Journée annuelle du CRAIM

L'EDS et le DIM

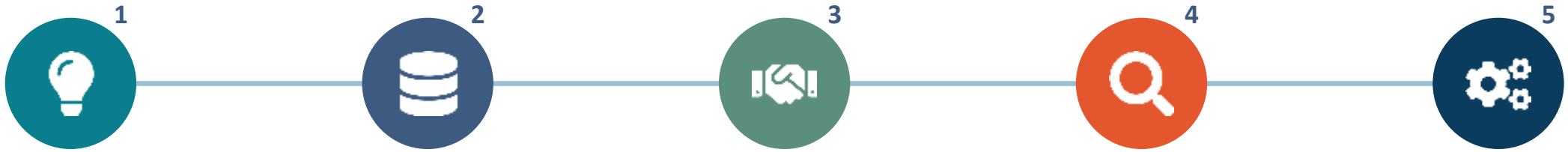
Comprendre l'entrepôt de données de santé hospitalier et ce qu'un DIM peut en faire



Paul-Antoine Beaudoin – Ingénieur EDS CHU Grenoble Alpes – pabeaudoin@chu-grenoble.fr
Frédéric Olive – Médecin DIM CHU Grenoble Alpes – folive@chu-grenoble.fr

PLAN

5 étapes pour comprendre et utiliser les EDS



Comprendre

Ce qu'est un EDS, ses données, son cadre CNIL, le rôle du DIM

Construire

Alimenter l'EDS
Exemple des flux PMSI

Définir

Traduire les concepts métiers avec les cliniciens

Chercher

Aller au-delà du PMSI : exploiter les comptes rendus

Outiller

Outils actuels, limites et apport de l'IA

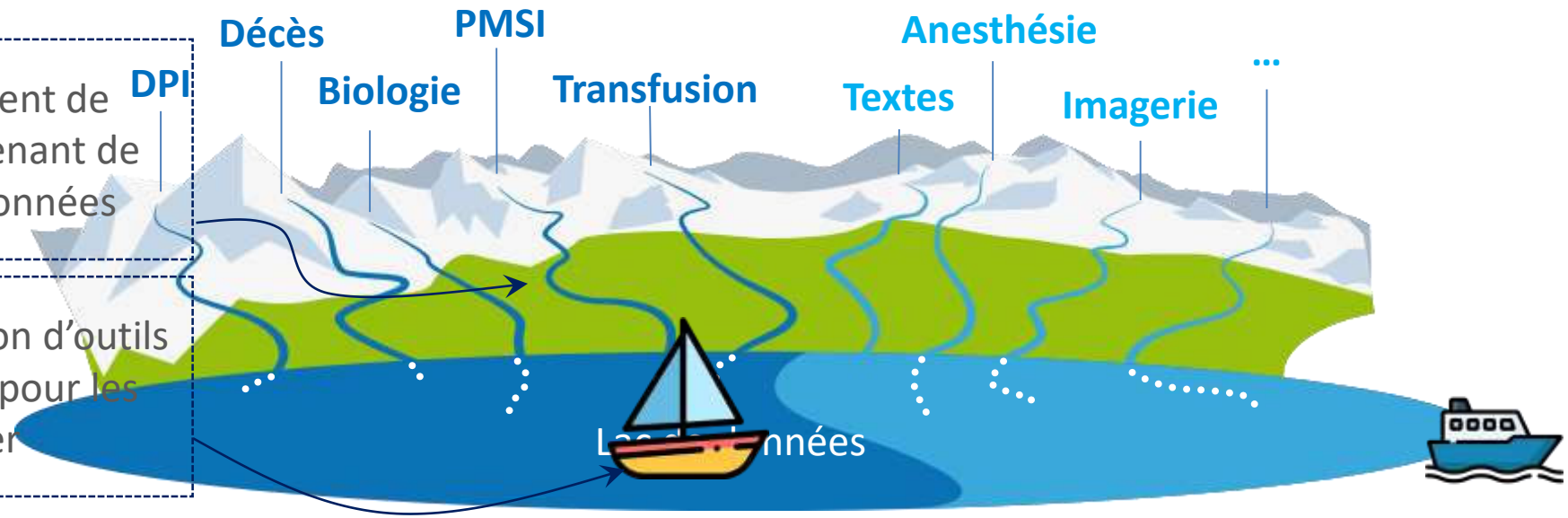
Question fil rouge : qu'apporte le DIM à l'EDS, et qu'apporte l'EDS au DIM ?

Qu'est-ce qu'un EDS ?

Un entrepôt de données de santé réutilise les données du soin pour d'autres usages que le soin : recherche , pilotage, qualité des soins.

Rassemblement de données provenant de sources cloisonnées

Mise à disposition d'outils de traitement pour les exploiter



1,1 M patients | **12 M séjours** | **752 k images** | **20 M CRs**

EDS encadré par autorisation CNIL depuis **Octobre 2019**
Responsable de traitement CHU Grenoble Alpes

Comment circulent les données ?

Les données quittent les logiciels de soin, sont préparées, puis stockées dans l'entrepôt.

Sources

DPI, PMSI, biologie, imagerie, prescriptions

→

Intégration

Flux automatiques, mise en forme, standardisation, liens entre les bases de données

→

Entrepôt

Stockage centralisé et pseudonymisé

→

Exploitation

Requêtes, extractions, analyses

Trois façons d'exploiter l'EDS

Pré-screening

Interroger et compter : combien de patients, quelles cohortes ?

Extraction

Obtenir un jeu de données pour un projet autorisé

Analyse

Fouille de textes, statistiques, intelligence artificielle dans un espace sécurisé, création de tableaux de bord

Quelles données dans l'EDS ?

Trois grandes familles de données alimentent l'entrepôt du CHUGA.

Données structurées

Codées, faciles à interroger

- PMSI
- Biologie : résultats d'examens
- Médicaments : prescriptions et administrations
- Administratif : mouvements et séjours
- …

Texte libre

Riche, plus difficile à interroger

- Comptes rendus d'hospitalisation
- Comptes rendus opératoires
- Courriers et consultations
- Comptes rendus d'imagerie

Imagerie médicale

- Images DICOM
  - CT scan
  - IRM
  - RX …

Plusieurs sources complémentaires permettent de remonter et consolider une même information.

CNIL et gouvernance : le cadre

Chaque usage de l'EDS est cadré, justifié et tracé.



Un cadre réglementaire

- RGPD et loi Informatique et Libertés
- Référentiel CNIL dédié aux entrepôts de données de santé
- **Autorisations CNIL spécifiques**



Droits des patients

- Information des patients
- Droit d'opposition
- Données pseudonymisées, accès habilité



Une gouvernance

- Comité d’éthique
- Évaluation de chaque projet avant extraction (COMOR)
- Traçabilité des accès

Sans autorisation et sans cadre, pas d'extraction : la confiance des patients est la condition de l'EDS.

La place du DIM dans l’EDS

Le DIM ne se contente pas d'utiliser l'EDS : il contribue à le concevoir et à le piloter.

Expertise sur l’information médicale:

- Connaissance des nomenclatures, du codage et de ses règles
- Proximité du terrain où s’élabore la production de l’information médicale
- A chaque étape d’un projet: élaboration, processus et cohérence des résultats

Interface : entre le soin, l'information médicale et les ingénieurs

Gouvernance : Participation aux instances et à la validation des projets (orientation des projets)

Principes d'autonomie

Dans le cadre d'un EDS indépendant du DIM

Cas d'utilisation 1

1

Poser ses propres questions sans lancer un projet d'ingénierie à chaque fois

- Réponses aux demandes d'analyses faites au DIM : fast-track de demandes pour thèses, projets de recherche, activité, indicateurs …
- Qualité, optimisation du codage: intégration d'outils déjà élaborés (processus permanent entre médecin et ingénieur au sein du DIM de creation de CQ

2

Pouvoir retourner aux dossiers médicaux patients

- Condition MAJEURE, encore sous réserve d'autorisation CNIL
- Vérification de résultats, simplicité des contrôles-recodage

3

En bénéficiant des capacités d'un EDS

- Volume des données extraites et retraitées
- Automatisation de l'alimentation
- Utilisation d'algorithmes pour le traitement des faux positifs

4

En bénéficiant d'un environnement sécurisé réglementaire

- Encadré par autorisation CNIL – oct 2019

Comprendre le métier pour alimenter l’EDS

L’exemple des données PMSI avec le DIM

Séjour et soins

Le patient est pris en charge

→

Codage par le DIM

Diagnostics CIM-10, actes CCAM

→

Groupage

Production du GHM et du RSA

→

Flux vers l'EDS

Données PMSI intégrées à l'entrepôt

Ce que le DIM apporte

- Le codage est finalisé après la sortie : le DIM sait quand la donnée est fiable
- Formation des ingénieurs de l’EDS au PMSI
- Collaboration sur les évolutions (Tables, etc…)

Ce que l'EDS gagne

- Données codées reliées aux comptes rendus et à la biologie
- Séjours retrouvables par diagnostic, acte ou GHM
- Base fiable pour constituer des cohortes

Du besoin clinique à la requête

Trois métiers autour d'une même question : chacun détient une partie de la réponse.

Cas d'utilisation 2



Le clinicien pose la question et reconnaît le bon patient



Le DIM fixe les définitions : quels codes, quels termes, quelle période



L'ingénieur traduit la définition en requête ou en algorithme

Une boucle de travail

1. Question clinique
2. Définition partagée
3. Requête
4. Relecture et affinage

On recommence jusqu'à validation du résultat

Sans définition partagée, deux requêtes « identiques » donnent deux résultats différents.

Deux exemples concrets de définition

Rachis dégénératif et hypotonie périphérique : des termes cliniques qui n'ont pas de définition évidente.

Rachis dégénératif

Questions à trancher

- Quelles pathologies inclure : arthrose, hernie discale, canal lombaire étroit ? -> en collaboration avec le clinicien
- Quels actes CCAM ?

Hypotonie périphérique

Questions à trancher

- Un terme clinique sans code unique : où le retrouver ?
- Comment la distinguer de l'hypotonie centrale ?
- Quelle définition : examen clinique, causes neuromusculaires, mots du compte rendu ?

Une constante : il faut bien définir ce que l’on cherche !

Quand le PMSI ne suffit pas

Le code dit l'essentiel du séjour ; le texte raconte le patient.

Typologie	Exemple	Stratégie
Recherches infracodes	AVC cérebelleux / avec thrombus flottant	**Recherche combinée PMSI puis textuelle**
Absence de code CIM10 (ou CCAM) satisfaisant	Épidurite rachidienne	**Recherche textuelle**
Variabilité de codage, precisions existantes mais peu opérantes ou complexes	Dysfonctionnements neurovésicaux, epilepsies, … (on peut plus ou moins prédire l'exhaustivité et la qualité d'un code CIM10 ou CCAM …)	**Selon la recherche**
Précisions non présentes sur la CIM10	Anapath, stade (cancers,autres), latéralité, … *Ex: cancer colo-rectal des patients de plus de 80 ans non opérés (pmsi) mais selon une stratégie "wait and watch"*	**Selon la recherche**

Information absente ou imprécise dans le codage : il faut aller lire les comptes rendus textuels.

La recherche textuelle : des pièges

Chercher un mot n'est pas chercher un diagnostic.

Piège	Exemple	Conséquence
Homonymie	« goutte » : la maladie, le goutte-à-goutte ou un collyre. « IMC » : indice de masse corporelle ou infirmité motrice cérébrale. BAV, SAM … terme identique au médecin traitant ou à l'adresse (Dr Prader, rue du muguet …)	**Faux positif**
Négation / Ambiguïté	« pas de thrombus flottant à l'échographie », «suspicion de …», "recherche de …"	**Faux positif**
Contexte	« père goutteux », « AVC en 2012 » : famille ou antécédent ancien.	**Faux positif**
Formulations	« thrombus flottant », « thrombus mobile », « thr. flottant » : plusieurs façons de dire la même chose.	**Faux négatif**

Faux positif : le mot est là, pas le diagnostic.

Faux négatif : le diagnostic est là, sous d'autres mots.

Messages: avoir une bonne connaissance du lexique utilisé par les soignants / on peut minimiser les faux+ par un découpage des CR en catégories et limiter la recherche sur ces catégories

Outils actuels : utiles mais limités

Quickview au CHUGA permet de chercher dans les textes, au prix de nombreux faux positifs.

Quickview ou d'autres logiciels du marché

Usages

- Interroger les comptes rendus par mots-clés
- Repérer vite des séjours candidats
- Gagner du temps sur un premier tri grâce à une interface de visualisation
- Accéder à un historique important sur des documents multiples avec classifications

Limites

- Nombreux faux positifs : homonymes, négations, antécédents
- Ergonomie logicielle à améliorer (script, interface)
- Besoin couvert en partie seulement (CQ, optimisation codage)

La recherche textuelle brute par mots-clés atteint certaines limites, il y a un besoin de développement d'autres algorithmes

Vers l'IA : l'exemple d'AMEDEE.IA

Un modèle de langue lit la phrase entière : il peut distinguer ce que les mots-clés confondent.

Ce que change un modèle de langue

- Lit la phrase entière, pas seulement le mot
- Distingue négation, antécédent, famille
- Reconnaît synonymes et formulations

Exemple sur le rachis dégénératif

- Définition du concept validée par le DIM et le clinicien
- Recherche des séjours concernés dans les comptes rendus
- Disctinction rachis traumatique / métastatique /dégénératif

Ce qu'il reste à construire

Ergonomie
un outil simple d'usage pour le DIM

Intégration
branché à l'EDS et aux outils existants

Appropriation métier
formation, validation, usage quotidien

Tout algorithme d’IA doit être validé sur données réelles, exemple du projet PARTAGES en collaboration avec le DIM

Trois messages clés

1

S'impliquer dès la construction

Les DIM doivent participer à la construction des EDS à l'échelle des GHT pour obtenir une autonomie d'exploitation.

2

Des synergies dans les deux sens

De l'EDS vers le DIM : données et outils. Du DIM vers l'EDS : flux PMSI, terminologies, expertise.

3

Consolider les outils

Les outils actuels répondent partiellement aux besoins de recherche textuelle : à consolider dans le projet EDS-D2H en AURA.

Paul-Antoine Beaudoin – Ingénieur EDS CHU Grenoble Alpes – pabeaudoin@chu-grenoble.fr
Frédéric Olive – Médecin DIM CHU Grenoble Alpes – folive@chu-grenoble.fr