Common Data Elements.Artificial Intelligence CDE.AI

Nicolas Garcelon

- *Institut imagine – Plateforme data science*

Guillaume Assié, Anne-Sophie Jannot,
Arnaud Sandrin, Xavier Tannier

BNDMR
Banque Nationale de Donnée
Maladies Rares
bndmr.fr

filière de santé
maladies rares

FIRENDO
FILIÈRE MALADIES RARES ENDOCRINIENNES

Filnemus
Filière Neuromusculaire

fai2r

SORBONNE UNIVERSITÉ

institut imagine
GUÉRIR LES MALADIES GÉNÉTIQUES

Université de Paris

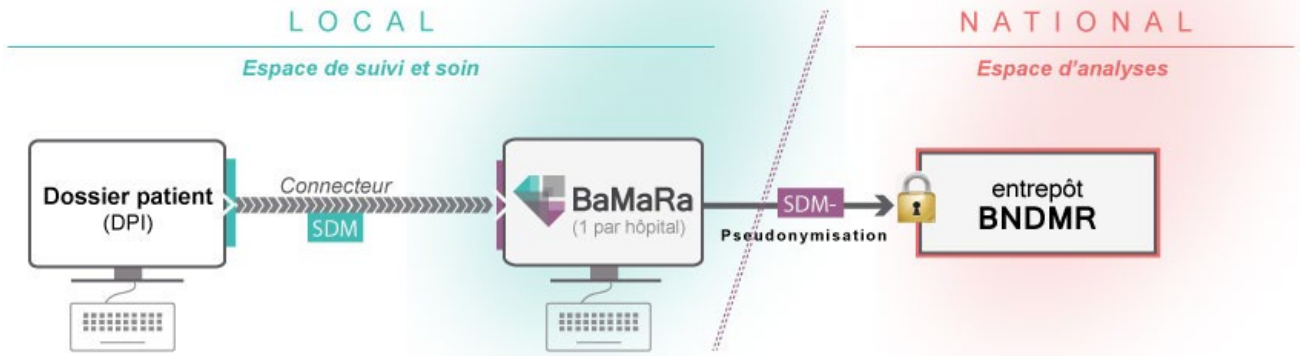
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DES AFFAIRES SOCIALES, DE LA SANTÉ ET DES DROITS DES FEMMES

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Contexte

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

- Banque Nationale de Données Maladies Rares (BNDMR) : 60 items à saisir (dont ~20 obligatoires), 1,2M de patients



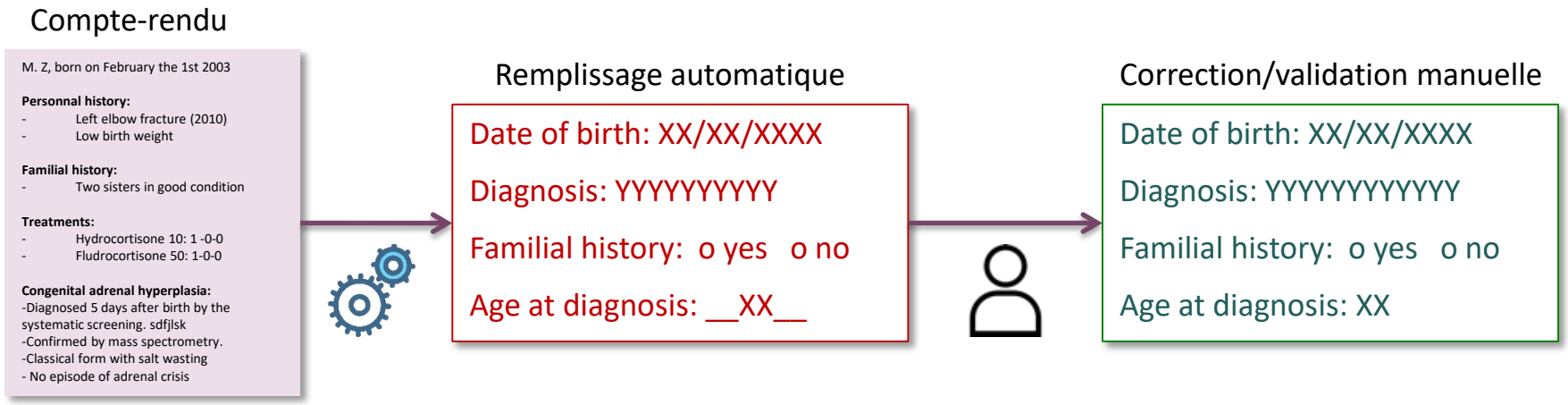
- Difficulté pour :
  - Exhaustivité des données à des fins de recherche (ex: 60% patients n'ont pas de signe clinique en HPO car donnée non obligatoire)
  - Qualité des données
  - Charge clinique : temps de saisie (création d'un dossier = 4 minutes par patient)

- Même problème pour toutes les bases recherches / registres

Objectifs

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

- Faciliter et accélérer la saisie des données pour la BNDMR
  - A partir du compte rendu d'un patient



- Reproductibilité sur une autre base :
  - Complications des glucocorticoïdes

Méthode

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

Rare diseases networks / expert centers

fai2r

FIRENDO
FILIÈRE MALADIES RARES ENDOCRINIENNES

Filnemus
Filière Neuromusculaire

>2000 unstructured medical records

Rare Diseases Common Data Elements (RD-CDE)

CDE.ai

Pre-filled RD-CDE collection form

Human Validation

BNDMR
Rare Diseases registry

BNDMR
Banque Nationale de Données Maladies Rares

BaMaRa

France Cohortes

Inserm

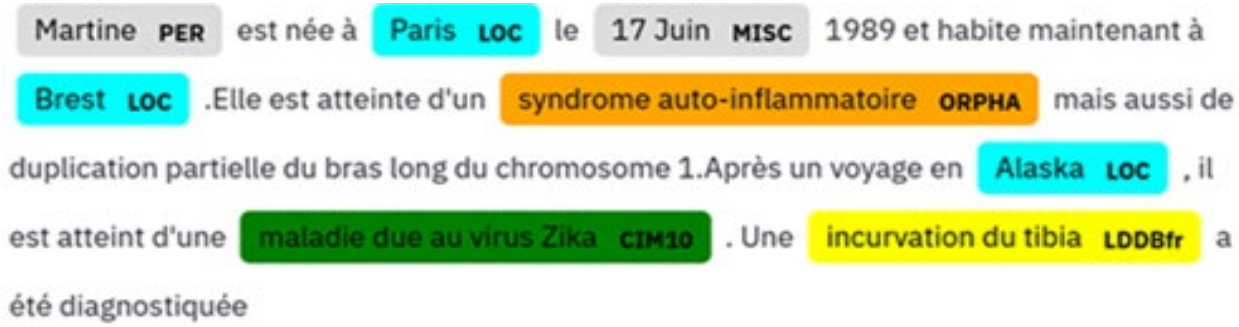
Secured data flow

Méthode pour l'extraction

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

2 approches :

- Extraction par un modèle d'apprentissage profond
  - Annotation manuelle de comptes rendus pour « apprendre » la représentation d'une variable.
  - Application du modèle sur un nouveau texte pour identifier cette variable



- Extraction à partir d'un pattern (expression régulière)
  - Définition d'un pattern pour extraire une donnée homogène :

```
N[ée]e? *le *([0-9][0-9][-/][0-9][0-9][-/][0-9][0-9][0-9][0-9])
```

The Glucocorticoids Complication database

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

Glucococorticoids Complications (Cushing syndrome)

- Endogenous
- Exogenous

Source: Wikipedia

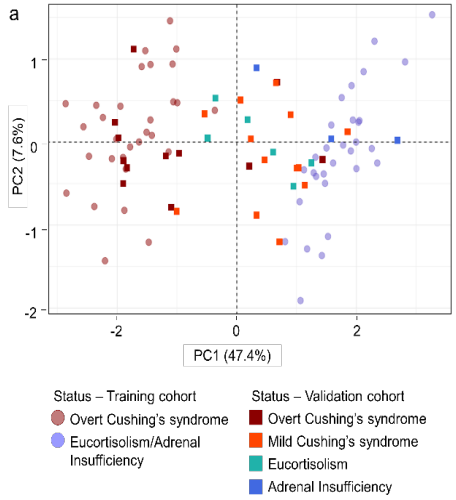
- Weight gain with faciotroncular adiposity
- Cutaneous and muscular atrophy
- Hypertension
- Diabetes mellitus
- Osteoporosis
- ….

A need for molecular markers of Glucococorticoids Complications

- Hormone assays not usable for exogenous Cushing
- Hormone assays do not predict individual complications
- Mild Cushing: should we treat or not?

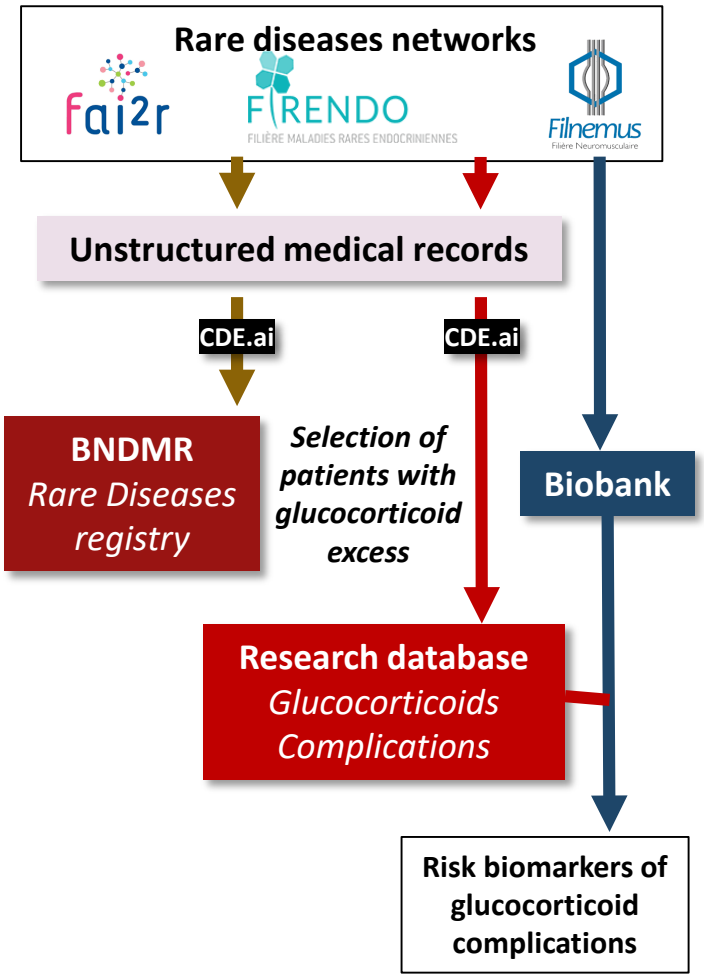
Markers:

- GR variants
- Whole blood methylome



Armignacco et al, unpublished

GC-CDE Flowchart



Inserm

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Méthode

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

Rare diseases networks / expert centers

fai2r

FIRENDO
FILIÈRE MALADIES RARES ENDOCRINIENNES

Filnemus
Filière Neuromusculaire

>2000 unstructured medical records

Rare Diseases Common Data Elements (RD-CDE)

Glucocorticoids Complications Common Data Elements (GC-CDE)

CDE.ai

Pre-filled RD-CDE collection form

Human Validation

BNDMR
Rare Diseases registry

Pre-filled GC-CDE collection form

Human Validation

Research database
Glucocorticoids Complications

BNDMR
Banque Nationale de Données Maladies Rares

BaMaRa

France Cohortes

Inserm

Secured data flow

Evaluation

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

3 Niveaux d'évaluation

Annotation des comptes rendus

Martine PER est née à Paris LOC le 17 Juin MISC 1989 et habite maintenant à Brest LOC .Elle est atteinte d'un syndrome auto-inflammatoire ORPHA mais aussi de duplication partielle du bras long du chromosome 1.Après un voyage en Alaska LOC , il est atteint d'une maladie due au virus Zika CIM10 . Une incurvation du tibia LDDBfr a été diagnostiquée

Précision de l'extraction des concepts

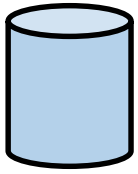
Remplissage du formulaire

Date de naissance : **17/06/1989**
Diagnostic : **syndrome auto inflammatoire**
Histoire familiale: o oui **X** non
Age au diagnostic : **6 ans**

Qualité du remplissage des formulaires

Collecte de la cohorte

RD-CDE et GC-CDE enregistrements

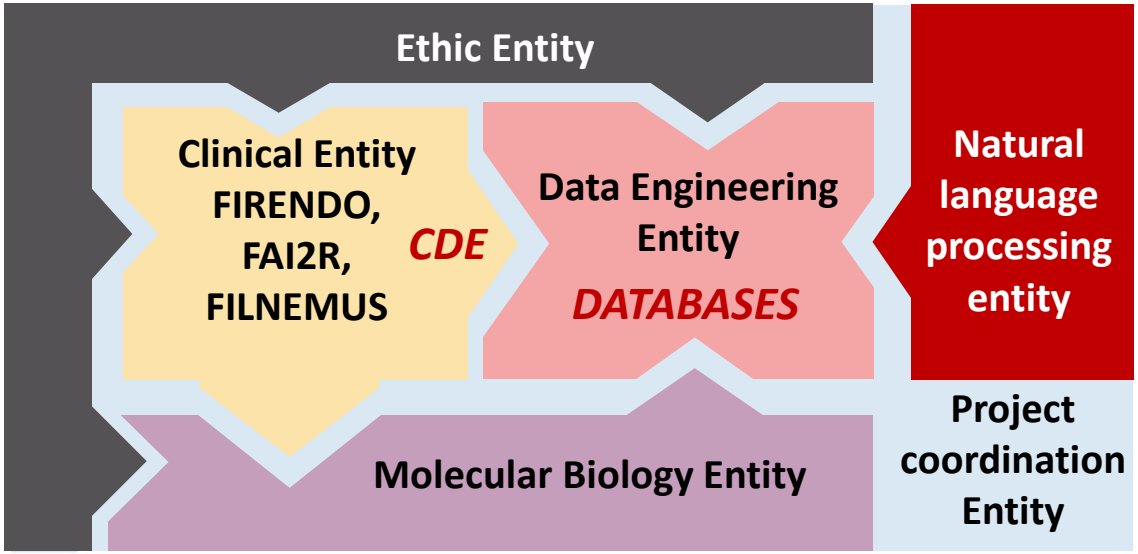


Base de données

Gain de temps

Consortium

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr



Un consortium multidisciplinaire

Team	Name	Affiliation
1	Assié G	Endocrine Dpt (Cohin APHP)
2	Jannot AS	Med Comput (HEGP APHP)
3	Sandrin A	BNDMR
4	Tannier X	LIMICS (Sorbonne U)
5	Garcelon N	Imagine, institut des maladies génétiques
6	Bertherat J	FIRENDO (APHP)
7	Belot A	FAI2R (HCL)
8	Attarian S	FILNEMUS (APHM)
9	Parfait B	CRB (Cochin, APHP)
10	Ragazzon B	Institut Cochin (INSERM)
11	Mamzer MF	Ethics lab ETREs (U Paris)

Conclusion

BNDMR
Banque Nationale de Données Maladies Rares
bndmr.fr

- **Conception et entraînement de CDE.AI**
  - Basé sur le **Rare Disease – Common Data Element**
  - Utilisation de la BNDMR, **infrastructure nationale** pour les maladies rares
  - Avec des spécialistes du **Traitement Automatique du Langage**
- **CDE.AI pour accélérer les bases recherche**
  - Sur les complications des glucocorticoïdes
  - Collecter des informations de biologie moléculaire avec GC-CDE
  - Réutilisable pour toutes les maladies rares et les questions de recherche.
- **L'importance de France Cohortes pour**
  - Héberger les dossiers médicaux pour l'entraînement de CDE.AI
  - Héberger la base de données des complications liées aux glucocorticoïdes
  - Favoriser l'avenir des collections de données sur les maladies rares avec CDE.AI
- **Impliquer les patients pour**
  - La réutilisation des données médicales hospitalières