



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

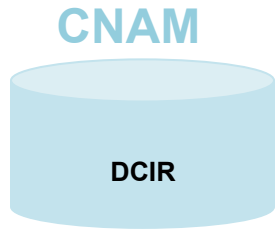
- Outil puissant et incontournable en pharmaco-épidémiologie
 - Analyse d'impact (bénéfices/risques) des médicaments en conditions réelles
 - Evaluation de parcours de soins
 - Suivi de l'évolution des maladies et des traitements
- Architecture centrée prestation / séjour → usage à des fins de recherche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

Architecture du SNDS

6 univers

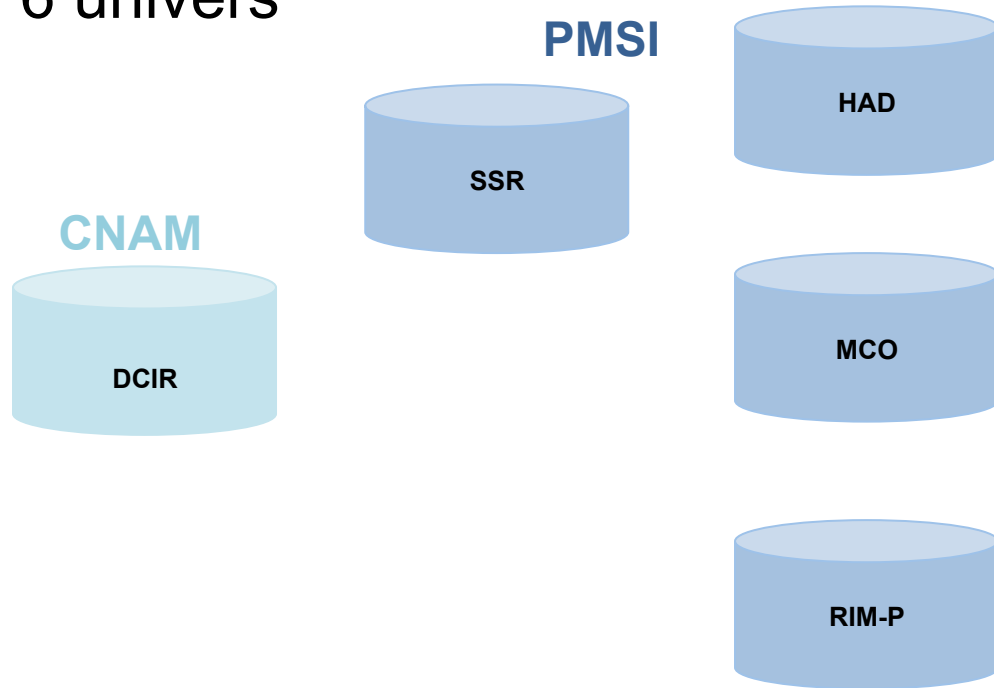
Architecture du SNDS

6 univers



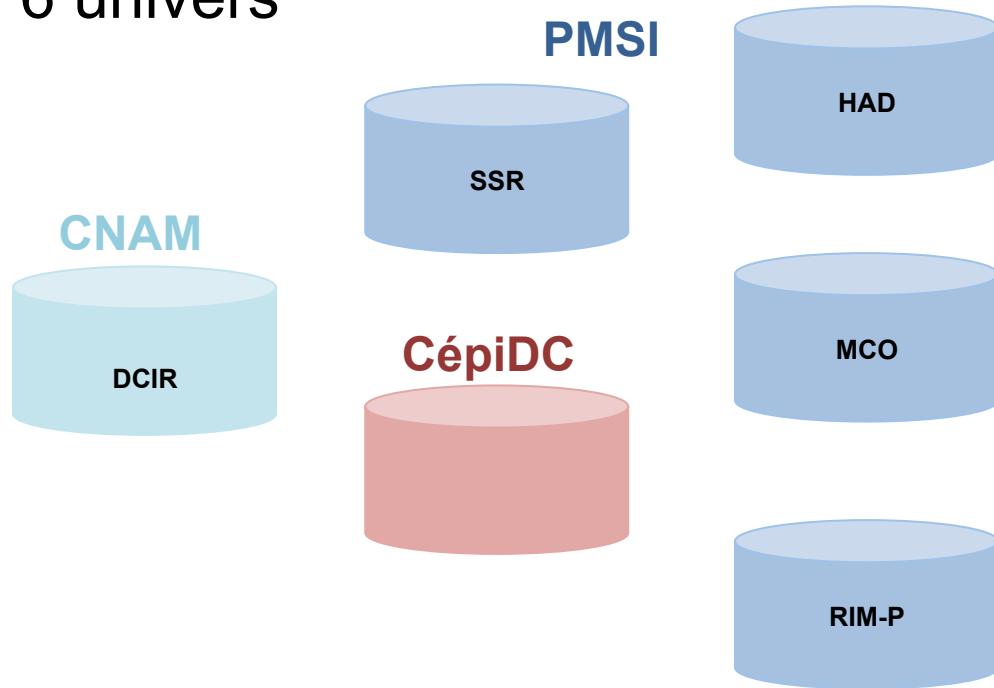
Architecture du SNDS

6 univers



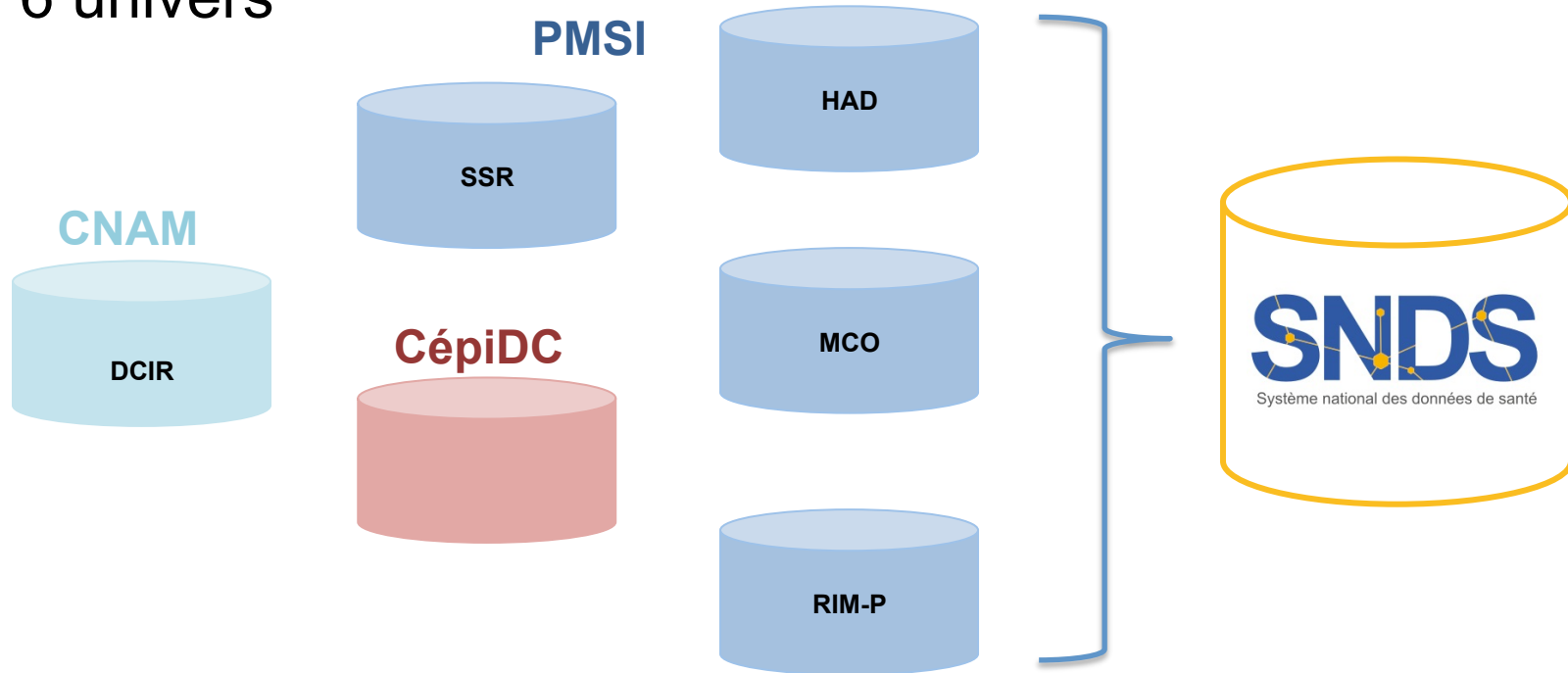
Architecture du SNDS

6 univers



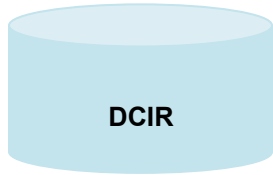
Architecture du SNDS

6 univers



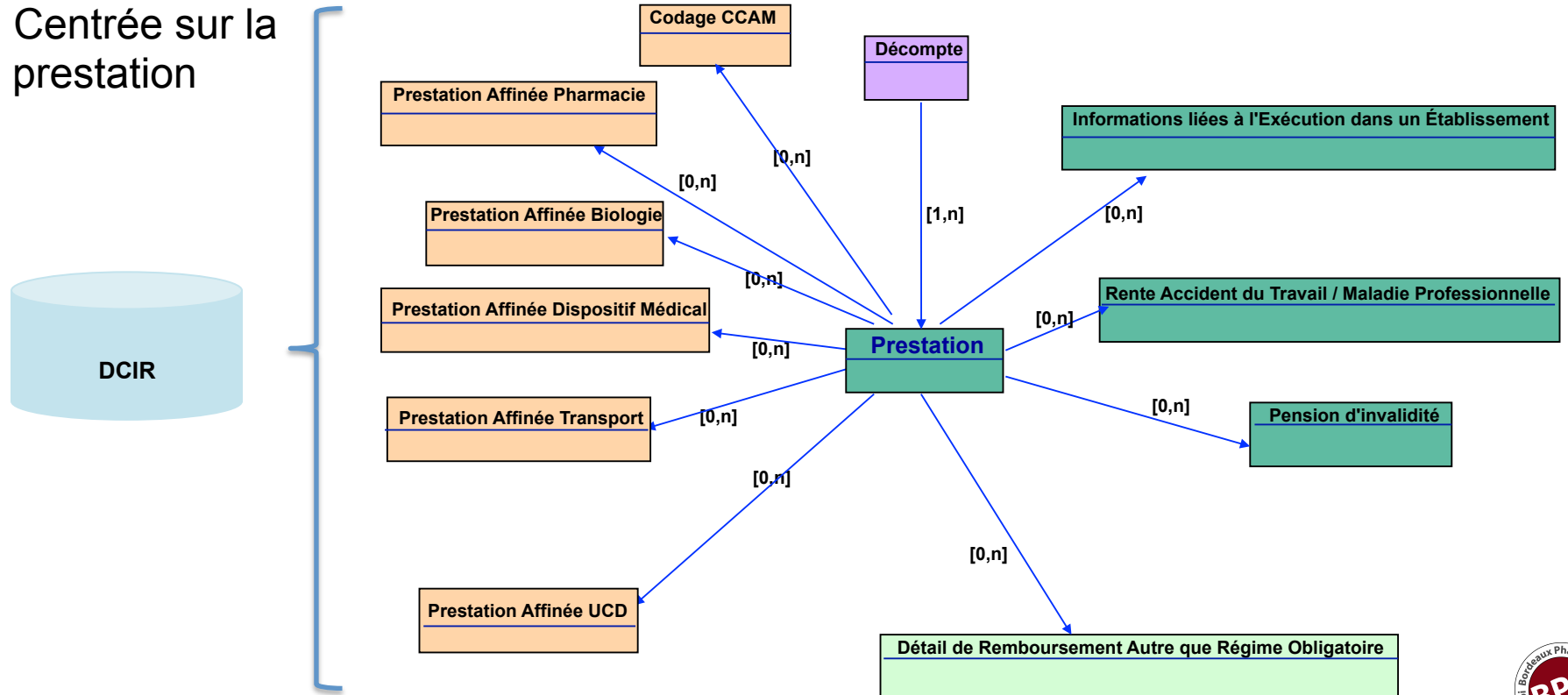
Architecture du SNDS

- 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

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

cam12_enq_ano	er_ete_f_2013	t_had12grc	t_had15d	t_mco07fb	t_mco09fbstc	t_mco10um	t_mco12fb	t_mco13fm	t_mco14act	t_mco15sup_bphnc	t_mco16sup_bphnc	t_mco16sup_bphnp	t_rip11r3a
cam12_enq_ano_liv	er_ete_f_2014	t_had12gre	t_had15dmpp	t_mco07fc	t_mco09fct	t_mco11a	t_mco12fbstc	t_mco13fmcstc	t_mco14ale	t_mco15sup_bphnc	t_mco16sup_ces	t_mco16sup_lact	t_rip11r3ad
er_aro_f_2007	er_ete_f_2015	t_had09e	t_had12med	t_mco07fh	t_mco09fhct	t_mco11b	t_mco12fc	t_mco13fmstc	t_mco14med	t_mco15sup_fffm	t_mco16sup_lect	t_mco16sup_lum	t_rip11rsad
er_aro_f_2008	er_ete_f_2016	t_had09med	t_had12medatu	t_mco07fi	t_mco09fhct	t_mco11bphna	t_mco12fcstc	t_mco13fmcstc	t_mco14medatu	t_mco15sup_fffm	t_mco16sup_prg	t_mco16sup_lum	t_rip11r3ad
er_aro_f_2009	er_tip_f_2008	t_had09stc	t_had12stc	t_mco07fm	t_mco09fhct	t_mco11bphnc	t_mco12fh	t_mco13gv11e	t_mco14medthrombo	t_mco15sup_lact	t_mco16sup_fffm	t_mco16sup_lum	t_rip11rsad
er_aro_f_2010	er_tip_f_2009	t_had09grc	t_had12grc	t_mco07ff	t_mco09fct	t_mco11bphnp	t_mco12fhstc	t_mco13gv11f	t_mco14or	t_mco15sup_fffm	t_mco16sup_lum	t_mco16sup_prg	t_rip11stc
er_aro_f_2011	er_tip_f_2010	t_had09grp	t_had12mz	t_mco07fctc	t_mco09fctc	t_mco11c	t_mco12fci	t_mco13alact	t_mco14pie	t_mco15sup_gv	t_mco16sup_fffm	t_mco16sup_lum	t_rip12c
er_aro_f_2012	er_inv_v_2009	t_had09med	t_had12medatu	t_mco07h	t_mco09fh	t_mco11bphna	t_mco12fcstc	t_mco13pstc	t_mco14medatu	t_mco15sup_fffm	t_mco16sup_loa	t_mco16sup_loa	t_rip12e
er_aro_f_2013	er_inv_v_2010	t_had09stc	t_had12stc	t_mco07fi	t_mco09fhct	t_mco11bphnc	t_mco12fh	t_mco13gv11e	t_mco14medthrombo	t_mco15sup_lact	t_mco16sup_fffm	t_mco16sup_lum	t_rip12fa
er_aro_f_2014	er_inv_v_2011	t_had13a	t_had15a	t_mco07fm	t_mco09fct	t_mco11bphnp	t_mco12fhstc	t_mco13gv11f	t_mco14or	t_mco15sup_fffm	t_mco16sup_loa	t_mco16sup_loa	t_rip12fb
er_aro_f_2015	er_inv_v_2012	t_had13b	t_had15fb	t_mco07fctc	t_mco09fctc	t_mco11c	t_mco12fci	t_mco13alact	t_mco14pie	t_mco15sup_gv	t_mco16sup_fffm	t_mco16sup_lum	t_rip12fc
er_aro_f_2016	er_inv_v_2013	t_had13c	t_had15fc	t_mco07fh	t_mco09fh	t_mco11cstc	t_mco12fi	t_mco13med	t_mco14porg	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fh
er_bio_f_2007	er_inv_v_2014	t_had13d	t_had15hd	t_mco07g9	t_mco09fh	t_mco11d	t_mco12fhstc	t_mco13medatu	t_mco14stc	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2008	er_inv_v_2015	t_had13e	t_had15fd	t_mco07g10b	t_mco09fhct	t_mco11diap	t_mco12fm	t_mco13medthrombo	t_mco14um	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2009	er_trs_f_2007	t_had13f	t_had15ff	t_mco07g11c	t_mco09g10c	t_mco11dmip	t_mco12fmstc	t_mco13or	t_mco14sup_gv	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2010	er_trs_f_2008	t_had13g	t_had15fdmpp	t_mco07g	t_mco09g11b	t_mco11e	t_mco12f	t_mco13pie	t_mco14valo	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2011	er_tot_f_2011	t_had13h	t_had15gre	t_mco07um	t_mco09g11b	t_mco11e	t_mco12f	t_mco13pie	t_mco14valo	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2012	er_tot_f_2012	t_had13i	t_had15grc	t_mco08a	t_mco09med	t_mco11fa	t_mco12fpcstc	t_mco13porg	t_mco14valo	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2013	er_tot_f_2013	t_had13j	t_had15grp	t_mco08b	t_mco09pie	t_mco11ffa	t_mco12g11d	t_mco13stc	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2014	er_tot_f_2014	t_had13k	t_had15leg	t_mco08c	t_mco09porg	t_mco11fb	t_mco12g11e	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2015	er_tot_f_2015	t_had13l	t_had15med	t_mco08d	t_mco09stc	t_mco11fbstc	t_mco12lact	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_bio_f_2016	er_tot_f_2016	t_had13m	t_had15med	t_mco08e	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2007	er pha_f_2013	t_had13n	t_had15m	t_mco08f	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2008	er pha_f_2014	t_had13o	t_had15s	t_mco08g	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2009	er pha_f_2015	t_had13p	t_had15stc	t_mco08h	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2010	er pha_f_2016	t_had13q	t_had15t	t_mco08i	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2011	er pha_f_2017	t_had13r	t_had15t	t_mco08j	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2012	er pha_f_2018	t_had13s	t_had15t	t_mco08k	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2013	er pha_f_2019	t_had13t	t_had15t	t_mco08l	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2014	er pha_f_2020	t_had13u	t_had15t	t_mco08m	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2015	er pha_f_2021	t_had13v	t_had15t	t_mco08n	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_cam_f_2016	er pha_f_2022	t_had13w	t_had15t	t_mco08o	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2007	er prs_f_2007	t_had13x	t_had15t	t_mco08p	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2008	er prs_f_2008	t_had13y	t_had15t	t_mco08q	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2009	er prs_f_2009	t_had14a	t_had15t	t_mco08r	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2010	er prs_f_2010	t_had14b	t_had15t	t_mco08s	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2011	er prs_f_2011	t_had14c	t_had15t	t_mco08t	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2012	er prs_f_2012	t_had14d	t_had15t	t_mco08u	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2013	er prs_f_2013	t_had14e	t_had15t	t_mco08v	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2014	er prs_f_2014	t_had14f	t_had15t	t_mco08w	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2015	er prs_f_2015	t_had14g	t_had15t	t_mco08x	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_dtr_f_2016	er prs_f_2016	t_had14h	t_had15t	t_mco08y	t_mco09um	t_mco11fct	t_mco12med	t_mco13um	t_mco14val	t_mco15sup_loa	t_mco16sup_loa	t_mco16sup_loa	t_rip12fi
er_ete_f_2007	er rat_f_2008	t_had14i	t_had15t	t_mco09a	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2008	er rat_f_2009	t_had14j	t_had15t	t_mco09b	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2009	er rat_f_2010	t_had14k	t_had15t	t_mco09c	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2010	er rat_f_2011	t_had14l	t_had15t	t_mco09d	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2011	er rat_f_2012	t_had14m	t_had15t	t_mco09e	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2012	er rat_f_2013	t_had14n	t_had15t	t_mco09f	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2013	er rat_f_2014	t_had14o	t_had15t	t_mco09g	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2014	er rat_f_2015	t_had14p	t_had15t	t_mco09h	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2015	er rat_f_2016	t_had14q	t_had15t	t_mco09i	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2016	er rat_f_2017	t_had14r	t_had15t	t_mco09j	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2017	er rat_f_2018	t_had14s	t_had15t	t_mco09k	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2018	er rat_f_2019	t_had14t	t_had15t	t_mco09l	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2019	er rat_f_2020	t_had14u	t_had15t	t_mco09m	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2020	er rat_f_2021	t_had14v	t_had15t	t_mco09n	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2021	er rat_f_2022	t_had14w	t_had15t	t_mco09o	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2022	er rat_f_2023	t_had14x	t_had15t	t_mco09p	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a
er_ete_f_2023	er rat_f_2024	t_had14y	t_had15t	t_mco09q	t_mco10fct	t_mco11a	t_mco12a	t_mco13a	t_mco14a	t_mco15a	t_mco16a	t_mco16a	t_rip12a

979
tables .sas7bdat

473 Go de données

979 élément(s)

473 Go

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

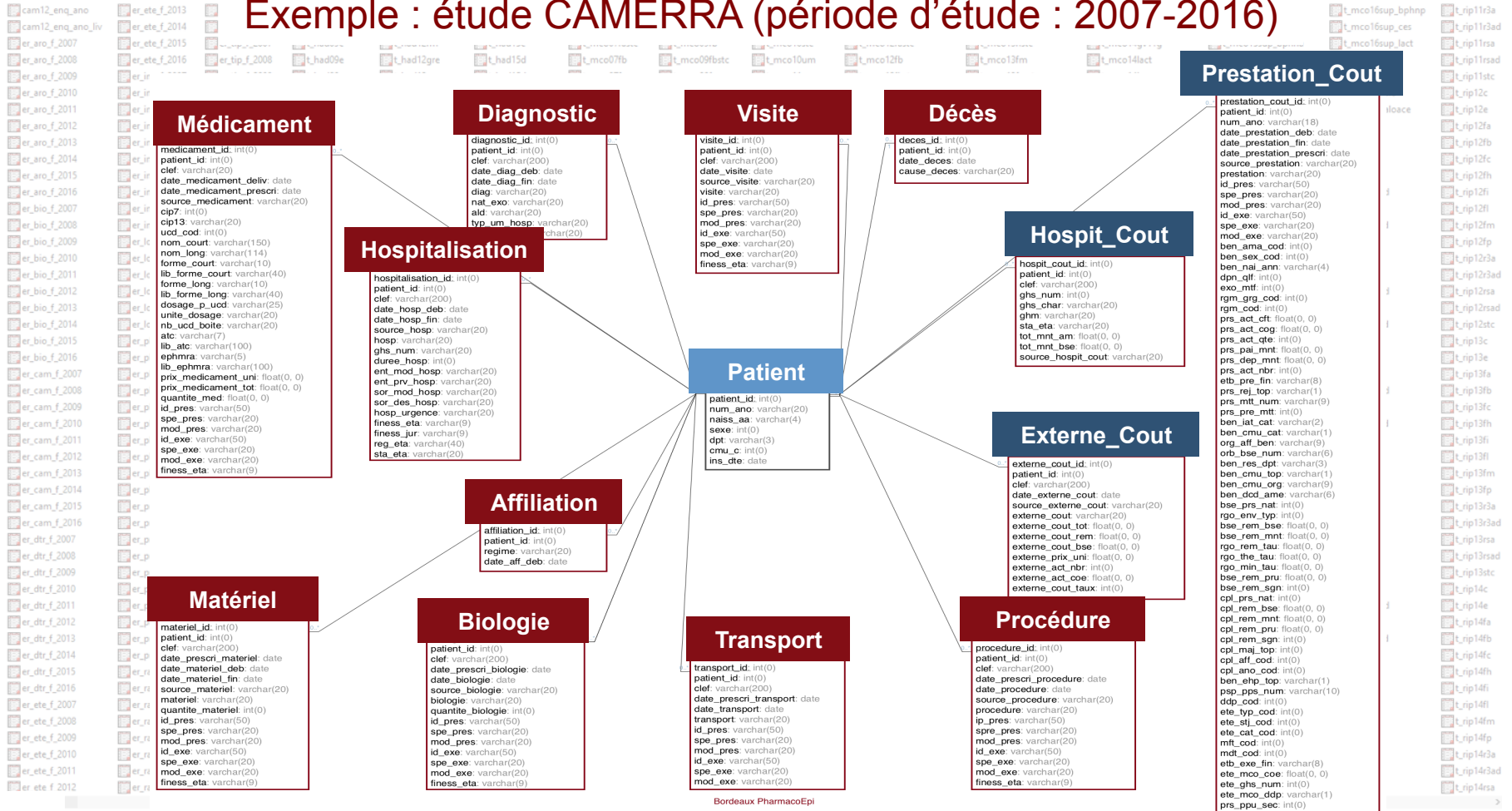
Démarche

- 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