FACULTE MIXTE DE MEDECINE ET DE PHARMACIE DE ROUEN

ANNEE 2011 N°

THÈSE POUR LE DOCTORAT EN MÉDECINE

(Diplôme d'État)

PAR

Nicolas GRIFFON
Né le 03 août 1982 à Fontenay Aux Roses

PRESENTÉE ET SOUTENUE PUBLIQUEMENT
LE 13 OCTOBRE 2011

MODÉLISATION ET ÉVALUATION D'UN OUTIL DE RECHERCHE D'INFORMATIONS AU SEIN DES DOSSIERS PATIENTS INFORMATISÉS

PRÉSIDENT DU JURY : Monsieur le professeur Jean François GEHANNO

DIRECTEUR DE THÈSE : Monsieur le professeur Stéfan J. DARMONI

ANNÉE UNIVERSITAIRE 2010 - 2011

U.F.R. DE MÉDECINE-PHARMACIE DE ROUEN

~~~~~~~~~~~~~~~~~~~~~~~~~

DOYEN :	Professeur Pierre FREGER
ASSESSEURS :	Professeur Michel GUERBET Professeur Benoit VEBER Professeur Pascal JOLY Professeur Bernard PROUST
DOYENS HONORAIRES :	Professeurs J. BORDE - Ph. LAURET - H. PIGUET – C. THUILLEZ
PROFESSEURS HONORAIRES :	MM. M-P AUGUSTIN - J.ANDRIEU-GUITRANCOURT - M.BENOZIO-J.BORDE - Ph. BRASSEUR - R. COLIN - E. COMOY - J. DALION -. DESHAYES - C. FESSARD – J.P FILLASTRE - P.FRIGOT -J. GARNIER - J. HEMET - B. HILLEMAND - G. HUMBERT - J.M. JOUANY - R. LAUMONIER – Ph. LAURET - M. LE FUR – J.P. LEMERCIER - J.P LEMOINE - Mlle MAGARD - MM. B. MAITROT - M. MAISONNET - F. MATRAY - P.MITROFANOFF - Mme A. M. ORECCHIONI - P. PASQUIS - H.PIGUET - M.SAMSON – Mme SAMSON-DOLLFUS – J.C. SCHRUB - R.SOYER - B.TARDIF - .TESTART - J.M. THOMINE – C. THUILLEZ - P.TRON - C.WINCKLER - L.M.WOLF

I – **MÉDECINE**

PROFESSEURS

M. Frédéric **ANSELME**	HCN	Cardiologie
M. Bruno **BACHY**	HCN	Chirurgie pédiatrique
M. Fabrice **BAUER**	HCN	Cardiologie
Mme Soumeya **BEKRI**	HCN	Biochimie et Biologie Moléculaire
M. Jacques **BENICHOU**	HCN	Biostatistiques et informatique médicale
M. Eric **BERCOFF**	HB	Médecine interne (gériatrie)
M. Jean-Paul **BESSOU**	HCN	Chirurgie thoracique et cardio-vasculaire
Mme Françoise **BEURET-BLANQUART**	CRMPR	Médecine physique et de réadaptation
M. Guy **BONMARCHAND**	HCN	Réanimation médicale
M. Olivier **BOYER**	UFR	Immunologie
M. Jean-François **CAILLARD**	HCN	Médecine et santé au Travail
M. François **CARON**	HCN	Maladies infectieuses et tropicales
M. Philippe **CHASSAGNE**	HB	Médecine interne (Gériatrie)
M. Alain **CRIBIER (*Surnombre)***	HCN	Cardiologie
~~~~~~~~~~~~~~~~~~~~~~~~~

M. Antoine **CUVELIER**	HB	Pneumologie
M. Pierre **CZERNICHOW**	HCH	Epidémiologie, économie de la santé
M. Jean-Nicolas **DACHER**	HCN	Radiologie et Imagerie Médicale
M. Stéfan **DARMONI**	HCN	Informatique Médicale/Techniques de communication
M. Pierre **DECHELOTTE**	HCN	Nutrition
Mme Danièle **DEHESDIN**	HCN	Oto-Rhino-Laryngologie
M. Philippe **DENIS (Surnombre)**	HCN	Physiologie
M. Jean **DOUCET**	HB	Thérapeutique/Médecine – Interne – Gériatrie.
M. Bernard **DUBRAY**	CB	Radiothérapie
M. Philippe **DUCROTTE**	HCN	Hépato-Gastro-Entérologie
M. Frank **DUJARDIN**	HCN	Chirurgie Orthopédique – Traumatologique
M. Fabrice **DUPARC**	HCN	Anatomie - Chirurgie Orthopédique et Traumatologique
M. Bertrand **DUREUIL**	HCN	Anesthésiologie et réanimation chirurgicale
Mlle Hélène **ELTCHANINOFF**	HCN	Cardiologie
M. Thierry **FREBOURG**	UFR	Génétique
M. Pierre **FREGER**	HCN	Anatomie/Neurochirurgie
M. Jean François **GEHANNO**	HCN	Médecine et Santé au Travail
Mme Priscille **GERARDIN**	HCN	Pédopsychiatrie
M. Michel **GODIN**	HB	Néphrologie
M. Philippe **GRISE**	HCN	Urologie
M. Didier **HANNEQUIN**	HCN	Neurologie
M. Philippe **HECKETSWEILER (surnombre)**	HCN	Hépato - Gastro/Policlinique
M. Fabrice **JARDIN**	CB	Hématologie
M. Luc-Marie **JOLY**	HCN	Médecine d'urgence
M. Pascal **JOLY**	HCN	Dermato - vénéréologie
M. Jean-Marc **KUHN**	HB	Endocrinologie et maladies métaboliques
Mme Annie **LAQUERRIERE**	HCN	Anatomie cytologie pathologiques
M. Vincent **LAUDENBACH**	HCN	Anesthésie et réanimation chirurgicale
M. Alain **LAVOINNE**	UFR	Biochimie et biologie moléculaire
M. Joël **LECHEVALLIER**	HCN	Chirurgie infantile
M. Patrick **LE DOSSEUR**	HCN	Radiopédiatrie
M. Hervé **LEFEBVRE**	HB	Endocrinologie et maladies métaboliques
M. Xavier **LE LOET**	HB	Rhumatologie
M. Jean-François **LEMELAND (Surnombre)**	HCN	Bactériologie
M. Eric **LEREBOURS**	HCN	Nutrition
Mlle Anne-Marie **LEROI**	HCN	Physiologie
M. Hervé **LEVESQUE**	HB	Médecine interne

Mme Agnès **LIARD-ZMUDA**	HCN	Chirurgie Infantile
M. Bertrand **MACE**	HCN	Histologie, embryologie, cytogénétique
M. Eric **MALLET *(Surnombre)***	HCN	Pédiatrie
M. Christophe **MARGUET**	HCN	Pédiatrie
Mlle Isabelle **MARIE**	HB	Médecine Interne
M. Jean-Paul **MARIE**	HCN	ORL
M. Loïc **MARPEAU**	HCN	Gynécologie - obstétrique
M. Stéphane **MARRET**	HCN	Pédiatrie
M. Pierre **MICHEL**	HCN	Hépato - Gastro - Entérologie
M. Francis **MICHOT**	HCN	Chirurgie digestive
M. Bruno **MIHOUT**	HCN	Neurologie
M. Pierre-Yves **MILLIEZ**	HCN	Chirurgie plastique, reconstructrice et esthétique
M. Jean-François **MUIR**	HB	Pneumologie
M. Marc **MURAINE**	HCN	Ophtalmologie
M. Philippe **MUSETTE**	HCN	Dermatologie - Vénéréologie
M. Christophe **PEILLON**	HCN	Chirurgie générale
M. Jean-Marc **PERON**	HCN	Stomatologie et chirurgie maxillo-faciale
M. Christian **PFISTER**	HCN	Urologie
M. Jean-Christophe **PLANTIER**	HCN	Bactériologie - Virologie
M. Didier **PLISSONNIER**	HCN	Chirurgie vasculaire
M. Bernard **PROUST**	HCN	Médecine légale
M. François **PROUST**	HCN	Neurochirurgie
Mme Nathalie **RIVES**	HCN	Biologie et médecine du développement et de la reproduction
M. Jean-Christophe **RICHARD**	HCN	Réanimation Médicale, Médecine d'urgence
M. Jean-Christophe **SABOURIN**	HCN	Anatomie - Pathologie
M. Michel **SCOTTE**	HCN	Chirurgie digestive
Mme Fabienne **TAMION**	HCN	Thérapeutique
Mlle Florence **THIBAUT**	HCN	Psychiatrie d'adultes
M. Luc **THIBERVILLE**	HCN	Pneumologie
M. Jacques **THIEBOT**	HCN	Radiologie et imagerie médicale
M. Christian **THUILLEZ**	HB	Pharmacologie
M. Hervé **TILLY**	CB	Hématologie et transfusion
M. François **TRON**	UFR	Immunologie
M. Jean-Jacques **TUECH**	HCN	Chirurgie digestive
M. Jean-Pierre **VANNIER**	HCN	Pédiatrie génétique
M. Benoît **VEBER**	HCN	Anesthésiologie Réanimation chirurgicale
M. Pierre **VERA**	C.B	Biophysique et traitement de l'image

M. Eric **VERSPYCK**	HCN	Gynécologie obstétrique
M. Olivier **VITTECOQ**	HB	Rhumatologie
M. Jacques **WEBER**	HCN	Physiologie

PROFESSEURS ASSOCIÉS A MI-TEMPS :

M. Jean-Loup **HERMIL**	UFR	Médecine générale
M. Alain **MERCIER**	UFR	Médecine générale
M. Philippe **NGUYEN THANH**	UFR	Médecine générale

MAÎTRES DE CONFÉRENCES

Mme Noëlle **BARBIER-FREBOURG**	HCN	Bactériologie – Virologie
M. Jeremy **BELLIEN**	HCN	Pharmacologie
Mme Carole **BRASSE LAGNEL**	HCN	Biochimie
M. Gérard **BUCHONNET**	HCN	Hématologie
Mme Mireille **CASTANET**	HCN	Pédiatrie
Mme Nathalie **CHASTAN**	HCN	Physiologie
Mme Sophie **CLAEYSSENS**	HCN	Biochimie et biologie moléculaire
M. Moïse **COEFFIER**	HCN	Nutrition
M. Vincent **COMPERE**	HCN	Anesthésiologie et réanimation chirurgicale
M. Manuel **ETIENNE**	HCN	Maladies infectieuses et tropicales
M. Guillaume **GOURCEROL**	HCN	Physiologie
Mme Catherine **HAAS-HUBSCHER**	HCN	Anesthésie - Réanimation chirurgicale
M. Serge **JACQUOT**	UFR	Immunologie
M. Joël **LADNER**	HCN	Epidémiologie, économie de la santé
M. Jean-Baptiste **LATOUCHE**	UFR	Biologie Cellulaire
Mme Lucie **MARECHAL-GUYANT**	HCN	Neurologie
M. Jean-François **MENARD**	HCN	Biophysique
Mme Muriel **QUILLARD**	HCN	Biochimie et Biologie moléculaire
M. Vincent **RICHARD**	UFR	Pharmacologie
M. Francis **ROUSSEL**	HCN	Histologie, embryologie, cytogénétique
Mme Pascale **SAUGIER-VEBER**	HCN	Génétique
Mme Anne-Claire **TOBENAS-DUJARDIN**	HCN	Anatomie
M. Eric **VERIN**	HCN	Physiologie

MAITRES DE CONFÉRENCES ASSOCIE A MI-TEMPS :

M. Pierre **FAINSILBER**	UFR	Médecine générale
M Emmanuel **LEFEBVRE**	UFR	Médecine générale
Mme Elisabeth **MAUVIARD**	UFR	Médecine générale

PROFESSEURS AGRÉGÉS OU CERTIFIÉS

Mme Dominique **LANIEZ**	UFR	Anglais
Mme Michèle **GUIGOT**	UFR	Sciences humaines - Techniques d'expression

II - PHARMACIE

PROFESSEURS

M. Thierry **BESSON**	Chimie Thérapeutique
M. Jean-Jacques **BONNET**	Pharmacologie
M. Roland **CAPRON** (PU-PH)	Biophysique
M. Jean **COSTENTIN** (PU-PH)	Pharmacologie
Mme Isabelle **DUBUS**	Biochimie
M. Loïc **FAVENNEC** (PU-PH)	Parasitologie
M. Michel **GUERBET**	Toxicologie
M. Olivier **LAFONT**	Chimie organique
Mme Isabelle **LEROUX**	Physiologie
M. Jean-Louis **PONS** (PU-PH)	Microbiologie
Mme Elisabeth **SEGUIN**	Pharmacognosie
M. Marc **VASSE** (PU-PH)	Hématologie
M Jean-Marie **VAUGEOIS (Délégation CNRS)**	Pharmacologie
M. Philippe **VERITE**	Chimie analytique

MAÎTRES DE CONFÉRENCES

Mme Dominique **ANDRE**	Chimie analytique
Mlle Cécile **BARBOT**	Chimie Générale et Minérale
Mme Dominique **BOUCHER**	Pharmacologie
M. Frédéric **BOUNOURE**	Pharmacie Galénique
Mme Martine **PESTEL-CARON**	Microbiologie
M. Abdeslam **CHAGRAOUI**	Physiologie
M. Jean **CHASTANG**	Biomathématiques
Mme Marie Catherine **CONCE-CHEMTOB**	Législation pharmaceutique et économie de la santé
Mme Elizabeth **CHOSSON (Délégation)**	Botanique
Mlle Cécile **CORBIERE**	Biochimie
M. Eric **DITTMAR**	Biophysique
Mme Nathalie **DOURMAP**	Pharmacologie
Mlle Isabelle **DUBUC**	Pharmacologie
Mme Roseline **DUCLOS**	Pharmacie Galénique
M. Abdelhakim **ELOMRI**	Pharmacognosie
M. François **ESTOUR**	Chimie Organique
M. Gilles **GARGALA** (MCU-PH)	Parasitologie
Mme Najla **GHARBI**	Chimie analytique

Mlle Marie-Laure **GROULT**	Botanique
M. Hervé **HUE**	Biophysique et Mathématiques
Mme Hong **LU**	Biologie
Mme Sabine **MENAGER**	Chimie organique
Mme Christelle **MONTEIL**	Toxicologie
M. Paul **MULDER**	Sciences du médicament
M. Mohamed **SKIBA**	Pharmacie Galénique
Mme Malika **SKIBA**	Pharmacie Galénique
Mme Christine **THARASSE**	Chimie thérapeutique
M. Rémi **VARIN** (MCU-PH)	Pharmacie Hospitalière
M. Frédéric **ZIEGLER**	Biochimie

PROFESSEUR ASSOCIÉ

M. Jean- Pierre **GOULLE**	Toxicologie

MAÎTRE DE CONFÉRENCE ASSOCIÉ

Mme Sandrine **PANCHOU**	Pharmacie Officinale

PROFESSEUR AGRÉGÉ OU CERTIFIÉ

Mme Anne- Marie **ANZELLOTTI**	Anglais

ATTACHÉES TEMPORAIRES D'ENSEIGNEMENT ET DE RECHERCHE

Mlle Virginie **SEGUIN**	Botanique
Mlle Sophie **RABEAU**	Chimie Générale et Minérale
Mlle Laetitia **LE GOFF**	Parasitologie

CHEF DES SERVICES ADMINISTRATIFS : Mme Véronique DELAFONTAINE

HCN - Hôpital Charles Nicolle *HB - Hôpital de BOIS GUILLAUME*
CB - Centre HENRI BECQUEREL *CHS - Centre Hospitalier Spécialisé du Rouvray*
CRMPR - Centre Régional de Médecine Physique et de Réadaptation

LISTE DES RESPONSABLES DE DISCIPLINE

Mlle Cécile **BARBOT**	Chimie Générale et Minérale
M. Thierry **BESSON**	Chimie thérapeutique
M. Roland **CAPRON**	Biophysique
M Jean **CHASTANG**	Mathématiques
Mme Marie-Catherine **CONCE-CHEMTOB**	Législation, Économie de la Santé
Mlle Elisabeth **CHOSSON**	Botanique
M. Jean **COSTENTIN**	Pharmacodynamie
Mme Isabelle **DUBUS**	Biochimie
M. Loïc **FAVENNEC**	Parasitologie
M. Michel **GUERBET**	Toxicologie
M. Olivier **LAFONT**	Chimie organique
M. Jean-Louis **PONS**	Microbiologie
Mme Elisabeth **SEGUIN**	Pharmacognosie
M. Mohamed **SKIBA**	Pharmacie Galénique
M. Marc **VASSE**	Hématologie
M. Philippe **VERITE**	Chimie analytique

ENSEIGNANTS MONO-APPARTENANTS

MAITRES DE CONFÉRENCES

M. Sahil **ADRIOUCH**	Biochimie et biologie moléculaire (Unité Inserm 905)
Mme Gaëlle **BOUGEARD-DENOYELLE**	Biochimie et biologie moléculaire (Unité Inserm 614**)**
M. Antoine **OUVRARD-PASCAUD**	Physiologie (Unité Inserm 644)

PROFESSEURS DES UNIVERSITÉS

M. Mario **TOSI**	Biochimie et biologie moléculaire (Unité Inserm 614)
M. Serguei **FETISSOV**	Physiologie (Groupe ADEN)

Par délibération en date du 3 mars 1967, la faculté a arrêté que les opinions émises dans les dissertations qui lui seront présentées doivent être considérées comme propres à leurs auteurs et qu'elle n'entend leur donner aucune approbation ni improbation.

Remerciements

A mon directeur de thèse, Stefan Darmoni, qui m'a embarqué dans l'aventure CISMeF avec énergie et bonne humeur,
A Jean François Géhanno, président du Jury et à Philippe Massari, Guillaume Savoye et Jean Philippe Leroy, membres du jury, pour l'intérêt qu'ils ont porté à mon travail,
A Pierre Czernichow pour la liberté qu'il laisse aux internes pendant leur internat, leur permettant d'explorer toutes les voies,
A toute l'équipe CISMeF, pour l'accueil chaleureux qu'ils m'ont réservé à bord,
Au docteur Dauchet pour m'avoir guidé dans mes débuts à Rouen et pour avoir aiguisé ma curiosité, affuté mon esprit critique,
A mes parents, mon frère et toute ma famille, responsables, pour une grande partie, de ce que je suis aujourd'hui. Vous êtes loin, mais pas tant que ça finalement,
A Basile, mon camarade d'infortune depuis nos débuts en médecine, une étape de plus vers la fin…
A tous les autres, qu'ils m'aient soutenu, subi ou juste accompagné pendant mes études de médecines,
Et enfin, à Claire qui me permet de voir le futur sous un jour meilleur.

PLAN

LISTE DES FIGURES

1 Introduction

Depuis plus de 5000 ans, les médecins rédigent des observations médicales [1]. Si, initialement, ces écrits avaient pour vocation primaire l'enseignement et la recherche, ils ont, au fil du temps, acquis de nombreuses autres fonctions. Ils ont pris, au XVIII^{ème} siècle, la forme qu'on leur connaît aujourd'hui : le dossier médical [2], guidant la pratique et aidant la mémoire des médecins [3]. D'abord succinct, il finira par s'intégrer, avec toutes les informations nécessaires pour une prise en charge partagée des patients, au dossier patient (DP). Aujourd'hui, ces DP contiennent, en plus du dossier médical, les informations nécessaires à la facturation [4], à la traçabilité et à l'évaluation de la qualité des soins [5] et ont une valeur probante devant la justice en cas de contentieux [6].

Bien souvent, l'informatisation du DP a été menée pour l'automatisation de certaines de ses fonctions [7], [8] : résultats de laboratoire et facturation notamment pour s'étendre aujourd'hui à l'ensemble du DP : comptes-rendus de séjour, d'acte, documents d'imageries… Les informations sont ainsi disponibles plus rapidement et plus sûrement pour l'ensemble des personnels prenant en charge un patient (radiologue, chirurgien, infirmier, biologiste…) au sein d'une équipe, d'un hôpital, d'un réseau de santé… Cependant, dans une large majorité des cas, les dossiers patient informatisés (DPI) sont lacunaires et ne correspondent qu'à une dématérialisation de fragments du dossier papier. Le gain en terme de volume et de papier est peut être important, mais les informations contenues dans ces nouveaux dossiers médicaux ne sont pas nécessairement plus exploitables que ne l'étaient les informations contenues dans les épais dossiers papier les ayant précédés [9], [10], [11]. Ainsi, les DPI représentent une somme d'informations colossale qui est difficilement utilisable autrement que pour le patient pour lequel elle a été produite et au moment de sa production.

Les intérêts possibles de l'informatisation du DP, basés sur l'exploitation informatique des données qu'il contient, sont nombreux et ont été largement étudiés [12]. La recherche d'informations (RI) fait partie de ceux-ci. La mise en place d'outils de RI au sein des DPI n'est pas récente. Dès 1985, Safran [13] mettait en place un moteur de recherche au sein des dossiers des patients du Boston's Beth Israel Hospital facile à utiliser. En 1995, l'étude de l'historique des événements de cet outil de recherche mettait en évidence plusieurs cas d'utilisations [14]:

“to display information about an individual patient (results reporting); to find data on a patient with similarities to one being seen (case finding): to describe a group of patients with at least one attribute in common (cohort description)”

Ces cas d’utilisation correspondent à un usage actif par l’utilisateur de l’outil de recherche. Il apparaît toutefois que les utilisateurs ne vont pas toujours chercher les réponses aux questions qu’ils se posent et qu’ils ne sont pas toujours conscients de leurs besoins d’informations [15]. Les systèmes d’aide à la décision médicale (SADM) peuvent pallier ces difficultés, mais la plupart de leurs modalités d’intervention, recensées par Renaud-Salis et al [12], nécessitent de trouver les informations nécessaires dans les dossiers médicaux…

L’équipe CISMeF, forte de son savoir-faire en RI sur internet [16] (url : http://cismef.chu-rouen.fr) et sur les terminologies [17] (url : http://pts.chu-rouen.fr), a travaillé, et va travailler, dans le cadre du projet RAVEL (Recherche et Visualisation des informations dans le dossier patient électronique) financé par le programme TecSan de l’Agence Nationale de la Recherche (ANR), au développement d’un outil de RI : RIDoPI (Recherche d’Information dans le Dossier Patient Informatisé) qui combinerait les points forts des solutions déjà existantes. Cet outil devrait assister les médecins dans leurs multiples fonctions : clinique, recherche et enseignement.

Les objectifs de ce travail sont : (1) d’évaluer le travail déjà fait par l’équipe CISMeF, notamment la capacité du modèle créé à prendre en compte les données structurées des DPI, (2) de pallier les difficultés mises en évidences et (3) de travailler à l’interface de l’outil RIDoPI.

2 Un Outil de RI dans les DPI

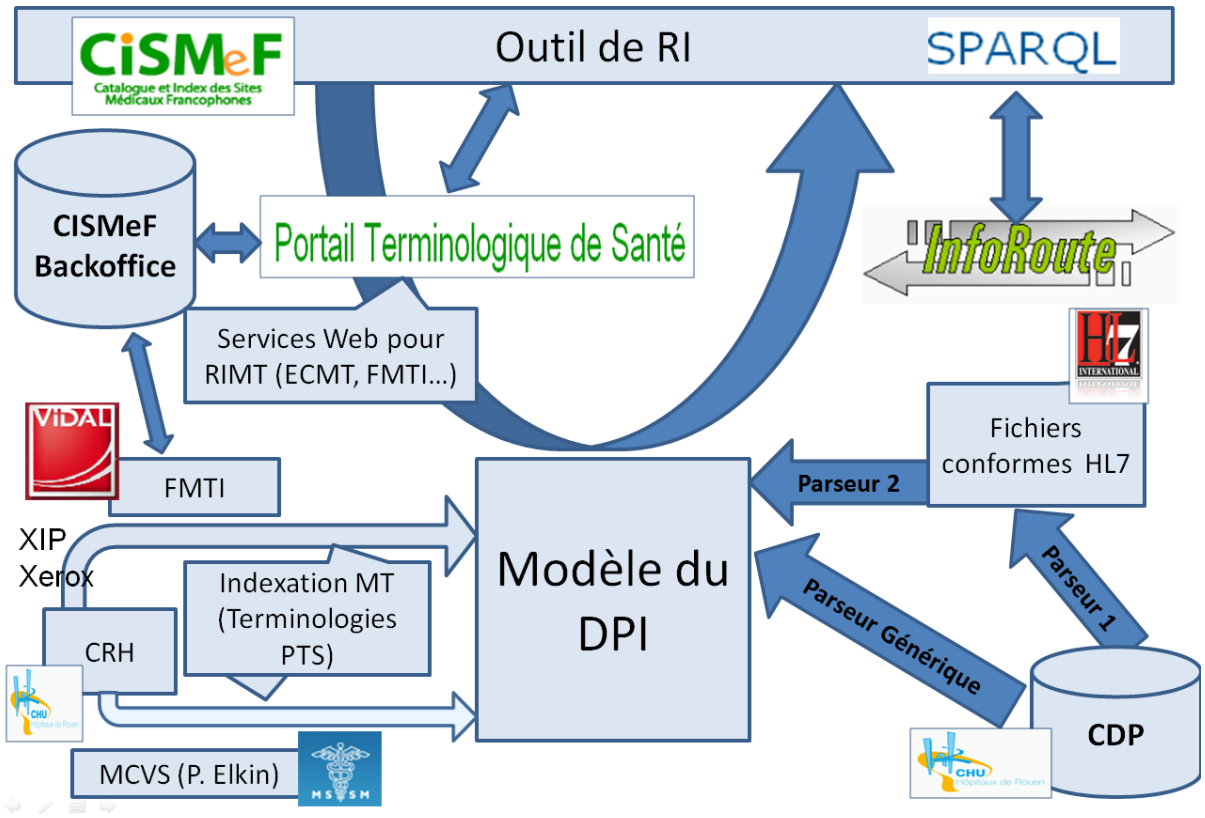
2.1 Définition

Les outils de recherche d'informations permettent de retrouver, au sein d'une base documentaire, ici les nombreux comptes-rendus qui composent les DPI, ceux qui répondent à une question précise, formulée par l'utilisateur à l'aide d'une interface d'édition de requête [18]. L'expérience de l'équipe CISMeF avec les moteurs de recherche booléens conduit naturellement à l'utilisation d'un moteur de recherche de ce type. Leur fonctionnement est basé sur les opérateurs booléens (ET, OU et SAUF) et sur l'indexation de chacun des documents de la base par des mots clés les représentant [18]. L'utilisateur doit formuler sa requête à l'aide des opérateurs : il veut les documents qui traitent de "techniques chirurgicales" ET d'"anévrysme de l'aorte abdominale". Le moteur de recherche lui ramène l'ensemble des documents indexés par ces deux termes.

La construction de l'outil RIDoPI nécessitera de nombreuses étapes (voir Figure 1). Dans un premier temps, les informations contenues dans les dossiers patients ("CDP") doivent être répliquées dans la base RIDoPI ("modèle du DPI") à l'aide d'un outil de réplication générique ("parseur"). Ensuite, les documents doivent être indexés : on crée des représentations des documents, exploitables informatiquement. Plusieurs outils permettent théoriquement de réaliser cette indexation ("XIP", "FMTI" ou "MCVS"). Enfin, il est nécessaire de mettre au point un outil d'édition de requête ("outil de RI") susceptible « d'interpréter » les demandes des utilisateurs et de les traiter de manière à sélectionner et renvoyer à l'utilisateur les documents correspondants.

Les outils de RI et d'indexation reposent sur des vocabulaires contrôlés déjà gérés par le système d'information de l'équipe CISMeF ("CISMeF Backoffice" et "Portail terminologique de santé" [17]). D'autres outils ("inforoutes" [19] et "CISMeF" [16]), déjà développés, permettront d'accéder facilement aux recommandations, conférences de consensus, publications scientifiques… correspondant à la situation du patient dont le DPI est consulté. Ce travail de thèse a porté sur le "modèle du DPI" et sur l'interface de "l'outil de RI".

Figure 1 : Schéma général du futur outil RIDoPI



2.2 Intérêts

2.2.1 RI mono-patient

2.2.1.1 Prise en charge d'un patient

Dans la pratique quotidienne les médecins ont souvent besoin, dans le cadre de leur démarche diagnostique ou de la prescription, d'informations normalement présentes dans le dossier patient. Christensen et Grimsmo [10] ont mis en évidence que plus d'un tiers des médecins ne cherchait pas d'information dans leurs dossiers patients du fait de la chronophagie de cette activité. Des outils permettant de faciliter cette recherche ne sont toujours pas disponibles en 2010 chez des médecins pourtant désireux d'en disposer [20].

Les questions posées peuvent être de différentes natures : questions sur l'histoire de la maladie du patient (informations ophtalmologiques dans le dossier d'un patient suivi depuis longtemps pour un diabète), questions sur les caractéristiques précises d'un médicament (adaptation de la posologie d'un médicament selon la fonction rénale), questions sur la prise en charge d'une pathologie, d'un symptôme (Quelles sont les recommandations pour la prise en charge d'un accès palustre ? d'une adénopathie cervicale ?)… et peuvent être très complexes (Que faire pour une femme de 88 ans, soufrant de dysphagie à cause d'un antécédent de cancer du larynx, qui est en détresse respiratoire du fait de sa sonde naso-gastrique alors que la famille nourrit des espoirs beaucoup trop optimistes ?), notamment pour les patients atteints de maladie chronique. Des aides existent pour y répondre : conférences de consensus des sociétés savantes, recommandations francophones (accessibles en particulier via CISMeF bonnes pratiques [21]), Vidal Recos [22], PubMed… mais plus de 50% des questions [23] n'ont de réponses qu'au sein du dossier du patient et sont difficilement accessibles. Des outils de RI mono-patient permettraient donc de faciliter grandement deux des trois cas d'utilisation du dossier médical identifiés par Nygren et Henriksson [24] : *« to search for specific details, and to prompt or explore hypotheses ».* Cela limiterait les situations ou le médecin identifie un besoin d'informations mais ne le satisfait pas. Couplé à des mécanismes d'alerte, les outils de RI peuvent même permettre de pallier les besoins non reconnus (le médecin n'est pas conscient d'avoir un besoin

d'informations) ou considérés comme insolubles (le médecin pense que l'information n'existe pas).

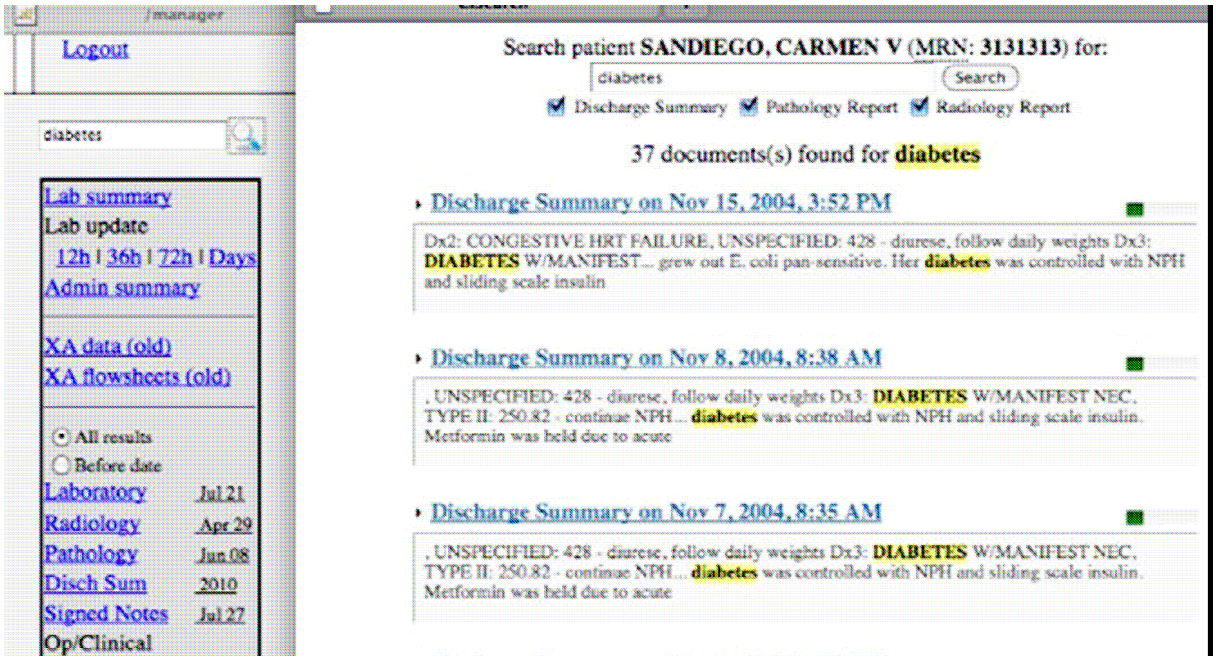
Quelques expérimentations ont déjà été menées :

- CISearch [25] est l'outil développé et implémenté au sein du DPI de l'hôpital universitaire de Columbia. Il permet à l'utilisateur d'effectuer des recherches dans l'ensemble des notes en texte libre (comptes-rendus de radiologie, d'anatomo-pathologie, résumés de sortie, notes de soins…) du dossier médical qu'il consulte. Il exploite quelques fonctionnalités de Lucene [26], un ensemble d'outils génériques pour l'indexation automatique et la recherche d'informations : utilisation de caractères spéciaux (joker, troncature, exact match…), aperçu des résultats, avec surbrillance des termes cherchés, et affichage des résultats du plus récent au plus ancien. L'interface utilisateur est présentée en Figure 2.
- MIRS (medical information retrieval system) [27] est également fondé sur Lucene. Il permet d'effectuer des recherches plein-texte dans l'ensemble des documents du DPI : comptes-rendus opératoires, lettres de sortie… Ces documents bénéficient toutefois de traitements particuliers permettant de les assigner aux domaines de la médecine (correspondant grossièrement aux spécialités médicales) desquels ils relèvent. Cette métadonnée est exploitable par l'utilisateur, au moment de sa requête, pour favoriser le rappel des documents relevant d'une ou plusieurs spécialités.

Toutefois, la dissémination de ces travaux en dehors des établissements des équipes qui les développent reste exceptionnelle. Ainsi, la plupart des applications de gestion des DPI, y compris et en particulier les principales solutions industrielles (McKesson, Cerner, Agfa…), ne disposent pas d'outils de recherche, ce qui, paradoxalement, complique parfois la tâche par rapport au dossier papier [28]. On est loin de l'objectif théorique du DPI : mettre les informations concernant un patient à disposition des professionnels de santé où et quand ils en ont besoin [29].

Au-delà d'une équipe soignante ou de confrères au sein d'un cabinet, le Dossier Médical Partagé (DMP) devrait permettre un partage des informations concernant un patient entre tous les professionnels de santé le prenant en charge. Les informations devraient y être renseignées puisque le code de la santé publique dispose que : *"chaque professionnel de santé […] reporte dans le dossier médical personnel, à l'occasion de chaque acte ou consultation, les éléments diagnostiques et thérapeutiques nécessaires à la coordination des soins de la personne prise en*

Figure 2 : Interface de CISearch



L'interface de CISearch permet essentiellement d'effectuer des recherches en texte libre au sein des documents présent dans le dossier du patient d'intérêt. Il n'y a pas de traitement du langage naturel.

charge" [30], le maintien de la convention étant subordonné à cette obligation [31]. Toutefois, si retrouver une information dans un dossier que l'on a soi-même constitué est compliqué, la retrouver dans un dossier renseigné par de multiples intervenants, aux points de vue, organisations et pratiques différents, est d'un tout autre niveau de difficulté. L'intérêt du DMP serait considérablement amoindri en l'absence d'outils permettant de trouver dans un délai acceptable l'information que l'on y cherche.

2.2.1.2 Inclusion dans les essais cliniques

Les promoteurs et investigateurs d'essais thérapeutiques sont confrontés à de nombreux problèmes pour réaliser leurs études [32]. L'Europe dépense 16 000 000 d'euros sur 4 ans pour le projet EHR4HC [33], projet qui vise à faciliter l'utilisation des données des dossiers médicaux pour la recherche clinique. Un pan de ce projet consiste à permettre de rechercher les patients qui répondent aux critères d'inclusion et d'exclusion des essais cliniques au sein de plusieurs systèmes d'information médicaux. En effet, l'une des difficultés de la recherche clinique est l'inclusion d'un nombre suffisant de participants dans un temps limité [32]. La moitié des patients éligibles n'est en effet pas incluse dans les essais cliniques correspondants [34].

Ohman [35] a décomposé l'inclusion d'un patient dans un essai clinique en trois phases : d'abord, le praticien doit être au courant de l'existence de l'essai, ensuite, le patient doit satisfaire aux critères d'inclusion/exclusion de l'essai, et enfin, phase la plus délicate, il faut une décision du médecin et de son patient pour participer. S'il est peu envisageable d'améliorer le rendement de cette troisième phase, dépendante des croyances des médecins et des patients, des actions sur les deux premières phases ont été étudiées [36], [37] et semblent assez prometteuses.

Le principe de base est de rapprocher les informations concernant un patient des critères d'inclusions/exclusions des essais cliniques et de signaler au clinicien toute éligibilité potentielle. Cela permet de contourner totalement la première phase, le système d'information se chargeant de connaître tous les essais en cours et de simplifier la seconde, selon la qualité des informations disponibles au sein du DPI [38].

2.2.2 RI multi-patient

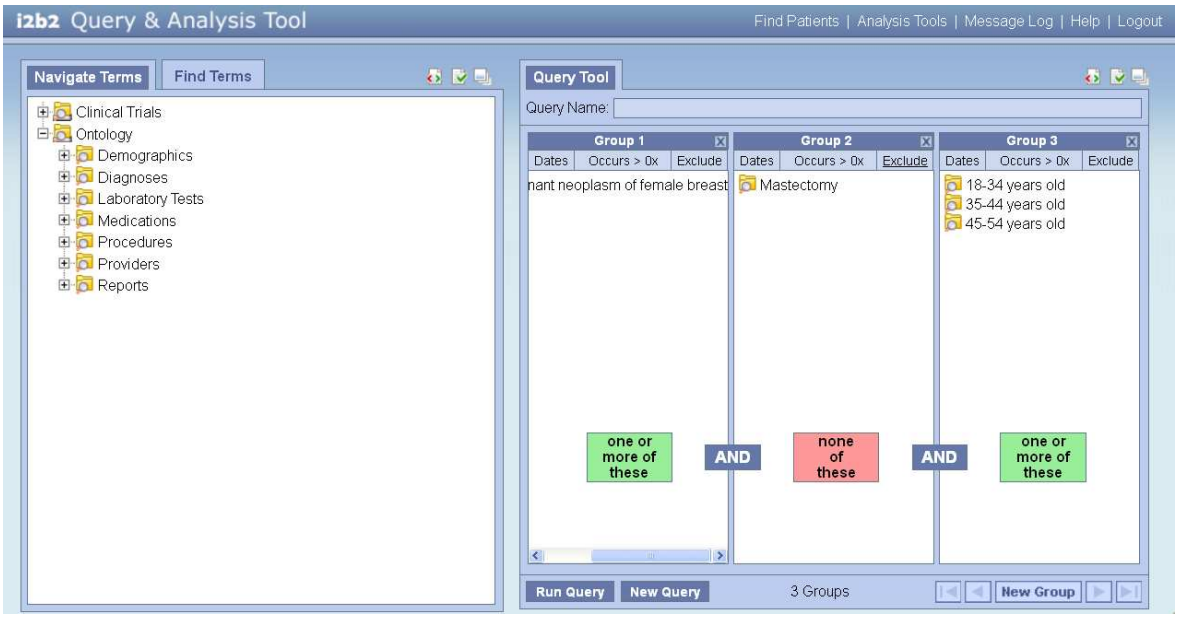
Les outils de RI multi-patient, permettant de sélectionner un groupe de patient aux caractéristiques similaires, existent depuis longtemps [13]. Leur utilisation a pourtant

été longtemps réservée aux spécialistes de l'information au prétexte que ces outils étaient compliqués à maîtriser. Les cliniciens sont demandeurs de tels outils, comme le souligne l'usage qui est fait des outils de RI mono-patient existants : Natarajan [25] a détecté, parmi les utilisateurs de CISearch, des chercheurs répétant les mêmes requêtes dans de multiples dossiers. L'enquête a révélé qu'ils essayaient d'identifier les patients respectant les critères d'inclusions de leur projet de recherche, palliant difficilement l'absence d'outil de recherche multi-patient.

Là encore de nombreux outils sont en cours de développement ou déjà disponibles. Ils sont aujourd'hui beaucoup plus simples à utiliser mais reposent encore sur des données qu'il faut bien souvent interpréter avec prudence :

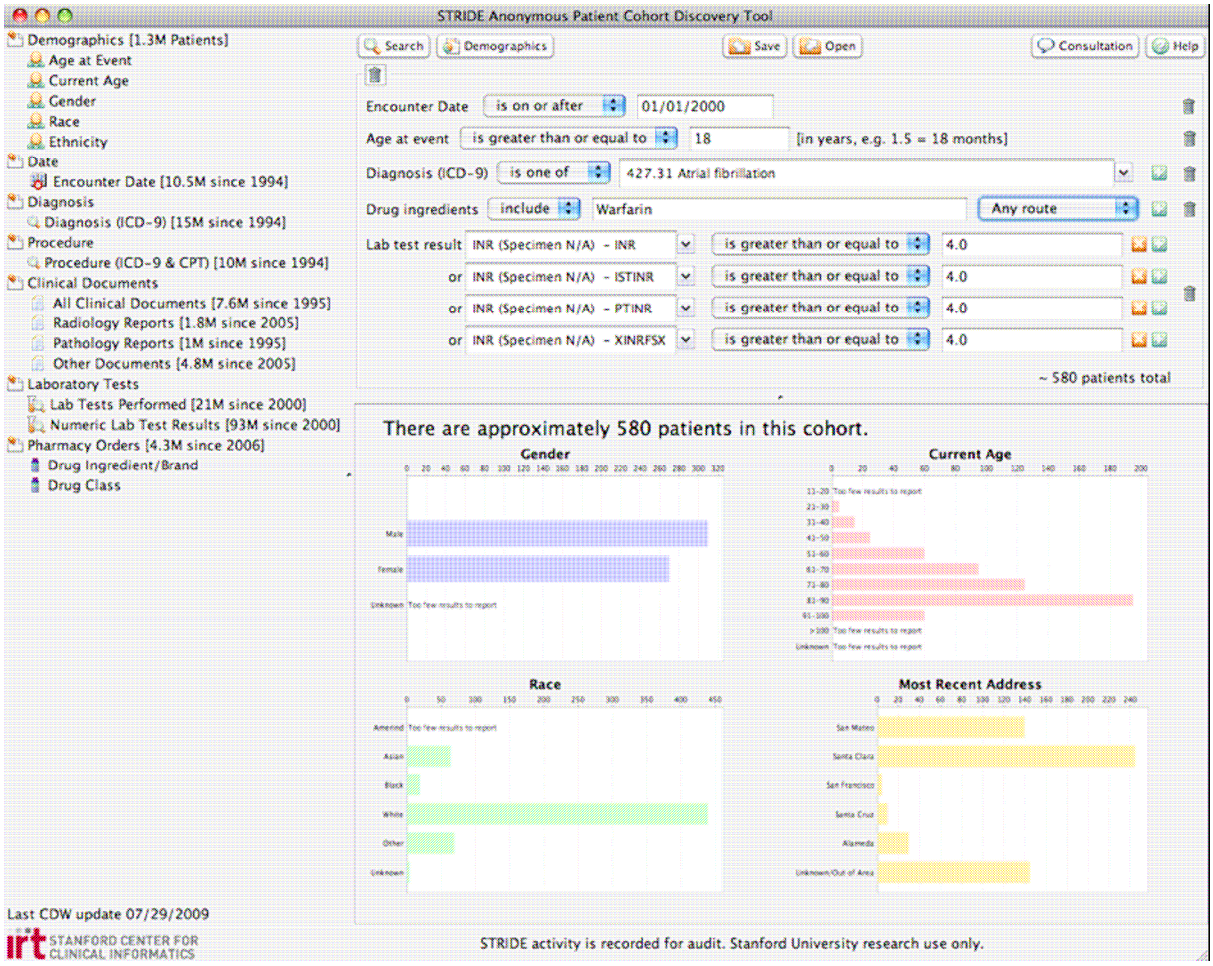
- Le projet I2B2 (Informatics for Integrating Biology and the Bedside) [39], [40] vise à faciliter l'utilisation des données cliniques et génomiques des DPI pour les chercheurs. Cet outil permet, dans un premier temps, d'effectuer des sélections de patients (voir Figure 3) qu'il sera possible d'analyser plus finement dans un deuxième temps, à l'aide de plug-ins ou manuellement. Ce projet, très avancé, commence à disséminer dans le monde [41], [42] et pourrait devenir un standard de fait.
- STRIDE (Stanford Translational Research Integrated Database Environment) est aussi un outil ayant pour vocation de faciliter la recherche translationnelle [43]. Il est construit de la même manière qu'I2B2 : un éditeur de requête permet de sélectionner des cohortes de patients (voir Figure 4) qui devront ensuite être revues manuellement pour leur inclusion/exclusion de la cohorte finale.
- Quoique non spécifique de la RI au sein des documents médicaux, Business Object (BO) est utilisé dans de nombreux établissements de santé pour effectuer des opérations de sélection de patients. Cet outil permet d'accéder aux données structurées des patients hospitalisés (données du programme de médicalisation des systèmes d'information (PMSI), des laboratoires de biologie, données administratives), mais il nécessite un gros travail de paramétrage pour être utilisable simplement par les cliniciens et fournir des résultats fiables.
- Roogle [45] est le seul outil en français existant à l'heure actuelle à notre connaissance. Il offre deux modes d'interrogation : plein texte et recherche structurée. L'indexation est fondée sur Lucene [26], avec une expansion sémantique : les documents sont indexés par les termes qu'ils contiennent, mais aussi, lorsqu'un de ces termes appartient à un vocabulaire contrôlé, par les

Figure 3 : Interface d'I2B2.



L'interface d'I2B2 permet de composer des requêtes relativement complexes : il suffit de glisser/déposer des termes des hiérarchies (à gauche) dans les groupes (à droite) pour composer des requêtes combinant ET, OU et SAUF (Ici : l*es patientes atteintes de tumeur maligne du sein, sauf celles ayant bénéficié d'une mastectomie, et ayant entre 18 et 34 ans ou 35 et 44 ans ou 45 et 54 ans.)*

Figure 4 : Interface du module de requête de STRIDE



A gauche les terminologies utilisables pour construire les requêtes. Il est aussi possible d'effectuer des recherches plein texte au sein des "clinical documents". En haut à droite, un exemple de requête. En bas à droite, les résultats, anonymes, de la requête. Voir [44] pour une description détaillée.

synonymes et les pères de ce terme au sein de l'UMLS (Unified Medical Language System) [46]. Ceci permet d'améliorer le rappel des recherches plein texte et des recherches au sein des données structurées (codage des actes (CCAM)[i], données biologiques (LOINC) et indexation automatique des comptes-rendus (MeSH)).

- Ainsi que de nombreux autres : StarTracker [47], CaFE [48], ClinQuery [13], EMERSE [49]…

On peut distinguer deux situations où ces outils peuvent s'avérer utiles : la recherche épidémiologique ou constitution de cohorte rétrospective et la mesure de la qualité des soins. Chacun de ces cadres a des particularités que nous allons détailler ci-dessous.

2.2.2.1 Recherche épidémiologique

La construction de cohorte rétrospective est un exercice qui peut être très chronophage pour les médecins qui s'y risquent dans les établissements de santé. Pour effectuer un premier tri dans l'ensemble des dossiers, les données administratives et les données du PMSI peuvent être utilisées. Les limites des bases de données médico-administratives vont induire selon les cas :

- Un manque de sensibilité [50] : l'investigateur risque alors d'avoir des difficultés à inclure suffisamment de sujets, au détriment de la puissance de son étude.
- Un manque de spécificité [51] : beaucoup de temps devra être consacré à une étude plus approfondie des dossiers médicaux (informatisés ou non) de chacun des patients identifiés pour n'en inclure finalement qu'une partie.

Des outils de RI plus poussés au sein des DPI permettraient aux utilisateurs d'affiner leurs requêtes : utilisant des mots clés plus larges s'ils désirent augmenter le rappel ou au contraire des mots clés plus précis s'ils veulent limiter le bruit. Ces outils ne peuvent bien évidemment pas trouver d'informations dans les documents papiers, mais leur introduction, en montrant aux soignants l'intérêt qu'ils peuvent avoir à l'informatisation des DP, pourrait faciliter la transition des DP aux DPI.

2.2.2.2 Mesure de la qualité des soins

La qualité des soins, son amélioration, sa mesure et l'amélioration de sa mesure sont devenues des préoccupations majeures des patients, des politiques et des

[i] Qui n'est pas (encore) incluse dans UMLS

établissements de santé [52]. Des outils de RI multi-patient permettraient de simplifier grandement la mesure de la qualité des soins : la vérification de la conformité aux recommandations des prescriptions d'AVK en cas de FA est d'autant plus simple à réaliser que les patients concernés et leurs traitements sont facilement identifiables.

L'effet d'outils de RI sur la qualité des soins au sein des établissements de santé pourrait même être double : (1) même si cela n'a pas été démontré, on peut penser que l'existence d'outils permettant de trouver des informations simplement au sein des DPI améliore directement la qualité des soins prodigués, le clinicien ayant plus d'informations pertinentes pour traiter son patient au moment où il doit prendre sa décision, en moins de temps et (2) une amélioration indirecte de la qualité des soins par une facilitation de la mesure de la qualité des soins (dégageant du temps ou des moyens pour agir sur la qualité des soins).

2.2.3 Recherche de cas similaire

Des outils de recherche d'information au sein des DPI devraient permettre aux cliniciens de retrouver des patients similaires. Cette fonction peut a priori être utile dans deux situations :

2.2.3.1 Pour l'enseignement

Pour limiter les difficultés du passage de la théorie à la pratique pour les étudiants en médecine, l'enseignement est souvent réalisé à l'aide de cas concrets. En clinique, en imagerie et en anatomo-pathologie par exemple, il est important de pouvoir illustrer ses cours à l'aide de cas, d'images ou de lames. Il est toutefois nécessaire de disposer de banques de cas importantes et dans lesquelles on peut trouver ce que l'on cherche.

Ces banques existent, dans chaque hôpital : ce sont les dossiers patients. Des solutions dédiées ont déjà été mises en place en radiologie et en anatomo-pathologie [53] mais sont encore peu répandues. Ainsi, il est toujours difficile de trouver un patient atteint d'une forme typique de polyarthrite rhumatoïde : le patient va avoir des particularités qui ne correspondent pas vraiment à la définition classique de la maladie, l'imagerie va être de mauvaise qualité… Le plus souvent, les enseignants/services se constituent une banque de cas au fil du temps, ce qui

constitue une charge de travail supplémentaire tant pour la recherche que pour le classement de ces cas !

La possibilité d'accéder facilement à l'ensemble des dossiers de patient atteints d'une maladie dans la forme sous laquelle l'enseignant le désire pour son cours devrait largement simplifier son travail.

2.2.3.2 Pour la pratique

Le raisonnement clinique des médecins est essentiellement fondé sur deux types de processus : hypothético-déductif et reconnaissance de cas [54], [55]. Le premier type consiste, pour le médecin, à émettre des hypothèses, en fonction de ses connaissances sur les pathologies et le patient qu'il prend en charge, puis à les valider ou les confondre jusqu'à trouver l'hypothèse la plus robuste. Dans le second type de processus, le médecin va effectuer un diagnostic et/ou mettre au point sa démarche thérapeutique par similarité par rapport à des situations cliniques qu'il connaît pour y avoir déjà été confronté. Cette situation est génératrice d'erreurs dans la mesure ou le praticien va se rappeler plus facilement les cas récents ou ceux qui l'ont marqué [56], ces derniers n'étant pas nécessairement les plus pertinents. La recherche d'information au sein du DPI peut limiter ce biais de rappel en proposant au médecin, non pas les patients les plus récents ou les plus marquants, mais bien ceux qui sont le plus similaire. Le clinicien pourra alors revoir dans les dossiers les prises en charge qui ont été effectuées ainsi que leurs réussites et leurs complications… Cette revue de cas ne doit pas prendre le pas sur les recommandations ou les conférences de consensus, mais les compléter ou les suppléer dans les cas ou elles sont lacunaires, datent un peu ou sont inexistantes.

3 Méthodes

3.1 Une modélisation du dossier médical adaptée à la RI

La modélisation est un élément fondamental qui conditionne le fonctionnement et les performances d'un système informatique. De nombreux travaux ont été consacrés à la modélisation des données médicales [57], [58]. La complexité des données et des organisations, l'inhomogénéité de ces dernières en fonction du type d'exercice (hospitalier ou extra hospitalier en particulier), du pays, ont fait privilégier, selon les cas, le contexte, la chronologie, l'usage des outils informatiques ou la prise en compte des différentes phases d'un processus de soins. Ces travaux ont permis d'aboutir à deux modèles normatifs basés sur les activités et les processus de soins : HISA (Health Informatics – Service Architecture) [59] et RIM (reference information model) [60]. Ils permettent d'utiliser des outils standards pour partager, communiquer et raisonner sur ces données. Ces modèles, sur lesquels sont basées, au moins pour partie, la plupart des applications récentes de gestion du DPI, ne sont pas adaptés à la RI du fait des nombreuses tables contenant des données ce qui nécessitent des requêtes complexes.

3.1.1 Le dossier patient au CHU de Rouen

Le DPI au CHU de Rouen répond à la norme HISA [59]. Il contient, sous forme structurée, les données médico-administratives. Depuis 1992, certains éléments des dossiers médicaux sont informatisés et intégrés au DPI [61] : les résultats d'analyses biologiques (réalisées en interne) sont stockés sous forme structurée, des documents d'imagerie (radio, scanner et IRM) sont stockés sous forme d'images (non structurées) et les courriers, comptes-rendus médicaux (d'hospitalisation, opératoire, d'imagerie ou d'anatomo-pathologie) et ordonnances de sorties sont stockés sous forme de texte libre (non-structurées). Une grande partie des informations ne figurent que dans les données non structurées.

Ce DPI est actuellement géré par l'application C Page Dossier Patient (CDP). Il contient plus de 4 millions de comptes-rendus et autres courriers pour 800.000 dossiers informatisés. La base de données du système CDP contient de nombreuses tables (et de nombreuses jointures) gérant les données du DPI (voir Figure 5). Comme dit précédemment, cette modélisation n'est pas adaptée à la RI.

Figure 5 : Modèle CDP simplifié

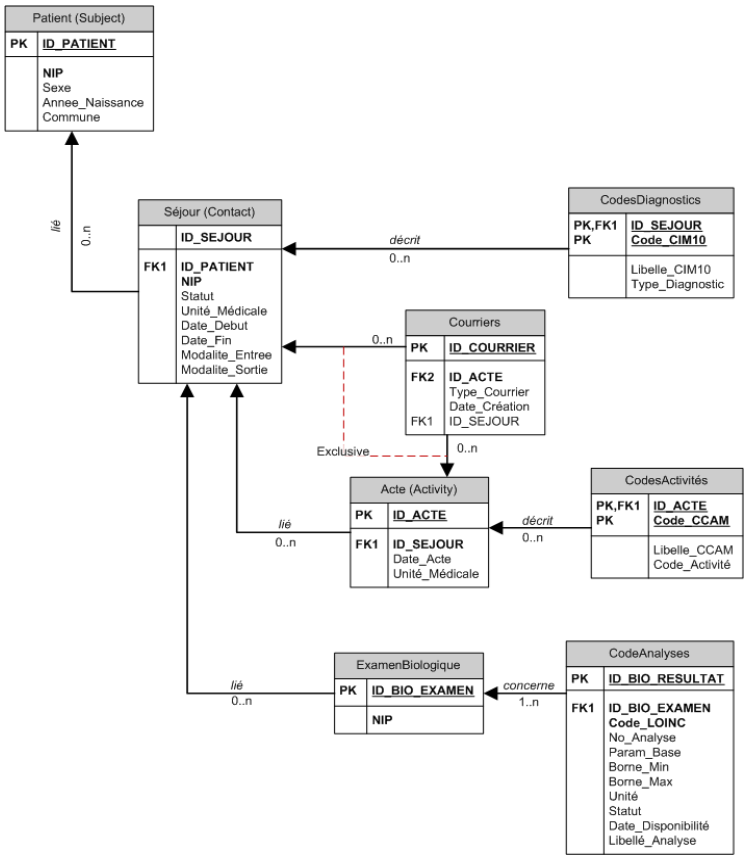
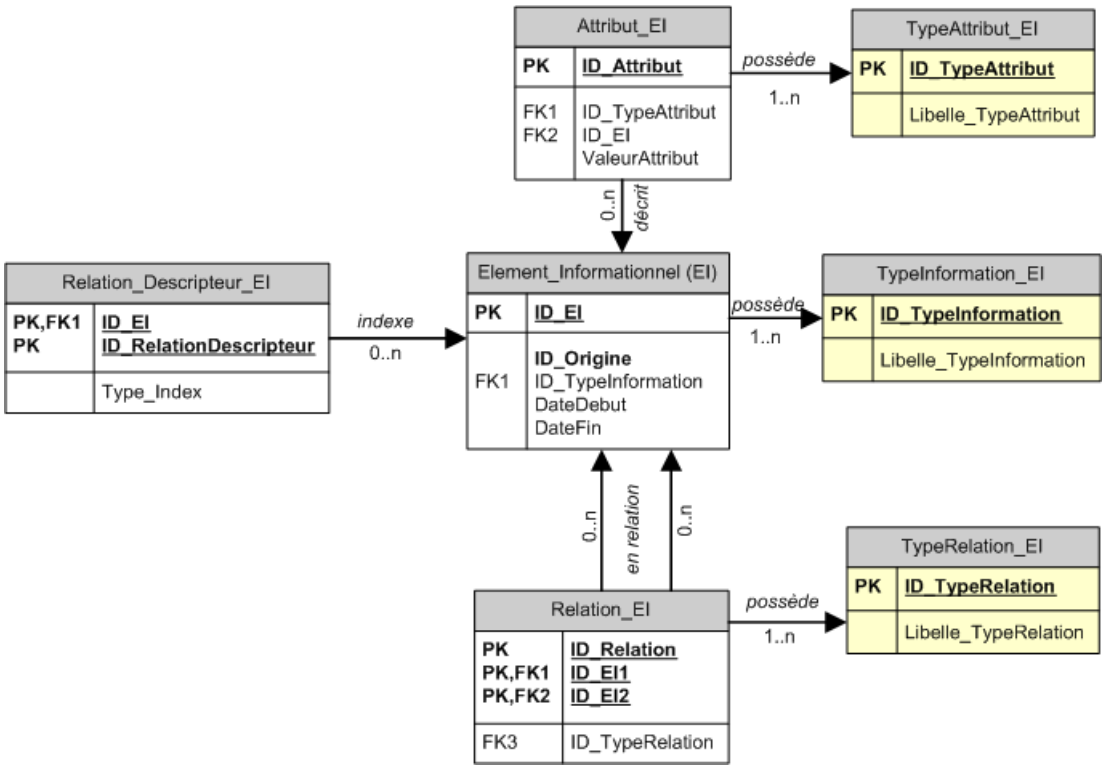


Figure 6 : Modèle du DPI adapté à la RI



3.1.2 Un modèle concis, adapté à la recherche d'information

La modélisation du dossier médical a essentiellement été réalisée par Ahmed-Diouf Dirieh Dibad, doctorant en informatique au sein de l'équipe CISMeF, et le docteur Philippe Massari.

Des modèles spécifiques, dédiées à la RI, ont été proposés [39], [53], [62] mais ne nous ont pas paru adaptés aux techniques du Web sémantique que nous prévoyons d'utiliser. Ceci nous a conduit à concevoir un modèle générique permettant : (1) d'intégrer de nouveaux types d'information sans ajouter de nouvelles tables et (2) de trouver l'information en un temps compatible avec les contraintes de l'exercice médical. Cela a aboutit au modèle présenté en Figure 6.

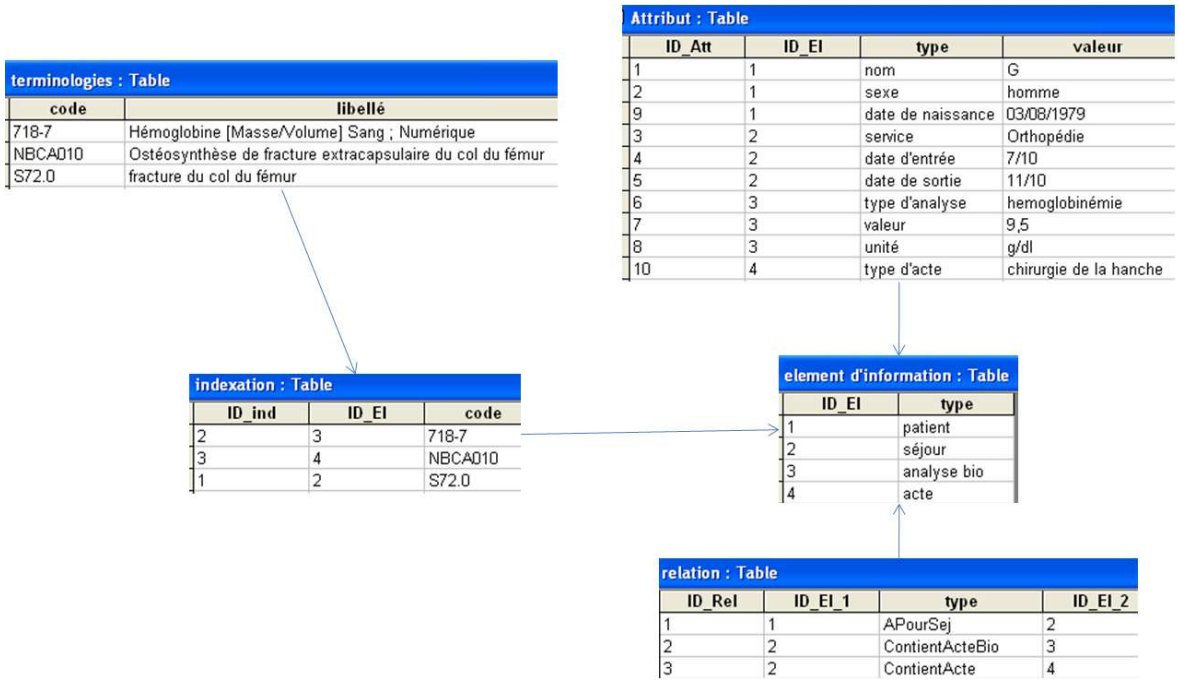
L'entité de base du modèle est «Element_Informationnel». Cette entité prend en compte tous les éléments informationnels de CDP. Un élément informationnel peut être un patient, un séjour d'hospitalisation pour une fracture de jambe, une radiographie de la cheville, un résultat d'analyse de l'hémoglobine… L'entité «Relation_EI» décrit les relations conceptuelles entre deux éléments informationnels : le patient P "a bénéficié" du séjour S1 (ou le séjour S1 "a concerné" le patient P), l'acte A "a été accompli durant" le séjour S. Les attributs spécifiques de chaque élément informationnel sont enregistrés dans l'entité «Attribut_EI», par exemple : la valeur de l'hémoglobine est de 9,5g/dl.

Les autres entités «TypeInformation_EI», «TypeAttribut_EI», «TypeRelation_EI» constituent les métadonnées de notre modèle et permet de définir les types utilisables. L'entité «Relation_Descripteur_EI» permet de gérer l'indexation d'un élément informationnel par un ou plusieurs descripteurs appartenant à une ou plusieurs terminologies médicales. Ce modèle permet ainsi d'utiliser les nombreuses terminologies nécessaires à l'indexation d'informations aussi variées que des médicaments, des examens biologiques, des pathologies, des symptômes, des actes chirurgicaux, des histologies… dans une seule table. La Figure 7 présente une instanciation de ce modèle.

3.2 Evaluation

Pour évaluer la capacité du nouveau modèle à répondre aux questions possibles des médecins, 20 dossiers patients du CHU y ont été intégrés. Ces dossiers, rassemblant 2075 prises en charge et 2377 actes, ont été choisis parmi une base de dossiers complexes anonymisés (voir Annexe A pour plus de détails quant à la

Figure 7 : Instanciation du nouveau modèle



terminologies : Table

code	libellé
718-7	Hémoglobine [Masse/Volume] Sang ; Numérique
NBCA010	Ostéosynthèse de fracture extracapsulaire du col du fémur
S72.0	fracture du col du fémur

Attribut : Table

ID_Att	ID_El	type	valeur
1	1	nom	G
2	1	sexe	homme
9	1	date de naissance	03/08/1979
3	2	service	Orthopédie
4	2	date d'entrée	7/10
5	2	date de sortie	11/10
6	3	type d'analyse	hemoglobinémie
7	3	valeur	9,5
8	3	unité	g/dl
10	4	type d'acte	chirurgie de la hanche

indexation : Table

ID_ind	ID_El	code
2	3	718-7
3	4	NBCA010
1	2	S72.0

element d'information : Table

ID_El	type
1	patient
2	séjour
3	analyse bio
4	acte

relation : Table

ID_Rel	ID_El_1	type	ID_El_2
1	1	APourSej	2
2	2	ContientActeBio	3
3	2	ContientActe	4

Monsieur G, 32 ans, est admis en orthopédie le 7/10 pour une fracture de hanche. Aucun antécédent connu, aucun facteur de risque particulier, une NFS est réalisée en préopératoire. Les résultats reviennent normaux, sauf l'hémoglobine qui est de 9,5g/dl. La chirurgie se passe bien. Le patient sort le 11/10 avec une ordonnance de rééducation.

La table des terminologies n'est pas incluse dans le modèle du dossier patient mais a été ajoutée sur le schéma pour plus de lisibilité. Les tables de métadonnées ont été enlevées pour la même raison.

complexité). Seules les données structurées : données démographiques, données de prises en charge hospitalières, codage des diagnostics et des actes (PMSI) et résultats d'analyses biologiques ont été utilisés pour l'évaluation.

Trente et un cas tests ont été construits par le docteur Philippe Massari et moi-même. Chacun de ces cas tests consiste en un cours résumé d'un dossier patient, un ou plusieurs critères de recherche et des résultats attendus. Ils permettent d'explorer la capacité du modèle à répondre à différents types d'interrogation : (1) Dans 12 cas, il fallait trouver les séjours correspondant à une pathologie ou un type de pathologie concernant un patient : épisodes infectieux, kystes ovariens… ; (2) dans 6 cas, un ou plusieurs actes étaient cherchés chez un patient : imagerie cérébrale, enregistrement polysomnographique… ; (3) dans 2 cas, il fallait retrouver des prises en charges particulières d'un patient : séjours aux urgences par exemple ; (4) dans 5 cas, la recherche mettait en jeu les relations temporelles entre plusieurs évènements survenus chez un patient : dernier hématocrite avant et premier hématocrite après chaque transfusion sanguine ; (5) Dans 2 cas, il fallait trouver des patients atteints de certaines pathologies et ayant certaines caractéristiques démographiques : les patients de moins de 50 ans ayant fait un infarctus du myocarde ; (6) dans 4 cas il fallait sélectionner des patients en fonction d'acte ou de diagnostic : ceux ayant eu des effets indésirables des médicaments, ceux au stade SIDA ayant eu une prostatite. La liste complète des cas tests est disponible en Annexe B.

Ahmed-Diouf Dirieh Dibadh a simulé, manuellement, le traitement de ces requêtes par un moteur de recherche. Les résultats ont été classés selon une échelle à trois modalités : (1) conformes à ceux attendus, (2) incomplets ou (3) erronés. Les cas qui renvoyaient des résultats incomplets ou faux étaient explorés en profondeur et des stratégies pour remédier aux difficultés rencontrées ont été élaborées.

3.3 Création de l'interface

La principale préoccupation lors de la création de l'interface a été de permettre l'édition d'un maximum de requêtes. Pour aboutir à ce résultat, on a vérifié qu'elle permettait de répondre aux cas d'utilisation créés pour l'évaluation, à d'autres issus de mon expérience et à certains issus d'une publication sur le sujet [63]. L'interface obtenue n'a pas été évaluée en termes de simplicité d'utilisation par des utilisateurs

finaux. Le fonctionnement pratique de l'outil RIDoPI sera présenté sur deux exemples distincts : l'un concernant la RI mono-patient, l'autre la RI multi-patient.

4 Résultats

4.1 Résultats de l'évaluation

Pour 22 des 31 cas d'utilisation (71% ; $IC_{95\%}$ = [55%-87%]), les résultats des recherches étaient conformes aux résultats attendus (voir Tableau 1). Pour les 9 autres cas d'utilisation, les résultats étaient soit incomplets, soit faux. L'analyse de ces erreurs a permis de mettre en évidence trois types d'erreurs : (1) Six de ces erreurs sont dues à des difficultés dans l'étape d'interprétation de la requête utilisateur (alignement avec les terminologies) : difficultés pour retrouver les codes correspondant à "maladies infectieuses", "accidents médicamenteux", "accidents des anticoagulants", "pneumopathies" ou "imageries cérébrales". (2) Deux erreurs sont dues à l'absence des libellés des unités de prises en charge : impossible dès lors de retrouver les prises en charge aux urgences d'un patient. (3) Enfin, une erreur est le fruit d'un codage PMSI discutable. Une pathologie était codée pour plusieurs séjours après celui de sa survenue, alors même qu'elle n'était plus prise en charge : en diagnostic principal d'abord, puis en diagnostic associé significatif.

4.2 Présentation de l'outil

L'outil RIDoPI est encore en cours de développement : de nombreuses fonctionnalités ne sont donc encore pas disponibles et les interfaces présentées ne sont pas toutes fonctionnelles. La forme définitive que l'interface prendra n'est pas non plus fixée. Son développement a été, et sera, essentiellement réalisée par Saoussen Sakji, post-doctorante en informatique au sein de l'équipe CISMeF.

Il existe pour le moment deux interfaces différentes pour les recherches multi et mono-patient, qui seront donc présentées séparément. Ces différences sont dues d'une part à la volonté d'avoir un démonstrateur fonctionnel à un stade précoce de développement et d'autre part pour illustrer une interface de recherche complexe (recherche mono-patient), nécessaire au formalisme des fonctionnalités de recherche, ainsi qu'une interface de recherche simple (recherche multi-patient) qui soit conviviale mais qui permette d'interpréter un maximum de requêtes.

Les fonctionnalités mises en place au sein de l'interface complexe permettent de construire toutes les requêtes des cas tests et de nombreuses autres requêtes. L'interface simple quant à elle ne permet de construire que les requêtes simples, soit 22 sur 31.

Tableau 1 : Evaluation des résultats RI

Type de requête	n	Résultats conformes	Résultats incomplets	Résultats erronés
Pathologie/Type de pathologies	12	7	2	3
Acte	6	5	1	0
Prise en charge	2	0	0	2
Relation temporelle	5	5	0	0
Données démographiques et diagnostiques (multi-patient)	2	2	0	0
Acte ou diagnostic (multi-patient)	4	3	0	1
TOTAL	31	22	3	6

4.2.1 Recherche mono-patient

L'interface de RI au sein d'un dossier médical est divisée en deux parties : la première partie rappelle l'identité du patient et ses caractéristiques (nom, prénom, date de naissance, sexe) tandis que la seconde est constituée de trois onglets, de complexité croissante.

Le premier onglet (voir Figure 8) permet d'effectuer des recherches sur les données d'indexation des séjours et des actes (à l'aide, pour le moment, de la CIM10 et de la CCAM respectivement) ainsi que sur un nombre restreint de métadonnées concernant les séjours (service, durée, date d'entrée et date de sortie). Il est possible d'associer plusieurs conditions à l'aide d'opérateurs Booléens (ET, OU) et de préciser si ces différentes conditions doivent concerner un seul séjour. Sans détailler le fonctionnement de l'ensemble des champs proposés pour effectuer des requêtes au sein du dossier médical du patient d'intérêt, notons que l'on peut rechercher des termes parmi l'indexation d'un acte ou d'un séjour d'une part et que l'on peut faire des recherches en affectant des valeurs aux métadonnées d'autre part (par exemple : *"date d'entrée = 03/08/2011"*). Pendant la saisie, un mécanisme d'auto-complétion guide l'utilisateur dans le choix des codes qu'il désire chercher (similaire à celui de la Figure 10). C'est le seul volet qui soit pleinement fonctionnel aujourd'hui.

Le second onglet est très similaire au premier, mais il est dédié à la recherche relative aux examens biologiques et à leurs résultats. Il permet donc de retrouver l'ensemble des analyses biologiques réalisées, selon leurs résultats. Ces derniers pourront être exprimés de plusieurs manières : normal/anormal, supérieur/inférieur à une valeur absolue ou positif/négatif. Le même mécanisme d'auto-complétion permettra de faciliter la saisie des termes.

Le troisième onglet permet d'effectuer des recherches complexes (contraintes temporelles relatives) avec l'ensemble des données structurées, et de leurs métadonnées, du DPI de Rouen (pathologies, actes et biologies) (voir Figure 9). Cet onglet regroupe les fonctionnalités des deux précédents volets, plus la possibilité d'ordonner les éléments cherchés : l'acte A avant le diagnostic D, par exemple la requête suivante peut être modélisée : afficher l'ECG qui a été réalisée juste après une cholécystectomie. Les recherches complexes ne nous semblent intéressantes à réaliser au sein d'un dossier médical que dans les cas des patients ayant des

Figure 8 : Interface RIDoPI pour la RI mono-patients : volet "séjour et acte"

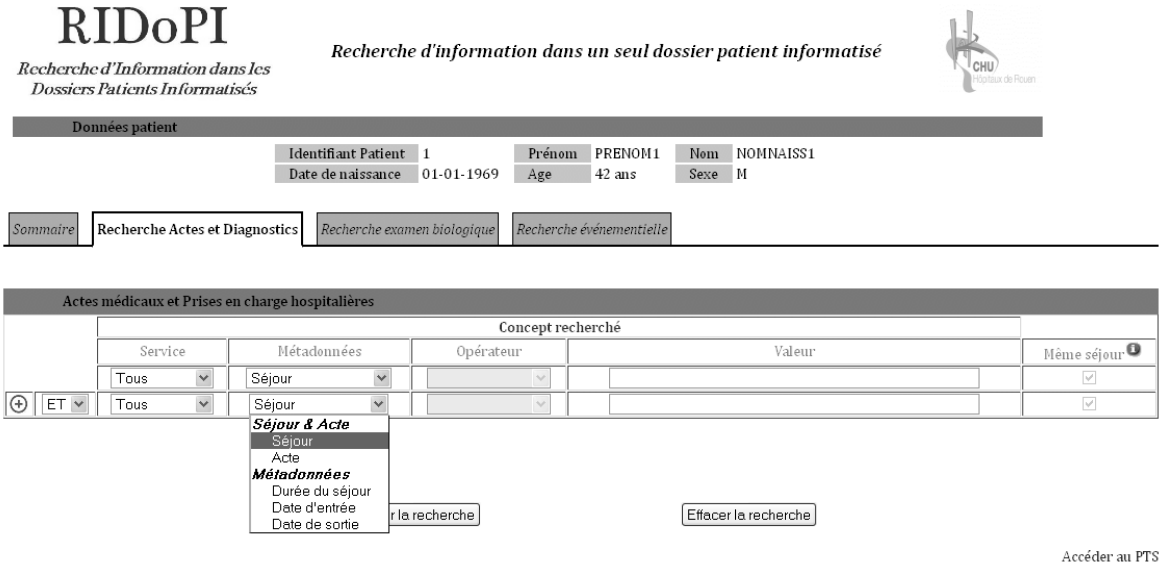
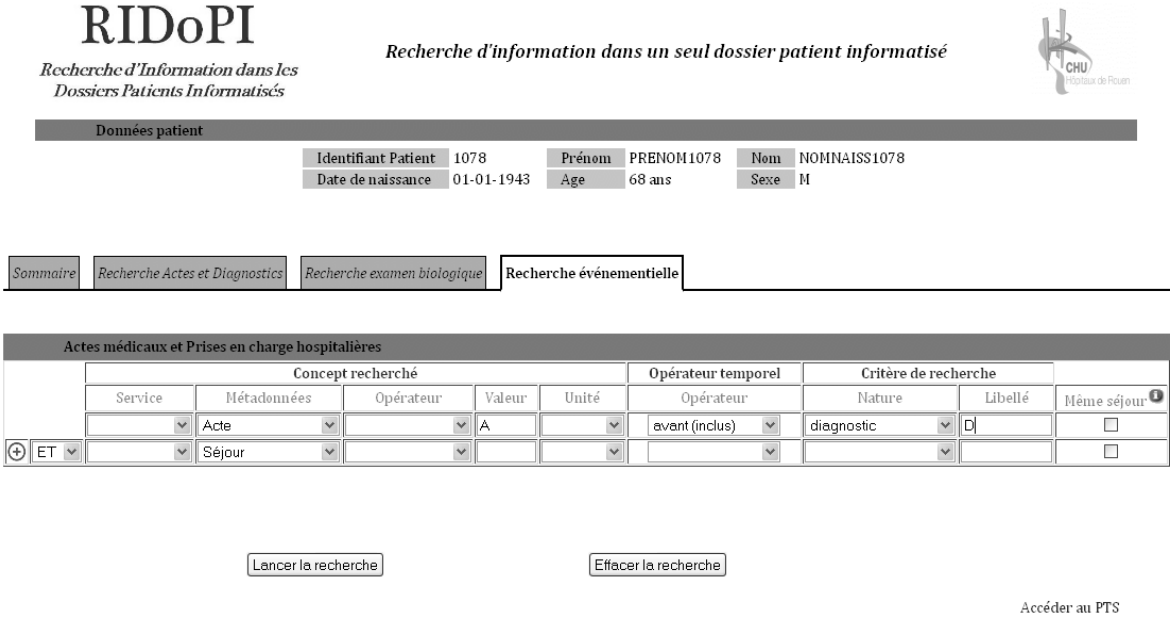


Figure 9 : Interface RIDoPI pour la RI mono-patients : volet "recherche événementielle"



histoires compliquées et suivis depuis longtemps (dossier médical riche !). Il convient de rappeler que près de 20% des dossiers médicaux contenus dans CDP contiennent plus de 50 prises en charge.

Les résultats des recherches sont renvoyés sous forme de liste d'actes ou de séjours avec quelques informations similaires à celles présentées en Figure 11 ainsi qu'avec des liens pointant vers les comptes-rendus correspondants.

4.2.2 Recherche multi-patient

L'interface de recherche multi-patient est constituée de deux champs d'interrogations dédiés, l'un à la recherche d'actes, l'autre à la recherche de diagnostics (voir Figure 10). La recherche peut, au sein de chaque champ, s'effectuer de deux manières différentes : (1) Soit on saisit un code (CCAM ou CIM10) auquel cas l'outil renvoie l'ensemble des patients dont un acte, ou un séjour, aura été indexé par ce code. (2) Soit on effectue une saisie en texte libre auquel cas l'outil RIDoPI retrouve tous les codes (CCAM ou CIM10) contenant ces mots et lance une recherche sur ces codes. Le mécanisme d'auto-complétion est toujours là pour guider l'utilisateur (voir Figure 10). Les possibilités offertes par cet outil sont encore très limitées en terme d'édition de requête, toutefois, il est prévu d'y intégrer la plupart des subtilités déjà intégrées dans la version "mono-patient" : possibilité de mettre plusieurs conditions, sur les actes, les séjours et les examens biologiques, navigation au sein des terminologies d'indexations, contraintes temporelles absolues ou relatives…

Les séjours ou actes satisfaisant à la condition de la recherche sont alors listés, regroupés par patient et par séjour, avec quelques renseignements : date d'entrée et de sortie et service d'hospitalisation (voir Figure 11). Il est prévu d'ajouter des informations sur le patient (âge et sexe). Un lien hypertexte permet d'accéder à l'ensemble du dossier de chaque patient, d'autres liens, vers les comptes-rendus de séjour et d'acte existants devraient être ajoutés prochainement.

Figure 10 : Interface RIDoPI pour la RI multi-patients

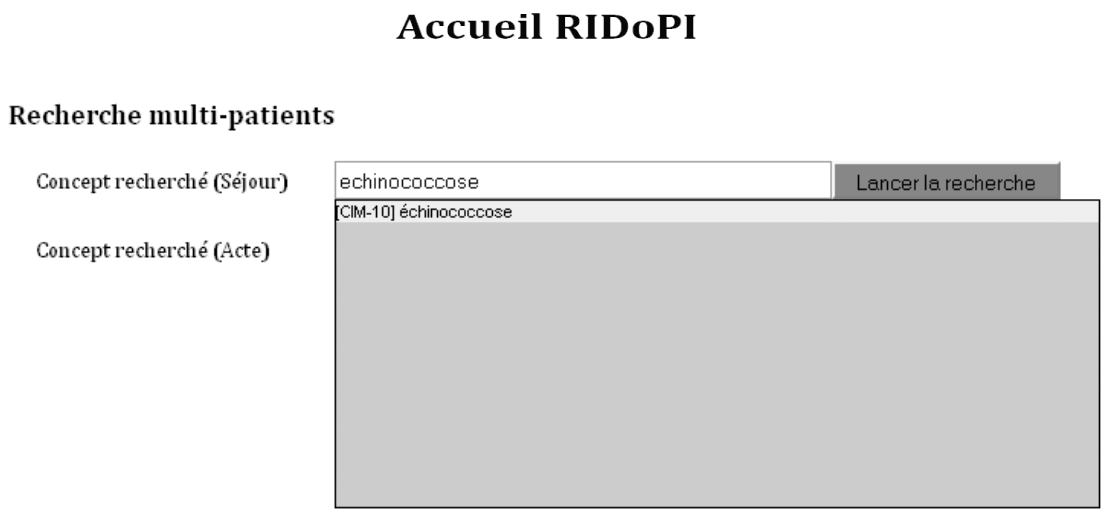


Figure 11 : Affichage des résultats d'une recherche multi-patient (RIDoPI)

Liste des séjours (10)

Identifiant Patient ▲▼					
Patient n° 474					
	Séjour n° 116631	*Date IN :* 12-02-2007	*Date OUT :* 18-02-2007	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 118133	*Date IN :* 16-03-2007	*Date OUT :* 31-03-2007	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 56451	*Date IN :* 24-12-2001	*Date OUT :* 27-12-2001	*Service :* Ortho Traumato Chir Plastique	infections à Echinococcus, autres et sans précision
	Séjour n° 87941	*Date IN :* 29-12-2004	*Date OUT :* 24-01-2005	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 97988	*Date IN :* 19-10-2005	*Date OUT :* 14-11-2005	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 89079	*Date IN :* 01-02-2005	*Date OUT :* 08-02-2005	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 97808	*Date IN :* 15-10-2005	*Date OUT :* 15-10-2005	*Service :* Maladies Infectieuses et Tropicales	infection hépatique à Echinococcus multilocularis
	Séjour n° 87776	*Date IN :* 25-12-2004	*Date OUT :* 29-12-2004	*Service :* Neurochirurgie	infection hépatique à Echinococcus, sans précision
Patient n° 1708					
	Séjour n° 64135	*Date IN :* 31-10-2002	*Date OUT :* 08-11-2002	*Service :* Chirurgie Digestive	infection hépatique à Echinococcus granulosus
Patient n° 474					
	Séjour n° 126506	*Date IN :* 28-09-2007	*Date OUT :* 28-09-2007	*Service :* Maladies Infectieuses et Tropicales	infections à Echinococcus, autres et sans précision

4.3 Cas d'utilisation illustrés

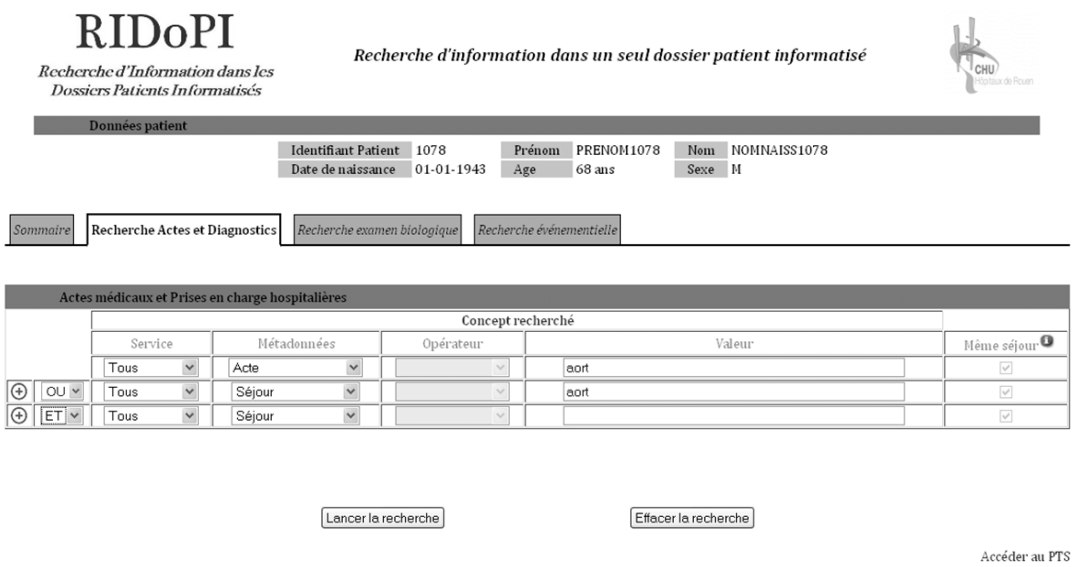
4.3.1 Cas n°1

Monsieur T, est hospitalisé en urgence pour des douleurs abdominales. L'interrogatoire révèle un antécédent de chirurgie sur l'aorte (le patient ne se souvient plus très bien, mais nous montre une cicatrice de laparotomie xypho-pubienne), réalisée dans ce même CHU quelques années auparavant, et un arrêt des matières et des gaz depuis plusieurs heures. L'imagerie révèle une occlusion haute. Avant d'opérer, le chirurgien souhaiterait avoir plus d'informations sur l'opération précédente et cherche donc dans RIDOPI l'ancien compte-rendu opératoire de Mr T. à l'aide des quelques informations qu'il a (voir Figure 12). Après moins d'une seconde d'attente, les résultats sont affichés (voir Figure 13) et le chirurgien peut accéder directement au compte-rendu qui l'intéresse.

4.3.2 Cas n°2

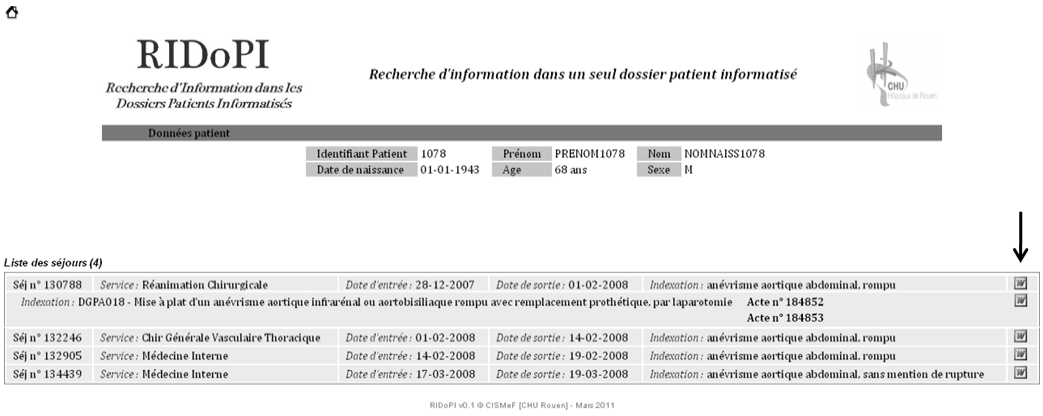
Un pneumologue suspecte que la consommation d'un additif alimentaire EXXX induirait une augmentation du risque de développer un asthme. Une brève revue de la littérature l'a conforté dans son idée sans qu'il n'ait pu voir que la relation avait déjà été mise en évidence chez l'homme. Il projette donc de réaliser une étude cas-témoin pour répondre à cette question. Les calculs de nombre de sujets nécessaires lui ont révélé qu'il devait inclure 185 cas et 370 témoins dans son étude… Il cherche donc dans l'outil RIDoPI les patients atteints d'asthme pris en charge au CHU (voir Figure 14). L'outil, après seulement quelques secondes, lui révèle la liste des patients correspondant à ses critères de sélection (voir Figure 15).

Figure 12 : L'outil RIDoPI en pratique : cas n°1 : requête



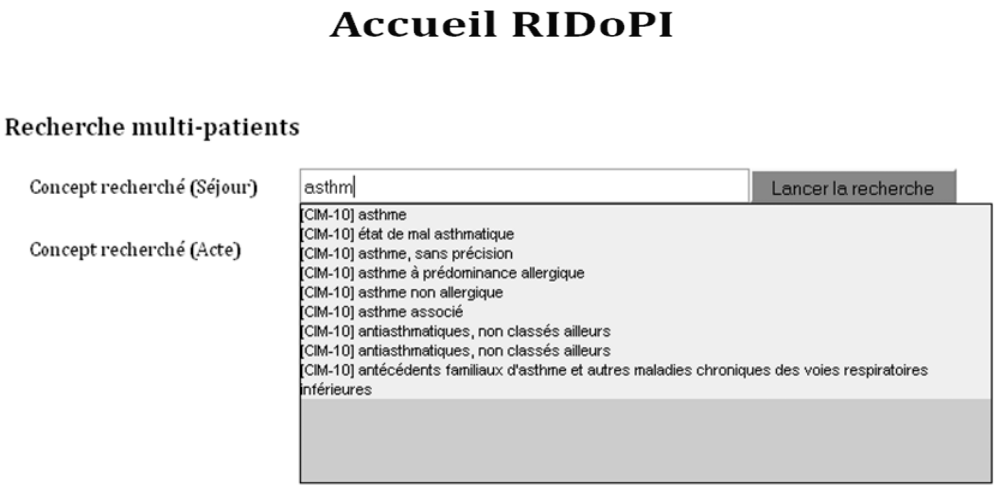
Le chirurgien cherche toute pathologie ou tout acte sur l'aorte. Comme certains actes/pathologies contiennent "aorte" tandis que d'autre contiennent "aortique", il décide d'utiliser comme mot clé "aort" qui recouvrira ces deux hypothèses.

Figure 13 : L'outil RIDoPI en pratique : cas n°1 : résultats



Rapidement, le chirurgien a à sa disposition l'ensemble des séjours et des actes concernant la pathologie aortique de son patient. Il peut accéder directement aux comptes-rendus en cliquant sur les liens à droite (flèche)

Figure 14 : L'outil RIDoPI en pratique : cas n°2 : requête



Le pneumologue cherche tous les patients atteints d'une forme quelconque d'asthme. L'auto-complétion lui permet de constater que certains codes contient "asthme" et d'autre contiennent "asthmatique". Aussi il a décidé d'utiliser comme mot clé "asthm" qui recouvrira les deux hypothèses.

Figure 15 : L'outil RIDoPI en pratique : cas n°2 : résultats

Ainsi, le pneumologue a, quasiment instantanément, la liste des patients qu'il peut inclure dans son étude.

5 Discussion

Le nouveau modèle de données a permis de retrouver aisément la majorité des informations que nous y avons cherchées. Les données que nous n'avons pas pu retrouver sont autant de pistes pour l'amélioration de la performance de la RI au sein de ce modèle. Les interfaces créées permettent de construire toutes les requêtes des cas tests, en n'ayant recours à l'interface complexe que dans un tiers des cas.

5.1 Evolutions à venir

5.1.1 Interprétation des requêtes utilisateurs

Six erreurs étaient dues à une mauvaise interprétation de la requête utilisateur. Cette étape est centrale dans la RI basée sur des vocabulaires contrôlés [64], [65]. L'équipe CISMeF a déjà été confrontée à ces problèmes pour la RI au sein du Catalogue et Index des Sites Médicaux en Français. Pour y pallier, elle a mis au point et implémenté de nombreuses méthodes dans Doc'CISMeF, le moteur de recherche Booléen dédié au CISMeF, parmi lesquelles :

1) La création de synonymes au sein des terminologies permet de limiter les situations où une recherche n'aboutit pas du fait que plusieurs expressions désignent une même notion (échocardiographie et échographie cardiaque, pneumonie franche lobaire aiguë et pneumopathie à pneumocoque…) ;
2) Les méta-termes [66] sont utilisés pour regrouper les mots clés de multiples terminologies. Ils correspondent globalement à une spécialité médicale (ex. cardiologie). Ainsi, toutes les pathologies infectieuses de la CIM sont sémantiquement liées au méta-terme "infectiologie", de même pour les codes de la CCAM spécifiques d'une pathologie infectieuse (ex. QAPA002 - Mise à plat de lésion infectieuse du cuir chevelu) ;
3) La racinisation et la lemmatisation [26], sont des techniques permettant de s'affranchir des variations flexionnelles des mots en les résumant à leur racine ou à leur lemme. Grâce à elles, l'utilisateur cherchant les "douleurs du thorax" trouvera la "douleur thoracique" ;
4) Les stratégies de recherche [67] sont des requêtes préparées par les documentalistes scientifiques de la bibliothèque médicale pour des mots clés souvent employés par les utilisateurs mais n'existant pas ou existant sous une autre forme au sein des terminologies.

Doc'CISMeF n'a pas été utilisé lors de l'évaluation. Sans permettre d'éviter toutes les erreurs, son implémentation aurait toutefois permis d'en limiter le nombre et l'importance. Il est prévu d'intégrer toutes les fonctionnalités développées au sein de CISMeF dans le futur outil RIDoPI, notamment dans le cadre du projet RAVEL.

5.1.2 Intégration des libellés

L'absence de libellés des unités d'hospitalisation est la plus triviale des 3 types d'erreurs mis en évidence par cette étude. Il suffit, pour remédier à ce problème, d'intégrer ces informations dans le modèle qui est tout à fait apte à les recevoir. Ce problème a ainsi déjà été corrigé dans la version de test de l'outil RIDoPI que nous utilisons.

5.1.3 Indexation

Le dernier type d'erreurs est dû à la qualité de l'indexation qui est utilisée : le PMSI. Il a été introduit dans les établissements de santé français par la réforme hospitalière du 31 juillet 1991 [68]. Il impose le codage des diagnostics et des actes, fournissant aux cliniciens et aux chercheurs des informations médicales structurées. Ce codage est aujourd'hui la principale base de RI dans un ou plusieurs DPI, ayant supplanté les systèmes l'ayant précédé depuis 20 ans maintenant. La production de cette information est lourde : au CHU de Rouen en 2010, 12 équivalents temps plein sont consacrés au codage des diagnostics alors que le codage des actes est assuré par les soignants eux-mêmes, réduisant la part de leur temps de travail consacrée aux soins. Aussi ce travail est-il ré-exploité pour de nombreuses autres applications : suivi d'activité, constitution de cohortes, études qualité… pour lesquelles il n'est pas envisageable de dégager autant de moyens.

De nombreuses études ont mis en évidence les limites de ces données. Le plus souvent, ces limites sont applicables à la recherche d'informations. Certaines d'entre elles tiennent aux terminologies qui sont utilisées [69]. La CIM-10 et la CCAM sont toutes deux des terminologies dites mono-hiérarchique. Cela signifie que chacun des concepts qui les compose (respectivement ≈19500 et ≈7500 termes) n'est présent qu'une fois dans la hiérarchie. Par exemple, les codes CIM10 correspondant aux infections urinaires sont situés dans le chapitre des maladies de l'appareil génito-urinaire mais absents du chapitre des maladies infectieuses. Cela n'empêche pas l'utilisateur avisé de chercher, et trouver, tous les patients atteints de maladies

infectieuses, mais cela peut conduire l'utilisateur peu au fait des terminologies à formuler des requêtes incomplètes, ramenant des résultats tronqués… Par ailleurs, l'évolution des terminologies conduit à la création ou à la disparition de certains codes, l'évolution des consignes de codage induit des modifications dans la manière de coder un évènement. En RI, on réindexe les documents dont l'ancienne indexation ne correspond plus aux règles ou aux terminologies actuelles. Ce n'est pas le cas du PMSI, aussi, l'exploitation et l'interprétation des codes sont malaisées pour les non-spécialistes [70]. Enfin, la granularité des terminologies utilisées dans le cadre du PMSI n'est pas très satisfaisante puisqu'elle varie énormément d'un terme à l'autre. Ainsi, les actes CCAM décrivant des radiographies du rachis vont préciser les niveaux radiographiés alors que ceux correspondant aux IRM du rachis vont seulement préciser le nombre de niveaux.

D'autres erreurs sont plus fonction des processus de codage. O'Malley a montré [71] que de nombreuses étapes de ce processus pouvaient être source d'une différence entre les pathologies diagnostiquées et les codes utilisés pour les décrire. Les erreurs peuvent survenir lors de la transmission des informations des médecins aux Techniciens d'Information Médical (TIM) : utilisation de synonymes pour désigner une même pathologie, erreur de transcription des comptes-rendus… On observe une grande variabilité du codage entre les codeurs [72] et intra-codeurs.

Certaines erreurs sont liées à la finalité du codage : depuis 2005, le but premier du codage PMSI est budgétaire : dans le cadre de la Tarification À l'Activité (T2A), les établissements de santé sont rémunérés en fonction des pathologies des patients qu'ils prennent en charge. Les budgets des hôpitaux dépendent donc très largement du codage de leur activité. Cet objectif comptable du PMSI amène les TIM (1) à ne pas coder les affections non rémunératrices. Les antécédents, les comorbidités et les complications légères seront moins codées et moins rattrapées par les contrôles qualité ; (2) à optimiser le codage pour favoriser l'orientation des patients dans des Groupes Homogènes de Malades (GHM) plus rémunérateurs. Ceci peut aussi bien conduire à des "exagérations" qu'à des "oublis" de code.

Ainsi le codage PMSI induit, par nature, des erreurs [72], [73] et plusieurs études [74], [75] ont montré que son utilisation induit des biais de sélection ou de classement. Il semble donc illusoire de réaliser des recherches au sein du DPI en se fondant uniquement sur ces données. Il est nécessaire de recourir à d'autres indexations réalisées : (a) à l'aide de terminologies répondant mieux aux besoins de

la RI ; (b) sur les données des comptes-rendus médicaux (d'hospitalisation, d'acte…), plus complets.

5.1.3.1 Indexation alternative

Une indexation de qualité est d'abord une indexation consistante [76] : deux indexations d'un même document devraient être identiques. C'est malheureusement loin d'être le cas dans les meilleures bases de données bibliographiques [77], [78]. L'indexation à visée de RI n'est donc pas exempte d'erreurs, mais sera moins biaisée que le codage PMSI. De nombreuses équipes étudient des moyens de favoriser l'indexation à la source - par les professionnels de santé qui diagnostiquent, prescrivent et réalisent des actes [79], [80], [81]. L'ANR TecSan 2011 finance ainsi le projet SIFaDo (Saisie Informatique Facile de Données médicales)[ii] qui vise à "développer des méthodes et [à] implémenter des outils basés sur ces méthodes conduisant à des modalités de saisie de données immédiatement utiles, faciles à utiliser même en consultation et éventuellement ludiques". De fait, les médecins sont toujours réticents à saisir des informations codées dans les dossiers médicaux. La principale cause de ce non codage est la surcharge cognitive que cette saisie impose aux médecins (interruption dans la démarche de travail, nécessité de connaître le/les référentiels terminologiques, réponse à de multiples questions pour orienter vers le choix d'un terme, sélection d'un terme dans des listes parfois longues). Par ailleurs, une telle solution ne permet pas de récupérer les vingt dernières années de dossiers médicaux non structurés et n'est donc pas suffisante.

La seule voie raisonnable de ce point de vue est l'indexation automatique, basée sur le traitement automatique du langage naturel (TAL). Ces méthodes ont largement été étudiées ces dix dernières années afin de les adapter au langage médical. Plusieurs projets et compétitions dont l'objectif est d'améliorer la recherche d'information ont été développés (Aladin DTH [82] , Biocreative [83], BioMITA [84], BioNLP shared task [85], BOOTStrep [86], CLEF [87], I2B2 [88], TREC Genomics [89]...). Le défi de ces projets est une indexation automatique, précise et de qualité de documents, mais peu concernent les dossiers patients ou le français, et la gestion des négations complique significativement cette tâche [90], [91]. De nombreuses études ont comparé les performances des indexations TAL et

[ii] Auquel participera également l'équipe CISMeF à partir de janvier 2012

PMSI [50], [51], [38], [92], [93], [94], [95]. Leurs résultats sont assez hétérogènes et ne permettent pas de conclure à la supériorité d'une stratégie par rapport à l'autre. On peut toutefois penser que l'amélioration des connaissances en TAL permettra d'améliorer les performances des outils d'indexation automatique et, subséquemment, des outils de recherche d'informations.

C'est cette voie qui est favorisée par l'équipe CISMeF, d'autant que, du fait de l'augmentation continue du nombre de ressources en santé sur internet, elle a déjà beaucoup travaillé sur l'indexation automatique de ces ressources [96], [97], [98] pour continuer à alimenter le CISMeF. Le travail de l'équipe a aboutit à un Extracteur de Concept Multi-Terminologique (ECMT) (URL : ecmt.chu-rouen.fr) [99]. Cet outil permet, comme son nom l'indique, d'extraire d'une phrase l'ensemble des concepts (= descripteurs) auxquels elle fait référence. L'intégration, déjà prévue, de ces concepts au sein du modèle permettra de les exploiter facilement en RI.

L'évaluation de cet outil pour l'indexation de comptes-rendus médicaux [100] a mis en évidence de nombreuses insuffisances, aussi ne l'a-t-on pas implémenté lors de cette étude. Une expérimentation sera réalisée avec le FMTI (qui est une version améliorée par la société VIDAL de l'ECMT) dans le cadre du projet RAVEL. Cette étape a déjà été étudiée pour le CHU de Rouen : environ 50 jours de traitement informatique pour les 4 millions de comptes-rendus !!!

5.1.3.2 Terminologies d'indexation et de RI

La plupart des terminologies, y compris celles dédiées à la recherche d'informations, n'ont qu'une couverture partielle de l'ensemble des domaines concernés par la RI au sein des DPI (La CIM-10 couvre essentiellement les pathologies, la CCAM les actes…). De la même manière, toutes les terminologies n'ont pas la même finesse concernant les domaines qu'elles couvrent : la CIM10 et ses 20.000 termes dédiés à la pathologie semble plus fine que le MeSH qui, avec guère plus de descripteurs (n=26.000), doit couvrir l'ensemble de la médecine. L'utilisation de plusieurs terminologies est donc nécessaire pour maximiser à la fois la couverture et la finesse de l'indexation [101]. Cela permet d'identifier plus de concepts dans les documents à indexer.

Le modèle, tel qu'il est conçu, ne limite, ni en qualité ni en quantité, les terminologies qui peuvent être utilisées. De nombreuses terminologies de santé sont déjà incluses dans le PTS (Portail Terminologique de Santé) [17]. Certaines sont

spécifiques d'un champ : la maladie (CIM-10), les maladies rares (ORPHANET), les actes (CCAM), la biologie (LOINC, NABM), le médicament (ATC, DCI), l'anatomie (FMA). D'autres ont une couverture plus importante de la médecine (SNOMED, MeSH). La plupart de ces terminologies ne sont pas adaptées aux utilisateurs finaux : elles peuvent être trop ou pas assez complexes, techniques, précises… Aussi, il nous semble préférable de prévoir des terminologies adaptées au vocabulaire et à la pratique clinique, dites terminologies d'interface [102].

De telles terminologies ont été créées dans le cadre du projet d'informatisation des prescriptions du CHU de Rouen. La conception "orientée médecin" de ces terminologies et leur utilisation quotidienne pour la prescription devraient assurer, à moyen terme, une bonne connaissance de ces terminologies par les cliniciens. L'intégration de ces terminologies "de prescription" ou "d'interface" dans l'outil de recherche permettra donc une capacité d'interrogation accrue au sein des dossiers patients informatisés. Pour s'adapter plus encore au mode de fonctionnement des médecins [25], on peut rajouter à ces terminologies des dictionnaires d'abréviations médicales [103],[104],[105], souvent ambiguës, qui peuvent être gérées par des outils avancés (c'est le cas dans le moteur de recherche Doc'CISMeF). Cela permet de faciliter la saisie pour l'utilisateur final : il saisit "IRA", l'outil lui propose "insuffisance rénale aiguë[CIM-10]" et "insuffisance respiratoire aiguë[CIM-10]") tout en limitant l'ambiguïté.

5.1.4 Intégration de nouvelles données

L'évaluation présentée ici a porté sur un faible nombre d'informations (PMSI et données de laboratoire). Les cas tests ayant été construits en connaissance de ces limites, ils n'ont pas mis en évidence la pauvreté d'indexation des DPI utilisés. Les données du PMSI et de biologie ne correspondent qu'à 50% des requêtes effectuées dans CiSearch [25]. Cette pauvreté des données limite l'intérêt de l'outil RIDoPI. L'indexation automatique des comptes-rendus devrait permettre d'enrichir les informations disponibles, mais nombre d'entre elles seront toujours indisponibles : médicaments prescrits et dispensés à l'hôpital, actes réalisés en dehors du CHU, courrier papier…

Intégrer ces données à notre modèle est théoriquement faisable, pour peu qu'elles soient informatisées. L'indexation automatique permet de structurer les données qui ne l'étaient pas et il n'y a alors plus qu'à intégrer cette indexation au

sein du modèle. Il peut être nécessaire d'intégrer un nouveau référentiel et de l'aligner à un référentiel déjà existant. La promotion, par l'Agence des Systèmes d'Information Partagés de santé (ASIP santé) [106], de référentiels terminologiques nationaux devrait limiter la survenue de ce type de problème.

5.2 Adaptation de l'outil aux différents cas d'utilisation

Même si les cas d'utilisation définis en introduction ont en commun les problématiques d'indexation et de modélisation, il existe aussi de nombreuses différences qui justifient des réglages différents pour chacun d'entre eux.

5.2.1 Interfaces

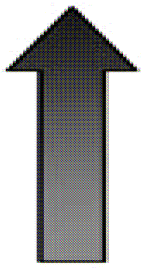
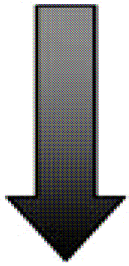
Pour être utile, un outil de RI doit être adapté aux utilisateurs finaux. Cela signifie d'une part qu'ils doivent pouvoir comprendre aisément son fonctionnement et être capable de s'en servir intuitivement et d'autre part qu'il peut répondre aux questions qu'ils se posent. Ces contraintes : complexité de l'outil d'interrogation et complexité des questions possibles, ne sont pas les mêmes en fonction du cas d'usage.

5.2.1.1 De recherche

Terry [107] distingue 5 possibilités pour interroger les DPI (voir Tableau 2). La mise en place d'un mécanisme d'autocomplétion et d'un lien vers le PTS (qui permet à l'utilisateur de naviguer au sein des terminologies) devrait diminuer la charge cognitive reposant sur l'utilisateur, quelle que soit la possibilité retenue. A terme, l'objectif est de disposer d'un outil de recherche simple et d'un outil de recherche avancée qui puisse satisfaire un maximum d'utilisateurs, quel que soit le cas d'utilisation (RI mono-patient et multi-patients).

L'interface complexe devrait ressembler à celle qui est proposée ici pour l'outil de RI mono-patient. Cette interface est certes difficile à appréhender, mais elle devrait permettre de répondre à la plupart des questions que peuvent se poser les cliniciens, s'approchant des outils existants les plus riches (STRIDE). L'interface simple quant à elle devrait se limiter à un champ d'interrogation en texte libre. S'il n'est pas nécessaire d'implémenter dans cette interface de l'outil toutes les fonctionnalités mises en place pour la version complexe, il est en revanche important de permettre des recherches relativement fines. Il sera donc nécessaire d'introduire des opérateurs permettant, par exemple, de spécifier :

Tableau 2 : Possibilités pour interroger des DPI

	Facilité d'utilisation	Niveau de complexité
Requêtes prédéfinis		
Editeur de requêtes simple		
Editeur de requêtes complexe		
Outils SQL simples		
Outils SQL avancés		

- Si la recherche s'effectue au sein du texte ou au sein des concepts l'indexant
- L'absence ou la présence d'un concept/mot : pour distinguer les situations ou le patient a de la fièvre de celles où il n'en a pas
- Le champ dans lequel doit être recherché le concept : s'agit-il d'un antécédent ou d'une pathologie active.
- …

La liste des opérateurs devra être déterminée en concertation avec les cliniciens et évoluera avec les usages. Natarajan [25] concluait, alors que son outil est bien plus simple que RIDoPI, qu'il était nécessaire de former les utilisateurs. Il semble difficile de faire autrement si l'on veut permettre aux utilisateurs de tirer parti des nombreuses fonctionnalités offertes.

5.2.1.2 De visualisation des résultats

De nombreux travaux ont aussi été réalisés sur les interfaces graphiques à proposer aux utilisateurs [108], [109], [110], des plus simples – ordonner les comptes-rendus dans le temps, tracer des courbes en fonction du temps pour les résultats biologiques – aux plus compliqués – mise en relation des différentes informations ayants traits au même problème de santé (relier la température et la NFS d'un patient à sa radio de poumon et à son diagnostic de pneumopathie, pendant que son HbA1C est associé à son diabète et à sa glycémie). Les besoins sont très différents selon les cas d'utilisation, et seront traités dans le cadre du projet RAVEL.

5.2.2 Balance rappel/précision

Le rappel quantifie la capacité d'un outil de recherche à ramener tous les documents pertinents, la précision quantifie celle à ne ramener que des documents pertinents. L'outil de recherche parfait n'existant pas (rappel = précision = 100%), il est nécessaire de faire un compromis entre ces deux valeurs. L'augmentation de l'une induisant le plus souvent une diminution de l'autre. Ce compromis semble dépendant du cas d'utilisation : un investigateur cherchant à inclure le maximum de patients dans sa cohorte tolérera sans doute plus de faux négatifs alors qu'un individu qui cherche des cas pour l'enseignement aura juste besoin de quelques vrais-positifs, sans chercher l'exhaustivité. L'idéal est donc de disposer d'un outil qui permette de favoriser le rappel ou la précision selon la volonté de l'utilisateur ou qui puisse, en

cas d'absence de résultat en mode « précision », basculer en mode « rappel » pour ne pas laisser l'utilisateur sur sa faim.

5.2.3 Problème de rafraîichissement des données

Du fait de contraintes techniques : disponibilité des bases de données, temps nécessaire à l'indexation [111],… il est peu envisageable de mettre à jour la base de données pour la RI en temps réel. Cette mise à jour aura lieu au mieux toutes les 24h. Ceci ne devrait pas avoir d'impact sur la RI multi-patients, mais peut être trompeur pour la RI mono-patient. Dans ces cas, il faudra réserver l'usage de cet outil à l'exploration du dossier pour la compréhension de l'histoire clinique, non à la recherche des derniers résultats de laboratoire.

5.2.4 Gestion de la confidentialité

Selon les buts servis, les données doivent être nominatives (prise en charge du patient), anonymes tout en permettant l'identification des cas (recrutement pour une étude) ou au contraire totalement anonymes (assurance qualité). Le but de ce travail était de mettre en place un prototype permettant d'effectuer le plus large type de recherches possible, remplissant au mieux les intérêts identifiés. Pour ne pas avoir à se soucier de la gestion de la confidentialité, les dossiers médicaux utilisés étaient entièrement anonymisés. Il est évident que, pour l'industrialisation de cet outil de recherche, beaucoup de travail devra encore être fourni, la gestion de l'anonymat des données n'en représente qu'une partie.

L'anonymisation automatique est explorée depuis plus de quinze ans [112] et de nombreux outils ont été proposés pour dé-identifier automatiquement des DPI [113]. Les meilleurs d'entre eux ayant plus de 95% de rappel et de précision pour les données identifiantes. Ces outils doivent encore s'affranchir de deux limites importantes : l'acceptabilité éthique et légale d'outils qui ne permettent pas d'anonymiser les DPI à 100% d'une part, et l'extrapolation de ces résultats à d'autres corpus et à d'autres langues d'autre part…

5.2.5 Utilisation de données de clinique courantes pour la recherche ?

Certains auteurs rappellents les difficultés à se servir des données recueillies lors de la pratique clinique pour la recherche [107], [114]. De fait, on ne peut être certain de leur exhaustivité : d'une part, les données qui n'ont pas d'intérêt en pratique clinique

ne seront pas renseignées alors qu'elles seront peut être nécessaires pour la recherche. D'autre part, les informations codées (PMSI) ou écrites dans les comptes-rendus ne sont pas les mêmes selon les médecins, spécialités, services, organisations… La non-présence d'une information dans un dossier médical n'est pas nécessairement préjudiciable pour le patient, mais peut être très dommageable lorsqu'informatisée et réutilisée à de multiples reprises dans des travaux de recherche. Par ailleurs, un certain nombre d'informations ne sont qu'approchées : dans le PMSI, on ne dispose pas de la date précise du diagnostic, juste des dates de séjour.

Ces biais, déjà présents avec les outils existant actuellement, doivent être gardés à l'esprit par tout investigateur. Les bénéfices attendus, en termes de simplification du travail de recrutement pour les essais cliniques ou les cohortes épidémiologiques, pour la mise en place de cet outil seront alors supérieurs aux risques.

5.3 Au-delà de la recherche d'information

5.3.1 Recherche contextuelle

L'ensemble des technologies mises en place pour la RI au sein du DPI peut également être utilisé, combiné à d'autres outils (PTS, info-routes), pour permettre l'accès à d'autres informations adaptées aux problématiques du patient. On pourrait ainsi aisément parvenir, à partir du DPI d'un patient diabétique de type II hospitalisé pour hypoglycémie sévère chez lequel on découvre une tumeur non-résécable productrice d'IGF-II, aux recommandations concernant la prise en charge de l'hypoglycémie chez le diabétique, aux dernières publications scientifiques concernant diabète et tumeur IGF-II sécrétante, aux interactions médicamenteuses entre chimiothérapie et insuline…

5.3.2 Réutilisation des données de santé

La mise au point d'un outil de RI performant et respectant la confidentialité et la sécurité des informations constitue une avancée importante vers la réutilisation des données de santé [115]. Les domaines possibles de réutilisations des données de santé, outre ceux traités ci- dessus, sont multiples selon Safran [116] :

“analysis, research, quality and safety measurement, public health, payment, provider certification or accreditation, marketing, and other business applications, including strictly commercial activities.”

À l’heure actuelle, cette réutilisation est quasiment inexistante, chaque nouvelle utilisation nécessitant un effort supplémentaire à la base : introduction du codage pour la facturation (PMSI et T2A), recueil spécifique pour les essais cliniques (recours aux ARC), les études épidémiologiques (Registres, enquête permanente cancer), les procédures d’accréditation (IPAQSS), la veille sanitaire (MDO, réseau OSCOUR)…

L’intégration et l’exploitation de l’ensemble des données disponibles au sein d’une base de données exploitable par les différents acteurs, avec une gestion des droits adaptée, permettrait de faire considérablement avancer l’ensemble de ces champs en réduisant d’autant le travail de collecte de données nécessaire à chacune d’elle.

6 Conclusion

Nous avons réalisé une validation de notre modèle, confirmant son aptitude à gérer les données du DPI et son adaptation à la RI tant au sein d'un seul dossier patient que dans une base multi-patients. Ce modèle a été conçu pour pouvoir aussi gérer les informations provenant de l'indexation multi-terminologique des courriers et des comptes-rendus médicaux (métadonnées, valeurs numériques et symboliques, etc.), qui seront de même format que les données codées du DPI.

Nous avons montré que des stratégies de RI (non actuellement implémentées) seraient nécessaires pour retrouver les informations médicales facilement. Elles devront être intégrées dans l'outil RIDoPI final.

L'interface complexe qui a été mise au point permet de poser toutes les questions issues des cas tests, et bien d'autres. Le juste compromis entre niveau de complexité et facilité d'utilisation ne sera trouvé qu'avec l'aide des cliniciens.

Il reste encore beaucoup de travail à fournir pour aboutir à un outil de RI convivial, simple, rapide et pratique. Il sera fourni, avec le concours d'autres laboratoires de recherche, d'équipes de recherche hospitalières et d'industriels (VIDAL, U936, LESIM, MESHS - STL et MEDASYS), lors du projet ANR TecSan 2011 RAVEL.

7 Références

[1] Breasted JH, The Edwin Smith surgical papyrus, University of Chicago Press, 1930.

[2] Article R1112-2 du code de la santé publique

[3] Moutel G. L'évolution du dossier médical et les nouvelles demandes des patients : quel impact sur la relation médecin-malade ? *La lettre du pneumologue.* X(3):99-106

[4] Tarification à l'activité. Dernier accès en août 2011. http://www.sante.gouv.fr/tarification-a-l-activite.html

[5] IPAQSS 2011 - MCO : itération de la généralisation du recueil. Dernier accès en août 2011. http://www.has-sante.fr/portail/jcms/c_627698/ipaqss-2010-mco-iteration-de-la-generalisation-du-recueil

[6] Article L1111-2 du code de la santé publique

[7] Degoulet P, Fieschi M. Informatique Médicale. Abrégé Masson 1998.

[8] Clayton PD, van Mulligen E: The economic motivations for clinical information systems. Proc. AMIA Annu. Fall Symp.1996;660-8

[9] Smith CA. Information retrieval in medicine: the electronic medical record as a new domain. *Proceedings of the American Society for Information Science and Technology*, 43: 1–30. doi: 10.1002/meet.1450430190

[10] Christensen T, Grimsmo A. Instant availability of patient records, but diminished availability of patient information: A multi-method study of GP's use of electronic patient records. *BMC Med. Inform. Decis. Making* 8 (2008) 12. doi:10.1186/1472-6947-8-12

[11] Payne TH, tenBroek AE, Fletcher GS, Labuguen MC. Transition from paper to electronic inpatient physician notes. *J Am Med Inform Assoc*. 2010 Jan-Feb;17(1):108-11. doi: 10.1197/jamia.M3173

[12] Renaud-Salis JL, Lagouarde P et Darmoni SJ. Etude des systèmes d'aide à la décision médicale. Haute Autorité de Santé. Juillet 2010

[13] Safran C, Herrmann F. ClinQuery: A program to search Boston's Beth Israel Hospital's large clinical database for patient care and clinical research. *14th Annual Symposium on Computer Applications in Medical Care*. 1990;965-6.

[14] Safran C & Chute G. Exploration and exploitation of clinical databases. *Int J Biomed Comput*. 39 (1995) 151. I56.

[15] Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ and Pifer EA. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. BMJ. 2002 March 23; 324(7339): 710.

[16] Darmoni SJ, Leroy J-P, Baudic F, Douyère M, Piot J & Thirion B. CISMeF: a structured health resource guide. Methods Inf Med. Mar 2000;39(1):30-5.

[17] Grosjean J; Merabti T; Dahamna B; Kergourlay I; Thirion B; Soualmia LF & Darmoni SJ. Health Multi-Terminology Portal: a semantics added-value for patient safety. Studies in Health Technology and Informatics. 2011;166:129-38.

[18] Shatkay H. Hairpins in bookstacks: Information retrieval from biomedical text. Brief Bioinform. 2005 Sep;6(3):222-38.

[19] Darmoni SJ, Pereira S, Névéol A, Massari P, Dahamna B, Letord C, Kedelhué G, Piot J, Derville A and Thirion B. French Infobutton: an academic and... business perspective. IOS Press. AMIA Symp 2008. Page 920.

[20] Dinakarpandian D, Williams AR, Dinakar C. Physician needs in health informatics: just ask the docs. *J Allergy Clin Immunol*. 2010 Jun;125(6):1401-1403.e4. doi:10.1016/j.jaci.2010.02.030

[21] CISMeF Bonnes Pratiques. Dernier accès en aout 2011. http://doccismef.chu-rouen.fr/servlets/CISMeFBP

[22] Vidal Recos. L'essentiel sur les recommandations thérapeutiques. Dernier accès en aout 2011. http://www.vidalrecos.fr

[23] Osheroff JA, Forsythe DE, Buchanan BG, Bankowitz RA, Blumenfeld BH and Miller RA. Physicians' Information Needs: Analysis of Questions Posed during Clinical Teaching. *Ann Intern Med*. 1991 Apr 1;114(7):576-81.

[24] Nygren E, Henriksson P. Reading the medical record. I. Analysis of physicians' ways of reading the medical record. *Comput Methods Programs Biomed*. 1992 Sep-Oct;39(1-2):1-12.

[25] Natarajan K, Stein D, Jain S, Elhadad N. An analysis of clinical queries in an electronic health record search utility. *Int J M Inform*. 2010 Jul;79(7):515-22. doi: 10.1016/j.ijmedinf.2010.03.004

[26] Hatcher E and Gospodnetic O. Lucene in Action. Manning Publications, 2004.

[27] Spat S, Cadonna B, Beck P, et al. Enhanced Information Retrieval from Narrative German-language Clinical Text Documents using Automated Document Classification. *Studies in Health Technology and Informatics.* 2008;136:473-8.

[28] Payne TH, Perkins M, Kalus R and Reilly D. The Transition to Electronic Documentation on a Teaching Hospital Medical Service. *AMIA Annu Symp Proc. 2006*;2006:629–33

[29] Ondo K, Wagner J & Gale K. The electronic medical record: Hype or reality? *Journal of Healthcare Information Management.* 2002;17(4): p2.

[30] Article L1111-15 du code de la santé publique

[31] Article L1111-14 du code de la santé publique

[32] Sung NS, Crowley WF Jr, Genel M, et al. Central challenges facing the national clinical research enterprise. *JAMA* 2003;289(10):1278–87.

[33] Electronic Health Record for Health Care. Dernier accès en aout 2011. URL : http://www.ehr4hc.eu/

[34] Ellis PM. Attitudes towards and participation in randomised clinical trials in oncology: A review of the literature. *Ann Oncol*. 2000 Aug;11(8):939-45.

[35] Ohmann C & Kuchinke W. Meeting the Challenges of Patient Recruitment - A Role for Electronic Health Records. *Int J Pharm Med* 2007;21 (4):263-70

[36] Séroussi B, Bouaud J. Using OncoDoc as a computer-based eligibility screening system to improve accrual onto breast cancer clinical trials. *Artif Intell Med*. 2003 Sep-Oct;29(1-2):153-67. doi:10.1016/S0933-3657(03)00040-X

[37] Embi PJ, Jain A, Clark J, Bizjack S, Hornung R and Harris CM. Effect of a Clinical Trial Alert System on Physician Participation in Trial Recruitment. *Arch Intern Med*. 2005 October 24; 165(19): 2272–7. doi: 10.1001/archinte.165.19.2272.

[38] Li L, Chase HS, Patel CO, Friedman C et Weng C. Comparing ICD9-encoded diagnoses and NLP-processed discharge summaries for clinical trials pre-screening: a case study. *AMIA Annual Symposium Proceedings.* 2008: 404-8.

[39] Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, Gainer V, Berkowicz D, Glaser JP, Kohane I, Chueh HC. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. AMIA Annu Symp Proc. 2007 Oct 11:548-52.

[40] Informatics for Integrating Biology & the Bedside. Dernier accès en juillet 2011. https://www.i2b2.org/index.html

[41] Mate S, Bürkle T, Köpcke F, Breil B, Wullich B, Dugas M, Prokosch HU and Ganslandt T. Populating the i2b2 Database with Heterogeneous Emr Data: a Semantic Network Approach. Stud Health Technol Inform. 2011;169:502-6.

[42] Takai-Igarashi T, Akasaka R, Suzuki K, Furukawa T, Yoshida M, Inoue K, Maruyama T, Maejima T, Bando M, Takasaki M, Sakota M, Eguchi M, Konagaya A, Matsuura H, Suzumura T, Tanaka H. On experiences of i2b2 (Informatics for integrating biology and the bedside) database with Japanese clinical patients' data. Bioinformation. 2011 Mar 26;6(2):86-90.

[43] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE–An Integrated Standards-Based Translational Research Informatics Platform. Dans: AMIA Annual Symposium Proceedings. 2009:391-5.

[44] STRIDE Cohort Discovery Tool User Guide. Dernier accès en juillet 2011. https://clinicalinformatics.stanford.edu/projects/cohort_discovery_tool_user_guide.html

[45] Garcelon N, Cuggia M, Bernicot T, Laurent JF, Garin E, Happe A et Duvauferrier R. R-oogle : système de recherche d'information du dossier patient informatisé combinant méthodes de recherches sémantiques et recherche plein texte. *In progress.*

[46] Lindberg DAB, Humphreys BL, McCray AT. The unified medical language system. Methods Inf Med. 1993;32:281–91.

[47] Gregg W, Jirjis J, Lorenzi NM, Giuse D. StarTracker : An Integrated, Web-based Clinical Search Engine. *AMIA 2003 Symposium Proceedings* p855

[48] Hanauer DA, Miela G, Chinnaiyan AM, Chang AE et Blayney DW. The registry case finding engine: an au-tomated tool to identify cancer cases from unstructured, free-text pathology reports and clinical notes. *Journal of the American College of Surgeons*. Nov. 2007 205:690-7.

[49] Seyfried L, Hanauer D, Nease D, Albeiruti R, Kavanagh J and Kales HC. Enhanced Identification of Eligibility for Depression Research Using an Electronic Medical Record Search Engine. *Int J Med Inform*. 2009 December ; 78(12): e13–e18. doi:10.1016/j.ijmedinf.2009.05.002.

[50] Gundlapalli AV, South BR, Phansalkar S, Kinney AY, Shen S, Delisle S, Perl T, Samore MH. Application of Natural Language Processing to VA Electronic Health Records to Identify Phenotypic Characteristics for Clinical and Research Purposes. *Summit on Translat Bioinforma*. 2008 Mar 1;2008:36-40

[51] Friedlin J, Overhage M, Al-Haddad MA, Waters JA, Aguilar-Saavedra JJR, Kesterson J, Schmidt M. Comparing Methods for Identifying Pancreatic Cancer

Patients Using Electronic Data Sources. *AMIA Annu Symp Proc. 2010*; 2010: 237–41

[52] Or Z, Com-Ruelle L (2008). La qualité des soins en France: comment la mesurer pour l'améliorer ? Paris, Institut de recherche et documentation en économie de la santé (document de travail 19). Dernier accès en aout 2011. http://www.irdes.fr/EspaceRecherche/DocumentsDeTravail/DT19QualiteDesSoinsEnFrance.pdf.

[53] Rubin DL et Desser TS. A Data Warehouse for Integrating Radiologic and Pathologic Data. *J Am Coll Radiol*. 2008;5:210-7.

[54] Guilbert JJ. Comment raisonnent les médecins. Genève, Médecine et Hygiène, 1992

[55] Montani M and Bellazzi R. Intelligent knowledge retrieval for decision support in medical applications. Medinfo. 2001;10(Pt 1) :498-502.

[56] Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C, Schmidt HG. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. JAMA. 2010 Sep 15;304(11):1198-203.

[57] Rector A. The interface between information, terminology, and inference models. Stud Health Technol Inform. 2001;84(Pt 1):246-50.

[58] Goble CA, Bechhofer SK, Solomon WD, Rector AL, Nowlan WA and Glowinski AJ. Conceptual, semantic and information models for medicine. In Proceedings of the 4th European-Japanese Seminar on Information Modelling and Knowledge Bases, pages 257-86, Stockholm, Sweden, 31st May- 3rd June 1994.

[59] Klein GO. Health Informatics – Service Architecture (HISA) – The intended role of the EN ISO 12967 standard – An informal guide. Septembre 2009. Dernier accès en aout 2011. http://www.consorzioedith.it/public/HISA%20-%20InformalGuideto%20its%20role-v1-3.pdf

[60] Schadow G, Mead CN, Walker DM. The HL7 reference information model under scrutiny. Stud Health Technol Inform. 2006;124:151-6.

[61] Massari P, Smuraga I, Froment L, Boudehent S, Czernichow P, Streiff J, Baldenweck M, Hecketsweiler P. Application de gestion des dossiers informatisés du CHU de Rouen. Cinquièmes Journées Francophone d'Informatique Médicale, Genève. 9-10 juin 1994.

[62] Choquet R, Daniel C, Boussaid O et Jaulent M. Etude méthodologique comparative de solutions d'entreposage de données de santé à des fins décisionnelles. IX^{ème} conférence internationale sur la science des systèmes de santé. 2008:1-6.

[63] Deshmukh VG, Meystre SM and Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. BMC Med Res Methodol. 2009 Oct 28;9:70.

[64] Lu Z, Kim W, Wilbur WJ. Evaluation of Query Expansion Using MeSH in PubMed. *Inf Retr Boston*. 2009;12(1):69-80.

[65] Thirion B, Robu I & Darmoni SJ. Optimization of the PubMed Automatic Term Mapping. *Stud Health Technol Inform*, 2009;150:238-42.

[66] Griffon N, Soualmia LF, Névéol A, Massari P, Thirion B, Dahamna B & Darmoni SJ. Evaluation of Multi-Terminology Super-Concepts for Information Retrieval. XXIII International Conference of the European Federation for Medical Informatics. Oslo, Norway, August 28th-31st 2011.

[67] Douyère M, Soualmia LF, Névéol A, Rogozan A, Dahamna B, Leroy JP, Thirion B, Darmoni SJ: Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. Health Info. Libr. J. 2004;21(4):253-261

[68] Article 1 de la Loi n°91-748 du 31 juillet 199 1 portant réforme hospitalière

[69] Cimino JJ. High-quality, Standard, Controlled Healthcare Terminologies Come of Age. *Methods Inf Med*. 2011 Mar 17;50(2):101-4.

[70] Jetté N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Paoin W, Sundararajan V, Gao S, Jakob R, Ustün B, Ghali WA; IMECCHI Investigators. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Med Care*. 2010 Dec;48(12):1105-10. doi: 10.1097/MLR.0b013e3181ef9d3e

[71] O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. Health Serv Res. 2005; 40 (5 Pt 2): 1620–1639. doi: 10.1111/j.1475-6773.2005.00444.x

[72] Vergnon P, Morgon E, Dargent S, Benyamine D, Favre M, Perrot L, Pradat E, Colin C. Évaluation de la qualité de l'information médicale des Résumés de Sortie Standardisés. Rev Epidemiol Sante Publique. 1998 Feb;46(1):24-33.

[73] Laroche M-L, Vergnenegre A, Druet-Cabanac M, Boutros-Toni F, Salamon R, Preux P-M. Qualité des données P.M.S.I. au CHU de Limoges: application de la méthode LQAS. Revue d'épidémiologie et de santé publique. 2002;50(5):433-9

[74] Lieberman D. Pitfalls of Using Administrative Data for Research. *Dig Dis Sci* 2010; 55:1506–8. doi: 10.1007/s10620-010-1246-x

[75] Keating NL, Landrum MB, Landon BE, Ayanian JZ, Borbas C and Guadagnoli E. Measuring the Quality of Diabetes Care Using Administrative Data: Is There Bias? *Health Serv Res*. 2003; 38(6 Pt 1): 1529–46. doi: 10.1111/j.1475-6773.2003.00191.x.

[76] Leonard LE. Inter-Indexer Consistency Studies, 1954-1975: A Review of the Literature and Summary of Study Results. Champaign, IL: University of Illinois Graduate School of Library Science; December 1977 (Occasional Papers, No. 131).

[77] Funk ME, Reid CA. Indexing consistency in MEDLINE. Bull Med Libr Assoc. 1983 Apr;71(2):176-83.

[78] Leininger K. Interindexer consistency in PsycINFO. Journal of Librarianship and Information Science 2000; 32; 4.

[79] McCullagh PJ, McGuigan J, Fegan M, Lowe-Strong A. Structure data entry using graphical input: recording symptoms for multiple sclerosis. Stud Health Technol Inform. 2003;95:673-8.

[80] Musser RC, Tcheng JE. Quantitative and qualitative comparison of text-based and graphical user interfaces for Computerized Provider Order Entry. AMIA Annu Symp Proc. 2006:1041.

[81] Nagy M, Hanzlicek P, Zvarova J, Dostalova T, Seydlova M, Hippman R, Smidl L, Trmal J, Psutka J. 5 Voice-controlled data entry in dental electronic health record. Stud Health Technol Inform. 2008;136:529-34.

[82] ALADIN-DTH. Dernier accès en aout 2011. http://www.aladin-project.eu/index.html

[83] Camon EB, Barrell DG, Dimmer EC, Lee V, Magrane M, Maslen J, Binns D, Apweiler R. An evaluation of GO annotation retrieval for BioCreAtIvE and GOA. BMC Bioinformatics 2005;6(Suppl): 17-27

[84] Krauthammer M, Nenadic G. Term identification in the biomedical literature. J Biomed Inform 2004;27(6):512-26

[85] Tsujii J. Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task. ACL WS:3-5

[86] BootStrep. Dernier accès en aout 2011. http://text0.mib.man.ac.uk/projects/bootstrep/

[87] Rector A, Rogers J, Taweel A, Ingram D, Kalra D, Milan J, Singleton P, Gaizauskas R, Hepple M, Scott D, Power R. CLEF: joining up healthcare with clinical and post-genomic research. In Proc of UK e-Science All Hands Meeting. 2003;264-7

[88] Uzuner O. Second i2b2 workshop on natural language processing challenges for clinical records. In Proc AMIA Annu Symp Proc 2008;1252-3

[89] Hersh W, Cohen A, Roberts P. TREC 2007 Genomics Track Overview.

[90] Chapman, W., Bridewell, W., Hanbury, P., Cooper, G. and Buchanan, B. Evaluation of negation phrases in narrative clinical reports. In Proc AMIA 2001

[91] Ceusters W, Elkin P, Smith B. Negative findings in electronic health records and biomedical ontologies: a realist approach. Int J Med Inform. 2007;Suppl 3:326-33.

[92] Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, Cai T, Goryachev S, Zeng Q, Gallagher PJ, Fava M, Weilburg JB, Churchill SE, Kohane IS, Smoller JW. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol Med. 2011 Jun 20;1-10.

[93] Love TJ, Cai T, Karlson EW. Validation of psoriatic arthritis diagnoses in electronic medical records using natural language processing. Semin Arthritis Rheum. 2011 Apr;40(5):413-20.

[94] Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. 2011 Aug 24;306(8):848-55.

[95] Cuggia M, Bayat S, Garcelon N , Sanders L, Rouget F, Coursin A, Pladys P. A full-text information retrieval system for an epidemiological registry. *Stud Health Technol Inform*. 2010;160(Pt 1):491-5. doi: 10.3233/978-1-60750-588-4-491

[96] Pereira S, Névéol A, Kerdelhué G, Serrot E, Joubert M & Darmoni SJ. Using multi-terminology indexing for the assignment of MeSH descriptors to health resources in a French online catalogue. AMIA Annu Symp Proc. 2008;586-90.

[97] Pereira S, Sakji S, Névéol A, Kergoulay I, Kerdelhué G, Serrot E, Joubert M & Darmoni SJ. Abstract multi-terminology indexing for the assignment of MeSH descriptors. AMIA Annu Symp Proc. 2009;521-5.

[98] Darmoni SJ, Sakji S, Pereira S, Merabti T, Prieur E, Joubert M & Thirion B. Multiple terminologies in an health portal: automatic indexing and information retrieval. Artificial Intelligence in Medicine. July 2009;255-9

[99] Sakji S, Gicquel Q, Pereira S, Kergourlay I, Proux D, Darmoni SJ,Metzger MH. Evaluation of a French Medical Multi-Terminology Indexer for the Manual Annotation of Natural Language Medical Reports of Healthcare-Associated Infections. *Stud Health Technol Inform*. 2010;160(Pt 1):252-6. doi: 10.3233/978-1-60750-588-4-252

[100] Pereira S, Massari P, Joubert M, Serrot E and Darmoni SJ. Exploring Multi-terminology Indexing of Discharge Summaries. In Proc MIE 2008 eHealth beyond the horizon ? get IT there. Göteborg, Sweden May 26-28, 2008

[101] Wagner MM. An automatic indexing method for medical documents. Proc Annu Symp Comput Appl Med Care. 1991;1011-7.

[102] Rosenbloom ST, Miller RA, Johnson KB, Elkin PL and Brown SH. Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems. JAMIA 2006;13:277-288 doi:10.1197/jamia.M1957

[103] www.remede.org. Dictionnaire des abréviations médicales. Dernier accès en aout 2011. http://www.remede.org/projets/dico/dico.html

[104] Lexique de terminologie médicale. Dernier accès en aout 2011. http://georges.dolisi.free.fr/Terminologie/Menu/terminologie__medicale_menu.htm

[105] Wikipédia. Liste d'abbréviations en médecine. Dernier accès en aout 2011. http://fr.wikipedia.org/wiki/Liste_d%27abr%C3%A9viations_en_m%C3%A9decine

[106] Agence des Systèmes d'Information Partagé en santé. Dernier accès en aout 2011. http://esante.gouv.fr/

[107] Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN and Cejic S. Using your electronic medical record for research:a primer for avoiding pitfalls. Family Practice 2010;27:121–6. doi:10.1093/fampra/cmp068

[108] Massari P, Pereira S, Thirion B, Derville A et Darmoni SJ. Use Of Super-Concepts To Customize Electronic Medical Records Data Display. Stud Health Technol Inform. 2008;136:845-50.

[109] Roque FS, Slaughter L et Tkatšenko A. A Comparison of Several Key Information Visualization Systems for Secondary Use of Electronic Health Record Content. Proc NAACL HLT 2010;76-83.

[110] Klimov D, Shahar Y et Taieb-Maimon M. Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med. 2010 May;49(1):11-31.

[111] Ehrler F, Ruch P, Geissbuhler A, Lovis C. Challenges and methodology for indexing the computerized patient record. Stud Health Technol Inform. 2007;129(Pt 1):417-21.

[112] Sweeney L. Replacing Personally-identifying Information in Medical Records, the Scrub System. *J Am Med Inform Assoc.* 1996;3:333–7.

[113] Uzuner O, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic deidentification. *J Am MedInform Assoc*. 2007 Sep-Oct;14(5):550-63.

[114] de Lusignan S, Metsemakers JFM, Houwink P, Gunnarsdottir V, van der Lei J. Routinely collected general practice data: goldmines for research? MIE2006, Maastricht, The Netherlands. Inform Prim Care 2006; 14: 203–9.

[115] Prokosch HU, Ganslandt T. Perspectives for Medical Informatics: Reusing the Electronic Medical Record for Clinical Research. *Methods Inf Med.* 2009;48:38–44. doi: 10.3414/ME9132

[116] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE, Expert Panel: Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc* 2007, 14(1):1-9. doi: 10.1197/jamia.M2273.

8 Abréviations

ANR : Agence Nationale de la Recherche

ASIP santé : Agence des Systèmes d'Information Partagés de santé

ATC : Anatomical Therapeutic Chemical Classification System

BO : Business Object

CCAM : Classification Commune des Actes Médicaux

CDP : C Page Dossier Patient

CISMeF : Catalogue et Index des Sites Médicaux de langue française

DCI : Dénomination Commune Internationale

DMP : Dossier Médical Partagé

DP : Dossier Patient

DPI : Dossier Patient Informatisé

ECMT : Extracteur de Concepts Multi-Terminologiques

EHR4HC : Electronic Health Record for Health Care

FA : Fibrillation Auriculaire

I2B2 : Informatics for Integrating Biology and the Bedside

IPAQSS : Indicateurs Pour l'Amélioration de la Qualité et de la Sécurité des Soins

LOINC : Logical Observation Identifiers Names and Codes

MDO : Maladie à Déclaration Obligatoire

MeSH : Medical Subject Headings

PMSI : Programme de Médicalisation des Systèmes d'Information

PTS : Portail Terminologique de Santé

RAVEL : Recherche et Visualisation des informations dans le dossier patient électronique

RI : Recherche d'information

RIDoPI : Recherche d'Information dans le Dossier Patient Informatisé

T2A : Tarification À l'Activité

TAL : Traitement Automatique du Langage naturel

TecSan : Technologies pour la Santé et l'autonomie - Edition 2011

TIM : Technicien d'Information Médical

9 **Annexes**

Annexe A : Critères de sélection des DP pour la base anonymisée

Age du patient > 40 ans et

(Nombre minimum d'hospitalisations dans le CHU > 2 ou

Au moins un séjour en hospitalisation d'une durée > 3 semaines ou

Nombre de documents contenus dans le dossier du patient >= 15)

Annexe B : Cas tests

Cas n°1

Num Patient	1
Sexe	M
Age	41
Nb prises en charge	93
Pathologies codées	Epilepsie Diabète Retard mental AVP – trauma crânien – hémorragie – triventriculaire – fracture bimalléolaire janvier 2009 Infections cutanées en janvier 2010
RI 1	Retrouver le dernier EEG
RI 2	Rechercher les épisodes infectieux

Cas n°2

Num Patient	2
Sexe	F
Age	96
Nb prises en charge	26
Pathologies codées	HTA Hypercholestérolémie Insuffisance coronarienne, infarctus, IVG Troubles du rythme cardiaque (FA 1er épisode 2004) Cœur pulmonaire chronique Phlébite superficielle Hémorragie digestive (06 02 2004) Diverticulose colique Zona Pneumopathie -> décès
Pathologies non codées	Cancer du sein mammectomie en 1989 Cure prolapsus Embolie pulmonaire en 1978 Artérite stade 2
RI 1	Retrouver le premier épisode de fibrillation auriculaire
RI 2	Y a t-il eu une hémorragie digestive ? Si oui quand ?

Cas n°3

Num Patient	4
Sexe	
Age	77
Nb prises en charge	
Pathologies codées	Poumons des oiseleurs – hémoptysie – pneumopathie à pseudomonas – pneumopathie à mycobactéries – Insuffisance respiratoire chronique Cholestéatome – surdité Diverticulose colique perforée (22/10/2001) – colectomie gauche – mélaena – occlusion (04/01/2005) HTA Infarctus du myocarde – IVG – insuffisance cardiaque globale – insuffisance mitrale – FA Hyperthyroïdie (cordarone) Hernie diaphragmatique Anémie réfractaire Accident des anticoagulants Effets indésirables des céphalosporines et bétalactamines
Pathologies non codées	Cause de l'hyperthyroïdie
RI 1	Lister les accidents médicamenteux
RI 2	Recherche d'une infection à mycobactéries
RI 3	Rechercher les consultations de cardiologie

Cas n°4

Num Patient	5
Sexe	M
Age	56
Nb prises en charge	36
Pathologies codées	Alcoolisme – tabagisme Stéatose – dysphagie Dépression HTA Dysphonie
Pathologies non codées	AVC Hernie hiatale
RI 1	Rechercher une IRM ou un scanner du crane
RI 2	Recherche des prises en charge aux urgences

Cas n°5

Num Patient	6
Sexe	F
Age	78
Nb prises en charge	18
Pathologies codées	Séquelles AVC – épilepsie HTA – IVG Embolie pulmonaire – IRC Anémie ferriprive Cystite (19/06/2001)
Pathologies non codées	Appendicectomie – cholécystectomie allergie à la PENICILLINE
RI 1	Rechercher le séjour pendant lequel il y a eu une cystite
RI 2	Rechercher une radio du bassin

Cas n°6

Num Patient	7
Sexe	F
Age	65
Nb prises en charge	30
Pathologies codées	Alcoolisme – tabagisme Pneumopathie virale – pneumopathie à pneumocoque – IRC Stéatose – RGO – œsophagite Hypothyroïdie
Pathologies non codées	Absence de reflux !!!
RI 1	Rechercher les endoscopies digestives
RI 2	Rechercher les séjours pendant lesquels il y a eu une pneumopathie

Cas n°7

Num Patient	8
Sexe	F
Age	58
Nb prises en charge	168
Pathologies codées	Maladie mitrale – insuffisance aortique – valves mécaniques – hétérogreffe – Insuffisance cardiaque – flutter auriculaire HTAP Kyste de l'ovaire en 1999 Anémie par carence en G6PD – beta thalassémie – sphingolipidose Effets indésirables des sulfamides Patient sous AVK – accident des AVK Cystite 04/04/2006 – insuffisance rénale chronique Infection après acte – complication des actes médicaux Intoxication par les opioïdes
Pathologies non codées	Nombreux codes CIM9 et Méhari
RI 1	Rechercher la date d'apparition de l'insuffisance ventriculaire gauche
RI 2	Rechercher quel séjour est dû à une complication d'un kyste de l'ovaire ?
RI 3	Rechercher s'il y a eu des complications des anticoagulants
RI 4	Rechercher les hospitalisations postérieures à 1999 pour lesquelles le déficit en Glucose-6-phosphate déshydrogénase n'est pas codé

Cas n°8

Num Patient	9
Sexe	M
Age	94
Nb prises en charge	
Pathologies codées	Insuffisance cardiaque – cardiopathie dilatée – Insuffisance aortique – insuffisance mitrale – fibrillation auriculaire – infarctus du myocarde ancien Apnée du sommeil Patient sous AVK
Pathologies non codées	Surdité – cécité
RI 1	Rechercher le dernier échocardiogramme
RI 2	Recherche d'enregistrement polysomnographique
RI 3	Rechercher le premier épisode de fibrillation auriculaire

Cas n°9

Num Patient	8
Sexe	F
Age	58
Nb prises en charge	168
Pathologies codées	Maladie mitrale – insuffisance aortique – valves mécaniques – hétérogreffe – Insuffisance cardiaque – flutter auriculaire HTAP Kyste de l'ovaire en 1999 Anémie par carence en G6PD – beta thalassémie – sphingolipidose Effets indésirables des sulfamides Patient sous AVK – accident des AVK Cystite 04/04/2006 – insuffisance rénale chronique Infection après acte – complication des actes médicaux Intoxication par les opioïdes
Pathologies non codées	Nombreux codes CIM9 et Méhari
RI 1	Rechercher le dernier électrocardiogramme (ECG) précédant la destruction d'un foyer arythmogène atrial
RI 2	Rechercher le premier électrocardiogramme (ECG) suivant la destruction d'un foyer arythmogène atrial
RI 3	Rechercher si il y a eu des complications des anticoagulants (Y44.2) et si il y a eu un TQ <20% dans les 10 jours précédents la date de codage.
RI 4	Rechercher les hématocrites (NUMPARAMBASE = NGR-HTE) avant et après les transfusions effectuées pendant le séjour du 26/02/2010 au 2702/2010

Cas n°10

Num Patient	10
Sexe	F
Age	56
Nb prises en charge	199
Pathologies codées	Maladie de Crohn – fistule anale – fistule ano-vaginale – colectomie – amputation abdomino-périnéale – iléostomie Cholécystectomie Psoriasis Hépatite médicamenteuse Spondylarthrite ankylosante – ostéomyélite Transfusion Désunion de cicatrice
Pathologies non codées	
RI 1	Rechercher l'acte d'anapath correspondant à l'examen de la pièce d'exérese du rectum (amput. Rectum laparo +ab périné HJFA007)

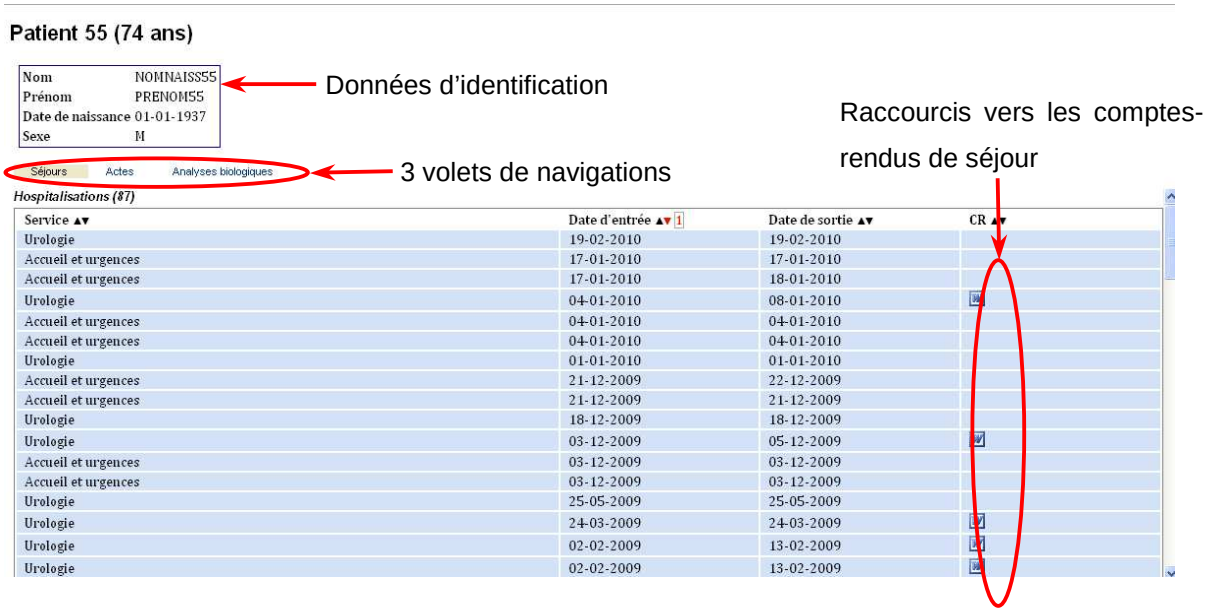
Cas N°11

Recherche multi-patients

RI 1	Rechercher les patients pour lesquels pendant le même séjour sont codés prostatite et SIDA
RI 2	Rechercher les patients de moins de 50 ans ayant fait un infarctus du myocarde
RI 3	Rechercher les patients ayant eu des effets indésirables des médicaments
RI 4	Rechercher les hommes hospitalisés pour un cancer du colon
RI 5	Rechercher les patients ayant bénéficié d'une chirurgie bariatrique.
RI 6	Rechercher les patients souffrant de pemphigoïde bulleux.

Annexe C : Interface de navigation au sein du dossier médical

Les résultats des requêtes présentent un certain nombre d'informations sur les patients, séjours, actes… qui permettent d'évaluer la pertinence du résultat par rapport à la requête mais qui n'est pas nécessairement suffisant. De nombreux liens permettent d'accéder, directement, aux comptes-rendus des différents éléments informationnels retrouvés mais il est aussi possible de retourner vers le dossier médical d'un patient pour pouvoir naviguer dedans. L'interface est alors divisée en 2 : la partie supérieure correspond aux données d'identification du patient et la seconde constitue le dossier médical proprement dit. Trois volets permettent d'accéder aux prises en charge, aux actes et aux analyses biologiques. Pour chacun de ces éléments, des liens permettent de lire les comptes-rendus (de séjour ou d'acte) ou des graphiques en fonction du temps (pour les données biologiques).



Patient 55 (74 ans)

Nom	NOMNAISS55
Prénom	PRENOM55
Date de naissance	01-01-1937
Sexe	M

Séjours | Actes | Analyses biologiques

Raccourcis vers les comptes rendus d'actes médicaux

Actes de chirurgie (13)

Acte ▲▼	Service ▲▼	Date ▲▼ 1	CR ▲▼
JELA002 - Pose d'une prothèse sphinctérienne urinaire périurétrale pénienne ou bulbomembranacée [bulbomembraneuse], par abord direct	ANESTHESIE REA CHIRURGICALE	04-02-2009	
JELA002 - Pose d'une prothèse sphinctérienne urinaire périurétrale pénienne ou bulbomembranacée [bulbomembraneuse], par abord direct	Urologie	04-02-2009	
JEGA002 - Ablation d'une prothèse sphinctérienne urinaire périurétrale pénienne ou bulbomembranacée [bulbomembraneuse], par abord direct	ANESTHESIE REA CHIRURGICALE	30-04-2008	
JEGA002 - Ablation d'une prothèse sphinctérienne urinaire périurétrale pénienne ou bulbomembranacée [bulbomembraneuse], par	Urologie	30-04-2008	

Actes d'imagerie (49)

Acte ▲▼	Service ▲▼	Date ▲▼ 1	CR ▲▼
ZCQK002 - Radiographie de l'abdomen sans préparation	Imagerie Médicale	17-01-2010	
ZCQK002 - Radiographie de l'abdomen sans préparation	Imagerie Médicale	21-12-2009	
ELQH002 - Scanographie des vaisseaux de l'abdomen et/ou du petit bassin [Angioscanner abdominopelvien]	Imagerie Médicale	03-12-2009	
ZBQK002 - Radiographie du thorax	Imagerie Médicale	03-12-2009	
ZCQK002 - Radiographie de l'abdomen sans préparation	Imagerie Médicale	03-12-2009	
JAQM004 - Échographie transcutanée unilatérale ou bilatérale du rein et de la région lombale, avec échographie transcutanée de la vessie	Urologie	25-05-2009	

Autres actes (2)

Acte ▲▼	Service ▲▼	Date ▲▼ 1	CR ▲▼
JCLE002 - Pose d'une endoprothèse urétérale, par endoscopie rétrograde	ANESTHESIE REA CHIRURGICALE	07-01-2010	
JCLE002 - Pose d'une endoprothèse urétérale, par endoscopie rétrograde	Urologie	07-01-2010	

Patient 55 (74 ans)

Nom	NOMNAISS55
Prénom	PRENOM55
Date de naissance	01-01-1937
Sexe	M

Séjours | Actes | Analyses biologiques

Libéllés des analyses biologiques avec liens vers leurs évolutions

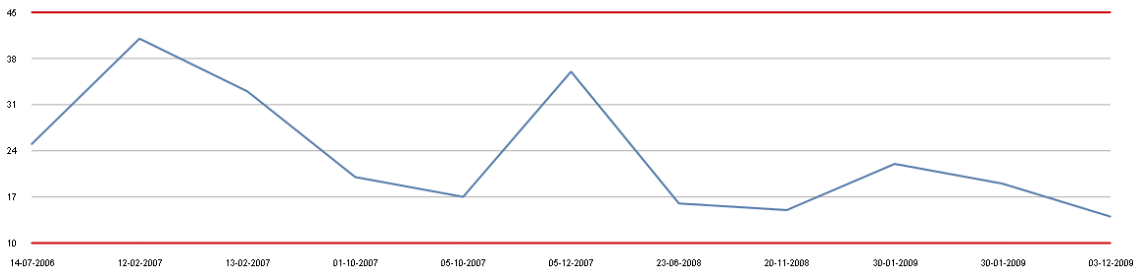
Analyses biologiques (1174)

Type d'analyse ▲▼ 2	Date--Heure ▲▼ 1	Dernier résultat ▲▼
Protéine C Réactive (22)	30-04-2008 -- 00:00:00	11 [-5]
1ère heure (8)	30-01-2009 -- 00:00:00	14 [3-10]
ALAT (TGP) (11)	30-01-2009 -- 00:00:00	19 [10-45]
ALAT (TGP) (11)	30-01-2009 -- 00:00:00	22 [10-45]
Amylase pancréatique (3)	30-01-2009 -- 00:00:00	23 [10-45]
ASAT (TGO) (10)	30-01-2009 -- 00:00:00	19 [10-35]
Bilirubine totale (13)	30-01-2009 -- 00:00:00	8 [2-18]
Bilirubine totale (13)	30-01-2009 -- 00:00:00	11 [2-18]
Calcium (30)	30-01-2009 -- 00:00:00	2.29 [2.25-2.60]
CCMH (31)	30-01-2009 -- 00:00:00	34.0 [32-36]
Chlore (30)	30-01-2009 -- 00:00:00	104 [96-106]
Cholestérol total (3)	30-01-2009 -- 00:00:00	6.3 [3.0-6.0]
Créatinine (30)	30-01-2009 -- 00:00:00	149 [65-120]
GGT (9)	30-01-2009 -- 00:00:00	25 [10-55]
Globules Blancs (31)	30-01-2009 -- 00:00:00	8.7 [4-10]
Globules rouges (31)	30-01-2009 -- 00:00:00	4.46 [4.5-5.8]
Glycémie (30)	30-01-2009 -- 00:00:00	6.4 [4.0-6.0]
Hématocrite (31)	30-01-2009 -- 00:00:00	0.38 [0.40-0.50]

Patient 55 (74 ans)

Nom	NOMNAISS55
Prénom	PRENOM55
Date de naissance	01-01-1937
Sexe	M

ALAT (TGP)



Annexe D : Poster MIE 2011

Information Retrieval in Electronic Health Record: a feasibility study

Ahmed-Diouf DIRIEH DIBAD[a], Nicolas GRIFFON[a], Saoussen SAKJI[a] (Ph.D.), Suzanne PEREIRA[b] (Ph.D.), Philippe MASSARI[a] (M.D.), Stéfan DARMONI[a1] (M.D., Ph.D.)

[a]*CISMeF, Rouen University Hospital & TIBS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France*

[b]*VIDAL, Issy les Moulineaux, France*

Abstract. Background: To allow Electronic Health Records (EHRs) being useful for medical decision making or research, the information must be easily found in it, even in voluminous EHRs. This requires to develop search capabilities for information retrieval (IR). Methods: To perform this, we propose an adapted concise model to IR. The data analysis of EHRs of Rouen University Hospital has led us to consider EHRs as being a set of events linked by conceptual relationships. After implementation, we have evaluated the capacity of the adapted concise model to take into account all data from the EHRs and its accommodation to IR. Results: We performed 31 queries on EHR. The results in 22 cases were considered successful, although mistakes are avoidable. These results confirm the ability of the adapted concise model to take into account all relevant data of EHR in IR. Conclusion: Based on the preliminary evaluation of the adapted concise model, we have demonstrated its accommodation to IR in EHR. Nevertheless, further work on larger sets is required to confirm our preliminary results.

Keywords. information retrieval; electronic health record; modeling;

Introduction

With the spread of information and communication technologies in the medical domain, an increasingly amount of health information is computerized into the Electronic Health Record (EHR). To make this information available for clinical use, information retrieval (IR) tools are needed. We present here a new model and an evaluation of its capacity to retrieve information.

1. Method

The Rouen University Hospital information system (IS) gathers clinical data in dozens of table. Only trained end-users are able to query EHR using specific codes namely CCAM and ICD10.

Based on this IS, we built a model in which: one table gathered all the events (e.g. patients, hospital visits, surgical procedure, biology test…), one table summed up all

[1] Corresponding author: Stefan J. Darmoni, CISMeF, Rouen University Hospital, 1 rue de Germont, 76031 Rouen Cedex, France; E-mail: stefan.darmoni@chu-rouen.fr.

the relations between events (surgical procedure *A* takes place during stay 1), one table contains events' attributes (biology test *B* was performed at date *T*, found value *V*), one table contained all the indexing data (diagnosis *D* was indexed by ICD-10 code I20.0).

To ensure the model effectiveness for IR, a physician (PM) created 31 test cases on 20 anonymous EHR, totalizing 2,075 hospital visits and 2,377 procedures. These test cases are clinical queries concerning only structured data (procedures, diagnoses and biology) in order to evaluate specific issues: mono vs. multi-patient search, chronology events, specific diagnoses or kind of pathology, procedures…

The results of the test cases were manually judged by an expert and classified as *relevant* when all pertinent information was restituted. In the other case, test cases results were classified as *irrelevant* and were explored manually to understand IS limits and to improve it.

2. Results

Results were relevant for 71% ($IC_{95\%}$ = [55%-87%]) of queries. Six irrelevant results were due to query interpretation: terms used for querying did not belong to terminologies used for indexing and the IS did not match them with the correct controlled terms. Two irrelevant results were due to data incompleteness introduced in the model and one irrelevant result was explained by manual indexing error.

3. Discussion & Conclusion

We proposed a new model to optimize information retrieval in EHR for individual or cohort data. Its structure allows simplified querying that is conceptually better for scalability e.g. less computing response time.

Manual exploration of results showed that six of the irrelevant results should be avoided using natural language processing tools. They have not yet been integrated in our model. Therefore, the end-user has to be very cautious in writing the queries. However, using tools developed elsewhere, i.e. predefine queries [1], super concept [2] stemming and lemmatization and synonymies will facilitate querying for end user and improve information retrieval performance of our model.

As RUH EHR follows HISA norms, this may allow the use of our adapted concise model into other EHR respecting this norm. The next step is to make our model HL7 compliant to allow its use for most EHR commercial solutions.

We described here an adapted concise model of EHR. Evaluation showed its efficacy for IR. However, future series will have to corroborate this and specify limits which remain unrecognized.

References

[1] M Douyère, L Soualmia, A Névéol, A Rogozan, B Dahamna, JP Leroy, B Thirion, SJ Darmoni. Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J* 2004: 21(4):253-261

[2] P Massari, S Pereira, B Thirion, A Derville SJ Darmoni. Use of Super-Concepts to Customize Electronic Medical Records Data Display. *In: Proc. MIE-2008. Göteborg, Sweden, May, Studies in Health Technology and Informatics*, 2008: 136:845-50.

Annexe E : Poster KR4HC 2011

A Model for Information Retrieval in Electronic Health Records

Nicolas Griffon[a], Saoussen Sakji[a], Ahmed-Diouf Dirieh Dibad[a], Julien Grosjean[a], Philippe Massari[a], Stefan Darmoni[a]

[a]CISMeF, Rouen University Hospital & TIBS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France
{nicolas.griffon, saoussen.sakji, ahmed-diouf.dirieh-dibad, julien.grosjean, philippe.massari, Stefan.darmoni}@chu-rouen.fr

Background: Information retrieval (IR) tools are required to extract knowledge from Electronic Health Records (EHR). The state of the art shows that many approaches (unstructured or structured data search, use of NLP, of semantic links…) were used for IR in EHR. **Results:** We propose an adapted concise model and interface to query EHR. Until now, the information system (IS) allows querying indexed data (diagnosis, procedures, biologic tests and hospital visits) and some useful metadata: numerical value of lab tests and temporal data of medical events (relative or absolute time-scale). **Conclusion:** Indexing unstructured data and integrating interface terminologies in our IS are still in progress. Further work is required to evaluate the performance of the system.

Keywords: information storage and retrieval, electronic health records, modeling

1 Introduction

With the spread of information and communication technologies in the medical domain, an increasingly amount of health information is computerized into the Electronic Health Record (EHR). The latter can include much information about patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, surgical and radiology reports… [1]

Historically, health information systems were programmed for administration (finance, planning or audit) purposes and then, subsequently, they were managed for clinical purposes [2]. This led to an anarchically spreading of their databases, gathering further information without necessarily allowing their use. Physician can encounter difficulties in finding information in the EHR of his patient [3]. The same problem affects the reuse of health data - analysis, research, quality and safety measurement, public health, payment, provider certification or accreditation, marketing and other business applications [4]. Extracting pieces of knowledge from this cumbersome amount of data will involve information retrieval (IR) tools in EHR [5].

This paper is organized as follows: Section 2 reviews briefly some of the numerous studies on IR tools; section 3 describes the current EHR model of Rouen University

Hospital (RUH) and the new model, established for IR purposes; section 4 presents tools improving search capacities in our model and section 5 includes discussions and conclusions.

2 State of the Art

Many different approaches were used in the different published information retrieval tools: search of structured data or free-text document, use of semantic – e.g. terms explosion or synonymy – or just graphical approach, allowing complex queries, searching for information in patient EHR or for patient in hospital EHR…

I2B2 [6] is an information system (IS) that has been developed for research and analysis purpose. A dimensional modeling approach [7] was used to build its data warehouse, centered on medical observations. It is based on structured information. iSMART [8] worked also on structured data, but formatted on CDA (Clinical Document Architecture). It identifies relevant documents thanks to semantics relationship between indexing terms.

EMERSE [9] is a tool performing free text search in medical report, as stored in hospital database. It is designed for cohort inclusion or epidemiology purpose. Users build bundle of expression that are looked for in medical report. This method is also used by CaFE [10] and Roogle [11]. Natarajan [12] had implemented a search tool based on LUCENE [13], initially oriented for daily clinical practice. Log analysis shows that physicians used it for clinical research, repeating the same search in many EHR.

Many other works have been performed in this domain but an integral review of the literature is beyond the scope of this article.

3 EHR at Rouen University Hospital

3.1 Transactional Model of EHR

The medical record computerization was performed in 1992 at Rouen University Hospital (RUH) [14]. The EHR vendor is CDP (C Page Dossier Patient). In RUH, around 800,000 unique patients are included in the EHR, which follows the highly structured HISA norm [15]. It gathers administrative, demographics, clinical and biological data. Some of these data are structured (demographics, procedures, biological results and diagnosis code) in an Oracle database but are dispersed among dozens of tables. However, most of the clinical information (discharge reports) are yet unstructured in text files.

Currently, in RUH only trained end-users are able to query structured data available for diagnostic related groups (DRG) – indexed with CCAM (French nomenclature for medical procedures) or the tenth revision of the International Classification of Diseases (ICD10) with French modifications [16].

3.2 Adapted Concise Model of EHR to IR

We designed an adapted concise model for the EHR. This allows a great simplification of the database without loss of information. Moreover, the addition of metadata can be done without modifying the model. It contains only seven tables for scalability e.g. minimization of the information retrieval response time (see figure 1). This model is broadly defined as follows:

- All the medical events are gathered in a single table,
- All the metadata concerning these events are summed up in few tables,
- All the indexing data, whatever the controlled vocabulary used – ICD10, CCAM and Logical Observation Identifiers Names and Codes (LOINC) [17], are available in a single table,

To achieve this simplification, we considered each medical event as a simple informational element (IE) e.g. patient, hospital visits, procedures, biological exams and reports. Then, the EHR is a set of IE which are linked by conceptual (e.g. all stays of one patient) and temporal relations (e.g. a procedure was performed during one stay). Each IE is identified by metadata and described by concepts belonging to controlled vocabularies.

The main class of the model is "***Informational Element***" which contains all the IE (structured and unstructured) of EHR. The class "***Relation_IE***" describes the conceptual and temporal relationships between a couple of IE. The three classes "***TypeRelation_IE***", "***TypeAttribut_IE***" and "***TypeInformation_IE***" define the metadata type. The class "***Relation_Descriptor-IE***" describes the indexing of the IE using the medical terminologies. This indexing represents structured data (e.g. visits and procedures), using DRG encoding, or unstructured data (e.g. medical reports), which will be done, in future, with the F-MTI [18]. The specific attributes for each IE are stored in "***Attribute_IE***". Health terminologies, used for indexing, are stored in RDF database.

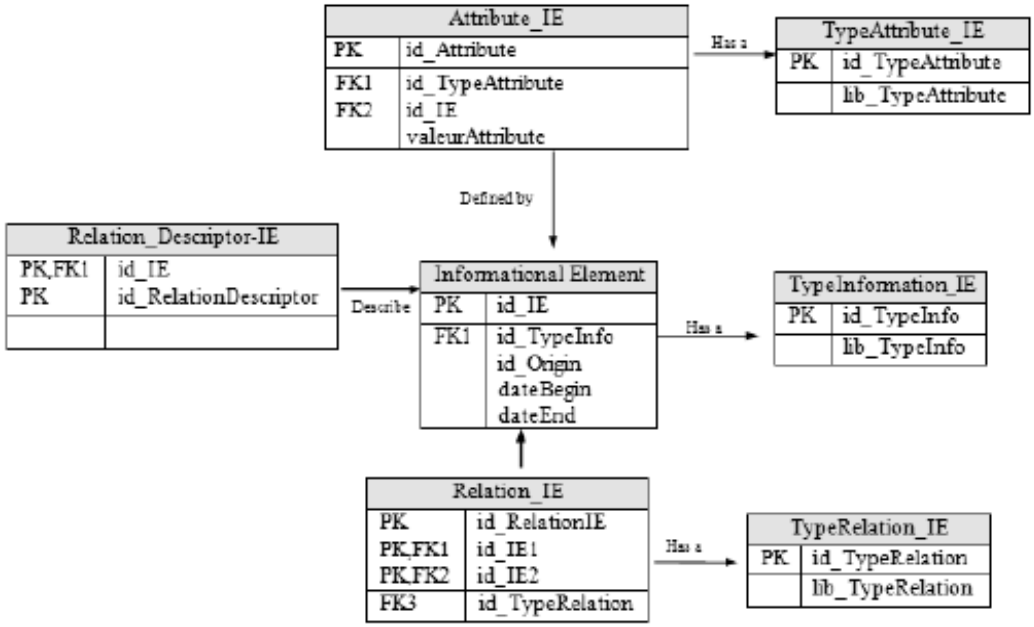


Fig. 1. New model of the Electronic Health Record

4 Tools for Search in RUH EHR

Querying in this model using a specific interface will soon be possible for the end user (see figures 2, 3 and 4). Our search tool is based on Doc'CISMeF search engine and its enhancement [19]. As this tool was initially developed for medical literature IR, it allows only basic searches for visits, procedures, diagnoses and laboratory test orders indexed with references terminologies (ICD10, LOINC, CCAM):

The test cases, performed for a preliminary study [20], showed that IR in EHR is far more complex than in internet health resources. To allow the use of Doc'CISMeF in EHR, we had to add inequalities operators (<, >). This addition allows us to answer quite complex queries, using metadata, in one or many EHR:

- Visits metadata queries e.g. searching neurology visits during more tan ten days (see figure 2);
- Numerical value of laboratory tests e.g. searching all patient with γGT > 150 U/l or γGT > 3N (see figure 3)
- Relative temporal relationship e.g. searching for the electrocardiogram just before and after patient flutter ablation (see figure 4), searching for patient with surgery and infection during the same hospital visit;

Test cases revealed also that references terminologies may not fit well to end-users knowledge. To facilitate use of the tool and, by the way, improve IR, we plan to add: (1) a health multi-terminology portal (HMTP) [21], (2) an auto-completion method based on HMTP and (3) interface terminologies [22] for biology, radiology and nursing.

Fig. 2. Search tool in patient EHR. Example of visit metadata query.

Fig. 3. Search tool in patient EHR. Example of lab result query.

Fig. 4. Search tool in patient EHR. Example of query using temporal relationship

5 Discussion

This new user interface is far more friendly than the old one. Besides more data are available to query in which can be useful for RUH physicians. We believe it will also interest clinicians in other hospitals. As RUH EHR follows HISA norms [15], this may allow the use of our IS for other EHR respecting this norm. The next step is to make our model HL7-compliant to allow its use for most EHR commercial solutions.

A preliminary study performed on this model, with basics search tools, showed quite good relevance [20]. Performances of the new IS have to be evaluated precisely. Recall and relevance needs can be different according to use cases (clinical trial, quality insurance, daily practice…). This will probably lead to different queries interpretation depending on user aims. The presented model ability to retrieve necessary data for time oriented, problem oriented or summarized views have also to be evaluated.

Searching in unstructured data – the medical reports – is not yet possible, but their semantic indexing with an automatic tool, like F-MTI [18], had been anticipated: the model is generic enough to allow their integration .

We described here an adapted concise model of EHR that theoretically answer to many physician questions. The next step is to evaluate our IS performances.

6 References

1. Health Information Management Systems Society's, http://www.himss.org/ASP/topics_ehr.asp
2. Clayton, P D, van Mulligen, E: The economic motivations for clinical information systems. Proc. AMIA Annu. Fall Symp. 660--668 (1996)
3. Payne, T H, tenBroek, A E, Fletcher, G S, Labuguen, M C. Transition from paper to electronic inpatient physician notes. J Am Med Inform Assoc. 17(1):108-11. (2010 jan-feb) doi: 10.1197/jamia.M3173
4. Safran, C, Bloomrosen, M, Hammond, W E, Labkoff, S, Markel-Fox, S, Tang, P C, Detmer, D E, Expert Panel: Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007, 14(1):1-9. doi: 10.1197/jamia.M2273.
5. Lau, A. Next generation of electronic patient record: moving from information to knowledge-based. Advances in Computer Science and IT. ISBN: 978-953-7619-51-0. (2009)
6. I2B2, https://www.i2b2.org/software/projects/datarepo/CRC_Design_15.pdf
7. Kimbal, R. A dimensional modeling manifesto. DBMS Magazine. August 1997.
8. Liu, S, Ni, Y, Mei, J, Li, H, X, G, Hu, G, Liu, H, Hou X, Pan, Y: ISMART: ontology-based semantic query of CDA documents. Proc.AMIA Annu. Symp.375-9 (2009)
9. Seyfried, L, Hanauer, D, Nease, D, Albeiruti, R, Kavanagh, J, Kales, H C: Enhanced Identification of Eligibility for Depression Research Using an Electronic Medical Record Search Engine. Int. J. Med. Inform. 78(12): e13–e18 (2009)
10. Hanauer, D A, Miela, G, Chinnaiyan, A M, Chang, A E et Blayney, D W. The registry case finding engine: an automated tool to identify cancer cases from unstructured, free-text pathology reports and clinical notes. Journal of the American College of Surgeons. 205:690-7. (Nov 2007)
11. Cuggia, M, Bayat, S, Garcelon, N, Sanders, L, Rouget, F, Coursin, A, Pladys, P. A full-text information retrieval system for an epidemiological registry. *Stud Health Technol Inform.* 160(Pt 1):491-5 (2010). doi: 10.3233/978-1-60750-588-4-491

12. Natarajan, K, Stein, D, Jain, S, Elhadad, N. An analysis of clinical queries in an electronic health record search utility. Int J M Inform. 79(7):515-22. (2010 Jul) doi: 10.1016/j.ijmedinf.2010.03.004
13. Hatcher, E, Gospodnetic, O. Lucene in Action. Manning Publications, 2004.
14. Massari, P, Smuraga, I, Froment, L, Boudehent, S, Czernichow, P, Streiff, J, Baldenweck, M, Hecketsweiler, P: Application de gestion des dossiers informatisés du CHU de Rouen. Cinquièmes Journées Francophone d'Informatique Médicale. 321-320 (1994)
15. HISA norm, http://www.consorzioedith.it/public/HISA - InformalGuideto its role-v1-3.pdf
16. Agence Technique de l'Information sur l'Hospitalisation, http://www.atih.sante.fr/index.php?id=000020000000
17. Logical Observation Identifiers Names and Codes, http://loinc.org/
18. Pereira, S, Sakji, S, Névéol, A, Kergourlay, I, Kerdelhué, G, Serrot, E, Joubert, M, Darmoni, S J: Multi-terminology indexing for the assignment of MeSH descriptors to medical abstracts in French. Proc. AMIA Annu. Fall Symp.521-5 (2009)
19. Douyère, M, Soualmia, L F, Névéol, A, Rogozan, A, Dahamna, B, Leroy, J P, Thirion, B, Darmoni, S J: Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. Health Info. Libr. J. 21(4):253-261 (2004)
20. Dirieh Dibad, A D, Griffon, N, Sakji, S, Pereira, S, Massari, P, Darmoni, S J. Information Retrieval in Electronic Health Record: a feasibility study. MIE symp. (2011)
21. Health multy-terminology server, http://pts.chu-rouen.fr
22. Rosenbloom, S T, Miller, R A, Johnson, K B, Elkin, P L, Brown, S H. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006;13:277–288. doi: 10.1197/jamia.M1957.

RÉSUMÉ

Introduction : Les dossiers patients informatisés constituent une source de données médicales importante. Cette manne d'information pourrait être réutilisée de multiples manières : pour les soins, la recherche clinique, la recherche épidémiologique et l'enseignement. L'absence d'outil de recherche d'informations au sein des dossiers patients est un des facteurs limitant cette réutilisation.

Objectifs : Les objectifs de ce travail sont : (1) d'évaluer un modèle de base de données dédié à la recherche d'informations, (2) de pallier les difficultés mises en évidences et (3) de travailler à l'interface d'un outil de recherche d'informations.

Méthodes : 20 dossiers patients anonymisés ont été intégrés au modèle. 31 cas tests ont été construits à partir de ces dossiers. Ils consistent chacun en un cours résumé d'un dossier patient, un ou plusieurs critères de recherche et des résultats attendus. Le moteur de recherche a été simulé pour réaliser ces recherches. Les résultats ont été classés en trois catégories : conformes, incomplets ou erronés. L'interface doit permettre de formuler un maximum de cas tests.

Résultats : Les cas tests ont été conformes dans 71% des cas ($IC_{95\%}$ = [55%-87%]). Les résultats incomplets ou erronés sont tous susceptibles d'être corrigés par des méthodes déjà connues : racinisation, synonymie, multi terminologie… Deux interfaces ont été créées : la première permet de formuler tous les cas test mais est relativement complexe, la seconde, beaucoup plus simple, permet de formuler 22 des 31 cas tests.

Conclusion : L'évaluation préliminaire confirme l'aptitude du modèle à gérer les informations des dossiers patients informatisés.