Anonymization server system for DICOM images

H. Suzuki*[a], M. Amano[a], M. Kubo[b], Y. Kawata[b], N. Niki[b], H. Nishitani[c]

[a] Systems Innovation Engineering, Graduate School of Advanced Technology and Science, The University of Tokushima, 2-1 Minami-Josanjima-cho, Tokushima, 770-8506 Japan
[b] Institute of Technology and Science, The University of Tokushima Graduate School,
[c] Institute of Health Biosciences, The University of Tokushima Graduate School

ABSTRACT

We have developed an anonymization system for DICOM images. It requires consent from the patient to use the DICOM images for research or education. However, providing the DICOM image to the other facilities is not safe because it contains a lot of personal data. Our system is a server that provides anonymization service of DICOM images for users in the facility. The distinctive features of the system are, input interface, flexible anonymization policy, and automatic body part identification. In the first feature, we can use the anonymization service on the existing DICOM workstations. In the second feature, we can select a best policy fitting for the Protection of personal data that is ruled by each medical facility. In the third feature, we can identify the body parts that are included in the input image set, even if the set lacks the body part tag in DICOM header. We installed the system for the first time to a hospital in December 2005. Currently, the system is working in other four facilities. In this paper we describe the system and how it works.

Keywords: DICOM image, protection of personal data, anonymization

1. INTRODUCTION

Digital Imaging and Communications in Medicine(DICOM) specifies data structure and data exchange protocol of digital image [1]. DICOM standard is now widely-recognized in Japan. Medical information like digital medical image is used for clinical practice. And it is important information for research and education. Development of Computer-Aided Diagnosis(CAD) system is a project which makes full use of medical information [2]. The project needs to construct massive database of medical information.

In secondary use of medical information, recognition by ethics committee is essential. It requires consent from the patient, and safety research environment. A guideline for constructing the environment has been published. But, the number of information leak cases have increased in recent years. The cause of almost cases is stolen and lost of recording medium. Anonymization is necessary method to protect the personal data of the information. DICOM image not only has pixel data, but has patient and study data as character information. It is necessary to develop a system to anonymize personal data in DICOM image.

It is easy to operate DICOM image using DICOM viewers. There are many commercial and free DICOM viewers. A user can download some free DICOM viewers easily. The recent software provides simple image processing and anonymization of DICOM image. The advantage of the anonymization function is that the operation is simple using Graphical User Interface(GUI). But it has problems in Japan. The anonymization method can not flexibly comply with the policy that specified by ethics committee in each medical facility. And the workloads increase.

In this paper, we develop an anonymization system based on the following 2 points.

1. We develop an anonymization method which can flexibly comply with the policy that specified by each medical facility. In Japan, the law about protection of personal information that administrated in 2005 is widely-recognized. But there is not a public guideline which is described what is personal data and appropriate processing. We call this policy anonymization policy. Each medical facility has implemented an individually anonymization policy. There are differences between the anonymization policies. An anonymization method is necessary to fit the differences.

Medical Imaging 2007: PACS and Imaging Informatics, edited by Steven C. Horii, Katherine P. Andriole,
Proc. of SPIE Vol. 6516, 65160Z, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.709947

2. We develop an anonymization system which implements smoothly anonymization of massive DICOM images in a medical facility. DICOM viewers that provide anonymization function are already used to anonymize few DICOM images. In order to implement anonymization of massive DICOM images automatically, a new system is required.

Section 2 describes a system which implements anonymization of massive DICOM imgages automatically. Section 3 describes operation of the anonymization system.

2. ANONYMIZATION SYSTEM

We develop an anonymization system that enables automatic anonymization work. The system allows to smoothly collect of massive DICOM images shown Fig.1.

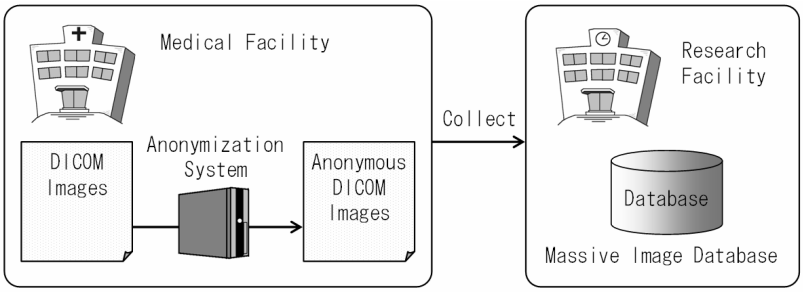

Fig. 1. Overview of anonymization system.

2.1 Anonymization method

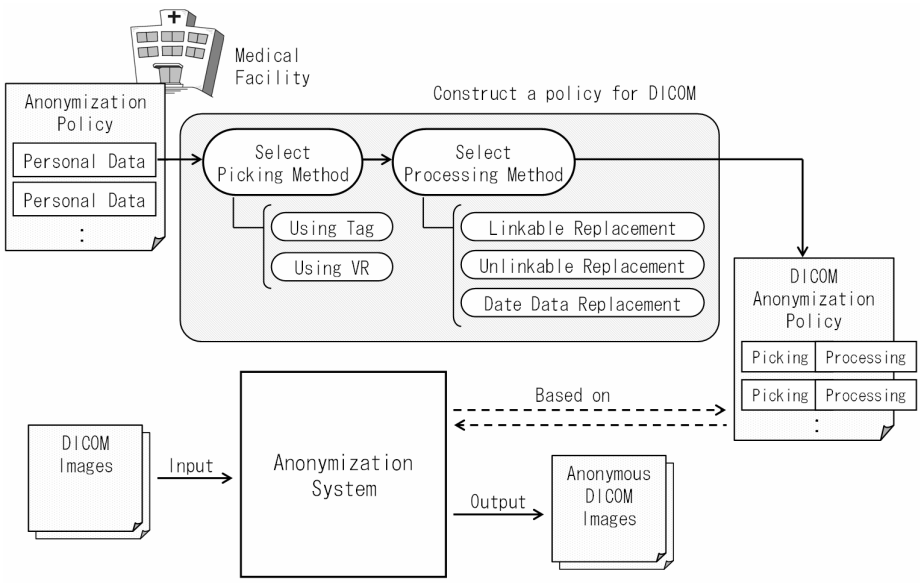

Fig. 2. Procedure of anonymization method.

Fig.2 shows the procedure of anonymization method. The DICOM anonymization policy is constructed based on the anonymization policy in the medical facility. The system anonymize based on the DICOM anonymization policy.

Here is the flow of the construction. Firstly, the picking method of the data elements that fit the anonymization policy is selected. Secondary, the processing method is selected. The picking of data elements use the tag and Value Representation(VR). The tag allows to pick one data element. The VR allows to pick same formed data elements. For example, the all date data in the DICOM image is selected by DA, and the all person name data is selected by PN. The picked personal data is classified to 2 types. One is suited to linkable replacement, the other is suited to unlinkable replacement. The date data has needed for researchers, and the data has been not anonymized. The processing method library is composed of 3 replacement methods, linkable replacement, unlinkable replacement and replacement for date data. Fig.3 shows flow of the replacement for date data. The method is available if the Patient's Age is permitted to use, and if the Patient's Birth Date exists in the DICOM image.

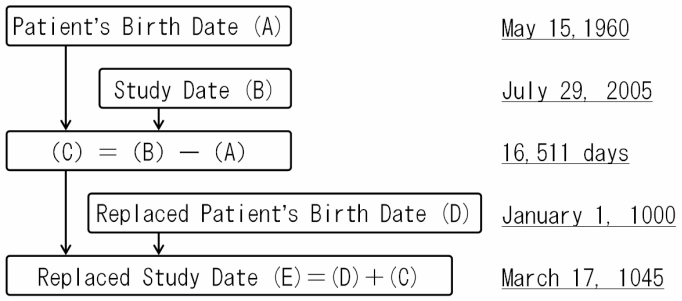

Fig. 3. Flow of replacement for date data.

2.2 Automatic anonymization specification

DICOM standard is widely used by medical information system in department of radiology. The systems are composed primarily of Radiology Information System(RIS) and Picture Archiving and Communication System(PACS). DICOM standard interface is available to smooth data exchange between machines complying with the standard.

The anonymization system implements automatic anonymization work using DICOM standard interface. Fig.4 is the overview of automatic anonymization. A study using Computed Tomography(CT) was implemented, the DICOM images that were created by the device are stored to DICOM image server. Then It is possible to transfer the DICOM images to the anonymization system from the server automatically. The automatic anonymization flow can be processed in parallel with usual routine work.

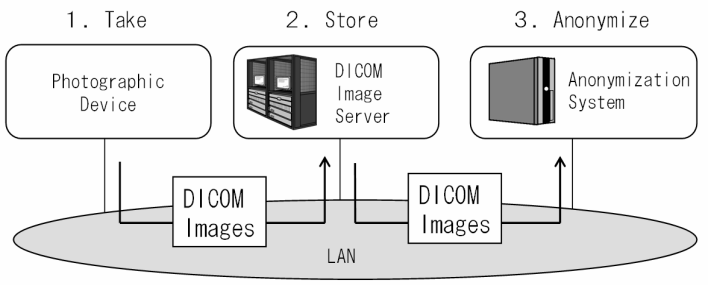

Fig. 4. Overview of automatic anonymization.

2.3 System architecture

Fig.5 is the system configuration diagram of the anonymization system.

The roles of the components are as follows.

(a) Input interface: This component receives DICOM images by data exchange between machines complying with DICOM standard. In the anonymization of few DICOM images, the GUI is useful.

(b) Preprocessing module: This component selects specific DICOM images that fit the purpose.

(c) Anonymization engine: This component anonymizes DICOM images based on the anonymization policy. This anonymizes the IDs if the text files is put into.

(d) Table operation interface: This component manages the operation of tables.

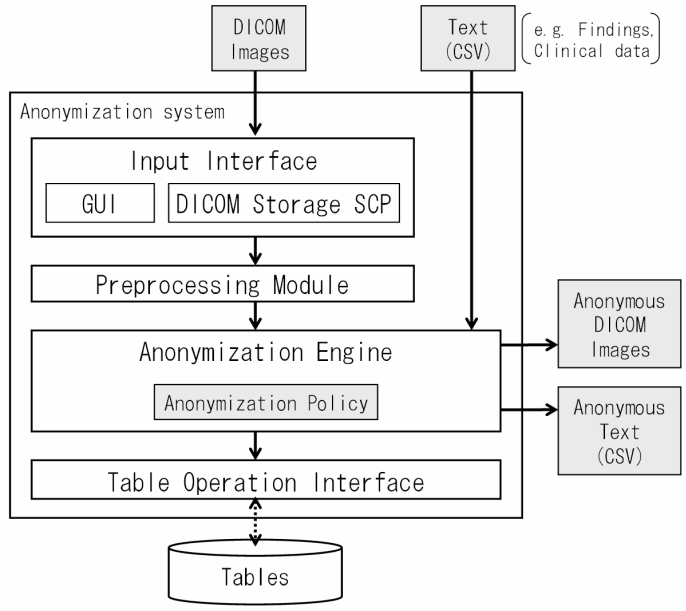

Fig. 5. Anonymization system architecture.

2.4 Input interface

The input interface is composed of 2 interfaces. One is a simple GUI that allows the users to anonymize DICOM images that stored in disk space of the computer. The other is DICOM Storage Service Class. The service class enables the system to receive DICOM images from photographic devices, DICOM image servers and DICOM work stations. The anonymization system supports Service Class Provider(SCP) as shown in Fig.6.

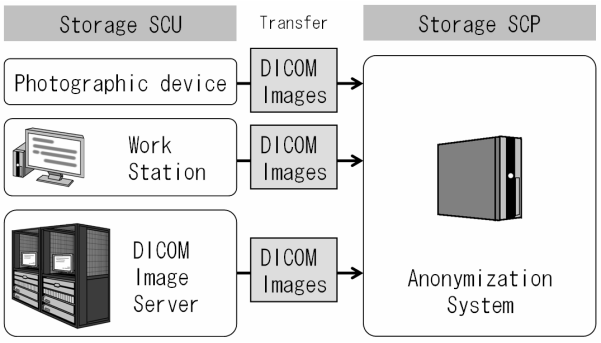

Fig. 6. Input interface using DICOM Storage SCP.

2.5 Preprocessing module

The anonymization of specific studies that fit the purpose has been needed. The anonymization preprocessing module allows to select the specific studies using the header information and simple image processing. The examples of header information conditions are as follows.

1. Body Part Examined (0018,0015) is chest.

2. Slice Thickness (0018,0050) is 1mm.

The Body Part Examined is Type 3, optional data element. The method of body part identification allows to indicate the body part of CT images. Bones and organs are depended on the body part. Some organs have air regions. We have constructed the method based on the patterns of 3 regions, body, bone and air [3]. The regions are extracted using threshold value. Fig.7 shows the overview of the method.

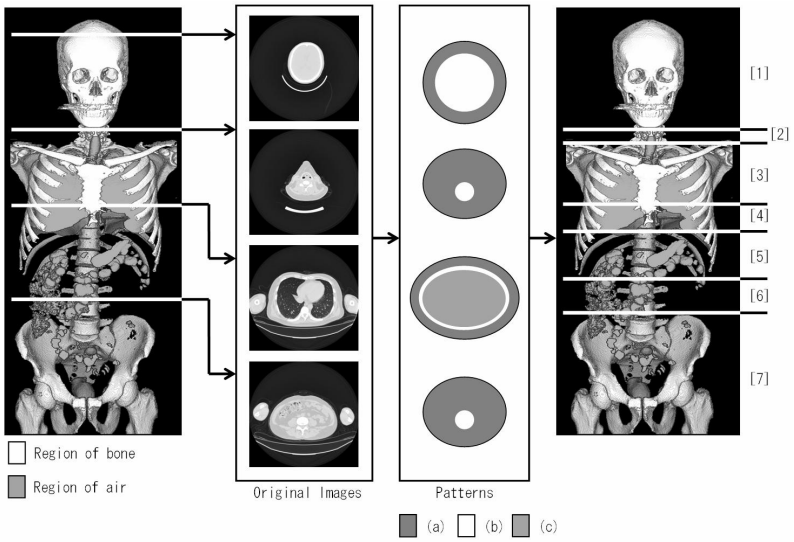

Fig. 7. Overview of body part identification method. (a) Region of body, (b) Region of closed convex hull of bone, (c) Region of closed convex hull of air. Image set is segmented into 7 parts using the patterns.

2.6 Anonymization engine

The anonymization engine has built-in two converters. One is DICOM image converter, the other is Comma Separated Values(CSV) converter. The DICOM image converter anonymizes DICOM images. The processing flow is as follows.

1. Load the DICOM image.

2. Get a data element of the data set.

3. Get the processing method that is indicated by the anonymization policy if the tag or VR of the data element corresponds with a picking condition of the policy.

4. Compute the processing method. In the linkable replacement, table operation interface is used.

The CSV converter encodes or decodes the identifiers of the CSV files. The feature allows to replace the patient IDs of the finding files to the anonymized IDs. The DICOM images and the findings are linkable.

2.7 Table operation interface

Software to manage tables has to provide the features, data consistency, validity and security. Database Management System(DBMS) provides the features. There are many DBMS that has built-in high reliability, facile introduction and extensive support tools. The anonymization manages tables using existing DBMS. Table operation interface manages the

operations of DBMS. The interface provides the feature that converts the requests by anonymization engine to SQL messages. The feature allows to separate development of the anonymization engine from the data model of tables.

2.8 Tables

A table is composed of identifier pairs. The constraints of attributes are PRIMARY KEY, UNIQUE and NOT NULL. The data types are based on the VR of the data elements. If the attribute is patient ID with LO, the data type is VARCHAR(64).

2.9 Implementation

The implementation equipment of the anonymization system is as follows.

OS: Windows Server 2003

Memory: 1024MB DDR PC-3200

Networks: 1,000 Mbit/s Ethernet

DBMS: MySQL 4.0

The development has used C/C++ language. The input interface and anonymization engine applied DICOM Toolkit(DCMTK) [4]. The DBMS is MySQL [5].

2.10 Anonymization test

We tested the anonymization system using original DICOM test data. The test used the DICOM anonymization policy shown Table.1.

Table. 1. DICOM anonymization policy for test.

Picking Condition	Processing Method
Tag is (0010,0020)	Linkable replacement
VR is DA	Date data replacement
VR is PN	Unlinkable replacement

The check of the header information used Osiris [6]. Fig.8 shows the header information of the test data and the result data. The anonymized ID was formed by an initial character and number. The checked result was normal as follows.

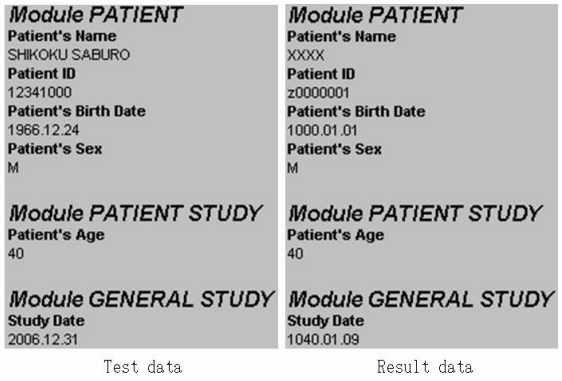

Fig. 8. Header information of test data and result data.

(a) Patient's Name was replaced to "XXXX" from "SHIKOKU^SABURO".

(b) Patient ID was replaced to "z00000001" from "12341000".

(c) Patient's Birth Date was replaced to "1000101" from "19661224".

(d) Study Date was replaced to "10400109" from "20061231".

3. RESULTS

The anonymization has operated in five medical facilities(facility A – E). The system has been used to anonymize and collect DICOM CT images for researches. In the facility A and B, The system has installed the PACS. The DICOM image server has always transferred DICOM images to the anonymization system automatically. If the past cases that stored in the DICOM image server are needed, the user selects the studies, and reserves tasks of transfer. Fig.9 shows anonymization flow in the facility A and B. The activities are as follows.

(A-1) The photographic device sends the DICOM image set to the DICOM image server.

(A-2) The DICOM image server stores the DICOM image set, then transfers the set to the anonymization system.

(B-1) The administrator selects the cases in the work station, then reserves task of transfer.

(B-2) The DICOM image server transfers the DICOM image sets to the terminal.

(B-3) The terminal receives the sets, then edits the header information and transfers to the anonymization system.

(C-1) The anonymization system receives and anonymizes the DICOM images. Finally, backup files of the tables are created.

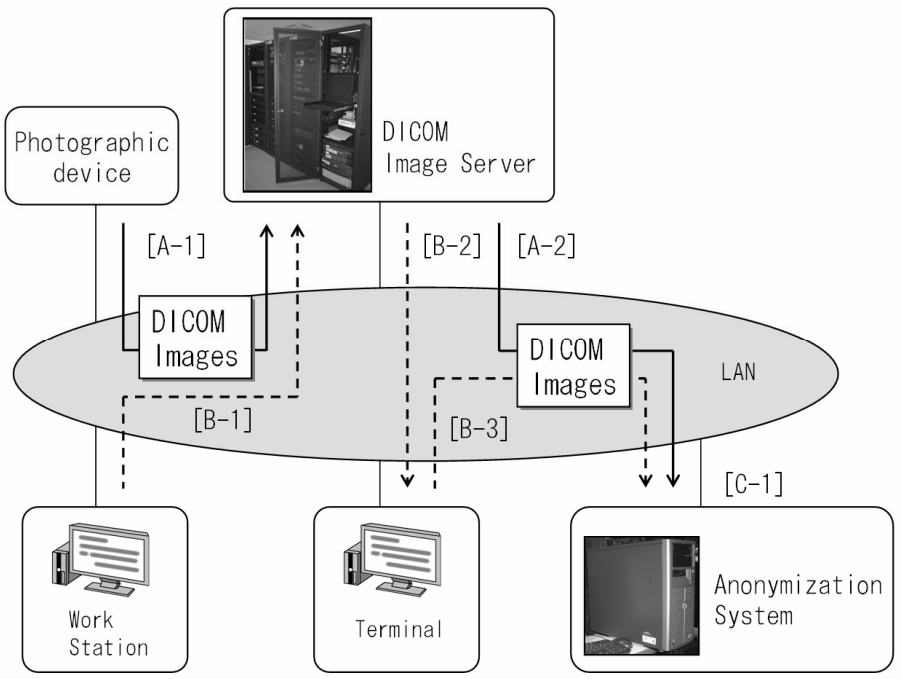

Fig. 9. Flow of anonymization in facility A and B.

The anonymization system has received DICOM images that were taken by one CT device in the facility A. In the facility B, the system has received the DICOM images that were taken by two CT devices. The anonymization system has finished the works before the day is out. Table.2 shows the number of anonymized studies, series and data volume a month. In the facility B, the preprocessing module has allowed to select studies of chest automatically.

Table. 2. Amount of anonymized data a month.

Facility	Study	Series	Data volume[GB]
A	446	1,081	158
B	225	378	37

Table. 3. Common anonymization policy in all facilities. The "Not anonymized" means explicit unanonymized data. The 4th row means that the group number is 0010. The 5th rows means private data elements.

Picking Condition	Processing Method
Tag is (0010,0040)	Not anonymized
Tag is (0010,1010)	
Tag is (0010,0020)	Linkable replacement
Tag is (0010,****)	Unlinkable replacement
Tag is odd	
VR is UN	
VR is PN	
VR is ST,LT,UT	

The other 3 facilities have implemented anonymization of the DICOM images that stored in the data storage mediums, external hard disk drive or Digital Versatile Disc(DVD). The simple GUI has allowed the users to anonymize easily.

A part of the anonymization policy has been same in all facilities shown Table.3.

4. CONCLUSIONS

In this paper, we reported an anonymization server system for DICOM images. The results proved that the system solved 2 problems. Firstly, the flexible anonymization method solved the differences of anonymization policies in each medical facility. Secondary, the system solved smoothly anonymization for massive DICOM images. The system creates useful anonymity DICOM images for researches and educations smoothly.

REFERENCES

1. National Electrical Manufacturers Association, *DICOM standard*, 2006.
2. Grant-in-Aid for Scientific Research on Priority Areas, Intelligent Assistance in Diagnosis of Multi-dimensional Medical Images, http://www.future-cad.org/fcad/
3. M. Nohara, H. Suzuki, S. Saita, M. Kubo, Y. Kawata, N. Niki, H. Nishitani, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, N. Moriyama, "Body part classification Algorithm based on Spatial Structure that uses Multi-detector row CT images, " *IEICE Technical Report*. MI2006-89(2007-1).
4. Oldenburg Research and Development Institute for Information Technology Tools and Systems, DICOM Toolkit, http://dicom.offis.de/index.php.en
5. MySQL AB, http://www.mysql.com
6. Osiris, http://www.expasy.ch/UIN
7. H. Suzuki, M. Amano, M. Kubo, Y. Kawata, N. Niki, H. Nishitani, "Anonymization for Secondary Use of Medical Images," *IEICE Technical Report*. PRMU2006-31(2006-5).
8. H. Suzuki, M. Amano, M. Kubo, Y. Kawata, N. Niki, H. Nishitani, "Operation and Performance Evaluation of Anonymization System in Secondary Use of DICOM Images," *IEICE Technical Report*. MI2006-121(2007-1).