Estimate of Total Volume of Medical Data in a Year in Japan

Hiroshi Nagata[a,b], Hiroshi Tanaka[b]

[a]*Medical Informatics Project, KDDI R&D Laboratories Inc.*
[b]*Institute of Medical Science, Tokyo Medical and Dental University*

Abstract

In this study we estimated the total volume of medical information in a year in Japan, and we could conclude that it could become 3 petabyte at most in digital. In Japan, the broadband Internet has been rapidly widespreading since 2001. The new IT infrastructure enables application service provider (ASP) of hospital information systems. It is expected that the ASPs will greatly reduce the hospital information investment, but to achieve the medical ASPs, it will be necessary the internet data centers (IDC) to store medical information. However, we haven't known how much medical information is generated in the entire country. Thus we tried to estimate it in this study. Our result will be useful for not only domestic venders but also foreign venders who aim at Japanese medical markets.

Keywords:

Medical Image, Application Service Provider, Internet Data Center, PACS

Introduction

Application service provider (ASP)[1,2] is considered to become a powerful business solution in the Broadband Internet age. Banks and securities firms already start to use ASPs as data processing tools of e-trades and customer services. ASPs will become to be used for various purposes in all industries in the near future for cost reductions and service improvements. The current also could have a big influence on the medical information industry. For instance, hospitals and clinics will be probable to save investments to hospital information systems if they use medical ASPs of electronic medical record (EMR) systems, picture archive and communication systems (PACS), and receipt systems via the Internet. Medical facilities will not preserve even their medical data for themselves because all of the data will be save at the Internet Data Centers (IDC) which are linked with the ASPs. It will be also desirable in the development of telemedicine, because hospitals and clinics become easily exchange EMRs and medical images on the servers in the IDCs.

After the turn of the century, the Broadband Internet starts rapid spreading in Japan, strongly promoted by the e-Japan strategy that the government announced in January 2001[3]. The government planned to spread the ADSL communication lines to 30 million homes and offices, and the optical fibers to 10 million homes and offices by the end of 2005. The subscribers of the ADSL, CATV and low price optical fiber reached 2.4 millions, 1.5 millions, and 26 thousands respectively in the March 2002[4]. It can be expected that many hospitals and clinics will be connected with the Broadband Internet within 2-3 years so that it is certain that medical ASP businesses will start soon. Actually, some local doctor associations obtained the government funds to research and develop small-scale ASPs in the last year. Also several venders develop medical ASPs and provide them to trial uses.

However, it is necessary to obtain some basic data to establish the ASPs for business usage. Especially the amount of medical data is very important to design the ASPs and IDCs.

Several estimates of the volume of medical data generated in a hospital have been done to establish intra-hospital PACS and EMR systems[5-7]. However the total volume in the entire Japan, which is necessary to develop nation-wide medical ASPs and IDCs, has not been country hasn't been estimated yet.

In this study, we challenged this problem. We used several statistic data prepared and opened by the Ministry of Health, Labor and Welfare (MHLW), analyzed them under some reasonable assumptions, and succeeded to obtain an appropriate estimate. We obtained the result that the medical data generated in one year in the entire Japan was, if all of them are digitalized, only 3peta byte (PB) at most. It is as small volume as to preserve and operate using the existing Internet technologies. Our result will encourage both medical side and the vender side that wish to promote developing and wide spreading of the medical ASPs/IDCs in Japan.

Materials and Methods

In Japan, medical practices are classified into several thousand details, and medical insurance is paid to the every individual detail. Each medical facility totals the details every end of the month to claim money to the medical insurance organizations. The MHLW generalizes the insurance organizations in the whole country. It totals what and how many details were done and opens the total results in every June two years ago over the Internet (the most resent data is of June 2000)[8]. We thought that we could estimate the volume of medical information generated in one year in the whole country by using this statistic data.

Examinatios	Contents	Number of image taken in an examination	Data size of an image
X-Ray Radiography	Plain, macro, and contract radiography	2	8MB 2000×2000 12bit
CT	Head and Neck, trunk, and limbs	100	256KB 512×512 8bit
MRI	Head and Neck, trunk, and limbs	100	64KB 256×256 8bit
Endoscopy	Otolaryngology, bronchus, upper and lower gastrointestinal tracts, urethra and bladder, and colposcopy	20	920KB 640×480:RGB24bit

Table 1. Numbers of images and data sizes of examinations.

Imaging	Number of examinations done in June in 2000	Estimate number of examinations in 2000	Estimate number of images taken in 2000	Data volume in a year (Tela Byte)
X-Ray Radiography	9,950,000	119,400,000	238,800,000	1,910
CT	1,320,000	15,840,000	1,584,000,000	396
MRI	520,000	6,240,000	624,000,000	39
Endoscopy	1,150,000	13,800,000	276,000,000	247
Total				2592

Table 2. Data volumes of each examinations in a year in digital.

However, it is difficult to precisely evaluate how much digital data is generated by the every individual detail. It is probable that the most of the detail generates a short text or some numerical values, but the volume of data individually changes. Only the volume of medical imaging can be accurately estimated because the data size of all images of X-ray radiography, CT, MRI, and endoscopy are standardized. Medical facilities should preserve these images with clinical records for a period legally provided. Table 1 shows the number of images generated in an examination and the data size of an image. The data size of a usual X-ray image is standardized to 8MB if it is digitalized, and about 2 radiograms are taken in an examination in Japan. Thus the data size generated by each radiography is about 16MB in digital.

At the CT and the MRI examinations, numbers of images are taken at once. Especially in recent years, taking images have been increasing because multi-slice CT and helical CT have been widely spreading. In this study, we assumed 100 images are taken at once in the each CT and MRI examination. The data sizes of a CT and an MRI images are 256KB and 64KB respectively. Thus the amount of digital data generated by CT and MRI examinations are assumed as 25.6MB and 6.4MB.

In an endoscopy, it is assumed about 20 digital images are captured, because number of shots of endoscope's film is just 20. The size of a digital endoscopic image is 900MB, so the data size of an endoscopy is 18MB. The preservation terms of images of CT, MRI, and endoscopy are the same as the radiograms.

Results

General Organization of the Paper ←10pt, Times bold

Table 2 shows numbers of each imaging examination and the estimate amount of data in 2000 in the entire Japan. Though the MHLW discloses only the statistics in June, it is able to estimate the total numbers in a year by just multiplying 12 by them. It was estimated that X-ray radiography had been done more than 110million times in 2000. Total number of radiograms should reach more than 230 millions. Thus the total digital data of the radiograms is estimated as 1910TB (Terabyte). On the CT and the MRI examinations, it was estimated that 1584 million CT images and 624 million MRI images in total were generated in the year. Endoscopy was done about 13.8 million times a year, and more than 270 million still images had been taken.

The total volume of the medical images was estimated as 2592 TB in digital. Most of the other medical information seems texts and numerical data. Therefore, we can conclud that the total volume of medical information in Japan doesn't greatly excess 3000 TB, or 3 PB (Peta bites).

Discussions

We calculated the gross amount of medical images generated in one year in the entire Japan, and from the result, we estimated the total amount of medical information generated in one year in the entire Japan is probably about 3PB.

+

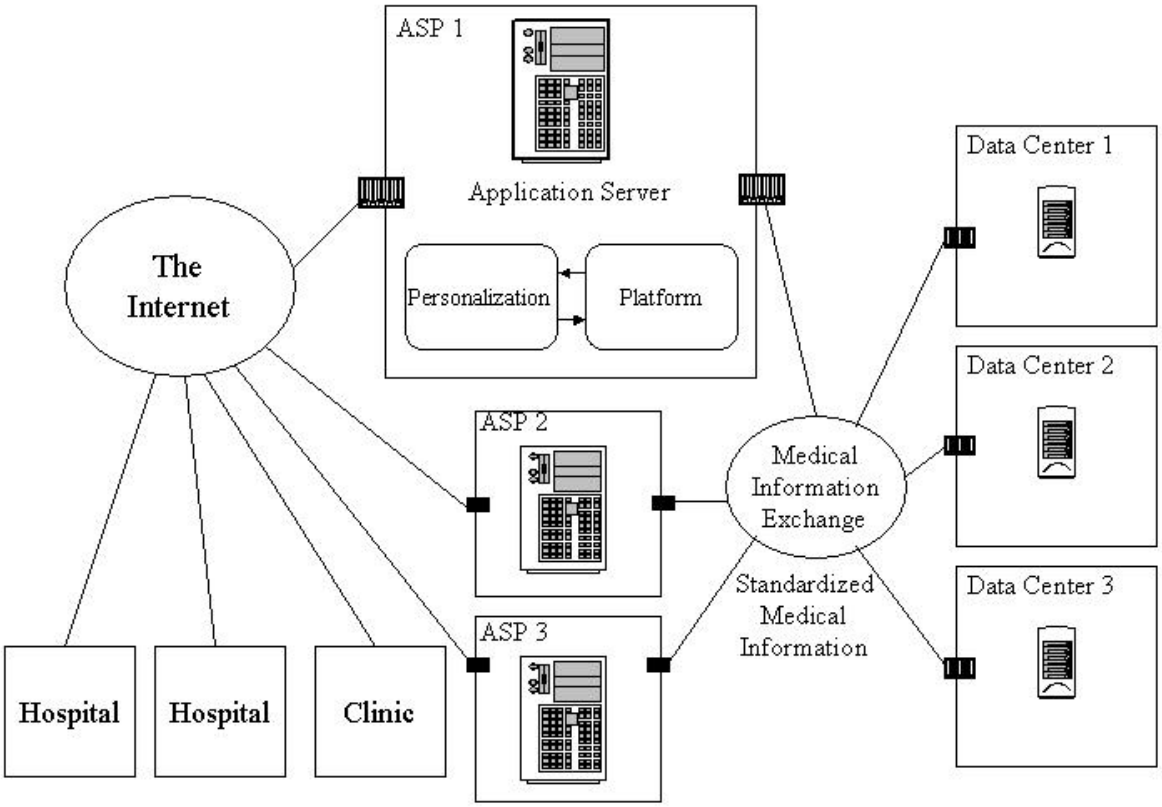

Figure 1. Conceptual image of future medical ASP/IDC system.

Doctors and engineers engaging in medical informatics in Japan have believed that it could be difficult to preserve all the medical information in digital because they believe, though there have been no certain evidence, the total volume of the data could be enormously huge. Our research overturned this common recognition.

Medical facilities have to legally preserve all the films or the digital data of radiograms for 2 years. However most of hospitals and clinics save them for more long terms because clinical records have to been preserved for 5 years. However, total data volume is only 15PB even if we preserve all the medical information for five years.

The volume is as small enough as to save all the data into only one or two thousand racks of high performance storages. Contrary, an IDC of private company can accommodate more than 1000 racks even by small-scale ones. The facts indicate that it is quantitatively possible to digitally preserve all the medical information in ASP/IDCs.

It is expected that many medical ASPs and IDCs will be constructed in near future. The next problem is how to achieve data exchanges among them. One of the most preferable solutions seems to perfectly separate the ASPs and IDCs. Figure 1 illustrates the conceptual image of our medical ASP/IDC model. The ASPs provide only medical applications to the client hospitals and clinics. The ASPs don't keep any medical data but immediately transmit the processed data to IDCs. The model can make each ASP exchange medical data between two or more IDCs. Also the model can allow the clients use many applications provided by two or more ASPs. To achieve the model, it is necessary to standardize description rules of medical information. Also it seems necessary to establish some organizations to administrate and control medical data exchange among all the ASPs and IDCs. The standardization has been studied for many years, but there is few studies on the medical data exchange over the Internet. Therefore now we are planning to challenge this problem as the next step to realize medical ASP/IDC solution in Japan.

Conclusions

We estimated the total volume of medical images using statistic data of government. As the result, we obtained the total value is 2.5PB in a year in entire Japan. Because most of the other medical data is test and numeric data, volume of which seems not so large, we concluded that the total volume of medical data in Japan probably be about 3PB in digital. This result will be useful for not only domestic venders but also foreign venders who aim at Japanese medical markets.

References

1. J. Gunson, D. Booth, and S. Thurston, ASP Configuration Handbook, Syngress Media Inc. (2001)
2. J.W. Toigo, The Essential Guide to Application Service Providers, Prentice Hall PTR. (2001)
3. e-Japan: Summary, (2001), Prime Ministerof Japan, http://www.kantei.go.jp/foreign/it/network/0122summary.html
4. 2002 WHITE PAPER Information asd Communications in Japan, Ministry of Public Management, Home Affairs, Posts and Telecommunications, http://www.kantei.go.jp/foreign/it/network/0122summary.html
5. H.Kndoh, J.Ikezoe, Y.Mori, et al. PACS in Osaka University Hospital, Computer Methods Programs Biomed. 43(1-2), 57-63 (1994)
6. I. Takeda and T.Kudo, Network System: the integrated picture archive and communication system with the hospital information system, J Digit Imaging. 4(4 Suppl 1), 6-14 (1991)
7. H. Okaniwa, H. Tsuneyoshi, S. Kabata, et al. Hospital-wide OACS wuth a digital image intensifier TV system. Comput Methods Programs Biomed. 43(1-2), 75-9 (1994)
8. Syakai Iryo Shinryo Koui Betu Tyosa 2000 (Annual Report of Medical Insurance, 2000), Ministry of Health, Labor and Welfare, http://wwwdbtk.mhlw.go.jp/IPPAN/ipcart/scm_k_Ichiran (Japanese only)

Address for Correspondence

Hiroshi Nagata is a senior research scientist of KDDI R&D Laboratories Inc. Also he is a invited research scientist of Tokyo Medical and Dental University, and Communication Research Laboratory. He graduated Department of Chemistry of University of Tsukuba and got Master of Science Degree. Then he engaged bio-informatics and medical informatics research at Olympus Optical Co. Ltd. During 1996 to 2000 he had proceeded to National Cancer Center Research Institute. After that he changed occupation to the company now he works. His interested area is very broad, from e-health to bio-informatics. Interested readers may contact the author, via either nagata.com@mri.tmd.ac.jp or Medical Informatics Project, KDDI R&D Research Laboratories Inc., 2-1-15 Ohara Kamifukuoka Saitama 356-8502, Japan.

Dr.Hiroshi Tanaka is a professor of the Institute of Medical Science, Tokyo Medical and Dental University. Also he is the head of the Information Center of Medical Science of the University. He engages vice president of Japan Association of Medical Informatics. He will become the president of the JAMI in this November. Prof.Tanaka has been very active in both medical informatics and bio-informatics fields. He has many graduate students in his laboratory. Interested readers may contact the author, via either tanaka@cim.tmd.ac.jp or his laboratoty, Institute of Medical Science, Tokyo Medical and Dental University, 1-5-45 Yusshima Bunkyo-ku Tokyo 113-8510, Japan.