

GIS PAROLOTHEQUE :

Une plateforme de recherche sur des données médicales de parole

10eme journée du réseau Cogiter

« Infrastructure pour l'expérimentation en ligne, la science participative/collaborative ou la recherche inter-disciplinaire par les données. »

Jérôme Farinas

4 décembre 2017



Parolothèque

- Structure en création sur Toulouse par Pr Virginie Woisard : Groupement Intérêt Scientifique
- Base de données d'enregistrements vocaux sur le modèle de la thumorotheque
- La parole en cancérologie peut être considérée comme un échantillon biographique pouvant être conservée dans une "parolothèque"
- Ne pas perdre ces données et de les mettre à la disposition de la communauté scientifique

GIS

- Groupement d'Interet Scientifique (ou Contrat de Programme de Recherche)
- Un GIS est matérialisé par une convention constitutive dont le préambule décrit :
 - le cadre scientifique de ce GIS,
 - les raisons et motifs qui ont conduit les partenaires à collaborer
 - leurs compétences et expériences respectives

GIS : Objectifs

- Fédérer autour d'un projet commun de recherche, des compétences scientifiques et des moyens sur un thème donné
- Rendre visible une collaboration ou synergie entre parties partenaires, éventuellement pour plus facilement bénéficier d'un financement commun
- Souplesse et rapidité de réaction ou de projet

Description des traitements

- Stockage
 - Entretiens avec les patients
 - Enregistrements de voix altérées
- Traitement automatique du contenu de la parole à différents niveaux (acoustiques, linguistiques et métalinguistiques...)
 - Recherche en épidémiologie, en psychologie et en sociologie
 - Recherche en traitement automatique des langues

Epidémiologie

- Enregistrements de conversations avec les personnes atteintes de cancer
- Production automatisée de transcriptions textuelles
- Recherche sur l'ensemble de ces données de causes possibles de cancer

Traitements Automatique de la Parole

- Mesurer de manière objective l'intelligibilité et la compréhensibilité de patients traités pour des cancers de VADS (5^e cancers les plus fréquents en France)
- Traitements ont des conséquences sur la phonation et la déglutition avec un impact majeur sur la qualité de vie
- Production d'un indice de sévérité automatique

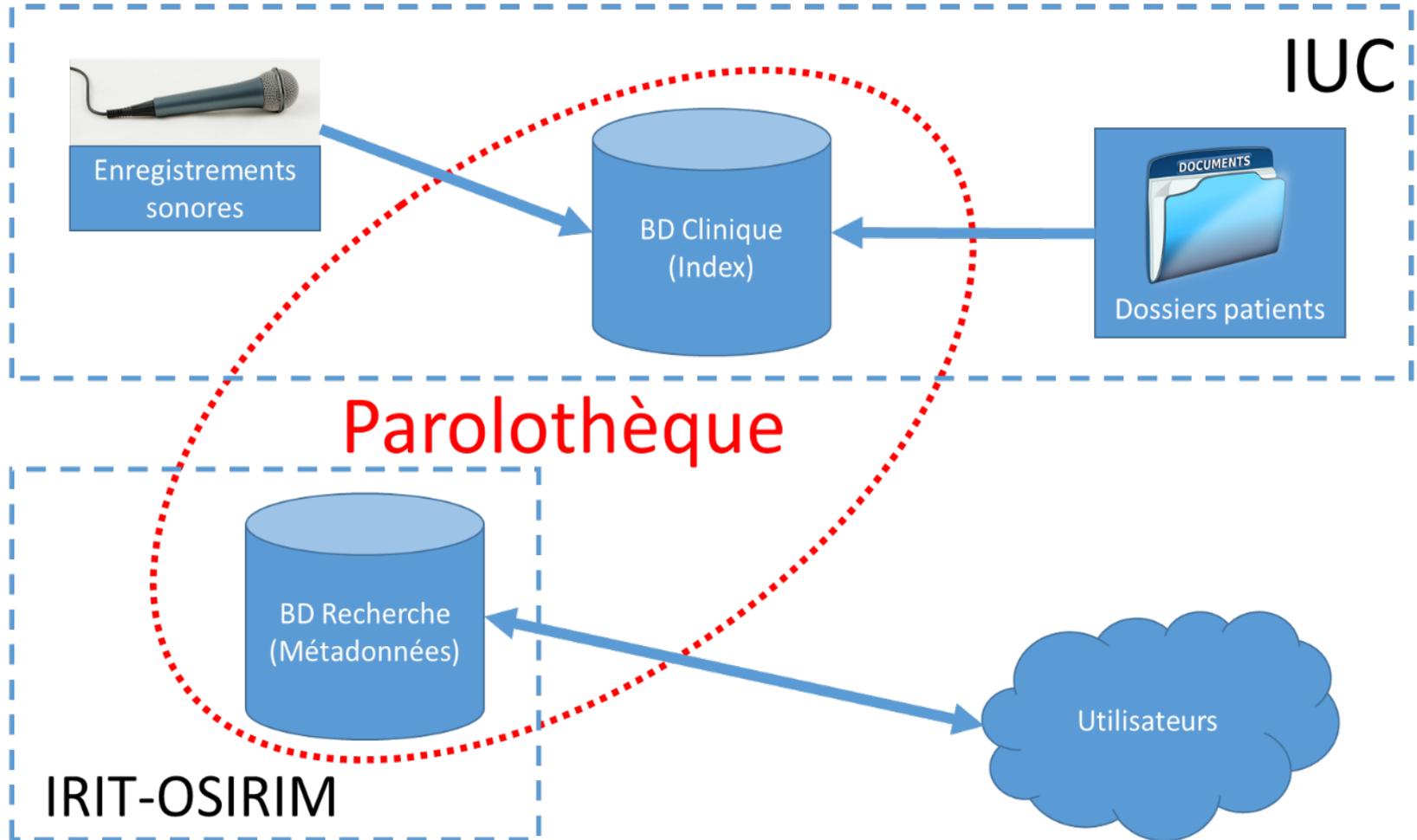
Projet C2SI

- Financement INCA 2014-2018
- Partenaires : CHU-T, UT2, UT3, UAPV, AMU
- Constitution d'un corpus de plus de 130 patients et témoins
- Mise en place d'une mesure Carcinologic Speech Severity Index par expériences perceptuelles
- Automatisation de la mesure

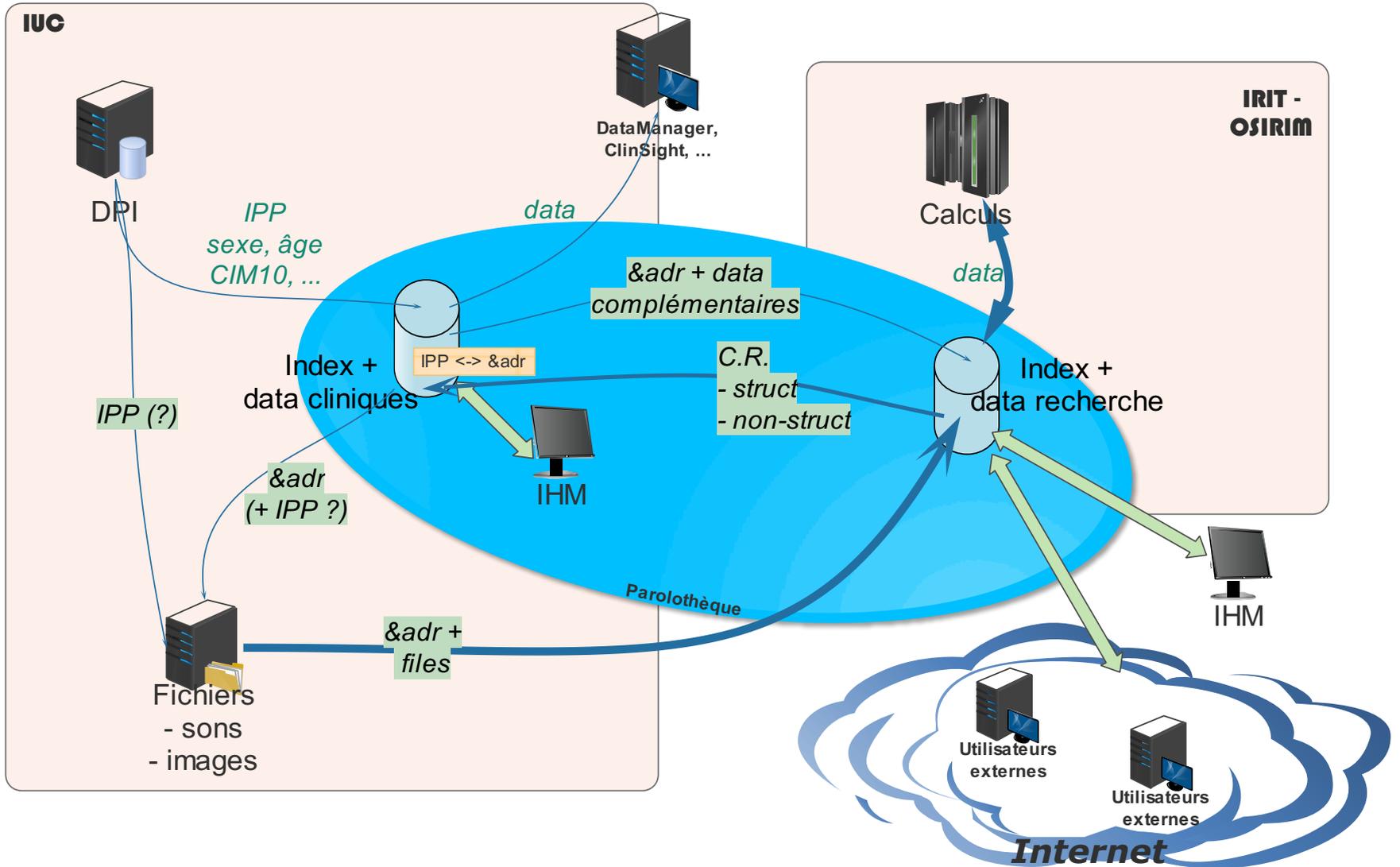
GIS : Fonctionnement

- Conception en 3 étapes :
 - Enregistrement de la parole et archivage,
 - Enrichissement et structuration de la base,
 - Traitement automatique des sources
- Analyse à toutes les étapes des aspects éthiques, juridiques et réglementaires
- Administration des projets de recherche et permettre l'accès aux données

Idée de départ



Infrastructure envisagée



OSIRIM

- Plateforme IRIT
 - Environnement pour l'indexation et la recherche d'informations dans les contenus multimedia (texte, son, image, vidéo, signal numérique)
 - Stockage : 1 Po
 - Calcul : cluster de 640 cœurs, 5 To RAM, 2x10 Gb, cluster 20 GPU GTX 1080 Ti
 - SLURM, Hadoop

Parolothèque : historique

- 2012 : idée de création de la parolothèque
- octobre 2013 : orientation vers la création d'un GIS
- 2017 : statuts du GIS parolothèque en cours de signature

- Partenaires initiaux : IRIT, IFERISS, OCTOGONE, CLLE, LIA, LPL, CHU-T

Défis

- étendre le champ disciplinaire à d'autres pathologies de la voix
- permettre l'accès à la recherche sur des données cliniques au niveau national, voire international
- faciliter l'exploitation de résultats recherche vers une exploitation clinique

Merci de votre attention. Des questions ?

- Contacts :
 - Pr Virginie Woisard, CHU Toulouse Woisard.Virginie@iuct-oncopole.fr
 - Pascal Gaillard, MSH-T Pascal.Gaillard@univ-tlse2.fr
 - Julien Pinguier, IRIT Julien.Pinguier@irit.fr
 - Jérôme Farinas, IRIT Jerome.Farinas@irit.fr
 - Alain Ghio, LPL alain.ghio@univ-amu.fr
 - Corinne Fredouille, LIA corinne.fredouille@univ-avignon.fr