



Standardisation de l'utilisation des données du SNDS à des fins de recherche médicale : présentation d'un modèle de données optimisé centré sur le patient

N.Thurin¹, R. Lassalle², J. Jové², R. Sylvestre², J. Dupont²,
C.Lacueille², P.Blin², N.Moore¹, C. Droz-Perroteau²

¹Bordeaux Pharmacoeconomics, INSERM CIC1401, INSERM U219, CHU de Bordeaux, Université de Bordeaux, Bordeaux, France

²Bordeaux Pharmacoeconomics, INSERM CIC1401, Université de Bordeaux, Bordeaux, France



Journées EMOIS 2019, 14-15 mars, Nancy, France



SNDS

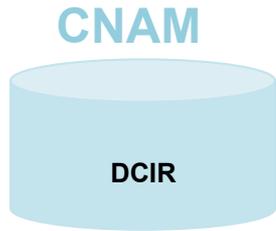
- Outil puissant et incontournable en pharmaco-épidémiologie
 - Analyse d'impact (bénéfices/risques) des médicaments en conditions réelles
 - Evaluation de parcours de soins
 - Suivi de l'évolution des maladies et des traitements
- Architecture centrée prestation / séjour → usage à des fins de recherche centrées sur le patient complexe

Architecture du SNDS

6 univers

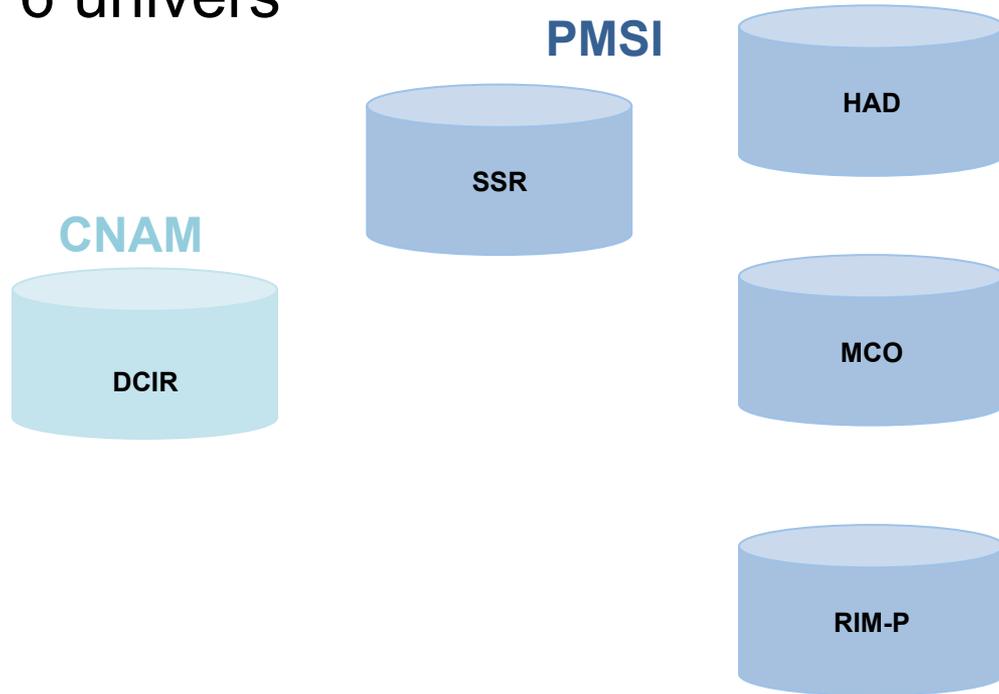
Architecture du SNDS

6 univers



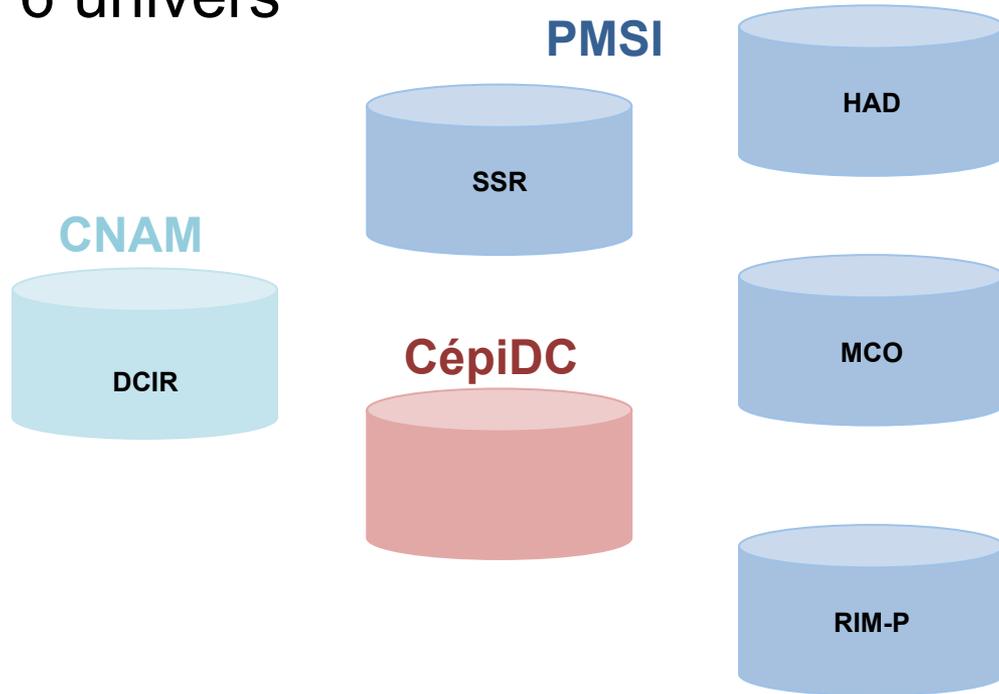
Architecture du SNDS

6 univers



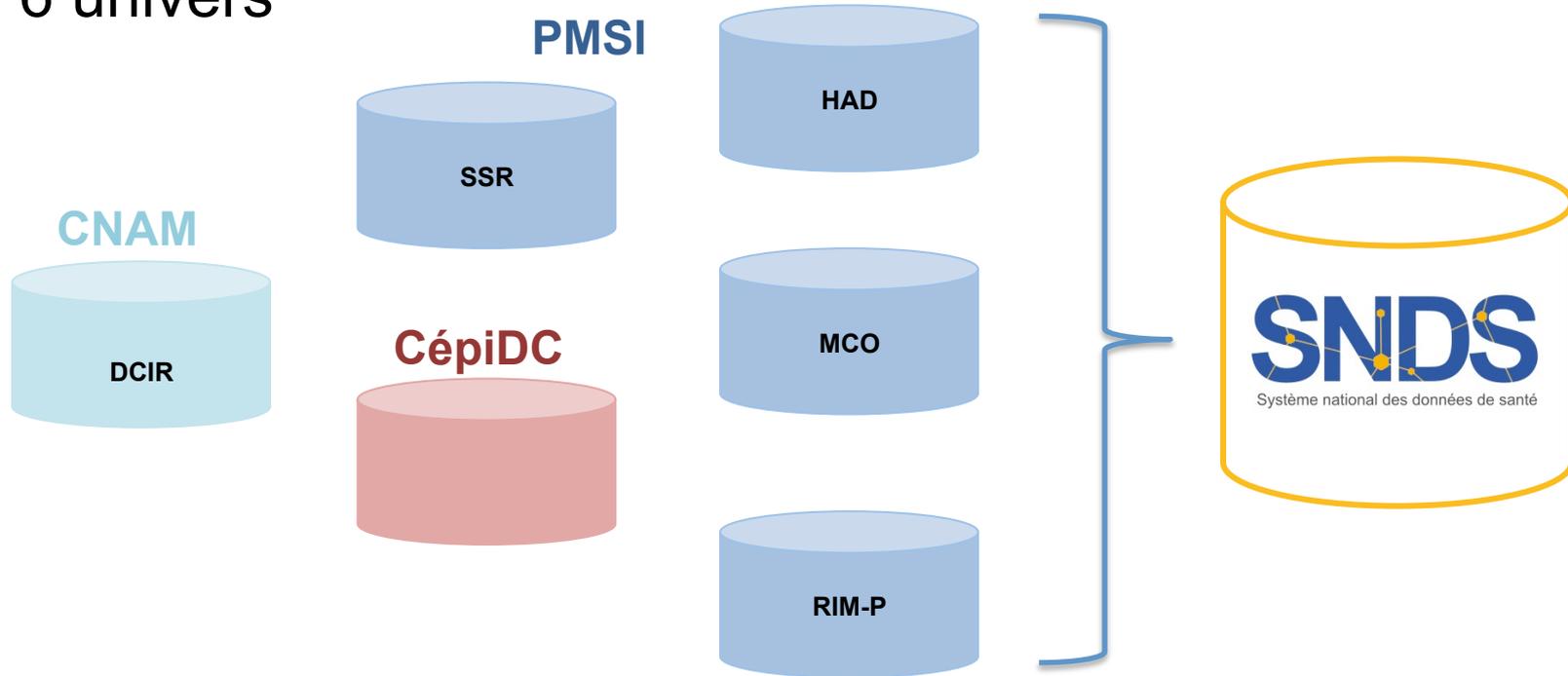
Architecture du SNDS

6 univers



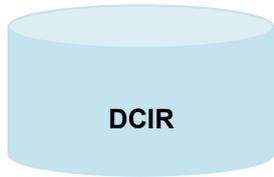
Architecture du SNDS

6 univers



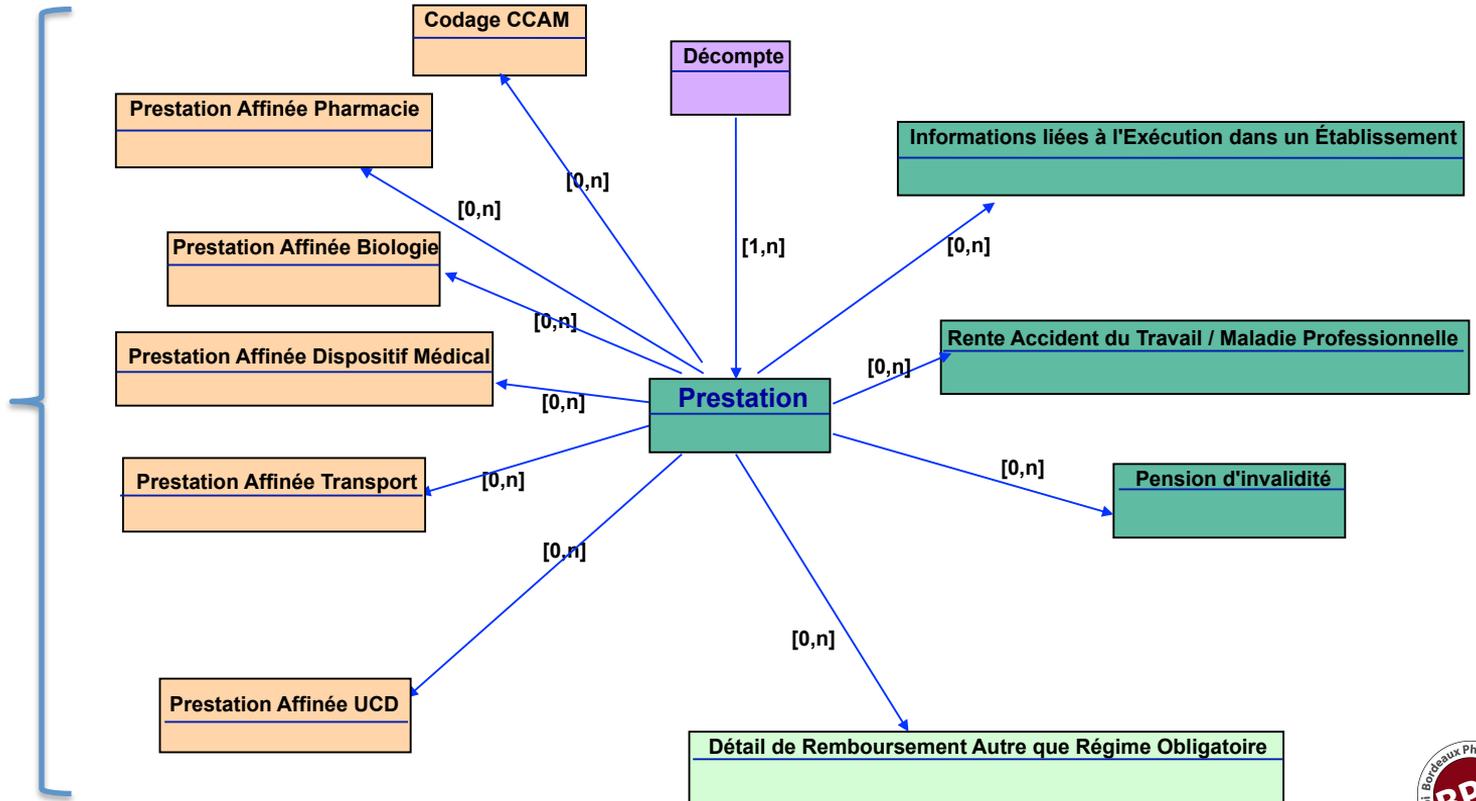
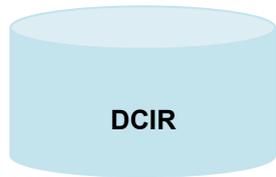
Architecture du SNDS

- Centrée sur la prestation



Architecture du SNDS

- Centrée sur la prestation



Difficultés d'exploitation pour la recherche

- Unité statistique = patient plutôt que prestation / séjour
- Nombre de tables annuelles et de variables
 - Très important
 - Croissant annuellement
- « Lourdeur » des traitements
 - Nombre élevé clés de jointure et jointures à effectuer
 - Volume de données
 - Temps machine

Volume annuel de tables / variables

| Univers | Nombre de tables/an ¹ | | Nombre de variables/an ¹ | |
|---------------------|----------------------------------|-----|-------------------------------------|-----|
| | Min | Max | Min | Max |
| DCIR | 12 | 13 | 375 | 394 |
| MCO | 19 | 54 | 342 | 859 |
| SSR | 14 | 27 | 260 | 519 |
| RIM-P | 6 | 16 | 92 | 361 |
| HAD | 6 | 22 | 114 | 437 |
| CépiDC ² | - | 2 | - | 61 |

¹ 2006 - 2016 ² 2013 - 2015



Ex : Nb total en 2015 = 134 tables et 2 631 variables

Objectifs

- Construire un modèle de données SNDS simplifié et optimisé pour conduire des études de pharmaco-épidémiologie
 - Centré autour du patient
 - Nombre de tables restreint (domaine)
 - Sélection des variables nécessaires

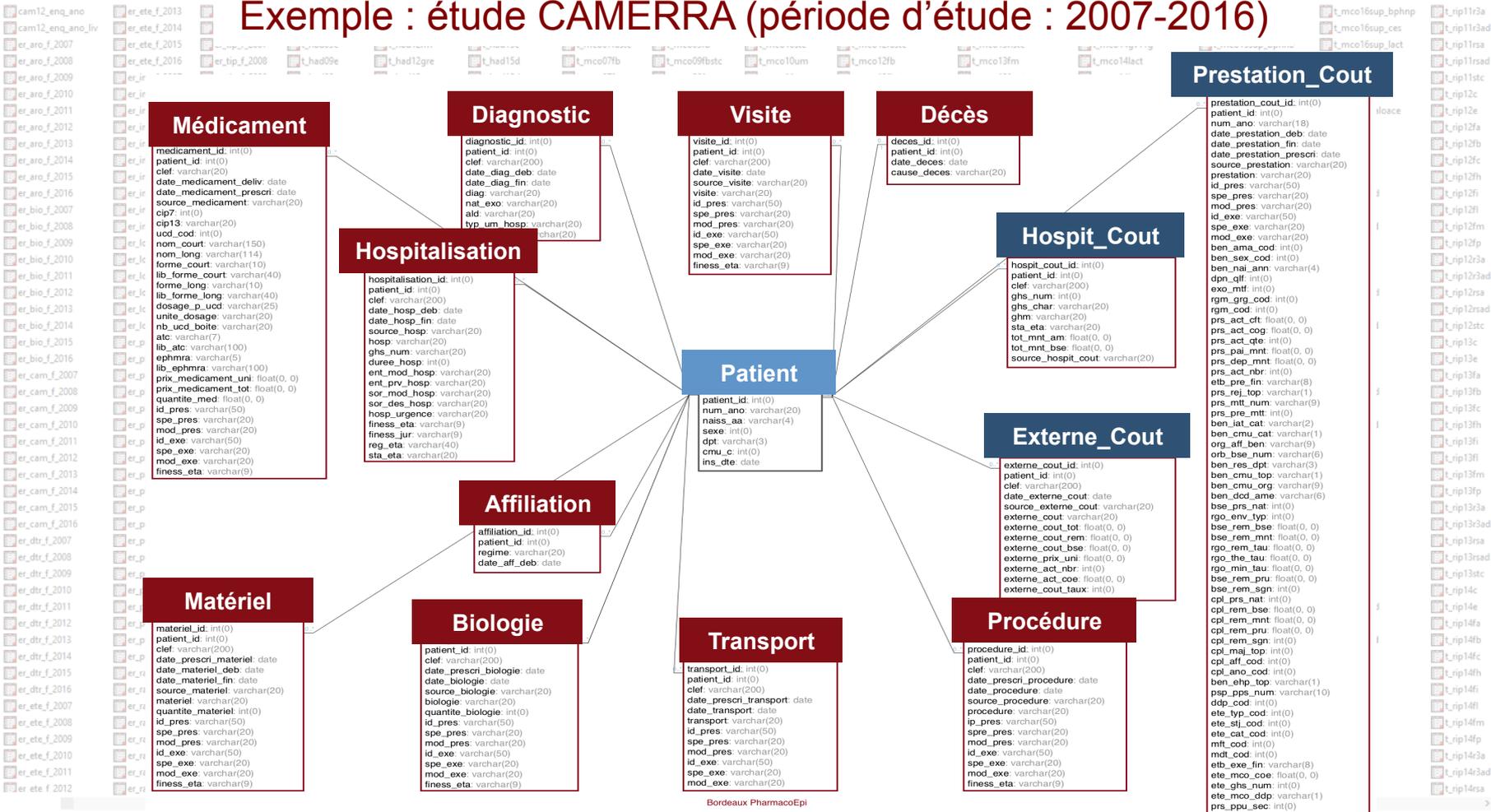
Démarche

- Définition du **modèle de données relationnel** par une équipe pluridisciplinaire
- Rédaction du cahier des charges
 - Sélection des variables pertinentes
 - Description des tables
 - **Cartographie des variables** sources vers le modèle final
 - Création du thésaurus associé
- **Programmation** et exécution sur données « test » (SAS®)
- **Validation** des programmes et des tables créées

Résultats

- Modèle de données relationnel simplifié
 - Clé de jointure numérique : ID_Patient
 - Structure des données définie avec 14 tables :
 - *Patient*
 - *Affiliation*
 - *Décès*
 - *Hospitalisation*
 - *Diagnostic*
 - *Médicament*
 - *Visite*
 - *Procédure*
 - *Biologie*
 - *Matériel*
 - *Transport*
 - *Prestation_cout*
 - *Hospit_cout*
 - *Externe_cout*

Exemple : étude CAMERRA (période d'étude : 2007-2016)



Résultats

- Structure des données : exemple Table *Diagnostic*

| Nom Variable | Libellé Variable | Type Variable | Format | Précision |
|---------------|--|---------------|--------|-------------------------|
| ID | Identifiant du diagnostic (clé primaire) | number | 11. | |
| PATIENT_ID | Identifiant du patient (clé étrangère) | number | 11. | |
| CLEF | Clef (pour relier avec table HOSPITALISATION, HOSPIT_COUT et autres données hospitalières) | char | \$200. | |
| DATE_DIAG_DEB | Date de début (séjour hospitalier / mise en ALD ou invalidité ou AT ou MP) | date | 8. | |
| DATE_DIAG_FIN | Date de fin (séjour hospitalier / mise en ALD ou invalidité ou AT ou MP) | date | 8. | |
| DIAG | Code CIM10 Diagnostic hospitalier / ALD / invalidité / AT / MP | char | \$20. | Vocabulaire CIM10 |
| NAT_EXO | Nature d'exonération | char | \$20. | Vocabulaire NAT_EXO |
| ALD | Code ALD | char | \$20. | Vocabulaire TYP_ALD |
| TYP_UM_HOSP | Type d'autorisation d'unité médicale | char | \$20. | Vocabulaire TYP_UM |
| SOURCE_DIAG | Source du diagnostic | char | \$20. | Vocabulaire SOURCE_DIAG |

- Données de 13 tables (*× nombre années*) des 6 univers synthétisées en 1 seule table et un nombre limité de variables
- 1 seule variable code CIM10

Résultats

- Pour chaque table de données
 - Documentation source initiale de la donnée
 - Calcul de variables dérivées clés : passage urgences, soins intensifs, *etc.*
 - Possibilité de relier les données d'un même univers présentes dans différentes tables (variable CLEF)
- Table *Médicaments* : liaison avec thésaurus IR_PHA_R
- Ajout des coûts hospitaliers environnés annuels (source : ENC ATIH)

Application du modèle simplifié

Exemple : étude CAMERRA (cohorte 2012 - N = 913 358 sujets)

| | Modèle initial | Modèle simplifié |
|------------------------|----------------|--|
| Nombre de tables | 973 | 14 |
| Volume des données SAS | 473 Go | 389 Go (82%) 52 Go (11%, sans la table Prestation_cout) |

Temps d'exécution : 5 jours

Conclusion / Perspectives

- Optimisation du travail des statisticiens
 - Standardisation des tables
 - Diminution de la variabilité inter-programmeurs
 - Gain notable en terme de temps
 - Meilleure reproductibilité des résultats et qualité scientifique
- Optimisation de l'exploitation des données
 - Utilisation simplifiée d'autres outils (macro HdPS...)
 - Simplification présentation structure des données
 - Programmes de recherche internationaux multi-pays multi-bases de données : interconnexion optimisée avec plateformes existantes
Exemples : Aetion, Pharmo, OMOP-OHDSI, etc.

Remerciements

- Equipe projet BPE :
 - Marie-Agnès Bernard – *Statisticienne référente*
 - Jérôme Dupont – *Data manager*
 - Jérémy Jové – *Statisticien référent*
 - Clémentine Lacueille – *Statisticienne-programmeuse*
 - Romain Sylvestre – *Data manager*
 - Nicolas Thurin – *Assistant médico-scientifique*



Merci pour votre attention

regis.lassalle@u-bordeaux.fr

Bordeaux
BPE
PharmacoEpi

Bordeaux PharmacoEpi - <http://www.pharmacoepi.eu>
Plateforme de recherche en Pharmaco-épidémiologie

CIC Bordeaux CIC1401

INSERM - Université de BORDEAUX - CHU de Bordeaux - Adera

Bâtiment Le Tondu - case 41 - 146 rue Léo Saignat - 33076 Bordeaux Cedex

Acc. +33 (0)5 57 57 46 75 – Fax +33 (0)5 57 57 47 40