

A Framework for the Evaluation of Healthcare Data's Value from the Perspective of Open Data

Ting Xu

School of Public Health, Shanghai Jiao Tong University School of Medicine

Danmai Wu

School of Public Health, Shanghai Jiao Tong University School of Medicine

Shu Wang

Medical information departments, Children's Hospital of Shanghai Jiao Tong University

Mingyue Wei

Medical information departments, Children's Hospital of Shanghai Jiao Tong University

Guangjun Yu (✉ yugj1688@shchildren.com.cn)

Children's Hospital of Shanghai Jiao Tong University

Research Article

Keywords: Open data, Data value evaluation, Delphi method

Posted Date: February 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-208649/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: This study is aimed at developing a framework for the evaluation of healthcare data's value, so as to provide a tool for data managers in making decisions on data openness.

Methods: In this study, the Delphi method was adopted. Firstly, a rudimentary framework was constructed following a literature review and focus group interviews, and an inquiry letter was designed. After handing out the inquiry letters to experts in related areas, the framework was modified according to the feedbacks. This process was repeated until a consensus was reached.

Results: For this inquiry 15 experts were invited; whose levels of activeness and authority were relatively high. This research produced a framework for the evaluation of healthcare data with 2 primary criteria, 7 secondary criteria and 21 tertiary criteria, after two rounds in Delphi method.

Conclusion: The framework established in this research lays a solid foundation for the identifying and evaluation of healthcare data's value, and is expected to drive the process of opening of valuable healthcare data.

Background

According to the definition by the Open Knowledge Foundation^[1], open data is "data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike". The Open Data Movement has swept throughout the world since December 2009 when U.S. president Obama issued Open Government Directive (M-10-06), ordering the federal agencies to publish at least 3 high-value data sets via Data.gov within 45 days. In June 6, 2013, 8 countries including the U.S. and U.K. issued *G8 Open Data Charter*, clarifying 5 principles and 3 joint actions, which are G8 National Action Plans, Release of high value data and Metadata mapping, to jointly drive the opening up of data in 14 areas of high value, including the area of health care^[2]. According to the information in the Datacatalogs.org run by the Open Knowledge Foundation, the website has incorporated 590 data portals in it, including UN Data, EU Open Data, Beijing Commission of Economy and Information Technology^[3].

It has been a universally acknowledged principle to identify and prioritize the open-up of high-value data in the open data movements in countries throughout the world. Not only did Open Government Directive and G8 Open Data Charter both highlighted the prioritizing of the open-up of high-value data, the World Bank also put forward the "80/20 principle", which means 20% of the data can contribute up to 80% of the public value and that people should identify and publish these data first^[4]. Some fast-moving countries and organizations have established principles for identifying and evaluating high-value data through practice. For instance, the *Open Data Prioritization Toolkit* published by the U.S. Federal CIO Council, by aggregating a list of questions, provided principles for the evaluation of value, cost and risk of open data^[5]. For another instance, in July 2016, Share-PSI of the EU published *Best Practice: Dataset Criteria* which, in the form of the "best Practice", sets out a number of criteria that can be used to prioritize the publication of some datasets ahead of others^[6].

In the healthcare and medical industry, data are highly concentrated and the need of big data in healthcare have extended far beyond the scope of diagnosis and treatment for patients. It is expected that the opening-up, sharing and application of data in the field of healthcare can create significant economic and social value. However, due to multiple characteristics inherent of it, including multiple users, privacy and complexity, there will be a lot more obstacles in the process of opening-up and sharing healthcare data. Especially, when the managers of the data are faced with requests of data from multiple parties, how should they make decisions concerning the topic of open data? Thus, this research is intended to, on the basis of complying with the laws and ethics, establishing a framework for the evaluation of healthcare data and accelerate the prioritizing of open-up of high-value healthcare data, in order to fully realize the potential of data.

Methods

This research has employed multiple qualitative research methods to establish a framework for the evaluation of healthcare data's value, which was completed through the following procedure:

1. By literature review, acquiring a basic understanding of the key factors determining the value of healthcare data. The author searched in both Chinese and English databases (including CNKI and Web of Science) using keywords including "Data Asset Valuation"/ "Data Value Assessment", before picking out from among the literature the key factors determining the value of healthcare data.
2. On the basis of literature review, the author invited experts in the relevant areas to conduct focus group interviews during which the factors were assessed, sorted out or supplemented by new factors, then established a rudimentary evaluation framework and designed an inquiry letter.
3. Sending out those inquiry letters to the experts in the relevant areas via email, then collecting, summarizing and analyzing the feedbacks of the experts. Then the author made modifications to the evaluation framework, designed an inquiry letter again, and collected advices from the experts for more times before the experts' opinions reached a consensus.
4. Data statistics and analytics. The data in this research were input via EXCEL and analyzed via SPSS 22.0.

1 The activeness of the experts participating in the research was measured through the ratio of effective inquiry letters, with the following formula: Active coefficient = (effective letters collected/ letters distributed) *100%.

2 The level of authority of the experts participating in the research was measured through the authority coefficient Cr, which is normally influenced by two factors: Ca (Expert judgment criteria) and Cs (Expert familiarity). The formula of Cr is as follows: $Cr = (Ca + Cs) / 2$. If $Cr > 0.7$, the authority of the expert is acceptable, the specific quantitative details of Cr are presented in Tables 1 and Table 2^[7].

3 The results of the expert consultation were described by mean, standard deviation, coefficient of variation and full score ratio. The higher CV is, the larger the divergence among the experts is, indicating that the key factors are subject to further modification; The higher full score ratio (ranging from 0-1, the ratio of one factor being attributed full scores) is, the more important this factor is.

Table 1
Expert judgement criteria

Judgment Basis	The degree of influence on expert judgment		
	Big	Medium	Small
Theoretical analysis	0.1	0.1	0.1
Practical experience	0.5	0.4	0.3
Peer understanding	0.3	0.2	0.1
Intuitive perception	0.1	0.1	0.1

Table 2
Expert familiarity

Degree of familiarity	Cs
Very familiar	0.9
quite familiar	0.7
fairly familiar	0.5
slightly familiar	0.3
minimally familiar	0.1
very unfamiliar	0.0

Results

1. Rudimentary framework and inquiry letter

After retrieval, 131 articles were found to be related to the research topic. Two researchers read and combed the literatures and extracted the key factors that affect the value of the data. This was followed by small-scaled focus group interviews involving 2 experts in healthcare management and 3 experts in computer information, which formed a rudimentary evaluation framework (please refer to Table 5 for details) based on the result of literature review.

An inquiry letter was designed based on the rudimentary framework, which include 3 segments: 1) an introduction to the research background, an overview of the framework and a guide as to how to answer the letter; and 2) a pool collecting the levels of agreement upon each factor which include 5 levels – “Disagree”, “Slightly disagree”, “slightly agree”, “”, “Definitely agree”, with a column in which experts can write down their advice after each of the factors; and 3) a survey about the experts’ basic information, familiarity with the research content, as well as experts’ judgement criteria (The familiarity are divided into 5 parts ranging from “unfamiliar” to “familiar”; the criteria include theoretical analysis, practical experience, peer understanding and intuitive perception, and the levels of each criteria’s weight are divided into “big”, “medium” and “small”).

2. Basic information about the experts enrolled in the Delphi method

In this research, the Delphi method involves 15 experts, whose professions include hospital managers, directors and frontline workers at information department, workers at governmental healthcare information bureau, technicians at information companies, as well as university researchers. The titles of these experts range from Senior (n=9), Sub-senior (n=4) to junior (n=1); their highest degrees range from doctor (n=4), master (n=9) to bachelor (n=2); their working age range from “≤10 years” (n=2), “10~20 years” (n=5), “20~30 years” (n=7), “≥30 years” (n=1); their specialties range from computer information (n=9), health care management (n=5) and public health (n=1) (As presented in Tables 3).

Table 3
Basic information about the experts enrolled in the Delphi method (N=15subjects)

Subject	Categories	Number	Percent (%)
Professional title	Senior	9	60.00
	Sub-senior	4	26.67
	Junior	2	13.33
Highest degree	Doctor	4	26.67
	Master	9	60.00
	Bachelor	2	13.33
Working age	≤10 years	2	13.33
	11~20 years	5	33.33
	21~30 years	7	46.67
	≥30 years	1	6.67
Specialty	Computer Information	9	60.00
	Health Care Management	5	33.33
	Public Health	1	6.67
Familiarity with the research content	Very familiar	6	40.00
	Familiar	5	33.33
	Quite familiar	4	26.67
	Fairly familiar	0	0.00

3. The level of activeness and authority of the experts

The level of activeness of the experts is measured by the active coefficient. In this research 2 rounds of inquiry letters are conducted, each round with 15 letters sent out. The letters reclaimed were both 15, resulting in an active coefficient of 100%.

The level of authority of the experts is measured by the authority coefficient. The judgement criteria adopted by the experts are listed in the table 4, and the results from the two rounds were the same. The experts' authority was proved to be at a rather high level, since the authority coefficient achieved 0.81 according to the formula, with a Ca of 0.9 and a Cs of 0.72.

Table 4
Expert judgment criteria (N=15 subjects)

Judgment Basis	Big		Medium		Small	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Theoretical analysis	5	33.33	10	66.67	0	0.00
Practical experience	10	66.67	5	33.33	0	0.00
Peer understanding	8	53.33	4	26.67	3	20.00
Intuitive perception	2	13.33	6	40.00	7	46.67

4. Results from the inquiry letters

After the first round of inquiry letters, the results of experts' appraisal of the tentative criteria are listed in table 5. According to the results, the average levels of agreement as well as consistency regarding primary and secondary criteria were relatively higher. Subsequently several changes were made to the tertiary criteria with low level of mean and full score ratio as well as a high level of CV: 1) C3, "conformity", was changed into "standard", meaning "Whether the data model, data elements, terminology are compliant with related national or regional standards"; 2) C4, "accessibility", was added, meaning "Whether there is a lag when accessing the data and volume capacity"; 3) C12, "professional title of the project leader", was removed; 4) C15, "The professional title of the decision maker", was changed into "the position of the decision maker"; 5) C18, "the data user's level of education" was changed into "the level of the individual's knowledge of healthcare", meaning "the ability of the individual to obtain and understand healthcare information, and the ability to sustain and promote health with the information"

The second round of inquiry was conducted by handing out the modified questionnaire which, after being filled in and collected, were again summarized and analyzed. From the results it can be seen that the levels of experts' agreement with the criteria as well as that of consistency among the opinions were enhanced, showing that a consensus among the experts was reached, given that the means of all tertiary criteria were larger than 4 and the CVs were smaller than 0.2, and the experts only offered suggestions concerning few details of the explanations of the criteria.

Table 5
The consistency among experts' appraisals (n=15 subjects)

Primary criteria	Secondary Criteria	Tertiary Criteria	Mean	Standard Deviation	CV	Full score ratio
A1 Inner value			5.000	0.000	0.000	100.00%
	B1 usable		4.867	0.352	0.072	86.67%
		C1 completeness	4.800	0.414	0.086	80.00%
		C2 faithfulness	4.733	0.458	0.097	73.33%
		C3 conformity	4.067	1.223	0.301	53.33%
	B2 easy to use		4.867	0.352	0.072	86.67%
		C4 accuracy	4.400	0.633	0.144	46.67%
		C5 coverage	4.467	0.640	0.143	53.33%
		C6 relevancy	4.733	0.458	0.097	73.33%
		C7 timeliness	4.533	0.640	0.141	60.00%
	B3 Important		4.667	0.617	0.132	73.33%
		C8 scale	4.733	0.458	0.097	73.33%
		C9 scarcity	4.400	0.910	0.207	60.00%
A2 value in use			5.000	0.000	0.000	100.00%
	B4 for scientific use		4.867	0.352	0.072	86.67%
		C10 level of scientific research project	4.733	0.594	0.125	80.00%
		C11 the level of the project's fund	4.533	0.640	0.141	60.00%
		C12 The professional title of the project leader	4.067	0.704	0.173	26.67%
		C13 the result of the ethical review over the project	4.467	0.834	0.187	60.00%
	B5 for managers to make decisions		4.867	0.352	0.072	86.67%
		C14 The scope of the decision to be made	4.667	0.488	0.105	66.67%
		C15 the professional title of the decision maker	3.667	0.900	0.245	20.00%
		C16 the purpose for making the decision	4.733	0.458	0.097	73.33%
	B6 for the purpose of searching by patients		4.933	0.258	0.052	93.33%
		C17 the individual's purpose for using the data	4.600	0.507	0.110	60.00%

	C18 the data user's level of education	3.733	0.799	0.214	13.33%
B7 for commercial use		4.667	0.617	0.132	73.33%
	C19 the overall strength of the entity	4.667	0.488	0.105	66.67%
	C20 the credit rating of the entity	4.333	0.900	0.198	53.33%
	C21 the commercial purpose	4.467	0.640	0.143	53.33%

5. The finalized framework

After the two rounds of inquiries were conducted under the Delphi method, the finalized framework for the evaluation of healthcare data was established, which includes 2 primary criteria, 7 secondary criteria and 21 tertiary criteria. What the specific criteria are and what they stand for are presented in table 6.

Table 6
The finalized framework for the evaluation of healthcare data's value

Primary criteria	Secondary Criteria	Tertiary Criteria	Explanation
A1 Inner value	B1 usable	C1 completeness	Whether there is anything missing in the data
		C2 faithfulness	Whether the data can faithfully represent the reality
		C3 standard	Whether the data model, data elements, terminology are compliant with related national or regional standards
		C4 accessibility	Whether there is lag when accessing the data and volume capacity
	B2 easy to use	C5 accuracy	Whether the data is granular
		C6 coverage	Whether the data can cover widely
		C7 relevancy	Whether the datasets are relevant with each other
		C8 timeliness	Whether the data is timely
	B3 important	C9 scale	The scale of the data's volume
		C10 scarcity	Whether the data is easy to obtain
A2 Value in use	B4 for scientific use	C11 the level of the scientific research project	The data is used for a project of the level of university/bureau/province/nation
		C12 The level of the project fund	The data is used for a project of which the fund is at the level of thousands/tens of thousands/hundreds of thousands/millions of RMB yuan
		C13 the result of the ethical review pertaining the project	Whether the project has passed an ethical review
	B5 for managers to make decisions	C14 The scope of the decision to be made	The data is used to support a decision for the need of an individual/this department/this unit/an upper unit
		C15 the position of the decision maker	The position of the decision maker who will use the data
		C16 the purpose for making the decision	The data is used in a decision making for internal management/enhancing efficiency/management innovation/other purposes
	B6 for the purpose of searching by patients	C17 the individual's purpose for using the data	The patients wants to search within the data for the purpose of treatment/insurance/health management
		C18 the level of the individual's knowledge of healthcare	The ability of the individual to obtain and understand healthcare information, and the ability to sustain and promote health with the information
	B7 for commercial use	C19 the overall strength of the entity	Overall strength of the entities such as a company or a plant, including R&D, innovation, scientific research
		C20 the credit rating of the entity	The credit rating of the entities such as a company or a plant

Discussion

1. The scientificity of the method

Delphi method is a research method used to appraise or predict with the help of experts' experience and knowledge, and a process in which the opinions of a panel of experts are collected through multiple rounds of questionnaires and put under effective control before a consensus is reached among the experts. The selection of experts and effective feedback lay the foundation of the scientific nature of the research, and the number of experts is recommended to be limited between 10-50^[8]. This research adopted Delphi method to establish a framework for the evaluation of healthcare data's value, with 15 experts in relevant areas invited, all of whom work in areas related with computer information or health care, and 85% of whom are of senior or sub-senior titles, with degrees of or higher than master, and whose working ages higher than 10. The authority coefficient is 0.81, demonstrating that the experts were highly representative and had solid knowledge as well as rich practical experiences. The experts were also highly active, given that the response rates of both two rounds of questionnaires reached 100%, and that the criteria less agreed upon received specific adjustment advice in the advice column. The final draft of the evaluation framework was proven to be scientific, given that the average degree of agreement and the full score ratio increased while the CV decreased to a relatively low level, after the criteria less agreed upon and inconsistently scored were modified based on the results of the questionnaires.

2. The framework and its significance

Following the two rounds Delphi inquiries, this research formulated the finalized framework for the evaluation of healthcare data's value, which includes two primary criteria, 7 secondary criteria and 21 tertiary criteria. The 2 primary criteria are "inner value" and "value in use". "Inner value", which includes 3 secondary criteria: "usable", "easy to use" and "important", means that the characteristics inherent of the data are related with the value of healthcare data. Taking one of the secondary criteria, B1 "usable", for instance, it means that the open data can more easily create bigger value in practice if the data is complete, faithful to the reality and is compliant with national or regional standards. The other primary criterion, "value in use", means that the value created by data open-up is related with the scenarios in which the data is used. This research summarized four major scenarios in which healthcare data could be used, including "for scientific use", "for managers to make decisions", "the individual's purpose for using the data" and "for commercial use". Taking one of the secondary criteria, B4 "for scientific use", for instance, it means that in scientific researches, provided that all ethical requirements are met, the data open-up can create more value if the research and its fund are of high level.

Currently, hospitals and public health institutions in China have accumulated a sea of healthcare data; yet the ability to make use of these data is relatively weak and one of the reasons of that is a lack of principles and guidance pertaining to the open-up and sharing of data. To address this problem, this research established a framework for the evaluation of healthcare data's value based on the idea of "prioritize the open-up of high-value data", drawing on the experience of previous researches and the opinions of experts. The final framework can help the managers of data to evaluate data's value from two aspects, namely, the characteristics and application of data, thus laying a foundation for a scientific and practical healthcare open data guidebook.

3. Limitations and future research

This research initiated an innovative research into the opening-up and sharing of healthcare data, and established a framework for the evaluation of data. Yet there are some limitations concerning this research. Firstly, this research did not

assign weights to the tertiary criteria after establishing the framework, so in future researches the author will proceed to assign weights to the current criteria so as to better guide data evaluation in practice. Secondly, data's value is not the only thing to be considered when opening up healthcare data, given that the risk and cost in the process also should not be ignored. With the basis of this research, the author will conduct further researches while aiming at establishing an open-data model for the evaluation of data from three dimensions, including value, risk and cost, so as to provide a theoretical guidance for the open data movement in the healthcare industry.

Conclusions

The application of healthcare big data will bring about profound revolution to healthcare models, yet the application is preconditioned by opening data in an orderly fashion and in accordance with laws and ethics. Prioritizing the high-value data has been one of the received principles in the open data movements across many countries throughout the world. Drawing insights from peer understanding and professional expertise, a framework for the evaluation of healthcare data's value was developed in a scientific method, and is expected to provide a theoretical tool with which data managers can implement this principle.

Abbreviations

UK: the United Kingdom; US: the United States; UN: the United Nations; EU: the European Union; CIO: Chief Information Officers; PSI: Public Sector Information; CNKI: China National Knowledge Internet; CV: coefficient of variation

Declarations

Acknowledgements

We would like to thank all the experts who participated in the inquiry and provided valuable advice.

Authors' contributions

Research conception and design: Yu, Xu and Wang; Literature retrieval: Xu and Wu; Carry out expert consultation and data collection: Xu, Wei and Wu; Write a draft article: Xu; Review and approval of the final draft of submitted articles: Yu, Wu, Wang, Wei and Xu.

Funding

This study was supported by National Natural Science Foundation of China (Grant ID: 71874110 and 91846302)

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

The study was performed in accordance to the ethical principle and norms for conducting medical research in China. It was approved by Department of Management Science, National Natural Science Foundation of China (Approval ID: 71874110). The study did not involve any medical examination, surgical procedure, or collection of personal health information. All experts participating in the study were informed in the preface of inquiry letter and provided informed consent.

Consent for publication

All the authors agree to publish this article.

Competing interests

All the authors have no conflict of interest to declare.

References

1. Open Knowledge Foundation. The Open Data Handbook, What is Open Data? <http://opendatahandbook.org/guide/en/what-is-open-data/>, Accessed 10 Dec 2020
2. Policy paper. GOV.UK. Open Data Charter. <https://www.gov.uk/government/publications/open-data-charter>, Accessed 9 Dec 2020
3. Data Portals. A Comprehensive List of Open Data Portals from Around the World. <http://www.dataportals.org/>, Accessed 10 Dec 2020
4. Word Bank Group. Open Data for Sustainable Development. <https://openknowledge.worldbank.org/bitstream/handle/10986/24017/Open0data0for0sustainable0development.pdf?sequence=4>, Accessed 10 Dec 2020
5. Federal CIO Council. Open Data Prioritization Toolkit. https://cio.gov/wp-content/uploads/filebase/cio_document_library/Open%20Data%20Prioritization%20Toolkit%20Summary.html. Accessed 11 Dec 2020
6. Share-PSI. Best Practice: Dataset Criteria. <https://www.w3.org/2013/share-psi/bp/dc/> Accessed 11 Dec 2020
7. Zeng G. Modern epidemiological methods and applications. 1st ed. Joint Press of Beijing Medical University and China Union Medical University, 1994.
8. Hill K Q , Fowles J . The methodological worth of the Delphi forecasting technique. *Technological Forecasting and Social Change*, 1975, 7(2):179–192.