open government data on different occasions. In this research, it is also found that local administrative leadership in Shanghai and Nanhai District of Foshan City also have declared in public their support for open government data. It is suggested that local administrators should deepen their understanding of the purpose and significance of open government data and strengthen their support for opening government data in important local conferences, public addresses and daily works, setting up a culture of open government data, increasing the awareness of open data, improving the capabilities, and providing guidance and support for departments in charge to advance open data.

3. Open high-value datasets and display data applications

It is recommended that open data initiatives should be oriented toward users' demands, and bring together data release departments and data demanders (such as enterprises, individual developers and reporters etc.) to communicate face to face on data requests in round-table meetings so as to enhance the quantity, quality, form and value of open data. Besides, data release departments should also collect public demands through mutual communications via websites and social media, to consistently open high-value data and display applications developed based on open data.

4. Improve data openness and machine-readability.

It is recommended that open government data initiatives, whether still under development and to be developed in the future, should improve the machine-readable proportion of open data, that is, to ensure that data are not released in formats like PDF, web pages or pictures etc. This requirement should not only be embodied local policy documents and work plans of open government data initiatives, but also be taken as a significant indicator to assess all data released in various initiatives. Trainings should be provided to explain the meaning and goals of machine-readable formats, introduce common machine-readable formats, and to provide corresponding tools to help transform non-machine-readable data into machine-readable forms. Furthermore, all initiatives should establish approval process to strictly examine data formats before they are released with both automatic and manual methods, so as to ensuring that released data satisfy the requirement of open formats.

5. Update data timely.

Open data initiatives should establish corresponding specifications and supervision mechanisms to ensure that data are updated timely. For example, automatic inspection procedures can be established in open data platforms to automatically send out updating alarms to corresponding departments, or reporting buttons could be set on data pages to allow the public to report data that are failed to be updated timely. Given the cost and pressures of operation and maintenance, data releasing departments may allow social organizations or enterprises to take the responsibility of maintaining data. Through public-private partnership, dynamic data can be consistently provided to the society.

6. Enable open license.

Open data initiatives should work out open license that are applicable to China. The license should be in line with existing law of China, but also explicitly ensure users the rights to freely obtain, use, and distribute data. Government should highlight the importance and necessity of open license and facilitate data releaser and data users to develop open license mutually.

7. Lower threshold for data accessibility.

Open data platforms should lower thresholds for data accessiblity and interaction with data holders, enabling users to acquire data and participate in interactions without having to register. Besides, the users' assessment on datasets, data requests and suggestions shall be timely reviewed and replied. Finally, the usage of social media tools should be strengthened to timely disseminate open data news and activities to the public, in order to fully popularize and publicize open data movement and enhance public attention, participation and support. Moreover, interactions could occur not only between government and the public, but also among data users themselves to discuss and communicate on matters related to open data and inspire more ideas and generate more applications, so as to construct an active, well-functioned and sustainable ecosystem of open data.

6. REFERENCES

[1] Davies T, Sharif R, Alonso J. 2015. *Open Data Barometer - Global Report Second Edition*. World Wide Web Foundation.

[2] Open Knowledge. Open Data Index, [2015-07-01]. DOI=http://index.okfn.org.

[3] The GovLab. Open Data 500. [2015-07-01], DOI=http://www.opendata500.com/.

[4] Caplan R, Davies T, Wadud A, et al. 2014. *Towards a Common Methods for Assessing Open Data： Work Shop Report & Draft Framework*. World Wide Web Foundation.

[5] Wikipedia, [2016-01-30]. https://en.wikipedia.org/wiki/Open_format