Project title: Implications of Medical Low Dose Radiation Exposure

Grant Agreement Number: 755523

Call identifier: NFRP-2016-2017

Topic: NFRP-9

Deliverable D2.3
Report on system integration and publication of the basic user guidelines

Lead partner:	UNIGE
Author(s):	UNIGE - Osman Ratib, Nicolas Roduit, Joël Spaltenstein
	b<>com - Eric Guiffard, Guillaume Pasquier, Emmanuel Cordonnier
Work Package:	WP2
Estimated delivery:	30/11/2018
Actual delivery:	30/11/2018
Type:	Report
Dissemination level:	Public

This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 755523.

Table of contents

List of figures

List of tables

Abbreviations

API	Application Programming Interface
BFO	Basic Formal Ontology
CID	(DICOM) Context Group Identifier
CT	Computed Tomography
CTDI	Computed Tomography Dose Index
DCM	DICOM Content Mapping resource
DICOM	Digital Imaging and Communications in Medicine
FHIR	Fast Healthcare Interoperability Resources
IOD	Information Object Definition
IRDBB	Image and Radiation Dose BioBank
IRI	Internationalized Resource Identifier
ITMI	Institute of Translational Molecular Imaging
LOINC	Logical Observation Identifiers Names and Codes
NM	Nuclear Medicine
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
RDF	Resource Description Framework
SNOMED	Systematized Nomenclature of Medicine
SPARQL	Simple Protocol And RDF Query Language
SPECT	Single Photon Emission Computed Tomography
SOP	(DICOM) Service Object Pair
SQL	Structured Query Language
SR	(DICOM) Structured Reporting
TID	(DICOM) Template Identifier
UML	Unified Modelling Language
W3C	World Wide Web Consortium
WADO	(DICOM) Web Access to DICOM Objects
XML	eXtensible Markup Language
XSD	XML Schema Definition

Technical terms glossary

Ansible	A computer program for descriptive server provisioning. Users provide an Ansible Playbook that describes the final state of a server - what software needs to be present for example - and the Ansible software will perform the necessary steps, including software installation, in order to ensure that the state of the server matches the state described in the Playbook.
DICOM Context Group Identifier (CGI)	DICOM Context Groups specify Value Set restrictions for the use of codes in the DICOM standard. DICOM Context Groups are referred to using Context Group Identifiers.
DICOM Content Mapping resource (DCM)	DICOM uses as much as possible existing terminology resources (such as SNOMED CT, LOINC, UCUM). However, DICOM manages its own terminology resource called DICOM Content Mapping resource.
Docker	A computer program that provides virtualization services.
Docker Container	A virtualized server, instantiated on the basis of a Docker Image, under the control of the Docker software. Multiple Docker Containers can run simultaneously on a single server.
Docker Image	A snapshot of a fully configured server with all dependent software installed. Can be instantiated into a running Docker Container by the Docker software.
Information Object Definition (IOD)	In DICOM, an Information Object Definition is an information object that can be exchanged between DICOM Application Entities.
Image and Radiation Dose BioBank (IRDBB)	The Image and Radiation Dose BioBank is a resource for managing image and dose data in an integrated way. IRDBB supports both DICOM data and non-DICOM data. The IRDBB software supports importation, internal management and query/retrieval of MEDIRAD research data.
Ontology Web Language (OWL)	The Ontology Web Language is one of the main standard languages developed by the World Wide Web Consortium (W3C) for representing ontologies. It is based on Description Logics (DL), a term that denotes a family of knowledge representation languages.
Resource Description Framework (RDF)	The Resource Description Framework (RDF) is a family of W3C specifications used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntaxes. In practice, RDF is a directed, labeled graph data format for representing information in the Web, that is also used in knowledge management applications. [adapted from Wikipedia]
Simple Protocol And RDF Query Language (SPARQL)	The SPARQL specification defines the syntax and semantics of the SPARQL query language for RDF. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs. [adapted from Wikipedia]
DICOM Service Object Pair (SOP)	A DICOM Service Object Pair is a specification of a service in the DICOM standard. It associates an object class (i.e. an Information Object Definition) to a set of services that can be applied to this object.
DICOM Structured Reporting (SR)	DICOM Structured Reporting is a specification of a class of objects in the DICOM standard, based on a hierarchical organization of observations (called SR tree). DICOM SR is used for representing a broad range of documents such as reports of imaging procedures,

	procedure logs, reports of Computed Assisted Detection (CAD) analyses, and radiation dose structured reports.
DICOM Template Identifier (TID)	A DICOM SR Template specifies the structure of a sub-graph of a DICOM SR tree. It is represented as a table describing the characteristics of each node (called content item) of the SR sub-graph, such as: nesting level, type of node (e.g. CODE, TEXT, PNAME, etc.), type of relation with parent node, multiplicity (e.g. unique or multiple), requirement type (i.e. mandatory or optional), restrictions on content (for CODE nodes). DICOM SR Templates are referred to by means of DICOM Template Identifiers, e.g. to specify their use in a particular Structured Report IOD.
Unified Modeling Language (UML)	Unified Modeling Language is a general-purpose, developmental, modeling language in the field of software engineering that is intended to provide a standard way to visualize the design of a system. Various kinds of models can be represented in UML, including Class Diagrams, Sequence Diagrams, Use Case Diagrams etc. [adapted from Wikipedia]
DICOM Web Access to DICOM Objects (WADO) or DICOMweb	DICOM Web Access to DICOM Objects (one of the serviced included in the RESTful communication also called DICOMweb™) is a term applied to the family of RESTful DICOM services defined for sending, retrieving and querying for medical images and related information. DICOMweb provide a light-weight mobile device and web browser friendly mechanism for accessing images, which can be implemented by developers who have minimal familiarity with the DICOM standard and which uses consumer application friendly mechanisms like http, JSON and media types (like "image/jpeg") to the maximum extent possible. The standard is formally defined in DICOM PS3.18 Web Services. [adapted from Wikipedia]
eXtensible Markup Language (XML)	The Extensible Markup Language (XML) is a markup language of the W3C that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures such as those used in web services. Several schema systems exist to aid in the definition of XML-based languages, while programmers have developed many application programming interfaces (APIs) to aid the processing of XML data. [adapted from Wikipedia]
XML Schema Definition (XSD)	XSD (XML Schema Definition) is a recommendation of the W3C that specifies how to formally describe the elements in an Extensible Markup Language (XML) document. It can be used by programmers to verify each piece of item content in a document. They can check if it adheres to the description of the element it is placed in. Like all XML schema languages, XSD can be used to express a set of rules to which an XML document must conform in order to be considered "valid" according to that schema. [adapted from Wikipedia]
b<>com [Rest DICOM Library]	The IRDBB Import software is developed on the basis of a software owned by b<>com, “b<>com [Rest DICOM Library]”, enabling transfer of medical imaging over the web, fully compliant with DICOMweb™, and “pure web, zero foot print”, that is to say deployed on the client side without any installation of software on the user’s computer.

1. References

Ref	Document name	Version	Date
[1]	Note to the Data Management Board, Consistent identification of entities in the IRDBB system		13/6/2018
[2]	Specification of DICOM import and central repository communication protocol	V1.0	28/9/2017

2. Introduction

This document reports the progress of system integration and delivers some basic user guidelines for the repository that will host the images and data of the *Image and Radiation Dose BioBank* (IRDBB).

This development of the team of the Institute of Translational Molecular Imaging (ITMI) of the University of Geneva (UNIGE) was undertaken in close collaboration with b<>com, who developed the Web Interface based on its in-house product [Rest DICOM Library], the INSERM team from Rennes who developed the Metadata repository, and the Vital-IT team from Geneva who assisted in the set up and implementation of the server [2].

This report first describes the integration process of different software components into the IRDBB system. Then it provides the basic description of the user interface to assist the users in their routine usage.

3. System integration

The IRDBB software development is done by b<>com at their offices in Rennes. b<>com packages the software, including the b<>com [Rest DICOM Library] as well as new modules developed specifically for MEDIRAD, into Docker Images which are then downloaded from a Docker Image repository hosted by b<>com. b<>com then provides an Ansible Playbook that, when executed, performs all necessary server configuration steps, including downloading the Docker Images and instantiating the Docker Containers.

ITMI, along with partner Vital-IT, provides server hardware on which the software designed by b<>com runs. Vital-IT further provides a relational database and file archive that are fully backed up on a daily basis. The Ansible Playbook configures the Docker Containers such that all persistent data is stored either to the database or to the file archive, depending on the needs of the specific Docker Container. This configuration separates two significant concerns, namely, the execution of server software programs, and the storage of data. In the event of software updates or software malfunction, the Docker Containers can be easily re-instantiated on the basis of the Ansible Playbook with minimal delay and without any further configuration or loss of data.

4. Basic user guidelines

The IRDBB-UI is an application developed to easily and securely upload de-identified imaging data and relevant dose information into a central biobank hosted by University of Geneva. In order to reach the purpose, the IRDBB-UI offers the possibility to import both DICOM and non-DICOM data with associated metadata. This document intends to deliver some basic user guidelines to help the user to properly use the application.

This basic user guidelines explain how to perform the following operations:

- User authentication
- Data importation
- Basic queries

The IRDBB-UI application is available at https://medirad.kheops.online and compatible with the following web browsers:

- Safari 12
- Google Chrome 70
- Internet Explorer 10, 11
- Firefox 63

4.1 User authentication

When connecting to the IRDBB-UI platform, the user shall authenticate using his/her credentials previously received by e-mail. Once authenticated, the user directly accesses to the IRDBB-UI application.

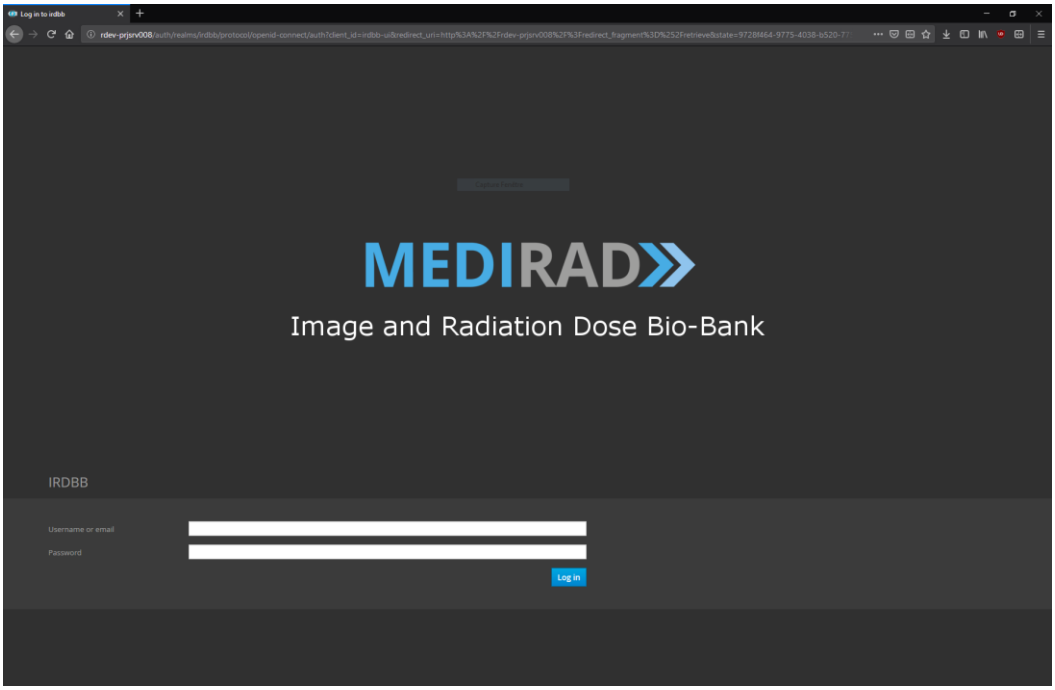

Figure 1: Authentication page

4.2 Data importation

The IRDBB-UI application offers the possibility to import DICOM and non-DICOM data. In case of non-DICOM data importation, the user shall have imported DICOM data first to create a related patient on the MEDIRAD platform. This created patient then allows non-DICOM data importation by establishing a relationship based on the Patient ID.

To perform any data importation, the user first needs to access the "STORE" page of the application which is the default home page of the application. This page is also accessible by clicking on the STORE button in the navigation bar located on top of any page of the application.

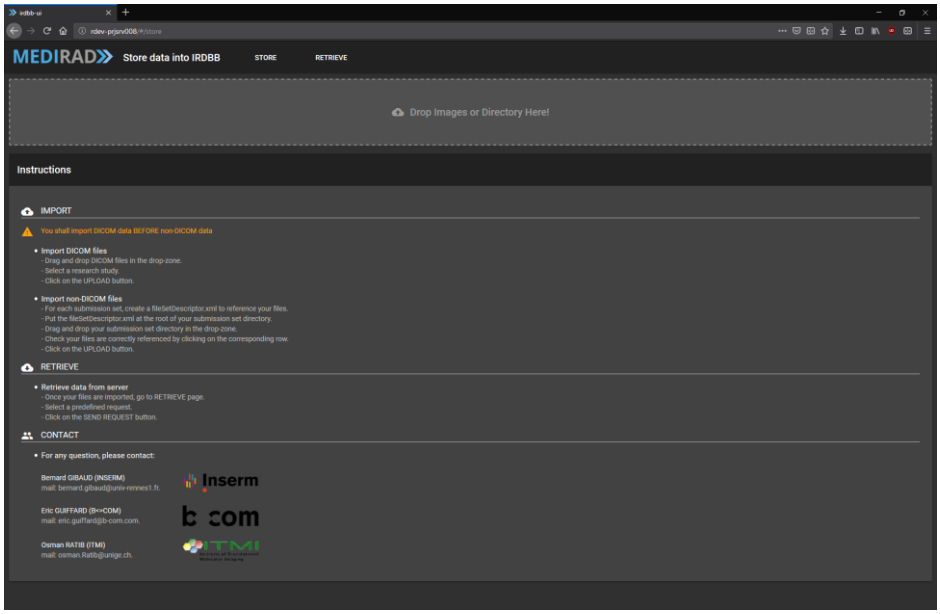

Figure 2: IRDBB STORE page

4.2.1 DICOM data importation

To perform DICOM data importation, the user has to drag and drop DICOM files over the "drop area". DICOM files can be dropped from a multiple selection or gathered into a directory.

The current version of IRDBB-UI only supports the following DICOM modalities:

- CT
- SR

DICOM data are then processed and displayed grouped by DICOM Study Instance UID. The user shall relate each of them to a *Patient Pseudo* and a MEDIRAD *Clinical Research Study* inside the MEDIRAD platform.

According to the rules defined by the MEDIRAD Data Management Board, the Patient Pseudo should be composed of four different information, separated by a hyphen [1]:

- MEDIRAD Project ID (i.e. 755523)
- Clinical Research Study ID (i.e. st232)
- Partner Number (i.e. UOC)
- Patient Number (i.e. 204)

A possible Patient Pseudo could then be 755523-st232-UOC-204.

The Clinical Research Study related to the imported DICOM data has to be selected among the provided list. More details on the provided Clinical Research Studies are available in the Basic queries part of these user guidelines.

In case of a patient involved in multiple Clinical Research Studies, the Patient Pseudo shall be defined during the first DICOM data import referring to this patient. This Patient Pseudo shall remain the same across the further imports, regardless of the Clinical Research Studies to which the data pertains.

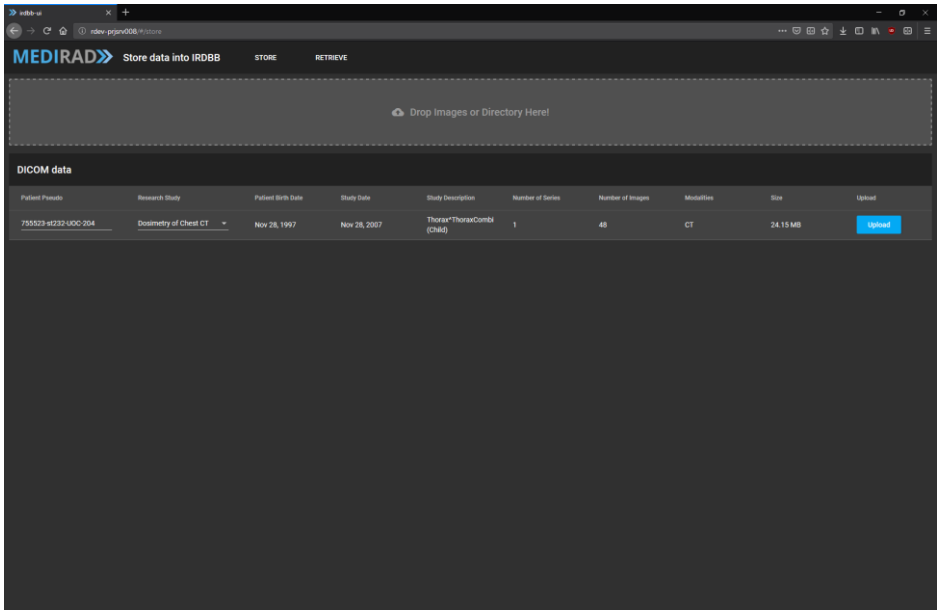

Figure 3: DICOM Data import

Finally, the user must click the UPLOAD button to start uploading to the MEDIRAD platform. Once the import is finished, a notification informs the user about the importation status. The Patient Pseudo is encoded into an *Encoded Patient Id* to ensure the confidentiality of inherent information. This Encoded Patient Id is displayed to the user in a dialog window.

It is the responsibility of each clinical site to maintain the relationship between the local patient identification (e.g. DICOM Patient ID) and the Encoded Patient Id used inside the MEDIRAD platform for all imported DICOM data. To do so, the Encoded Patient Id shall be saved and linked to the corresponding local patient identification.

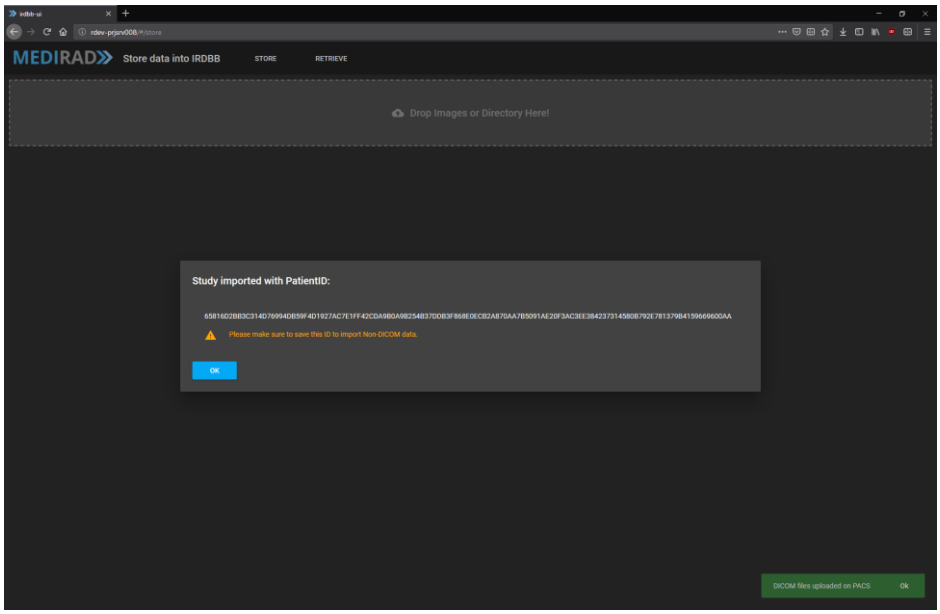

Figure 4: Encoded Patient Id displayed to the user

4.2.2 Non-DICOM data importation

Non-DICOM data importation applies to one or more *Non-DICOM File Sets*. Each Non-DICOM File Set shall be composed of a single *fileSetDescriptor* and an arbitrary number of non-DICOM files, referenced in the fileSetDescriptor. This fileSetDescriptor must be placed at the root of the Non-DICOM File Set.

The user should make sure that the fileSetDescriptor refers to the right patient by using the saved Encoded Patient Id from previous DICOM importation. As a reminder, a DICOM files importation related to the same patient must be performed before any non-DICOM importation.

For more details about how to write this fileSetDescriptor, please refer to the Non-DICOM metadata file part of these user guidelines.

To perform non-DICOM data importation, the user has to drag and drop Non-DICOM File Sets over the "drop area".

Non-DICOM data are then processed by the IRDBB-UI application and displayed to the user grouped by Non-DICOM File Sets. The Patient Pseudo and the Clinical Research Study are also displayed so that the user can check that the Non-DICOM File Sets refer to the right patients and the right Clinical Research Study. The user can also verify that all files are properly referenced in the fileSetDescriptor by clicking the row corresponding to a Non-DICOM File Set. The referenced files are highlighted in green and the non-referenced files highlighted in red. Only referenced files will be imported.

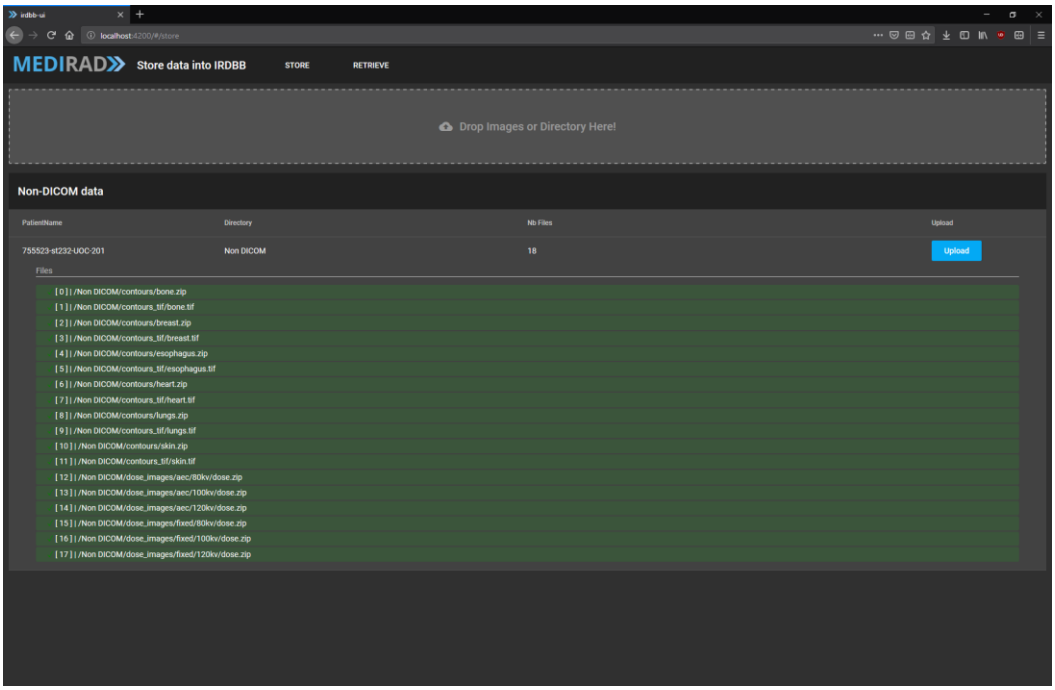

Figure 5: Non-DICOM data import

Finally, click the Upload button to start uploading to the MEDIRAD platform. Once the import is finished, a notification informs the user about the importation status.

4.2.3 Non-DICOM metadata file

The fileSetDescriptor file has been designed to provide a common metadata structure shared by every clinical site. Any Non-DICOM File Set must include a fileSetDescriptor XML file, containing metadata associated to each file. This fileSetDescriptor must be named **"fileSetDescriptor.xml"** and provided at the root of the Non-DICOM File Set to start the import.

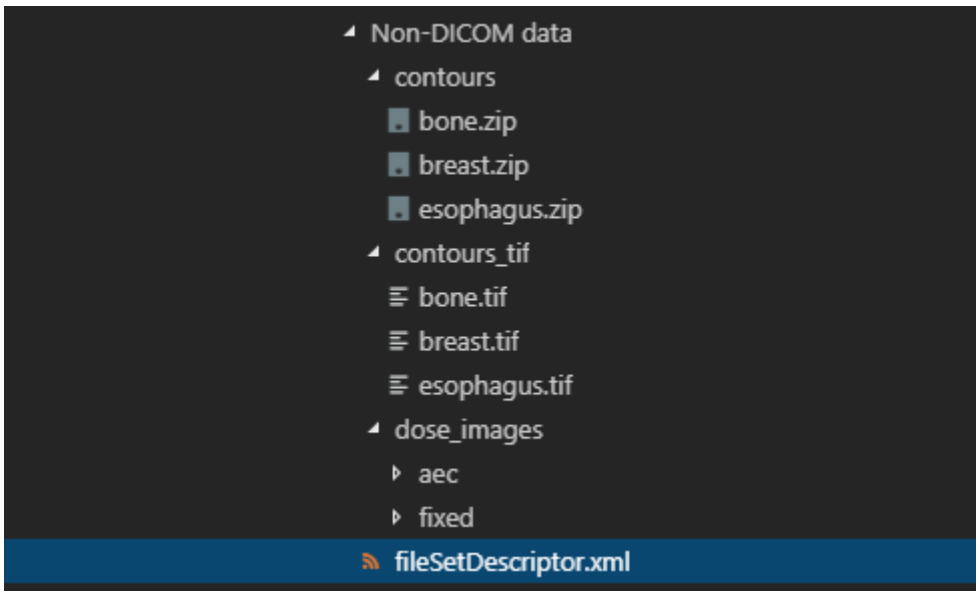

Figure 6: Example of Non-DICOM File Set hierarchy

Before non-DICOM data importation, the user must make sure that the fileSetDescriptor refers to the right patient by using the saved Encoded Patient Id from prior DICOM importation. To do so, copy the related Encoded Patient Id into the <irdbb:PatientId></irdbb:PatientId> tag.

```
<?xml version="1.0" encoding="UTF-8"?>
<irdbb:NonDicomFileSetDescriptor
    xmlns:irdbb="https://www.irdbb-medirad.com"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="./WP2MS17-V3.xsd">
    <irdbb:ReferencedClinicalResearchStudy>
        <irdbb:ClinicalResearchStudyID>Clinical Research Study subtask 2.1.2</irdbb:ClinicalResearchStudyID>
        <irdbb:ClinicalResearchStudyTitle>Dosimetry of Chest CT</irdbb:ClinicalResearchStudyTitle>
    </irdbb:ReferencedClinicalResearchStudy>
    <irdbb:PatientId>ENCODED PATIENT PSEUDO</irdbb:PatientId>
```

Figure 7: Encoded Patient Id reference in fileSetDescriptor.xml

It is the responsibility of each clinical site to generate the fileSetDescriptor files from its own data. The XSD schema of the fileSetDescriptor will be available through the IRDBB-UI application.

FileSetDescriptor files provide detailed information about the meaning and provenance of both image and dosimetric data. Therefore they are quite complex, and it is strongly recommended that each site develops a specific software application to automate the translation of their local data into these fileSetDescriptor files.

An alternative may also be a manual edition or generation using a XML editor software, such as XML Spear (http://www.donkeydevelopment.com/) that can be used to simplify edition and validation of the fileSetDescriptor file over the “fileSetDescriptor.xsd” schema.

4.3 Basic queries

The IRDBB-UI application offers the possibility to retrieve DICOM and non-DICOM imported data and query the related metadata using pre-defined requests based on the SPARQL protocol.

To perform any request, the user first needs to access the “RETRIEVE” page by clicking on the RETRIEVE button in the navigation bar located on the top of any page of the application.

The retrieve page includes a RETRIEVE section presenting the list of available requests with an associated description. To execute any request, simply select the request and click the SEND QUERY button.

Results are then presented in a table into the Response section. The user can filter results by using the three filter options located above the result table. The number of items per page can be configured and results can also be reordered by clicking the column headers.

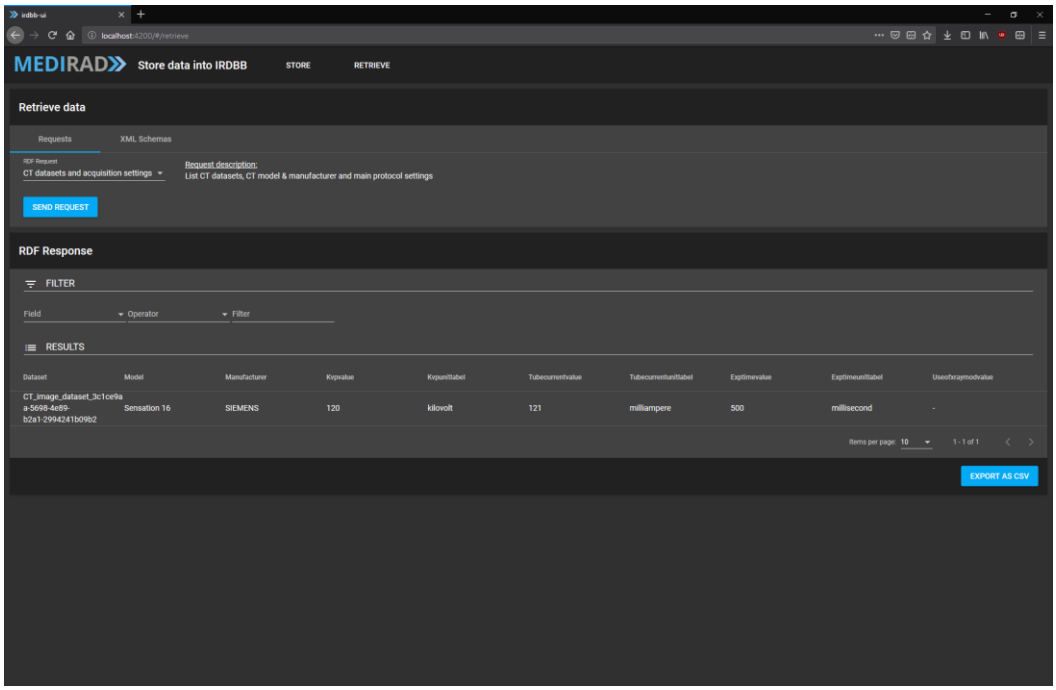

Figure 8: Request DICOM metadata

DICOM and non-DICOM data can respectively be retrieved using the “CT dataset and handles” and “Non-DICOM datasets and handles” queries, allowing the user to download results individually by clicking the DOWNLOAD button presented in the results table.

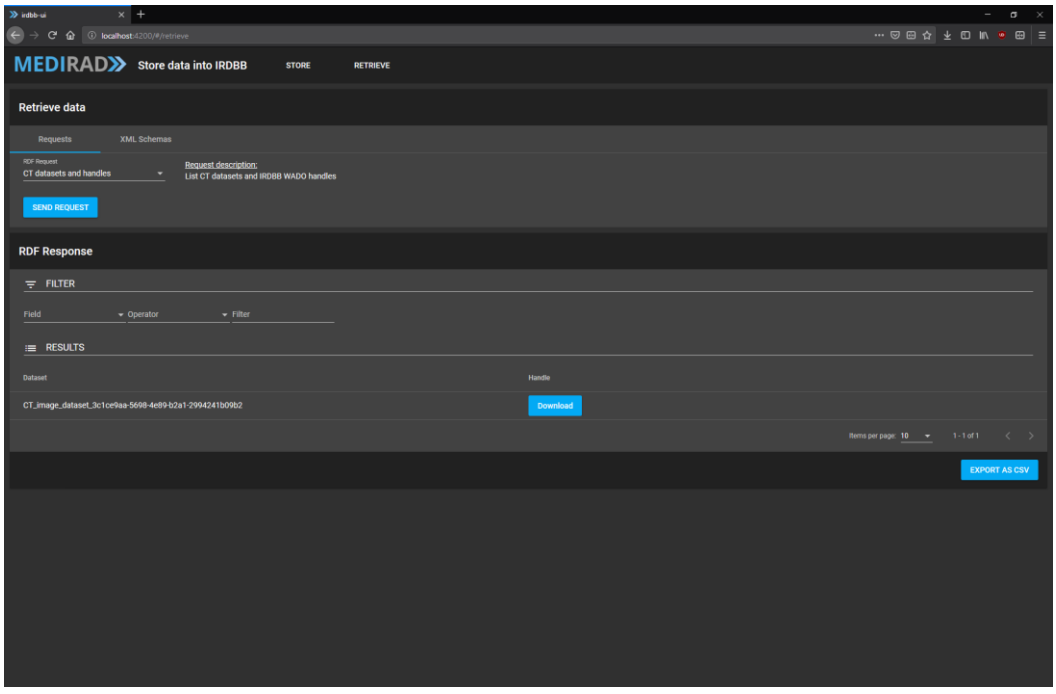

Figure 9: Retrieve DICOM data using “CT dataset and handles” request

Finally, the results can be exported as CSV to be saved by the user using the EXPORT AS CSV button located below the result table.

The currently available requests are listed hereinafter with the associated result fields and description. Please note that this list is preliminary and will probably be extended to better match the users' views.

Requests	Fields	Comments
CT Datasets and acquisition settings	Datasets, Model, Manufacturer, kvpvalue, kvpunitlabel, Tubecurrentvalue, tubecurrentunitlabel, exptimevalue, exptimeunitlabel, useofxraymodvalue	Allows to list datasets ID and related acquisition settings.
Patient and demographic data	Human, Id, Sex, BirthDate	Allows to list patients and their basic demographic data: Human field is an anchor towards Human resource is the RDF repository; ID is the Patient ID (used in datasets)
Processing steps	Process, Date, Time, Processclass Labelclassroleinstitut, nameinstitut	Allows to list the processes having led to some files or measurement data; it is part of what is called "provenance information", i.e. details of the process that led to its generation (e.g. an image acquisition or a data processing)
Mean absorbed dose in organs	PatientId, Classdoselabel, Processclass, Value, Unitclasslabel, Matentity, Matentityclasslabel	Allows to list Mean absorbed doses in organs and related organ and patient (Patient Id is the same piece of information than ID in "Patient and demographic data")
Clinical research studies	Study, Id, Name, Description	Allows to list the Clinical Research Studies currently available
Non-DICOM datasets and Handles	PatientId, Dataset, Datasetclasslabel, Fhirhandle	Allows to list the non-DICOM datasets, their associated type and a handle allowing to download their content
Volumes Of Interest and organs	Dataset, Filename, Voiclass, Matentity, Matentityclasslabel, Processclass, Date, Time	Allows to list the Volumes Of Interest (VOI), the corresponding material entity (i.e. organ) and details of their provenance (in principle segmentation process)
CT datasets and handles	Dataset, Handle	Allows to list the CT datasets and a handle allowing to download their content.
3D Dose Maps and provenance data	Dataset, Process, Softwarename, Inputct, Kvpvalue, Kvpunitlabel, Useofxraymodvalue	Allows to list the 3D Dose Maps and details of their provenance (simulation process, CT data in input, and basic simulation parameters)

Table 1: List of currently available requests

4.4 Support of additional DICOM modalities

The importation mechanism developed by b<>com, Inserm and ITMI allows in principle to import any kind of DICOM data. The limitation arises from the need of ensuring minimal indexing of this data in

the Semantic Database, so that this data can be retrieved. This translation of DICOM metadata into RDF requires significant programming work, this is why we asked the users to specify the essential metadata only, i.e. those that are critically important to query/retrieve the data (knowing that once they have retrieved the DICOM data, the users can access all the DICOM metadata).

In the short term, it was agreed with b<>com to modify the current implementation so that the extraction of DICOM metadata be completed directly in the Semantic Translator, rather than in IRDBB-UI. This should facilitate addressing the needs of the users to support new modalities in the future.

Concerning the support of additional DICOM metadata required by University of Crete (J. Damilakis and coll.) after the preliminary tests related to Milestone MS17 (M12), they will be integrated as part of the re-engineering of the extraction of DICOM metadata within the Semantic Translator. We will do our best to achieve that by the end of the year 2018.

ANNEX 1

Clinical research study ID	Clinical research study name	Responsible partner
755523_st212	Development of a novel method to estimate patient organ dose from chest CT	UOC
755523_st2321	Estimation of patient organ doses from chest CT used in multi-modality systems	UGENT
755523_st2322	Estimation of patient organ doses from two commonly used PET and SPECT tracers	INSERM
755523_t32	Biokinetic modelling and treatment planning – I-131 ablation of thyroïd	RMH
755523_t33	Dosimetry calculation - I-131 ablation of thyroïd	INSERM
755523_st531	Doses for cohort members from extension of follow-up	ISGLOBAL
755523_st532	Estimation of doses for sujects in the case-control study	ISGLOBAL

Table 2: Clinical Research Study List