



**HAL**  
open science

# Déploiement d'un Entrepôt de Données de Santé : exemple du Centre Hospitalier Régional de Metz-Thionville : aspects scientifiques, techniques et réglementaires

Vianney Guardiolle

## ► To cite this version:

Vianney Guardiolle. Déploiement d'un Entrepôt de Données de Santé : exemple du Centre Hospitalier Régional de Metz-Thionville : aspects scientifiques, techniques et réglementaires. Médecine humaine et pathologie. 2020. hal-03805943

**HAL Id: hal-03805943**

**<https://hal.univ-lorraine.fr/hal-03805943v1>**

Submitted on 7 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-thesesexercice-contact@univ-lorraine.fr](mailto:ddoc-thesesexercice-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

**THÈSE**  
pour obtenir le grade de  
**DOCTEUR EN MÉDECINE**

Présentée et soutenue publiquement  
dans le cadre du troisième cycle de Médecine Spécialisée  
par

**Vianney GUARDIOLLE**

le 14 septembre 2020

**Déploiement d'un Entrepôt de Données de Santé :**  
**exemple du Centre Hospitalier Régional de Metz-Thionville**  
Aspects scientifiques, techniques et réglementaires

**Membres du jury :**

Monsieur le Professeur N. Jay

**Président**

Madame le Docteur F. Claudot

**Juge**

Monsieur le Docteur J. Godet

**Juge**

Monsieur le Docteur C. Goetz

**Juge et Directeur**



**Professeur Marc**



**Président de l'Université de Lorraine Professeur  
Pierre MUTZENHARDT**

**Doyen de la Faculté de Médecine  
BRAUN**

**Vice-doyenne  
Pr Laure JOLY**

**Assesseurs :**

*Premier cycle* : Dr Nicolas GAMBIER  
*Deuxième cycle* : Dr Antoine KIMMOUN  
*Troisième cycle* : Pr Laure JOLY  
*Formation à la recherche* : Pr Nelly AGRINIER  
*Relations Grande Région* : Pr Thomas FUCHS-BUDER  
*CUESIM* : Pr Stéphane ZUILY  
*SIDES* : Dr Julien BROSEUS  
*Vie Facultaire* : Dr Philippe GUERCI  
*Etudiant* : Mme Audrey MOUGEL

**Chargés de mission**

*Docimologie* : Dr Jacques JONAS  
*Orthophonie* : Pr Cécile PARIETTI-WINKLER  
*PACES* : Pr Mathias POUSSEL  
*Relations internationales* : Pr Jacques HUBERT

*Présidente du Conseil de la Pédagogie* : Pr Louise TYVAERT  
*Président du Conseil Scientifique* : Pr Jean-Michel HASCOET

=====  
**DOYENS HONORAIRES**

Professeur Jean-Bernard DUREUX - Professeur Jacques ROLAND - Professeur Patrick NETTER - Professeur Henry COUDANE

=====  
**PROFESSEURS HONORAIRES**

Etienne ALIOT - Jean-Marie ANDRE - Alain AUBREGE - Gérard BARROCHE - Alain BERTRAND - Pierre BEY - Marc-André BIGARD - Patrick BOISSEL - Pierre BORDIGONI - Jacques BORRELLY - Michel BOULANGE - Jean-Louis BOUTROY - Serge BRIANÇON - Jean-Claude BURDIN - Claude BURLET - Daniel BURNEL - Claude CHARDOT - Jean-François CHASSAGNE - François CHERRIER - Henry COUDANE - Jean-Pierre CRANCE - Jean-Pierre DESCHAMPS - Jean-Bernard DUREUX - Gilbert FAURE - Gérard FIEVE - Bernard FOLIGUET - Jean FLOQUET - Robert FRISCH - Alain GAUCHER - Pierre GAUCHER - Jean-Luc GEORGE - Alain GERARD - Hubert GERARD - Jean-Marie GILGENKRANTZ - Simone GILGENKRANTZ - Gilles GROSDIDIER - François GUILLEMIN - Philippe HARTEMANN - Gérard HUBERT - Claude HURIET - Michèle KESSLER - François KOHLER - Henri LAMBERT - Pierre LANDES - Pierre LASCOMBES - Marie-Claire LAXENAIRE - Michel LAXENAIRE - Alain LE FAOU - Jacques LECLERE - Pierre LEDERLIN - Bernard LEGRAS - Jean-Pierre MALLIÉ - Philippe MANGIN - François MARCHAL - Jean-Claude MARCHAL - Yves MARTINET - Pierre MATHIEU - Michel MERLE - Daniel MOLÉ - Pierre MONIN - Pierre NABET - Patrick NETTER - Jean-Pierre NICOLAS - Francis PENIN - Claude PERRIN - Luc PICARD - François PLENAT - Jean-Marie POLU - Jacques POUREL - Francis RAPHAEL - Antoine RASPILLER - Denis REGENT - Jacques ROLAND - Daniel SCHMITT - Michel SCHMITT - Michel SCHWEITZER - Daniel SIBERTIN-BLANC - Claude SIMON - Danièle SOMMELET - Jean-François STOLTZ - Michel STRICKER - Gilbert THIBAUT - Paul VERT - Hervé VESPIGNANI - Colette VIDAILHET - Michel VIDAILHET - Jean-Pierre VILLEMOT - Michel WEBER - Denis ZMIROU - Faïez ZANNAD

=====

## PROFESSEURS ÉMÉRITES

Etienne ALIOT - Pierre BEY - Henry COUDANE - Serge BRIANÇON - Jean-Marie GILGENKRANTZ - Simone GILGENKRANTZ - Michèle KESSLER - Alain LE FAOU - Patrick NETTER - Jean-Pierre NICOLAS - Luc PICARD - François PLENAT - Jean-Pierre VILLEMOT - Faiez ZANNAD

=====

## PROFESSEURS DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

(Disciplines du Conseil National des Universités)

### 42<sup>e</sup> Section : MORPHOLOGIE ET MORPHOGENÈSE

#### 1<sup>re</sup> sous-section : (*Anatomie*)

Professeur Marc BRAUN R Professeure Manuela PEREZ

#### 2<sup>e</sup> sous-section : (*Histologie, embryologie et cytogénétique*)

Professeur Christo CHRISTOV

#### 3<sup>e</sup> sous-section : (*Anatomie et cytologie pathologiques*)

Professeur Guillaume GAUCHOTTE

### 43<sup>e</sup> Section : BIOPHYSIQUE ET IMAGERIE MÉDICALE

#### 1<sup>re</sup> sous-section : (*Biophysique et médecine nucléaire*)

Professeur Gilles KARCHER R Professeur Pierre-Yves MARIE R Professeur Pierre OLIVIER - Professeur Antoine VERGER

#### 2<sup>e</sup> sous-section : (*Radiologie et imagerie médicale*)

Professeur René ANXIONNAT - Professeur Alain BLUM - Professeur Serge BRACARD - Professeure Valérie CROISÉ - Professeur Jacques FELBLINGER - Professeur Benjamin GORY - Professeur Damien MANDRY - Professeur Pedro GONDIM TEIXEIRA

### 44<sup>e</sup> Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLÉCULAIRE, PHYSIOLOGIE ET NUTRITION

#### 1<sup>re</sup> sous-section : (*Biochimie et biologie moléculaire*)

Professeur Jean-Louis GUEANT - Professeur Bernard NAMOUR - Professeur Jean-Luc OLIVIER

#### 2<sup>e</sup> sous-section : (*Physiologie*)

Professeur Christian BEYAERT - Professeur Bruno CHENUUEL - Professeur Mathias POUSSEL

#### 3<sup>e</sup> sous-section : (*Biologie cellulaire*)

Professeure Véronique DECOT-MAILLERET

#### 4<sup>e</sup> sous-section : (*Nutrition*)

Professeur Didier QUILLIOT - Professeure Rosa-Maria RODRIGUEZ-GUEANT - Professeur Olivier ZIEGLER

### 45<sup>e</sup> Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE

#### 1<sup>re</sup> sous-section : (*Bactériologie – virologie ; hygiène hospitalière*)

Professeur Alain LOZNIIEWSKI R Professeure Evelyne SCHVOERER

#### 2<sup>e</sup> sous-section : (*Parasitologie et Mycologie*)

Professeure Marie MACHOUART

#### 3<sup>e</sup> sous-section : (*Maladies infectieuses ; maladies tropicales*)

Professeur Bruno HOEN - Professeur Thierry MAY - Professeure Céline PULCINI - Professeur Christian RABAUD

### 46<sup>e</sup> Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ

#### 1<sup>re</sup> sous-section : (*Épidémiologie, économie de la santé et prévention*)

Professeure Nelly AGRINIER - Professeur Francis GUILLEMIN

#### 4<sup>e</sup> sous-section : (*Biostatistiques, informatique médicale et technologies de communication*)

Professeure Eliane ALBUISSON - Professeur Nicolas JAY

### 47<sup>e</sup> Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE

#### 1<sup>re</sup> sous-section : (*Hématologie ; transfusion*)

Professeur Pierre FEUGIER

#### 2<sup>e</sup> sous-section : (*Cancérologie ; radiothérapie*)

Professeur Thierry CONROY - Professeur Frédéric MARCHAL - Professeur Didier PEIFFERT - Professeur Guillaume VOGIN

#### 3<sup>e</sup> sous-section : (*Immunologie*)

Professeur Marcelo DE CARVALHO-BITTENCOURT - Professeure Marie-Thérèse RUBIO

#### 4<sup>e</sup> sous-section : (*Génétique*)

Professeur Philippe JONVEAUX - Professeur Bruno LEHEUP

## **48<sup>e</sup> Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE**

### **1<sup>re</sup> sous-section : (Anesthésiologie-réanimation et médecine péri-opératoire)**

Professeur Gérard AUDIBERT - Professeur Hervé BOUAZIZ - Professeur Thomas FUCHS-BUDER  
Professeure Marie-Reine LOSSER - Professeur Claude MEISTELMAN

### **2<sup>e</sup> sous-section : (Médecine intensive-réanimation)**

Professeur Pierre-Édouard BOLLAERT - Professeur Sébastien GIBOT - Professeur Bruno LÉVY

### **3<sup>e</sup> sous-section : (Pharmacologie fondamentale ; pharmacologie clinique ; addictologie)**

Professeur Pierre GILLET - Professeur Jean-Yves JOUZEAU

### **4<sup>e</sup> sous-section : (Thérapeutique-médecine de la douleur ; addictologie)**

Professeur Nicolas GIRERD - Professeur François PAILLE - Professeur Patrick ROSSIGNOL

## **49<sup>e</sup> Section : PATHOLOGIE NERVEUSE ET MUSCULAIRE, PATHOLOGIE MENTALE, HANDICAP ET RÉÉDUCATION**

### **1<sup>re</sup> sous-section : (Neurologie)**

Professeur Marc DEBOUVERIE - Professeur Louis MAILLARD - Professeur Sébastien RICHARD - Professeur Luc TAILLANDIER  
Professeure Louise TYVAERT

### **2<sup>e</sup> sous-section : (Neurochirurgie)**

Professeur Thierry CIVIT - Professeure Sophie COLNAT-COULBOIS - Professeur Olivier KLEIN

### **3<sup>e</sup> sous-section : (Psychiatrie d'adultes ; addictologie)**

Professeur Jean-Pierre KAHN - Professeur Vincent LAPREVOTE - Professeur Raymund SCHWAN

### **4<sup>e</sup> sous-section : (Pédopsychiatrie ; addictologie)**

Professeur Bernard KABUTH

### **5<sup>e</sup> sous-section : (Médecine physique et de réadaptation)**

Professeur Jean PAYSANT

## **50<sup>e</sup> Section : PATHOLOGIE OSTÉO-ARTICULAIRE, DERMATOLOGIE ET CHIRURGIE PLASTIQUE**

### **1<sup>re</sup> sous-section : (Rhumatologie)**

Professeure Isabelle CHARY-VALCKENAERE - Professeur Damien LOEUILLE

### **2<sup>e</sup> sous-section : (Chirurgie orthopédique et traumatologique)**

Professeur Laurent GALOIS - Professeur Didier MAINARD - Professeur François SIRVEAUX

### **3<sup>e</sup> sous-section : (Dermato-vénéréologie)**

Professeure Anne-Claire BURSZTEJN - Professeur Jean-Luc SCHMUTZ

### **4<sup>e</sup> sous-section : (Chirurgie plastique, reconstructrice et esthétique ; brûlologie)**

Professeur François DAP - Professeur Gilles DAUTEL - Professeur Etienne SIMON

## **51<sup>e</sup> Section : PATHOLOGIE CARDIO-RESPIRATOIRE ET VASCULAIRE**

### **1<sup>re</sup> sous-section : (Pneumologie ; addictologie)**

Professeur Jean-François CHABOT - Professeur Ari CHAOUAT

### **2<sup>e</sup> sous-section : (Cardiologie)**

Professeur Edoardo CAMENZIND - Professeur Christian de CHILLOU DE CHURET - Professeur Yves JUILLIERE

Professeur Nicolas SADOUL

### **3<sup>e</sup> sous-section : (Chirurgie thoracique et cardiovasculaire)**

Professeur Juan-Pablo MAUREIRA - Professeur Stéphane RENAUD

### **4<sup>e</sup> sous-section : (Chirurgie vasculaire ; médecine vasculaire)**

Professeur Sergueï MALIKOV - Professeur Denis WAHL - Professeur Stéphane ZUILY

## **52<sup>e</sup> Section : MALADIES DES APPAREILS DIGESTIF ET URINAIRE**

### **1<sup>re</sup> sous-section : (Gastroentérologie ; hépatologie ; addictologie)**

Professeur Jean-Pierre BRONOWICKI - Professeur Laurent PEYRIN-BIROULET

### **2<sup>e</sup> sous-section : (Chirurgie viscérale et digestive)**

Professeur Ahmet AYAV - Professeur Laurent BRESLER - Professeur Laurent BRUNAUD - Professeure Adeline GERMAIN

### **3<sup>e</sup> sous-section : (Néphrologie)**

Professeur Luc FRIMAT - Professeure Dominique HESTIN

### **4<sup>e</sup> sous-section : (Urologie)**

Professeur Pascal ESCHWEGE - Professeur Jacques HUBERT

## **53<sup>e</sup> Section : MÉDECINE INTERNE, GÉRIATRIE ET MÉDECINE GÉNÉRALE**

### **1<sup>re</sup> sous-section : (Médecine interne ; gériatrie et biologie du vieillissement ; addictologie)**

Professeur Athanase BENETOS - Professeur Jean-Dominique DE KORWIN - Professeure Gisèle KANNY

Professeure Christine PERRET-GUILLAUME - Professeur Roland JAUSSAUD - Professeure Laure JOLY

### **3<sup>e</sup> sous-section : (Médecine générale)**

Professeur Jean-Marc BOIVIN - Professeur Paolo DI PATRIZIO

**54<sup>e</sup> Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE, ENDOCRINOLOGIE ET REPRODUCTION**

**1<sup>re</sup> sous-section : (Pédiatrie)**

Professeur Pascal CHASTAGNER - Professeur François FEILLET - Professeur Jean-Michel HASCOET  
Professeur Emmanuel RAFFO - Professeur Cyril SCHWEITZER

**2<sup>e</sup> sous-section : (Chirurgie infantile)**

Professeur Pierre JOURNEAU - Professeur Jean-Louis LEMELLE

**3<sup>e</sup> sous-section : (Gynécologie-obstétrique ; gynécologie médicale)**

Professeur Philippe JUDLIN - Professeur Olivier MOREL

**4<sup>e</sup> sous-section : (Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale)**

Professeur Bruno GUERCI - Professeur Marc KLEIN - Professeur Georges WERYHA

**55<sup>e</sup> Section : PATHOLOGIE DE LA TÊTE ET DU COU**

**1<sup>re</sup> sous-section : (Oto-rhino-laryngologie)**

Professeur Roger JANKOWSKI - Professeure Cécile PARIETTI-WINKLER

**2<sup>e</sup> sous-section : (Ophtalmologie)**

Professeure Karine ANGIOI - Professeur Jean-Paul BERROD

**3<sup>e</sup> sous-section : (Chirurgie maxillo-faciale et stomatologie)**

Professeure Muriel BRIX

=====

**PROFESSEURS DES UNIVERSITÉS**

**61<sup>e</sup> Section : GÉNIE INFORMATIQUE, AUTOMATIQUE ET TRAITEMENT DU SIGNAL**

Professeur Walter BLONDEL

**64<sup>e</sup> Section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE**

Professeure Sandrine BOSCHI-MULLER - Professeur Pascal REBOUL

**65<sup>e</sup> Section : BIOLOGIE CELLULAIRE**

Professeure Céline HUSELSTEIN

=====

**PROFESSEUR ASSOCIÉ DE MÉDECINE GÉNÉRALE**

Professeure associée Sophie SIEGRIST

Professeur associé Olivier BOUCHY

=====

**MAÎTRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS**

**42<sup>e</sup> Section : MORPHOLOGIE ET MORPHOGENÈSE**

**1<sup>re</sup> sous-section : (Anatomie)**

Docteur Bruno GRIGNON

**2<sup>e</sup> sous-section : (Histologie, embryologie, et cytogénétique)**

Docteure Isabelle KOSCINSKI

**44<sup>e</sup> Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLÉCULAIRE, PHYSIOLOGIE ET NUTRITION**

**1<sup>re</sup> sous-section : (Biochimie et biologie moléculaire)**

Docteure Shyue-Fang BATTAGLIA - Docteure Sophie FREMONT - Docteure Catherine MALAPLATE - Docteur Marc MERTEN - Docteur Abderrahim OUSSALAH

**2<sup>e</sup> sous-section : (Physiologie)**

Docteure Silvia DEMOULIN-ALEXIKOVA - Docteur Jacques JONAS

**45<sup>e</sup> Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE**

**1<sup>re</sup> sous-section : (Bactériologie – Virologie ; hygiène hospitalière)**

Docteure Corentine ALAUZET - Docteure Hélène JEULIN - Docteure Véronique VENARD

**2<sup>e</sup> sous-section : (Parasitologie et mycologie)**

Docteure Anne DEBOURGOGNE

**46<sup>e</sup> Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ**

**1<sup>re</sup> sous-section : (Epidémiologie, économie de la santé et prévention)**

Docteur Cédric BAUMANN - Docteure Frédérique CLAUDOT - Docteur Alexis HAUTEMANIÈRE

Docteur Arnaud FLORENTIN (stagiaire)

**2<sup>e</sup> sous-section (Médecine et Santé au Travail)**

Docteure Isabelle THAON

**3<sup>e</sup> sous-section (Médecine légale et droit de la santé)**

Docteur Laurent MARTRILLE

**47<sup>e</sup> Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE**

**1<sup>re</sup> sous-section : (Hématologie ; transfusion)**

Docteur Julien BROSEUS - Docteur Maud D'AVENI

**2<sup>e</sup> sous-section : (Cancérologie ; radiothérapie)**

Docteur Lina BOLOTINE

**3<sup>e</sup> sous-section : (Immunologie)**

Docteur Alice AARNINK (stagiaire)

**4<sup>e</sup> sous-section : (Génétique)**

Docteur Céline BONNET

**48<sup>e</sup> Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE**

**1<sup>e</sup> sous-section : (Anesthésiologie-réanimation et médecine péri-opératoire)**

Docteur Philippe GUERCI

**2<sup>e</sup> sous-section : (Médecine intensive-réanimation)**

Docteur Antoine KIMMOUN

**3<sup>e</sup> sous-section : (Pharmacologie fondamentale ; pharmacologie clinique ; addictologie)**

Docteur Nicolas GAMBIER - Docteur Françoise LAPICQUE - Docteur Julien SCALA-BERTOLA

**49<sup>e</sup> Section : PATHOLOGIE NERVEUSE ET MUSCULAIRE, PATHOLOGIE MENTALE, HANDICAP ET RÉÉDUCATION**

**2<sup>e</sup> sous-section : (Neurochirurgie)**

Docteur Fabien RECH (stagiaire)

**3<sup>e</sup> sous-section : (Psychiatrie d'adultes ; addictologie)**

Docteur Thomas SCHWITZER (stagiaire)

**50<sup>e</sup> Section : PATHOLOGIE OSTÉO-ARTICULAIRE, DERMATOLOGIE ET CHIRURGIE PLASTIQUE**

**4<sup>e</sup> sous-section : (Chirurgie plastique, reconstructrice et esthétique ; brûlologie)**

Docteur Lætitia GOFFINET-PLEUTRET

**51<sup>e</sup> Section : PATHOLOGIE CARDIO-RESPIRATOIRE ET VASCULAIRE**

**3<sup>e</sup> sous-section : (Chirurgie thoracique et cardio-vasculaire)**

Docteur Fabrice VANHUYSE

**4<sup>e</sup> sous-section : (Chirurgie vasculaire ; Médecine vasculaire)**

Docteur Nicla SETTEMBRE

**52<sup>e</sup> Section : MALADIES DES APPAREILS DIGESTIF ET URINAIRE**

**1<sup>re</sup> sous-section : (Gastroentérologie ; hépatologie ; addictologie)**

Docteur Anthony LOPEZ

**2<sup>e</sup> sous-section : (Chirurgie viscérale et digestive)**

Docteur Cyril PERRENOT

**53<sup>e</sup> Section : MÉDECINE INTERNE, GÉRIATRIE ET MÉDECINE GÉNÉRALE**

**3<sup>e</sup> sous-section : (Médecine générale)**

Docteur Kénora CHAU (stagiaire)

**54<sup>e</sup> Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE, ENDOCRINOLOGIE ET REPRODUCTION**

**4<sup>e</sup> sous-section : (Endocrinologie, diabète et maladies métaboliques ; Gynécologie médicale)**

Docteur Eva FEIGERLOVA

**5<sup>e</sup> sous-section : (Biologie et médecine du développement et de la reproduction ; gynécologie médicale)**

Docteur Mikaël AGOPIANTZ (stagiaire)

**55<sup>e</sup> Section : PATHOLOGIE DE LA TÊTE ET DU COU**

**1<sup>re</sup> sous-section : (Oto-Rhino-Laryngologie)**

Docteur Patrice GALLET

=====  
**MAÎTRES DE CONFÉRENCES**

**5<sup>e</sup> Section : SCIENCES ÉCONOMIQUES**

Monsieur Vincent LHUILLIER

**7<sup>e</sup> Section : SCIENCES DU LANGAGE : LINGUISTIQUE ET PHONÉTIQUE GÉNÉRALES**

Madame Christine DA SILVA-GENEST

**19<sup>e</sup> Section : SOCIOLOGIE, DÉMOGRAPHIE**

Madame Joëlle KIVITS

**64<sup>e</sup> Section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE**

Madame Marie-Claire LANHERS - Monsieur Nick RAMALANJAONA

**65<sup>e</sup> Section : BIOLOGIE CELLULAIRE**

Madame Nathalie AUCHET - Madame Natalia DE ISLA-MARTINEZ - Monsieur Christophe NEMOS

**66<sup>e</sup> Section : PHYSIOLOGIE**

Monsieur Nguyen TRAN

**69<sup>e</sup> Section : NEUROSCIENCES**

Madame Sylvie MULTON

=====

**MAÎTRES DE CONFÉRENCES ASSOCIÉS DE MÉDECINE GÉNÉRALE**

Docteur Cédric BERBE

=====

**DOCTEURS HONORIS CAUSA**

Professeur Charles A. BERRY (1982)  
*Centre de Médecine Préventive, Houston (U.S.A)*  
Professeur Pierre-Marie GALETTI (1982)  
*Brown University, Providence (U.S.A)*  
Professeure Mildred T. STAHLMAN (1982)  
*Vanderbilt University, Nashville (U.S.A)*  
Professeur Théodore H. SCHIEBLER (1989)  
*Institut d'Anatomie de Würzburg (R.F.A)*  
Professeur Mashaki KASHIWARA (1996)  
*Research Institute for Mathematical Sciences de  
Kyoto (JAPON)*

Professeure Maria DELIVORIA-  
PAPADOPOULOS (1996) *Université de  
Pennsylvanie (U.S.A)*  
Professeur Ralph GRÄSBECK (1996)  
*Université d'Helsinki (FINLANDE)*  
Professeur Duong Quang TRUNG (1997)  
*Université d'Hô Chi Minh-Ville (VIÊTNAM)*  
Professeur Daniel G. BICHET (2001)  
*Université de Montréal (Canada)*  
Professeur Marc LEVENSTON (2005)  
*Institute of Technology, Atlanta (USA)*

Professeur Brian BURCHELL  
(2007)  
*Université de Dundee (Royaume-  
Uni)*  
Professeur Yunfeng ZHOU (2009)  
*Université de Wuhan (CHINE)*  
Professeur David ALPERS (2011)  
*Université de Washington (U.S.A)*  
Professeur Martin EXNER (2012)  
*Université de Bonn (ALLEMAGNE)*

## **Dédicaces**

### **Au Président du Jury**

À notre maître et président de thèse Monsieur le Docteur Nicolas Jay  
Professeur des Universités - Praticien Hospitalier du CHRU de Nancy  
Université de Lorraine, Faculté de Médecine  
Service d'Evaluation et d'Information Médicales

Notre premier contact avec l'informatique médicale s'est déroulé pendant notre externat, lors d'un stage au Service d'Evaluation et d'Information Médicales du CHRU de Nancy. Initiation qui s'est par la suite concrétisée par le choix de la spécialité de santé publique, deux Master 2, cette thèse... et bien d'autres choses à venir !

Nous vous remercions d'avoir accepté la direction de cette thèse.

Soyez assuré de toute notre gratitude.

## Aux juges

À notre maître Madamele Docteur Frédérique Claudot

Maître de Conférence des Universités - Praticien Hospitalier du CHRU de Nancy

Université de Lorraine, Faculté de Médecine

École de Santé Publique

Notre premier contact avec le Droit a eu lors lieu d'un de vos cours de Sciences Humaines et Sociales de PCEM1. La partie réglementaire de cette thèse en est la continuation et nous espérons que vous la jugerez digne d'intérêt.

Nous vous remercions d'avoir accepté de juger cette thèse.

Soyez assurée de toute notre gratitude.

À notre maître Monsieur le Docteur Julien Godet

Maître de Conférence des Universités - Praticien Hospitalier des Hôpitaux Universitaires de Strasbourg

Université de Strasbourg, faculté de pharmacie

Groupe de Méthodes en Recherche Clinique

Nos premières discussions concernaient l'entrepôt de Strasbourg et ont alimenté en partie le contenu de cette thèse. Nous espérons que vous la jugerez digne d'intérêt.

Nous vous remercions d'avoir accepté de juger cette thèse.

Soyez assuré de toute notre gratitude.

À notre maître et directeur de thèse Monsieur le Docteur Christophe Goetz

Médecin de santé publique - Praticien Hospitalier du CHR de Metz-Thionville

Département d'Information Médicale

Plateforme d'Appui à la Recherche Clinique

Nous vous remercions pour votre encadrement et votre soutien depuis maintenant plus d'un an ; nous espérons pouvoir continuer à travailler ensemble au sein du CHR pour les prochaines années !

Nous vous remercions d'avoir accepté de diriger et de juger cette thèse.

Soyez assuré de toute notre gratitude.

## **À la famille**

À mes chers parents, merci pour votre amour et votre soutien infailibles. Plus spécifiquement, lors de ces études de médecine : des concours de PCEM1 à l'ECN en passant par les Masters et autres DU, les ascenseurs émotionnels n'auront pas manqué. Sans vous, rien de cela n'aurait été possible !

À mes frères et à ma sœur ainsi que leurs épouses et époux : merci également pour votre soutien et votre présence !

À mes neveux et nièces, dont le nombre a cru en même temps que le celui des années d'études : d'ici huit ans, ce sera à votre tour de vous lancer : je ne doute pas que la relève sera assurée !

## **Aux amis et collègues**

À tous les médecins et professionnels qui m'ont formé et soutenu pendant ces études : le Dr Paris, le Dr Dollet, le Dr Villeroy de Galhau, le Dr Bannay, le Dr Fresson, le Dr Tanguy, Le Pr Jay, le Dr Schockmel, le Dr Wack, le Dr Florentin, le Dr Renaudin, le Dr Vernier, le Dr Fiet, le Dr Yai, le Pr Meyer, le Pr Sauleau, le Dr Nisand, le Dr Goetz, le Dr Weisslinger, le Dr Faulon, le Dr Béthune, le Dr Fiore et bien d'autres encore.

À tous mes cointernes et collègues : Ngoc-Ha, Maxime, Yoann, Arnaud, Clément, Luc, Astiana, Asma, Magali, Caroline, Maud, Maxime, Marie et Victor.

À toutes les personnes ayant contribué à cette thèse : le Pr Jay, le Dr Goetz, le Dr Weisslinger, le Dr Faulon, le Dr Béthune, le Dr Fiore, Mme Fristot, le Dr Bannay, le Dr Muller, le Dr Godet, M. Kronz, M. Germain, Mme Striebel, le Pr Cuggia ; à mon père et à ma sœur pour la relecture.

À tous mes amis de Nancy, de Strasbourg, de Metz et d'autres horizons, rencontrés dans le cadre universitaire, associatif, professionnel ou fortuitement.

## Serment d'Hypocrate

«Aumomentd'êtréadmisàexercerlamédecine,jeprometset je jure d'être fidèle aux lois de l'honneur et de la probité. Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux. Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité. J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences. Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire. Admis dans l'intimité des personnes, je tairai les secrets qui me sont confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs. Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément. Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés. J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité. Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré et méprisé si j'y manque ».

## Table des matières

Liste des Professeurs.....	2
Dédicaces.....	8
Serment d’Hypocrate .....	12
Table des matières .....	13
Liste des Acronymes.....	16
I. Introduction.....	18
II. Systèmesd’organisation de connaissances .....	20
III. Système d’Information Hospitalier (SIH) du Centre Régional Hospitalier de Metz-Thionville (CHRMT) .....	24
1. Généralités .....	24
2. Architecture du réseau.....	24
3. Échange de données entre les applications : rôle de l’EAI (Enterprise Application Integration) 24	
4. Cartographie des flux au CHRMT .....	28
a. Périmètre de l’EDS (entrepôt de données de santé) du CHRMT .....	28
b. Flux des diagnostics médicaux PMSI .....	28
c. Flux des observations cliniques.....	29
d. Flux des actes de la Classification Commune des Actes Médicaux (CCAM).....	31
i. Généralités .....	31
ii. Actes CCAM réalisés en externe.....	31
iii. Actes CCAM réalisés en hospitalisation dans les autres champs PMSI.....	31
iv. Alimentation de l’EDS.....	32
e. Flux des actes du Catalogue Spécifique des Actes de Rééducation et Réadaptation (CSARR) et E.D.G.A.R. ....	32
f. Flux des identités et mouvements des patients.....	33
g. Flux des prescriptions et validation médicamenteuse.....	33
h. Flux des résultats de biologie. ....	34
IV. EDS : bases théoriques et application au cas du CHRMT .....	36
1. Entrepôts de Données : bases théoriques. ....	36
2. ehop :caractéristiques.....	37
3. Activités de traitement de l’EDS au sein du CHRMT .....	40

V. Article : quel cadre réglementaire pour les entrepôts de données de santé hospitaliers ? .....	41
1) Introduction.....	45
2) Définitions .....	46
a) Définition des données à caractère personnel .....	46
b) Définition des données “dites sensibles” et de santé.....	47
c) Définition des données “perçues sensibles” .....	47
e) Définition d’un traitement de données.....	48
3) Cadre réglementaire auquel sont soumis les EDS.....	48
a) Cadre général .....	48
b) Secret professionnel et dérogations .....	50
4) Impact du cadre réglementaire sur la demande d’autorisation .....	51
a) La gouvernance de l’EDS .....	51
b) La finalité .....	53
c) Catégories de données.....	58
d) Base légale de traitement et dérogations aux interdictions de traiter les données dites et perçues sensibles.....	58
e) Responsable de traitement et accès aux données.....	59
f) Droits des patients .....	60
i) Droit à l’information.....	60
1. Généralités sur le droit à l’information.....	60
2. Information du patient lors de l’inclusion de ses données dans l’entrepôt .....	60
Prise en charge avant date mise en place EDS.....	60
Prise en charge après la date de mise en place de l’EDS .....	61
3. Information sur les traitements .....	61
ii) Droit d’accès, de rectification à la portabilité .....	63
iii) Droit à l’effacement/oubli.....	64
iv) Droit d’opposition .....	64
v) Droit à la limitation de traitement .....	65
vi) Obligation de notification en cas de rectification ou effacement des données .....	65
g) Droits des soignants .....	67
h) Droits des utilisateurs de l’entrepôt.....	67
i) Protection des données.....	67
i) Durée de conservation .....	67
ii) Principe de minimisation des données .....	67
iii) Pseudonymisation et anonymisation .....	68
iv) Accès différencié aux données.....	69

v) Sécurité informatique.....	70
5) Conclusion .....	71
VI. Conclusions / Perspectives .....	76
VII. Références bibliographiques .....	77
Autorisation d'imprimer.....	78
Résumé de la thèse .....	79
Titre en anglais .....	79
Mots clés .....	79

## Liste des Acronymes

AIPD	Analyse d'Impact relative à la Protection des Données
ATIH	Agence Technique pour l'Information Hospitalière
CCAM	Classification Commune des Actes Médicaux
CHRMT	Centre Hospitalier Régional de Metz-Thionville.
CIM10	Classification Internationale statistique des Maladies et des problèmes de santé connexes
CSARR	Catalogue Spécifique des Actes de Rééducation et Réadaptation
CP	Code Pénal
CSP	Code de Santé Publique
CSE	Comité Scientifique et Ethique
DPI	Dossier Patient Informatisé
DIM	Départements d'Information Médicale
EAI	Enterprise Application Integration
EDS	Entrepôt de Données de Santé
GAM	Gestion Administrative des Malades
GME	Groupe Médico-Economique
GIRCI-EST	Groupement Interrégional de la Recherche Clinique et d'Innovation de l'Est
HAD	Hospitalisation A Domicile
HTML	Hypertext Markup Language
ISO	International Organization for Standardisation
IPP	Identifiant Permanent du Patient
MCO	Médecine, Chirurgie, Obstétrique et Odontologie
MR	Méthodologie de Référence de la CNIL
MR001	Méthodologie de Référence de la CNIL numéro 001
MR002	Méthodologie de Référence de la CNIL numéro 002
MR004	Méthodologie de Référence de la CNIL numéro 004
NABM	Nomenclature des Actes de Biologie Médicale

NIR	Numéro d'Inscription des personnes au Répertoire national d'identification des personnes physiques
OSE	Opérateurs de Services Essentiels
PAISaGE	Plateforme d'Accessibilité et d'Intelligence des données de Santé du Groupement Sanitaire Grand Est
PMSI	Programme de Médicalisation des Systèmes d'Information
PSY	Psychiatrie
SIH	Système d'Information Hospitalier
SOC	Systèmes d'Organisation de Connaissances
SSR	Soins de Suite et Réadaptation
XML	Extensible MarkupLanguage
UCD	Unités Communes de Dispensation
LPP	Liste des Produits et Prestations
TAL	Traitement Automatisé du Langage

## I. Introduction

La digitalisation de l'hôpital au cours de ces dernières décennies a facilité le recueil numérique de l'information médicale par la mise en place du Dossier Patient Informatisé (DPI) et son alimentation par les divers logiciels métiers spécialisés (informations administratives, observations cliniques, biologie, prescriptions, imagerie...) [1]. Les méthodes statistiques traditionnelles sont très efficaces pour analyser les données structurées (biologie, prescriptions...) mais beaucoup moins pour le texte libre (comptes-rendus médicaux, observations cliniques...), qui contient la majorité des informations médicales. La démocratisation des techniques d'apprentissage machine, tel que les réseaux de neurones, permet à présent d'exploiter plus aisément ces ressources [2]. La réutilisation de ces données peut se faire à des fins de recherche, de pilotage hospitalier et d'aide à la décision thérapeutique. Mais elle présente des contraintes : celles des 3 V du big data (contrainte de vitesse, de variété et de volume) et celles liées à leur nature confidentielle. C'est pourquoi le Pr Cuggia et son équipe du CHU de Rennes, précurseurs français en la matière, ont développé fin des années 2000 un outil dédié : un Entrepôt de Données de Santé (EDS), initialement nommé Roogle [3] et à présent renommé ehop (pour entrepôt de données de l'hôpital).

C'est cette solution qui a été retenue pour l'EDS du Centre Hospitalier Régional de Metz-Thionville (CHRMT). Il s'agit de l'établissement support du Groupement Hospitalier de Territoire 6 Lorraine-Nord. Il est composé de huit sites : l'hôpital de Mercy, l'hôpital de Bel-Air, l'hôpital Femme Mère Enfant de Metz, l'hôpital Femme Mère Enfant de Bel-Air, l'hôpital Félix Maréchal, l'hôpital de Hayange, l'EHPAD le Parc et l'EHPAD Saint-Jean.

En 2018, près de 200 000 hospitalisations (pour plus de 660 000 journées) et 500 000 consultations y ont été réalisées.

Le CHRMT est également très impliqué dans la recherche avec pour 2019 :

- l'accompagnement de 180 projets de recherche.
- l'inclusion de 1227 patients dans des protocoles expérimentaux.
- la publication de 267 articles.

Par ailleurs, le CHRMT est membre du Groupement Interrégional de la Recherche Clinique et d'Innovation de l'Est (GIRCI-EST), qui réunit également les CHU et les Centre de Lutte Contre le Cancer de Besançon, Dijon-Bourgogne, Nancy, Reims et Strasbourg. C'est dans ce cadre qu'a été décidée la création de la Plateforme d'Accessibilité et d'Intelligence des données de Santé du Groupement Sanitaire Grand Est (PAISaGE), un réseau d'EDS dont le CHRMT sera partie intégrante.

L'objectif de cette thèse est de décrire les aspects scientifiques, techniques et réglementaires de la mise en place et de l'exploitation d'un EDS au sein d'un hôpital français en s'appuyant sur l'exemple du CHRMT.

Nous aborderons d'abord les Systèmes d'Organisation de Connaissances (SOC) puis le Système d'Information Hospitalier (SIH) du CHRMT, les EDS, dont celui du CHRMT et enfin les considérations réglementaires.

## II. Systèmes d'organisation de connaissances

Pour un ordinateur, les mots ne sont que des chaînes de caractères qu'il manipule : il n'a pas de compréhension de leur sens. Par ailleurs, lors d'une communication entre deux parties, la signification des messages doit être la même pour le récepteur et l'émetteur. L'interopérabilité sémantique permet de répondre à cette double contrainte, par l'utilisation des SOC.

Un SOC est formé de concepts, les plus petites unités de connaissance, reliés entre eux par des relations sémantiques. Il représente un domaine de connaissance terminologiquement. [4]

Ces concepts sont dérivés du monde physique par l'abstraction et sont symbolisés dans le langage humain par des termes (ou mots), des codes, des abréviations ou des pictogrammes. Dans les SOC, les concepts sont identifiés par des codes (idéalement non signifiants) et décrits par des libellés.

Les liens entre terme, concept et réalité sont illustrés par le triangle d'Oden Richard. Le schéma 1 s'inspire de l'exemple du livre « informatique médicale, e-santé, fondements et applications » [1] :

- l'objet cœur est la réalité physique du cœur, un assemblage de muscles, de nerfs et de cartilages.
- le concept de cœur est ce que l'activité mentale produit quand il pense à l'objet cœur : un organe dont la fonction de pompe assure la circulation sanguine.
- le terme « heart » symbolise le concept de cœur dans la langue anglaise ; les termes « cœur » et « Herz » en font autant respectivement dans la langue française et allemande.

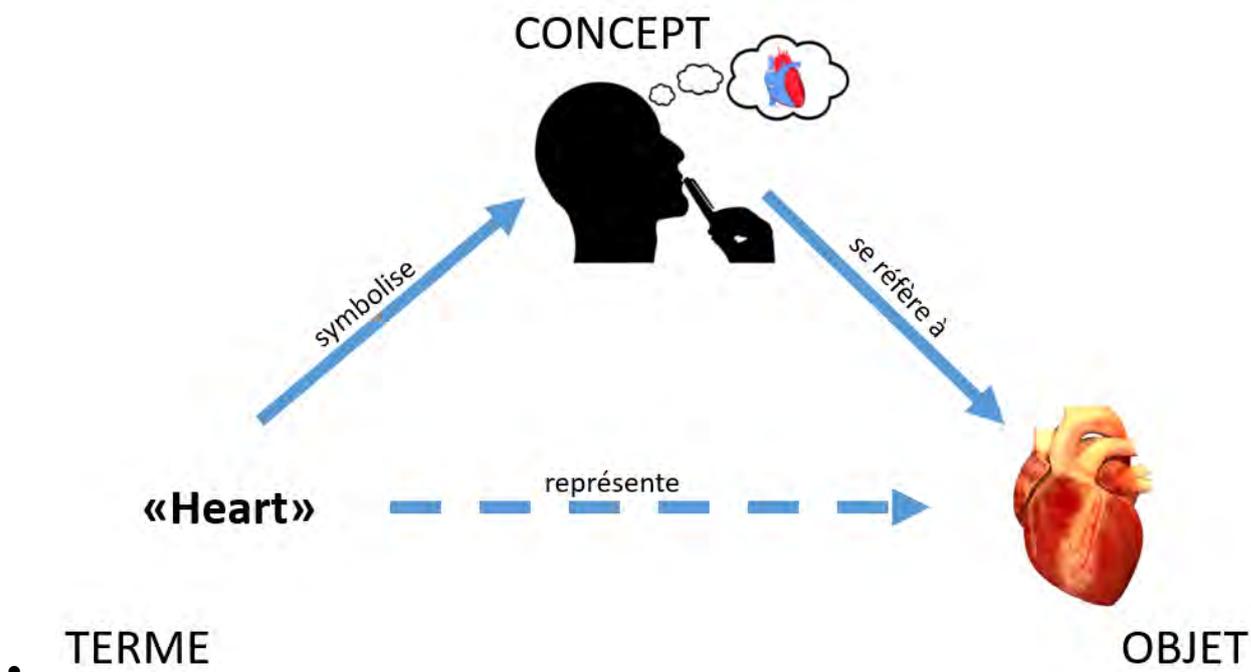


Figure 1 Triangle d'Oden Richard : exemple du cœur, inspiré de l'exemple du livre « informatique médicale, e-santé, fondements et applications »

À première vue, ces notions peuvent sembler superflues. Mais il existe des situations pour lesquelles il n'existe pas de terme pour un concept et pour lesquelles l'usage d'une périphrase est obligatoire. Par exemple en allemand, il existe un terme pour le concept « rassasié », à savoir « satt », mais pas pour le concept « désaltéré ». De même, la correspondance entre un concept médical et une nomenclature peut ne pas être parfaite et constitue une première limitation à la réutilisation des données. Par exemple, la dixième version de la Classification Internationale statistique des Maladies et des problèmes de santé connexes (CIM10) ne permet pas de coder la latéralité d'un problème de santé. Avec cette nomenclature, il n'est donc pas possible de différencier une fracture du premier métacarpien de la main droite et une fracture du premier métacarpien de la main gauche, les deux étant codés S62.2.

Une deuxième limitation, plus générale, est le fait que la réutilisation de données produites lors du soin est moins fiable que la production de données lors d'une étude prospective. Ainsi la Haute Autorité de Santé considère les études rétrospectives, dont les études sur EDS hospitalier, comme étant de « faible niveau de preuve scientifique » (grade C), alors qu'elle considère les études de cohorte comme « présomption scientifique » (grade B) et les analyses de décisions fondées sur des études bien menées comme « preuve scientifique établie » (grade A).[5]

Ces limitations étant explicitées, nous pouvons maintenant approfondir les manières de créer les concepts et quels types de relations sémantiques existent entre eux.

Les aspects de la formation du concept sont clarifiés par leurs définitions que Duclos et al définissent comme «une représentation d'un concept par un énoncé descriptif qui sert à le différencier des concepts associés. » [1]

Traditionnellement, il existe deux manières de définir un concept :

- Par intention, quand sa définition indique le concept plus générique que lui (concept superordonné) puis les caractères le distinguant. Un exemple historique est celui de Socrate qui définit l'Homme comme « un animal pensant ». Dans cet exemple, le concept superordonné est « animal » et le concept différenciant est « pensant ».
- Par extension, quand sa définition énumère tous concepts subordonnés selon un critère de division. Par exemple, les départements de la région Grand-Est sont les Ardennes, l'Aube, le Bas-Rhin, la Haute-Marne, le Haut-Rhin, la Marne, la Meurthe-et-Moselle, la Meuse, la Moselle et les Vosges.

Quand plusieurs concepts sont symbolisés par un même terme, on parle d'homonymie. Par exemple, l'acronyme IVG peut symboliser soit le concept d'insuffisance ventriculaire gauche soit le concept d'interruption volontaire de grossesse. La connaissance du contexte (service de gynécologie-obstétrique ou de cardiologie) permet le plus souvent de lever l'ambiguïté. [4]

Quand un même concept est symbolisé par plusieurs symboles, on parle de synonymie. Par exemple, attaque cérébrale et accident vasculaire cérébrale. Ce genre de situation peut être géré par l'utilisation d'un terme préféré. [4]

Parmi les multiples relations sémantiques existant entre les concepts, on retiendra surtout la relation hiérarchique [1], qui peut être générique ou partitive. Une relation sémantique entre

deux concepts est générique quand l'intention d'un concept inclut celle de l'autre avec au moins un caractère distinctif supplémentaire ; par exemple, infarctus du myocarde et infarctus du myocarde à la phase aiguë. Une relation sémantique entre deux concepts est partitive quand un des concepts constitue une partie de l'autre concept, par exemple «ensembles des os du pied » et « calcanéum ».

Traditionnellement on distingue comme typologie de SOC : les terminologies, les nomenclatures, les thésauri et les systèmes de classification. [1]

Une liste de termes se référant à des concepts d'un domaine particulier constitue une terminologie. Par exemple, la liste de noms du personnel d'un service. Si on rajoute une contrainte d'organisation et la gestion des synonymies, il s'agit d'un thésaurus ou vocabulaire contrôlé. Par exemple, Le MedicalSubjectHeadings est le thésaurus de référence dans le domaine biomédical.

Par ailleurs, si on ajoute des règles de construction pour les termes d'une terminologie, il s'agit alors d'une nomenclature. Par exemple, la nomenclature des groupes médicaux économiques, utilisés en pour le champ Soins de Suite et Réadaptation (SSR) du Programme de Médicalisation des Systèmes d'Information (PMSI) : chaque code de Groupe Médico-Economique(GME) est constitué de deux chiffres (catégorie majeure de diagnostic), deux chiffres (groupe nosologique), une lettre (racine de GME) et un chiffre (séjour de 0 jour ou plus, avec comorbidité associée ou non).

Enfin, dès lors qu'on organise des concepts d'un domaine de connaissance de manière hiérarchique dans des groupes appelés classes, il s'agit d'une classification. Dans le champ de la santé, une classification des maladies est également appelée nosologie. Un exemple est la CIM 10 : le code I50.1 (insuffisance ventriculaire gauche) est le code fils du code père I50 insuffisance cardiaque.

Des exemples de SOC utilisés dans le monde de la santé sont présentés dans le tableau 1 :

**Tableau 1 : exemples de systèmes d'organisation des connaissances utilisés en santé**

<b>Système d'organisation des connaissances</b>	<b>Objectif</b>	<b>Maintenu par</b>
dixième version de la Classification Internationale statistique des Maladies et des problèmes de santé connexes	Codage des diagnostics de maladies ou autres problèmes de santé	Organisation Mondiale de la Santé pour sa version internationale  Agence Technique de l'Information Hospitalière (ATIH) pour sa version descriptive à visée du Programme de Médicalisation des Systèmes d'Information (PMSI)
Nomenclature des Actes de Biologie Médicale	Codage des actes de biologie médicale	Assurance maladie
Classification Commune des Actes Médicaux (CCAM)	Codage des actes médicaux	Assurance maladie pour la version nationale à visée de facturation  ATIH pour version descriptive à visée de groupage PMSI
Grille E.D.G.A.R.	Codage des actes réalisés en psychiatrie ambulatoire (entretien, démarche, groupe, accompagnement, réunion)	ATIH
actes du Catalogue Spécifique des Actes de Rééducation et Réadaptation	Codage des actes réalisés soins de suite et réadaptation : <ul style="list-style-type: none"> <li>• Par les paramédicaux</li> <li>• Par les médicaux si non présent dans la CCAM</li> </ul>	ATIH
Nomenclature Générale des Actes Professionnels	Codage des actes cliniques médicaux et des actes des chirurgiens-dentistes, des sages-femmes et des auxiliaires médicaux  Codage des consultations externes pour les médicaux	Assurance Maladie
Unités Communes de Dispensation	Codage des médicaments	Assurance maladie
Liste des Produits et Prestations	Codage :des dispositifs médicaux, des articles diététiques et des véhicules pour handicapés.	Assurance maladie
Logical Observation Identifier Names and Codes	Analyses biologiques Résultats de tests cliniques ou comptes-rendus	LOINC Committee

Les bases théoriques des SOC étant clarifiées, nous allons avant d'aborder l'EDS du CHRMT, nous intéresser à la source des données qui l'alimentera : le SIH du CHRMT.

### **III. Système d'Information Hospitalier (SIH) du Centre Régional Hospitalier de Metz-Thionville (CHRMT)**

#### **1. Généralités**

Un système d'information hospitalier (SIH) est « destiné à faciliter la gestion de l'ensemble des informations médicales et administratives d'un hôpital [6] ». Il est par vocation intégrateur et l'on pourrait tout aussi bien parler de système intégré de communication et de traitement de l'information hospitalière. Par intégration, on entend « une Opération consistant à inclure un logiciel (progiciel ou développement spécifique), ou plus généralement tout élément informatique, au sein d'un système d'information. [7] ».

Le SIH doit :

- Être fiable et rapide.
- Être suffisamment ergonomique afin de faciliter son utilisation par les producteurs de données, qu'ils soient médicaux, paramédicaux, techniques ou administratifs.
- Permettre d'améliorer la qualité des données (cohérence de l'identité d'un même patient entre deux applications par exemple).

Au sein de ce SIH, nous nous intéresserons particulièrement au DPI, qui permet une coordination de la prise en charge d'un même patient entre plusieurs professionnels de santé ou services.

Avant de réutiliser ces données, il est nécessaire de bien connaître leur contexte de production. Ces flux d'informations reposent sur un réseau informatique, qu'il convient d'étudier préalablement.

#### **2. Architecture du réseau**

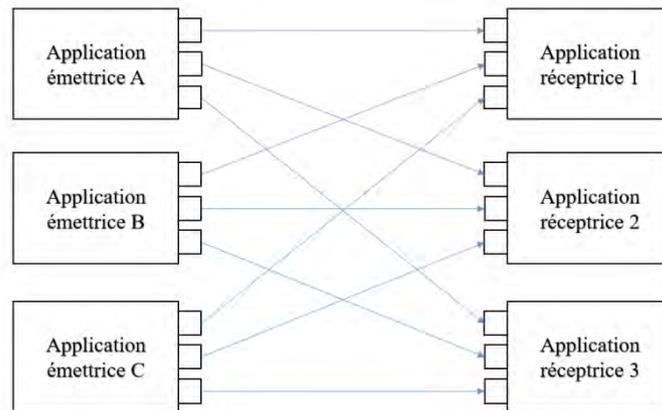
La communication entre les différents postes et applications du CHR se réalise via un réseau. Parmi les différentes architectures de réseau existantes, celle du SIH est basée sur les technologies web (protocole TCP/IP, FTP...). Elles offrent une grande souplesse dans la gestion des flux et sont de facto devenues un standard, facilitant l'intégration de l'EDS.

#### **3. Échange de données entre les applications : rôle de l'EAI (Enterprise Application Integration)**

Pour la gestion des échanges interapplications, deux approches sont possibles. Une première architecture réseau possible est celle où les applications échangent directement entre elles, dans un même format. Le demi-connecteur côté application émettrice interroge la base de

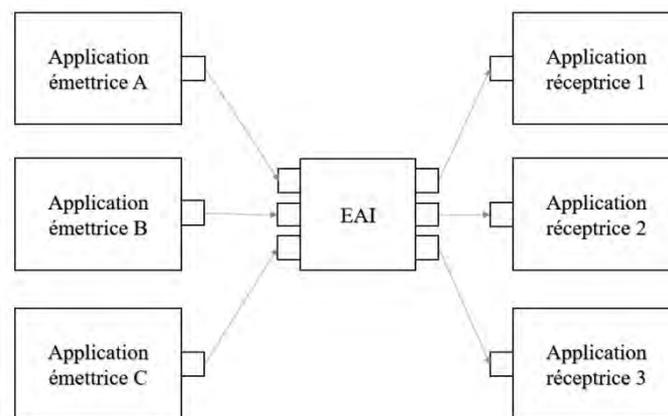
données et met en forme les données et le demi-connecteur côté application réceptrice réceptionne ces données et les intègre dans la base de données de l'application réceptrice.

La figure 2 présente la situation de trois applications émettrices communiquant chacune directement avec trois applications réceptrices. Par exemple, les applications émettrices de gestion des résultats de biologie, de mouvements du patient au sein de l'hôpital et des observations cliniques, envoyant des données vers un serveur d'archivage, un serveur de backup en temps réel et un entrepôt de données.



**Figure 2 : architecture d'un système d'information avec échanges interapplications directs.**

Une seconde architecture possible est celle dans laquelle un système d'Enterprise Application Intégration (EAI) gère les échanges interapplications. Un EAI peut se définir comme le « processus d'intégration d'application préexistante d'un système informatique »[8]. Ici, il désigne également le logiciel qui centralise les échanges entre les applications (Cloverleaf pour le CHR). Autrement dit, il assure un rôle de transport des données d'un point A à un point B, et possiblement du point B au point A. L'exemple mentionné à la figure 1 est repris à la figure 3, mais avec une architecture incorporant cette fois un EAI.



**Figure 3: architecture d'un système d'information avec échanges interapplications gérés par un système d'Enterprise Application Integration**

L'utilisation d'un EAI au sein d'un SIH apporte de nombreux avantages par rapport à une architecture où les applications communiquent directement entre elles ; ils sont listés dans le tableau 2.

**Tableau 2 : Comparaison des avantages selon le type d'architecture d'échanges interapplications : directs et gérés par un système d'Enterprise Application Integration (EAI)**

<b>Aspect Technique</b>	<b>Échanges interapplications directs</b>	<b>Échanges interapplications via l'EAI</b>
<b>Gestion des échanges</b>	Décentralisée, complexe	Centralisée, simplifiée
<b>Nombre total de demi-connecteurs à modifier en cas de modification d'une application</b>	Autant que d'applications connectées à cette application	Un seul
<b>Formats des données envoyées par l'application source et reçues par l'application réceptrice</b>	Doivent être identiques	Peuvent être identiques ou différents, si le format de l'application émettrice est plus complexe que le format de l'application réceptrice
<b>Nombre total de demi-connecteurs à développer/acheter pour n applications émettrices, envoyant chacune des données à n applications réceptrices</b>	$n^2$ ; croissance exponentielle	$4*n$ ; croissance linéaire
<b>Possibilité de supervision des échanges : alerte en cas de panne par exemple</b>	Impossible	Possible
<b>Possibilité d'enrichissement des données par des applications tierces</b>	Impossible	Possible

L'architecture du réseau du SIH étant à présent clarifiée, nous pouvons aborder la cartographie des flux d'informations du SIH au CHRMT

## 4. Cartographie des flux au CHRMT

### a. Périmètre de l'EDS (entrepôt de données de santé) du CHRMT

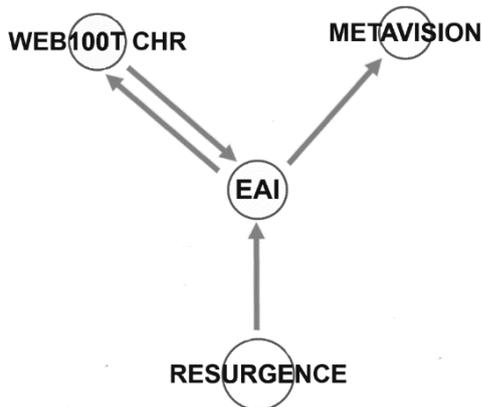
Nous aborderons dans cette partie uniquement la cartographie des flux d'informations qui seront inclus dans l'EDS, à savoir :

- les identités et mouvements du patient
- les observations médicales et paramédicales (compte-rendu, questionnaires...)
- les résultats de biologies
- le circuit des médicaments (prescription et délivrance)
- les résultats d'analyses biologiques
- les données de tous les champs du PMSI

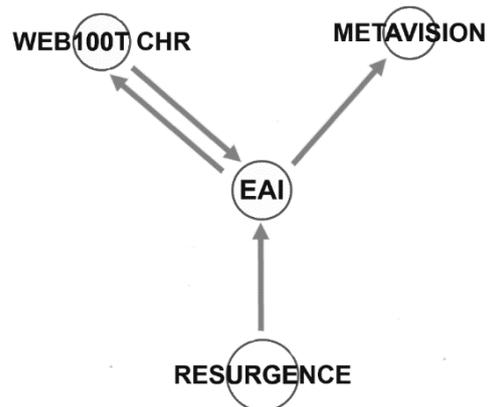
Par ailleurs, nous utiliserons une représentation simplifiée, qui pourrait faire croire à tort à la multiplicité des EAI, une communication directe entre les applications et l'unicité de logiciel pour la saisie. Comme indiqué dans le chapitre ci-dessus, ce n'est pas le cas. Sauf exception (système expert de validation automatique pour les examens de biologie les plus courants par exemple), les échanges entre applications sont centralisés et gérés par un seul et même EAI. Un premier exemple est la représentation graphique des flux des diagnostics médicaux PMSI

### b. Flux des diagnostics médicaux PMSI

Les figures 3 et 4 représentent le flux des diagnostics PMSI au CHR, de manière exacte puis de manière simplifiée. Les diagnostics sont saisis dans Résurgence pour le service des urgences et directement dans WEB100T pour le reste des services. Les diagnostics saisis dans Résurgence sont déversés dans WEB100T. Enfin l'intégralité des diagnostics PMSI est copiée de WEB100T vers Métavision, logiciel utilisé par les services de Réanimation et Surveillance Continue.



**Figure 3 : représentation exacte du flux des diagnostics PMSI dans le SIH du CHR de Metz-Thionville, 2020**



**Figure 4 : représentation simplifiée du flux des diagnostics PMSI dans le SIH du CHR de Metz-Thionville,**

Le format d'échange est l'Extensible MarkupLanguage (XML). Ce langage de description, basé sur un système de balises permettant de structurer du texte, est très utilisé pour les échanges d'informations entre applications. À l'inverse de son cousin le HypertextMarkupLanguage (HTML) utilisé essentiellement pour la structuration et l'affichage des pages web, il est générique, c'est-à-dire qu'il permet à l'utilisateur de créer ses propres balises.

La nomenclature utilisée est la CIM 10 dans sa version destinée au PMSI, maintenue par l'Agence Technique pour l'Information Hospitalière (ATIH).

L'EDS sera alimenté en diagnostics PMSI depuis WEB100T. Chaque diagnostic sera accompagné de l'Identifiant Permanent du Patient au sein de l'hôpital (IPP), du numéro de séjour, du numéro identifiant du professionnel l'ayant codé et du numéro de résumé d'unité médicale auquel il se rattache.

### c. Flux des observations cliniques

Les observations cliniques se répartissent en deux catégories : le texte libre (observations, compte-rendu opératoire, compte-rendu d'hospitalisation...) et les données structurées (questionnaires médicaux et paramédicaux : poids à l'entrée du patient...).

Pour la majorité des services de l'hôpital, ces données sont saisies directement dans le DPI DxCare. Certains services cependant utilisent des logiciels experts dédiés permettant la saisie de données qui seront en partie déversées dans DxCare. Ces logiciels sont : Résurgences (urgences), Aria (radiothérapie), Endoalpha (endoscopies), DxImage (imagerie), Anapath (anatomopathologie), Cardio Report (cardiologie interventionnelle), Ophtix (ophtalmologie),

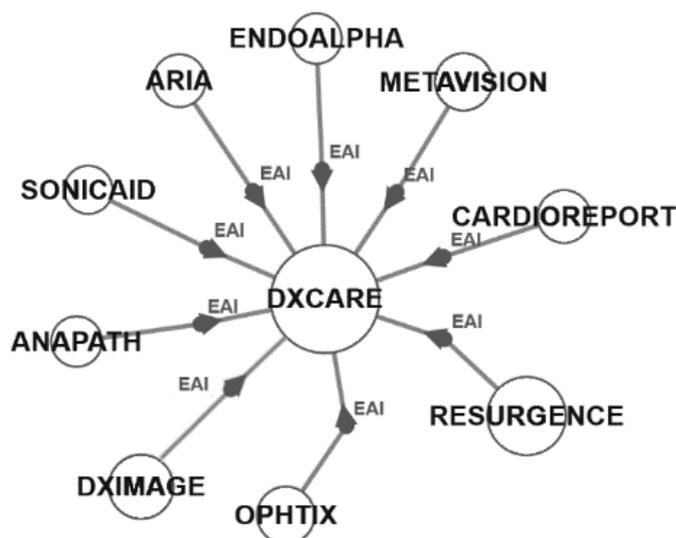
Metavision (réanimation et Surveillance Continue), et Sonicaid (échographie obstétrique). Ces flux sont représentés à la figure 4.

Le format pour des flux de questionnaires est le XML.

Les formats pour les flux des textes libres sont les suivants :

- HL7 : développé par l'organisation « HealthLevel 7 », il est spécifiquement utilisé pour les échanges de données de santé. La version utilisée est la version 2.7.
- HPRIM : acronyme de « Harmoniser et PRomouvoir l'Informatique Médicale », il s'agit d'un format propriétaire développé initialement pour transmettre des résultats de biologie mais qui dans sa troisième version permet également de véhiculer des images et des comptes-rendus divers.
- Le doc est un format texte propriétaire de Microsoft, le plus utilisé avant l'apparition du format docx.
- PDF est l'acronyme pour « Portable Document Format », un format initialement développé en 1993 par la société Adobe. L'objectif était alors de pouvoir afficher et imprimer un document de la même manière, quelque soit le logiciel ou le système d'exploitation utilisé. Il est certifié norme International Organization for Standardisation (ISO) 32000-1 depuis le premier janvier 2008 pour sa version 1.7 et ISO ISO 32000-2:2017 pour sa version 2.0. Cette dernière version est également très utilisée pour les échanges entre applications de données textuelles.

L'EDS sera alimenté en observations cliniques depuis DxCare. Chaque observation sera accompagnée : de l'IPP, du numéro de séjour, du numéro identifiant du professionnel de santé l'ayant rédigé et de la date et l'heure de sa saisie.



**Figure 4 : représentation simplifiée du flux des questionnaires et des observations médicales dans le SIH du CHR de Metz-Thionville, 2020**

#### d. Flux des actes de la Classification Commune des Actes Médicaux (CCAM)

##### i. Généralités

Les actes médicaux listés dans la Classification Commune des Actes Médicaux (CCAM), nationale et dans la CCAM descriptive, maintenues par l'assurance maladie d'une part et l'ATIH d'autre part, peuvent être réalisés en externe ou en hospitalisation. Ces flux sont représentés à la figure 5.

Le format de ces flux est le XML.

##### ii. Actes CCAM réalisés en externe

Les actes CCAM réalisés en externe sont codés directement dans Gestion administrative des Malades (GAM), un logiciel de facturation.

##### iii. Actes CCAM réalisés en hospitalisation dans les autres champs PMSI

Des actes CCAM sont réalisés en hospitalisation, dans tous les champs du PMSI, à savoir :

- Médecine, Chirurgie, Obstétrique et Odontologie (MCO)
- SSR
- Hospitalisation ADomicile (HAD)
- Psychiatrie (PSY)

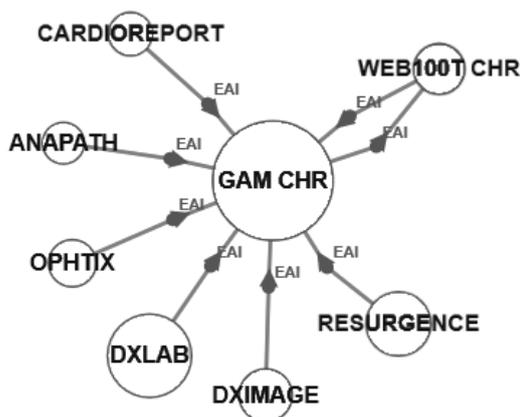
Ces actes :

- sont dans la plupart des services saisis dans le logiciel Web100T. Il s'agit d'un logiciel de référence pour la production des fichiers d'activités transmis à l'ARS, pour tous les champs du PMSI.
- pour les autres, sont saisis dans des logiciels experts dédiés : Résurgences (urgences), Anapath (anatomopathologie) Cardio Report (cardiologie interventionnelle), DxImage (imagerie), Ophtix (ophtalmologie).
- L'intégralité des actes CCAM réalisés en hospitalisation est ensuite déversée dans GAM-CHR.

À noter qu'il existe également un flux depuis GAM alimentant WEB100T avec les actes réalisés saisis initialement dans les logiciels experts.

iv. Alimentation de l'EDS

L'EDS sera alimenté en actes CCAM depuis GAM, du fait de son exhaustivité. Chaque acte sera accompagné de l'IPP, du numéro de séjour, du numéro identifiant du professionnel de santé l'ayant effectué et de la date et de l'heure de sa réalisation.



**Figure 5 : représentation simplifiée du flux des actes de la Classification Commune des Actes Médicaux dans le SIH du CHR de Metz-Thionville, 2020**

e. Flux des actes du Catalogue Spécifique des Actes de Rééducation et Réadaptation (CSARR) et E.D.G.A.R.

En parallèle des actes CCAM utilisables dans tous les champs PMSI, WEB100T gère d'autres nomenclatures d'actes réservées à certains champs PMSI : les actes du Catalogue Spécifique des Actes de Rééducation et Réadaptation (CSARR) en SSR et les actes E.D.G.A.R. en psychiatrie ambulatoire.

Ces actes CSARR et E.D.G.A.R. sont destinés à rester dans WEB100T et ne sont pas transmis à GAM. Ils sont nécessaires à la description de l'activité de l'établissement en SSR et en psychiatrie.

L'EDS sera alimenté en actes CSARR et E.D.G.A.R. depuis le logiciel WEB100T. Chaque acte sera accompagné de l'IPP, du numéro de séjour, du numéro identifiant du professionnel de santé l'ayant effectué et de la date et de l'heure de sa réalisation.

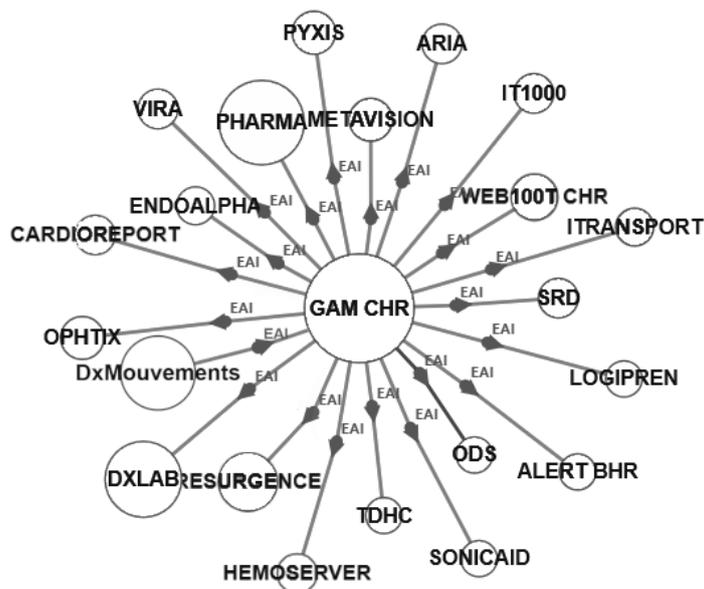
#### f. Flux des identités et mouvements des patients

Les identités des patients et les venues sont créées dans GAM. Les mouvements sont créés en temps réel dans DxMouvements, envoyés dans GAM puis retransmis aux autres logiciels du SIH. Ces flux sont représentés à la figure 6.

Ces flux sont au format HL7, version 2.3.1.

Les types de mouvement seront codés selon les instructions de l'ATIH [9]. Par exemple 01 pour « Retour prestation de soins dans une unité de soins de courte durée d'un autre établissement » ou 9 pour « sortie par décès ».

L'EDS sera alimenté en identités et en mouvements depuis GAM du fait de son exhaustivité. Les identités contiendront les noms, prénoms, adresses, dates de naissance et les IPP. Les codes de mouvements seront accompagnés des IPP, des numéros de séjour, des codes des unités fonctionnelles, des dates et heures d'arrivée et de sortie et des modes d'entrée et de sortie.



**Figure 6 : représentation simplifiée du flux des identités et mouvements dans le SIH du CHR de Metz-Thionville, 2020**

#### g. Flux des prescriptions et validation médicamenteuse

Au CHRMT, les prescriptions médicamenteuses se font au sein du logiciel Chimio pour les chimiothérapies anticancéreuses et dans DxCare pour les autres traitements et dispositifs médicamenteux implantables. Ces flux sont représentés à la figure 7.

Concernant les prescriptions effectuées dans DxCare : ces demandes sont envoyées au logiciel Pharma pour validation. Une fois le traitement préparé par la pharmacie, il est délivré au service qui se charge de la dispensation au patient puis cette information est saisie dans DxCare et enfin retransmise à Pharma.

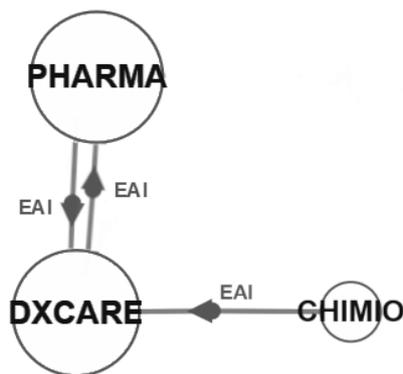
Pour les chimiothérapies anticancéreuses, les prescriptions et validations se font directement dans chimio ; ces informations sont ensuite déversées dans DxCare.

Ces flux sont au format PN13, qui permet de prendre en charge la totalité du circuit du médicament. Ils utilisent comme nomenclatures :

- les Unités Communes de Dispensation (UCD) pour les médicaments
- la Liste des Produits et Prestations (LPP) pour les DMI

Ces nomenclatures sont maintenues par l'assurance maladie avec une mise à jour au minimum hebdomadaire.

L'EDS sera alimenté par les données de prescriptions et de dispensation depuis DxCare. Chaque prescription et chaque dispensation seront accompagnées par l'IPP, le numéro de séjour, le numéro identifiant du professionnel de santé et par la date et l'heure de réalisation.



**Figure 7 : représentation simplifiée des flux de prescription et des plans de soin au CHR de Metz-Thionville, 2020**

#### h. Flux des résultats de biologie.

Différents automates produisent les résultats de biologie.

GEMweb est une application chargée de rapatrier les résultats des gaz du sang. IT1000 est une application chargée de rapatrier les autres biologies délocalisées (glucométries...). Biolink est une application permettant de faire le lien (middleware) entre GEMweb et IT1000 d'une part et DxLab d'autre part.

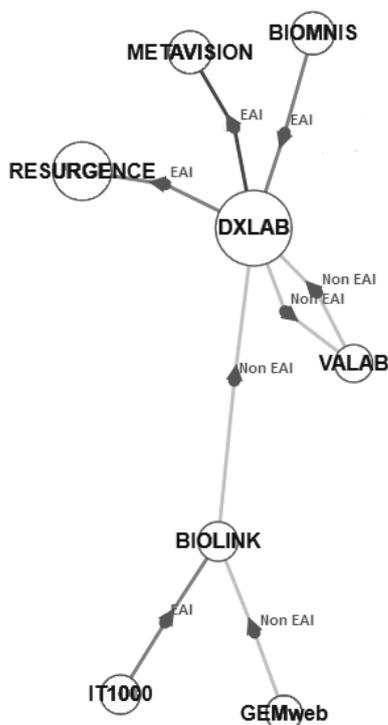
Les automates du laboratoire du CHRMT envoient leurs données directement dans DxLab. Une fois dans DxLab ces résultats sont validés manuellement par les médecins biologistes ou envoyés au serveur Valab (système expert de validation automatique pour les examens de biologie les plus courants).

D'autres résultats de biologie, provenant de laboratoires extérieurs au CHR, sont envoyés de l'application Biomnis vers Dxlab.

L'intégralité de résultats de biologie est ensuite envoyée de DxLab vers les dossiers patients de DxCare, Résurgence et Métavision. Ces flux sont représentés à la figure 8.

Les flux sont au format HL7 et HL2 ; la nomenclature utilisée est une nomenclature propre au CHR, maintenue par le laboratoire de biologie, avec une correspondance avec la nomenclature des actes de biologie médicale (NABM) maintenue par l'assurance maladie, pour la facturation.

L'EDS sera alimenté des résultats de biologie depuis DxCare. Chaque résultat sera accompagné de l'IPP, du numéro de séjour, de numéro identifiant du professionnel de santé prescripteur, de la date et de l'heure de réalisation, du numéro de l'automate et du statut de validation.



**Figure 8 : représentation simplifiée des flux de résultats biologiques au CHR de Metz-Thionville, 2020**

Les sources alimentant l'EDS étant définies, nous pouvons donc finalement aborder la partie sur les EDS : nous aborderons successivement les bases théoriques des EDS, la solution technique retenue par le CHRMT et enfin les activités de traitements prévues pour l'EDS du CHRMT.

## IV. EDS : bases théoriques et application au cas du CHRMT

### 1. Entrepôts de Données : bases théoriques.

Les outils de big data doivent permettre de répondre à une triple contrainte sur les données : contrainte de volume, contrainte de variété et contrainte de vélocité. Dans son ouvrage « Building the datawarehouse », [10] Inmon insiste sur la nécessité d'avoir un système opérationnel d'une part et un système dédié à la réutilisation des données qui y sont générées d'autre part. Les principales raisons de cette séparation sont :

- Les données utiles pour le système de production sont physiquement différentes des données utiles à l'analyse/l'information.
- Les technologies sur lesquelles s'appuient les processus opérationnels sont fondamentalement différentes de la technologie utilisée pour les besoins d'analyses/d'information.
- La communauté d'utilisateurs du système d'information opérationnel n'est pas la même que la communauté d'utilisateurs extrayant l'information et menant les analyses.
- Les caractéristiques des environnements opérationnels et d'analyse sont fondamentalement différentes.

Les entrepôts de données sont un type d'architecture logicielle optimisée pour la réutilisation et l'analyse de données. Mary Breslin les définit ainsi [11] :

- ils contiennent de grandes quantités de séries temporelles détaillées.
- le travail d'extraction des données opérationnelles, leur réconciliation et leur intégration dans les tables de la base de données relationnelles de l'EDS sont réalisés une fois pour toutes.
- ils permettent une réutilisation aisée des données des systèmes opérationnels par les utilisateurs finaux.

Dans le cadre hospitalier, ces séries temporelles concernent toutes les données générées lors de la prise en charge du patient : dosages biologiques, observations cliniques, données de dispositifs connectés (électrocardiogrammes, glucomètres...). Le travail d'extraction, de réconciliation et d'intégration est également complexe, du fait du grand nombre d'applications métiers présentes au sein du SIH. Mais puisqu'il n'est réalisé qu'une fois par du personnel spécialisé, il simplifie le travail de l'utilisateur final.

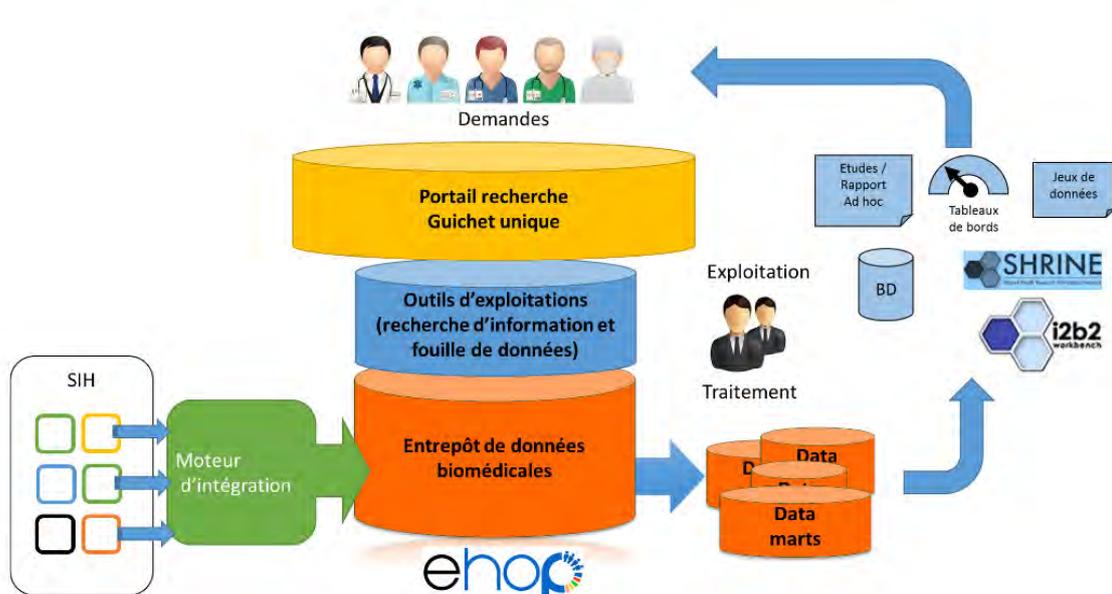
À ce jour, plusieurs solutions d'EDS sont disponibles sur le marché (ehop, Dr Warehouse..) ou en libre (I2B2...). Les membres de PAISaGE ont finalement opté pour l'EDS ehop pour les raisons suivantes :

- son intégration d'un moteur de TAL (Traitement Automatisé du Langage) avancé et de fonctions de déidentification de documents
- son faible coût d'exploitation, tant grâce aux coûts de licence que du faible nombre de postes équivalents temps plein nécessaire pour le maintenir en interne.
- il s'agit d'une solution déjà adoptée par le réseau d'EDS des Hôpitaux Universitaires du Grand Ouest, ce qui facilitera de futures coopérations.
- l'évolution imminente de son logiciel de gestion de base de données : passage de la technologie Oracle à postgresSQL, avec un gain de temps d'un facteur 5 à 10.
- l'expérience et la pérennité de l'entreprise chargée de sa commercialisation et de son intégration, Enovacom : crée en 2002 et actuellement propriété de la société Orange.

## 2. ehop : caractéristiques

Ehop est une solution développée par le laboratoire traitement du signal et de l'image rattaché au CHU de Rennes à partir du projet roogle (indexation des données patients issues du SIH) [3]; elle intègre à la fois un EDS et un portail de gestion des demandes [12].

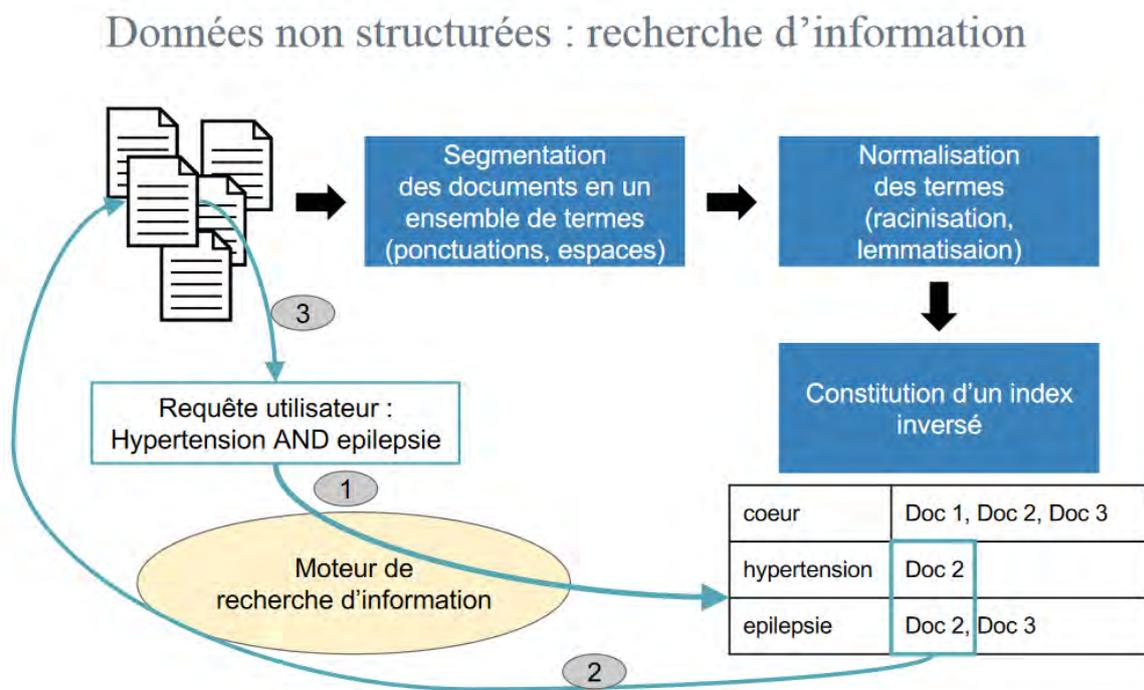
Les données sont intégrées dans l'entrepôt à partir du SIH par un moteur d'intégration. Après examen et validation des demandes, les données extraites (datamart) sont transmises au demandeur dans un cadre sécurisé, qui n'accède ainsi qu'aux données dont il a besoin et non à l'intégralité des données contenues dans l'EDS. Ceci est représenté à la figure 9 [12].



**Figure 9 : ehop, un entrepôt de données de santé avec portail de gestion des demandes intégré.[12]**

Cet entrepôt de nouvelle génération est capable de traiter les données structurées (PMSI, dosages biologiques...). Mais surtout, il intègre des outils de TAL, permettant une réutilisation plus poussée des données textuelles non structurées (compte-rendu, observations...) qui constituent la majeure partie des données médicales. Le moteur d'intégration découpe chaque phrase en token, les normalise (racinisation, lemmatisation...) et constitue un index inversé. Cette indexation permet une réduction drastique du temps de requêtage, par rapport à une approche consistant à parser l'intégralité des chaînes de caractères des documents contenus dans l'EDS.

Un exemple est présenté dans la figure 10 [12] : l'index inversé permet de retrouver rapidement le document 2, qui est le seul parmi les trois à faire référence à la fois aux concepts d'hypertension et d'épilepsie.



**Figure 10 : Constitution d'un index inversé du contenu des textes libres au sein d'ehop et recherche.[12]**

Par ailleurs ce moteur de TAL est capable de contextualisation, avec la gestion des antécédents médicaux et des négations, et également d'extraction de concepts (valeurs biologiques, posologie...). Des exemples sont donnés dans la figure 11 [12].

## Données non structurées : traitement automatique du langage

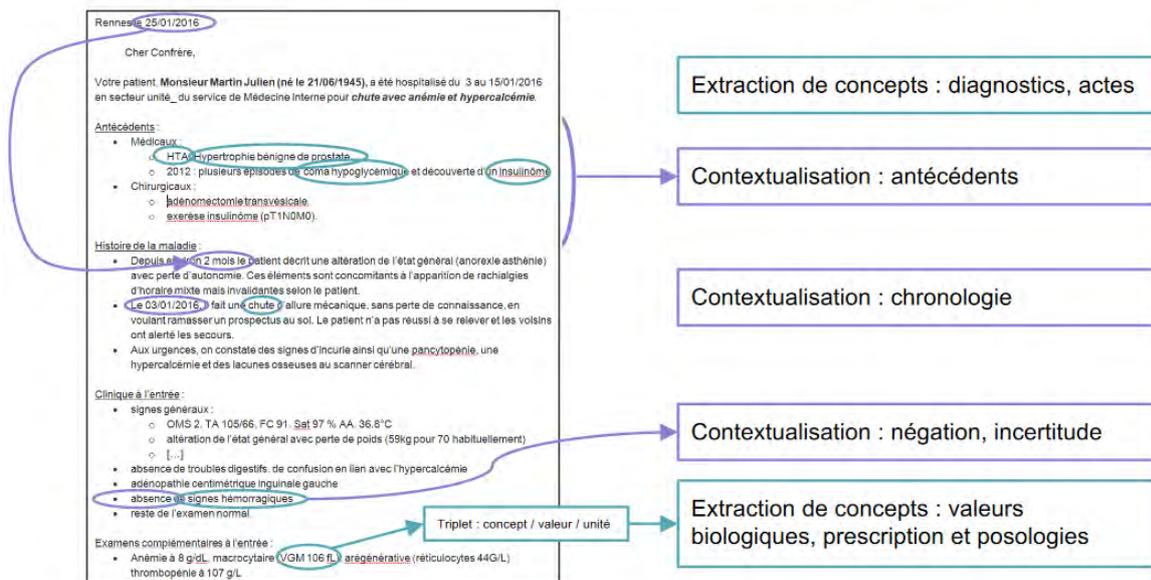


Figure 11 : exemple de contextualisation et d'extraction de concepts au sein d'ehop.[12]

Afin d'assurer au mieux la confidentialité des données, le schéma de base d'ehop est composé en trois parties :

- les documents nominatifs originaux.
- les documents déidentifiés pour les patients : ses données à caractère personnelles (nom, adresse...) sont masquées par un algorithme dédié.
- les identités des patients.

Les documents nominatifs originaux ne sont jamais, sauf exception, consultés.

Les requêteurs, sous autorité fonctionnelle médicale, n'ont accès en routine qu'aux documents déidentifiés. Si la réidentification du patient est nécessaire, par exemple pour lui proposer d'intégrer une étude, l'identité peut être révélée au cas par cas.

Le module de déidentification ne concerne pour l'instant que les patients ; un module de déidentification des professionnels de santé est en cours de développement.

Ayant à présent éclairci les aspects techniques du SIH et de l'EDS, nous pouvons aborder à présent les utilisations qui en seront faites au sein du CHRMT.

### 3. Activités de traitement de l'EDS au sein du CHRMT

Les données contenues dans l'EDS seront réutilisées à des fins de recherche et à des fins de gestion.

Les activités de traitements à finalité de recherche seront :

- les recherches relevant de la méthodologie de référence (MR) de la CNIL numéro 001 (MR001) : recherche dans le domaine de la santé avec recueil du consentement.
- les recherches relevant de la MR de la CNIL numéro 003 (MR003) : recherche dans le domaine de la santé sans recueil du consentement.
- les recherches relevant de la MR de la CNIL numéro 004 (MR004) : recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé, dont les études de faisabilité.
- les recherches ne relevant d'aucune MR, dont les recherches utilisant des données du SNDS.
- la sélection des patients à inclure dans les essais cliniques.

Les activités de traitements relevant de la gestion du CHRMT et de l'EDS seront :

- la gestion des utilisateurs d'ehop et de leurs droits d'accès.
- la traçabilité des accès aux données de l'entrepôt.
- l'exploitation du système d'information : débogage, mise à jour...
- le codage PMSI : optimisation et automatisation.
- le suivi des indicateurs de qualité, de sécurité des soins et des analyses de pratiques.
- le pilotage hospitalier : contrôle de facturation, tableau de bord...
- les vigilances : pharmacovigilance, matériovigilance, identitovigilance, hémovigilance...

## V. Article : quel cadre réglementaire pour les entrepôts de données de santé hospitaliers ?

# Quel cadre réglementaire pour les entrepôts de données de santé hospitaliers ?

Vianney GUARDIOLLE<sup>1</sup>, Aurélie BANNAY<sup>2</sup>, Julien GODET<sup>3</sup>, Benjamin BETHUNE<sup>1</sup>, Martine FAULON<sup>1</sup>, Pierre FIORE<sup>1</sup>, Corinne WEISSLINGER<sup>1</sup>, Joris MULLER<sup>1</sup>, Christophe GOETZ<sup>1,5</sup>

<sup>1</sup> Département d'Information Médicale, Centre Hospitalier Régional Metz-Thionville, 1 allée du château CS45001 57085 Metz Cedex 03

<sup>2</sup> Service d'Évaluation et d'Information Médicales, Centre Hospitalier Régional Universitaire de Nancy, 1 rue Foller 54035 NANCY

<sup>3</sup> Groupe Méthodes Recherche Clinique (GMRC) - Hôpitaux Universitaires de Strasbourg - Hôpital Civil - 1 pl. de l'hôpital 57000 Strasbourg France

<sup>4</sup> Service de Santé Publique, Hôpitaux Universitaires de Strasbourg, 1 place de l'hôpital 67000 Strasbourg France

<sup>5</sup> Plateforme d'Appui à la Recherche Clinique, Centre Hospitalier Régional Metz-Thionville, 1 allée du château CS45001 57085 Metz Cedex 03

### Résumé Français

Position du problème : Les entrepôts de données de santé (EDS), permettent une réutilisation plus aisée des données produites lors du soin et sont déployés dans les hôpitaux français. À ce jour (janvier 2021), il n'existe pas de méthodologie de référence de la Commission Nationale Informatique et Libertés (CNIL) spécifique à la mise en place des EDS dans les

établissements de santé. L'objectif de cet article est de faire la synthèse de la réglementation à laquelle les EDS hospitaliers nominatifs français sont soumis.

Méthodes : les sources utilisées étaient le Règlement Général de la Protection des Données (RGPD), la Loi Informatique et Libertés (LIL), le Code de Santé Publique (CSP), le Code Pénal (CP), le Code de la Sécurité Sociale et les ressources mises à disposition par la CNIL.

Résultats : des données à caractère personnelles étant traitées, le RGPD, intégré à la nouvelle LIL, s'applique. La base légale des activités de traitement n'étant pas exclusivement le consentement, une demande autorisation préalable à toute constitution et exploitation de l'EDS doit donc être adressée à la CNIL. Par ailleurs, des données de santé, dites sensibles, étant traitées à grande échelle, des risques plus importants pour les droits et libertés des patients existent : une analyse d'impact relative à la protection des données doit donc également être effectuée. La demande d'autorisation devra préciser la gouvernance, qui afin de garantir la protection et la confidentialité des données, intégrera le Département d'Information Médicale ; l'autorité fonctionnelle sur l'équipe requêtant les données au sein de l'entrepôt devant à minima être médicale. Par ailleurs, pour chacune des activités de traitement, la demande d'autorisation devra préciser notamment : la finalité, les catégories de données traitées, la base légale de traitement et les dérogations aux interdictions de traiter les données dites et perçues sensibles, le responsable de traitement, les mesures de sécurité et les droits des patients.

Conclusion : une demande d'autorisation préalable à toute constitution et exploitation devra être adressée à la CNIL. Elle intégrera une AIPD, le détail de chaque activité de traitement et les mesures prises pour assurer la confidentialité et la sécurité des données.

## **Résumé Anglais**

Context: Clinical Data Warehouses (CDW) allow for reuse of data obtained as part of previous medical care. They have been deployed in French hospitals for about fifteen years. To date, there is no specific reference guide (Méthodologie de Référence-MR) produced by the French data protection authority, the Commission Nationale Informatique et Libertés (CNIL) for clinical data warehousing in hospitals. This article aims at summarizing the existing legal constraints on data warehousing in French hospitals.

Method: our sources are the General Data Protection Regulation (GDPR), the Loi Informatique et Libertés (data protection act LIL), the Code de Santé publique (public health code RCSP), the Code Pénal (penal code - CP), the Code de la Sécurité Sociale (social security system code) and the documents produced by the CNIL.

Results: Since personal data are processed in data warehouses, the European GDPR, integrated in the new version of the French LIL, applies. Because explicit consent is not the only legal basis, an authorization delivered by the CNIL beforehand is mandatory. Furthermore, health data being sensitive and processed at a large scale, a data protection impact assessment (DPIA) is required. The Département d'Information Médicale (department of medical information - DIM) must be part of the governance process in order to guarantee confidentiality and the protection of data. The request for information from databases must thus be made specifically by a competent medical authority. The authorization request addressed to the CNIL must eventually include information on its purpose, the categories of personal data that are being processed, the legal basis used, the identity and the contact details of both the data controller and the data protection officer, the security measures implemented and the patients' rights. These rights apply differently depending on the legal basis used.

Conclusion: The risks to patients' rights and liberties are high in clinical data warehouses, hence the application of GDPR and the LIL, which aim at enforcing data protection and confidentiality. An authorization request, which includes a DPIA, must be submitted to the CNIL prior to the process of data.

## **Manuscrit**

### **1) Introduction**

Dans le cadre hospitalier, la transformation numérique de ces dernières décennies a facilité le recueil de l'information médicale. Ceci notamment par la mise en place du dossier patient informatisé, constitué de données issues de divers logiciels métiers spécialisés (informations administratives, observations cliniques, biologie, prescriptions, imagerie...). La réutilisation de ces données à des fins de recherche, de pilotage hospitalier et d'aide à la décision thérapeutique nécessite de les décloisonner et de prendre en compte leur variété, leur volume avec une grande vitesse de traitement. Les entrepôts de données de santé (EDS) sont des outils permettant avec une grande flexibilité et de hautes performances de répondre à ces exigences. En France, le CHU de Rennes a été précurseur dans la mise en place de ces EDS avec l'intégration d'un volume de données croissante. En 2010, il comptait 2,1 millions de documents médicaux contre 49,2 millions pour l'année 2018 [1]. Ces données concernent l'identité des patients et des professionnels de santé, la structure de l'établissement, les comptes-rendus, les prescriptions et administrations de traitements, la biologie et le Programme de Médicalisation des Systèmes d'Information (PMSI). L'EDS de l'hôpital européen Georges Pompidou comptait 20,8 millions d'observations cliniques en 2009 contre 122 millions en 2016 [2]. En parallèle, le nombre d'études référencées dans PubMed® se rapportant à un EDS (« clinical data warehouse ») a cru exponentiellement ces deux dernières décennies : 6 publications en 1999, 11 en 2004, 30 en 2009, 78 en 2014 et 193 en 2019.

À ce jour (janvier 2021), de nombreux hôpitaux sont engagés dans une démarche pour doter leur système d'information hospitalier (SIH) d'un EDS. L'exploitation de ces données personnelles sensibles exige un travail préliminaire important afin de prendre en compte les problèmes de sécurité, de confidentialité et de réglementation. Les obligations réglementaires

visant à limiter ces risques sont évolutives. La principale est la mise en place du règlement général de la protection des données européen [3] n°2016/679 (RGPD), applicable depuis le 25 mai 2018 puis transcrit en droit national dans la nouvelle loi informatique et liberté [4] (LIL), entrée en vigueur le premier juin 2019. La littérature spécialisée rends compte de ces changements dans le cadre général [5] ; la Commission Nationale Informatique et Liberté(CNIL) n'a néanmoins pas publié de méthodologie de référence (MR) portant spécifiquement sur les EDS hospitaliers nominatifs. Or les enjeux sont importants : la mauvaise prise en compte des obligations réglementaires en matière de sécurité et de confidentialité sur des données de santé à caractère personnel peut mener à de lourdes condamnations [6].

L'objectif de cet article est de faire la synthèse de la réglementation à laquelle les EDS hospitaliers nominatifs sont soumis. Nous aborderons successivement quelques définitions, le cadre réglementaire auquel sont soumis les EDS, l'impact de ce cadre sur la demande d'autorisation, avant de conclure. Les problématiques de patrimonialité des données, déjà traitées ailleurs [7], ne seront pas discutées dans cet article.

## **2) Définitions**

La CNIL identifie trois typologies de données à caractère personnel dans son guide « AIPD : les bases de connaissance » [8] : les données courantes, les données dites sensibles au sens RGPD et les données perçues sensibles au sens RGPD. Ces concepts doivent d'abord être définis, ainsi que celui de « traitement de données ».

### **a) Définition des données à caractère personnel**

L'article 4 du RGPD définit une donnée à caractère personnel comme « toute information se rapportant à une personne physique identifiée ou identifiable ». Une personne physique

identifiable étant « une personne physique qui peut être identifiée, directement ou indirectement ».

### **b) Définition des données “dites sensibles” et de santé**

Les données dites sensibles (ou “à caractère particulier”) au sens du RGPD sont des données relevant des opinions philosophiques, politiques, religieuses, syndicales, de la vie sexuelle, des données de santé et des origines raciales ou ethniques.

Les données à caractère personnel concernant la santé sont définies par la CNIL comme « les données relatives à la santé physique ou mentale, passée, présente ou future, d’une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l’état de santé de cette personne. » [9]. Elles peuvent par exemple correspondre à des diagnostics, des prescriptions, des actes techniques, des numéros de séjour, des numéros identifiants de patient (cf. section 3 du chapitre 3 du titre II de la LIL). Elles sont considérées par le RGPD comme des données dites « sensibles »

### **c) Définition des données “perçues sensibles”**

Les données à caractère personnel perçues sensibles sont les données bancaires, les données biométriques ou le numéro d’inscription des personnes au répertoire national d’identification des personnes physiques (NIR).

### **d) Définition des données à caractère personnel courantes**

Les données à caractère personnel courantes sont les données à caractère personnelles qui n’appartiennent ni à la catégorie des données sensibles au sens RGPD, ni à la catégorie des données perçues sensibles, comme l’état civil.

### **e) Définition d'un traitement de données**

Un traitement de données se définit par toute opération portant sur des données (collecte, conservation, copie, extraction, consultation, transmission, modification...), quel que soit le procédé utilisé (article 4 du RGPD). Sauf exception le RGPD, transcrit dans la nouvelle LIL, s'applique dès lors qu'un traitement est appliqué à des données à caractère personnel (article 2 de la LIL).

## **3) Cadre réglementaire auquel sont soumis les EDS**

### **a) Cadre général**

Le RGPD s'applique donc dans le cadre des EDS hospitalier et impose des bonnes pratiques visant à garantir le respect des droits et des libertés des patients.

La CNIL a publié des MR permettant de simplifier les démarches administratives lors la réutilisation des données, dont la méthodologie de référence 004 (délibération n°2018-155 du 3 mai 2018). Mais il n'existe à ce jour aucune MR concernant la constitution de l'entrepôt en elle-même. La base légale des activités de traitements n'étant pas systématiquement le consentement explicite, une demande d'autorisation à la CNIL devra donc être effectuée avant toute constitution et exploitation d'un EDS [10].

Un EDS contient par définition des données à caractère personnel concernant la santé pour un grand nombre de personnes. Du fait de ce caractère dit sensible au sens RGPD et de la grande échelle du traitement, celui-ci est « susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques ». Une analyse d'impact relative à la protection des données (AIPD) doit donc être intégrée à la demande d'autorisation à la CNIL (cf. articles 62 et 63 de la LIL et alinéa 3.b de l'article 35 du RGPD). Le RGPD et la LIL ne précisent pas le nombre de personnes à partir duquel un traitement est considéré comme à grande échelle. Néanmoins

par définition, un EDS est un outil de données massives, qu'il est raisonnable de considérer comme relevant de la grande échelle. Par ailleurs, la CNIL indique explicitement sur son site internet qu'une AIPD est obligatoire dans le cadre des EDS [10].

Les démarches administratives auprès de la CNIL doivent être effectuées préalablement à la mise en place de l'EDS, sous peine de s'exposer aux sanctions mentionnées l'article 226-16 du Code pénal (CP) : « Le fait, y compris par négligence, de procéder ou de faire procéder à des traitements de données à caractère personnel sans qu'aient été respectées les formalités préalables à leur mise en œuvre prévues par la loi est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende. ». Par ailleurs, en cas de viol du secret professionnel, l'article 226-13 du CP prévoit que « la révélation d'une information à caractère secret par une personne qui en est dépositaire soit par état ou par profession, soit en raison d'une fonction ou d'une mission temporaire, est punie d'un an d'emprisonnement et de 15 000 euros d'amende. ». La nécessaire conformité aux lois particulières du pays est rappelée au paragraphe 4 de l'article 9 du RGPD : « Les États membres peuvent maintenir ou introduire des conditions supplémentaires, y compris des limitations, en ce qui concerne le traitement des données génétiques, des données biométriques ou des données concernant la santé ». Le respect du RGPD passe donc aussi par le respect des spécificités de la réglementation française.

Ceci est cohérent avec la hiérarchie des normes qui se décline par ordre d'importance décroissante : du bloc constitutionnel (constitution française par exemple), du bloc de conventionnalité (RGPD par exemple), du bloc de légalité (Article L6132-3 du code de santé publique par exemple), des principes généraux du droit, des règlements (décrets et arrêtés, article R6113-11-3 du CSP par exemple) et des actes administratifs (circulaire et directives, par exemple l'instruction n° DGOS/R1/DSS/1A/2020/52 du 10 septembre 2020) et enfin de la jurisprudence.

## **b) Secret professionnel et dérogations**

Le colloque singulier médecin/patient ne peut se faire que dans un climat de confiance réciproque, dont le secret médical/professionnel est un des piliers. La confidentialité des données est donc essentielle pour une bonne prise en charge du patient. Cette confidentialité est donc également essentielle pour une bonne qualité des données potentiellement réutilisées, puisque le contenu du dossier patient, intégré à l'EDS, reflète de cette prise en charge. C'est pourquoi l'article L1110-4 du code de santé publique (CSP) garantit au patient « le droit au respect de la vie privée » lors de sa prise en charge dans un établissement de santé. Quelques dérogations nécessaires au secret professionnel ont été mises en place dans l'intérêt du patient et de la collectivité, dont :

- La dérogation pour l'équipe de soin, défini à l'article L1110-12 du Code de Santé Publique (CSP) comme « l'ensemble de professionnels qui participent directement au profit d'un même patient à la réalisation d'un acte diagnostique, thérapeutique, de compensation du handicap, de soulagement de la douleur ou de prévention de perte d'autonomie, ou aux actions nécessaires à la coordination de plusieurs de ces actes ». La finalité étant une coordination de la prise en charge médicale et paramédicale du patient.
- Les dérogations au médecin responsable de l'information médicale de l'établissement/du territoire et de son équipe (article L6113-7 du CSP et Article L6132-3 du CSP). Les finalités étant par exemple la production des données PMSI nécessaires à la tarification à l'activité ou la production de données agrégées nécessaires au pilotage hospitalier.
- Les dérogations prévues par l'article 226-14 du Code pénal. La finalité étant la protection des personnes subissant des sévices en incapacité de se protéger ou des personnes à caractère dangereux pour elle-même ou autrui.

Le cadre réglementaire étant fixé, nous pouvons aborder son impact sur la demande d'autorisation.

#### **4) Impact du cadre réglementaire sur la demande d'autorisation**

La gouvernance doit être détaillée dans la demande d'autorisation préalable. En outre, chaque activité de traitement doit explicitement y être décrite. Pour chacune, doivent notamment être précisés (article 33 de la LIL) : une finalité, les catégories de données nécessaires à cette activité de traitement, une base légale de traitement, le responsable de traitement, les droits assurés aux personnes dont les données sont traitées et les mesures de protection des données.

##### **a) La gouvernance de l'EDS**

Des mesures organisationnelles, préventives et d'audits, sont nécessaires afin de limiter le risque d'accès frauduleux, accidentel ou illégitime aux données de l'EDS, et ainsi d'en préserver la confidentialité. La structure ayant en charge cette mission sera d'autant plus efficace qu'elle disposera de tout ou partie de l'autorité fonctionnelle sur le personnel requêtant dans l'EDS. Cette autorité fonctionnelle, qui concerne le pouvoir d'organisation et d'encadrement du travail est à distinguer de l'autorité hiérarchique, qui elle concerne le pouvoir de nomination, d'avancement et de mutation. [11]

Au sein des hôpitaux français, les départements d'information médicale (DIM) traitent les données médicales nominatives à des fins médico-économiques et participent à leur protection depuis 1991 (circulaire DH/P.M.S.I. n° 303 du 24 juillet 1989). À ce jour, le personnel qui est sous autorité fonctionnelle du médecin responsable de l'information médicale de l'établissement/du territoire :

- bénéficie d'une dérogation au secret professionnel pour un grand nombre d'activités de traitements sur les données médicales nominatives : qualité des données, pilotage hospitalier, PMSI... Par opposition, le personnel de la direction du système

d'information ne peut traiter des données à caractère personnel que si la finalité de traitement est la mise en place et la mise à jour des composants du SIH. (article R6113-4, R6113-5, article R6113-8, article R6113-10, R6113-11-1, et Article R6113-11-3 du CSP).

- participe à la gestion des modalités d'attribution et de contrôle des autorisations d'accès et réalise des audits d'accès aux données (article R6113-6 et R6113-11-3 du CSP).
- bénéficie explicitement d'une série d'exemptions d'obligations administratives lors de la mise en place de traitements de données : par exemple, l'obligation de faire une demande préalable à la CNIL avant toute nouvelle activité de traitement en rapport avec ses missions (alinéa 4 de l'article 65 et alinéa 3 de l'article 66 de la LIL).

Par ailleurs, la jurisprudence confirme l'importance du rôle des DIM pour la protection des données médicales au sein des hôpitaux. En 2013, la direction du Centre Hospitalier de Saint-Malo a ainsi été mise en demeure pour avoir sous-traité des traitements de données à des fins du PMSI en dehors de l'autorité fonctionnelle du médecin responsable du DIM (décision n°2013-037 du 25 septembre 2013).

Il semble donc légitime que le DIM, fasse partie intégrante de la gouvernance des EDS hospitaliers. Avec à minima, une autorité fonctionnelle médicale sur le personnel requêtant l'EDS.

Une structure couramment utilisée dans la gouvernance des EDS est le Comité Scientifique et Ethique (CSE). Comme son nom l'indique, il est composé de personnes disposant de compétences éthiques ou scientifiques. Par exemple, des médecins, des chercheurs, des juristes, des experts en éthique, des méthodologistes, des statisticiens... Ses rôles sont de :

- vérifier et valider le caractère éthique ainsi que la pertinence scientifique des projets de réutilisation des données de l'EDS.
- faciliter la publication dans les revues anglophones, ces dernières exigeant le plus souvent l'accord d'un comité d'éthique.

Enfin, le dernier aspect de la gouvernance concerne les orientations stratégiques de l'exploitation de l'EDS. Elles sont couramment définies par un comité de pilotage. Ce comité est géré par la direction administrative, conseillée par des représentants médicaux et scientifiques. Son rôle est purement stratégique et ne concerne pas la gestion des accès aux données.

L'impact du cadre sur la gouvernance étant achevé, nous pouvons à présent aborder l'impact du cadre sur les activités de traitements.

## **b) La finalité**

La demande d'autorisation préalable à la création d'un EDS doit en préciser la ou les finalité(s) de traitement. La finalité de traitement est l'objectif principal de l'utilisation des données. Si le responsable de traitement souhaite utiliser les données à d'autres fins que celles prévues dans la demande d'autorisation, il doit faire une nouvelle demande d'autorisation. [12]. Dans le cas contraire, il s'agit d'un détournement de finalité, qui est passible de 300 000 euros d'amende et de cinq ans d'emprisonnement [13], cf article 226-21 du CP. Il convient donc d'anticiper au mieux les différentes réutilisations prévues des données de l'EDS afin de les intégrer d'emblée dans la demande d'autorisation initiale.

Une série d'activités de traitements génériques est disponible au tableau 1, avec pour chacune : une finalité, une base légale, une dérogation à l'interdiction de traiter les données dites sensibles et le NIR, les textes réglementaires et le mode d'information et de consentement.

Tableau 1 : cas génériques d'activités de traitements mis en œuvre dans le cadre d'un entrepôt de données de santé hospitalier, selon la réglementation en vigueur en novembre 2020.

Traitement	Base légale Article 6 RGPD	Textes justifiant la base légale	Dérogation à l'interdiction de traiter des données dites sensibles au sens RGD Article 9 RGDP	Dérogation à l'interdiction de traiter le NIR Décret NIR Article 30 CSP	Destinataires	Information et consentement	Autres Contraintes
<b>Recherche</b>							
Recherche couverte par la MR001	Mission d'intérêt public	délibération n° 2018-153 du 3 mai 2018 Articles 66 et 73 de la LIL	Recherche scientifique	Utilisation du NIR exclue de la MR001	Cités à l'alinéa 2.3 de la délibération n° 2018-153 du 3 mai 2018	Consentement du patient obligatoire par écrit	Passage par CSE Passage par CPP Catégorie 1 loi Jardé : autorisation ANSM Catégorie 2 loi Jardé : information ANSM Engagement de conformité MR001
Recherche couverte par la MR003	Mission d'intérêt public	Délibération n° 2018-154 du 3 mai 2018 Articles 66 et 73 de la LIL	Recherche scientifique	Utilisation du NIR exclue de la MR003	Cités à l'alinéa 2.3 de la délibération n° 2018-154 du 3 mai 2018	Consentement du patient obligatoire par oral	Passage par CSE Passage par CPP Information à l'ANSM Engagement de conformité MR003
Recherche couverte par la MR004	Mission d'intérêt public	Délibération n°2018-155 du 3 mai 2018 Articles 66 et 73 de la LIL	Recherche scientifique	Utilisation du NIR exclue de la MR004	Cités à l'alinéa 2.3 de la délibération n° 2018-155 du 3 mai 2018	Régime de non-opposition du patient par défaut	Engagement de conformité MR004 Passage au CSE
Recherche non couverte par une MR	Mission d'intérêt public	LOI n° 2009-879 du 21 juillet 2009	Recherche scientifique	Alinéa 2 de l'article 30 du CSP : finalité exclusive de recherche scientifique, sous réserve de cryptographie	Variable	Variable	Passage au CSE RIPH : +/- passage par CPP +/- ANSM

Pré-Screening	Mission d'intérêt public	Article L1110-4 du CSP	Recherche scientifique	Alinéa 2 de l'article 30 du CSP : finalité exclusive de recherche scientifique, sous réserve de cryptographie	Équipe de soins (art L1110-12 du CSP)	Régime de non-opposition du patient par défaut	Dépôt d'un dossier de demande à la CNIL hors RIPH : dépôt d'un dossier de demande au HDH  Passage au CSE Tous les traitements de recensements seront communiqués au DPO qui les ajoutera au registre dédié.
<b>Gestion de l'EDS</b>							
Gestion des utilisateurs de l'EDS et de leurs droits d'accès	Obligation légale du responsable de traitement	Article R6113-6 du CSP	Intérêt public dans le domaine de la santé publique	Non applicable	DIM	Information du patient via notice d'information remise lors de l'admission	
Traçabilité des accès aux données de l'EDS	Obligation légale du responsable de traitement	Article R6113-6 du CSP	Intérêt public dans le domaine de la santé publique	Non applicable	DIM	Information du patient via notice d'information remise lors de l'admission	
Administration et exploitation du Système d'Information	Intérêt légitime du responsable	Article R6113-1 du CSP	Intérêt public dans le domaine de la santé publique	Non applicable	DIM Équipe gérant l'entrepôt	Information du patient via notice d'information remise lors de l'admission	
<b>PMSI, qualité, pilotage, vigilances</b>							
Optimisation et automatisation du Codage PMSI	Obligation légale du responsable de traitement	Articles L6113-7, L6132-3 et R6113-10 du CSP	Intérêt public dans le domaine de la santé publique	DN,2.B.4 .Remontées d'informations nominatives vers les organismes d'assurance maladie	DIM	Information du patient via notice d'information remise lors de l'admission	
Suivi des indicateurs de	Obligation légale du	Article R162-36 du CSP	Intérêt public dans le domaine de la santé	DN,2.B.1. référencement des données de santé au	Équipe de soins DIM	Information du patient via notice	

qualité, de sécurité des soins et des analyses de pratiques Utilisation des données de l'EDS dans le cadre du pilotage hospitalier	responsable de traitement  Obligation légale du responsable de traitement	Articles R6113-8 et R6113-11-1 du CSP	publique  Intérêt public dans le domaine de la santé publique	moyen du numéro d'inscription au répertoire national d'identification des personnes physiques  DN,2.B.5 : opérations liées à la facturation et à la prise en charge financière des dépenses de santé	Tableaux de bord établissement (données agrégées) : DH, DF, DIM, chef de pôle, chef de service, cadres supérieurs Bilan d'activité du service : chef de service/praticien demandeur, DH, DIM, chef de pôle, chef de service, cadres. Accès différencié selon la hiérarchie. Dénombrement de cas pour les études de faisabilité d'essais cliniques : chef de service, praticien demandeur Données de facturation : DF, DIM	d'information remise lors de l'admission  Information du patient via notice d'information remise lors de l'admission
Vigilances	Obligation légale du responsable de traitement	Article R5121-170 du CSP Article R5212-4 du CSP Article R1112-3 du CSP Article R1221-43 du CSP Article R5212-14 du CSP	Intérêt public dans le domaine de la santé publique	DN,2.B.10 : gestion et suivi des alertes sanitaires	Équipe de soins DIM Réfèrent pharmacovigilance Réfèrent matériovigilance Cellule d'identitovigilance Correspondant hémovigilance	Information du patient via notice d'information remise lors de l'admission

---

ANSM : agence nationale de sécurité du médicament et des produits de santé

CPP : comité de protection des personnes

CSE : conseil scientifique et éthique

DN : Décret NIR, n° 2019-341 du 19 avril 2019

DF : direction des finances

DH : directeur/directrice de l'hôpital

DIM : Département d'Information Médicale

DPO/DPD : délégué à la protection des données

EDS : Entrepôt de données de santé

HDH : Health Data Hub

LIL : loi informatique et libertés

NIR : numéro d'inscription des personnes au répertoire national d'identification des personnes physiques, plus connu sous le nom de « numéro de sécurité sociale »

MR : méthodologie de référence de la CNIL

RIPH : recherche impliquant la personne humaine

RGPD : règlement général sur la protection des données

---

### **c) Catégories de données**

Pour chaque activité de traitement, doivent être décrites les catégories de données nécessaires, avec la date de début de recueil. Il est inutile de détailler variable par variable. Par exemple, « données de biologie issues du logiciel A » entre 2012 et 2019 et « données de biologie issues du logiciel B depuis 2020 » est suffisant : la liste exhaustive des examens biologiques concernés n'est pas nécessaire.

Pour chaque activité de traitement, une base légale doit être justifiée ; si des données dites ou perçues sensibles sont traitées, des contraintes supplémentaires s'appliquent.

### **d) Base légale de traitement et dérogations aux interdictions de traiter les données dites et perçues sensibles**

Chaque activité de traitement doit se fonder sur une seule base légale [14], pour certaines justifiées par la réglementation.

Pour les traitements de données à caractère personnel courantes, le respect d'une des bases légales listées aux articles 6 du RGPD et 5 de la LIL est une condition nécessaire et suffisante pour valider le critère de licéité.

Pour les traitements de données dites sensibles au sens RGPD, deux interprétations contradictoires existent.

Une première interprétation, issue du Comité des Ministres aux États membres en matière de protection des données relatives à la santé [15], stipule que pour les activités de traitement de données dites sensibles au sens RGPD :

- l'article 6 du RGPD ne s'applique pas

- le respect d'une des bases légales listées à l'article 9 du RGPD est une condition nécessaire et suffisante pour valider le critère de licéité.

Une deuxième interprétation soutenue par la CNIL stipule que pour valider le critère de licéité, chaque activité de traitement de données dites sensibles au sens RGPD doit cumuler deux conditions : [16]

- le respect d'une des bases légales de traitement, listées aux articles 6 du RGPD et 5 de la LIL.
- le respect d'une des dérogations à l'interdiction de traiter les données dites sensibles au sens RGPD, listées aux articles 9 du RGPD et 6 de la LIL.

En pratique, la CNIL étant l'organisme chargé de valider la conformité à la législation de la demande d'autorisation, les auteurs conseillent de suivre son interprétation.

Pour les traitements de données utilisant le NIR, chaque activité de traitement doit cumuler ces conditions :

- le respect d'une base légale de traitement, listées aux articles 6 du RGPD et 5 de la LIL.
- le respect d'une des dérogations à l'interdiction de traiter le NIR, dont certaines sous conditions de chiffrement, à l'article 30 de la LIL ou bien listées dans le décret NIR n°2019-341 du 19 avril 2019.

#### **e) Responsable de traitement et accès aux données**

Enfin, pour chaque activité de traitement, doit être désigné un responsable de traitement, défini par l'article 4 du RGPD comme « la personne physique ou morale, l'autorité publique, le service ou un autre organisme qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement ». Dans le cadre des établissements de santé en France, il

s'agit du directeur de l'établissement. À noter que le responsable de traitement et l'exécutant du traitement peuvent être deux personnes différentes, avec des droits et des devoirs différents. Ainsi dans le cadre des EDS hospitaliers, le responsable de traitement n'a pas le droit de réaliser lui-même les traitements ou d'accéder aux données.

## **f) Droits des patients**

### **i) Droit à l'information**

#### **1. Généralités sur le droit à l'information**

L'article 5 du RGPD indique que les données à caractère personnel doivent être traitées de « manière licite, loyale et transparente au regard de la personne concernée ».

L'article 71 de la LIL impose entre autres aux établissements de santé d'informer les patients sur les traitements de données les concernant. Cette information doit être concise, transparente, compréhensible et aisément accessible, par écrit, voie électronique ou tout autre moyen (article 12 du RGPD).

Le contenu de cette information est notamment précisé aux articles 13 et 14 du RGPD.

L'établissement a une double obligation d'information concernant les EDS vis-à-vis du patient : lors de l'inclusion de ses données et lors de leurs réutilisations.

#### **2. Information du patient lors de l'inclusion de ses données dans l'entrepôt**

Le régime de non-opposition du patient est celui qui s'applique pour l'inclusion de ses données dans l'EDS. L'information du patient est néanmoins obligatoire. Deux cas sont à distinguer : l'inclusion des données des patients pris en charge avant la date de mise en place de l'EDS et celle de ceux pris en charge à partir de la date de mise en place de EDS.

#### **Prise en charge avant date mise en place EDS**

Une information individuelle concernant l'inclusion dans l'EDS est obligatoire pour tous les patients pris en charge, y compris ceux pris en charge avant la mise en place de l'EDS.

Cependant, l'article 14.5.b du RGPD, précise que « L'obligation d'information individuelle de la personne concernée peut faire l'objet d'exceptions dans l'hypothèse où la fourniture d'une telle information se révélerait impossible, exigerait des efforts disproportionnés ou compromettrait gravement la réalisation des objectifs du traitement ». Une information collective peut être alors menée, par tous les moyens possibles : information sur page internet, campagne de presse, campagne sur les réseaux sociaux, envoi de courrier électronique [17]... Cette manière de procéder a été validée pour le projet INCLUDE du CHU de Lille (délibération n°2019-103 du 5 septembre 2019) et également précédemment pour le CHU de Nantes (Délibération n°2018-295 du 19 juillet 2018).

### **Prise en charge après la date de mise en place de l'EDS**

Pour les patients pris en charge après la date de mise en place de l'EDS, une information individuelle doit être effectuée. La CNIL indique que cette information doit être « individuelle, complète, claire et lisible, spécifiquement relative à la constitution de l'entrepôt. ». La CNIL indique également que le support est libre : oral, écrit ou tout autre moyen [17]. Néanmoins, la création d'une notice papier dédiée permet de s'assurer que tous les patients auront bénéficié de l'information. Elle peut en effet :

- être intégrée au livret d'accueil pour les patients hospitalisés.
- être remise en main propre pour les patients venant en consultation.
- être envoyée par courrier pour les autres particuliers, par exemple les patients admis aux urgences avant d'être hospitalisés.

### 3. Information sur les traitements

S'agissant des traitements de données dont la dérogation à l'interdiction de traiter les données dites sensibles au sens RGPD relève de la recherche scientifique, la CNIL indique que « Conformément à la loi, les personnes doivent être informées individuellement de chaque projet de recherche mené sur la base de l'entrepôt »[10]. Cette information peut par exemple

être réalisée par envoi de courrier postal ou électronique, ou sur un espace numérique individuel sécurisé de l'établissement. Si des données issues du système national des données de santé sont traitées, le patient doit explicitement en être informé.

S'agissant des traitements pour lesquels l'obligation d'informer la personne n'est pas individuelle, la CNIL indique [17] :

- que cette information peut être collective.
- mais qu'une information individuelle est tout de même à privilégier.
- et que sauf exception, une seule information est nécessaire.

Cette information peut être réalisée par exemple en même temps que celle concernant l'inclusion dans l'EDS.

Par ailleurs, la législation tient compte d'un certain nombre de cas particuliers.

L'article 48 de la LIL impose au responsable de traitement de prévoir une information adaptée pour les mineurs de moins de quinze ans.

L'article 70 de la LIL précise que l'information concerne également :

- les titulaires de l'exercice de l'autorité parentale pour les mineurs, sauf exception.
- la personne chargée d'une mission de représentation dans le cadre d'une tutelle, d'une habilitation familiale ou d'un mandat de protection future, pour les majeurs protégés.

Enfin en ce qui concerne les personnes décédées, le RGPD ne s'applique plus lors du traitement de leurs données (cf considérant 27 du RGPD). Pour le cadre réglementaire français, la situation est plus ambiguë : la Commission d'Accès aux Documents Administratifs, dans son avis 20182992 du 17 mai 2019, précise que « selon sa doctrine et celle de la CNIL, les données à caractère personnel ne concernent que les personnes vivantes », sauf dans le cas où « certaines informations relatives aux personnes décédées

pourraient comporter des conséquences sur la vie privée de personnes vivantes ». Néanmoins, le chapitre V du titre II de la LIL se nomme « dispositions régissant les traitements de données à caractère personnel relatives aux personnes décédées », ce qui implique que ces données sont toujours considérées comme à caractère personnel. Dans tous les cas, dans le cadre des EDS hospitaliers, les données des personnes restent soumises au secret professionnel même après leur décès. L'article 84 de la LIL indique que certains droits de la personne s'éteignent au moment de son décès, sauf directives écrites lors de son vivant (article 85 de la LIL). Dans le cadre de la recherche, l'article 86 que « Les informations concernant les personnes décédées, y compris celles qui figurent sur les certificats des causes de décès, peuvent faire l'objet d'un traitement à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé, sauf si l'intéressé a, de son vivant, exprimé son refus par écrit. ».

En pratique pour les EDS, cela signifie que :

- l'établissement doit mettre en place une information aux patients concernant les directives sur le traitement de leurs données après leur éventuel décès.
- les ayants droit d'une personne décédée ne peuvent pas s'opposer à ce que ses données soient réutilisées dans le cadre de la recherche, hors directives manuscrites rédigées de son vivant.

#### ii) Droit d'accès, de rectification à la portabilité

Le droit d'accès consiste pour le patient à obtenir la communication dans un format compréhensible des données le concernant faisant l'objet d'un traitement (article 49 de la LIL). Le cas échéant, le patient peut demander la rectification de données inexactes le concernant (article 50 de la LIL).

Le droit à la portabilité est une extension du droit d'accès. Il consiste au droit de récupérer une partie de ses données dans un format ouvert et lisible par une machine, pour un usage

personnel ou un transfert à un organisme tiers (article 55 de la LIL). Il ne s'applique que pour les activités de traitement dont la base légale est soit contractuelle soit le consentement (article 20 du RGPD).

### iii) Droit à l'effacement/oubli

Le droit à l'effacement permet à la personne, dans certaines conditions, d'exiger auprès du responsable de traitement l'effacement de ses données (article 51 de la LIL). Ce droit peut s'appliquer si la personne concernée retire le consentement sur lequel est fondé le traitement. Ce droit ne s'applique jamais si la base légale du traitement est (article 17 RGPD) le respect d'une obligation légale ou l'intérêt public dans le domaine de la santé publique. Si une des activités de traitements de l'EDS a une de ces bases légales, ce droit ne s'applique donc pas.

### iv) Droit d'opposition

Le droit d'opposition permet au patient de s'opposer à tout moment à ce que ses données fassent l'objet d'un traitement (article 56 de la LIL). L'article 21 du RGPD indique que ce droit s'applique si la base légale de traitement est soit l'exécution d'une mission d'intérêt public, soit l'intérêt légitime du responsable de traitement. L'article 56 de la LIL précise que ce droit ne s'applique pas aux traitements dont la base juridique est une obligation légale ou lorsque l'application de ces dispositions a été écartée par une disposition expresse de l'acte instaurant le traitement. Par exemple, dans le cadre des EDS hospitaliers, le patient peut donc s'opposer à la réutilisation de ses données si la finalité de l'activité de traitement est la recherche mais pas si la finalité de l'activité de traitement concerne le PMSI. La demande se fait auprès du délégué à la protection des données, dont les coordonnées font partie des informations à communiquer obligatoirement.

Le droit d'opposition est distinct du droit à l'effacement. Une personne peut faire valoir son droit d'opposition pour une finalité sans faire valoir son droit à l'effacement : les données seront toujours présentes mais ne seront pas exploitées.

#### v) Droit à la limitation de traitement

Le droit à la limitation de traitement complète les droits précédemment cités et consiste en la possibilité de geler l'utilisation des données par le responsable de traitement (article 53 de la LIL). Lorsque le patient constate qu'une des situations listées à l'alinéa 1 de l'article 18 du RGPD est effective, le droit à la limitation de traitement s'applique. Pendant ce délai, le patient a la possibilité de demander à l'établissement de geler l'utilisation de ses données. Inversement, le patient peut demander directement la limitation de traitement de certaines données dans le cas où l'établissement souhaite lui-même les effacer.

#### vi) Obligation de notification en cas de rectification ou effacement des données

L'article 19 du RGPD, implémenté dans la LIL sous l'article 54, précise que « Le responsable du traitement notifie à chaque destinataire auquel les données à caractère personnel ont été communiquées toute rectification ou tout effacement de données à caractère personnel ou toute limitation du traitement effectué [...], à moins qu'une telle communication se révèle impossible ou exige des efforts disproportionnés. Le responsable du traitement fournit à la personne concernée des informations sur ces destinataires si celle-ci en fait la demande».

Une matrice résumant les droits des patients en fonction de la base légale de traitement dans le cadre des EDS est disponible au tableau 2.

Tableau 2 : Matrice de l'applicabilité des droits conférés par le RGPD et la loi informatique et libertés aux personnes dont les données sont traitées, selon la base légale de traitement, dans le cadre des entrepôts de données de santé hospitaliers en France en novembre 2020.

Base légale du traitement effectué dans le cadre des entrepôts de données de santé hospitaliers				
Droit du patient	Obligation légale du responsable de traitement	Mission d'intérêt public	Consentement au traitement des données pour une ou plusieurs finalités spécifiques	Intérêt légitime du responsable de traitement
Droit à l'information	Applicable	Applicable	Applicable	Applicable
Droit à d'accès/de rectification	Applicable	Applicable	Applicable	Applicable
Droit à la portabilité	Non applicable	Non applicable	Applicable	Non applicable
Droit à l'effacement/l'oubli <sup>a</sup>	Non applicable	Non applicable si recherche scientifique	Applicable si retrait du consentement	Applicable
Droit à l'opposition	Non applicable	Applicable	Applicable si retrait du consentement	Applicable
Droit à la limitation de traitement	Applicable si une des conditions listées à l'alinéa 1 de l'article 18 du RGPD est remplie	Applicable si une des conditions listées à l'alinéa 1 de l'article 18 du RGPD est remplie	Applicable si une des conditions listées à l'alinéa 1 de l'article 18 du RGPD est remplie	Applicable si une des conditions listées à l'alinéa 1 de l'article 18 du RGPD est remplie
Obligation de notification en cas de rectification ou d'effacement des données	Applicable si application du droit à la rectification ou à la limitation de traitement.	Applicable si application u droit à la rectification, à l'effacement ou à la limitation de traitement.	Applicable si application du droit à la rectification ou à la limitation de traitement.	Applicable si application du droit à la rectification ou à la limitation de traitement.

a Si un même jeu de données à caractère personnel est simultanément concerné par une activité de traitement empêchant son effacement et par une activité de traitement permettant son effacement, la première activité de traitement prime et il ne sera pas effacé

### **g) Droits des soignants**

L'entrepôt contient également des données à caractère personnel concernant les professionnels de santé. Les droits décrits ci-dessus s'appliquent donc également à eux. Il faudra ainsi prévoir un deuxième circuit d'information et de recueil d'opposition.

### **h) Droits des utilisateurs de l'entrepôt**

Enfin, l'entrepôt contient des données à caractère personnel sur les utilisateurs, telles que leurs noms. Il convient de les informer et de recueillir leur consentement pour la gestion de l'entrepôt.

### **i) Protection des données**

#### **i) Durée de conservation**

L'article 1112-7 du CSP fixe la durée de conservation du dossier médical du patient au sein des établissements de santé. C'est cet article que la CNIL a pris comme référence dans son autorisation pour l'EDS « INCLUDE » du CHU de Lille en 2019, pour la durée de conservation des données. Concernant l'autorisation donnée à l'Assistance publique des Hôpitaux de Paris en 2017, toujours pour le même type d'EDS, la CNIL indique également que les données seront conservées dans l'EDS pendant la durée prévue par les dispositions légales et réglementaires applicables en matière de conservation des dossiers médicaux (délibération n°2017-013 du 19 janvier 2017).

#### **ii) Principe de minimisation des données**

Un autre élément participant à la protection des données est la minimisation des données. Cette pratique consiste à ne recueillir que les informations nécessaires pour un traitement, afin de réduire les risques pour la sécurité et la confidentialité en cas d'accès frauduleux ou accidentel à la base de données. La minimisation est simple à mettre en œuvre dans une étude

scientifique standard, où toutes les variables à recueillir sont décidées en amont, après revue de la littérature. Cette pratique est plus délicate à mettre en œuvre dans le cadre des EDS de données de santé, pour lesquels il n'est pas toujours possible au moment de leur mise en place de connaître les données qui seront utiles à l'avenir. Une technique qu'il est possible d'appliquer est la pseudonymisation, distincte de l'anonymisation.

### iii) Pseudonymisation et anonymisation

La pseudonymisation et l'anonymisation sont deux pratiques qui consistent à dégrader la granularité de l'information d'un jeu de données dans une optique de minimisation des données. Dans le cas d'une anonymisation, toute ré-identification d'une personne, par quelque moyen que ce soit, est impossible. Dans le cas d'une pseudonymisation, la ré-identification est possible en croisant la base avec des données tierces. C'est pourquoi les données pseudonymisées demeurent des données à caractère personnelles. [18]

Du point de vue pratique, travailler avec des données nominatives ou le NIR, permet de croiser des données de sources les plus diverses ; l'inconvénient étant un risque plus élevé de rupture de confidentialité. Afin de garantir au maximum la confidentialité, l'idéal est de ne travailler qu'avec des données anonymisées, celles-ci n'étant alors plus à caractère personnel et donc non soumises au RGPD ; l'inconvénient étant alors une granularité souvent trop grossière pour la recherche. La pseudonymisation peut donc apparaître comme un compromis raisonnable entre ces deux options.

La CNIL précise qu'«en pratique la pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, numéro de séjour, numéro de patient.) d'un jeu de données par des données indirectement identifiantes (alias, numéro dans un classement, etc.). »[18]. D'autres mesures peuvent être prises, comme le remplacement d'une adresse postale par un code postal/Insee, le remplacement de la date de naissance par l'année de naissance ou la mise en classe de variables quantitatives telle que l'âge par exemple.

Dans la délibération 2018-289 du 12 septembre 2018 de la CNIL portant sur l'EDS de données d'IQVIA, « La Commission rappelle toutefois que la pseudonymisation ne constitue qu'une mesure de sécurité visant à réduire les risques pour les personnes concernées, et n'exonère en rien le responsable de traitement de ses obligations. ».

De plus, la pseudonymisation présente des contraintes :

- il n'existe à ce jour aucune technique fiable à 100% pour dé-identifier un document/un compte-rendu contenant des données directement ou indirectement identifiantes.
- la pseudonymisation implique une plus grande complexité technique. Par exemple, il faut prévoir une modalité de pseudonymisation par type de fichier, en amont de son intégration dans l'EDS.
- la dégradation de la granularité des données nécessaires pour la pseudonymisation est telle que, même si elle n'est pas aussi importante que pour l'anonymisation, elle réduit le potentiel d'utilisation de l'EDS.

#### iv) Accès différencié aux données

Une autre méthode de protection des données est l'accès différencié aux données selon le profil de l'utilisateur. Dans l'article « Strategies for maintaining patient privacy in i2b2 » par Murphy et al [19], les auteurs proposent de moduler la granularité de l'information accessible en fonction du niveau de confidentialité de l'utilisateur. En pratique :

- les utilisateurs de niveau 1 n'ont accès qu'à des dénombrements obfusqués de patients. L'obfuscation est un procédé qui consiste à ajouter ou soustraire une certaine valeur delta à un résultat de dénombrement. Cela permet pour le destinataire de disposer d'une estimation suffisamment précise pour être utile mais pas suffisamment précise pour ré-identifier aisément un patient par recoupement de plusieurs requêtes.
- les utilisateurs de niveau 2 n'ont accès qu'à des données agrégées non obfusquées.

- les utilisateurs de niveau 3 n'ont accès qu'à certaines données individuelles déidentifiées.
- les utilisateurs de niveau 4 ont accès à toutes les données individuelles dé-identifiées.
- les utilisateurs de niveau 5 ont accès à toutes les données individuelles nominatives.

Cette organisation permet de trouver un compromis entre la sécurité et la confidentialité des données d'une part et les besoins occasionnels d'une granularité fine de données d'autre part.

#### v) Sécurité informatique

Les mesures organisationnelles ne suffisent pas à assurer la sécurité et confidentialité des données. Une politique de sécurité informatique, qui comporte des mesures de protection physique des locaux ainsi que des mesures logicielles, doit être menée. Dans ce sens, le décret n° 2018-384 du 23 mai 2018 établit les établissements de santé comme des opérateurs de services essentiels (OSE), c'est-à-dire des opérateurs tributaires des réseaux ou des systèmes d'information et qui fournissent un service essentiel dont l'interruption aurait un impact significatif sur le fonctionnement de l'économie ou la société. Parmi les obligations incombant aux OSE, celles d'appliquer une série d'obligation par décret (arrêté du 14 septembre 2018), afin de garantir un niveau de sécurité informatique élevé. Les EDS hospitaliers sont donc déployés dans des milieux de haute sécurité.

## 5) Conclusion

Pour mettre en place un EDS dans un établissement de santé, la réglementation actuelle impose la réalisation d'une demande d'autorisation préalable à la CNIL. Du fait du risque élevé pour les droits et libertés des personnes, une AIPD devra être incluse à cette demande. Plusieurs principes permettent de limiter les risques pour la sécurité et la confidentialité des données contenues dans les EDS :

- la participation du DIM à la gouvernance, celui-ci étant l'organe historique de gestion et de protection des données à caractère personnel au sein des établissements de santé. Avec à minima, une autorité fonctionnelle médicale sur le personnel requêtant l'EDS.
- la mise en place d'un CSE, vérifiant et validant le caractère éthique et la pertinence scientifique des projets de réutilisation des données de l'EDS.
- la mise en place d'un comité de pilotage transversal, impliquant les instances médicales, scientifiques et administratives, pour les orientations stratégiques.
- le respect des droits des patients, dont l'applicabilité varie en fonction de la base légale de traitement.
- la minimisation des données, bien que ce concept soit partiellement contradictoire avec le principe de l'EDS.
- la pseudonymisation des données.
- la mise en place de limites à la durée de conservation des données.
- la mise en place d'accès différenciés aux données en fonction du profil de l'utilisateur.
- la mise en place d'audits de traçabilité d'accès aux données.
- le respect des règles de sécurité imposées aux OSE.

## Références Bibliographiques

[1] Madec J, Bouzillé G, Riou C, Van Hille P, Merour C, Artigny M-L, et al. eHOP Clinical Data Warehouse: From a Prototype to the Creation of an Inter-Regional Clinical Data Centers Network. *Stud Health Technol Inform* 2019;264:1536-7. doi:10.3233/SHTI190522.

[2] Jannot A-S, Zapletal E, Avillach P, Mamzer M-F, Burgun A, Degoulet P. The Georges Pompidou University Hospital Clinical Data Warehouse: A 8-years follow-up experience. *International Journal of Medical Informatics* 2017. doi:10.1016/j.ijmedinf.2017.02.006.

[3] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) | Journal Officiel de l'Union Européenne, 2016, L119 [https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv:OJ.L\\_.2016.119.01.0001.01.FRA&toc=OJ:L:2016:119:FULL](https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.FRA&toc=OJ:L:2016:119:FULL) (accessed November 15, 2020).

[4] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Version Juin 2020 | Legifrance <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000886460> (accessed November 15, 2020).

[5] Benyahia N. Établissements de santé et start-up face à la régulation des innovations par la protection des données de santé. *Journal du Droit de la Santé et de l'Assurance - Maladie (JDSAM)* 2018;N° 20:21-4.

[6] Un hôpital allemand condamné à 105 000 euros d’amende pour non-respect du RGPD | Accompagnement Cybersécurité des Structures de Santé n.d. <https://cyberveille-sante.gouv.fr/cyberveille-sante/1535-un-hopital-allemand-condamne-105-000-euros-damende-pour-non-respect-du-rgpd> (accessed November 15, 2020).

[7] Mouron P. Pour ou contre la patrimonialité des données personnelles. *Revue européenne des médias et du numérique* 2018:90-106.

[8] Commission Nationale Informatique et Libertés. Analyse d’impact relative à la protection des données. Les bases de connaissances. [en ligne]. <https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-fr-basesdeconnaissances.pdf>. Accessed November 15, 2020.

[9] Qu’est-ce qu’une donnée de santé ? | CNIL n.d. <https://www.cnil.fr/fr/quest-ce-que-une-donnee-de-sante> (accessed November 15, 2020).

[10] Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences ? | CNIL n.d. <https://www.cnil.fr/fr/traitements-de-donnees-de-sante-comment-faire-la-distinction-entre-un-entrepot-et-une-recherche-et> (accessed November 15, 2020).

[11] Comité d’évaluation de la réforme de la gouvernance des établissements publics de santé. Rapport au parlement. [En ligne] [https://www.senat.fr/fileadmin/Fichiers/Images/evenement/loi\\_hpst/rapport\\_evaluation.pdf](https://www.senat.fr/fileadmin/Fichiers/Images/evenement/loi_hpst/rapport_evaluation.pdf) Accessed November 15, 2020.

[12] Définir une finalité | CNIL n.d. <https://www.cnil.fr/fr/definir-une-finalite> (accessed November 15, 2020).

[13] La finalité d'un fichier, c'est quoi ? | Besoin d'aide | CNIL n.d. <https://www.cnil.fr/fr/cnil-direct/question/la-finalite-dun-fichier-cest-quoi> (accessed November 15, 2020).

[14] Comment choisir la base légale d'un traitement ? Cas pratiques avec certains traitements mis en œuvre par la CNIL | CNIL n.d. <https://www.cnil.fr/fr/comment-choisir-la-base-legale-dun-traitement-cas-pratiques-avec-certains-traitements-mis-en-oeuvre> (accessed November 15, 2020).

[15] article 5 de Recommandation CM/Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé | Comité des ministres [https://www.apda.ad/sites/default/files/2019-03/CM\\_Rec%282019%292F\\_FR.pdf](https://www.apda.ad/sites/default/files/2019-03/CM_Rec%282019%292F_FR.pdf) (accessed November 15, 2020).

[16] La licéité du traitement : l'essentiel sur les bases légales prévues par le RGPD | CNIL n.d. <https://www.cnil.fr/fr/la-liceite-du-traitement-lessentiel-sur-les-bases-legales-prevues-par-le-rgpd> (accessed November 15, 2020).

[17] Traitement de données de santé : comment informer les personnes concernées ? | CNIL n.d. <https://www.cnil.fr/fr/traitement-de-donnees-de-sante-comment-informer-les-personnes-concernees> (accessed November 15, 2020).

[18] Identifier les données personnelles | CNIL n.d. <https://www.cnil.fr/fr/identifier-les-donnees-personnelles> (accessed November 15, 2020).

[19] Murphy SN, Gainer V, Mendis M, Churchill S, Kohane I. Strategies for maintaining patient privacy in i2b2. *J Am Med Inform Assoc* 2011;18:i103-108. doi:10.1136/amiajnl-2011-000316.

**Remerciements :**

Pr Nicolas JAY

Dr Frédérique Claudot

**Financements :**

Aucun financement externe n'a été apporté.

## VI. Conclusions / Perspectives

Nous allons conclure cette thèse, toujours d'un point de vue scientifique, technique et réglementaire.

D'un point de vue scientifique, la réutilisation des données générées par le SIH d'un établissement de santé ne peut se faire de manière efficiente que par la mise en place d'un outil dédié, l'EDS, distinct du système de production. La réutilisation de ces données doit se faire avec prudence, en ayant conscience des limites qu'elles présentent :

- Il s'agit certes de données générées en vie réelle ; elles sont néanmoins la plupart du temps de moins bonne qualité que des données saisies dans le cadre d'une étude prospective.
- Une partie des données est codée par des nomenclatures, créées à des fins médico-économiques (CCAM, NABM...). Il n'existe pas forcément de correspondance parfaite entre ces SOC et des concepts médicaux donnés.

D'un point de vue technique, la bonne intégration au sein du SIH des composants du DPI sont un prérequis à toute installation d'un EDS. L'utilisation d'un EAI en amont permet de simplifier ce processus. Le choix d'une solution technique pour l'EDS parmi toutes les offres disponibles sur le marché et en libre doit se faire en prenant en compte de multiples critères : capacité à gérer les données non structurées, plan d'action de mises à jour, coûts d'exploitation, interopérabilité avec les réseaux d'EDS déjà en place...

D'un point de vue juridique, dans l'attente de la publication d'une MR spécifique aux EDS, la réglementation actuelle impose aux établissements de santé la réalisation d'une demande d'autorisation à la CNIL préalablement à toute mise en place d'un EDS. Cette demande devra détailler l'intégralité des activités de traitements, avec notamment leur finalité, leur base légale, leur dérogation à l'interdiction de traiter les données dites/perçues sensibles et les catégories de données traitées. Du fait du risque élevé pour les droits et libertés des personnes, une AIPD devra être incluse à cette demande. Plusieurs principes permettent de limiter les risques pour la sécurité et la confidentialité des données contenues dans les EDS :

- la participation des DIM, organe historique de gestion et de protection des données à caractère personnel au sein des établissements de santé.
- la minimisation des données, bien que ce concept soit partiellement contradictoire avec le principe de l'EDS.
- lapseudonymisation des données.
- la mise en place de limites de durée de conservation des données.
- la mise en place d'accès différenciés aux données en fonction du niveau de confidentialité de l'utilisateur.
- la mise en place d'audits de traçabilité d'accès aux données.
- le respect des règles de sécurité imposées aux OSE dont les établissements de santé font partie.

Par ailleurs, le responsable de traitement est tenu de garantir des droits aux personnes, particulièrement le droit à l'information et le droit d'opposition.

## VII. Références bibliographiques

Ces références concernent les parties I à IV de cette thèse.

Les références de l'article se trouvent à la suite de son manuscrit.

[1]

[1] Venot A, Burgun A, Quantin C. Informatique médicale, e-Santé: Fondements et applications. 2013. <https://doi.org/10.1007/978-2-8178-0338-8>.

[2] Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc 2019;26:1297-1304. <https://doi.org/10.1093/jamia/ocz096>.

[3] Cuggia M, Garcelon N, Campillo-Gimenez B, Bernicot T, Laurent J-F, Garin E, et al. Roogle: an information retrieval engine for clinical data warehouse. Stud Health Technol Inform 2011;169:584-588.

[4] Stock WG. Concepts and semantic relations in information science. Journal of the American Society for Information Science and Technology 2010;61:1951-1969. <https://doi.org/10.1002/asi.21382>.

[5] Etat des lieux : Niveau de preuves et gradation des recommandations de bonnes pratiques. | Haute Autorité de Santé. 2013. [https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat\\_des\\_lieux\\_niveau\\_preuve\\_gradation.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat_des_lieux_niveau_preuve_gradation.pdf) (accessed June 15, 2020).

[6] Degoulet P, Fieschi M (1998) Informatique médicale, 3<sup>ème</sup> éd. Paris, Masson

[7] BAUMANN A. Définition de Intégration - lexique de droit de l'informatique. BAUMANN Avocats n.d. <https://www.baumann-avocats.com/definition/integration.php> (accessed July 28, 2020).

[8] Losavio F, Ortega D, Pérez M. Modeling EAI [Enterprise Application Integration], 2002, p. 195-203. <https://doi.org/10.1109/SCCC.2002.1173194>.

[9] Codes mouvements | Publication ATIH n.d. <https://www.atih.sante.fr/codes-mouvements> (accessed August 1, 2020).

[10] Inmon WH, (2002) Building the Data Warehouse, 3<sup>ème</sup> éd. Wiley Computer Publishing

[11] Breslin M. Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. Business Intelligence Journal 2004;9.

[12] Webinar QuantIM - Marc Cuggia | Sesstim n.d. <https://sesstim.univ-amu.fr/video-box/webinar-quantim-marc-cuggia> (accessed July 28, 2020).

## **Autorisation d'imprimer**

NANCY, le **14 août 2020**

Le Président de Thèse

**Professeur Nicolas JAY**

NANCY, le **26 août 2020**

Le Doyen de la Faculté de Médecine

**Professeur Marc BRAUN**

AUTORISE À SOUTENIR ET À IMPRIMER LA THÈSE/ **11284C**

NANCY, le **1<sup>er</sup> septembre 2020**

LE PRÉSIDENT DE L'UNIVERSITÉ DE LORRAINE,

## **RÉSUMÉ DE LA THÈSE**

Les entrepôts de données, permettant une réutilisation plus aisée des données produites lors du soin, sont déployés dans les hôpitaux français depuis la fin des années 2000. L'objectif de cette thèse est de décrire les aspects scientifiques, techniques et réglementaires de la mise en place et de l'exploitation d'un Entrepôt de données de santé (EDS) au sein d'un hôpital français en s'appuyant sur l'exemple du Centre Hospitalier Régional de Metz-Thionville.

D'un point de vue scientifique, la réutilisation des données est notamment limitée par l'imperfection de la correspondance entre les nomenclatures employées et les concepts médicaux. D'un point de vue technique, la bonne intégration au sein du système d'information hospitalier des composants du dossier patient informatisé sont un prérequis à toute installation d'un EDS. D'un point de vue réglementaire, des données à caractères personnelles étant traitées, le Règlement Européen sur la Protection des Données, implémenté dans la Loi informatique et libertés française (LIL) s'applique. Une analyse d'impact relative à la protection des données doit être incluse à la demande d'autorisation préalable adressée à la Commission Nationale Informatique et Libertés. Son contenu est détaillé aux articles 13 et 14 du la LIL.

Conclusion : l'exploitation d'un EDS hospitalier est d'autant plus pertinente que les sous-jacents scientifiques et techniques sont maîtrisés ; dans l'attente de la publication d'une méthodologie de référence spécifique aux EDS, ce travail synthétise la réglementation à laquelle ils sont soumis.

---

## **TITRE EN ANGLAIS**

Deployment of a clinical data warehouse: example of the French Metz-Thionville Regional Hospital. Scientific, technic and legal dimensions.

---

## **THÈSE : MÉDECINE SPÉCIALITÉ ANNÉE 2020**

---

### **MOTS CLÉS :**

Système d'Organisation des Connaissances, Nomenclature, Système d'Information Hospitalier, Entrepôt de Données de Santé hospitalier, ehop, Règlement Général sur la Protection des Données, Loi Informatique et Libertés, Gouvernance, Confidentialité et Sécurité des données, Traitement de Données de Santé

---

### **INTITULÉ ET ADRESSE :**

**UNIVERSITÉ DE LORRAINE**

**Faculté de Médecine de Nancy**

9, avenue de la Forêt de Haye

54505 VANDŒUVRE-LÈS-NANCY Cedex