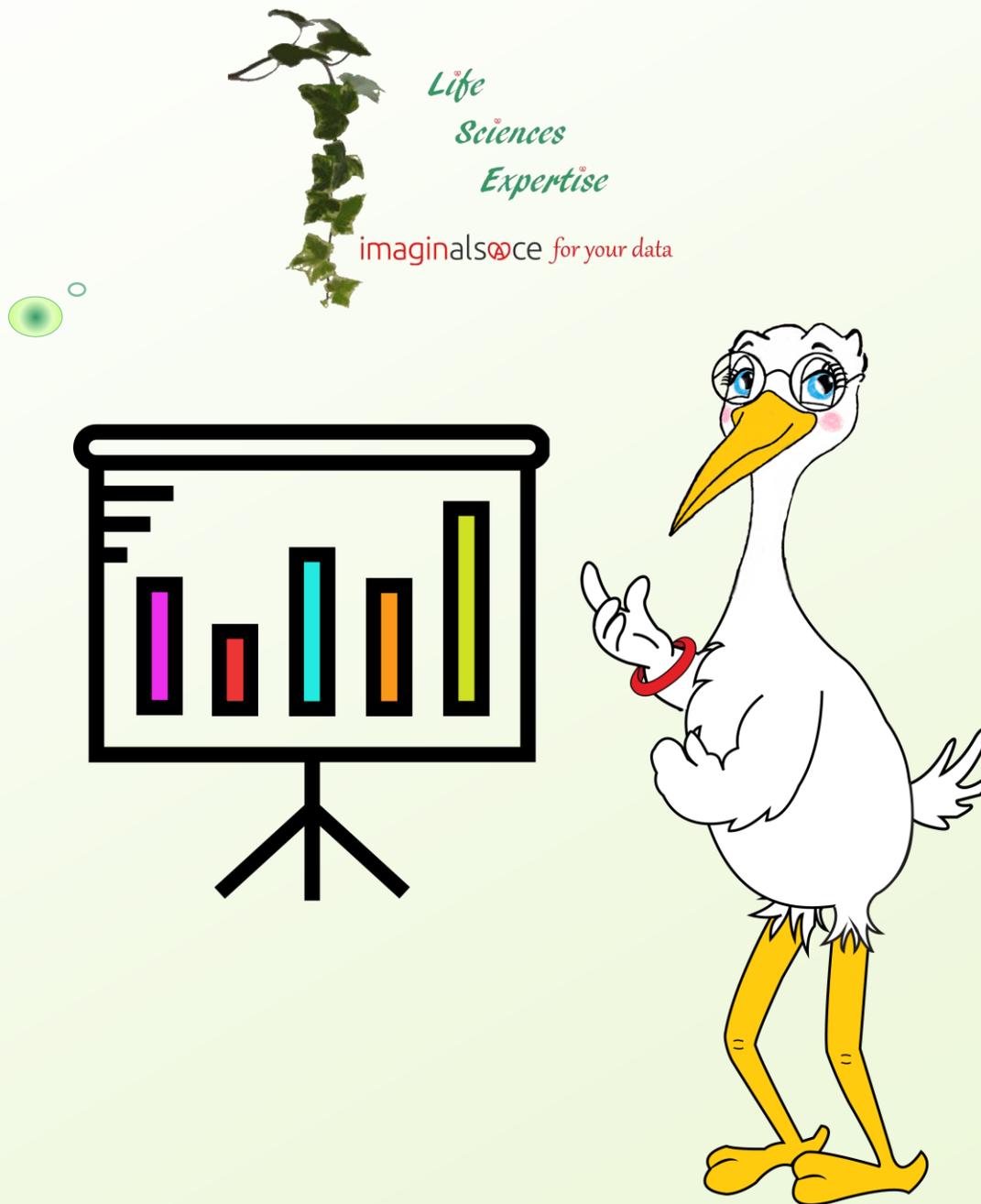


Vulgarisation sur les entrepôts de données

Isabelle Laugel & Pierre-Yves Lastic
Séminaire GUF CDISC - Tours
15 Juin 2023



Agenda



- Définition d'un entrepôt de données
- Recommandations CNIL
- Raisons d'utiliser un entrepôt de données
- Health Data Hub
- CDISC
- Exemples (Hubert Méchin)

Vulgarisation sur les entrepôts de données

✓ Définition



Entrepôts de données - Definition

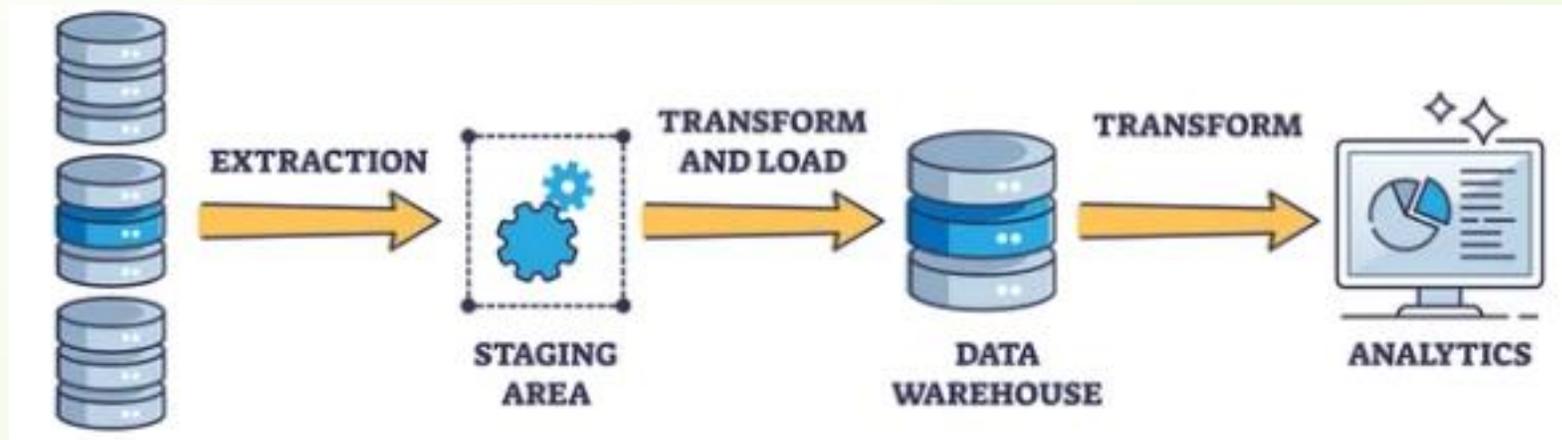
- ▶ Un entrepôt de données rassemble des données de sources différentes permettant d'effectuer des analyses et apporter une aide à la décision
 - ✓ Aussi nommé « Base de données décisionnelle » ou Data Warehouse (DWH)
- ▶ Les entrepôts de données sont exploités par des outils d'analyses statistiques, de « reporting » ou d'intelligence artificielle
- ▶ De plus, ils permettent également d'optimiser la performance des bases de données opérationnelles
 - ✓ les opérations gourmandes en ressources étant effectuées au niveau des entrepôts de données plutôt qu'au niveau des bases opérationnelles

Entrepôts de données - Contenu

- ▶ Les Entrepôts de données sont mis en place pour répondre aux besoins d'une organisation donnée, ils n'existent donc pas de modèle standard applicable à tous les entrepôts
 - ✓ Les Entrepôts de données contiennent généralement des données de formats différents (SAS, Excel, XML, ...)
- ▶ Pour intégrer toutes les données dans des analyses, il est donc important de gérer tant le contenu que la structure de tous ces fichiers
 - ✓ Une standardisation des métadonnées est donc nécessaire, ainsi que des contrôles qualité, afin de garantir l'intégrité des données

Entrepôts de données - Gestion

- ▶ Avant d'être stockées dans l'entrepôt, les données doivent être transformées et validées, via des outils ETL (Extract, Transform and Load)
- ▶ Les outils d'analyse des données sont en évolution permanente afin de rester au plus prêt des besoins des utilisateurs (statistiques, reports, dashboards, prévisions, ...)



Protéger les données personnelles, accompagner l'innovation, préserver les libertés individuelles

[MA CONFORMITÉ AU RGPD](#) | [THÉMATIQUES](#) | [TECHNOLOGIES](#) | [TEXTES OFFICIELS](#) | [LA CNIL](#) | 



 > [La CNIL adopte un référentiel sur les entrepôts de données de santé](#)



La CNIL adopte un référentiel sur les entrepôts de données de santé

17 novembre 2021

La création d'un entrepôt de données de santé nécessite le respect de certaines formalités. La CNIL a adopté un nouveau référentiel pour simplifier les procédures en proposant un cadre adapté aux pratiques.

Types d'entrepôts couverts par le référentiel

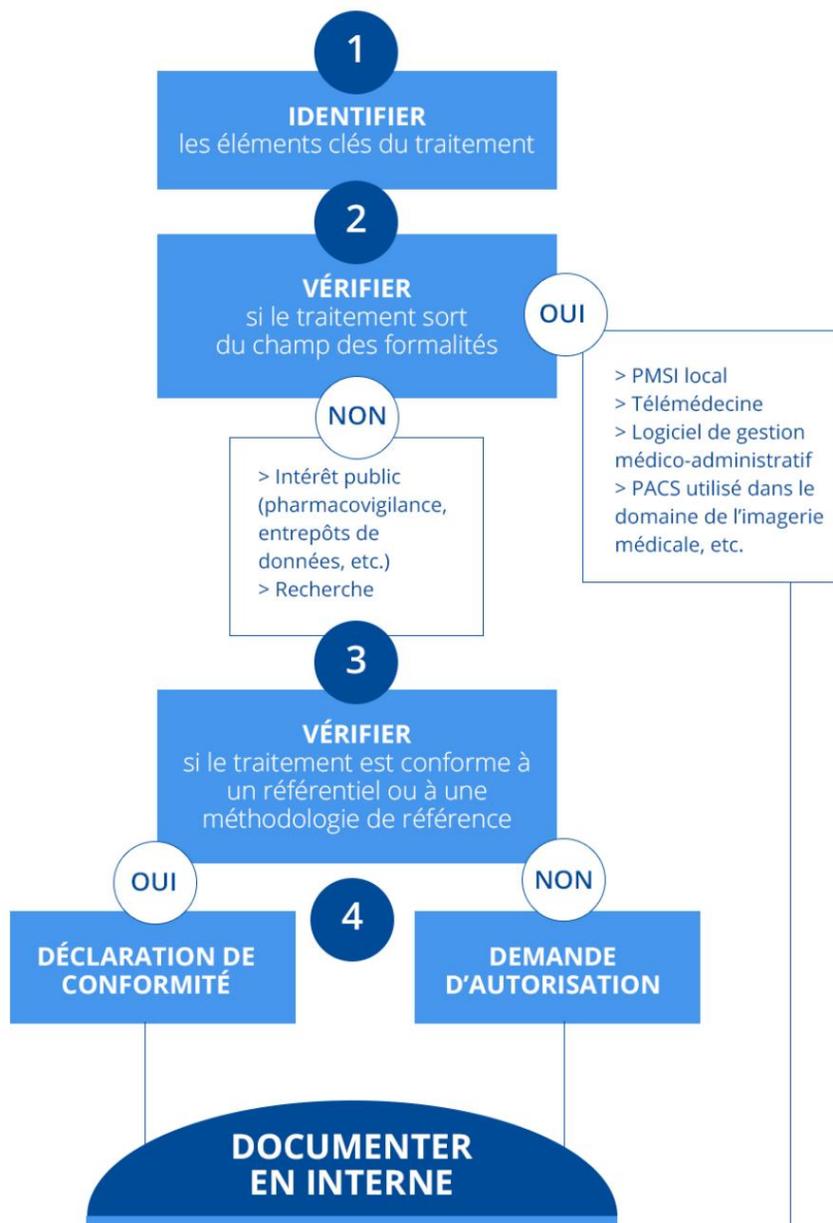
- ▶ Le référentiel entrepôt ne s'applique qu'aux entrepôts de données de santé reposant sur l'exercice d'une mission d'intérêt public, au sens de l'article 6.1.e du RGPD.
 - ✓ Un établissement privé de soins mettant en œuvre un entrepôt de données de santé nécessaire à l'exécution d'une mission d'intérêt public pourra être couvert par le référentiel.
 - ✓ Ce n'est pas le cas des entrepôts mis en œuvre par des entreprises privées à des fins de recherche sans exécution d'une mission d'intérêt public => ~~pharma~~ !
- ▶ Concernant les entrepôts de données de santé qui ne répondent pas aux critères du référentiel, les organismes doivent solliciter la CNIL afin d'obtenir une autorisation individuelle préalablement à la mise en œuvre de leur entrepôt.

○ Périmètre du référentiel entrepôt

- ▶ Seule la constitution de l'entrepôt est couverte par l'engagement de conformité au référentiel.
 - ✓ La réutilisation des données de l'entrepôt dans le cadre de recherches, d'études ou d'évaluations dans le domaine de la santé doit faire l'objet des formalités à part
 - ✓ Exemples : engagement de conformité aux méthodologies de référence ou demande d'autorisation recherche.



Quelles formalités pour mon traitement?



DOCUMENTER EN INTERNE

- > **Renseigner** le registre des traitements
- > **Mener** une analyse d'impact pour les traitements le justifiant
- > **Vérifier** l'effectivité de l'information délivrée aux personnes concernées et l'effectivité de leurs droits
- > **Formaliser** le rôle et les responsabilités des acteurs : responsables de traitements, co-responsables de traitements et sous-traitants. (ex : hébergeurs de données de santé)
- > **Renseigner** les actions menées pour garantir la sécurité des données
- > Etc.

Méthodologies de référence

- ▶ MR-001 RECHERCHES DANS LE DOMAINE DE LA SANTÉ AVEC RECUEIL DU CONSENTEMENT
 - ✓ Essais cliniques
- ▶ MR-002 ÉTUDES NON INTERVENTIONNELLES DE PERFORMANCES CONCERNANT LES DISPOSITIFS MÉDICAUX DE DIAGNOSTIC IN VITRO
- ▶ MR-003 RECHERCHES DANS LE DOMAINE DE LA SANTÉ SANS RECUEIL DU CONSENTEMENT
 - ✓ Recherches non interventionnelles et essais cliniques de médicaments par grappe. Intérêt public.
- ▶ MR-004 RECHERCHES N'IMPLIQUANT PAS LA PERSONNE HUMAINE, ÉTUDES ET ÉVALUATIONS DANS LE DOMAINE DE LA SANTÉ
 - ✓ Réutilisation de données. Intérêt public.
- ▶ MR-005 ÉTUDES NÉCESSITANT L'ACCÈS AUX DONNÉES DU PMSI ET/OU DES RPU PAR LES ÉTABLISSEMENTS DE SANTÉ ET LES FÉDÉRATIONS HOSPITALIÈRES
- ▶ MR-006 ÉTUDES NÉCESSITANT L'ACCÈS AUX DONNÉES DU PMSI PAR LES INDUSTRIELS DE SANTÉ

Analyse d'impact obligatoire?

- ▶ **Cas général :** analyse d'impact obligatoire si au moins deux des critères suivants sont réunis:
 - ✓ Le traitement comporte-t-il une évaluation relative à la personne concernée (ex : évaluation de l'état de santé) ?
 - ✓ Y a-t-il une décision automatique avec effet juridique ou affectant la personne de manière significative ?
 - ✓ Une surveillance systématique est-elle mise en place (ex : dispositif de géolocalisation utilisé pour surveiller des personnes âgées ou des nourrissons) ?
 - ✓ **Le traitement comporte-t-il des données sensibles (ex : données de santé) ?**
 - ✓ **Est-ce un traitement à grande échelle ?**
 - ✓ **Y a-t-il un croisement de données ?**
 - ✓ **Le traitement concerne-t-il des personnes vulnérables (ex : patients, personnes âgées, etc.) ?**
 - ✓ **S'agit-il d'un usage innovant ?**
 - ✓ Le traitement peut-il entraver l'exercice d'un droit ou l'exécution d'un contrat (ex : droit aux prestations sociales) ?
- ▶ **La constitution d'un entrepôt de données nécessite une étude d'impact!**



GDPR CHECKLIST TO FEED YOUR PIA

1. **General/Global Checks:**
 - Relevant policies & procedures, e.g. standardized content for Patient Informed Consent, Cybersecurity policies, Record Keeping policies
2. **Study specific checks:**
 - GDPR compliance of Protocol, Case Report Forms, Data Management Manual, etc.
3. **Risk Assessment and mitigation plans**



**Privacy Impact Assessment
(PIA)**

TEMPLATES



CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

February 2018 edition

PIA, templates February 2018 edition

Contents

Foreword.....2

1 Study of the context: templates.....4

1.1 Overview of the processing.....4

Description of the processing under consideration.....4

Sector-specific standards applicable to the processing.....4

1.2 Data, processes and supporting assets.....4

Data description, recipients and storage durations.....4

Description of the processes and supporting assets.....4

2 Study of the fundamental principles: templates.....5

2.1 Assessment of the controls guaranteeing the proportionality and necessity of the processing.....5

Explanation and justification of purposes.....5

Explanation and justification of lawfulness.....5

Explanation and justification of data minimization.....6

Explanation and justification of data quality.....6

Explanation and justification of storage durations.....6

Assessment of the controls.....6

2.2 Assessment of controls protecting data subjects' rights.....7

Determination and description of the controls for information for the data subjects.....7

Determination and description of the controls for obtaining consent.....8

Determination and description of the controls for the rights of access and to data portability.....8

Determination and description of the controls for the rights to rectification and erasure.....10

Determination and description of the controls for the rights to restriction of processing and to object.....11

Determination and description of the controls applicable to processors.....11

Determination and description of the controls on transfer of data outside the European Union.....12

Assessment of the controls.....12

3 Study of data security risks: templates13

3.1 Assessment of security controls.....13

Description and assessment of controls implemented for treating the risks related to data security.....13

Description and assessment of general security controls.....15

Description and assessment of organizational controls (governance).....18

3.2 Risk assessment: potential privacy breaches.....20

Analysis and assessment of risks.....20

Assessment of the risks.....20

4 Validation of the PIA: templates21

4.1 Preparation of the material required for validation21

Elaboration of the synthesis regarding compliance with [GDPR] of the controls selected to ensure compliance with the fundamental principles.....21

Elaboration of the synthesis regarding compliance with good security practices of controls implemented for treating the risks related to data security.....22

Mapping of risks related to data security.....23

Elaboration of action plan.....24

Documentation of the advice of the person in charge of "Data Protection" aspects.....24

Documentation of the view of data subjects or their representatives.....24

4.2 Formal validation of the PIA25

Documentation of the validation.....25

Please note: these templates may have to be adapted, and should be used as a complement to the guide "PIA, methodology".

CNIL. 1

1 Study of the context: templates

1.1 Overview of the processing

Description of the processing under consideration

Description of the processing ¹	Clinical Trial Protocol
Processing purposes	id.
Processing stakes	id.
Controller	id.
Processor(s)	id.

Sector-specific standards applicable to the processing²

Standards applicable to the processing	Consideration
Clinical Trials Regulation	EU Legal Obligation
CNIL MR-001	France Legal Obligation

1.2 Data, processes and supporting assets

Data description, recipients and storage durations

Data types	Recipients	Storage duration
Clinical Trial Protocol	IT Policies	Clinical Trials Regulation
	Access Rights SOP	Records Mgt Policy

Description of the processes and supporting assets

[insert a diagram of data flows and a detailed description of the processes carried out]

Processes	Detailed description of the process	Data supporting assets
Clinical Trial Protocol		
Clinical Development Standard Operating Procedures		

2 Study of the fundamental principles: templates

2.1 Assessment of the controls guaranteeing the proportionality and necessity of the processing

Explanation and justification of purposes

Purposes	Legitimacy
Clinical Trial Protocol	Health Authority Authorization Ethics Committee Approval

Explanation and justification of lawfulness

Lawfulness criteria	Applicable	Justification
The data subject has given consent ³ to the processing of his or her personal data for one or more specific purposes	✓	Data/Study Dependant
Processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract		
Processing is necessary for compliance with a legal obligation to which the controller is subject	✓	
Processing is necessary in order to protect the vital interests of the data subject or of another natural person		
Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller		
Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child ⁴	✓	

LINKING PIA WITH QUALITY DOCUMENTS

Privacy Impact Assessment (PIA)

TEMPLATES



CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

February 2018 edition

PIA, templates February 2018 edition

Contents	
Foreword.....	2
1 Study of the context: templates.....	4
1.1 Overview of the processing.....	4
Description of the processing under consideration.....	4
Sector-specific standards applicable to the processing.....	4
1.2 Data, processes and supporting assets.....	4
Data description, recipients and storage durations.....	4
Description of the processes and supporting assets.....	4
2 Study of the fundamental principles: templates.....	5
2.1 Assessment of the controls guaranteeing the proportionality and necessity of the processing.....	5
Explanation and justification of purposes.....	5
Explanation and justification of lawfulness.....	5
Explanation and justification of data minimization.....	6
Explanation and justification of data quality.....	6
Explanation and justification of storage durations.....	6
Assessment of the controls.....	6
2.2 Assessment of controls protecting data subjects' rights.....	7
Determination and description of the controls for information for the data subjects.....	7
Determination and description of the controls for obtaining consent.....	8
Determination and description of the controls for the rights of access and to data portability.....	8
Determination and description of the controls for the rights to rectification and erasure.....	10
Determination and description of the controls for the rights to restriction of processing and to object.....	11
Determination and description of the controls applicable to processors.....	11
Determination and description of the controls on transfer of data outside the European Union.....	12
Assessment of the controls.....	12
3 Study of data security risks: templates.....	13
3.1 Assessment of security controls.....	13
Description and assessment of controls implemented for treating the risks related to data security.....	13
Description and assessment of general security controls.....	15
Description and assessment of organizational controls (governance).....	18
3.2 Risk assessment: potential privacy breaches.....	20
Analysis and assessment of risks.....	20
Assessment of the risks.....	20
4 Validation of the PIA: templates.....	21
4.1 Preparation of the material required for validation.....	21
Elaboration of the synthesis regarding compliance with [GDPR] of the controls selected to ensure compliance with the fundamental principles.....	21
Elaboration of the synthesis regarding compliance with good security practices of controls implemented for treating the risks related to data security.....	22
Mapping of risks related to data security.....	23
Elaboration of action plan.....	24
Documentation of the advice of the person in charge of "Data Protection" aspects.....	24
Documentation of the view of data subjects or their representatives.....	24
4.2 Formal validation of the PIA.....	25
Documentation of the validation.....	25

Please note: these templates may have to be adapted, and should be used as a complement to the guide "PIA, methodology".

CNIL. 1

PRODUCT LIFE-CYCLE PROCESSES

RESEARCH

Research

LABORATORY STUDY

Laboratory Animal Management

Test & Reference Compounds Management

Laboratory Study Conduct

Bioanalysis & Pharmacokinetics

MEDICAL AND CLINICAL

Clinical Development Post-Authorization Studies and Programs

Pharmacovigilance

Regulatory Submission & Maintenance

Scientific and Medical Information & Ethics

MANUFACTURING AND DISTRIBUTION

Design & Development

Technology Transfer

Facilities, Utilities and Equipment

Qualification & Validation

Manufacturing & Packaging

Control & Release

Supply Chain

Product Discontinuation

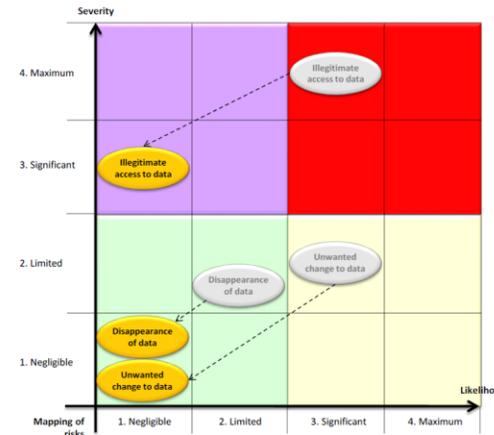


MEASURE RISKS & IMPROVE MITIGATION

Elaboration of the synthesis regarding compliance with good security practices of controls implemented for treating the risks related to data security

Controls implemented for treating the risks related to data security	Assessment
Controls bearing specifically on the data being processed	
Encryption	○○●
Anonymization	○○●
Data partitioning (in relation to the rest of the information system)	○○●
Logical access control	○○●
Traceability (logging)	○○●
Integrity monitoring	○○●
Archiving	○●○
Paper document security	○●○
General security controls regarding the system in which the processing is carried out	
Operating security	○○●
Clamping down on malicious software	○○●
Managing workstations	○○●
Website security	○○●
Backups	○○●
Maintenance	○●○
Security of computer channels (networks)	○○●
Monitoring	○○●
Physical access control	○●○
Hardware security	○○●
Avoiding sources of risk	○○●
Protecting against non-human sources of risks	○○●
Organizational controls (governance)	
Organization	○○●
Policy (management of rules)	○○●
Risk management	○○●
Project management	○●○
Management of incidents and data breaches	○○●
Personnel management	○●○
Relations with third parties	●○○
Supervision	○○●

Mapping of risks related to data security



Caption
 Symbol : ●●● ●○○ ○●○ ○○●
 Meaning : Non applicable Unsatisfactory Planned improvement Acceptable



Vulgarisation sur les entrepôts de données

- ✓ Pourquoi utiliser un entrepôt de données en recherche clinique ?



Entrepôts de données cliniques

- ▶ Les entrepôts de données cliniques sont intéressants pour de nombreuses raisons dans secteur de la recherche clinique
 - ✓ Analyses statistiques basées sur plusieurs études cliniques
 - ✓ Etudier les caractéristiques de certains groupes de patients à travers plusieurs études
 - ✓ Intégration de données de santé aux données d'une étude
 - ✓ Intégration de données émanant de réseaux sociaux, forums, objets connectés,...
 - ✓ Evaluation de nouvelles relations entre les causes et les effets de maladies
 - ✓ Identification de participants pour une étude clinique
 - ✓ Pharmacovigilance
 - ✓ Identification de signaux
 - ✓ Aide à la conception de nouveaux protocoles
 - ✓ Entraîner des modèles d'intelligence artificielle

Vulgarisation sur les entrepôts de données

- ✓ Health Data Hub



- ✓ <https://www.health-data-hub.fr/>

Health Data Hub

- ▶ Le « Health Data Hub » est une plateforme française de données de santé créée en 2019 pour faciliter le partage des données de santé issues de sources diverses dans le cadre de la recherche
- ▶ Les missions du « Health Data Hub » sont définies dans le code de la santé publique :
 - ✓ Mise à disposition des données issues du système national des données de santé
 - ✓ Innovation dans l'utilisation des données de santé
 - ✓ Information des patients
 - ✓ Elaboration des référentiels de la CNIL
 - ✓ Diffusion des normes de standardisation pour l'échange et l'exploitation des données de santé
 - ✓ Lancement d'appels à projets

Health Data Hub

- ▶ Les données du « Health Data Hub » peuvent être utilisées dans le cadre de projets de recherche (après accord)
- ▶ Il est important de noter que l'utilisation de données de la vie réelle est fortement encouragée par les autorités de santé, puisqu'elles permettent de diminuer le risque des patients, et d'évaluer une population plus large
- ▶ Cette plateforme de données met à disposition de la recherche les données de santé nécessaires au développement de modèles d'intelligence artificielle utilisés par exemple dans la recherche médicamenteuse, l'établissement de diagnostics, ou la prédiction de l'évolution de maladies

Vulgarisation sur les entrepôts de données

- ✓ CDISC et les entrepôts de données

cdisc

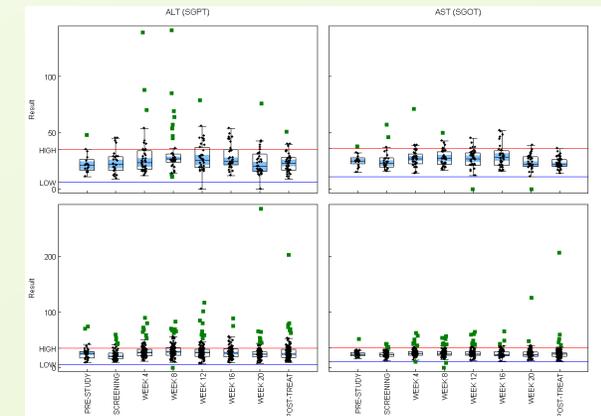
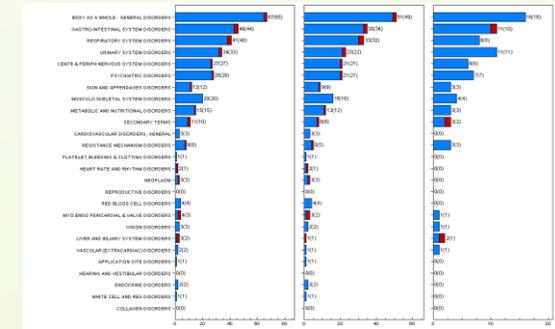
Entrepôts de données & CDISC

- Pour standardiser les données stockées dans un entrepôt de données, le modèle SDTM semble intéressant:

- ✓ Il permet une comparaison facile de données émanant de deux études distinctes

- ✓ Un programme créé pour un domaine peut souvent être adapté rapidement à un autre domaine (LB → VS par exemple)

- ✓ La présence de nombre variables -DY permet de s'affranchir des dates réelles (anonymisation)



Entrepôts de données & CDISC

- ▶ Néanmoins, l'utilisation du standard SDTM n'est pas simple, parce qu'il autorise de nombreuses souplesses :
 - ✓ Il existe beaucoup de variables définies par le promoteur
 - ✓ Certaines données peuvent se trouver dans deux domaines différentes
 - ECOG status: RS ou QS ?
 - ✓ Certains domaines ne sont pas encore clairement définis
 - Biomarqueurs ?
 - ✓ La méthode choisie pour coder les domaines en SDTM est parfois laissée libre
 - AE: une ligne par AE ou une ligne par changement de sévérité ?
 - ✓ De nombreuses variables se trouvent dans des tables SUPPxx

Entrepôts de données & CDISC

- ▶ Néanmoins, l'utilisation du standard SDTM n'est pas simple (suite):
 - ✓ Il est possible de créer des domaines spécifiques au promoteur
 - ✓ Le contenu des variables même si il est compatible SDTM n'est pas toujours identique ou comparable entre deux études
 - ✓ Le standard évolue avec le temps
 - Il faut aussi prendre en compte les changements au niveau des terminologies et des dictionnaires MedDRA et WHODD
 - ✓ Les données de santé sont plus souvent au standard HL7 FHIR (Fast Healthcare Interoperability Ressources)
 - Pas tout à fait compatible SDTM...

Entrepôts de données & CDISC

- ▶ Implémenter CDISC pour un entrepôt de données reste une bonne solution à condition de se préparer:
 - ✓ Choisir les domaines nécessaires
 - ✓ Définir des standards, même là où CDISC n'en prévoit pas ou n'en impose pas (variables utilisées, terminologies, ...)
 - ✓ Savoir si on planifie des métadonnées « 100% CDISC » ou « CDISC-like »
 - ✓ Déterminer comment gérer les variables de SUPPxx
 - ✓ Être conscient que les données provenant d'autres sources que des études cliniques ne sont pas validées de la même façon et peuvent admettre des éléments manquants

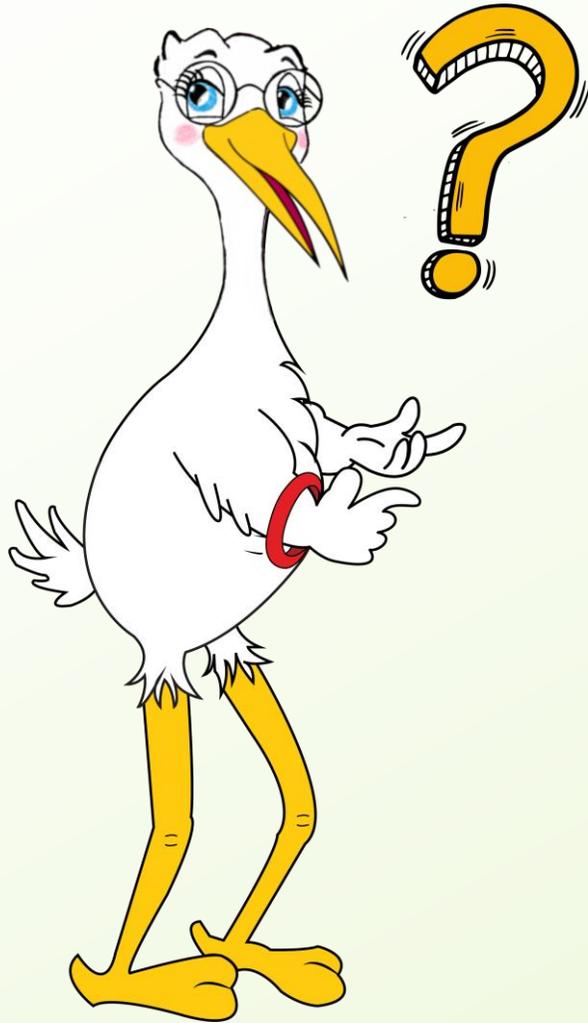
Entrepôts de données & CDISC

- ▶ Une fois que les règles d'implémentation ont été définies, il peut être judicieux de se servir de CORE (CDISC Open Rules Engine) pour valider les données

✓ <https://www.cdisc.org/core>



One Stop
Life Sciences
Solution



Life
Sciences
Expertise

imagin@lsoce for your data

Questions?

