

# Conception et développement d'une interface d'interrogation sur un entrepôt de données biomédicales multimodales au CHRU de BREST

## Mots-clés

Entrepôt de données - Fouille de données - Big Data – Ontologies – Données biomédicales – Recherche plein texte - Interface utilisateur – Knowledge Map -TSQL

## Contexte

Tous les hôpitaux sont confrontés au problème de la dispersion des données médicales numériques sur des supports multiples qui sont indispensables à la pratique du spécialiste mais qui n'offrent qu'une vision fractionnée et incomplète du patient et de son dossier médical. Le Centre de Données Cliniques (CDC) souhaite pallier cette carence en recueillant, organisant, standardisant et exploitant les données médicales issues des différents logiciels métiers du CHRU de Brest. Sa composante technique est un entrepôt colligeant des données multimodales (comptes rendus d'hospitalisation, biologie, données omiques, médicaments, imagerie, données administratives et financières,...) de manière historisée pour permettre la reconstitution des parcours patients. Les données recueillies sont soit structurées (CIM-10, CCAM, ATC ..) autorisant des recherches s'appuyant sur des nomenclatures internationales ou sous forme de textes libres permettant de fouiller le contenu textuel des comptes-rendus. L'outil doit permettre aux équipes de recherche de disposer des informations produites par les équipes médicales sous une forme facilement exploitable.

## Objectifs

Cet entrepôt est en cours de chargement et dispose d'une interface sommaire permettant d'interagir avec le système (voir annexe 1). Nous souhaitons refondre cette interface avec comme objectifs :

- Une épuration de l'ergonomie permettant un abord initial plus aisé des données tout en permettant un mode d'utilisation plus expert
- Une navigation plus conforme aux logiciels standards (menus, onglets,...)
- La création d'un module de type « Advanced Query Builder » permettant d'exploiter pleinement tous les niveaux de granularité des données structurées contenues dans

l'entrepôt et les fonctions natives du SGBD ORACLE (et notamment des fonctionnalités de recherche plein texte)

## **Matériels et Méthodes**

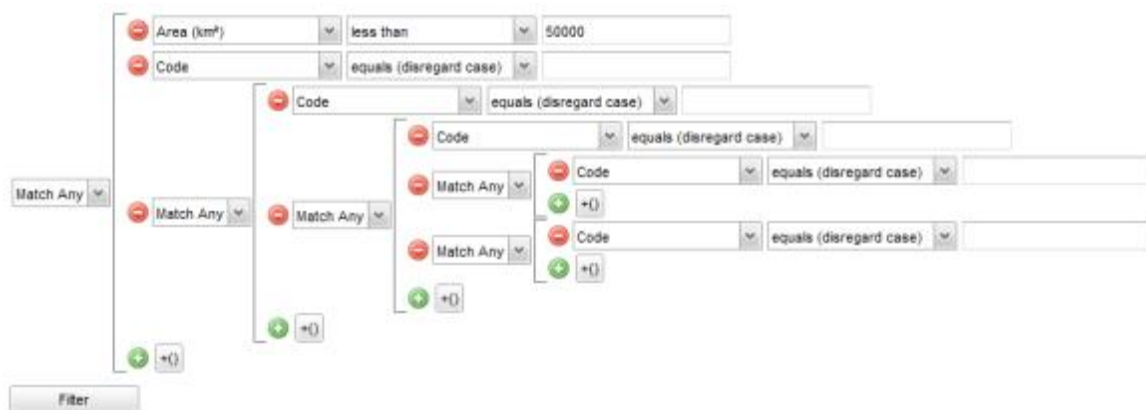
Le contexte technique de cet entrepôt privilégie les outils de type open source. La base de données en revanche s'appuie sur le SGBD ORACLE 11g2 et utilise notamment les modules ORACLE TEXT(recherche plein texte au sein des comptes-rendus) et ORACLE SPATIAL (stockage de coordonnées pour affichage cartographique) . Les développements des outils d'alimentation et de l'interface actuelle font appel à un serveur WEB sous Linux (RED HAT6) et le langage PHP. La couche ergonomique s'appuie sur AJAX.

Sur le plan méthodologique nous proposons 3 temps à ce stage :

- Une phase exploratoire de l'état de l'art en matière d'interface homme machine dans le domaine des entrepôts de données de santé (I2B2, Stride, KNAVE..)
- Une phase de conception qui prenne en compte cet état de l'art, les objectifs attendus et les contraintes techniques.
- Une phase de réalisation d'un prototype qui sera évalué ultérieurement sur des cas d'usage en comparaison avec l'interface actuelle

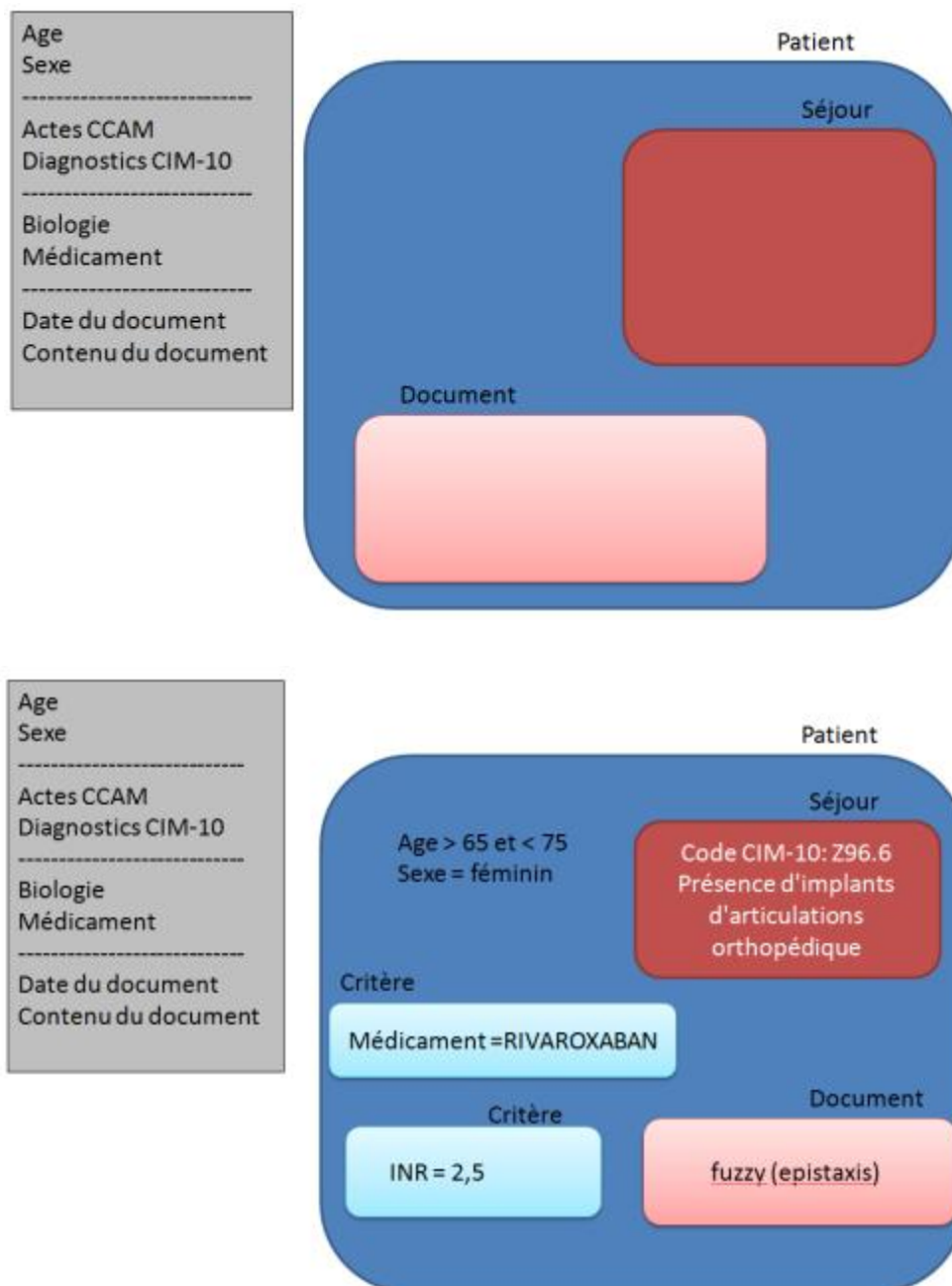
## **Quelques pistes**

Un module « Advanced Filter Builder » du type GWT (voir [http://www.smartclient.com/smartgwt/showcase/#featured\\_filter\\_builder\\_grid](http://www.smartclient.com/smartgwt/showcase/#featured_filter_builder_grid) ) nous paraît intéressant pour créer des filtres de sélection complexes. En revanche la liste des opérateurs doit être contextuelle (s'adapter au type de la donnée pour ne pas noyer l'utilisateur ni l'induire en erreur).



De même, la sélection des variables doit se faire en drag and drop. Enfin pour permettre de sélectionner des variables à différents niveaux de granularité (patient, séjour, document,..) un système de représentation ensembliste pourrait être utile.

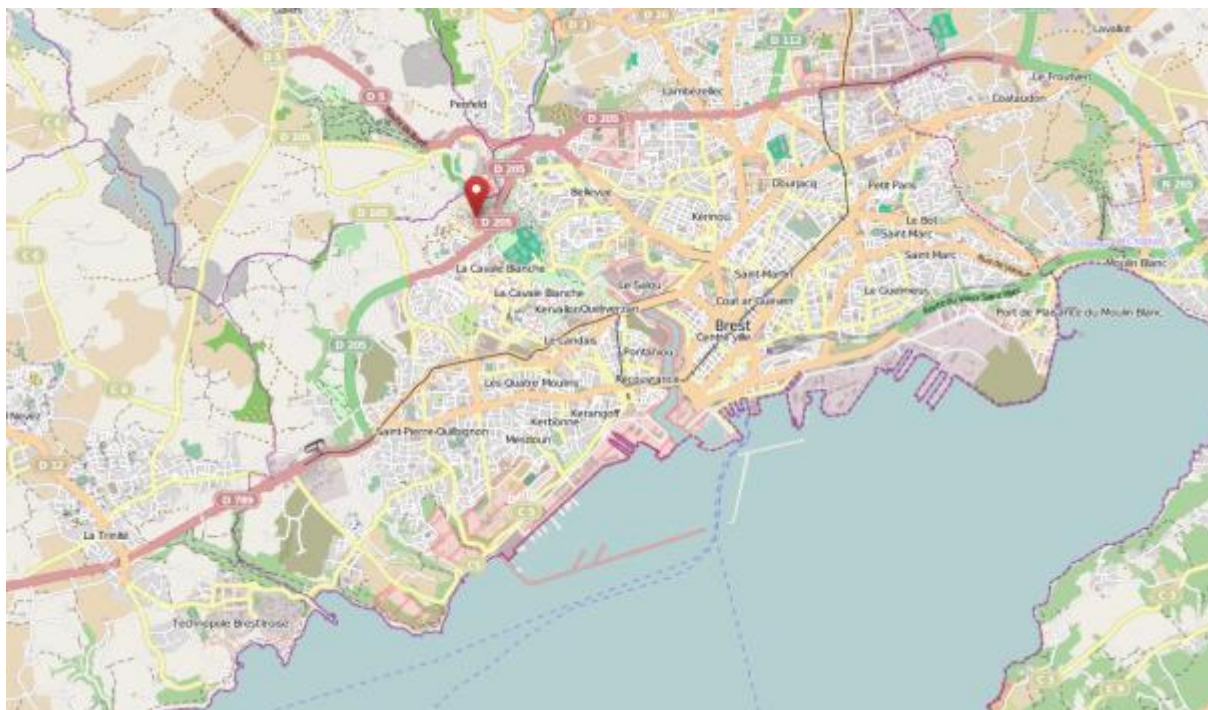
Un exemple d'interaction utilisant les diagrammes de Venn ( <http://creately.com/creately-start>). On peut imaginer que chaque niveau (le patient, le séjour, le document,..) soit représenté de manière graphique sous la forme d'un diagramme de Venn. Les variables de sélection sont « drag and droppées » sur le bon niveau de sélection (ex : age uniquement possible sur le patient, recherche plein texte uniquement sur le document, etc..) . Une fois le critère en place on peut le paramétrer avec le module « Advanced Filter Builder », on peut associer plusieurs critères sur un même niveau et les combiner. Une maquette est présentée ci-dessous.



De même il est possible de rajouter par exemple un deuxième document et lui appliquer des critères de sélection qui seront appliqués. Les différents éléments de sélection peuvent enfin être positionnés relativement les uns aux autres sur une échelle de temps en utilisant l'algèbre de Allen, exemple : l'INR est mesurée x heures après la délivrance du médicament. Un papier examine notamment ces aspects ici : « Visualizing Queries on Databases of Temporal Histories: New Metaphors and their Evaluation » par Luca Chittaro et Carlo Combi.

## **Modalités pratiques et contacts**

Le stage se déroulera au Centre Hospitalier Régional et Universitaire de Brest sur le site de la « [Cavale Blanche](#) » au sein du Département d'Information Médicale dirigé par le Dr Jean-Michel CAUVIN.



L'encadrement sera assuré sur place par :

- Clara LE GUILLOU [clara.leguillou@chu-brest.fr](mailto:clara.leguillou@chu-brest.fr)
- André HAPPE [andre.happe@chu-brest.fr](mailto:andre.happe@chu-brest.fr)

La gratification légale obligatoire sera versée au stagiaire, soit 436,05 €/mois pour un temps complet (35 heures par semaine).

Tout complément d'information peut-être obtenu en appelant au 02.98.34.75.56

Annexes

Interface actuelle

Formulaire de recherche : Intégralité de l'entrepôt (14660 patients)

+ Requetes enregistrees

- Criteres generaux

Date du document(JJ/MM/AAAA):

Unite(s) Fonctionnelle(s):

Age du patient pdt la consultation:

Age du patient maintenant:

30

80

Sexe :

H☒

F☐

Ne pas prendre en compte les patients decedes a l'hopital :☐

Le nombre de patients correspondants aux criteres generaux est de : 1072.

- Recherche en Full-text

ExclureType de doc : texteUF :

affections endocriniennes

Indure periphrases negatives :☐

Enrichissement semantique :☒

Pour cette expression : 865 patients

IndureType de doc : texteUF :

insuffisance renae

Indure penphrases negatives :☐

Enrichissement semantique :☒

Pour cette expression : 670 patients

ajouter requete

RequeteStatistiquesMap

Utilisez le formulaire a gauche pour creer votre requete

Age du patient : de 30 a 80

Sexe : M

Exclure les patients dont Les documents (avec enrichissement semantique) contenant l'expression "affections endocriniennes" X

Les documents (avec enrichissement semantique) contenant l'expression "insuffisance renae" X

+ Ajouter une contrainte temporelle sur les items ci-dessus

Formulaire de recherche : Intégralité de l'entrepôt (14660 patients)

+ Requetes enregistrees

+ Criteres generaux

+ Recherche en Full-text

- Recherche structuree

Filter : liste concepts - arbre

Actes

Diagnostics

Genes

+ agents antitumoraux Ajouter

+ encephalopathies Ajouter

+ affection Ajouter

- gene Ajouter

- structural gene Ajouter

- cytoskeletal gene Ajouter

> XLIS (3 patients) Ajouter

Les valeurs possibles : Toutes

IndureExclureToutes

MUTE

+ proto-oncogene, signaling f

+ enzyme gene Ajouter

+ ligand binding protein gene

+ transcription regulation gene

+ maladies neurologiques Ajouter

+ tissu Ajouter

> CYTOCHROME B (1 patients) Ajouter

> AP151 (1 patients) Ajouter

> AP152 (2 patients) Ajouter

> CASK (3 patients) Ajouter

> COL4A1 (1 patients) Ajouter

> DDX11 (1 patients) Ajouter

Laboratoire

+ ACMH ANTI TNF Ajouter

+ HEMOGLOBINE/OXYMETRIE Ajouter

+ MARQUEURS D ACTIVATION DE LA COAGULATION Ajouter

+ GAZ VEINEUX Ajouter

+ VENTILATION: Ajouter

+ BACTERIO-LCR Ajouter

+ BILAN MARTIAL Ajouter

+ THROMBOSE Ajouter

+ TYPAGE CD3 CD4 CD8 Ajouter

+ CHIMIE DE PONCTION Ajouter

- CHIMIE DES URINES Ajouter

> Hemoglobine A1C (8 patients) Ajouter

> Rapport ualb/proteines (4 patients) Ajouter

> Rapport Albumine/Creat. (4 patients) Ajouter

> PH Cristallurie (2 patients) Ajouter

> Densite corrigee/ pH (2 patients) Ajouter

> Nombre d'agregats (1 patients) Ajouter

> Taille moyenne des agregats (1 patients) Ajouter

> Taille maximale des agregats (1 patients) Ajouter

> Rapport Nbre d'agregats/ Nbre de cristaux (1 patients) Ajouter

> Creatinine enzymatique (23 patients) Ajouter

- Serum/Plasma Ajouter

> Alanine aminotransferase (34 patients) Ajouter

> Albumine (13 patients) Ajouter

Choisir un interval ou une valeur hors interval

Interval : de à :

Ou X fois la normale fois la normale

superieure : sup

Ou Valeur hors-interval : ☐

IndureExclureAnnuler

Les documents contenant l'expression "atrophie cereb%" X

Les documents contenant l'expression "atrophie optique" X

Les documents contenant l'expression "IRM" X

Absence de diabete sucre X

+ Ajouter une contrainte temporelle sur les items ci-dessus

l'evenement Les documents contenant l'expression "IRM"

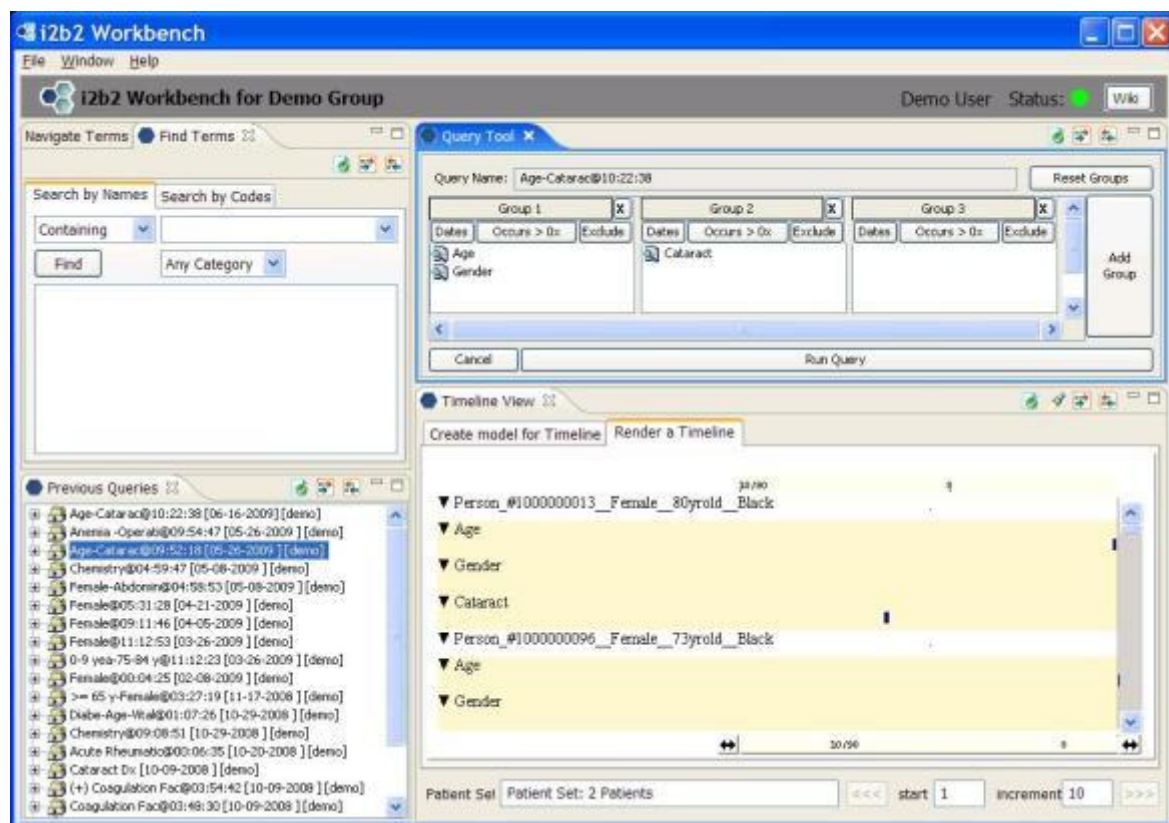
Precede de Maximum 2 mois

l'evenement Les documents contenant l'expression "atrophie cereb%"

ajouter annuler

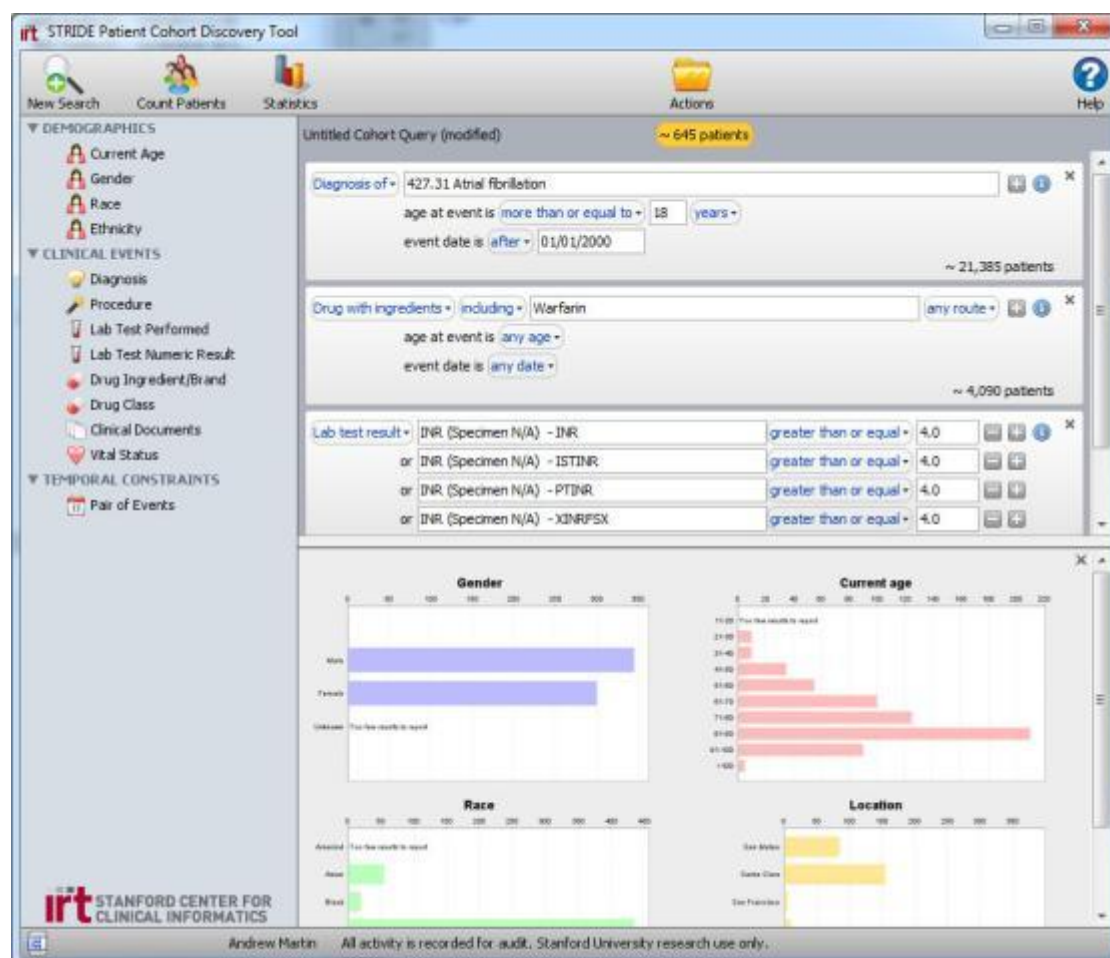


I2B2 (Informatics for Integrating Biology and the Bedside) voir <http://www.i2b2.org>



STRIDE (Stanford Translational Research Integrated Database Environment) voir

(<https://clinicalinformatics.stanford.edu/services/index.html>)



KNAVE (Knowledge-based Navigation of Abstractions for Visualization and Explanation) voir ([http://medinfo.ise.bgu.ac.il/medLab/ResearchProjects/RP\\_KNAVE.htm](http://medinfo.ise.bgu.ac.il/medLab/ResearchProjects/RP_KNAVE.htm))

