Informed Consent in Building Big Data in Healthcare:
The Essential Role of Hubs in Curating and Disseminating Knowledge

Shinto Teramoto[1*] and Yuriko Haga[2]

[1]Faculty of Law, Kyushu University, Japan
[2]Faculty of Law, Kanazawa University, Japan
[1]E-mail: jshin768@gmail.com; [2]E-mail: y-haga@staff.kanazawa-u.ac.jp

[*]Corresponding author

Abstract

This article discusses a practical way to secure informed consent from contributors of big data in healthcare. The prevailing method of obtaining patients' informed consent to their medical treatment is "process-based informed consent." That is, a patient's informed consent is built through continuous communication between the patient and their medical team, rather than a physician's explanation and a patient's written consent isolated from their continuous communication. In contrast, the contributors of data and the builders or users of big data have little opportunity for continuous communication between them. However, this hardly justifies the builders and users of big data in healthcare to undermine the informed consent of contributors. The authors attempted to find a practical way in which we can substantialize informed consent in the course of building big data in healthcare. In order to discuss this matter, the authors employed simplified social network models, as well as conducted intensive interviews with a group of practicing lawyers. By these means, the authors found that continuous diffusion of knowledge through hubs, which curate and disseminate newly developed knowledge, in information networks to citizens effectively assists contributors, builders and users of big data to give and gain substantial informed consent.

Keywords: big data, healthcare, informed consent

1. Introduction

The social benefit of building and using big data is widely recognized, and governments are encouraging the development of big data industries (*See e.g.*, METI, 2016; NITRD, 2016). Big data is considered highly beneficial in the medical and healthcare sector for several reasons, especially when artificial intelligence (AI) is employed in its analysis. First, there are much empirical knowledge concerning treatment, but little scientific evidence (*See e.g.*, Dutfield 2004). Big data analysis is expected to accelerate the finding of such evidence (*See e.g.*, IBM, 2017). Second, big data analysis of genes, lifestyle, diet, etc. is expected to be able to make new discoveries, such as the causes of the different impact of specific diseases on patients (*See e.g.*, Newsweek, 2015). Third, big data analysis is expected to be able to identify a specific piece of pertinent knowledge from among numerous pieces of knowledge, such as the identification and diagnosis of a tumor (*See e.g.*, Engadget, 2016).

A major part of big data used in the healthcare sector is comprised of the personal health data of contributors who are ordinary citizens. Accordingly, it is impractical to build big data without obtaining the consent of the contributors. Moreover, in order to prevent the frustration, resentment or objections from the contributors that may occur later, and is likely to cause disputes between contributors and the builders or users of big data, the contributors' consent must be satisfactorily informed consent. "Informed consent" in medicine and medical research has been incrementally developed through years of discussion (*See e.g.*, Berg et al., 2001). In contrast, "informed consent" in the context of building big data in healthcare is not much discussed, or even deemed as being detrimental to building big data, although such a cynical view is criticized by some authors (*See e.g.,* Ioannidis, 2013). In light of the aforementioned background, there is an attempt to find a way in which the builders or users of big data in healthcare can secure informed consent from the contributors of data.

2.   The Development of Informed Consent in the Medical Sector

The concept of informed consent has been intensively discussed concerning the patient- physician relationship. Therefore, it is very helpful to examine the experience of medical and legal practitioners concerning legal disputes arising from unsuccessful informed consent when we consider informed consent in building big data in healthcare. In the legal disputes raised by patients against medical institutions, the patients often base their claims on the fact that they are imperfectly informed, or that the information given by their physicians misled them to have expectations that substantially deviated from the actual results of their treatments (*See*, *e.g.*, Osaka district ct., Sakai branch, 19 Dec. 2001, 1994 (wa) 1021, and Tokyo high ct., 18 July 2001, 2000(ne)3379 in Japan; Canterbury v. Spence, 464 F.2d 772 (D.C. 1972), and Schloendorff v. Society of NY Hospital, 105 N.E. 92 (N.Y. 1914) in the U.S.). In most cases, it is not possible for a physician to provide sufficient information to their patient by means of a one-time verbal explanation. A substantial informed consent is difficult to secure by means of a limited set of events, namely a physician's verbal explanation and the patient's signature on a consent form. However, a patient can be incrementally informed through the process of continuous and interactive communication with their medical team including the physicians in charge and paramedics. Thus, contemporary medical practice emphasizes not only "event-based informed consent" but also "process-based informed consent" (*See* e.g., Berg et al., 2001).

3.   Hypothesis

According to the practice of social network analysis, a society comprised multiple actors represented by a network having multiple nodes, each of which corresponds to an actor (*See e.g.*, Scott, 2017; Prell, 2011). Assuming that the society surrounding and including a contributor of data and a builder or user of big data is represented by such a network, the relationship between a contributor and a builder or user is denoted by a pair of nodes chosen from such network. Obviously, an informed consent substantialized between such parties is characterized as cooperative behaviors between them. Also, it is known that cooperative behaviors between the nodes of a social network are most likely to occur in a homogeneous network, that is, a regular network, among three types of networks: regular, random, and scale-free, and least likely to occur in a heterogeneous network, that is, a scale-free network (*See* Konno, 2011).
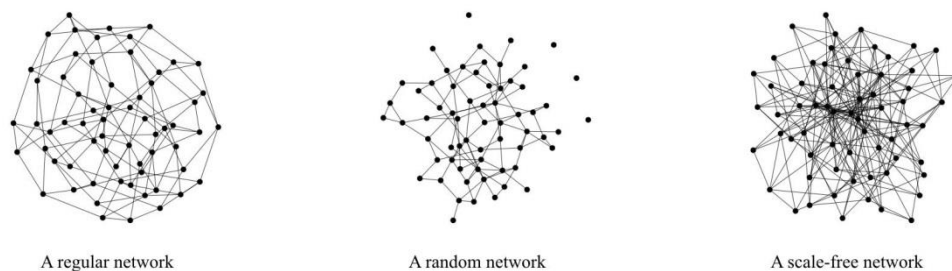

Figure 1 A Regular, Random, and Scale-Free Network

As outlined in Part 2 above, continuous and interactive communication between a patient and their team makes their own social network more homogeneous by filling the gap of knowledge between them, and substantializes the patient's informed consent. Figure 2 depicts a social network surrounding and including a patient, their physician (representing their medical team), and the sources of medical information. In the left graph of Figure 2, only the physician connected to the various sources of medical knowledge. This means the degree centrality of the physician is very high while that of the patient is very low. In the middle graph of Figure 2, the physician liaises the patient with multiple sources of medical knowledge. By this way, in Figure 3, the patient is also connected with such sources of medical knowledge. Finally, the degree centrality of the patient is increased and the heterogeneity of the network is alleviated. This seems to coincide with the knowledge concerning the cooperative behaviors between the nodes in a social network.
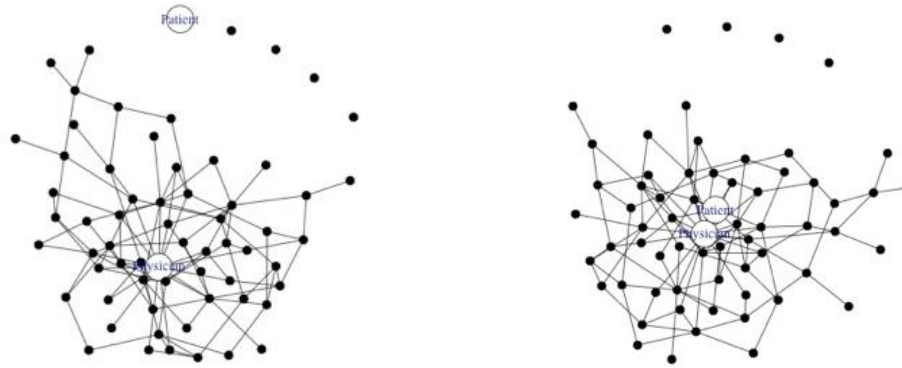
Figure 2 A Physician Helps Patients to Connect Themselves with the Sources of Medical Information

According to such understanding, we would be able to help contributors and builders or users of big data to substantialize informed consent between them, by making the social network surrounding and including them more homogeneous. However, in contrast with the patient-physician relationship, there is little opportunity for the contributors and the builders or users of big data to have continuous and interactive communication because often their relationship is only through the Internet or even through electronic gadgets that automatically transmit the contributors personal healthcare data to the builder. Such limited (or even lack of) communication between the contributors and the builders or users requires the authors to consider another way, which is an alternative to their continuous and interactive communication, but is able to make the relevant social network more homogeneous.

The authors hypothesized that, the informed consent of the contributors to the builders or users of big data can be substantialized, even when a third party other than the builders or users provides the contributors with satisfactory knowledge about the attributes and behaviors of the builders and users, as well as the scope of their usage of big data.

4. Method - Intensive Interviews with a Group of Practicing Lawyers

In order to assess the viability of the said hypothesis, we need to survey human emotional behavior in consenting to contributing personal health data in building big data under diversified conditions, including the contributors' knowledge concerning the builders or users of big data, the usage of the data, the attributes of the builders or users, etc.

The authors deemed impractical to design a fair and transparent method to conduct an experiment on such emotional behavior because the knowledge of the examinees who play the role of contributors in experiments cannot be fixed or frozen. The knowledge of the examinees incrementally increases even during the experiment unless they are isolated from the actual society. Moreover, it should be noted that such isolation infringes the rights of the examinees.

A longitudinal survey on a very large number of people, each of which is a possible contributor of data, would be desirable, at least from a purely academic perspective. However, such survey would require the authors to become a big data builder and gain substantial informed consent from the interviewees. One of the reasons why this research is being conducted is to find a way to substantialize contributors' informed consent. It is believed that, at present, there is no means of gaining satisfactory informed consent from interviewees. It would cause our own moral risk to conduct a big data survey while we are unable to gain satisfactory informed consent from interviewees. Accordingly, such survey method should not be hastily employed.

In light of these considerations, multiple use-cases were prepared and each of them represents a specific case of collecting personal health data to build and use big data in healthcare, and conducted intensive interviews with a group of practicing lawyers to examine the concerns they raised with each of the said use-cases. Such method does not necessarily produce objective results. However, it is useful because the practicing lawyers have experience in predicting people's emotional response to specific conduct of the

industries involved. Also, the risk of gaining unsatisfactory informed consent from the examinees can be avoided because they also have expertise in protecting the rights of their clients, as well as of themselves.

5. Results

The authors prepared multiple use-cases combining the respective attributes of a builder, user and the usage of big data as itemized in Table 1. Each use-case describes a situation in which the personal healthcare data of the contributors is collected by builders of big data and utilized for academic, medical or industrial purposes. It is also assumed that any data is anonymized by the builder when it is disclosed to any party, unless explicitly granted by the respective contributors.

Interviews were conducted with eight practicing lawyers, who are members of the sub-committee for the research of trust law at the Legal Research Commission of the Daiichi Tokyo Bar Association of Japan (http://www.ichiben.or.jp/english/). They were requested to discuss and give their respective opinions on whether they had any concerns with the collection and use of data in the various use-cases, assuming that they themselves were the contributors of data. Also, they were requested to inform about their respective predictions on the probability that disputes between the contributors and the builders or users of big data would arise in each use-case. The interviews were conducted a total of 9 times (13th October, 10th November, and 8th December of 2016; and the 12th January, 19th February, 9thMarch, 13th April, 11th May, and 14th September of 2017), for a period of around 90 to 120 minutes each. The opinions of the interviewees did not divert from one another although it should be noted that their opinions could have been affected through their collective discussion.

According to the consensus of the interviewees, they had little concern and estimated a low probability of disputes arising when the identity and practice of both the builder and user of big data was well known to the contributors of data. The usage of big data was well within the contributors' reasonable expectation, and the contributors could trust that the users of big data would not go beyond such expected usage. Also, their consensus indicated that the items marked with (A) in Table 1 would usually correspond to such acceptable builders, users and usage of big data. In addition, the interviewees reported that they were familiar with the attributes of or had substantial knowledge of the items marked with (A) through their personal or professional experience, as well as from newspapers, legal or other types of magazines, TV programs, and various websites. In contrast, they expressed much concern and estimated a high probability of disputes arising concerning use-cases that included any of the items marked with (C). For example, they expressed concern about the probability that a newly established fitness club would use big data in healthcare beyond the expected usage of big data. They pointed out that it was likely that the builders or users would notify the contributors of their scope of usage of big data in very general term, such as "any and all medical or healthcare research purposes." They predicted a considerable probability of disputes arising between the contributors and the builders or users due to their different understanding about whether or not a specific usage of data fell within the notified scope of usage.

Table 1 The Attributes of the Builders, Users and Usage of Big Data Included in the Use-Cases

The Builder of Big Data	The User of Big Data	The Usage of Big Data
- A health check-up company continuously employed by the contributor's employer (A) - A PHR (Personal Healthcare Record) service provider contracted directly by the contributors - A well-known academic research institution (A)	- A health check-up company continuously employed by the relevant contributor's employer (A) - A PHR service provider contracted directly by the contributors - A well-known academic research institution (A) - An established pharmaceutical company (A) - A food company - A fitness club with a national brand - A newly established fitness club (C) - An insurance company - A personal healthcare device production company - A health insurance union to which the contributor belongs (A) - The contributor's primary care physician (A)	- To provide health guidance or medical care directly to the contributor (A) - To provide health guidance to the contributor's employer (A) - To conduct academic research (A) - To publicize the product of academic research in an academic journal (A) - To publicize the product of academic research in a newspaper - To develop medical drugs (A) - To develop health food - To develop and provide fitness programs - To develop insurance products

6. Discussion

The responses of the interviewees were summarized as follows:

1) They had less concerns and estimated a lower probability of disputes arising between the contributors and the builders or users of big data, when the contributors had substantial knowledge about the attributes and behaviors of the contributors and the builders, as well as the scope of usage of the data.

2) They had greater concerns and estimated a higher probability of disputes arising, when the gap in the knowledge level was substantial between the contributors and the builders or users. These findings suggest that the contributors informed consent is likely to be substantialized, if the relevant information is diffused through the society surrounding and including the contributors, builders and users, and, in this way, the society is more homogeneous from the perspective of the respective parties' knowledge level.

The interviewees also indicated that their knowledge about builders, users and usage of data was possibly derived from a source other than the builders or users. This suggests that such knowledge is likely to alleviate contributors concerns, irrespective of whether the knowledge is provided by the builders or users, or some other sources. This suggestion generally coincides with the hypothesis that the informed consent of the contributors of data to the builders or users of big data can be substantialized, even when a third party other than the builders or users provides the contributors with satisfactory knowledge about the attributes and behaviors of the builders and users, as well as the scope of their usage of big data. The understanding on cooperative behaviors in a social network, which the authors employed to present the said hypothesis, did not take into consideration the possibility that the respective nodes would deem the same society as having different geometric structures. However, the interviewees pointed out the substantial probability that the society surrounding and including the contributors, builders and users of big data would be deemed heterogeneous by the contributors, but homogeneous by the builders and users. According to this observation by the interviewees, the probability that the builders and users would provide the contributors with only unsatisfactory information is estimated as very high. In order for the contributors to become satisfactorily informed, the role of third parties diffusing qualitative information through the society is essential. Therefore, the authors consider that our hypothesis should be adjusted to emphasize the role of such third parties to enhance the knowledge level of the citizens concerning the attributes and behaviors of the builders and users of big data, as well as the usage of data and its advantages and disadvantages to the society and the respective contributors of data.

7. Conclusion and Future Work

The discussion above clarifies the problems that the authors have to overcome in order to substantialize contributors informed consent. The authors have to enhance the knowledge level of citizens, every one of whom is very likely to become a contributor to big data in healthcare. However, it is not practical to expect the builders and users of data to convey satisfactory information to citizens. Accordingly, the authors have to depend on other means to diffuse information through the society.

It is known that the spillover of knowledge is most likely to occur in a heterogeneous network (that is, a scale-free network), and least likely to occur in a homogeneous network (that is, a regular network among the three types mentioned) (*See* Konno, 2016). In other words, knowledge is likely to spillover from a node with higher degree centrality (that is, a node connected with a greater number of nodes) to nodes, each of which has lower degree centrality (that is, a node connected with a smaller number of nodes). A new piece of knowledge created by a node with lower degree centrality is not likely to reach most of the nodes, until it is once curated by a node with higher degree centrality, which, in turn, diffuses the same piece of knowledge through the social network. This understanding of the mechanism of the spillover of knowledge coincides with ordinary experience. That is, most knowledge is derived from schools, publishers, newspapers, TV programs, curation websites, search engines and Wikipedia, or articles at SNS attracting much attention of followers. Most people have very low degree centrality while each of these sources of information has very high degree centrality. In order to substantialize the contributors' informed consent, the authors have to rely on such sources of information, or, in other words, hubs in the network to curate and diffuse knowledge (Figure 3). The irony is that the authors have to depend on a "heterogeneous" network to diffuse knowledge, in order to make the society surrounding and including the contributors, builders and users of big data more "homogeneous."

Unfortunately, the quality of information disseminated by hubs is not necessarily warranted. Objections to the appropriateness of such information can be raised by many persons, but most of them have much lower degree centrality. So, it cannot be expected that such objections will prevail and overwhelm the diffusion of incorrect information, unless such objection is curated and diffused by some other hubs. In Japan, a website called "WELQ," which had curated and disseminated healthcare information, was closed by its operating company, DeNA Co. Ltd., in November 2016 (http://dena.com/jp/press/2016/12/01/1/), after criticism of the quality of such information became prevalent through other curation sites and social network services.
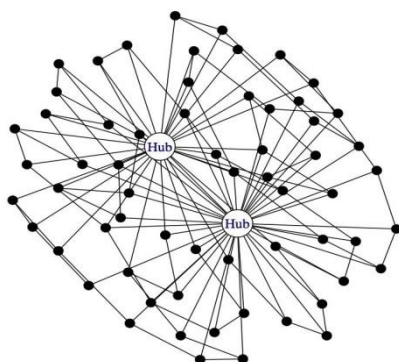

Figure 3 Hubs in a Network

However, such self-cleansing mechanism is not necessarily warranted to work. Assuming that competition among hubs fails to promote dissemination of accurate information, the intervention by law to nudge hubs to disseminate more qualitative information is justified. The next problem to be solved by the authors is to design a legal framework to encourage the hubs to curate and disseminate qualitative information.

8. Acknowledgements

9. References
Berg, J.W., Appelbaum, P.S., Lidz, C.W., & Parker, L.S. (2001). Informed consent: Legal theory and clinical practice. New York: Oxford University Press.
Dutfield, G. (2004). *Intellectual property, biogenetic resources, and traditional knowledge*. Oxford: Routledge.
Engadget. (2016). IBM's Watson AI saved a woman from leukemia. Retrieved from https://www.engadget.com/2016/08/07/ibms-watson-ai-saved-a-woman-from-leukemia/
IBM. (2017). IBM Internet things blog, Jen Clark, Is Watson the best medicine? The impact of big data analysis on healthcare. Retrieved January 3, 2017, from https://www.ibm.com/blogs/internet-of-things/iot-and-healthcare/.
Ioannidis, J. (2013). Informed consent, big data, and the Oxymoron of Research that is not research, Am I *Bioeth*, *13*(4), 40-2.
Konno, T. (2011). A condition for cooperation in a game on complex networks. *Journal of Theoretical Biology*, *19*(3), Issue 1, 224–233.
Konno, T. (2016). Knowledge spillover processes as complex networks. *Physica*, A 462, 1207-1214.

METI. (2016). Approach of Ministry of Economy, Trade and Industry of Japan about IoT, AI, and Big Data
(*IoT, AI, Big Data ni kansuru Keizaisangyosyo no Torikumi ni Tuite*). Retrieved from
https://www.itscj.ipsj.or.jp/hasshin_joho/hj_forum/files/meti20160523.pdf
NITRD (The Networking and Information Technology Research and Development Program). (2016). The
Federal Big Data Research and Development Strategic Plan. Retrieved from
https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf
Prell, C. (2011). *Social network analysis*. London: Sage Publications
Scott, J. (2017). *Social Network Analysis* (4th ed.). London: Sage Publications.