

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### ORTHOGEN

De Nizza, Damien; Ortiz Vega, James Jerson; Meurisse, Hubert; Schobbens, Pierre-Yves

*Published in:*  
INFORSID 2013

*Publication date:*  
2013

*Document Version*  
Première version, également connu sous le nom de pré-print

#### [Link to publication](#)

*Citation for pulished version (HARVARD):*

De Nizza, D, Ortiz Vega, JJ, Meurisse, H & Schobbens, P-Y 2013, ORTHOGEN: Système d'Information Intégré pour la Traçabilité et la Gestion Multi-Paramètres des Infections Orthopédiques. dans *INFORSID 2013: Informatique des Organisation et Systèmes d'Information et de Décision*. Paris, pp. 312-328.

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

---

# ORTHOGEN : Système d'information Intégré pour la Traçabilité et la Gestion Multi-Paramètres des Infections Orthopédiques

**Damien De Nizza<sup>1</sup>, James Ortiz<sup>2</sup>, Hubert Meurisse<sup>1</sup>,  
Pierre-Yves Schobbens<sup>2</sup>, Jean-Paul Trigaux<sup>1</sup>**

<sup>1</sup> Radiology Unit RDGN/CUMG  
Catholic University of Louvain  
{damien.denizza, hubert.meurisse, jean-paul.trigaux }@uclouvain.be

<sup>2</sup> Computer Science Faculty  
University of Namur  
{jor, pys}@info.fundp.ac.be

---

**RÉSUMÉ.** Diagnostiquer une infection orthopédique avec ou sans présence de prothèse (essentiellement hanche et genou) reste, de nos jours, très délicat. Principalement, la difficulté du diagnostic différentiel entre descellement mécanique aseptique et descellement septique d'une prothèse résulte du caractère torpide de l'infection. Un diagnostic inadéquat, tardif, voire un traitement inadapté dû à un mauvais diagnostic initial, peuvent être la cause de douleurs invalidantes, de séquelles et de complications pour le patient, sans compter les coûts directs et indirects importants pour le patient et la société. Dans cet article, nous présentons une plateforme complète de diagnostic des infections de prothèses ostéo-articulaires intégrant des modules novateurs tant au niveau imagerie qu'au niveau biologie moléculaire, un système d'organisation intelligente des données et des interfaces utilisateur multidisciplinaires et adaptatives, tout en assurant une traçabilité complète du diagnostic.

**ABSTRACT.** Diagnosing orthopedic infection with or without the presence of prosthesis (mainly hip and knee) remains today, very delicate. Primarily, the difficulty of differential diagnosis between aseptic loosening and mechanical septic loosening of a prosthesis follows from the inherently indolent infection. Diagnosis inadequate, late or inadequate treatment of a bad initial diagnosis, may be the cause of disabling pain, and sequelae of complications for the patient, not to mention the significant direct and indirect costs to the patient and society. In this paper, we present a platform full diagnosis of infections of osteoarticular prostheses modules integrating it both innovative imaging at molecular biology level, a system of intelligent organization of data and user interfaces and adaptive multidisciplinary while ensuring full traceability of diagnosis.

**MOTS-CLÉS :** Infection orthopédique, Système d'information, Biologie moléculaire, Imagerie, Ontologies .

**KEYWORDS:** Infection orthopedic, Information System, Molecular Biology, Imaging, Ontologies.

---

## 1. Introduction

Le diagnostic des infections orthopédiques [Trampuz et Widmer2006] est un processus complexe, long, fastidieux et coûteux. Sa complexité vient du fait que beaucoup d'infections peuvent prendre plusieurs mois pour se développer. Dans le cas d'infections aiguës, une réaction rapide est nécessaire pour endiguer le mal. Cependant, une réponse antibiotique précoce non adaptée peut engendrer des bactéries multi-résistantes. En imagerie, étant donné la similitude des caractéristiques observées dans le cas de descellements sceptiques et aseptiques, un recoupement avec l'histoire du patient et des ponctions locales sont nécessaires pour une discrimination correcte. Le diagnostic d'infections orthopédiques implique une approche pluridisciplinaire [Osmon *et al.*2013]. Toute fois, à l'heure actuelle, il existe peu (voire pas) d'outils informatiques intégrés entre ces différentes disciplines, ce qui complexifie et freine le travail d'élaboration de guidelines diagnostiques visant à optimiser la détection ainsi que la prise en charge thérapeutique du patient.

La détection d'une infection constitue une première étape. En effet, dans le souci de permettre un traitement efficace, il convient de pouvoir identifier le germe responsable de la maladie. Les méthodes conventionnelles [Zimmerli et Ochsner2003] de culture présentent des inconvénients qui peuvent entraver la démarche d'identification. Premièrement, certaines bactéries sont réputées pour avoir une croissance lente qui peut aller de quelques jours à plusieurs semaines ; un délai qui peut constituer un risque pour le patient infecté. Deuxièmement, la manipulation de bactéries vivantes induit un risque de contamination non négligeable. Enfin, la spécificité de la technique de culture n'est pas toujours suffisante car beaucoup de résultats s'avèrent être des faux négatifs. Enfin, Le coût de la mise au point du patient est en lien direct avec la durée de prise en charge médicale.

Dans cet article, nous présentons un système d'information intégré destiné à l'aide au diagnostic des infections orthopédiques, où : (1) Ce système sera le point de jonction entre les différentes disciplines. Il sera capable de s'interconnecter avec l'environnement hospitalier, d'en extraire des données, de les formater dans un modèle informatique à partir duquel seront réalisées des inférences diagnostiques. (2) Les différents acteurs gravitant autour du patient pourront faire des recherches spécifiques sur ces informations en fonction de leur domaine d'expertise et bénéficieront de présentations spécialisées.

Dans cet article, nous commençons par nous positionner par rapport aux travaux relatifs à la modélisation des problèmes décisionnels. Notre problématique étant médicale, nous étudions les contraintes relatives à un tel environnement et mettons en évidence les critères que doit respecter le modèle pour être valide. Ensuite, nous présentons l'approche choisie pour la construction du modèle et la structuration des données. S'en suit un résumé de l'étude menée concernant les outils et méthodes utilisés dans la modélisation du système d'information ORTHOGEN<sup>1</sup>. Enfin, nous concluons et présentons les travaux à venir.

---

1. Subvention accordée par la Région wallonne à une unité de recherche universitaire ou de niveau universitaire à travers de la DGTRE : WALEO 3

## **2. Analyse du Problème**

### **2.1. Problème de Décision**

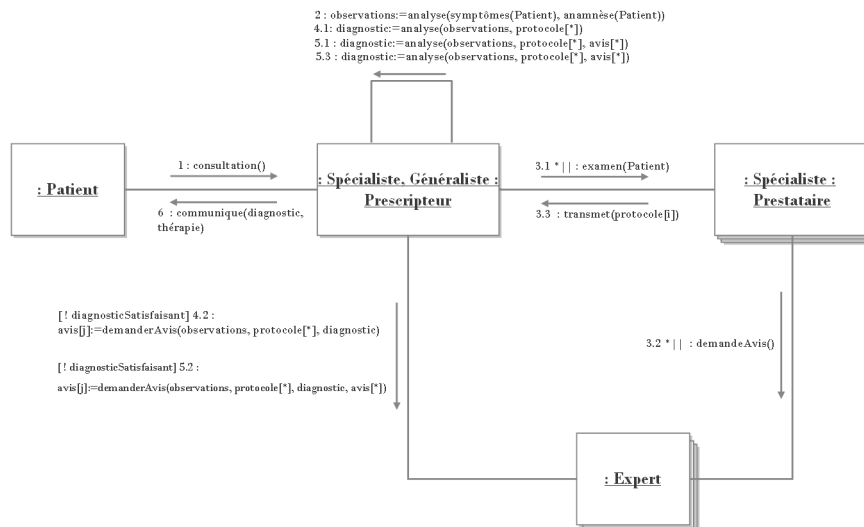
La problématique de la détermination d'un diagnostic (ou d'une suite thérapeutique adaptée à donner en réponse à une pathologie) revient à un problème décisionnel. Plus précisément, à un problème de "Pattern Classification" [Duda *et al.*2007] : un pattern de caractéristiques (symptômes cliniques, observations issues de l'imagerie, informations laboratoire, anamnèse du patient) est associé à une classe de pathologie. Typiquement, pour traiter ce type de problème de classification, on recourt à une formalisation mathématique où chaque pathologie est définie en tant que classe/modèle à partir de caractéristiques qui en détaillent les propriétés fonctionnelle et/ou descriptives. C'est sur base de ces jeux de données et des à priori sur des modèles connus que l'on peut réaliser la classification. Le processus de décision qui intervient au sein de la démarche diagnostique peut être structuré par 3 étapes [Degoulet et Fieschi1998] : l'identification du problème décisionnel, la structuration du problème décisionnel et le choix de la solution.

L'étape d'identification du problème décisionnel détermine le domaine de connaissance concerné. Lorsqu'un prescripteur se trouve au sein d'une démarche diagnostique, la quantité d'information est généralement abondante. Il convient de faire abstraction d'une partie de celle-ci de manière à prendre en compte uniquement les données qui sont pertinentes pour le domaine et le problème à décider. Cette étape peut être fortement influencée par l'expérience du décideur. Ce qui démontre l'intérêt d'une collaboration (avec les prestataires) ou d'une expertise extérieure. Il cherche à s'appuyer sur les caractéristiques les plus discriminantes (qui entraînent la meilleure classification/décision). L'étape de structuration du problème renvoie à la définition d'hypothèses diagnostiques. C'est à partir de celles-ci que pourra se faire le raisonnement. En effet, de nombreuses interprétations peuvent être faites à partir des mêmes informations. Ces hypothèses focalisent les réflexions et les inférences sur le domaine concerné par le problème. L'implication des experts du domaine est importante à cette étape et repose essentiellement sur une traduction (sous forme de règles structurées) de l'expertise qu'ils ont accumulé au fil des diagnostics. Enfin, l'étape du choix de la solution revient à substituer les règles structurées établies ci-dessus avec les informations issues des données filtrées à l'étape d'identification du problème décisionnel. C'est ce qui constitue le processus de raisonnement à proprement parler. En plus des règles structurées, ce processus fait généralement appel à une connaissance apprise. Les décisions antérieures peuvent influencer les décisions courantes. L'apprentissage permet au décideur d'adapter les règles au fur et à mesure des décisions qui sont prises.

### **2.2. Environnement Complexe**

La pose d'un diagnostic, en particulier dans le cadre des infections, nécessite une étude pluridisciplinaire qui peut prendre plusieurs mois, voire plusieurs années [Osmon *et al.*2013]. Ce processus tend généralement à générer une quantité importante d'information. La nature de ces informations est généralement hautement variée et hétérogène. Il s'agit donc de favoriser les canaux de communication entre les diffé-

rents intervenants et de gérer de manière optimale et adaptée l'information. "Adaptée" signifie qu'il convient de pouvoir segmenter le volume abondant d'information et d'en faire des présentations spécialisées en fonction du domaine d'expertise des acteurs. Cette démarche de prospection de l'information (segmentation dans le volume conséquent de données) n'est pas une tâche aisée et il n'existe pas de méthode universelle ; chaque groupe d'expert fonctionne selon ses propres guidelines. "Optimale" renvoie à une formalisation des échanges et l'assurance d'une traçabilité complète des éléments ayant servi aux inférences qui ont amené au diagnostic. Pour permettre la circulation de l'information, le milieu hospitalier s'appuie sur des standards de communication et de formatage des données. Ce type de gestion a également pour conséquence de limiter le morcellement et la redondance de l'information, favorisant la mise en relation plus rapide d'indicateurs pertinents face à une pathologie.



**Figure 1.** Illustration des rôles dans la démarche diagnostique

Le schéma présenté à la Figure 1 montre le flux de collaboration amenant à la pose d'un diagnostic. (1) Le patient initie ce flux en se présentant au prescripteur qui réalise ses premières observations sur base des symptômes et de l'anamnèse du patient. (3.1) Il planifie ensuite une série d'examens complémentaires auprès de prestataires<sup>2</sup>. Ceux-ci réalisent leurs examens et protocolent les résultats. (3.2) Ils peuvent faire appel à un expert pour appuyer leurs conclusions et (3.3) les transmettre au prescripteur. (4.1) Lorsque celui-ci les récupère, il poursuit son travail d'analyse. (4.2) S'il juge qu'il est nécessaire d'avoir des avis supplémentaires, il a alors également la possibilité de faire expertiser ses évaluations. Ces experts rendent des avis qui vont permettre au prescripteur de réévaluer son diagnostic (5.1). Cette possibilité est représentée par le flux 5.2 qui indique bien que le prescripteur peut transmettre à un expert un ou plusieurs avis émanant d'autres experts. (5.3) Sur base de ses observations, des protocoles re-

2. Le symbole "\*" sur le schéma précise que plusieurs examens sont demandés. Le symbole "||" signifie que ces examens peuvent s'effectuer en parallèle.

mis par les prestataires et des avis transmis par les experts, le prescripteur valide un diagnostic, détermine une thérapie et (6) les communique au patient.

### 2.3. Contraintes d'ORTHOGEN

Lorsque les prestataires ou le prescripteur rendent leurs conclusions, ils ne sont pas entièrement manichéens. Il y a toujours une part d'incertitude qui persiste et se propage au fil des étapes. On cherche néanmoins à la limiter le plus tôt possible. Le diagnostic est considéré comme établi lorsque la conviction est suffisamment appuyée par des éléments tangibles qui vont dans le même sens. Dans le projet ORTHOGEN, il convient donc d'intégrer cette notion d'incertitude [Duda *et al.*2007]. De plus, bien que nous soyons dans un premier temps centrés sur les prothèses de hanches, le projet vise les infections ostéo-articulaires dans leur ensemble. Pour permettre une évolution dans ce sens, on souhaite que le système d'information à mettre en place puisse répondre à des critères d'adaptabilité à d'autres domaines. Ceci en visant à limiter le recodage de l'application ou la génération d'une nouvelle ontologie adaptée. L'objectif est d'avoir un modèle informatique assez général pour favoriser sa réutilisabilité.

## 3. Approche choisie

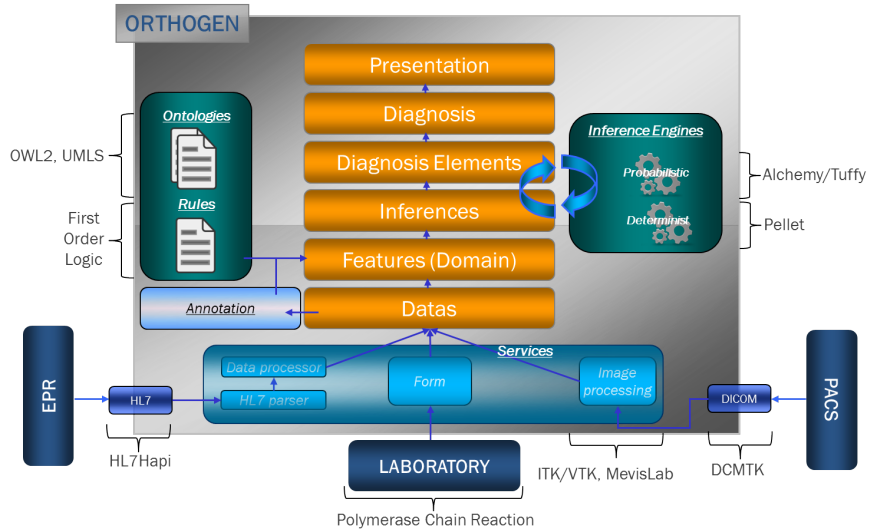
L'approche qui est proposée pour remplir les objectifs fixés se présente sous la forme d'un modèle par couches d'abstraction. Cette approche permet de répartir la complexité du système sur plusieurs niveaux. Chacun des niveaux a pour objectif d'assurer la gestion d'une partie du système et de rendre des outputs exploitables pour la couche supérieure. La Figure 2 présente la transformation des données (de bas en haut) jusqu'au diagnostic. Le niveau le plus bas matérialise l'environnement hospitalier avec les différentes sources d'informations. On y retrouve principalement l'imagerie, le dossier médical et les informations résultant d'examen<sup>3</sup> laboratoires.

Le premier niveau du système ORTHOGEN en tant que tel concerne les connecteurs standards. Ceux-ci permettent d'interagir avec les entités présentes dans le milieu hospitalier et assurent les moyens de communication nécessaires à la circulation des informations (de l'hôpital vers ORTHOGEN et inversement). La couche supérieure va s'appuyer sur les fonctionnalités offerte par les connecteurs pour extraire les données relatives au patient et qui pourront être intéressantes pour le diagnostic. Ces données vont faire l'objet d'une formalisation et être transformées en variables du système. Cette tâche est assurée par la couche d'annotation sémantique. C'est un préalable indispensable pour la tâche d'inférence : les données doivent être structurées logiquement pour permettre à une machine d'interpréter l'information et faire des déductions diagnostiques. Ces inférences vont aboutir sur des éléments de diagnostic à partir desquels sera constitué le diagnostic final. Enfin, une couche de transformation<sup>3</sup> des données (contenu et forme) permettra d'adapter la présentation des informations en fonction du contexte<sup>4</sup> de visualisation.

---

3. Filtrage, cryptage, mise en page, compression, etc.

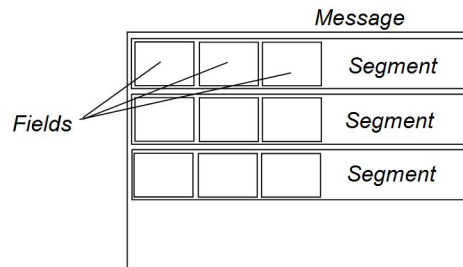
4. Spécialité de l'utilisateur, consultation sur un serveur distant, etc.



**Figure 2.** Schéma directeur du modèle ORTHOGEN

### 3.1. Connecteurs

Pour permettre au système ORTHOGEN de s’interconnecter avec son environnement, deux types de connecteurs principaux ont été envisagés : l’un pour interagir avec le dossier médical informatisé et l’autre pour interagir avec le PACS (Picture Archiving and Communication System) [PACS] de l’hôpital. Concernant le dossier médical, le standard d’échange informatisé communément adopté est HL7 (Health Level 7) [HL7]. La version v2 de ce protocole est utilisé aux Cliniques Universitaires de Mont-Godinne [Mont-Godinne]. Son fonctionnement est basé sur l’échange de messages formatés. Chaque message est constitué de segments, eux-mêmes composés de champs comme l’illustre la Figure 3. Pour l’élaboration de ce connecteur,



**Figure 3.** Structure d’un message HL7 v2

nous avons donc besoin de pouvoir générer et parser cette structure. Il existe cependant une multitude de messages différents qui peuvent être transmis et reçus en HL7. Pour ORTHOGEN, nous avons besoin de pouvoir transmettre et récolter de manière fiable

des résultats/observations, identifier un patient et identifier un examen. Nous avons identifié deux messages pour remplir ces fonctions. Le premier est un message de type "ORUR01" constitué notamment des segments MSH (pour le formatage du header du message), OBR (pour l'identification de l'examen), PID (pour l'identification du patient) et OBX (pour le transport des observations). Le second message identifié pour assurer une communication suffisamment fiable de l'information est un message de type ACK permettant de valider la réception d'un message précédemment envoyé. Il est constitué notamment des segments MSH et MSA (pour accuser réception).

Il est à noter que l'information qui se trouve dans le dossier médical est généralement du texte pur. Ce qui introduit une difficulté à la réception des messages : comment parser les informations pertinentes qu'ils contiennent ? En effet, le langage naturel est par définition ambigu et nous avons besoin d'extraire des informations non équivoques du dossier médical. Plusieurs stratégies ont été investiguées. La première, la plus simple, consiste à fouiller dans les messages des mots-clés correspondant au type d'information recherché. Mais cette approche n'est pas suffisante car on ne peut garantir que l'information extraite est bien conforme au sens qui est précisé dans le texte. En effet, si nous prenons, par exemple, un filtrage sur le mot clé "descellement" pour obtenir des informations sur la présence de descellement de prothèse chez le patient, cette approche retournera indistinctement une confirmation de descellement pour des phrases telles que "Constatation de descellement", "pas de descellement" ou "aucun signe de descellement". Il est possible de rendre plus robuste ce filtrage par mots-clés en recourant aux expressions régulières. Celles-ci prennent la forme de chaînes de caractères formatées selon une syntaxe bien définie. Elles vont permettre de détecter des patterns de chaîne de caractères spécifiques dans un texte. Grâce à elles, il devient possible de distinguer des phrases plus complexes<sup>5</sup>. Néanmoins, cette approche reste limitée car il faut considérer une quantité importante de patterns possibles pour exprimer l'information que l'on souhaite extraire. Une autre approche consiste à s'accorder avec l'équipe de gestion du dossier médical sur la définition de champs spécifiques dans leur base de données. De cette manière, lorsque l'on souhaite interroger les antécédents du patient sur des thèmes bien précis, il suffit d'interroger la base de données sur ces champs spécifiques. Cette méthode lève toute ambiguïté quant à l'information extraite. C'est la méthode qui est privilégiée pour le projet<sup>6</sup>. Cette problématique met en évidence l'intérêt que peuvent offrir des systèmes tels que celui visé par le projet ORTHOGEN en matière de structuration de l'information médicale. En effet, comme nous le verrons par la suite, nous avons besoin d'organiser l'information de telle manière qu'elle puisse être informatiquement traitable. Ainsi, ce que le système renvoi n'est plus du texte pur mais un contenu structuré qui pourra

---

5. Par exemple, l'expression régulière suivante ((P)pas del(Ala)ucun).\*descellement(s) ? permet de détecter la présence de phrases telles que "pas de signes tangibles de descellement" ou "aucun signe évident de descellement". Bien que le mot-clé "descellement" soit présente, ce pattern exclut la présence de cette lésion et cette information ne sera pas ajoutée aux informations diagnostiques.

6. Par exemple, une application de radio-pharmacie (IBC-606, Veenstra) fonctionne de la sorte aux Cliniques de Mont-Godinne [Mont-Godinne]



être dé-sérialiser ultérieurement. Ceci permet de lever la problématique de traitement du langage naturel que nous venons d'évoquer <sup>7</sup>.

L'environnement PACS déployé aux Cliniques Universitaires de Mont-Godinne est assuré par la société Télémis S.A [Telemis] et repose sur l'implémentation DCMTK [DCMTK] de la norme DICOM. Pour la plateforme ORTHOGEN, nous avons besoin de deux fonctionnalités : pouvoir récupérer/envoyer des images du/vers le PACS. Pour la récupération, des essais de Query Retrieve ont été réalisés en jouant sur les primitives C-FIND et C-MOVE via les processus findscu et movescu de DCMTK. Cependant, pour la mise en place du démonstrateur ORTHOGEN, le plus simple est d'exploiter les fonctionnalités de de Query Retrieve de Télémis (dont les clients sont dotés d'un browser PACS complet et facile d'utilisation) qui permet de pousser des images du réseau et de doter la plateforme d'un processus storescp pour la réception des images <sup>8</sup>. En ce qui concerne l'envoi des images vers le PACS, un simple processus de type movescu fera l'affaire.

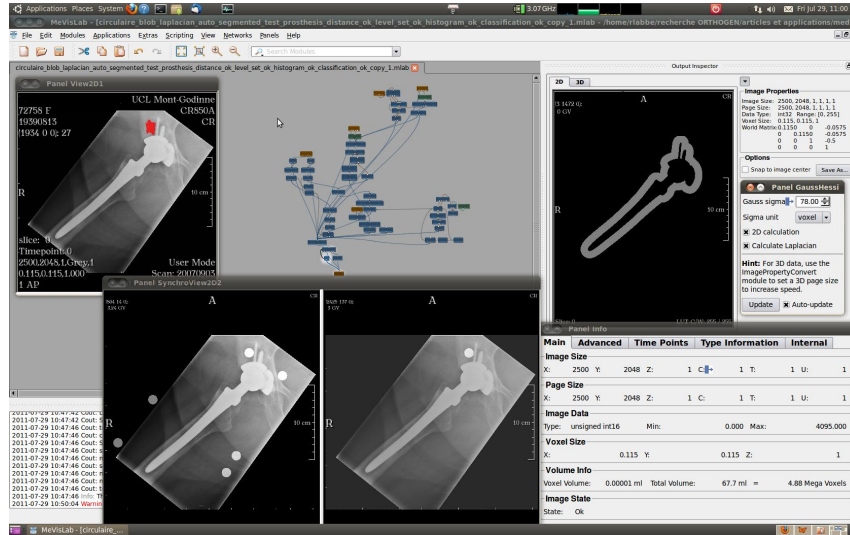
### **3.2. Extraction de données**

Lors de la prise en charge d'un patient, une quantité importante d'informations hétérogène va être générée. Surtout dans des cas tels que les infections de prothèses où la prise en charge peut durer plusieurs mois voir plusieurs années. On y retrouve des protocoles textes, des résultats de laboratoire ou encore des images provenant des différentes modalités disponibles sur le site hospitalier. Grâce aux connecteurs, nous souhaitons récupérer les données de l'environnement pour les transmettre au système afin que celui-ci puisse les intégrer dans le modèle et faire des raisonnements logiques. Cependant, nous ne souhaitons naturellement pas envoyer dans le système le volume entier d'informations relatives à un patient. En effet, seule une partie de ces informations seront pertinentes pour le cas courant. Nous allons donc segmenter l'information dans ce volume. Typiquement, dans le dossier médical, nous cherchons (à l'aide de requêtes HL7) des informations à propos des antécédents infectieux du patient, de son âge, s'il y a prise d'antibiotiques, etc. Concernant les médias, nous ne pouvons envisager d'envoyer les images directement au module de raisonnement. Nous allons avoir recours à des algorithmes de traitement d'image pour isoler des régions d'intérêt et mettre en évidence leurs propriétés physiques. Ce sont ces données extraites qui seront ensuite analysées par le système et recoupées entre-elles pour faire des déductions logiques quant au diagnostic. L'imagerie recense l'ensemble des méthodes ayant jusqu'à présent fait l'objet de recherches fondamentales mais souvent appliquées, touchant de près ou de loin à l'extraction de caractéristiques propres au domaine du placement de prothèse (voir Figure 4).

---

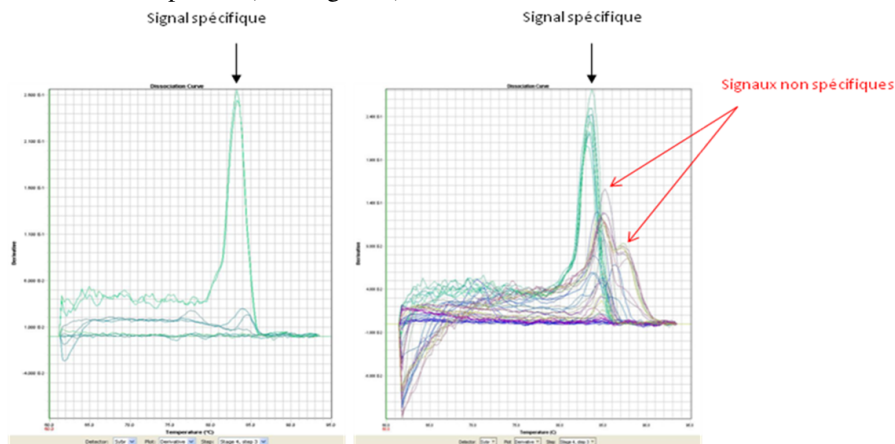
7. Du moins, pour les diagnostics ayant déjà été établis par le système. Naturellement, pour les premiers diagnostics, la consultation des informations dans le dossier médical nous renvoi à du texte pur.

8. Le Query Retrieve fait partie des opérations de base disponibles dans tous les PACS modernes. De plus, grâce à la norme DICOM et DCMTK, cette solution reste fonctionnelle pour d'autres solutions que Télémis.



**Figure 4.** Tentative fructueuse de détection d'un granulome infection de 22mm

Un volet génétique vient compléter le panel de données collectées. A l'instar de l'imagerie, nous recourons à une technique particulière d'amplification génique, appelée PCR (Polymerase Chain Reaction), pour extraire les informations relatives à la présence et l'identification de germes potentiellement responsables de douleurs invalidantes chez le patient (voir Figure 4).



**Figure 5.** Exemple d'amplification spécifique et aspecificque.

Ces données seront transmises et couplées aux autres données pour inférence. Le kit génétique vise à développer une méthode moléculaire permettant une identification rapide et non ambiguë des agents infectieux dans les échantillons orthopédiques. Le

objectif est de développer un outil d'analyse multigénomique permettant l'identification spécifique d'un microorganisme sur base de la présence simultanée de plusieurs marqueurs moléculaires. Le dossier médical informatisé, l'imagerie et la PCR constituent les sources d'informations principales à partir desquelles nous extrayons des données d'intérêt.

### 3.3. *Caractérisation*

#### 3.3.1. *Annotation Sémantique*

Grâce aux pipelines d'extraction développés en imagerie et grâce au kit génétique, de nouveaux éléments d'informations peuvent être associés aux symptômes du patient dont on veut déterminer la pathologie. Ces données sont assez diverses, par exemple : 54 ans, 80 kg, prise de novobiocine, présence de prothèse, compatibilité avec staphylocoque blanc, ponction de liquide articulaire, image RX, région d'intérêt (image, coordonnées), 1.5mm diamètre, 0.8 suv, etc. Pour permettre à une machine de raisonner sur ces données, nous devons les organiser. Pour ce faire, il faut les rattacher à une sémantique précise, les associer à des concepts clairs et non ambigus. Ces concepts sont ensuite classés selon une taxonomie bien précise et mis en relations pour définir le domaine médical visé (ici, les infections ostéo-articulaires). Cette association de termes à des concepts d'un domaine correspond à de l'annotation. Il existe plusieurs types d'annotation :

- 1) Structurelle : description des propriétés physiques ou logiques,
- 2) Linguistique : description à l'aide de tag ou d'une description textuelle,
- 3) Sémantique : description selon des concepts définissant un domaine.

Pour ce type d'annotation, nous devons définir formellement le domaine des infections de prothèse de hanche. L'approche choisie consiste à utiliser une ontologie [Gruber1993] qui va nous permettre d'établir une taxonomie de concepts, de les mettre en relation, de les instancier et de réaliser des inférences [Charlet *et al.*2003].

#### 3.3.2. *Ontologies Médicales*

Les objectifs du projet ORTHOGEN sont, d'une part la définition d'une ontologie spécifique au domaine et, d'autre part, son intégration au sein d'un système d'information qui permettra l'intégration verticale des données relatives aux infections et aux prothèses ostéo-articulaires. Le développement de l'ontologie été réalisé en collaboration avec les spécialistes médicaux afin d'assurer un niveau de qualité lors de l'inférence. Ce niveau de qualité sera évalué d'après les caractéristiques suivantes :

- 1) L'utilisation d'une ontologie médicale pour l'intégration complète de toutes les informations utiles au diagnostic.
- 2) La mise en place d'une interface intuitive d'aide au diagnostic basée sur l'ontologie et des protocoles de diagnostic permettant une vision spécifique du dossier selon la discipline de l'utilisateur.
- 3) Souplesse du modèle d'information, permettant la mise à niveau en fonction des avancées technologiques futures.

#### 4) Adaptabilité de la plateforme à d'autres domaines de diagnostic.

Cette ontologie se base sur le standard UMLS[UMLS] et est écrite à l'aide du langage OWL2 [OWL]. L'éditeur utilisé pour créer l'ontologie est Protégé [Protégé] du centre de recherche en informatique biomédicale de Stanford. En outre, nous avons commencé à concevoir l'architecture d'un moteur d'inférence probabiliste, son lien avec la Logique de Markov [Domingos *et al.*2008] et son intégration avec le système ORTHOGEN.

#### 3.3.3. *Ontologies et Ressources Terminologiques Médicales*

Comme dans la plupart des domaines de recherche, le domaine médical est un domaine très complexe, caractérisé par un vocabulaire extrêmement riche, en termes de quantité d'informations médicales importantes véhiculées entre les différents professionnels de santé, et difficile à manipuler. De ce fait, la communauté médicale a visé la nécessité de modéliser leurs connaissances et de les rendre explicites pour des besoins de partage et réutilisation. Par conséquent, plusieurs Ressources Terminologiques et Ontologiques (RTO) ont été proposées. Chacune répond à des besoins précis et divers et modélise une partie du domaine.

Les RTO ci-dessous sont les principaux :

– **ICD** : L'appellation complète de la Classification internationale des maladies est " Classification statistique Internationale des Maladies et des problèmes de santé connexes" [ICD].

– **FMA** : Le Foundational Model of Anatomy FMA est une ontologie de référence pour le domaine de l'anatomie [FMA].

– **GALEN** : Galen (General Architecture for Language, Encyclopedia and Nomenclature) est un projet européen, développé par l'organisation Open Galen depuis 1992-1999 [GALEN].

– **MENELAS** : Est un projet européen, avait pour but de proposer une approche d'accès aux dossiers médicaux rédigés en différents langues naturels [MENELAS].

– **UMLS** : UMLS (Unified Medical Language System), développé par le US National Library of Medicine, est une grande base de données conçue pour intégrer un grand nombre de termes biomédicaux provenant de diverses sources (plus de 123 sources dans l'édition de 2008) comme vocabulaires cliniques ou des classifications. UMLS est structuré en trois parties : Metathesaurus, Semantic Network, Specialist Lexicon.

- **Metathesaurus** : Le Metathesaurus représente la base du vocabulaire d'UMLS contenant l'ensemble de tous les concepts biomédicaux ainsi que de nombreuses informations sur ceux-ci. Tous les concepts repris dans le Metathesaurus sont assignés à au moins un type sémantique provenant du Semantic Network.

- **Semantic Network** : Le réseau sémantique est un graphe composé de noeuds représentant des concepts ainsi que des arcs traduisant les relations sémantiques pouvant exister entre ces concepts. Ainsi, le réseau sémantique correspond à une méta-hiérarchie à laquelle tous les concepts du Metathesaurus sont assignés.

- **SPECIALIST Lexicon** : Le SPECIALIST Lexicon est un lexique général anglais incluant de nombreux termes biomédicaux.

Dans le cadre du processus de construction de l'ontologie ORTHOGEN, nous nous sommes appuyés sur le Metathesaurus UMLS pour définir l'ontologie du domaine des infections de prothèses ostéo-articulaires. Cette approche offre l'avantage de moins solliciter les interventions d'experts (dont l'emploi du temps peut-être chargé, surtout en routine clinique) durant l'élaboration de l'ontologie.

Comparaison de différentes ontologies dans le domaine des systèmes médicaux et parce que nous avons choisi UMLS :

	ICD	FMA	GALEN	MENELAS	UMLS
Complétude	Non	Non	Oui	Oui	Oui
Non ambiguïté	Non	Non	Oui	Non	Oui
Non redondance	Oui	Oui	Oui	Oui	Oui
Synonymes	Oui	Oui	Oui	Oui	Oui
Relation explicites	Oui	Oui	Oui	Oui	Oui
Concepts/Termes	50/200 mil	75/120 mil	85/300 mil	9/15 mil	2,1/8 million

**Tableau 1.** *Ontologies et Ressources Terminologiques Médicales*

### 3.3.4. *L'incertitude et Ontologies*

Un autre point lié avec le domaine des ontologies médicales est l'incertitude. Presque toutes les ontologies médicales contiennent des concepts incertains, bien que généralement modélisés de façon inappropriée. Ces dernières années, des efforts ont été réalisés dans la représentation et le raisonnement de l'incertitude dans le Web sémantique [Declerck et Charlet2011]. Ces travaux sont principalement axés sur la façon d'étendre les logiques du Web sémantique à l'approche probabiliste, ou sur la façon de combiner ces langages avec les formalismes probabilistes [Yang et Calmet2005][Haase et Völker2008]. L'imprécision peut-être de différente nature. Par exemple,

- L'incertitude : "70% de virus sont résistants aux antibiotiques",
- Vague : "Les antibiotiques sont inefficaces dans 70% des cas",
- Subjectivité : "Je crois que le germe Mycobacterium Smegmatis est résistant au Rifampicin",
- Ambiguïté : Le terme virus a plusieurs sémantiques (c'est à la fois une entité, un logiciel malveillant, un film, etc).

De plus, tant les relations que les instances peuvent être incertaines. Par exemple, "Les virus sont généralement résistants aux antibiotiques", "Les antibiotiques sont inefficaces généralement dans 70% des cas", "Mycobacterium Smegmatis est résistant à Rifampicin avec une probabilité supérieure à 90%". Deux approches s'affrontent pour la modélisation de l'incertitude. L'une favorisant l'intégration des probabilités directement au sein de l'ontologie, alors que l'autre préconise la séparation entre les deux. La stratégie d'intégration implique donc que la connaissance du domaine et l'incertitude

(les probabilités) soient fusionnés. Concrètement, il s'agit de constituer des ontologies augmentées par des assertions de probabilités : incertitude dans les relations, dans les individuals et présence de constructeurs pour représenter les instructions conditionnelles. Ceci revient donc à étendre le langage utilisé pour la définition d'ontologies. En conséquence, les applications qui manipulent ce type d'ontologies doivent requérir certaines adaptations : il faut un moteur d'inférence adapté (par exemple, Pronto [PRONTO] qui est une extension de Pellet [PELLET]) et des opérateurs supplémentaires pour la manipulation de l'ontologie (vu que le langage OWL est étendu). A l'inverse, la stratégie de séparation vise à conserver une transparence dans la structure des ontologies classique en n'altérant pas le langage OWL. Les informations de probabilités se trouvent donc dans des structures isolées des ontologies. De ce fait, on peut toujours utiliser un moteur classique (dit déterministe) pour inférer sur les données, mais on peut également faire des inférences probabilistes à l'aide d'un moteur adapté. Les théories classiques sur lesquelles s'appuient ces moteurs sont les réseaux Bayésiens [Tighe et Tawfik2008] ou la Logique Markovienne [Domingos *et al.*2008]. Pour le projet ORTHOGEN, nous avons choisi l'approche de séparation ontologie-probabilités avec la théorie Markovienne [Domingos *et al.*2008]. Les avantages sont les suivants : compatibilité avec OWL2 (DL et règles), apports de la Logique Markovienne et bénéfice des possibilités offertes par les algorithmes d'inférence probabiliste [Duda *et al.*2007].

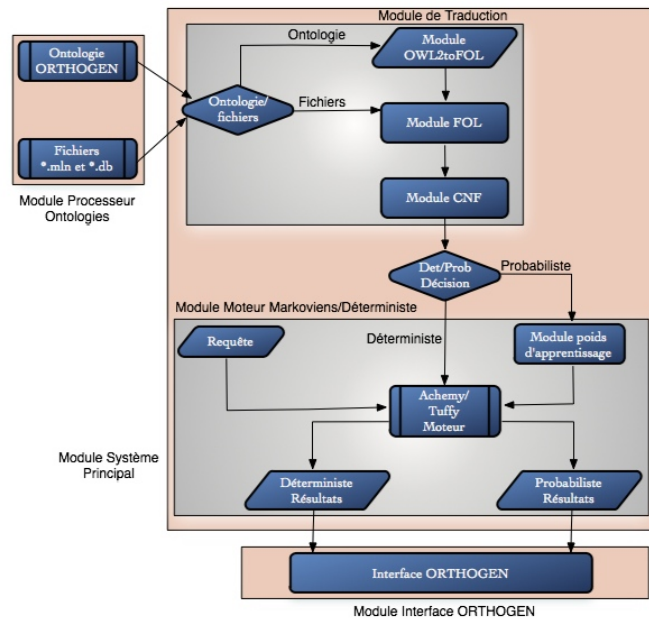
### 3.4. *Raisonnement et Incertitude*

Si nous résumons la situation jusqu'ici, grâce aux connecteurs, les outils d'extraction (imagerie, PCR) permettent de récupérer des données d'intérêt depuis l'environnement hospitalier (PACS, dossier médical informatisé). Ces données sont ensuite annotées à l'aide d'une ontologie définissant la sémantique du domaine. Ces données étant organisées logiquement selon un langage formel, il est désormais possible de pratiquer des inférences diagnostiques. Pour le projet ORTHOGEN, nous souhaitons compléter les inférences déterministes avec des inférences probabilistes, de manière à intégrer dans les raisonnements la notion d'incertitude évoquée au point précédent. L'approche probabiliste est un domaine qui essaie de trouver des mécanismes efficaces pour modéliser le raisonnement, tenant compte de l'incertitude de certaines connaissances. Dans ce domaine, les modèles graphiques probabilistes permettent de fournir un outil compact et expressif pour modéliser l'incertitude et la complexité. Il existe deux principaux types de modèles probabilistes graphiques : ceux dirigés (Réseaux Bayésiens [Tighe et Tawfik2008]) et ceux non-orientés (Réseaux Markoviens [Domingos *et al.*2008]). Le raison qui ont motivé de choisir les Réseaux Markoviens est le suivante : Les réseaux Markoviens, à la différence des réseaux Bayésiens, permettent la gestion des cycles dans les graphes. Une relation de réciprocité entre deux nœuds est dès lors possible.

#### 3.4.1. *Architecture des moteurs d'inférence*

L'architecture du moteur d'inférence probabiliste pour l'ontologie de domaine est construite à partir des modules suivants : Module Interface ORTHOGEN, Module Processeur Ontologies, Module Système Principal (Module Moteur Markovien (Al-

chemy/Tuffy [Alchemy/Tuffy]), Module de Traduction. La Figure 6 nous présente l'architecture du moteur d'inférence.



**Figure 6.** Architecture du moteur d'inférences

L'architecture est constituée d'un système principal avec lequel interagissent deux composants externes. Premièrement, le Module Processeur d'Ontologie responsable de la traduction des ontologies OWL en logique du premier ordre. Deuxièmement, le Module Interface avec la plateforme ORTHOGEN. Troisièmement, le Module Système Principal est constitué de deux composants : Module Moteur Markovien, responsable du raisonnement et du processus d'apprentissage. Le processus d'apprentissage de poids est responsable de l'application des techniques d'apprentissage probabiliste (apprentissage automatique ou itératif). Dans le processus de raisonnement, le système utilisera deux informations indispensables : un ensemble de formules pondérées et une requête. Les résultats du raisonnement sont retournés par le Markov Logic Engine à la couche d'interface. Au sein du module de traduction, dans le processus de traitement de l'ontologie, l'ontologie OWL est transformée en une représentation efficace de calcul en utilisant un processeur ontologie.

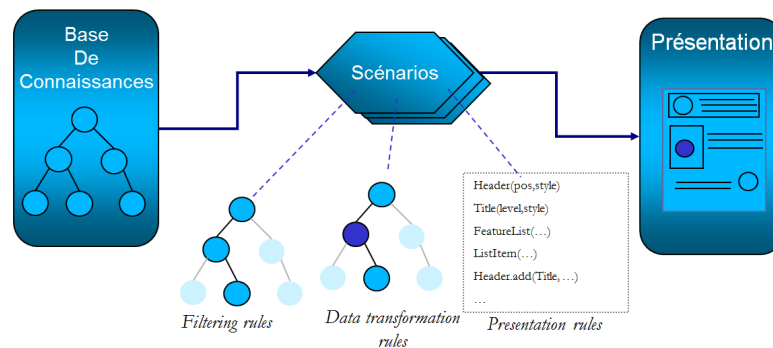
### 3.5. Présentations Spécialisées

L'ultime couche du système concerne les transformations de contenu à des fins de présentations spécialisées. Comme les données sont stockées au sein des ontologies sont logiquement organisées, il est possible de faire des requêtes sur la base de connaissance pour récupérer un certain type d'information précis. En fonction de l'uti-

lisation que l'on souhaite en faire, il est ensuite possible de transformer ces données et d'en faire des présentations adaptées. Ceci est possible car nous avons privilégié une séparation entre le fond et la forme : le contenu (fond) est indépendant des formes qu'il peut prendre. Ainsi, en fonction du contexte de visualisation il est possible d'adapter le contenu et de l'organiser spatialement ou temporellement conformément aux desideratas ou aux habitudes des utilisateurs, par exemple en fonction de sa spécialité. Etant donné que nous travaillons avec des ontologies, plusieurs mécanismes sont à notre portée pour :

- Filtrer l'information (sélection d'une information spécifique),
- Adapter le contenu (formater, transformer l'information sélectionnée),
- Créer une manifestation des informations (présentation).

Ces mécanismes sont les opérateurs classiques pour travailler sur des fichiers XML/OWL. Il s'agit de XPath, XSLT, XSL-FO, SPARQL, OWLAPI. Ces outils offrent les primitives de manipulation des données selon les besoins du contexte de visualisation. Afin de guider le processus de transformation (de contenu et de forme) du diagnostic vers une présentation spécialisée, des scénarios contextuels sont envisagés. Deux grandes classes de scénarios sont identifiées : des présentations dynamiques pour l'interface graphique utilisateur et des exportations spécialisées du diagnostic (comme par exemple l'envoi d'un résumé du diagnostic vers un serveur HTTP/FTP, la création d'un document PDF pour une étude ou encore la sérialisation des informations à destination du dossier médical informatisé) (voir Figure 7).



**Figure 7.** Schéma général de transformation de contenu et de présentation

#### 4. Conclusion

Dans cet article, nous proposons une stratégie de modélisation d'un système d'information basé sur une structuration de données en couches d'abstraction qui leur attribue une sémantique de plus en plus riche. Cette modélisation permettant d'aborder la question de l'élaboration multidisciplinaire du diagnostic des infections orthopédiques qui fait appel à des concepts hétérogènes puisés dans l'anamnèse du patient, dans l'imagerie et dans l'analyse moléculaire. Pour extraire les données de l'environnement (imagerie, dossier médical, résultats de laboratoire), des algorithmes de



traitements d'images ont été mis en place et, en ce qui concerne la biologie moléculaire, nous utilisons la PCR pour extraire des informations génétiques. Ces données sont ensuite caractérisées et annotées sémantiquement au sein d'une ontologie. Cette dernière s'appuie sur le méta-thésaurus UMLS afin de rester conforme à une nomenclature standard et pouvoir s'adapter plus facilement à d'autres domaines médicaux. Pour assurer les raisonnements logiques aidant à la détermination du diagnostic, nous utilisons un moteur d'inférence. Afin d'intégrer la notion d'incertitude dans ces inférences, le fonctionnement est basé sur la logique Markovienne, ce qui permet de réaliser à la fois des raisonnements déterministes ainsi que probabilistes. Enfin, la structuration logique des informations au sein du modèle ORTHOGEN permet des transformations de présentation et de contenu sur les données dont le but de proposer, d'une part, des vues spécialisées aux utilisateurs et, d'autre part, envisager des scénarios d'exportation du diagnostic final sous une forme structurée vers des entités tierces telles qu'un serveur HTTP/FTP (accès à distance au résumé du diagnostic) ou le dossier médical informatisé du patient.

Nous sommes actuellement en phase de développement d'un prototype qui vise à intégrer les développements des différents partenaires du projet au sein d'un système d'information unique. L'objectif est de déployer la plateforme au sein de l'hôpital Universitaire de Mont-Godinne pour une mise en situation réelle. Un groupe d'expert se chargera d'utiliser l'outil et de formuler une critique objective sur les résultats obtenus dans le but d'établir une validation de l'approche et de la solution globale. Ce groupe hétéroclite est constitué de radiologues, nucléaristes et orthopédistes présents sur le site de Mont-Godinne. Cette présence favorisera leur disponibilité et permettra une utilisation suffisante de la plateforme. En effet, le moteur d'inférence Markovien fait appel à des notions d'apprentissage et il est important que la base de connaissance puisse évoluer vers un état stable et proche de la réalité pour obtenir des résultats le plus objectifs possibles.

## 5. Bibliographie

- [Alchemy/Tuffy] « Tuffy is an open-source Markov Logic Network inference engine », <http://hazy.cs.wisc.edu/hazy/tuffy/doc/>.
- [Charlet *et al.*2003] Charlet J., Bachimont B., Troncy R., « Ontologies pour le Web sémantique. », *Web sémantique*, , 2003.
- [DCMTK] « Dicom ToolKit », <http://dicom.offis.de/dcmtdk.php.en>.
- [Declerck et Charlet2011] Declerck G., Charlet J., « Intelligence artificielle, ontologies et connaissances en médecine Les limites de la mécanisation de la pensée », *Revue d'Intelligence Artificielle*, , 2011.
- [Degoulet et Fieschi1998] Degoulet P., Fieschi M., *Informatique Médicale - 3ème 'édition.*, MASSON, 3e édition édition, 1998.
- [Domingos *et al.*2008] Domingos P., Kok S., Lowd D., Poon H., Richardson M., Singla P., « Markov Logic », *Probabilistic Inductive Logic Programming*, , 2008.
- [Duda *et al.*2007] Duda R. O., Hart P. E., Stork D. G., *Pattern Classification - 2ème 'édition.*, A Wiley-Interscience Publication, 2e édition édition, 2007.
- [FMA] « Foundational Model of Anatomy », <http://ww.fma.biostr.washington.edu/>.

- [GALEN] « General Architecture for Language », <http://hsl.virginia.edu/Galen/>.
- [Gruber1993] Gruber T. R., « A translation approach to portable ontology specifications », *Knowl. Acquis.*, , 1993.
- [Haase et Völker2008] Haase P., Völker J., « Uncertainty Reasoning for the Semantic Web I », Springer-Verlag, 2008.
- [HL7] « Health Level Seven Implementation Support Guide for HL7 Standard Version 2.3 », <http://www.hl7.org>.
- [ICD] « International Statistical Classification of Diseases and Related Health Problems », <http://www.icd10.ch/>.
- [MENELAS] « Une ontologie couvrant le domaine des maladies coronariennes », <http://www-test.biomath.jussieu.fr/Menelas/>.
- [Mont-Godinne] « Le CHU Mont-Godinne », <http://www.uclmontgodinne.be/ghome.php>.
- [Osmon *et al.*2013] Osmon D. R., Berbari E. F., Berendt A. R., Lew D., Zimmerli W., Steckelberg J. M., Rao N., Hanssen A., Wilson W. R., « Diagnosis and management of prosthetic joint infection : clinical practice guidelines by the infectious diseases society of america. », *Clin Infect Dis.*, , 2013.
- [OWL] « OWL Web Ontology Language », <http://protege.stanford.gov/>.
- [PACS] « PACS », <http://www.connectingforhealth.nhs.uk>.
- [PELLET] « OWL 2 Reasoner for Javas », <http://pellet.owldl.com/>.
- [PRONTO] « Probabilistic Semantic Web Ontologies », <http://pellet.owldl.com/pronto>.
- [Protégé] « OWL Web Ontology Language », <http://protege.stanford.edu/>.
- [Telemis] « Télémis S.A », <http://www.telemis.com/>.
- [Tighe et Tawfik2008] Tighe C. A., Tawfik A. Y., « Using causal knowledge to guide retrieval and adaptation in case-based reasoning about dynamic processes », *Int. J. Know.-Based Intell. Eng. Syst.*, , 2008.
- [Trampuz et Widmer2006] Trampuz A., Widmer A. F., « Infections associated with orthopedic implants. », *Curr Opin Infect Dis.*, , 2006.
- [UMLS] « Unified Medical Language System », <http://www.nlm.nih.gov/>.
- [Yang et Calmet2005] Yang Y., Calmet J., « OntoBayes : An Ontology-Driven Uncertainty Model », *International Conference on Computational Intelligence for Modelling, CIMCA '05*, 2005.
- [Zimmerli et Ochsner2003] Zimmerli W., Ochsner P., « Management of infection associated with prosthetic joints. », *Infection Conference.*, , 2003.