



## Détection des doublons patients et identification des professionnels

F Laborde, A Dugué, E Pinon, L Digue, V Bousser



## Contexte

Les Réseaux Régionaux de Cancérologie (RRC) d'Aquitaine (RCA), du Limousin (ROHLim) et de Poitou-Charentes (Onco-Poitou-Charentes) sont en cours de mutualisation de leur Dossier Communicant de Cancérologie (DCC) vers le nouvel outil de la Nouvelle-Aquitaine K-Process.

Les informations des **11 bases de données** de l'outil historique du RCA déployé depuis 2003, le Dossier Patient Réseau de Cancérologie (DPRC), devaient être migrées.

**L'objectif était de fusionner les doublons d'identités patients et d'identifier les professionnels de santé pour opérer une migration de qualité vers K-Process.**

**1** Chaque base de données a fait l'objet d'un **export** des identités patient, données structurées, identités professionnels et documents.

## Détection algorithmique

**3** La recherche de rapprochements potentiels d'identités utilise des méthodes différentes selon qu'il s'agisse de patients ou de professionnels

**Identités patients : 11 méthodes** de détection de rapprochements d'identités se basant sur un réglage des seuils de similarité à partir des éléments de l'identité patient

Nom d'usage / Nom de naissance / Prénom / Ville / Code Postal / Date de naissance / Sexe / Lieu de résidence

**Identités professionnels : 6 méthodes** de détection de rapprochements d'identités entre les professionnels DPRC et l'annuaire national RPPS ([annuaire.sante.fr](http://annuaire.sante.fr)) en se basant sur :

Nom / Prénom / Lieu d'exercice

### Calcul des similarités

- Noms, prénoms, date de naissance : algorithme de **distance Jaro-Winkler** après **normalisation** des chaînes de caractères
- Lieu de résidence ou d'exercice : **géocodage** via la *Base Adresse Nationale* ([adresse.data.gouv.fr](http://adresse.data.gouv.fr)) puis fonction de **haversine** pour la distance en km avec score en % (100% étant le cas de 2 patients résidant <de 1500m l'un de l'autre)

## Volumétrie

**5** Sur 165 000 identités patients

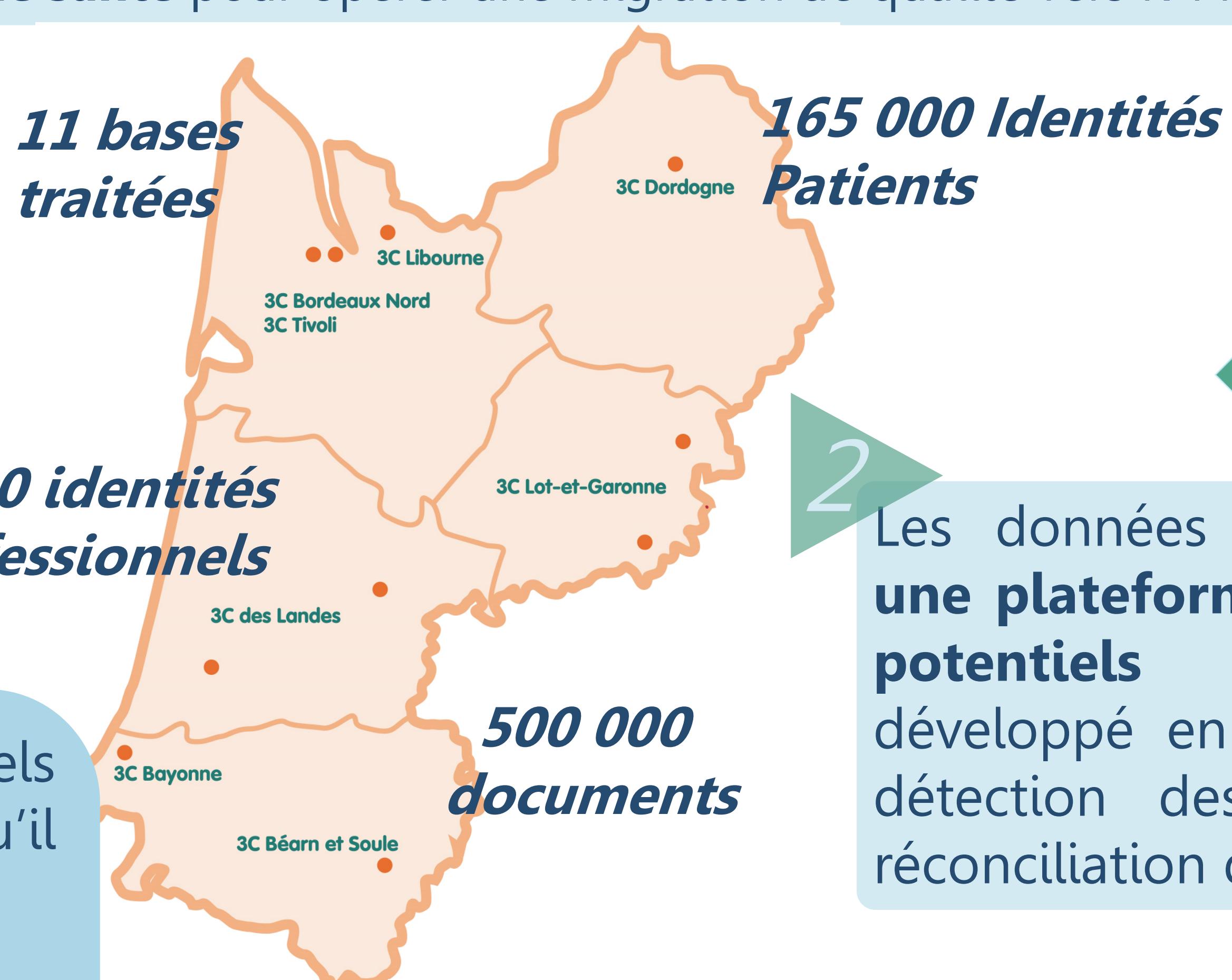
- 8000 (4,8%) rapprochements potentiels par l'algorithme
- 5500 (3,3%) invalidations manuelles
- 2500 (1,5%), soumis au second tour de traitement

→ **1300 (0,78%) rapprochements effectués**, 300 invalidés et 900 restent à traiter

→ **Au final, estimation de 1,3% de rapprochements d'identités**

**Sur 9500 identités professionnels**

- **3400 (36%) rapprochés automatiquement** par le logiciel
- **3600 (38%) nécessitant une action de contrôle ou de recherche manuelle**
- **2500 (26%) non retrouvées** (*déménagement ou cessation d'activité*)



## Données

**2** Les données exportées soumises à une **plateforme de rapprochements potentiels d'identités** (outil développé en interne) en vue de la détection des doublons et de la réconciliation d'identités

**Exemple** d'ajustement des seuils de similarité pour détecter les rapprochements pour les **femmes ayant changé de nom** (mariage ou divorce)

- **Seuil de similarité entre les noms d'usage et/ou noms de naissance > 40%**
- **Prénoms > 40%**
- **Similarité entre les lieux de résidence > 90%**
- **Date de naissances et sexes : 100%**

## Instruction des cas

4

La plateforme de rapprochements a permis de **prétraiiter** les identités **patients** en **invalidant les rapprochements** (cas manifestes de rapprochements concernant 2 identités patients distinctes à ne pas fusionner).

Un **second tour** de traitement a consisté à la **lecture des fiches RCP** ou à un retour aux **dossiers patients** (par RCA ou équipes 3C).

L'algorithme a validé **automatiquement les rapprochements professionnels** en cas de concordance parfaite des traits d'identités des professionnels avec le RPPS.

## Une base « propre »

### Un souhait d'exhaustivité :

Le RCA a fait le choix d'un nombre relativement élevé de rapprochements proposés par l'algorithme, ceci afin de maximiser les chances de repérage de doublons.

### Un gage de confiance :

Le rapprochement des identités patients a permis de repérer de nombreux cas qui seraient restés doublonnés avec une stratégie de comparaison stricte de traits d'identités.

La consolidation des identités professionnels a permis de se séparer des coordonnées obsolètes.

Ce travail permet une migration du DPRC vers K-Process de qualité visant à créer un dossier réellement partagé et opérationnel.