

Raisonner sur les données médico-administratives pour reconstruire et analyser les parcours de soins

Séminaire INOVIVE – 21/09/2021

T. Guyet, Inria Lyon

(travail majoritairement mené à l'IRISA)

Chaire AI-RACLES (Inria/APHP/CS)



Thomas.guyet@inria.fr

Contextes

- Consortium PEPS – Pharmaco-Epidemiologie des produits de Santé (Financement ANSM)
 - CHU Rennes / IRISA/Inria / EHESP
 - Consortium multi-disciplinaire avec un programme de R&D sur les outils
 - **Objectif : Exploiter les données du SNDS pour répondre à des questions de pharmaco-épidémiologie**
- Chaire APHP (E. Audureau) / Inria / CS (L. Le Brusquet, A. Tannehaus)
 - **Objectif : Exploiter les données de l'Entrepôt de l'APHP par des méthodes l'IA**

Exemple : Etude GENEPI (E. Polard et al.): *“évaluer l’association entre la substitution Princeps-Génériques et les hospitalisations pour épilepsie”*

→ exploitation des données du SNDS

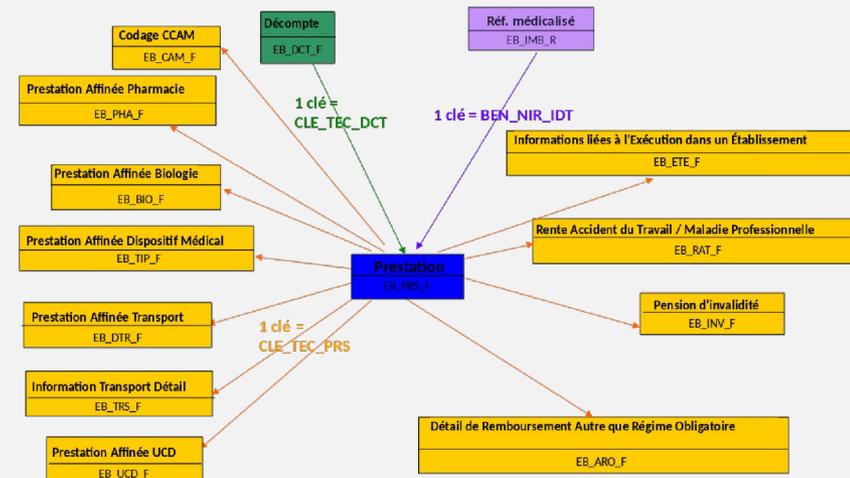
Données médico-administratives et études épidémiologiques

- Sources de données médico-administratives

- Données SNDS – SNIIRAM
- Données hospitalières (eg. APHP)

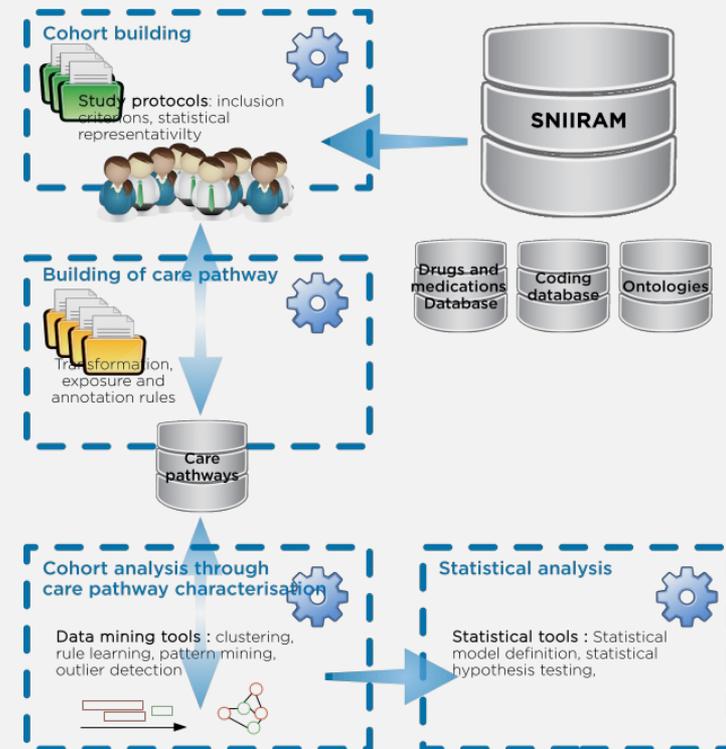
- Avantages/Inconvénients du SNDS

- + Données numériques disponibles sans délai de collecte
- + Population large
- + Taux de couverture de la population
- + Données structurées
- + Données historisées
- + Possibilité de chaînage avec des données de cohortes, données de capteurs, etc
- Des informations partielles
- Information médicale limitée
- *Fossé sémantique entre la donnée et l'information recherchée*



Données médico-administratives et études épidémiologiques

- Sources de données médico-administratives
 - Données SNDS – SNIIRAM
 - Données hospitalières
- *Épidémiologie numérique* : Alternative aux études épidémiologiques classiques
 - Les étapes de l'étude épidémiologique deviennent des étapes d'une analyse de données
 - Sélection de patients
 - Identification d'événements d'intérêts
 - Analyse des liens entre événements d'intérêts
- Dans ce travail : notion centrale de « **parcours de soins** »



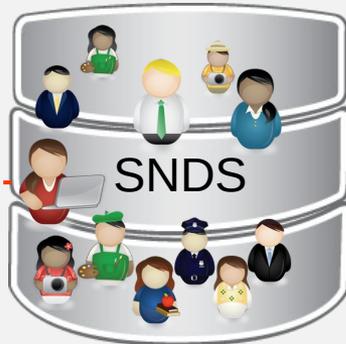


Question de l'étude
 Association entre la
 substitution Princeps-
 Génériques et les
 hospitalisations pour
 épilepsie ??



Exemple : Etude GENEPI (E. Polard et al.): "évaluer l'association entre la substitution Princeps-Génériques et les hospitalisations pour épilepsie"

Quels patients choisis ?



Question de l'étude
Association entre la substitution Princeps-Génériques et les hospitalisations pour épilepsie ??



Quels événements utiliser pour décrire le parcours de soins ?

Comment le formaliser, les représenter ?



Quelle question poser ?

Comment traduire la question ?

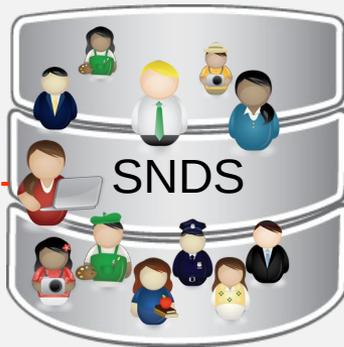
Quelles questions peuvent être efficacement posées ?



Problème majeur : Fossé sémantique

Champs sémantique administratif

Quels patients choisis ?



Quels événements utiliser pour décrire le parcours de soins ?

Champs sémantique médical



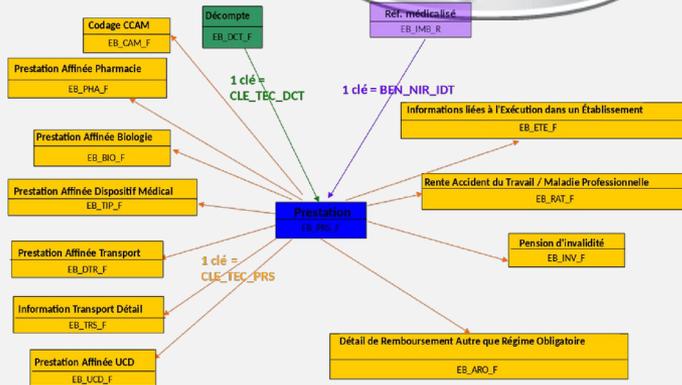
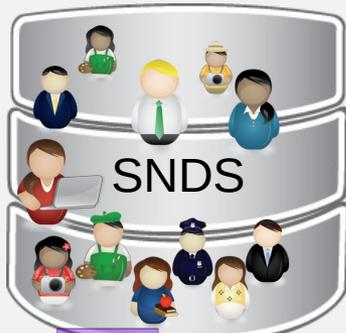
Question de l'étude
Association entre la substitution Princeps-Génériques et les hospitalisations pour épilepsie ??



Quelle "question" poser ?

Problème majeur : Fossé sémantique

Champs sémantique administratif



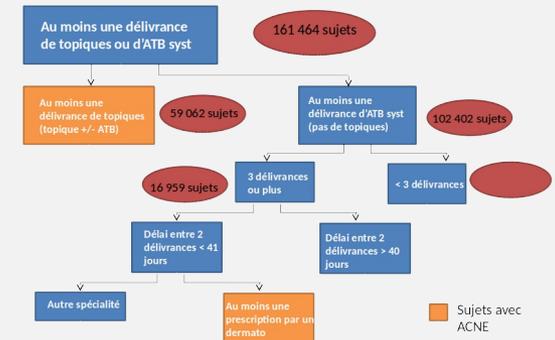
Champs sémantique médical



Question de l'étude Association entre la substitution Princeps-Génériques et les hospitalisations pour épilepsie ??

```
SELECT
  EXE_S01_DTD, PHA_PRS_IDE, PFS_PRS_NUM, PFS_EXE_NUM,
  PSE_SPE_CDD, PSP_SPE_CDD, PSE_ACT_MAT, PSP_ACT_MAT
FROM
  ER_PRS_F as prs,
  ER_PHA_F as pha
WHERE
  (prs.FLX_D15_DTD = pha.FLX_D15_DTD) AND (prs.FLX_TRT_DTD = pha.FLX_TRT_DTD) AND
  (prs.FLX_DMT_TYP = pha.FLX_DMT_TYP) AND (prs.FLX_DMT_NUM = pha.FLX_DMT_NUM) AND
  (prs.FLX_DMT_ORD = pha.FLX_DMT_ORD) AND (prs.ORG_CLE_NUM = pha.ORG_CLE_NUM) AND
  (prs.DCT_ORD_NUM = pha.DCT_ORD_NUM) AND (prs.PRS_ORD_NUM = pha.PRS_ORD_NUM) AND
  (prs.REM_TYP_AFF = pha.REM_TYP_AFF) AND (BEN_NIR_PSA = '123456789')
ORDER BY
  EXE_S01_DTD,
  PHA_PRS_IDE;
```

Requête SQL



■ Sujets avec ACNE

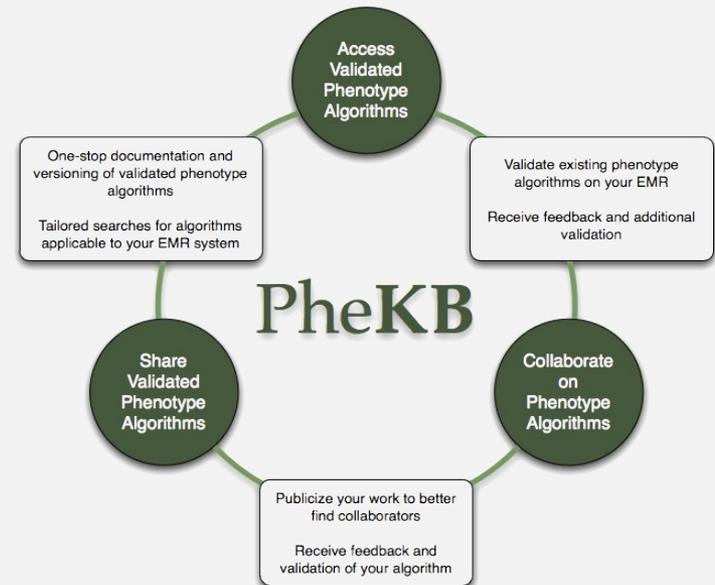
Fossé sémantique – illustrations pratiques

- Enrichir à l'aide d'**information phénotypique**
 - Données disponibles : des prescriptions de soins
 - Information médicale pertinente : le patient est atteint de telle maladie ou de tel syndrome (sur telle période)
 - Enjeux de précision des algorithmes d'identification

Exemple 1 : Patient épileptique dans le SNDS : *“patient ayant eu au moins 10 délivrances d'un médicament anti-épileptique sur 1 an.”*

ReDSiam

Réseau pour mieux utiliser les Données
du Système national des données de santé



Kirby, J. C., Speltz, P., Rasmussen, L. V., Basford, M., Gottesman, O., Peissig, P. L., ... & Ellis, S. B. (2016). PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association*, 23(6), 1046-1052.

Fossé sémantique – illustrations pratiques

- Enrichir à l'aide d'**information phénotypique**
 - Données disponibles : des prescriptions de soins
 - Information médicale pertinente : le patient est atteint de telle maladie ou de tel syndrome (sur telle période)
 - Enjeux de précision des algorithmes d'identification

Exemple 2 VTE (Venous Thrombo-Embolicism) : In clinical practice facing a suspicion of VTE physicians first prescribe antithrombotics and then confirm or not the diagnosis through specific medical procedures: e.g. **Doppler ultrasonography** or CT-scan. Patients with suspected **Pulmonary Embolism** are often hospitalized whereas patients with suspected Deep Vein Thrombosis (DVT) are managed on an outpatient basis. On the one side, if the DVT suspicion is confirmed, **antithrombotic deliveries** continue **for 3 to 12 months (once per month)**. Hence, the diagnosis (through the same medical procedures as above) is **preceded or followed** by initiating an **anticoagulant treatment within a time window** of **at most 0 to 2 days**. On the other side, Pulmonary Embolism suspicion leads to hospitalization **during** which medical procedures are performed to confirm the diagnosis and **then** anticoagulant delivery is observed only **after** the patient comes back home.

Problématique et approche

- Proposer des **outils pour faciliter l'enrichissement des données du SNDS** vers de données exploitables par des méthodes d'analyse de données pour en tirer des informations médicales
- Notre approche
 - Modélisation de « parcours de soins » : conserver la **richesse temporelle** et descriptive du SNDS
 - Exploiter des **connaissances du domaine** : intégrer des connaissances formalisées (taxonomies ATC, CIM10, CCAM, ...)
 - **Approche versatile** : fournir un cadre pour la conception d'outils répondant à un grand nombre d'études
- Nos propositions
 - 1) Langage de requête hybride dédié aux études de pharmaco-épidémiologie
 - 2) Exploration des outils du web sémantique (Thèse J. Bakalara)

Langage de requêtes dédié aux études de pharmaco- épidémiologie

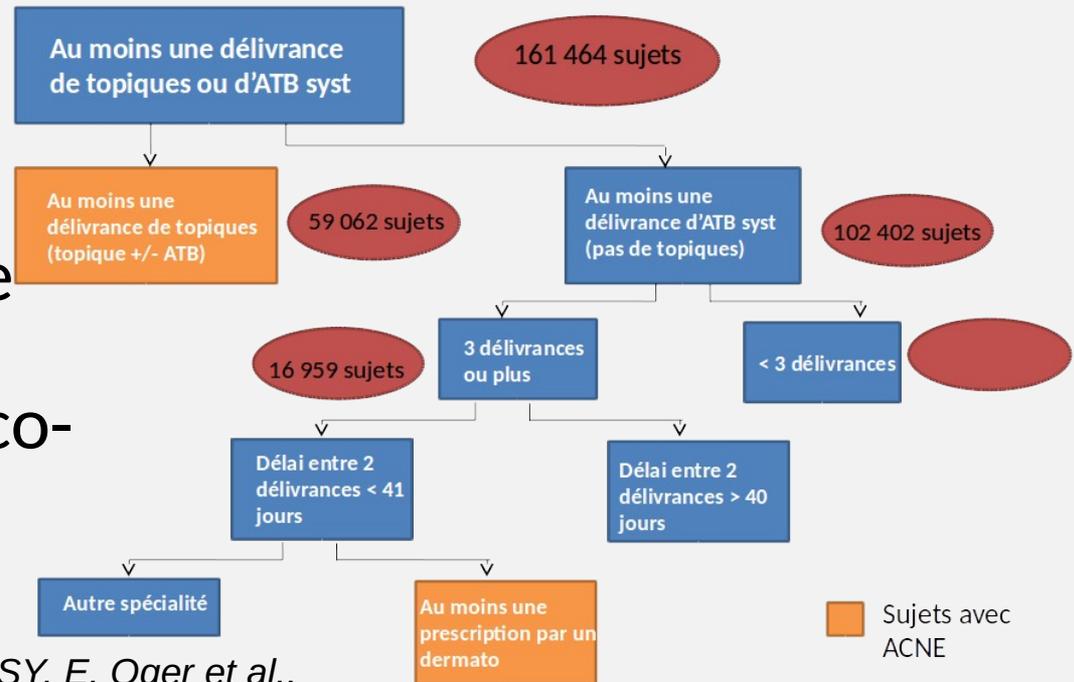
Langage de requête dédié

- Langage SQL

```

SELECT
  EXE_SOI_DTD, PHA_PRS_IDE, PFS_PRE_NUM, PFS_EXE_NUM,
  PSE_SPE_COD, PSP_SPE_COD, PSE_ACT_NAT, PSP_ACT_NAT
FROM
  ER_PRS_F as prs,
  ER_PHA_F as pha
WHERE
  (prs.FLX_DIS_DTD = pha.FLX_DIS_DTD) AND (prs.FLX_TRT_DTD = pha.FLX_TRT_DTD) AND
  (prs.FLX_EMT_TYP = pha.FLX_EMT_TYP) AND (prs.FLX_EMT_NUM = pha.FLX_EMT_NUM) AND
  (prs.FLX_EMT_ORD = pha.FLX_EMT_ORD) AND (prs.ORG_CLE_NUM = pha.ORG_CLE_NUM) AND
  (prs.DCT_ORD_NUM = pha.DCT_ORD_NUM) AND (prs.PRS_ORD_NUM = pha.PRS_ORD_NUM) AND
  (prs.REM_TYP_AFF = pha.REM_TYP_AFF) AND
  (BEN_NIR_PSA = '123456789')
ORDER BY
  EXE_SOI_DTD,
  PHA_PRS_IDE;
    
```

- « Langage » dédié pour la réalisation de la préparation des données de pharmaco-epidemiologie



ISOPSY, E. Oger et al.,

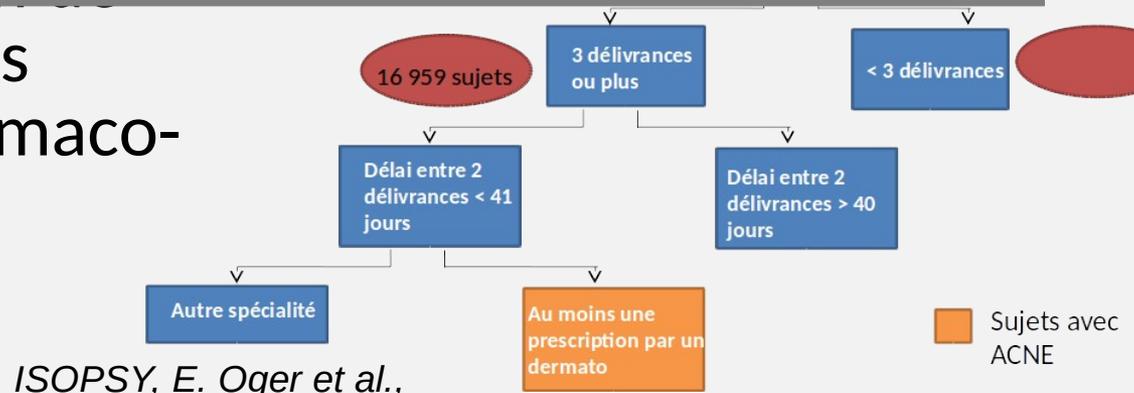
Langage de requête dédié

- Langage SQL

```
SELECT
  EXE_SOI_DTD, PHA_PRS_IDE, PFS_PRE_NUM, PFS_EXE_NUM,
  PSE_SPE_COD, PSP_SPE_COD, PSE_ACT_NAT, PSP_ACT_NAT
FROM
  ER_PRS_F as prs,
  ER_PHA_F as pha
WHERE
  (prs.FLX_DIS_DTD = pha.FLX_DIS_DTD) AND (prs.FLX_TRT_DTD = pha.FLX_TRT_DTD) AND
  (prs.FLX_EMT_TYP = pha.FLX_EMT_TYP) AND (prs.FLX_EMT_NUM = pha.FLX_EMT_NUM) AND
  (prs.FLX_EMT_ORD = pha.FLX_EMT_ORD) AND (prs.ORG_CLE_NUM = pha.ORG_CLE_NUM) AND
```

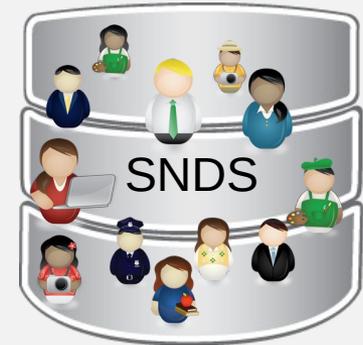
Peut-on proposer un langage “naturel” à utiliser par les épidémiologistes pour qu’ils représentent leurs études ??

- « La... pour la préparation des données de pharmaco-epidemiologie



ISOPSY, E. Oger et al.,

Éléments d'une étude de pharmaco-épidémiologie

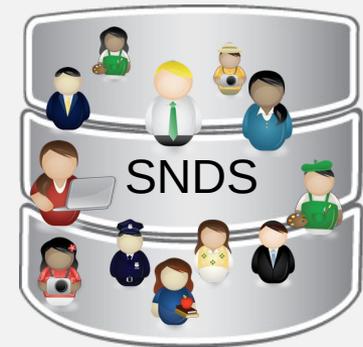


- Algèbre d'opérations « élémentaires » sur des « parcours de soins »
 - Sélection de patients sur un critère $\sigma_{\varphi}(D)$
 - eg. sélection des **patients épileptiques**
 - Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$
 - eg. sélection des **antiépileptiques**
 - Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$
 - eg. événements **pas après une hospitalisation**
 - Induction d'événements $\iota_{\varphi}(D)$
 - eg. **exposition à un antiépileptique**
 - Labellisation d'un patient sur un critère
 - eg. patients avec switches

$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie

- Algèbre d'opérations « élémentaires » sur des « parcours de soins »



- Sélection de patients sur un critère $\sigma_{\varphi}(D)$

- eg. sélection des **patients épileptiques**

- Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$

- eg. sélection des **antiépileptiques**

- Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$

- eg. événements **pas après une hospitalisation**

- Induction d'événements $\iota_{\varphi}(D)$

- eg. **exposition à un antiépileptique**

- Labellisation d'un patient sur un critère

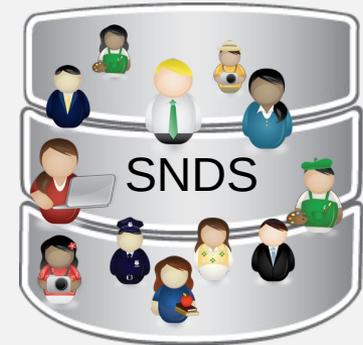
- eg. patients avec switches



$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie

- Algèbre d'opérations « élémentaires » sur des « parcours de soins »



- Sélection de patients sur un critère $\sigma_{\varphi}(D)$

- eg. sélection des **patients épileptiques**

- Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$

- eg. sélection des **antiépileptiques**

- Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$

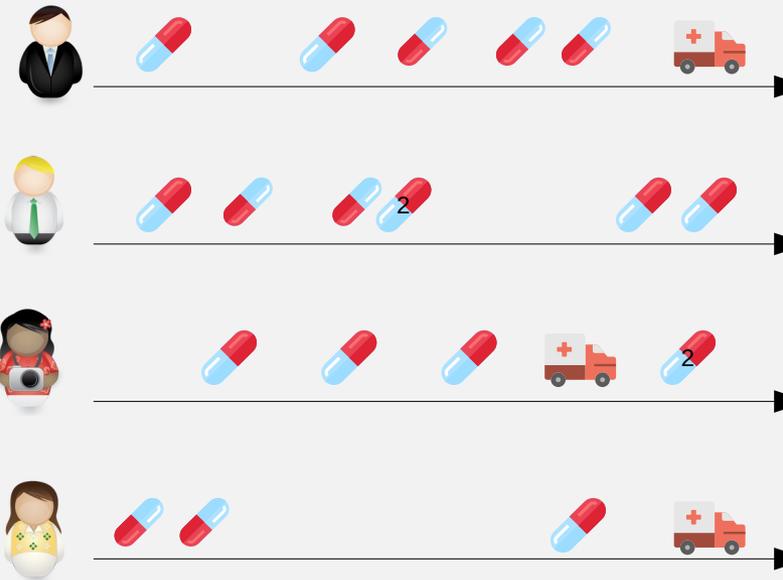
- eg. événements **pas après une hospitalisation**

- Induction d'événements $\iota_{\varphi}(D)$

- eg. **exposition à un antiépileptique**

- Labellisation d'un patient sur un critère

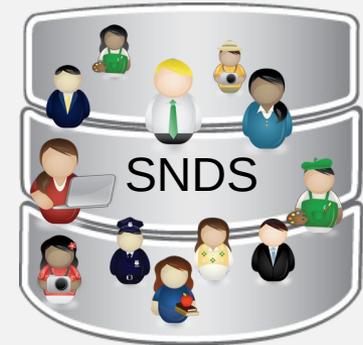
- eg. patients avec switches



$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie

- Algèbre d'opérations « élémentaires » sur des « parcours de soins »



- Sélection de patients sur un critère $\sigma_{\varphi}(D)$

- eg. sélection des **patients épileptiques**

- Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$

- eg. sélection des **antiépileptiques**

- Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$

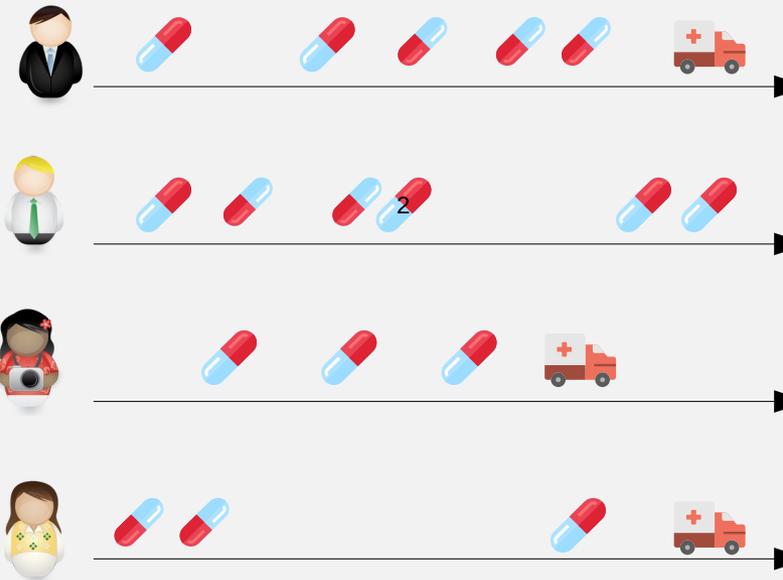
- eg. événements **pas après une hospitalisation**

- Induction d'événements $\iota_{\varphi}(D)$

- eg. **exposition à un antiépileptique**

- Labellisation d'un patient sur un critère

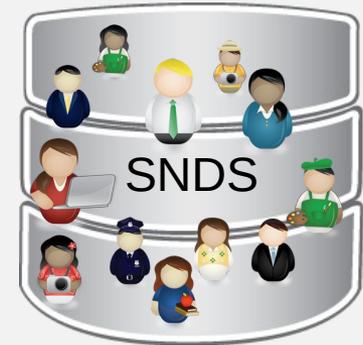
- eg. patients avec switches



$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie

- Algèbre d'opérations « élémentaires » sur des « parcours de soins »



- Sélection de patients sur un critère $\sigma_{\varphi}(D)$

- eg. sélection des **patients épileptiques**

- Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$

- eg. sélection des **antiépileptiques**

- Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$

- eg. événements **pas après une hospitalisation**

- Induction d'événements $\iota_{\varphi}(D)$

- eg. **exposition à un antiépileptique**

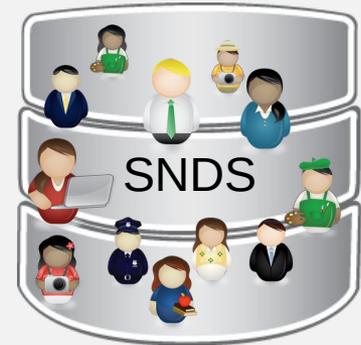
- Labellisation d'un patient sur un critère

- eg. patients avec switches



$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie



- Algèbre d'opérations « élémentaires » sur des « parcours de soins »

- Sélection de patients sur un critère $\sigma_{\varphi}(D)$

- eg. sélection des **patients épileptiques**

- Projection d'événements sur un critère attributaires $\pi_{\varphi}(D)$

- eg. sélection des **antiépileptiques**

- Sélection d'événements sur critère temporels $\tau_{\varphi}(D)$

- eg. événements **pas après une hospitalisation**

- Induction d'événements $\iota_{\varphi}(D)$

- eg. **exposition à un antiépileptique**

- Labellisation d'un patient sur un critère

- eg. patients avec switches

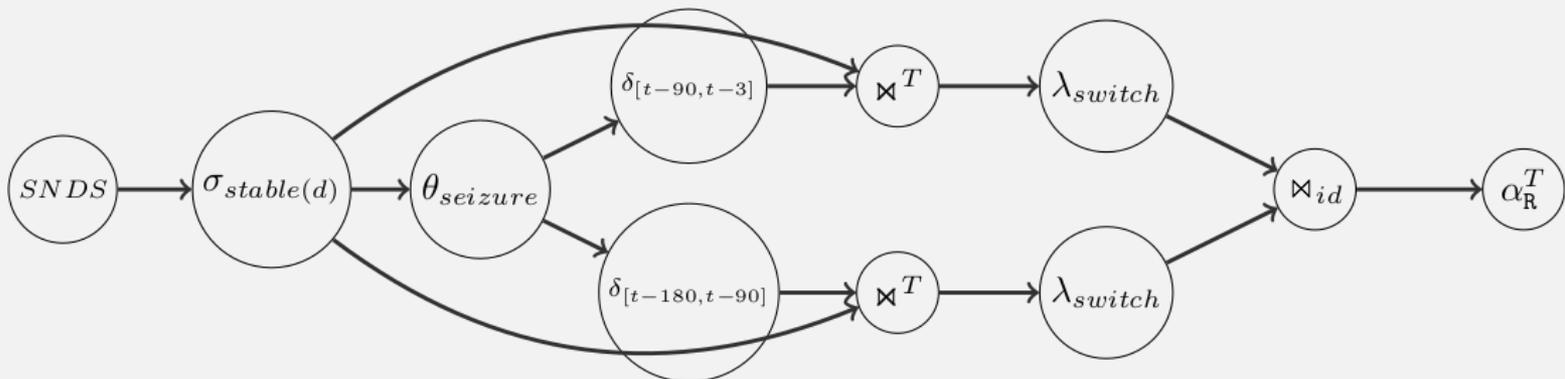
$$\lambda(\iota_{\varphi}(\tau_{\varphi}(\pi_{\varphi}(\sigma_{\varphi}(D))))))$$

Éléments d'une étude de pharmaco-épidémiologie

- Deux niveaux d'abstraction
 - **Opérateurs** : définitions d'opérations génériques (classiques, adapté pour les parcours)
 - **Critères phénotypiques** : richesse sémantique requise

$$\sigma_{\varphi}(D) = \{d \in D \mid d, \mathcal{M} \models \varphi\}$$

- Exemple de l'étude GENEPI
 - Représentation d'une étude de type case-crossover



Opérationnalisation ?

- Besoins
 - Représenter la notion de **parcours** de soin
 - Implémenter les quelques **opérateurs** de base
 - Donner des possibilités d'avoir des **critères « riches » pour les opérateurs**
- Plusieurs approches d'opérationnalisation
 - Implémentation « en dur » : efficace, mais *peu versatile*
 - Implémentation déclarative : versatile, mais peu efficace
 - En ASP (Guyet et al., AIME 2017)
 - À l'aide d'outils du web sémantique
 - **Implémentation hybride**
 - opérateurs de base « en dur »
 - critères exprimés de manière déclarative

Modèle de parcours de soins

- Modèle de données orienté parcours de soins
 - Collection de patients
 - Pour chaque patient
 - Caractéristiques propres à un patient (eg. statut ALD, sexe, date de naissance)
 - Ensemble d'évènements datés et attribués
 - Modèle de temps : continu, avec intervalle
 - Modèle attributs : relationnel, il a des paramètres
 - Modèle de données flexible et assez générique
- Représentation algébrique

Definition (Attributed sequence in context $\mathcal{C} = \langle \mathcal{C}, \mathcal{R} \rangle$). Let $\mathcal{C} = \{c_1, \dots, c_n\}$ be the set of n sequence attribute definitions, then an **attributed sequence** is a pair $(id, \vec{v}, \mathbf{s}^1, \dots, \mathbf{s}^r)$ such that:

- ▷ $id \in \mathbb{N}$ is an identifier of the individual,
- ▷ \vec{v} is a tuple of values such that v_i is a value of type τ_i in the domain δ_i for all $i \in [n]$, where $c_i = \langle \nu_i, \tau_i, \delta_i \rangle$ is the attribute definitions of the i -th attribute,
- ▷ $\mathbf{s}^i = \langle s_1^i \dots s_{k^i}^i \rangle$ is a sequence of attributed events of type \mathcal{R}_i , i.e. s_i is an attributed event in context \mathcal{R}_i .

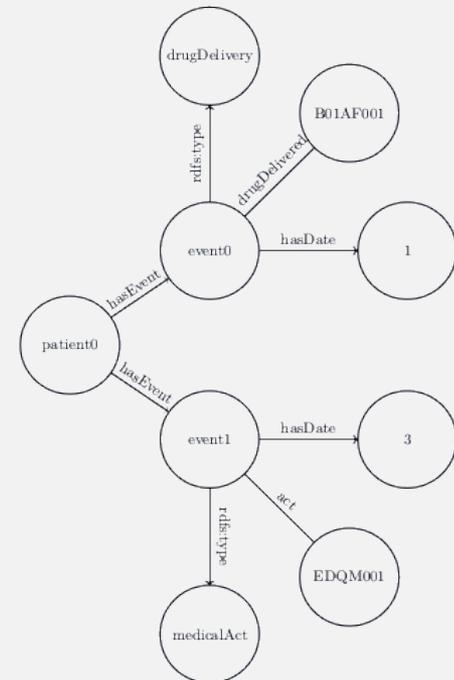
Modèle de parcours de soins

- Modèle de données orienté parcours de soins
 - Collection de patients
 - Pour chaque patient
 - Caractéristiques propres à un patient (eg. statut ALD, sexe, date de naissance)
 - Ensemble d'évènements datés et attribués
 - Modèle de données très générique et flexible
- Représentation formelle en logique du premier ordre

```
– ASP      sex(p1, 1).
           birthYear(p1, 1979).
           event(p1, 0, 0, drugDelivery(3430126, 1) ).
           event(p1, 3, 5, procedure(ECGH011) ).
           event(p1, 12, 12, drugDelivery(4991401, 1) ).
           event(p1, 23, 23, drugDelivery(3588755, 2) ).
           event(p1, 38, 38, procedure(YYYY600) ).
           event(p1, 45, 45, drugDelivery(3723093, 1) ).
           event(p1, 62, 66, procedure(GAQE001) ).
```

Modèle de parcours de soins

- Modèle de données orienté parcours de soins
 - Collection de patients
 - Pour chaque patient
 - Caractéristiques propre à un patient (eg. Statut ALD, Sexe, date de naissance)
 - Ensemble d'évènements datés et attribués
 - Modèle de données très générique et flexible
- Représentation formelle en logique du premier ordre
 - ASP
 - RDF



Opérateurs hybrides

- Exemple de l'opérateur de sélection

$$\sigma_{\varphi}(D) = \{d \in D \mid d, \mathcal{M} \models \varphi\}$$

- Critères complexes
(phénotypiques,
informationnels ou inductifs)

φ : - patient avec VTE
- patient épileptique

- Besoin d'intégrer des raisonnements complexes sur les parcours de soins
- Besoin de flexibilité pour exprimer des contraintes variées
- Résolution hybride : critères exprimés sous une forme de logique du premier ordre (ASP/RDF)
 - Expressivité élevée
 - Lisibilité et flexibilité de contraintes

Exemple d'expression d'un critère (inductif)

- Période de stabilité d'un traitement antiépileptique : une année sans crise et au moins 10 délivrances d'antiépileptique

```
#const nb_delivery=10.
#const p_duration=365.

aed(AED) :- AED=(n03ax09;n03ax14;n03ax11;n03ag01;n03af01;n03af02).

%% choice of the index date as a crisis event
1{ end(T) : seq(T, 9999), T>p_duration } 1.

begin(TB) :- end(T), TB=T-p_duration.

%% no crisis within the period
:- 1 { seq(T,9999) : T<E, T>=B }, end(E), begin(B).

%% number of deliveries of class AED
nbprescript(AED,N) :- N=#sum{T:seq(T,C), T<E, T>=B, cip(C,AED) },
begin(B), end(E), aed(AED).

stable :- N>=nb_delivery, nbprescript(AED,N), aed(AED).
:- not stable.
```

φ :

Exemple d'expression d'un critère (inductif)

- Période de stabilité d'un traitement antiépileptique : une année sans crise et au moins 10 délivrances d'antiépileptique

```
#const nb_delivery=10.
#const p_duration=365.

aed(AED) :- AED=(n03ax09;n03ax14;n03ax11;n03ag01;n03af01;n03af02).

%% choice of the index date as a crisis event
1{ end(T) : seq(T, 9999), T>p_duration } 1.

begin(TB) :- end(T), TB=T-p_duration.

φ : %% no crisis within the period
:- 1 { seq(T,9999) : T<E, T>=B }, end(E), begin(B).

%% number of deliveries of class AED
nbprescript(AED,N) :- N=#sum{T:seq(T,C), T<E, T>=B, cip(C,AED) },
begin(B), end(E), aed(AED).

stable :- N>=nb_delivery, nbprescript(AED,N), aed(AED).
:- not stable.
```

Limites

- les solveurs ASP sont peu à l'aise avec le raisonnement ontologique !
- fonctionnement en mémoire (efficacité à améliorer)

Exploitation des outils du web sémantique pour l'identification de patients dans le SNDS

Thèse en cours J. Bakalara

Supervision : O. Dameron, A. Happe, E. Oger

Alternative du web sémantique

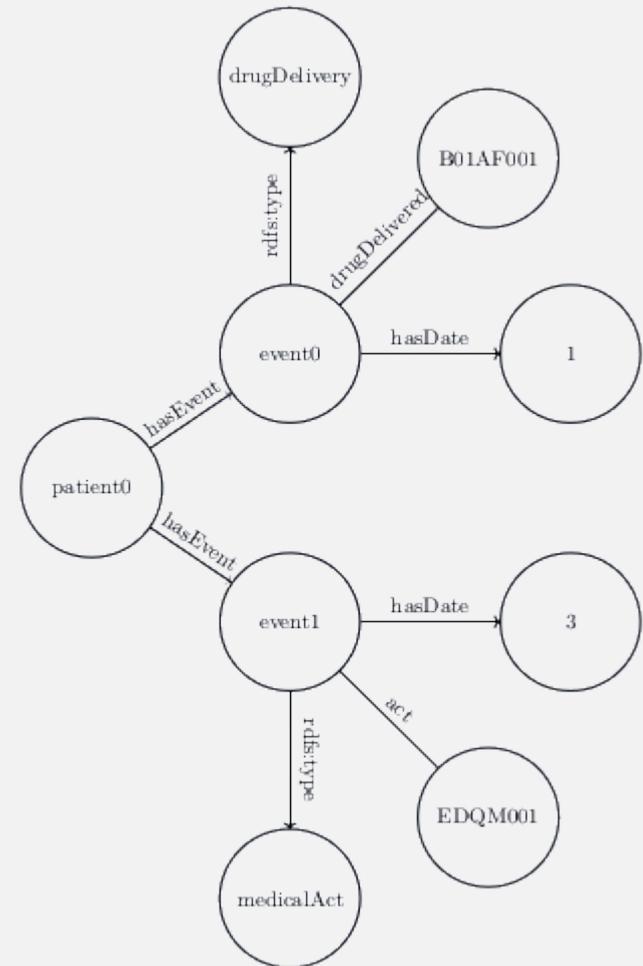
- Web sémantique = alternative à ASP dans l'outil précédent
- Besoin pour sélectionner des patients : spécifier des critères complexes (**phénotypes**)
 - eg. phenotype d'un patient épileptique
 - nécessite :
 - Des connaissances du domaine : concept de “médicament anti-épileptique” (à partir de la taxonomie ATC)
 - Des raisonnements ontologiques et temporels
- Web sémantique
 - Adapté pour représenter des données et raisonner sur des connaissances formalisées
 - Peu adapté pour manipuler des contraintes temporelles

Proposition

- 1) Représentation des parcours de soins en RDF
- 2) Spécification d'un phénotype comme une "chronique"
- 3) Mécanisme de requêtage efficace pour l'identification des patients adhérents au phénotype

Représentation RDF des parcours de soins

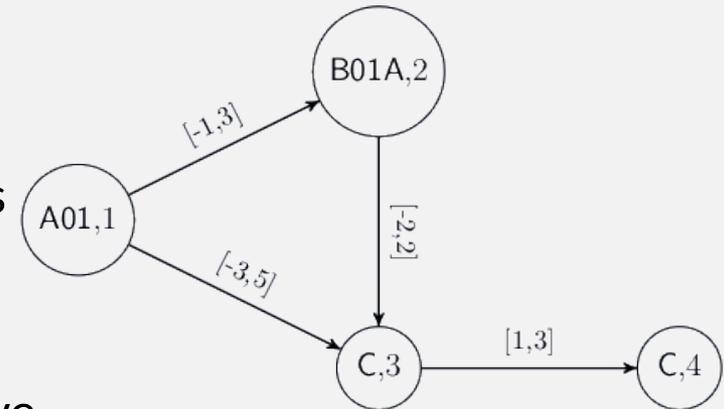
- Proposition d'un modèle de représentation des parcours de soins en RDF
 - Adaptation du travail de Y. Rivault
 - Modélisation plus complète des notions du SNDS
- Création de “moulinettes” pour transformer automatiquement les données SNDS en format RDF
 - Expérimentées sur données synthétiques
 - En cours sur données réelles



Spécifier des phénotypes par des “chroniques”

- Chronique

- Modèle temporel de l'état de l'art des systèmes à événements discrets
- Ensemble d'événements contraints par des délais entre leurs occurrences



- Phénotype = chronique

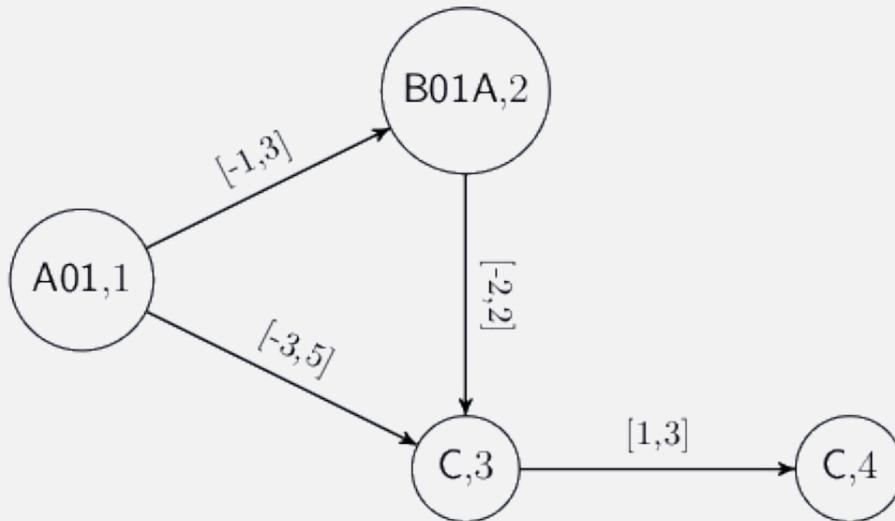
- Offre une représentation graphique intuitive
- Permet de donner des **contraintes temporelles riches**
- **Algorithmes efficaces** pour retrouver les occurrences d'une chronique
- Mais ... **événements un peu simple** : pas d'utilisation de notre ontologie !

- Chronique : exprimable en SPARQL

- Permet de faire des dépôts facilement réutilisables de phénotypes
- mais très lente !

Exemple 1 : Patient épileptique dans le SNDS : “patient ayant eu au moins 10 délivrances d'un médicament anti-épileptique sur 1 an.”

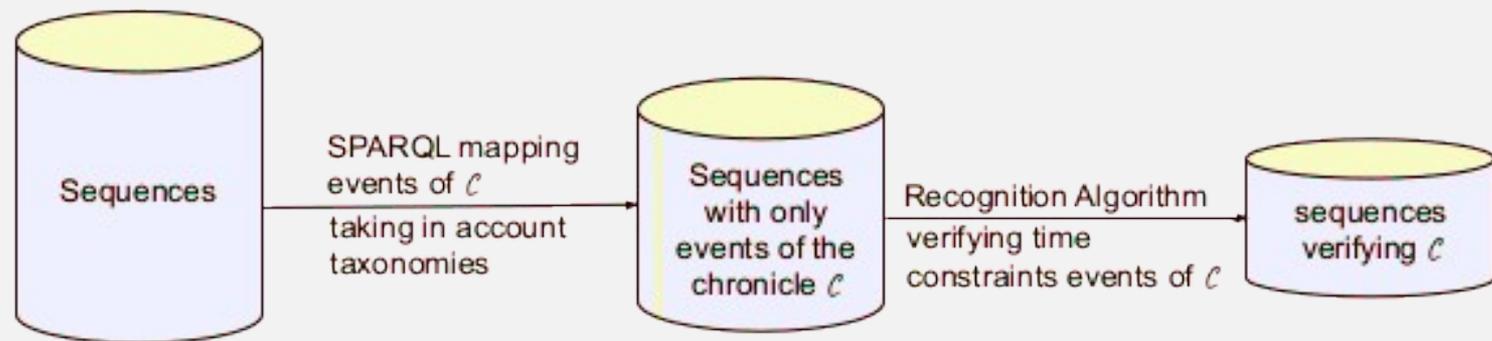
Chronique en SPARQL



```
SELECT DISTINCT * WHERE{
  ?sequence patdb:hasEvent ?evt1 .
    ?evt1 seq:evtDate ?date1 .
    ?evt1 seq:evtlabel ?atc1 .
    ?atc1 rdfs:subClassOf* atc:A01 .
  ?sequence patdb:hasEvent ?evt2 .
    ?evt2 seq:evtDate ?date2 .
    ?evt2 seq:evtlabel ?atc2 .
    ?atc2 rdfs:subClassOf* atc:B01A .
  ?sequence patdb:hasEvent ?evt3 .
    ?evt3 seq:evtDate ?date3 .
    ?evt3 seq:evtlabel ?atc3 .
    ?atc3 rdfs:subClassOf* atc:C .
  ?sequence patdb:hasEvent ?evt4 .
    ?evt4 seq:evtDate ?date4 .
    ?evt4 pseq:evtlabel ?atc4 .
    ?atc4 rdfs:subClassOf* atc:C .
  FILTER ( ?date2 - ?date1 >= -1)
  FILTER ( ?date2 - ?date1 <= 3)
  FILTER ( ?date3 - ?date1 >= -3)
  FILTER ( ?date3 - ?date1 <= 5)
  FILTER ( ?date3 - ?date2 >= -2)
  FILTER ( ?date3 - ?date2 <= 2)
  FILTER ( ?date4 - ?date3 >= 1)
  FILTER ( ?date4 - ?date3 <= 3)
}
```

HyCor : Recherche efficace à base de chroniques « sémantiques »

- Chronique « sémantique »
 - Nœud = requête de sélection d'événements
 - Permet d'exploiter toute la richesse de définition des événements représentés dans le graphe
 - On conserve les contraintes temporelles sur les délais
- Méthode de requêtage hybride (encore!)



Quelques résultats

- Exemple de requête :
identification de patient atteints d'une VTE (Venous Thrombo-Emboliem)
- Performances en temps

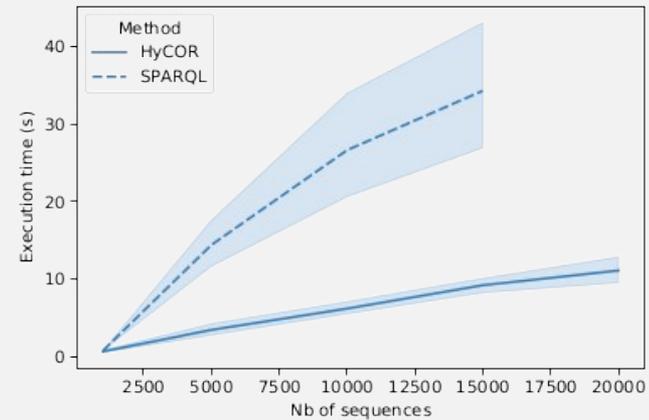
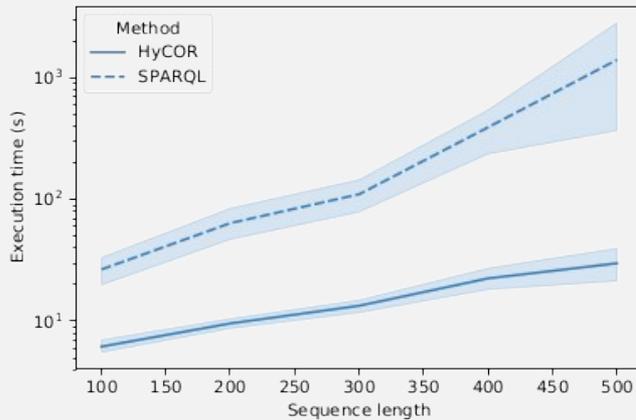
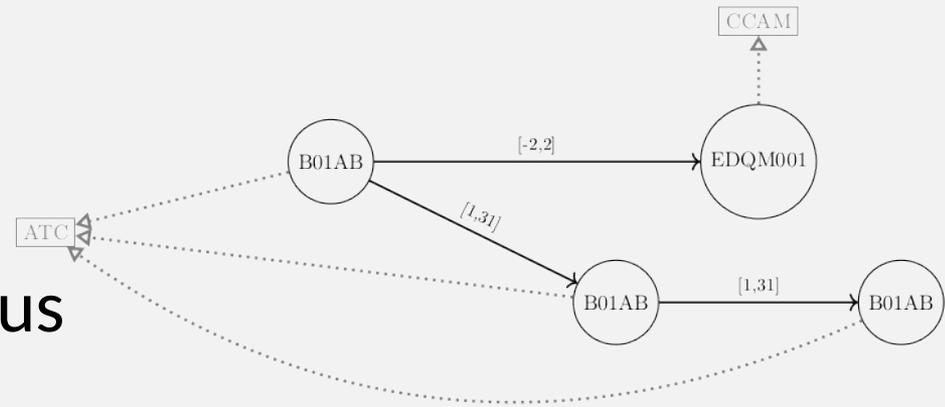


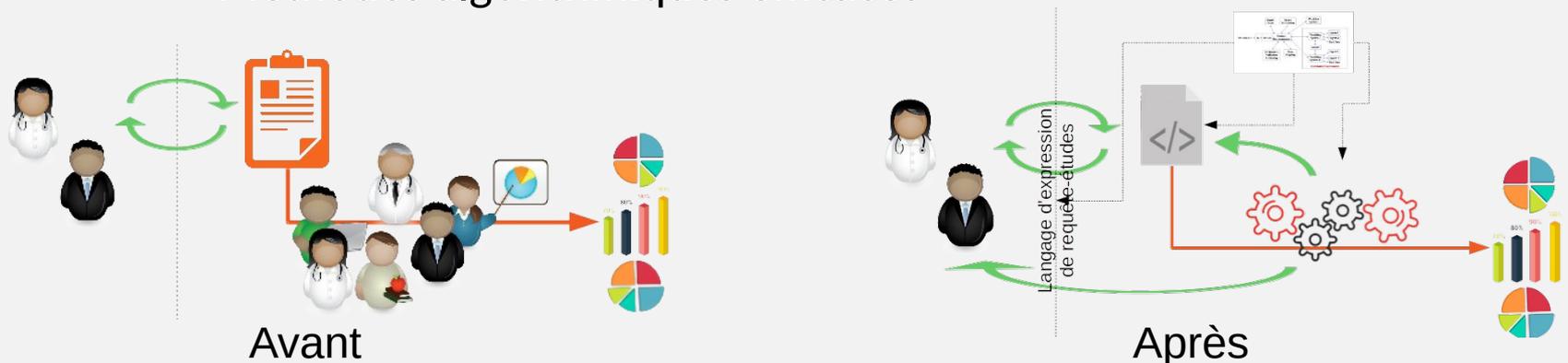
Figure 5.5 – Execution times (in seconds) of SPARQL and HyCOR wrt sequences length on 10 000 sequences (on the left) and wrt number of sequences (with length 100).

Conclusions

- SNDS/données hospitalières : source pour la *pharmaco-épidémiologie numérique*
 - source de données intéressante pour certaines études
 - données complexes et usage secondaire => fossé sémantique
- **Modéliser et raisonner sur les parcours de soins pour combler (en partie) le fossé sémantique**
 - Besoin d'outils pour faciliter ces manipulations
 - Besoin de méthodologie informatique pour augmenter les possibilités de la pharmaco-épidémiologie numérique
 - Flexibles et expressives : pour exploiter au mieux les données
 - Réutilisables et automatisables : pour focaliser les épidémiologistes sur les questions difficiles
 - Étape indispensable pour utiliser des méthodes d'apprentissage automatique

Conclusions

- L'approches suivies
 - Représenter de manière riche et exploitable les données sous forme de parcours de soin
 - Dimension Ontologique
 - Dimension Temporelles
 - Formaliser les études épidémiologiques :
 - Proposition d'un langage dédié permettant la représentation bout en bout d'une étude
 - Opérationnaliser la réalisation des études en combinant
 - Raisonnement automatique pour les capacités de raisonnement
 - Méthodes algorithmiques efficaces

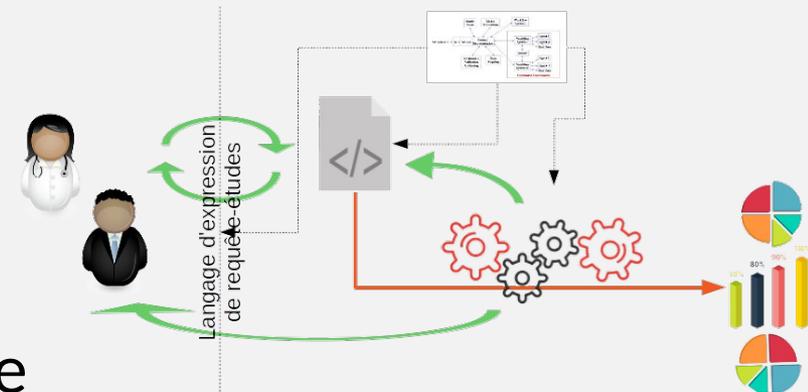


Conclusions

- Web sémantique
 - Exploite les connaissances formalisées du domaine
 - Facilite la réutilisation des définitions de phénotype
 - Une piste intéressante pour combler le fossé sémantique :
 - Ontologie Mediated Query Answering adapté à des données temporelles
 - Combinaison Modèles temporels + Web sémantique fertile
 - Thèse de Q. Cao, LITIS et Icube (Chroniques également)
- Une perspective intéressante pour l'exploitation des données du SNDS
 - Modélisation du SNDS à poursuivre
 - Champs d'expérimentation pour les technologies du web sémantique : **jeux synthétiques possibles**
 - Perspectives de développements d'outils au sein du HdH

Perspective 1 : Vers la (semi-)automatisation des études ?

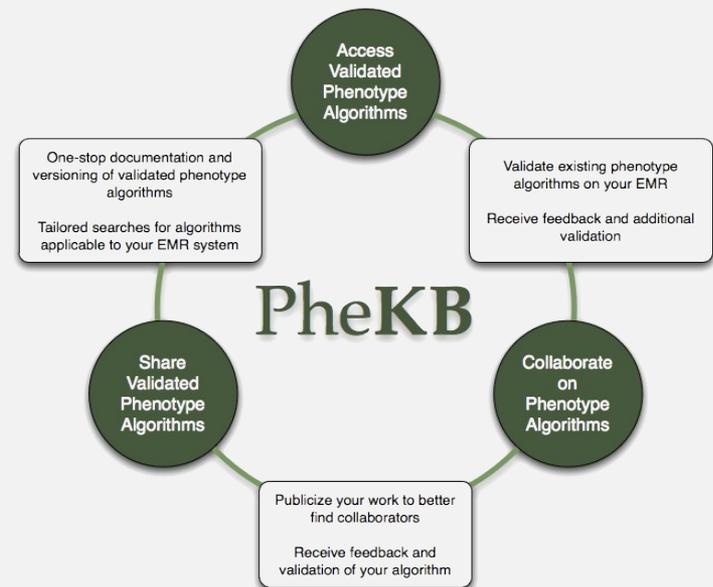
- Pourquoi (semi-)automatiser ?
 - Concentrer les épidémiologistes sur les tâches à valeur ajoutée
 - Faciliter la **réutilisation** et la reproductibilité des études
- Automatisation des études : Déterminer les workflows qui répondent le mieux à une étude
 - Ajout de variables dans les workflow
 - Optimiser les workflows : *Automatic Data Science*



Automating Data Science, Tijl De Bie, Luc De Raedt, Holger H. Hoos, Padhraic Smyth, Dagstuhl Seminar, 2018

Perspective 2 : PheKB pour le SNDS

- Structurer les connaissances du SNDS
 - Partager et réutiliser des phénotypes
 - Donner la possibilité de définir des phénotypes sur des parcours de soins
 - Contraintes ontologiques et temporelles
- connexion forte avec INOVIVE



Merci de votre attention

Remerciements

REPERES

- Pr. E. Oger
- A. Happe
- PharmD. E. Polard
- N. Le Meur

- Equipes Inria/IRISA

D. Gross-Amblard, O. Dameron, J. Bakalara, Y. Dauxais,