Eur Radiol (2015) 25:3685–3695
DOI 10.1007/s00330-015-3794-0

COMPUTER APPLICATIONS

Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy

K. Y. E. Aryanto[1] · M. Oudkerk[1] · P. M. A. van Ooijen[1]

Received: 14 July 2014 / Revised: 18 December 2014 / Accepted: 7 February 2015 / Published online: 3 June 2015

Abstract

Purpose To compare non-commercial DICOM toolkits for their de-identification ability in removing a patient's personal health information (PHI) from a DICOM header.

Materials and Methods Ten DICOM toolkits were selected for de-identification tests. Tests were performed by using the system's default de-identification profile and, subsequently, the tools' best adjusted settings. We aimed to eliminate fifty elements considered to contain identifying patient information. The tools were also examined for their respective methods of customization.

Results Only one tool was able to de-identify all required elements with the default setting. Not all of the toolkits provide a customizable de-identification profile. Six tools allowed changes by selecting the provided profiles, giving input through a graphical user interface (GUI) or configuration text file, or providing the appropriate command-line arguments. Using adjusted settings, four of those six toolkits were able to perform full de-identification.

Conclusion Only five tools could properly de-identify the defined DICOM elements, and in four cases, only after careful customization. Therefore, free DICOM toolkits should be used with extreme care to prevent the risk of disclosing PHI, especially when using the default configuration. In case optimal security is required, one of the five toolkits is proposed.

Key Points

- *Free DICOM toolkits should be carefully used to prevent patient identity disclosure.*
- *Each DICOM tool produces its own specific outcomes from the de-identification process.*
- *In case optimal security is required, using one DICOM toolkit is proposed.*

Keywords DICOM Freeware Tools · Patient Data Privacy · Anonymisation and Pseudonymisation · Data Protection · Anonymous Testing

Introduction

The Digital Imaging and Communication in Medicine (DICOM) standard [1] has been commonly used for storing, viewing, and transmitting information in medical imaging [2]. Because of its structure and open character it can be easily adapted and upgraded to accommodate changes in medical imaging technology [3]. DICOM was developed to ease the exchange of data between different manufacturers, but it also enables data sharing between institutions or enterprises for clinical research or clinical practice.

A DICOM file not only contains a viewable image that holds all of the pixel values but it also contains a header with a large variety of data elements. Each data element is represented by a unique tag with specific values and data types. The tag of an element is written with two hexadecimal numbers indicating its group and element number. These meta-data elements include identifiable information about the patient, the study, and the institution. Sharing such sensitive data demands proper protection to ensure data safety and maintain patient privacy.

✉ K. Y. E. Aryanto
k.y.e.aryanto@umcg.nl

[1] University of Groningen, University Medical Center Groningen, Center for Medical Imaging - North East Netherlands (CMINEN), Department of Radiology, Hanzeplein 1, Postbus 30001, 9700 RB Groningen, The Netherlands

There are two methods to de-identify patient-related information in a DICOM header. The first method is anonymization which removes information carried by header elements or replaces the information with random data such that the remaining information cannot be used to reveal the patient identity at all. The other method, pseudonymization, is implemented by replacing the most identifying fields within a data record using one or more artificial identifiers that could be used by authorized personnel to track down the real identity of the patient. This method is most frequently used in clinical analysis, processing, and research [4–6] since good clinical practice requires that, should additional findings be encountered that are essential for the well-being of the patient, it should be possible to somehow track the real identity of the patient in order to inform him or her about these findings.

Numerous tools have been built to perform the task of DICOM data de-identification in order to fulfil the requirements of patient data protection. Each tool introduces its own de-identification profiles to remove or replace a selection of header elements and, therefore, produces its own specific outcomes from the data de-identification process. In this work, ten non-commercial (free) DICOM toolkits were selected and tested for their de-identification effectiveness and completeness to determine the tools' ability to remove a patient's personal health information (PHI) from the DICOM header. This work also provides further consideration of DICOM toolkits that could perform data de-identification to meet regulatory requirements.

Methods

Various applications, libraries, and frameworks have been developed for handling, viewing, transmitting, and processing DICOM data. These toolkits offer many features useful for clinical practice or clinical research purposes such as DICOM data validation, image viewing and analysis, PACS server, and converting and modifying, including de-identifying, DICOM data. Similar work examining seven free DICOM software toolkits and their ability to de-identify 38 tags that contain patient or study information using their default and modified configurations has been previously presented [7, 8].

Several DICOM toolkits were selected to be compared for their de-identification capabilities. The candidates were gathered through an internet search to obtain as many free toolkits as possible using a number of dedicated information sources on the web [9–12] and also through a web search engine with the search term “DICOM anonymizer” or “free DICOM anonymizer”. Main inclusion criteria were the ability of the applications or frameworks to perform de-identification and availability as freeware or an open source tool that can be downloaded and installed or is accessible as an on-line, web-based, anonymization service. Other inclusion criteria were based on how commonly the toolkits were used in practice, by noting practitioner toolkit preferences via direct discussion or via answers posted in online discussion forums or the like. The continuity of a toolkits' development was also considered as inclusion criteria; it was determined by the update history of the software and active communication about the software. Selected toolkits were not only end-user applications but also several frameworks, providing features allowing users to perform de-identification directly.

All selected tools were evaluated on a workstation running Microsoft Windows XP Service Pack 3 and tested to de-identify the elements of a “dummy” DICOM file header. Fifty header elements were chosen to be de-identified since they contained data that could be used to reconstruct a patient's real identity individually or in combination with other elements (Table 1).

Two scenarios were defined to perform the de-identification. First, the default setting of the tools were used, meaning that the installed tools were used to perform the process as is, without any customization. Then, customized settings were defined to obtain the best possible configuration to perform the de-identification process. For each test, the unchanged elements were observed to determine whether any of the potential identifying information was retained. The test was performed using a dummy DICOM image (Fig. 1).

The DICOM header elements of the dummy DICOM file were filled with the string “Should anonymized” when possible, except for those containing date or time values. Using this dummy DICOM file, the de-identification process was performed according to the two scenarios. The de-identified DICOM files were checked to determine whether they still contained elements as listed above with the original value or the given string. Figure 2 describes the workflow of the method.

Results

Ten tools were selected, namely Conquest DICOM software [13], RSNA Clinical Trial Processor (CTP) [14], DICOM library [15], DICOMworks [16], DVTK DICOM anonymizer [17], GDCM [18], K-Pacs [19], PixelMed DICOMCleaner [20], Tudordicom [21], and YAKAMI DICOM tools [22]. Table 2 shows the general features offered by the selected tools. Several of them have been previously introduced, implemented and reported on individually in the literature [23–26]. There are also several frameworks which have features to perform the de-identification but which were not included in this comparison since they cannot be used directly as a stand-alone application.

All selected tools are easy to install by following a step-by-step installation wizard. Additionally, some require other supporting applications, frameworks, or runtime

Table 1 Fields in the DICOM header defined to be de-identified

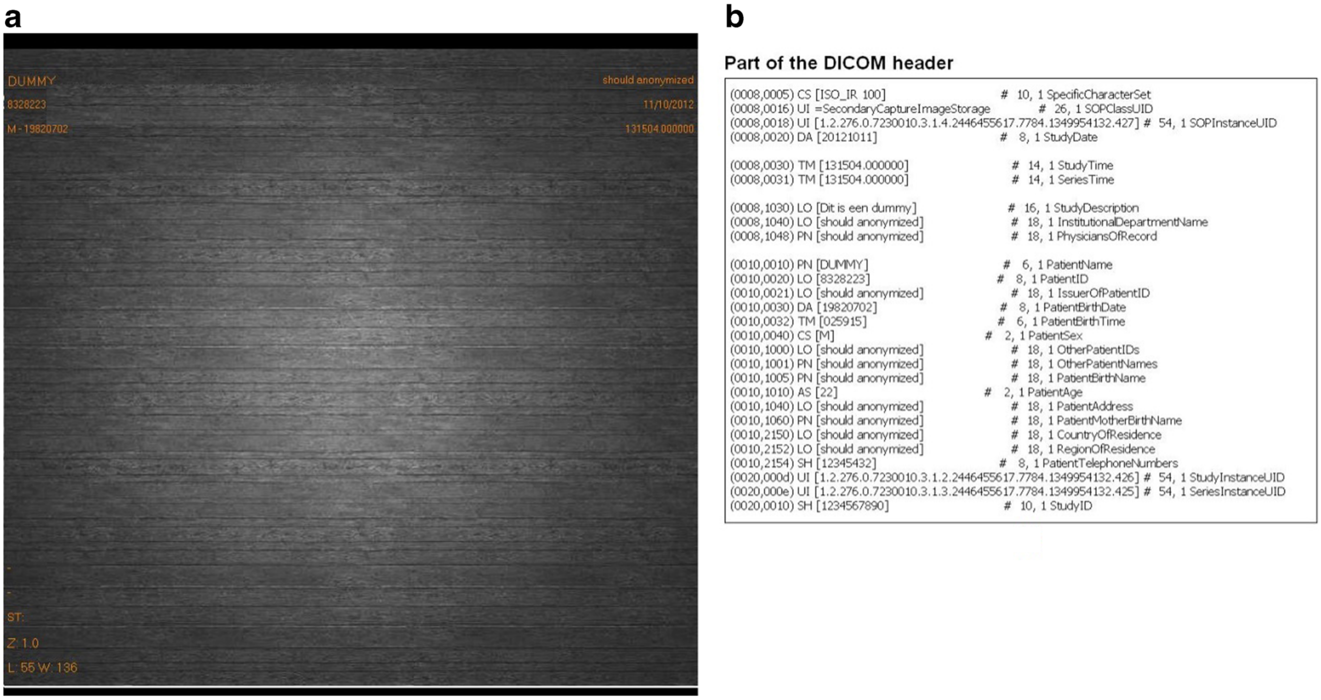
Tag ID	Tag Name
0008,0020	StudyDate
0008,0021	SeriesDate
0008,0022	AcquisitionDate
0008,0023	ContentDate
0008,0024	OverlayDate
0008,0025	CurveDate
0008,002A	AcquisitionDatetime
0008,0030	StudyTime
0008,0031	SeriesTime
0008,0032	AcquisitionTime
0008,0033	ContentTime
0008,0034	OverlayTime
0008,0035	CurveTime
0008,0050	AccessionNumber
0008,0080	InstitutionName
0008,0081	InstitutionAddress
0008,0090	ReferringPhysiciansName
0008,0092	ReferringPhysiciansAddress
0008,0094	ReferringPhysiciansTelephoneNumber
0008,0096	ReferringPhysicianIDSequence
0008,1040	InstitutionalDepartmentName
0008,1048	PhysicianOfRecord
0008,1049	PhysicianOfRecordIDSequence
0008,1050	PerformingPhysiciansName
0008,1052	PerformingPhysicianIDSequence
0008,1060	NameOfPhysicianReadingStudy
0008,1062	PhysicianReadingStudyIDSequence
0008,1070	OperatorsName
0010,0010	PatientsName
0010,0020	PatientID
0010,0021	IssuerOfPatientID
0010,0030	PatientsBirthDate
0010,0032	PatientsBirthTime
0010,0040	PatientsSex
0010,1000	OtherPatientIDs
0010,1001	OtherPatientNames
0010,1005	PatientsBirthName
0010,1010	PatientsAge
0010,1040	PatientsAddress
0010,1060	PatientsMothersBirthName
0010,2150	CountryOfResidence
0010,2152	RegionOfResidence
0010,2154	PatientsTelephoneNumbers
0020,0010	StudyID
0038,0300	CurrentPatientLocation
0038,0400	PatientsInstitutionResidence
0040,A120	DateTime
0040,A121	Date
0040,A122	Time
0040,A123	PersonName

environments to be pre-installed, depending on what type of programming language in which they were developed. Toolkits developed using Java will need a Java Runtime to be pre-installed. A NET framework is needed for applications that are developed using C#. Some toolkits require other, more specific, applications to be pre-installed to support the complete process of reading or processing the DICOM files. For example, Tudordicom and CTP also require additional Java ImageIO Tools [27] to be present on the system to be able to read and process the compressed DICOM files. The GDCM installation under Microsoft Windows requires a Win32 OpenSSL [28] to be pre-installed, while YAKAMI needs DirectX to be present. All required pre-installations are available freely from the web from their respective manufacturers.

A modifiable setting, in this case the ability to adjust the de-identification profiles, is important for an application to meet a user's more specific needs. Six of the ten toolkits have customizable de-identification profiles. DVTk provides two profile selections to perform the de-identification, in a simple or complete way. In the other five tools, customization can be done using the GUI provided by the applications, inserting scripts into text file, or using command-line arguments. However, not all toolkits provide customizable de-identification profiles. Conquest, DICOM Library, DICOMWorks, and KPACS have a fixed profile for the de-identification process.

Using both default and customized configurations, two scenarios were performed to determine to what extent the profiles could provide a secure de-identification by observing the remaining original values of the defined 50 elements. These elements were selected based on their likelihood of being the cause of a data breach when exposed to a third party, either by the element itself or combination with other elements.

From the tested applications, only DICOM Library can de-identify all of the defined elements using its default setting, while another four can perform this task using user-customized profiles. These four tools are CTP, GDCM, Tudordicom, and Yakami Dicom Tools. In addition to the header de-identification, Yakami DICOM Tools, Pixelmed DICOM Cleaner and CTP provide the ability of removing information "burned in" into the image pixels by blacking out a certain region of the image. The summary of the comparison is shown in Table 3. The list of changed tag elements are shown in Table 4. The success rate in de-identifying the



Part of the DICOM header

```
(0008,0005) CS [ISO_IR 100]                              #  10, 1 SpecificCharacterSet
(0008,0016) UI =SecondaryCaptureImageStorage              #  26, 1 SOPClassUID
(0008,0018) UI [1.2.276.0.7230010.3.1.4.2446455617.7784.1349954132.427] #  54, 1 SOPInstanceUID
(0008,0020) DA [20121011]                                #   8, 1 StudyDate

(0008,0030) TM [131504.000000]                             #  14, 1 StudyTime
(0008,0031) TM [131504.000000]                             #  14, 1 SeriesTime

(0008,1030) LO [Dit is een dummy]                          #  16, 1 StudyDescription
(0008,1040) LO [should anonymized]                         #  18, 1 InstitutionalDepartmentName
(0008,1048) PN [should anonymized]                         #  18, 1 PhysiciansOfRecord

(0010,0010) PN [DUMMY]                                   #   6, 1 PatientName
(0010,0020) LO [8328223]                                #   8, 1 PatientID
(0010,0021) LO [should anonymized]                         #  18, 1 IssuerOfPatientID
(0010,0030) DA [19820702]                                #   8, 1 PatientBirthDate
(0010,0032) TM [025915]                                 #   6, 1 PatientBirthTime
(0010,0040) CS [M]                                    #   2, 1 PatientSex
(0010,1000) LO [should anonymized]                         #  18, 1 OtherPatientIDs
(0010,1001) PN [should anonymized]                         #  18, 1 OtherPatientNames
(0010,1005) PN [should anonymized]                         #  18, 1 PatientBirthName
(0010,1010) AS [22]                                   #   2, 1 PatientAge
(0010,1040) LO [should anonymized]                         #  18, 1 PatientAddress
(0010,1060) PN [should anonymized]                         #  18, 1 PatientMotherBirthName
(0010,2150) LO [should anonymized]                         #  18, 1 CountryOfResidence
(0010,2152) LO [should anonymized]                         #  18, 1 RegionOfResidence
(0010,2154) SH [12345432]                               #   8, 1 PatientTelephoneNumbers
(0020,000d) UI [1.2.276.0.7230010.3.1.2.2446455617.7784.1349954132.426] #  54, 1 StudyInstanceUID
(0020,000e) UI [1.2.276.0.7230010.3.1.3.2446455617.7784.1349954132.425] #  54, 1 SeriesInstanceUID
(0020,0010) SH [1234567890]                               #  10, 1 StudyID
```

Fig. 1 Dummy DICOM image. **a**) A generated DICOM file consisting of header data and image pixels. **b**) Part of the header. The 50 tag elements to be de-identified by various selected DICOM toolkits were filled with dummy information or the string "should be anonymized"

DICOM header using the default setting provided by the toolkits is shown in Fig. 3, while Fig. 4 shows the success rate using the advance setting.

Only two toolkits provided a high success rate of de-identification when using the default setting (CTP and DICOM Library), while an additional four achieved a high success rate after careful customization (GDCM, PixelMed, TudorDICOM, and Yakami DICOM tool). DICOM Library is the only tool that achieves a 100 % success rate at its default setting. The success rate of the CTP to de-identify the DICOM header using its default profile is 98 %, which increases to a complete de-identification of the specified elements under custom settings. Pixelmed could deliver a high success rate of 98 % using its advance setting while it failed to do so in its default setting (only 64 %). Meanwhile, DVTK provided less than a 44 % success rate using its default setting and the optimization capabilities did not allow much improvement, resulting in a success rate of 48 %.

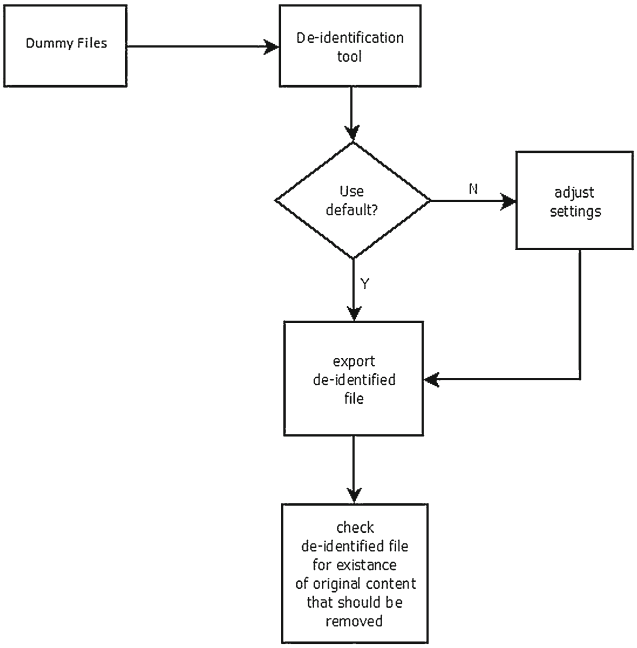

Dummy Files
De-identification tool
Use default?
N
adjust settings
Y
export de-identified file
check de-identified file for existance of original content that should be removed

Fig. 2 Flowchart of the method to test DICOM de-identification tools

Only five out of ten selected free DICOM toolkits could de-identify all of the defined DICOM elements properly with a 100 % success rate. Four of them could only achieve this after improvement using advance settings with user controlled de-identification protocols. One toolkit achieved a 98 % success rate after manual improvement of the de-identification settings. Only two out of ten toolkits were able to give a success rate above 90 % using the default setting, with all remaining tools performing at less than 65 %, of which four even achieved success rates of 26 % or less.

Discussion

Various toolkits have been built to de-identify DICOM data, either as free or paid applications. Paid toolkits have advantages such as customer support and development updates, while free versions less likely to have consistent updates. However, the free versions are not necessarily of poorer

Table 2 Selected DICOM toolkits

Name	Platform	Type of Distribution	User Interface	Function	Source Avail.	Programming Language	Year update	Requirements	Doc/ User Manual
DICOMWorks	Windows	Freeware	GUI	Application	N	N/A	2007	*OS : Microsoft Windows systems	Y
KPacs	Windows	Freeware	GUI	Display, PACS Client, Server	N	N/A	2009	* OS : Windows 2000/XP * Processor : >=Pentium III (800 MHz) * Monitor with 1024x768 pixel resolution	Y
Conquest Dicom Server	Windows, Linux	Open Source	GUI	Library, PACS Server	Y	C/C++	2010	* OS : Windows NT/ 2000/XP/Vista/Win7/ Linux * 1024x768x256 display. * TCP/IP functioning	Y
DVTk DICOM Anonymizer	Windows	Open Source	GUI	Library, Application	Y	C#	2011	* OS : Microsoft Windows XP/Vista/Windows7 * .NET 2.0 Framework	Y
DICOM library	Windows, Macintosh, Linux	Free Online	GUI	Library	N	N/A	2013	N/A	Y
PixelMed DICOMCleaner	Windows, Macintosh, Linux	Open Source	Command-line utility	Display, Library, Utility	Y	Java	2013	* Java Runtime (JRE) 1.5 or newer * Microsoft Windows XP/ 2000/Windows 7/Linux/ Mac OS X	Y
Tudordicom	Windows, Macintosh, Linux	Open Source	GUI	Utility/ Application, Processor	Y	Java	2013	* Java Runtime (JRE) 1.5 or newer * Java ImageIO	Y
CTP	Windows, Macintosh, Linux	Open Source	GUI	Utility/ Application, Processor	Y	Java	2013	* Java Runtime (JRE) 1.5 or newer * Java Advanced Imaging ImageIO Tools	Y
GDCM	Windows, Macintosh, Linux	Open Source	Command-line utility	Utility, Library	Y	C#, C++, Python	2013	* OpenSSL	Y
YAKAMI DICOM Tools	Windows	Freeware	GUI	Utility/ Application, PACS Client	N	N/A	2013	* OS: Windows7/ Vista/ XP/2000 * .NET 2.0 Framework * DirectX®	Y

Table 3 Summary of comparison of the de-identification toolkit

Name	De-identification Profiles			De-identification features			De-identify 50 Elements	
	Customizable	Profiles	Configuration	Multiple Files	Automatic	Pixel Blackout	Default	Customized
Conquest Dicom Server	N	Fixed	N/A	Y, study/series	N	N	N	N/A
CTP	Y	Defined,Element or Group selection	GUI or text file input	Y, directory	Y	Y	N	Y
Dicom Library	N	Fixed	N/A	Y, directory	Y	N	Y	N/A
DICOMWorks	N	Fixed	N/A	Y, study/series	Y	N	N	N
DVTK DICOM Anonymizer	Y	Fixed profiles selection	GUI	Y, directory	Y	N	N	N
GDCM	Y	Defined, Element selection	Command options/ arguments	Y, directory	Y	N	N	Y
KPacs Anonymizer	N	Fixed	N/A	Y, directory	Y	N	N	N/A
PixelMed DICOMCleaner	Y	Group selection	GUI	Y, files/study/series	N	Y	N	N
Tudordicom	Y	Element selection	GUI	Y, directory	Y	N	N	Y
YAKAMI Dicom Tools	Y	Element selection	GUI or text file input	Y, files or directory	Y	Y	N	Y

quality. Many of the free toolkits are provided in an open source version, which means that the tools are open for improvements either by users or related communities.

The elements to be de-identified in this work were chosen based on their potential for being the cause of a data breach when exposed to a third party, either by the element itself or in combination with other elements. Even though all of those elements will not be filled in a daily routine, a recommendation for removal or modification of those elements is still required due to the possibility of practitioners giving values to the elements, as determined via our observation of several cases where those elements contained certain values. The values are most likely the appropriate values required by the elements and could possibly reveal a patient's identity.

The selection of 50 DICOM tags was made based on a careful inspection of possible fields containing sensitive information in combination with the information of Supplement 142 of the DICOM standard. This selection was, therefore, based on experience of the authors which could influence the quality score.

The selection of software packages included in this work was based on a number of parameters. It would be impossible to review all available software. Therefore, a possible bias could be introduced by the selection of the software packages. However, to obtain the most relevant results, software packages were selected on criteria that would identify their frequency of download and use. Based on these criteria, the software packages most frequently used and, thus, probably with the highest impact in daily practice, were selected.

A default configuration of a de-identification profile allows users to quickly run a required task as intended without in-depth knowledge of the tool itself. Nevertheless, the default configuration does not always provide de-identification of sensitive patient-related information within the DICOM data for a specific research project or for educational purposes. For such reasons, a customizable configuration is required to perform the intended task. The customizable settings will provide more flexibility and improved tool performance, especially if the image data are needed for a specific research project or for educational purposes.

The selection of element tags was done by considering two kinds of elements, direct and indirect patient information fields, consisting of 17 and 33 elements, respectively. Direct patient information fields have information that directly points to patient identity, including PatientsName, PatientID, IssuerOfPatientID, PatientsBirthDate, PatientsBirthTime, PatientsSex, OtherPatientIDs, OtherPatientNames, PatientsBirthName, PatientsAge, PatientsAddress, PatientsMothersBirthName, CountryOfResidence, RegionOfResidence, PatientsTelephoneNumbers, CurrentPatientLocation, PatientsInstitutionResidence. The remaining elements are indirect patient information fields. The elements listed above are recommended for de-identification

Table 4 The results of DICOM header element de-identification by ten DICOM toolkits

Fields/Tags	Tag Name	Conquest	CTP	DICOM Library	DICOM works	DVTK	GDCM	KPACS	Pixelmed	Tudor	Yakami
0008, 0020	Study Date	N	Y	Y	N	N	N	N	N	N	N
0008, 0021	Series Date	N	Y	Y	N	N	N	N	N	N	N
0008, 0022	AcquisitionDate	N	Y	Y	N	N	N	N	N	N	N
0008, 0023	ContentDate	N	Y	Y	N	N	N	N	N	N	N
0008, 0024	OverlayDate	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
0008, 0025	CurveDate	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
0008, 002A	AcquisitionDatetime	Y	Y	Y	N	N	N	N	N	N	N
0008, 0030	StudyTime	N	Y	Y	N	N	N	N	N	N	N
0008, 0031	SeriesTime	N	Y	Y	N	N	N	N	N	N	N
0008, 0032	AcquisitionTime	N	Y	Y	N	N	N	N	N	N	N
0008, 0033	Content Time	N	Y	Y	N	N	N	N	N	N	N
0008, 0034	Overlay Time	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
0008, 0035	CurveTime	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
0008, 0050	Accession Number	N	Y	Y	N	Y	Y	N	Y	N	N
0008, 0080	Institution Name	N	Y	Y	Y	Y	Y	N	N	Y	N
0008, 0081	Institution Address	N	Y	Y	N	Y	Y	N	N	Y	N
0008, 0090	Referring Physicians Name	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
0008, 0092	Referring Physicians Address	N	Y	Y	N	Y	Y	N	Y	Y	N
0008, 0094	Referring Physicians Telephone Number	N	Y	Y	N	Y	Y	N	Y	Y	N
0008, 0096	Referring PhysicianIDSequence	N	Y	Y	N	N	N	N	Y	Y	N
0008, 1040	InstitutionalDepartmentName	N	Y	Y	N	Y	Y	N	N	Y	N
0008, 1048	PhysicianOfRecord	N	Y	Y	N	Y	Y	N	Y	Y	N
0008, 1049	PhysicianOfRecordID Sequence	N	Y	Y	N	N	N	N	Y	Y	N
0008, 1050	Performing Physicians Name	N	Y	Y	N	Y	Y	N	Y	Y	N
0008, 1052	Performing PhysicianID Sequence	N	Y	Y	N	N	N	N	Y	Y	N
0008, 1060	NameOf Physician Reading Study	N	Y	Y	N	Y	Y	N	Y	Y	N
0008, 1062	Physician Reading StudyID Sequence	N	Y	Y	N	N	N	N	Y	Y	N
0008, 1070	Operators Name	N	Y	Y	N	Y	Y	N	Y	Y	N
0010, 0010	Patients Name	Y	Y	Y	Y	Y	Y	N	Y	Y	N
0010, 0020	PatientID	Y	Y	Y	Y	Y	Y	N	Y	Y	N
0010, 0021	IssuerOf PatientID	N	Y	Y	N	N	N	N	Y	N	N
0010, 0030	Patients BirthDate	Y	Y	Y	N	N	Y	N	Y	N	N
0010, 0032	Patients BirthTime	N	Y	Y	N	Y	Y	N	Y	Y	N
0010, 0040	PatientsSex	N	Y	Y	N	Y	Y	N	N	Y	N
0010, 1000	OtherPatientIDs	N	Y	Y	N	Y	Y	N	Y	Y	N
0010, 1001	OtherPatientNames	N	Y	Y	N	Y	Y	N	Y	Y	N

Table 4 (continued)

Fields/Tags	Tag Name	Conquest	CTP	DICOM Library	DICOM works	DVTK	GDCM	KPACS	Pixelmed	Tudor	Yakami
0010, 1005	Patients BirthName	N	Y	Y	N	N	N	N	Y	Y	N
0010, 1010	PatientsAge	N	Y	Y	N	Y	Y	N	N	Y	N
0010, 1040	Patients Address	N	Y	Y	N	N	N	N	Y	Y	N
0010, 1060	Patients Mothers BirthName	N	Y	Y	N	N	N	N	Y	Y	N
0010, 2150	CountryOf Residence	N	Y	Y	N	N	N	N	Y	N	N
0010, 2152	RegionOf Residence	N	Y	Y	N	N	N	N	Y	N	N
0010, 2154	Patients Telephone Numbers	N	Y	Y	N	N	N	N	Y	Y	N
0020, 0010	StudyID	N	Y	Y	Y	N	Y	N	Y	N	N
0038, 0300	Current Patient Location	N	Y	Y	N	N	N	N	Y	N	N
0038, 0400	Patients Institution Residence	N	Y	Y	N	N	N	N	N	Y	N
0040, A120	DateTime	Y	Y	Y	N	N	N	N	N	N	N
0040, A121	Date	Y	Y	Y	N	N	N	N	N	N	N
0040, A122	Time	Y	N	Y	N	N	N	N	N	N	N
0040, A123	PersonName	Y	Y	Y	N	N	N	N	Y	Y	N

to prevent the elements containing date or time related to patients, data acquisition, or other process being used, alone or in combination with others, to reveal the real patient identity that may lead to the breach of a patient's important data. In order to de-identify the elements, dummy date or time values are set to the appropriate elements to replace the original values. These dummy values vary depending on the aim of the study or research.

The support of configurable profiles should provide options to the user to perform a specific de-identification process more freely. Several methods were introduced by the different toolkits, such as adding, modifying or removing header elements one element at a time or using a list of actions, defined by the tools or manually, to be conducted on several elements simultaneously. Some tools require script files to be manually written or adapted using a text file editor or employ a user interface to generate these script files from within the application.

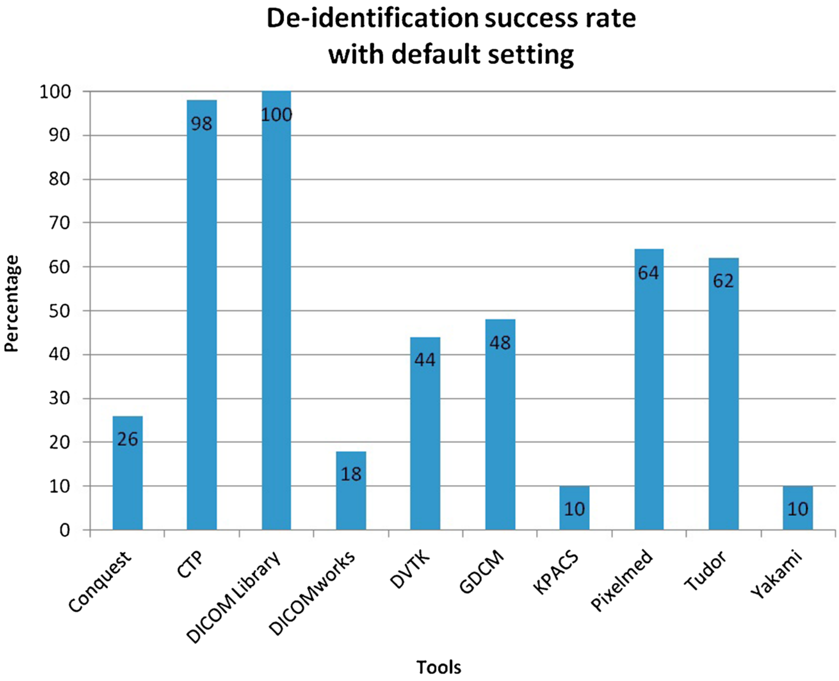
The ability of a tool to de-identify multiple files automatically can be a significant advantage. This feature will ease the de-identification process for a set of images which is usually required when de-identifying data from cross-section-based modalities such as computed tomography (CT) and magnetic resonance imaging (MRI). Tools lacking this capability would require one to manually perform the task one file at a time, resulting in a more time consuming method which is cumbersome for the user and more prone to errors. Customizable or user-defined selection of de-identification profiles will be a major advantage when compared to standard settings, because otherwise nobody will check which of these DICOM tags will be de-identified.

Supplement 142 in the DICOM standards provides a profile within clinical trials de-identification that has become the standard of DICOM data security. Nevertheless, to have the full list of the tags in supplement 142 to be de-identified would still be difficult to do manually. Instead, we provided 50 elements considered to be the minimum requirements for a third party to reveal the identity of a patient. Furthermore, the recommended software has also provided a configuration claiming to conform to Supplement 142 in the DICOM standard.

The ability to blackout the embedded information written on the images is an advantage in identity protection. In some cases, patient information can be included in the DICOM image data as "burned in" information, for example, in the case of storage of secondary capture images or with frame-grabbed ultrasound examinations. A de-identification of the DICOM header could become meaningless when such information is still present within the image itself. This feature is only supported by Yakami DICOM Tools, Pixelmed DICOM cleaner and CTP.

Another potential risk is the use of private tags. These private tags can be used by the manufacturer to provide

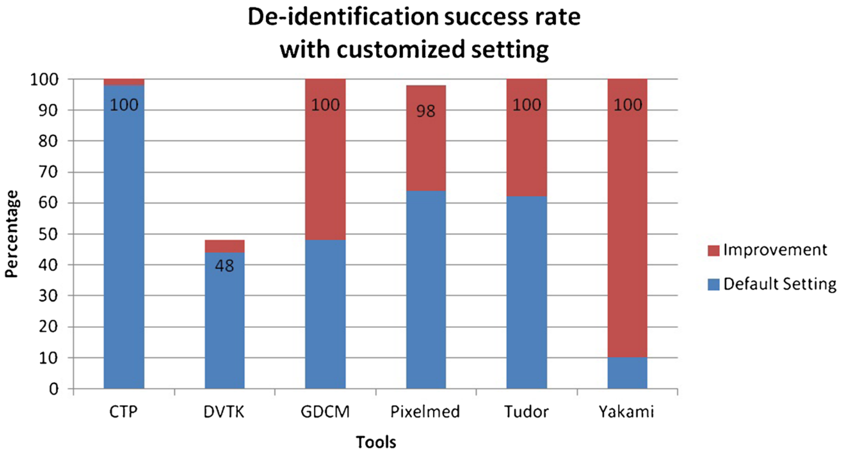
Fig. 3 Success rate of the toolkit to de-identify fifty DICOM header elements using the default settings. The numbers presented in the bars are the total score using the default de-identification setting



additional, proprietary information within the DICOM header. These tags may contain sensitive data regarding a patient's PHI. However, not all private elements consist of sensitive data. Therefore, unless the tags contain important information for further processing, it is recommended that those elements should be removed. Private tags are typically documented to provide additional information related to the device/manufacturer. However, the additional data which may contain patient related information can also be added manually or automatically, for example, when private tags are not displayed in the DICOM viewer. However, as mentioned above, private DICOM tags may also provide sensitive patient data. Although these data are not visible through the DICOM viewer, they are available for viewing using the tag reader and may be used by other parties to reveal the patients identity.

The utilization of a framework or of library tools such as GDCM is limited since those tools are intended to be used for advanced purposes, integrated into another application as a toolkit. However, the provided functionality is sufficient for practical use. Other known frameworks that provide a de-identification process are DCM4CHE [29, 30] and DCMTK [31]. DCM4CHE is a framework developed using the Java programming language that is claimed to have better functionality compared to the others [32]. However, this framework is not directly suitable for practical use, but can be used by a software developer to be integrated into new software tools. The RSNA Clinical Trial Processor (CTP) tested in this study

Fig. 4 Success rate of the toolkit to de-identify fifty DICOM header elements using the advanced settings. The numbers presented in the bars are the maximum success rate obtained after customization of the de-identification settings



is one of the toolkits that use this framework as part of the software.

The low de-identification performance of several applications might be caused by the main role of the application itself. For example, the tools that were intended to be an image viewer are likely to have low priority for development and implementation of the image de-identification process. On the other hand, an application that is addressed as a DICOM data processor will have more advanced options to perform the de-identification task since that is one of its intended uses.

The DICOM Library is an online service to share images. It is developed mainly for educational and scientific purposes [15]. Its output data were well de-identified and downloadable. However, the uploading of images to be de-identified by the service should be considered further since the process is done outside the domain of the sender. This means that even though the source files are claimed to be de-identified at the client side, the implementation of an unsupervised process involving uploading to a third party should be utilized with care and checked with hospital security regulations. Using this kind of service may cause a security breach due to the possibility that unmodified parts of data still contain sensitive information. It might, thus, not be allowed according to the security policies of most institutions since it is unknown what exactly happens with the uploaded files at the server side. Furthermore, the files could be retained at the server for some unknown period of time without the uploading party being aware of this storage. Even though online, web-based anonymization services are not ideal for the transfer of such confidential data using standard transfer protocols, there are still possibilities to make such methods acceptable, either by moving the services to a more secure line or transfer only data without burnt-in information within the images. However, although the transfer is claimed to be secure, information that is not processed by such service, i.e., burnt-in information within the images themselves, can still reveal patient identity. We suggest that the use of online services without full control from the user should be avoided as much as possible.

The challenge with the blackout of regions is that it is a fully manual process. When annotations are made on the image, e.g., in ultrasound, the location of this information will vary and in some cases manually entered annotations could be positioned at several places or on top of the actual image. Therefore, default settings to overcome this problem are not available. This calls for extra attention when ultrasound images are involved and instructing imagers involved in studies not to include annotations that are ‘burned’ into the images.

Conclusion

Only two out of ten free available DICOM de-identification toolkits had a success rate of de-identification higher than 90 % using the default setting. All remaining tools performed with a success rate lower than 65 %, of which four only achieved a success rate of 25 % or less.

Free DICOM toolkits should, therefore, be used with extreme care when de-identifying sensitive data since they have a high risk of disclosing personal health information, especially when using the default configuration. Four out of ten tools are not recommended to be used in de-identifying DICOM data since they could cause serious threats to patient privacy.

In case optimal security is required, RSNA CTP is recommended for its high level of customization to perform de-identification to exactly meet the regulatory requirements [33].

Acknowledgments The scientific guarantor of this publication is Prof. Dr. Matthijs Oudkerk. The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. This study received funding by the ZonMw Innovative Medical Devices Initiative (IMDI) under project registration number 104002003. No complex statistical methods were necessary for this paper. Institutional Review Board approval was not required because our study was not onhuman subjects. Methodology:prospective, experimental, performed at one institution.

References

1. N. E. M. A. (NEMA), “The DICOM Standard.” [Online]. Available: http://medical.nema.org/
2. O. Pianykh, “What Is DICOM?,” in in Digital Imaging and Communications in Medicine (DICOM), Springer Berlin Heidelberg, 2012, pp. 3–5
3. Mustra M, Delac K, Grgic M (2008) *Overview of the DICOM standard*. IEEE 1:10–12
4. Noumeir R, Lemay A, Lina J-M (2007) Pseudonymization of Radiology Data for Research Purposes. J Digit Imaging 20(3): 284–295
5. Neubauer T, Riedl B (2008) Improving patients privacy with Pseudonymization. Stud Health Technol Info 136:691–696
6. Neubauer T, Heurix J (2011) A methodology for the pseudonymization of medical data. Int J Med Inform 80(3):190–204
7. Lakhani, P, Chen, J, Nagy, P, Safdar, N, “Protecting Your Patient's Privacy: Is Your DICOM Anonymizer Working for You?”, *Radiological Society of North America 2009 Scientific Assembly and Annual Meeting*, *November 29 - December 4, 2009 ,Chicago IL*.http://archive.rsna.org/2009/8011488.html Accessed September 10, 2014
8. National Institutes of Health, “I Do Imaging,” 2013. [Online]. Available: http://www.idoimaging.com/
9. W. Schöch, “Diploma thesis ‘Using DICOM SR in Pathology’,” 2012. [Online]. Available: http://www.schoech.de/diploma/toolkits.html

10. D. A. Clunie, “David Clunie’s Medical Image Format Site,” 2013. [Online]. Available: http://www.dclunie.com/medical-image-faq/html/part8.html#DICOMDeidentifiers
11. Plastimatch development team, “DICOM anonymizer comparison, ” 2013. [Online]. Available: http://plastimatch.org/dicom_comparison.html
12. Marcel van Herk, “Conquest DICOM software.” [Online]. Available: http://ingenium.home.xs4all.nl/dicom.html
13. RSNA, “CTP-The RSNA Clinical Trial Processor.” [Online]. Available: http://mircwiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor
14. D. Library, “DICOM Library - Anonymize, Share, View DICOM files ONLINE.” [Online]. Available: http://www.dicomlibrary.com
15. Dicomworks project, “DicomWorks - Free DICOM software.” [Online]. Available: http://www.dicomworks.com
16. DVTk, “DVTk Project.” [Online]. Available: http://www.dvtk.org/
17. GDCM, “GDCM: Grassroots DICOM library.” [Online]. Available: http://gdcm.sourceforge.net/wiki/index.php/Main_Page
18. Andreas Knopke, “K-Pacs.” [Online]. Available: http://k-pacs.net/
19. P. Publishing, “PixelMed Java DICOM Toolkit.” [Online]. Available: http://www.pixelmed.com
20. C. de R. P. H. Tudor, “The Tudor Dicom Tools.” [Online]. Available: http://santec.tudor.lu/project/optimage/dicom/start
21. Masahiro YAKAMI, “YAKAMI DICOM Tools.” [Online]. Available: http://www.kuhp.kyoto-u.ac.jp/~diag_rad/intro/tech/dicom_tools.html
22. Puech PA, Boussel L, Belfkih S, Lemaitre L, Douek P, Beuscart R (2007) DicomWorks: software for reviewing DICOM studies and promoting low-cost teleradiology. J Digit Imaging Off J Soc Comput Appl Radiol 20(2):122–130
23. Potter G, Busbridge R, Toland M, Nagy P (2007) “Mastering DICOM with DVTk. J Digit Imaging Off J Soc Comput Appl Radiol 20(Suppl 1):47–62
24. Rodríguez González D, Carpenter T, Hemert J, Wardlaw J (2010) An open source toolkit for medical imaging de-identification. Eur Radiol 20(8):1896–1904
25. Aryanto KYE, Broekema A, Oudkerk M, a van Ooijen PM (2012) Implementation of an anonymisation tool for clinical trials using a clinical trial processor integrated with an existing trial patient data information system. Eur Radiol 22(1):144–151
26. Oracle, “Java Advanced Imaging Image I/O Tools Installation.” [Online]. Available: http://www.oracle.com/technetwork/java/install-jai-imageio-1-0-01-139659.html
27. Shining Light Production, “Win32 OpenSSL.”
28. dcm4che, “dcm4che, a DICOM Implementation in JAVA.” [Online]. Available: http://www.dcm4che.org/
29. Warnock MJ, Toland C, Evans D, Wallace B, Nagy P (2007) Benefits of using the DCM4CHE DICOM archive. J Digit Imaging Off J Soc Comput Appl Radiol 20(Suppl 1):125–129
30. O. computer science Institute, “DCMTK - DICOM Toolkit.” [Online]. Available: http://dicom.offis.de/dcmtk.php.en
31. OBA Vasquez, S Bohn, M Gessat “Evaluation of Open Source DICOM Frameworks”
32. Freymann JB, Kirby JS, Perry JH, Clunie DA, Jaffe CC (2012) Image data sharing for biomedical research–meeting HIPAA requirements for De-identification. J Digit Imaging Off J Soc Comput Appl Radiol 25(1):14–24