INSIDE DB Specification

Version: 0.1 (incomplete)

Date: 3/4/2018

Authors: Giancarlo Sportelli (GS)

Introduction

The INSIDE PET/Tracker system has been used in the last few years to monitor hadron beams and treatment plans at the Centro Nazionale di Adroterapia Oncologica (CNAO) synchrotron facility in Pavia. Its in-beam PET was tested during 2016 with polymethyl methacrylate (PMMA) and anthropomorphic phantoms [Piliero et al., 2016; Bisogni et al., 2017] and its response was thorougly characterized. The same setup was also used in December 2016 to perform the first online monitoring test during a conventional patient treatment session.

Early results obtained with the INSIDE project have raised clinical interest in the possibility of monitoring in real time the anatomical modification in the context of the planning target volume (PTV) and near organs at risk (OAR). The envisaged clinical impact comes from having a feedback on the goodness of the treatment plan, i.e., on the coverage completeness of the target, for each treatment fraction. In particular, the system would allow to intervene in time on the PTV in case of unexpected monitoring feedback, thus preventing any undesired collateral effects.

At the state of the art, clinicians compare the simulated CT (baseline) with the re-evaluation CT acquired during the radio-therapy (RT) in order to find any possible anatomical modification that could impact the distribution of the planned dose. This comparison is not done at each treatment fraction. Modifications could be caused by patient weight loss, mucus stagnation or inflammatory (e.g., radio-induced) processes. The same response to the disease and to the treatment could induce anatomical modifications in unexpected ways. For example, there are new clinical evidences that show how the regression of adenoid cystic carcinoma and squamous-cell carcinoma, is particularly rapid when the two diseases are treated with carbon and proton ions, respectively. Their fast regression must be monitored during each treatment fraction. Therefore, these kinds of disease are good candidates for testing the enhanced monitoring features provided by the INSIDE system.

The CNAO is planning for conducting a longitudinal study to assess systematically the early response of disease during the treatment and any anatomical modifications that have significant degrading effects on the goodness of the dose distribution. The study will be conducted on conventional clinical treatments in 2018.

The CNAO longitudinal study is not the first one conducted with in-vivo PET imagers, see, e.g., the MIRANDA study [Nischwitz et al., 2015]. However, it presents the following unique features: 1) it is the first one that uses in-beam PET data, which can provide additional statistics and hence better signal-to-noise ratios without any additional room occupancy; 2) it will be cross-validated with the integrated particle tracker; 3) it will use TOF-grade PET data for reduced artefact due to incomplete angular coverage.

Such a study will involve professionals and researchers from different research bodies for several months. A large amount of data will be collected and analysed by the involved partners. These data will have to

be stored reliably and efficiently, and queried effectively. Since it will have references to patients' data, it will also have to be managed in a way that guarantees privacy and allows controlled sharing.

As of today, no integrated data management systems have been used for such purposes.

Objectives of the project

This project aims at designing and deploying a database management system (DBMS) for collecting, processing and sharing the data collected in the framework of the INSIDE project and with special focus on the longitudinal clinical study planned for 2018 at CNAO.

The DBMS will provide a software front-end application and a back-end server that will support data collection during the treatments and data revision and sharing after it. This will be achieved upon the completion of the following sub-objectives:

- Definition of a data collection protocol
- Design and development of a server architecture
- Design and development of a database schema
- Design and development of a Create/Read/Update/Delete (CRUD) software front-end application
- Design and development of an automated procedure for safeguarding patients' privacy and respecting local ethics regulation
- Ensure the safety and reliability of the acquired data
- Allow fine control on access rights to the data

Implementation

Overview of data collection and management methodology

A data acquisition procedure will be described in the form of a protocol document (i.e., the acquisition protocol). The protocol will be available to the researchers in charge to operate the PET/Tracker system during the treatments. Correspondingly, a software application will be available to the researchers, which will guide them through the steps of the acquisition protocol. We refer to the raw data coming from the PET/Tracker system, simulated acquisitions and results of processing algorithms as simply the *data*. All information that describe how such *data* have been acquired, simulated or calculated are referred to at *metadata*. Examples of *data* are: files containing lists of photon-pairs that will be later used for image reconstruction (e.g., list-mode coincidence files); reconstructed images; simulated PET data. Examples of *metadata* are: configuration files fed to the acquisition hardware; log files; descriptive notes of who was operating the hardware; hardware/firmware and software versions of the acquisition systems. All data and metadata will be shared remotely to the institutions participating to the project according to a predefined set of rules that will take into account the needs for privacy, safety and reliability of the database.

We refer to *data taking* as the process of collecting data and metadata during the treatment using the provided software application. The result of a data taking session is a *data draft*. A data draft can be saved and revised by its authors as needed. Once the data draft is ready, its authors may *submit* it along with the data to the central servers that will be physically located at INFN Sezione di Pisa. Once submitted, the data will be integrated within the central database. INFN will take care of data storage safety and reliability. Once in the central servers, data and metadata will be available to the other participants to the project according to the predefined access rules. Further revisions on the data will be

performed directly on the central database. All the insertion, update and delete operations made on the central database during such revisions will be logged and associated with the revising user.

All the data and metadata available on the central server can be used in order to produce *derived data*. All the derived data shall be reproducible by the partners using only the data available in the central server.

Data taking workflow

Data taking can be divided into the following sequential stages:

- [STAGE 1] Creation of the data draft
- [STAGE 2] Status data retrieval before the treatment
- [STAGE 3] Acquisition data retrieval after the treatment
- [STAGE 4] Clinical data retrieval
- [STAGE 5] Data draft revision and submission

[STAGE 1] Creation of the data draft

The user creates a new draft corresponding to a new treatment fraction on patients or a scientific test on phantoms. This may happen, e.g., the day before the treatment. (S)he then inserts the following metadata, or at least all those that are available:

- a unique reference to the treatment target
- CT scans of the target, possibly in DICOM format
- clinical information on the target type
- treatment planning data/beam delivery data
- simulated data relating to the target/treatment
- system status logs (with timestamp)
- system calibration data
- system acquisition configurations

[STAGE 2] Status data retrieval before the treatment

The user opens the draft and adds/updates the following metadata, then saves and closes the draft (this stage may happen, e.g., a few hours before the treatment):

- system status logs (with timestamp)
- detectors/tracker/bed position (note: these positions in our case may be always the same, hence might be automatically pre-filled)

[STAGE 3] Acquisition data retrieval after the treatment

The user opens the draft and uploads the following metadata (this stage should happen right after the treatment):

- time reference to the start of the treatment
- time reference to the start of the PET/Tracker acquisitions
- acquisition log files
- PET master_board and TX_boards configuration files

- reference calibration
- treatment duration
- acquisition duration
- treatment/beam delivery log files (made available by CNAO servers, modality TBD)

The user also selects the following locally available data for off-line transfer to the central server, then saves and closes the draft:

- raw acquisition files (binary files, size from some MB to some GB)

[STAGE 4] Clinical data retrieval

The user opens the draft and adds/updates the following data, then saves and closes the draft (this stage may happen, e.g., a few hours after the treatment):

- Reconstructed DICOM 3D images of the acquisition

[STAGE 5] Data draft revision and submission

The user opens the draft and checks its data and metadata. (S)he then saves and submits the draft (this stage may happen, e.g., the day after the treatment). Once submitted, the draft is transferred off-line to the central server and its local copy is eliminated. Local copy elimination is required at this stage in order to facilitate data coherence checks later on.

Data analysis and processing workflow

All data and metadata are available to the project partners for download and analysis. The results obtained by analysis and processing can be uploaded as related derived data. A schema is defined below that defines the aggregations of data, metadata and derived data, referred to as *models*, and the associations between them.

Data models

A schematic diagram of the data models is reported in Figure 1[TO BE FINISHED]. A technical description of the contents type of each model is reported in the following subsections.

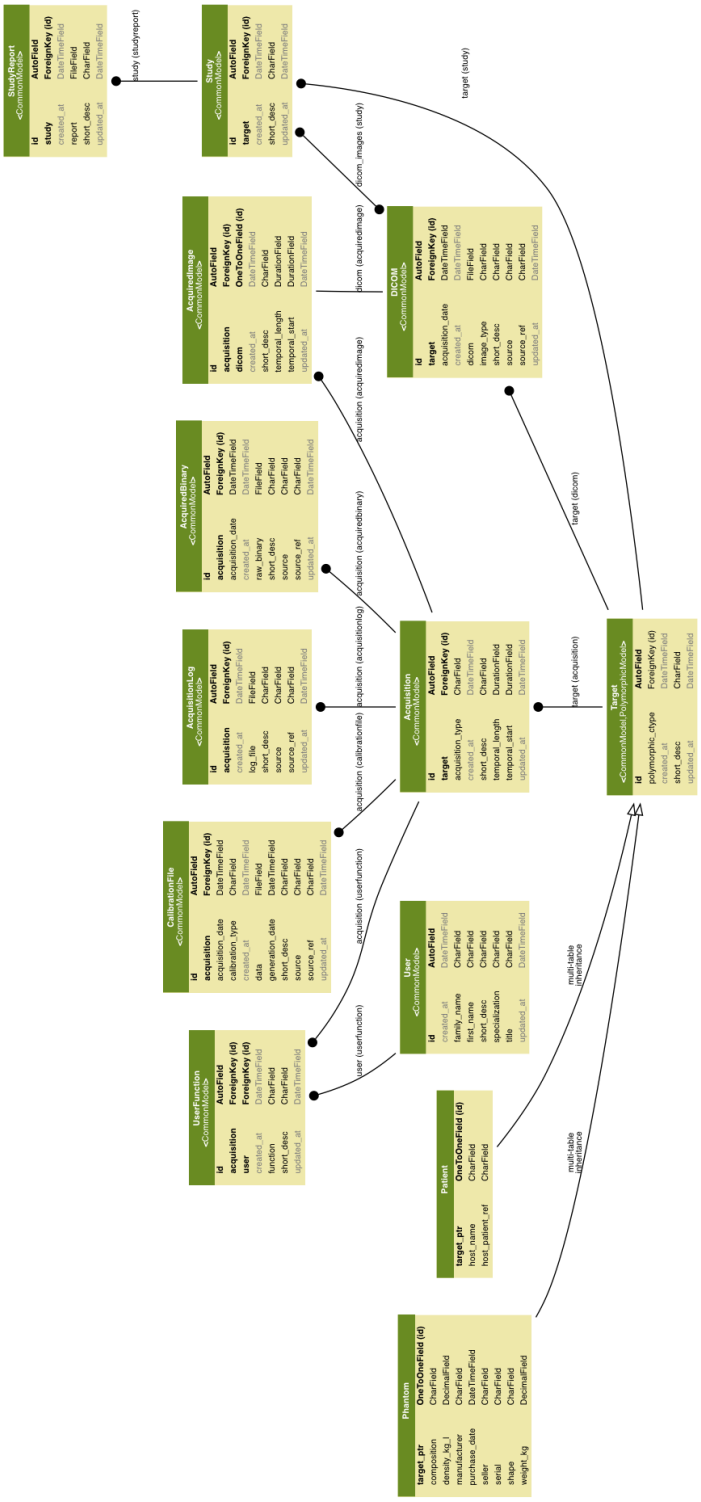

Figure 1: INSIDE database schema

Model Target

Field	Relation	Type
DICOM	many	Model
Acquisition	many	Model
Study	many	Model

Sub-model Phantom:

Field	Relation	Type
serial	attribute	string
shape	attribute	string
composition	attribute	string
weight_kg	attribute	decimal
density_kg_l	attribute	decimal
manufacturer	attribute	string
seller	attribute	string
purchase_date	attribute	datetime

Sub-model Patient:

Field	Relation	Type
host_patient_ref	attribute	string
host_name	attribute	string

Model DICOM

Field	Relation	Type
acquisition_date	attribute	datetime
source_ref	attribute	string
source	attribute	string
image_type	attribute	choice (1)
dicom	attribute	file
AcquiredImage	one	Model
Study	many	Model

(1) image_type choices are: ct, pet, mri, spect, treatment_plan, range_verification.

Model User

Field	Relation	Type
first_name	attribute	string
family_name	attribute	string
title	attribute	string
specialization	attribute	string

Model Acquisition

Field	Relation	Type
temporal_start	attribute	decimal (1)
temporal_length	attribute	decimal
acquisition_type	attribute	choice (2)
AcquiredImage	many	Model
AcquiredBinary	many	Model
CalibrationFile	many	Model
AcquisitionLog	many	Model
User	many	Model (3)

(1) temporal_start represents the seconds from the treatment start.
(2) acquisition_type choices are: measured, simulated.
(3) relationship through UserFunction.

Model UserFunction

Field	Relation	Type
function	attribute	string
User (1)	one	Model
Acquisition (1)	one	Model

(1) reference pairs (User,Acquisition) must be unique.

Model AcquiredImage

Field	Relation	Type
DICOM	one	Model
temporal_start	attribute	decimal (1)
temporal_length	attribute	decimal

(1) temporal_start represents the seconds from the acquisition start.

Model AcquiredBinary

Field	Relation	Type
acquisition_date	attribute	datetime
source_ref	attribute	string
source	attribute	string
raw_binary	attribute	file

Model CalibrationFile

Field	Relation	Type
acquisition_date	attribute	datetime
generation_date	attribute	datetime
source_ref	attribute	string
source	attribute	string

calibration_type	attribute	choice (1)
data	attribute	file

(1) calibration_type choices are: energy, time, threshold, tac_mask, baseline

Model AcquisitionLog

Field	Relation	Type
source_ref	attribute	string
source	attribute	string
log_file	attribute	file

Model Study

Field	Relation	Type
DICOM	many	Model
StudyReport	many	Model

Model StudyReport

Field	Relation	Type
report	attribute	file

Adopted technologies

An application software prototype has been built using the Django 2.0 open-source web framework, which follows the model-view-template (MVT) architectural pattern on top of the structured query language database engine sql-lite3.

Network architecture

The INSIDE DBMS runs on a network of servers and workstations schematically illustrated in

Figure 2.

Data draft will be created and populated by the user workstation, which needs to be connected to all the other network nodes inside the CNAO. When the user submits the draft, the latter will be transferred to the INSIDE DB server for off-line transmission to the central server at INFN. The INSIDE DB server will also host the software web application that will be used by the user workstation. A description of the tentative characteristics of each network node follows in Figure 2.

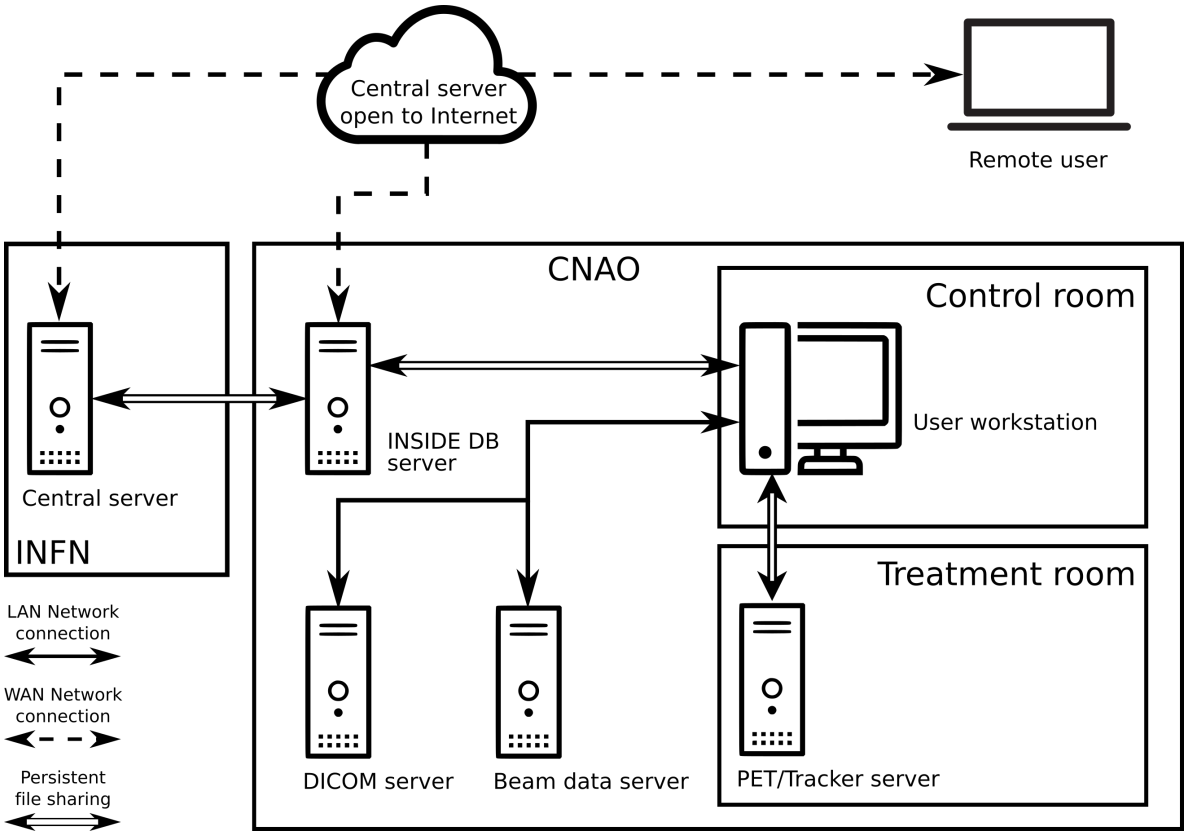

Figure 2: Schematic representation of the network architecture of the machines used by the INSIDE DBMS

Table 1: Tentative specifications of the network nodes

Name	Operating system	Place	Subnet	CPU	RAM	Disk
Central server	Linux	INFN	?	high-end	16 GB	2 TB
INSIDE DB server	Linux	CNAO	?	mid-end	8 GB	10 GB
DICOM server	N/A	CNAO	N/A	N/A	N/A	N/A
Beam data server	N/A	CNAO	N/A	N/A	N/A	N/A
User workstation	Any (Mac?)	CNAO	?	mid-end	8 GB	10 GB
PET/Tracker server	Windows	CNAO	?	high-end	?	?
Remote user	Any	Any	N/A	Any	8 GB	10 GB

Access control and user rights

TBD

Who has access

TBD

Ethics, privacy and regulations

TBD