

Développement d'une ressource de stockage intégrée pour les images et les données dosimétriques (projet MEDIRAD)

Bernard Gibaud¹, Marine Brenet¹, Guillaume Pasquier², Gêrôme Pasquier³, Nicolas Roduit³, Joël Spaltenstein³, Cédric Moubri-Tournes², Nicolas Van Dooren³ et Osman Ratib³

¹Laboratoire du Traitement du Signal et de l'Image, Inserm 1099, Rennes

²Institut de Recherche Technologique b<>com, Rennes

³Institute of Translational Molecular Imaging, Genève

Adresse : Faculté de Médecine, 2 avenue du Pr Léon Bernard 35043 Rennes

Mail : bernard.gibaud@univ-rennes1.fr

1. CONTEXTE ET OBJECTIFS

L'imagerie médicale constitue une source non négligeable d'exposition à des rayonnements ionisants. Par exemple, en France l'exposition « médicale » moyenne est estimée à 1,6 mSv/an, ce qui constitue 35% de l'exposition totale moyenne, estimée à 4,5 mSv/an [1]. Minimiser ces expositions et mieux comprendre leurs effets biologiques sont donc des objectifs importants au regard de la santé des populations. Dans le cadre du programme EURATOM, le projet européen MEDIRAD (Grant Agreement no. 755523) porte sur différents aspects relatifs aux faibles doses occasionnées par des actes d'imagerie médicale, abordés sous des angles multiples (physique, biologique, épidémiologique, relatif à la qualité d'image). Ce projet comporte plusieurs volets expérimentaux au cours desquels des images seront acquises avec différents imageurs en radiologie et en médecine nucléaire et pour lesquelles des doses absorbées seront calculées avec différentes méthodes.

L'objectif du développement présenté ici était de concevoir un système informatique permettant le partage au sein du consortium MEDIRAD des images et des données dosimétriques relatives aux expositions à des rayonnements ionisants liées à ces procédures d'imagerie médicale. L'originalité du travail découle : 1) du souhait de gérer ces données dans un environnement intégré, 2) de faire en sorte que ce développement informatique puisse être réutilisé pour couvrir des besoins similaires dans d'autres environnements, 3) que le système facilite au maximum la réutilisation ultérieure des données, conformément aux concepts de biobanque d'images [2] et de principes F.A.I.R. de partage de données [3].

2. METHODES

2.1. Données à gérer

On peut distinguer deux grands types de données à gérer :

- Les données produites par les équipements d'imagerie médicale à l'origine de ces expositions : il s'agit des scanners X, des systèmes de médecine nucléaire – Tomographie d'Emission Mono-Photonique (TEMP) et Tomographie par Emission de Positons (TEP), et le cas échéant les systèmes hybrides. Ces données comprennent d'une part les images elles-mêmes et d'autre part des documents appelés comptes rendus de dose (en anglais Radiation Dose Structured Report ou RDSR).

Ces données ont la caractéristique d'être représentées selon le standard DICOM, donc leur récupération à partir des systèmes imageurs est facile.

Les données produites par les chercheurs du MEDIRAD, qui quantifient les doses absorbées par les tissus du fait des expositions liées à ces procédures d'imagerie médicale, principalement :

- Exposition aux rayons X lors de Scanner X en pédiatrie
- Injection de radio-pharmaceutiques en imagerie hybride TEMP-CT et TEP-CT
- Injection de ^{131}I pour le traitement du cancer de la thyroïde (radiothérapie interne).

2.2. Architecture

L'architecture du système intégré - appelé *Image and Radiation Dose Biobank* (IRDBB) - est représentée Figure 1.

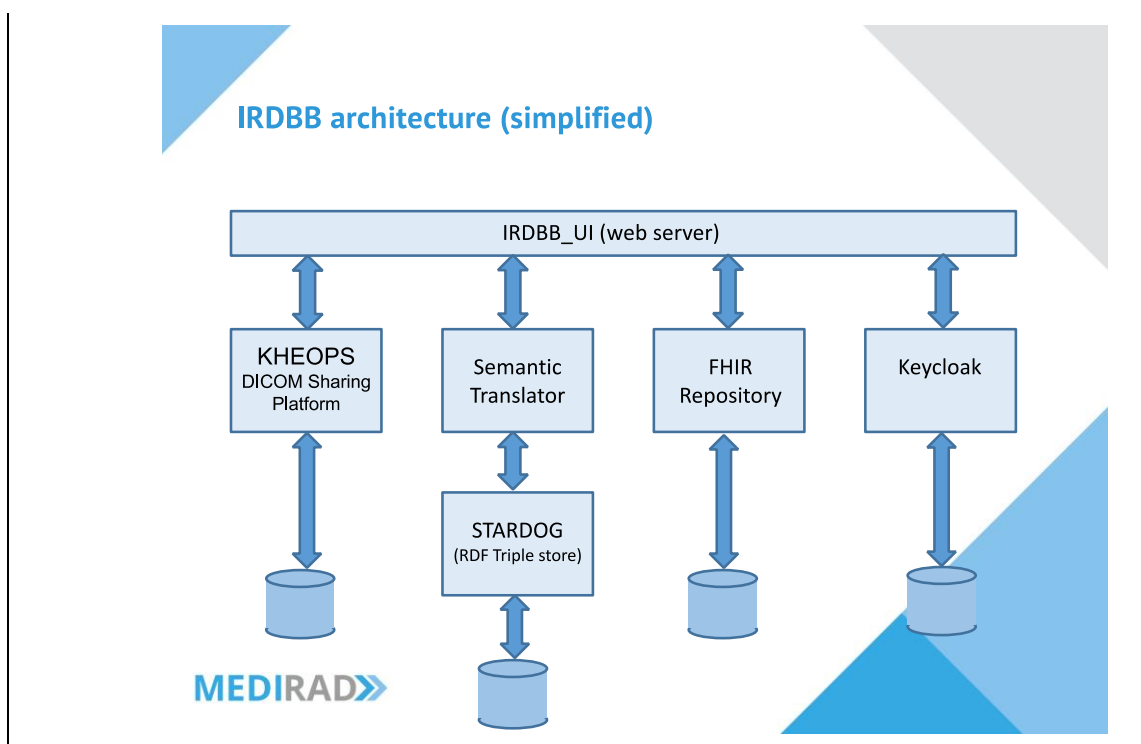


Figure 1 : Architecture of the IRDBB system

Elle met en évidence six composants principaux :

1. Un serveur web, appelé *IRDBB_UI*, gérant l'interaction avec les utilisateurs, notamment pour l'import des données et le requêtage de la base de données ;
2. Un gestionnaire de données DICOM, appelé *KHEOPS*: il est basé sur le logiciel DCM4CHE [4] et constitue le socle d'un système de stockage partagé dédié à la recherche en imagerie à l'Hôpital Cantonal de Genève ;
3. Un gestionnaire de fichiers, appelé *FHIR repository*, dédié au stockage des données non-DICOM, basé sur le standard FHIR [5] ;
4. Un gestionnaire des données sémantiques, appelé *Semantic Translator*, chargé de peupler une base de données sémantique décrivant les images et les données dosimétriques qui en sont dérivées ainsi que les informations de provenance associées ; ces données sémantiques sont des instances des classes d'une

ontologie appelée OntoMEDIRAD, conçue pour définir de façon explicite les informations mises en jeu dans le projet MEDIRAD ;

5. Une base de données sémantique, ou *Semantic Database*, implémentée sur un Triple Store RDF STARDOG [6] ;
6. Un gestionnaire des accès utilisateurs, basé sur le logiciel *KEYCLOAK de Redhat*, assurant le contrôle d'accès.

Chacun de ces composants s'exécute dans un container de type DOCKER [7]. La communication entre les composants est assurée par des services REST (communication via le protocole https).

2.3. Fonctionnement

L'utilisateur accède aux services du système IRDBB au moyen d'un simple navigateur internet. Après un enregistrement préalable auprès de l'administrateur du système, l'utilisateur peut importer ou consulter des données.

Le processus d'import de données se fait patient par patient et inclut les étapes suivantes : 1) Drag & Drop par l'utilisateur d'un répertoire contenant les données DICOM ou non-DICOM à importer ; 2) Saisie par l'utilisateur du pseudo du patient concerné, et de l'identifiant de l'étude MEDIRAD concernée par ces données ; 3) Dé-identification automatique des données DICOM selon un profil conforme aux recommandations du standard DICOM ; 4) Insertion dans le PACS ou le FHIR repository des données à importer ; 5) Traduction en RDF et insertion dans la base de données sémantique des métadonnées essentielles associées à ces documents (type, identifiant, patient concerné, données de provenance, etc.).

La consultation des données est réalisée à partir d'un navigateur internet : le service de consultation propose une liste de requêtes SPARQL pré-enregistrées. Les réponses sont fournies sous la forme d'un tableau qui peut ensuite être exporté vers un tableur EXCEL. Le cas échéant, ces réponses incluent des liens de téléchargement vers les images ou les fichiers non-DICOM.

La sécurité est assurée grâce à une identification par login/mot de passe complétée par la fourniture d'un code temporaire (FreeOTP authenticator) pouvant être obtenu avec un téléphone mobile.

3. RESULTATS

Le système a été conçu pour couvrir les besoins de différentes tâches du projet MEDIRAD. A ce jour, il supporte actuellement l'imagerie CT, TEP et TEMP.

En ce qui concerne les données non DICOM, deux workflows de calcul dosimétriques sont considérés.

Le premier concerne l'estimation des doses absorbées dans différents organes ou tissus lors de scanners du thorax (en pédiatrie et chez des adultes). Le second concerne l'estimation des doses absorbées par les tissus lors du traitement par radiothérapie interne du cancer de la thyroïde par ¹³¹I. Cette estimation se fonde sur des explorations en TEMP de la biodistribution du radio-pharmaceutique et sur une estimation pharmaco-cinétique, suivie d'une estimation des doses absorbées par simulation de Monte Carlo.

Pour chacun de ces workflows, une structure de données ad-hoc a été définie sous la forme d'un schéma XML. Ce dernier permet de décrire sous la forme d'un fichier XML les jeux de données à insérer dans le repository FHIR, ainsi que leurs informations de provenance. Par exemple une carte 3D de doses absorbées se verra décrite par des informations comme : son type, le nom du fichier, un indicateur de format, une référence à la méthode de calcul utilisée, le nom du logiciel utilisé. Tous les éléments de vocabulaire nécessaires ont été spécifiés dans l'ontologie OntoMEDIRAD (disponible à l'adresse <https://github.com/OsiriX-Foundation/MediradOnto/tree/master/ontology>).

4. CONCLUSION

Le système doit être mis en exploitation au début de l'année 2020. Il permettra l'import des images DICOM au fur et à mesure de leur acquisition, ainsi que leur récupération par les équipes de chercheurs travaillant sur l'estimation des doses absorbées. Les développements concernant l'import de données non-DICOM doivent être finalisés au cours du premier trimestre 2020, ce qui permettra l'import des données dosimétriques dans le système IRDBB, à la fois sous la forme de cartes voxéliques, et sous la forme de doses moyennes dans des régions d'intérêt associées à des zones anatomiques.

5. REFERENCES

- [1] Exposition de la population française aux rayonnements ionisants. Rapport IRSN/2015-00001
- [2] ESR Position paper on imaging biobank. *Insights Imaging* (2015) 6(4):403-10, 2015. DOI 10.1007/s13244-015-0409-x
- [3] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).
- [4] <https://www.dcm4che.org/>
- [5] <https://www.hl7.org/fhir/>
- [6] <https://www.stardog.com/>
- [7] <https://www.docker.com/>
- [8] <https://github.com/OsiriX-Foundation/MediradOnto/tree/master/ontology>