

**DONNÉES DE SANTÉ**

# Intelligence artificielle : le projet RUBY

RUBY EST UNE APPLICATION INFORMATIQUE DÉVELOPPÉE PAR LE DÉPARTEMENT D'ÉPIDÉMIOLOGIE, DE BIostatistique ET DES DONNÉES DE SANTÉ (DEBDS) QUI A POUR BUT DE STRUCTURER LES DONNÉES DE SANTÉ DES PATIENTS AFIN DE POUVOIR LES UTILISER POUR LA RECHERCHE BIOMÉDICALE.

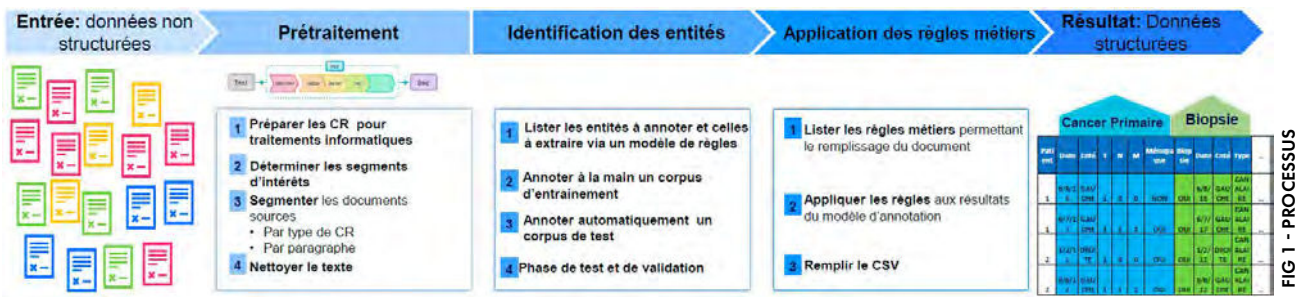


FIG 1 - PROCESSUS

Le Centre Antoine Lacassagne dispose d'un grand volume de données patients informatisées. Malheureusement, la majorité de ces données (80%) sont dans des comptes rendus sous forme de textes non structurés et donc non exploitables. La structuration des données dans des bases de données est très chronophage et doit se faire manuellement par des assistant(e)s de recherche clinique. Chaque dossier médical nécessite au moins une heure de travail pour être saisi dans une base de données. Pour les 400 000 patients qui ont été pris en charge au Centre Antoine Lacassagne depuis 25 ans, cela représenterait 156 ans de travail !

Une structuration automatique des données des patients n'avait encore jamais été réalisée au Centre, ni en France à notre connaissance, ce pourquoi, au DEBDS nous nous sommes tournés vers IBM pour nous accompagner dans la réalisation de ce projet qui fait partie intégrante du projet d'établissement 2020-2024.

La première tentative de structuration automatique a été réalisée sur les données de patientes atteintes de cancer du sein. Pour ce projet nommé RUBY, nous avons « appris » à un algorithme d'intelligence artificielle de deep learning (nommé Réseaux de neurones à convolution) à reconnaître les éléments de langage d'une cinquantaine de données de 7000 comptes rendus médicaux afin qu'il puisse les traduire automatiquement en données structurées (cf figure ci-dessus).

Les résultats sont très encourageants puisque RUBY est plus précis que la saisie manuelle dans près d'un cas sur 2. Mais surtout, RUBY est beaucoup plus rapide, il lui faut une heure pour structurer 5000 comptes rendus de chirurgie et 10 000 comptes rendus d'anatomopathologie là où il aurait fallu un an de saisie manuelle.

Nous avons pour ambition d'étendre RUBY à toutes les données de la prise en charge du cancer du sein

et, en collaboration avec le Dr Dorian Culié en thèse d'Université sur le sujet, nous allons reproduire cette expérience sur les données des patients traités pour des tumeurs de la thyroïde puis les données des patients pris en charge pour les autres types de cancers ORL.

Nous comptons également améliorer les performances de RUBY en renforçant l'entraînement des algorithmes puis les intégrer dans un processus totalement automatisé de création d'une véritable plateforme de données de santé qui regroupera toutes les données informatisées des patients du Centre.

**Renaud SCHIAPPA**  
Responsable de l'unité de Data management du DEBDS

**Dr Dorian CULIÉ**  
Chirurgien ORL

**Dr Emmanuel CHAMOREY**  
Chef du DEBDS