

Born digital data and its application to corporate management

Masayuki Kudo, Toshihiro Kaino

School of Business Administration
Aoyama Gakuin University
4-4-25, Shibuya, Shibuya-ku, Tokyo 150-8366, Japan
kaino@busi.aoyama.ac.jp

Abstract-On a network, various information is increasing with remarkable development of the Internet every day. While having established the status as media with a new medium called the Internet, as for the present condition, there are also many problems, such as the informational quality, the method of use, etc. of flowing there. This is a problem that is not avoided in process in which new media are born. The problem that is attracting attention in recent years is existence of “born digital” data. Although the present condition is that the agreement about the clear definition of born digital data is not made, a general interpretation is “A data which is sent on the Internet and exists only on the Internet.” The purpose of this research tends to point out the necessity for saving that information by checking existence of born digital data in the information currently exhibited on the web-page of a company useful for stake holders, such as a researcher of management and economics, and an investor, a creditor, and investigating what information exists as born digital data.

.Introduction

With developing internet technology, many kinds of information exist on the network. While internet is generalized as a new media, some problems show up. These are a quality of information, the way to use the information and so on. It is the problem in which is inevitable the process

of appearing a new media. For an instance, same problem came up to the incunabula of TV(Television). At the age, however, TV programs were not conserved systematically. That was why it was not able to be seen the TV programs in ancient age. There are various problems on the internet, meanwhile “born digital data” might be one of the most conspicuous problem now. Although there is not precise definition of “born digital data”, this is the general interpretation of the data which is sent on Internet and exists only on the internet. There was an institution of conserving information of newspaper as a media with government. As described, television stations themselves currently try to preserve the contents as valuable products for the future. An obstacle of conserving “born digital data” is a typical problem of internet which is caused by the advantage of internet enable information to be renewal simultaneously. Then the data before being renewed would be disappeared, if the information were not intended to be conserved. For example, at the search engine like “Google” might be able to get the data before renewal temporarily. But it is not preserved forever. It is mainly started to advocate necessity that each national library conserves “born digital data”. For example, this is the aim of national library of Australia. The aim of the digital services project is to provide the infrastructure for long term management of digital material in the library through provision of hardware and software systems supporting integrated collection management in a digital environment and provide the library with cost-effective technical solutions for efficient development

and delivery of digital services to provide enhanced access to the library's collections and other documentary resources, and provide shared access to digital collections in cooperation with other institutions(see Table. 1). There are two problems of saving digital data. One is how, second is what kind of data is going to be saved because of a huge amount of information on the web. It would be impossible that all data on the web are gathered and preserved by a cost and system. In addition, it is supposed to be a large amount of cost gathering, keeping and managing data on the internet by using only a search engine. There are totally 7500 terabytes' data on the web. Also there must be more than 3300 million web-pages to be hit at the end of 2003 by using "Google"[1]. It could be essential that "born digital data" is advocated how to gather and preserve what kind of data on the web for user.

. Overview of the research

The Goal of our research is that "born digital data" is found on the some web-pages to which scholar of business and economics and stakeholder pay attention and looked for what kind of the information is necessary to be preserved. It is be tried to find "a born digital data" by two ways. One is comparing each recent papers of the journal of Japan society for management information systems for finding that some web-pages from the reference could be disappeared. That means other researcher can not see it again. That will be missing important information of the people in the future. Second is looking around some corporations' web-pages from popular company top 20 of Yahoo Japan's recruiting site[2]. This is the reason that it is a main stream corporations establish their web-pages, and generally show the corporation's information to customer, client and stakeholder. Some information would be published by papers on which are changing executives, balanced sheet, and profit and loss statement and so on. But some would be showed only on the web-pages, for example, it is each branch's reports. Therefore, in that case, the data on the web-page would be disappeared, unless company tried to keep it on. The data of the company is bankrupted or merged have never been able to be seen.

Table.1 The project of each countries' and company's reserving digital information[1]

Country Or Company	The name Of project	The concept of the Project
USA	Minerva (2000 ~)	The reserving the web-site of specific topic
EU	NEDLIB (1997 ~)	Experiment and establishment of the reserving technology of informational resources on the internet
Japan	WARP (2002 ~)	Reserving the nation's web-site of specific topic
Australia	PANDORA (1996 ~)	Reserving the nation's web-site of specific topic
Sweden	Kulturarw3 (2000 ~)	Reserving and showing the nation's web-site
Internet archive alexa internet	Wayback Machine(1996 ~)	Gathering and reserving web-sites all over the world

. The basic idea of this research

We decide to be as a user in this research, because the significance of preserving the "born digital data" must be in the situation no matter who can access. It is not impossible for expert or people who connect with a corporation to be able to get information disappeared after renewal unless the corporation bankrupt. Also they might be able to get original data of its web-page in some way, which could be kept in the corporation. On the other hand, it is difficult to have the information in public. That is why this research is needed to be the sight as user. The hypothesis in this research is that

each different web-page has different quality and quantity of “born digital data”, because each different web-pages have different user as well as each user do. Some companies use their web-pages as just an advertisement. This way is useful for a company to advertise the new product instead of putting a lot of information. On the other hand, some companies think their web-pages for bringing publicity. It makes sense that they put more the IR information, strategy and so on.

. How to research

The information or contents on the web-page, which might vanish and nobody can access, are looked for. Each web-pages from reference of papers in the Journal of Japan society for management information systems is compared to. Then some digital data already disappear are noticed. We also research to distinguish each different type of corporation’s web-pages’. Then the information and contents on the web are estimated by “A”, “B”, “C”, “D”, and “E”. A means almost all information is composed of born digital data(see Table 3).

. Results of analysis

As a result of our research, it can be said that there are the born digital data. The older referred web links from papers in the Japan society for management information already disappear. Concretely 11(31%) of 36 referred web links disappear in our research which is analyzed theses from 2000 to 2003. In fact, generally there are not many times researcher refers from web-page. It is believes that saving born digital data has some importance; if our conjecture could be true that referring from web-page on a research would be increased. This is another result from our research comparing referred web links of same occupation. It is picked up manufactures because there is many web pages that includes born digital data. For example, each company makes renewal so fast, therefore people can not see the

information of old product and support again. Even though, it can not be compared the web-page with the old. Then the research can not be the conclusion which the information researched is born digital data.

Table.2 Born digital data on the journal of Japan society of management information systems

year	Vol. No.	Total number of referred papers	No. of referred web links	No. of dead referred web links
2000	No.1	42	3	3
	No.2	98	0	0
	No.3	55	4	1
Subtotal		195	7	4
2001	No.1	132	0	0
	No.2	41	0	0
	No.3	84	17	7
	No.4	117	0	0
Subtotal		374	17	7
2002	No.1	95	1	0
	No.2	53	0	0
	No.3	73	1	0
Subtotal		221	2	0
2003	No.1	74	1	0
	No.2	102	9	0
Subtotal		176	10	0
Total		966	36	11

Table.3 Evaluation index of born digital data

Index	Criteria
A	Most of web pages are born digital data.
B	Many web pages include born digital data.
C	Some web pages include born digital data.
D	Few web pages are born digital data.
E	All web pages doesn’t include born digital data.

Table.4 Analysis of manufacturer's home pages[3]-[7]

	Sony	Suntory	Shiseido	IBM Japan	NTT Docomo
*1	D	D	D	D	D
*2	B	B	C	B	C
*3	D	B	C	D	D
*4	D	D	D	D	D
*5	C	D	C	C	C
*6	C	B	B	C	C
*7	C	E	E	A	B
*8	C	C	C	C	C

*1 Information of the company

*2 Information of material of product

*3 Information of accessory of product's material

*4 News release

*5 Recruiting

*6 Service

*7 Supporting user

*8 Total born digital limits

. Conclusions

“Born digital data” is verified by our research. Although it is true that comprehensions of concept of “born digital data” actually depend on the producer of the web-page, the data which is put on the web-page for the first time is absolutely the one. If it is renewal or deleted, it is high possibility that people can not see it again. So our research could affirm that there are valuable data, which is supposed to be preserved. However, it is really difficult to say what kind of information must be saved, because our research focuses on only the theses and corporation's web-page, there is a lot of kind web-page on the internet though. Anyway it must be obligation of human history to descend information, which is valuable and have a possibility to be vanished. This is the reason why “born digital data” reflect the present clearly. And it is called for development of the technology to collect

the web site's information that cannot be gathered by the current technologies. Especially, it is required the data mining technology that gathers the data according to the purpose by soft-computing technologies.

References

- [1] T.Nagatsuka: "Long-term preservation of information sources and access to digital materials on the internet", Journal of Information Processing and Management, Vol.45, No.7, pp.466-476, 2002.
- [2] <http://fresh.employment.yahooo.co.jp/fresh/ranking>
- [3] <http://www.sony.co.jp>
- [4] <http://www.suntory.co.jp>
- [5] <http://www.shiseido.co.jp>
- [6] <http://www.ibm.com>
- [7] <http://www.nttdocomo.co.jp>