



Actes de la journée du 20 novembre 2018

# Cahier de l'atelier 5

## Les données collectées : le potentiel de l'accès sécurisé



Co-animé par les établissements de santé

### INTERVENANTS

ATIH  
Damien MURE  
Brice SAUVAJON  
Sandra STEUNOU

Hospices civils de Lyon  
Antoine DUCLOS  
Cécile PAYET  
Stéphanie POLAZZI

CHU Montpellier  
Dr Michel LEHMANN

I. Présentation et contexte	2
II. Une vision d'établissement	6
III. Utilité pour la recherche	8



### Damien MURE

Bonjour à tous, je suis ingénieur en informatique et adjoint à la responsable de sécurité à l'ATIH. Bienvenue à ces journées des données hospitalières. L'atelier d'aujourd'hui va se dérouler en trois temps. La première partie sera animée par Sandra Steunou, de l'ATIH, qui vous présentera l'accès sécurisé et le contexte de la création de cette plateforme. Nous aurons ensuite deux interventions de personnes extérieures à l'ATIH. Le premier intervenant sera Michel Lehmann, du CHU de Montpellier, qui livrera une vision d'établissement, pour savoir ce que les DIM

peuvent faire pour leur établissement avec la plateforme. La dernière présentation sera portée par trois représentants des HCL. Ils reviendront sur l'utilité de la plateforme pour la recherche.

Nous essayerons d'entrecouper ces présentations de moments d'échanges avec la salle.

*Un diaporama est projeté.*

## I. Présentation et contexte

### Sandra STEUNOU

Bonjour à tous. Je suis statisticienne à l'ATIH. Je vais vous parler des raisons pour lesquelles l'ATIH a mis en place cette plateforme d'accès sécurisé aux données, mais également de la méthode d'accès et des données disponibles.

La loi de modernisation de notre système de santé, et notamment son article 193, a nécessité de renforcer

la sécurité de la diffusion des données du PMSI, qui sont considérées comme des données sensibles. Les CD qui étaient auparavant distribués pour tous les ordinateurs font désormais partie du passé.

L'ATIH a mis en place un serveur sécurisé aux données hospitalières, dont l'adresse est la suivante: <http://acces-securise.atih.sante.fr>.

### 1. Méthodes d'accès

Cet accès sécurisé est centralisé et géré par l'ATIH. Il est accessible via un système d'authentification forte. Tous les traitements réalisés sur la plateforme, tout comme les accès, sont tracés. La plateforme est accessible 24 heures sur 24 et 5 à 6 jours sur 7.

L'accès à la plateforme est gratuit et simple, via un navigateur web. La plateforme dispose du logiciel SAS Enterprise Guide, qui permet d'avoir une multitude de possibilités. Vous avez également la possibilité de demander R Studio Serveur sur la plateforme.

La plateforme dispose également d'un explorateur de fichiers, du Pack Office avec Excel, Word et Access, mais également de douze tutoriels qui permettent de découvrir les grandes notions liées à l'utilisation des données.

Un support dédié à l'accès sécurisé a également été mis en place. Il est joignable à l'adresse suivante: [support@atih.sante.fr](mailto:support@atih.sante.fr).

### 2. Méthodologie de référence

La CNIL a publié en juillet dernier une nouvelle méthodologie de référence, la MR-005, spécifique pour les établissements de santé et les fédérations. Elle permet un accès simplifié et vous donne la possibilité d'accéder, sur la plateforme, aux quatre champs du PMSI (médecine, chirurgie et obstétrique, soins de suite et réadaptation, hospitalisation à domicile et psychiatrie), en base nationale et avec chaînage sur neuf années d'historique en plus de l'année en cours. Vous avez également accès aux résumés de passage aux urgences (RPU).

La MR— 005 couvre de nombreuses finalités, notamment la valorisation de l'offre de soin, la planification, les études épidémiologiques, médico-économiques, etc. Si vous entrez dans les finalités décrites dans la méthodologie de référence, une simple déclaration de conformité sur le site de la CNIL suffit pour accéder au serveur. En revanche, l'appariement de données entre le PMSI et des registres ou des cohortes doit faire l'objet d'un processus classique en passant par l'INDS, le CEREES et la CNIL.

### 3. Données disponibles

Au-delà des données d'activités (PMSI, RPU) des référentiels communs à tous les utilisateurs de la plateforme sont accessibles: des référentiels généraux, mais également des référentiels plus spécifiques pour le PMSI avec l'historique de la CCAM, de la CIM-10, du CSARR, etc.

La SAE est également disponible depuis 2008 jusqu'à 2017, ainsi qu'un espace de documentation, comprenant notamment tous les dictionnaires de variables et des tables qui composent le serveur.

Les avantages de ce dispositif sont les suivants :

- simplicité d'accès via un navigateur web ;
- structuration et indexation des tables, ce qui permet une optimisation des temps de traitement sur un serveur assez rapide ;

- mise à jour hebdomadaire des données ;
- données retravaillées, avec par exemple des regroupages de GHM en version constante sur plusieurs années, des bases de valorisation de l'activité, les bases de l'algorithme Inca, etc. ;
- possibilité d'imports de données sans restriction (projets, programmes, base de données, nomenclatures, etc.) et d'exports (chaque utilisateur de la plateforme s'engageant à ne pas restituer de données brutes, les données qui sortent du serveur doivent en effet être présentées sous forme de statistiques agrégées) ;
- partage de programmes, avec des noms de variables communs qui permettent une uniformisation sur plusieurs années.

### 4. Perspectives

Dans les évolutions à venir, nous prévoyons de mettre en place un espace documentaire, sous forme de site Web, pour remplacer la partie documentation qui existe aujourd'hui en fichier statique Word ou Excel. Il rassemblerait toute la documentation utile à l'utilisation des bases de données disponibles sur le serveur. Cet espace permettrait également une réelle fluidité dans l'accès aux actualités et aux dernières mises en ligne.

Nous prévoyons également d'y inclure un forum d'échanges pour que vous puissiez, entre établissements, dialoguer sur cet espace autour de bonnes

pratiques, de méthodes de travail, etc.

Nous souhaitons enfin simplifier la démarche pour accéder au serveur en utilisant une procédure dématérialisée avec la signature électronique. Ce dispositif est en cours de test. Actuellement, la procédure est intégralement réalisée par courrier, ce qui demande des délais importants. Nous gagnerions vraiment du temps en passant par la signature électronique.

FIGURE 1

FIGURE 1

**Journée ATIH**  
des données hospitalières

**Projet d'amélioration en cours**

- Espace documentaire :
  - Documentation utile à l'exploitation des bases
  - Actualités et dernières mises en ligne
  - Echange
- Simplification administrative :
  - Signature électronique

Atelier 5

7

Aujourd'hui, le serveur compte environ 600 utilisateurs, dont plus de 300 en établissement de santé, dans 130 établissements.

Je vous remercie pour votre attention. Je vous propose un temps d'échanges, M. Brice Sauvajon étant également présent pour répondre aux questions plus techniques ou informatiques.

Lors des inscriptions, nous avons noté que nombre d'entre vous n'ont pas encore accès à la plateforme.

#### De la salle

Pourriez-vous résumer la demande sur le MR-005 en quelques mots ?

#### Sandra STEUNOU

Vous pouvez trouver la méthodologie de référence sur le site de la CNIL, qui décrit toutes les finalités. La démarche nécessaire pour pouvoir accéder au serveur est de se déclarer conforme à cette MR— 005, ce qui vous engage à avoir, dans votre établissement, un délégué à la protection des données qui tient à jour, au sein du registre des activités de traitement, la liste des études mises en œuvre dans le cadre de cette méthodologie.

#### De la salle

Le médecin DIM doit-il faire la déclaration ? Une signature du chef d'établissement est-elle nécessaire ?

#### Sandra STEUNOU

Normalement, l'engagement de respecter la méthodologie doit être pris par le responsable juridique.

#### De la salle

Quelle est la démarche exacte à suivre ?

#### Sandra STEUNOU

Il suffit de se rendre sur le site de la CNIL et de cliquer sur «Se déclarer conforme à la méthodologie de référence». La MR— 005 est vraiment spécifique aux établissements de santé pour l'accès au serveur de l'ATIH. Vous recevez ensuite un mail comprenant votre déclaration de conformité, que vous nous envoyez. Toute la procédure est ensuite décrite sur le site de l'Agence. Nous fonctionnons aujourd'hui sur un système de convention à signer en version papier.

Une fois que tous les documents sont validés, nous vous envoyons les dispositifs par courrier. Lorsque nous recevons votre accusé de réception, nous vous transmettons le mot de passe pour vous connecter à la plateforme. Le mot de passe doit bien évidemment être modifié dès la première utilisation.

#### De la salle

La déclaration doit-elle être réalisée individuellement ou par établissement ?

#### Sandra STEUNOU

La MR— 005 est une déclaration pour l'établissement. Ensuite, vous avez une convention qui engage le responsable juridique de votre structure. Enfin, vous avez une annexe spécifique à chaque utilisateur, pour chaque demande de compte. Dans cette annexe, un engagement individuel doit être signé pour ne pas rétrocéder de données sous forme brute.

#### De la salle

Des médecins DIM ont-ils déjà réalisé cette procédure ?

#### Sandra STEUNOU

Je pense que quelques-uns l'ont fait.

#### De la salle

J'espère en effet ne pas être le seul à l'avoir fait. La procédure est très simple et très peu engageante. Pour nous, dans le secteur privé, chez Elsan, il s'agit d'un retour à la normalité, puisque nous avons été privés de base nationale pendant deux ans. Les techniciens de l'Agence font preuve d'une grande collaboration.

#### De la salle

De notre côté, nous avons toujours eu accès aux bases. Cette année, nous nous sommes vus refuser la MR-005 parce que nous sommes un groupe de santé, et non un établissement de santé. Cette subtilité nous a obligés à déposer de nouveau un dossier INDS, CEREES, etc. Il suffirait pourtant que nous déclarions les accès auprès de n'importe lequel de nos établissements. Je regrette que la procédure nous soit complexifiée, alors que le dispositif visait au contraire à une simplification. Je trouve dommage que les groupes coopératifs autour des établissements ne puissent pas avoir l'accès.

#### Sandra STEUNOU

Je crois que vos demandes consécutives n'ont pas été prises en compte. La MR— 005 est vraiment spécifique aux établissements et aux fédérations.

#### De la salle

Nous utilisons tout de même le serveur depuis son ouverture. Un accès SAS et R Studio a effectivement été ouvert. J'ai cependant transmis une remarque à laquelle je n'ai jamais eu de réponse. Nous n'avons aucun accès aux bases via R Studio, mais uniquement via SAS. Notre groupe n'utilise pas SAS, et souhaiterait s'en affranchir, ce qui n'est pas possible aujourd'hui. Une évolution est-elle prévue pour répondre à cette demande ?

### Sandra STEUNOU

Nous avons ajouté R Studio après SAS, en complément, suite à plusieurs demandes. Le lien entre les deux n'est pas possible actuellement. Pour SAS, nous passons en effet par un serveur de métadonnées qui gère tous les droits d'accès. La connexion avec R Studio n'est donc pas possible. Nous sommes cependant en train de réfléchir à une architecture qui permettrait d'accéder aux données avec d'autres logiciels que SAS.

Actuellement, pour ceux qui souhaitent travailler sur R Studio, la procédure est de faire une extraction sur SAS Enterprise Guide pour pouvoir avoir des données sous le format SAS, en CSV ou en TXT. Ces données peuvent ensuite être intégrées dans R Studio.

### De la salle

Je souhaiterais revenir rapidement sur la notion de secret statistique et données agrégées. Qu'entendez-vous par données agrégées ? Nous avons par exemple des données agrégées qui ne respectent pas le secret statistique et permettent de retrouver la personne.

### Sandra STEUNOU

Aucune définition précise n'existe à ce jour pour les données agrégées. La DREES a produit un rapport intitulé « SRisque faible », qui permet d'aiguiller un peu les utilisateurs. La question reste toutefois en suspens et il est compliqué de définir les données agrégées. Il est cependant certain qu'une extraction de tables, voire une table, ne peut pas être utilisée. Les identifiants de séjour, les identifiants de patients, les codes géographiques de résidence, l'âge du patient, etc., font partie des données sensibles.

### De la salle

J'étais auparavant dans un hôpital public, j'avais donc accès aux données. Aujourd'hui, je suis dans un hôpital privé. La convention est-elle payante ?

### Sandra STEUNOU

Non, tout est gratuit, que ce soit pour un établissement public ou privé.

### De la salle

Bonjour, je suis DIM à la Fondation Santé des Étudiants de France. En tant que DIM de fondation qui gère douze établissements sans médecin DIM, les accès m'ont été refusés. En revanche, j'ai un DIM non médecin qui a reçu un Token. Je n'en sais pas plus pour le moment.

Des formations à SAS sont-elles prévues et par qui seront-elles organisées ?

### Sandra STEUNOU

Nous avons organisé une série de sessions en 2016 et nous avons formé environ 100 personnes. Aujourd'hui, des tutoriels sont à votre disposition sur le serveur et reprennent tous les éléments vus pendant les formations. L'EHESP de Rennes propose également des formations sur les données du serveur. Plusieurs sessions sont organisées chaque année. Des organismes de formation indépendants proposent également des formations spécifiques au logiciel SAS Enterprise Guide.

### De la salle

Bonjour, je suis médecin DIM à l'hôpital Saint-Joseph à Paris. Nous évoquons tout à l'heure les RPU. Qu'est-il possible de faire avec les RPU ? Des expériences d'analyses ont-elles été réalisées sur ces aspects ?

### Sandra STEUNOU

Pour l'instant, les RPU ne sont pas chaînés avec le PMSI. Seule une analyse de l'activité des urgences est donc possible. Aucun chaînage n'est possible, il s'agit d'une base annexe. Nous avons accès à tous les passages aux urgences des années 2016 et 2017. Près de 20 millions de passages sont donc recensés, avec les diagnostics et les actes.

La qualité des données s'avère correcte du point de vue global. En revanche, la situation est moins satisfaisante lorsque nous entrons dans le détail. Nous sommes en train de mettre en place une procédure pour améliorer la qualité des données et leur exhaustivité en faisant des relances auprès des gestionnaires d'établissements Syrius.

Une réflexion est également menée pour inclure le chaînage entre les RPU et le PMSI.

### De la salle

Quel est le délai moyen entre le moment où nous commençons la démarche et le moment où nous recevons le Token ?

### Sandra STEUNOU

Par la Poste, actuellement, le délai moyen est de 30 jours. Nous espérons le réduire avec la procédure simplifiée, qui limitera les échanges.

### De la salle

Quels sont les facteurs de ralentissement des VM qui sont installés ? Lorsque nous travaillons sur des volumes importants, nous avons l'impression, à certains moments de la journée, nous avons des difficultés sur des territoires étendus.

### Brice SAUVAJON

Nous n'avons jamais vraiment observé ces situations au niveau de l'utilisation des VM, que ce soit sur la

mémoire ou le CPU. Le problème provient peut-être de Guide. Nous pourrions tester sur une autre machine.

Les VM sont des machines virtuelles. Lorsque nous avons une infrastructure importante, nous faisons tourner plusieurs serveurs virtuels sur un seul serveur physique.

Je vous invite à transmettre un ticket au support avec les observations que vous avez faites, et nous essayerons d'enquêter sur cette problématique.

#### Sandra STEUNOU

En général, les machines sont loin d'être surutilisées. Nous sommes loin d'atteindre les ressources maximums des machines. Nous pouvons toutefois parfois rencontrer des dysfonctionnements lors de la saturation d'un espace temporaire par un utilisateur suite à une mauvaise manipulation.

#### Damien MURE

Je vous propose de passer à la présentation suivante.

## II. Une vision d'établissement

#### Dr Michel LEHMANN

Bonjour à tous, je suis médecin DIM au CHU de Montpellier à mi-temps. J'utilise les bases nationales depuis très longtemps maintenant, puisque j'ai travaillé dans différentes institutions. Les premières utilisations que j'en ai remontent à 1997 ou 1998. J'ai vu passer différentes générations d'échanges avec l'ATIH. Personnellement, je ne peux que me féliciter des progrès que nous constatons. Ces progrès s'accompagnent d'interactions croissantes avec les personnes de l'ATIH qui mettent nos propres données à notre disposition, et je les en remercie beaucoup.

Ma présentation va être très rapide pour prendre plus de temps sur une illustration en direct du dispositif.

Je précise que je travaille avec d'autres établissements. Je suis médecin responsable du DIM, ce qui me confère plusieurs accès. Cette situation ne me simplifie pas la vie du point de vue des déclarations, d'enregistrement et de traçabilité des études que je peux être amené à réaliser.

Nous avons déjà parlé de la composition de l'accès sécurisé : tables SAS et données PMSI, dans lesquelles tous les champs sont représentés et qui sont mises à jour de manière hebdomadaire, ce qui nous donne une vision sur l'activité des établissements en cours, même si nous savons que cette activité est susceptible de changer.

Les actualisations plus tardives, réalisées par Lamda, manquent. Cependant, nous avons déjà échangé sur le sujet avec l'ATIH.

Je vais peut-être répéter certains aspects déjà évoqués. Nous trouvons également des données complémentaires dont les statistiques d'activité par établissement, mises à disposition par la DREES, des

données de population mises à disposition par l'INSEE et un ensemble d'informations sur les établissements. Il manque cependant une information concernant la filiation entre les établissements. Lorsque trois établissements fusionnent ou lorsqu'un nouvel établissement naît, nous avons toujours des difficultés à identifier les filiations, ce qui peut aboutir à de petites anomalies.

Le dispositif nous donne également la possibilité de stocker localement nos propres référentiels. Nous pouvons donc stocker des programmes informatiques et nos propres tables.

L'utilisation des bases de données est indiquée pour toute étude comparative nécessitant de mieux connaître les données d'hospitalisation des établissements tiers. Lorsque vous travaillez sur vos propres données, vous n'avez pas besoin d'accéder au serveur de l'ATIH. Cependant, nous nous prenons très vite au jeu de la comparaison avec des établissements tiers. Il est extrêmement enrichissant de savoir ce que font nos concurrents ou nos pairs. De nombreux éléments intéressants peuvent ressortir de ces comparaisons interétablissements.

L'utilisation du serveur est également indiquée pour répondre à toute question ponctuelle visant à faire évoluer des pratiques de codage (mesures d'incidence, repérage d'établissements atypiques, sollicitations directes éventuelles, etc.). Dans les années 2006-2012, nous avons beaucoup travaillé sur le sujet à Montpellier. Il nous est même arrivé d'interroger d'autres établissements sur leurs pratiques de codage, grâce aux bases nationales.

Enfin, le serveur nous permet de recourir aux divers référentiels disponibles à des fins internes.



FIGURE 2

FIGURE 2

**Indications**

- toute étude comparative nécessitant de mieux connaître les activités d'hospitalisation des établissements tiers.
- toute question ponctuelle visant à faire évoluer des pratiques de codage
  - mesure d'incidence
  - repérage d'établissements atypiques
  - sollicitation directe éventuelle.
- recours aux divers référentiels disponibles à des fins internes.

Données de Santé ATIH, Lyon, 20/11/2018

Vous retrouvez la méthodologie en un clic sur le site de la CNIL.

Tout a été dit sur l'accès et les modalités d'administration. Mon Token a été renouvelé cet été en moins de quinze jours. Une fois que vous l'avez, il est important de veiller à ne pas le perdre. Le seul petit bémol dont je souhaite parler est que, étant un utilisateur multicasquettes puisque j'interviens pour différents établissements, je souhaiterais pouvoir accéder à mon environnement de travail avec une seule clé, ce qui n'est pas possible techniquement.

Le partage des traitements et données peut être réalisé par plusieurs utilisateurs sur un même site client.

Nous verrons un petit exemple du langage SAS un peu plus loin.

Les exports sont possibles sous divers formats. Je pense que nous serons rapidement dépassés par la richesse de reporting offerte par Enterprise Guide. Je vous propose ensuite quelques mises en garde et précautions d'emploi.

Pour commencer, les facilités d'accès et temps de réponse peuvent provoquer un phénomène d'accoutumance et induire un abus, voire une addiction à l'usage des données.

Cette facilité ne doit pas faire oublier que le propre d'un résultat statistique est d'être le plus souvent faux (si le calcul réalisé est erroné) ou de répondre à une autre question que celle posée. Face à la facilité d'usage, nous perdons parfois de vue l'intérêt et la pertinence des résultats.

Je conseille donc un travail le plus collégial possible. À Montpellier, je délègue les traitements à une tierce personne pour pouvoir discuter avec lui ou elle des résultats. Il est beaucoup plus rentable et économique d'utiliser plusieurs intelligences.

Dans le cadre des modalités de suivi, parmi les suggestions, nous avons évoqué la possibilité de développer les interactions entre utilisateurs, pour que nous puissions, comme sur n'importe quel forum, échanger nos bonnes pratiques, nos manières d'utiliser les statistiques que nous générons, etc. C'est la raison pour laquelle, en plus d'une foire aux questions, j'ai proposé la notion de foire aux idées. Plutôt que toujours être dans la doléance, être dans la proposition peut être intéressant.

Nous avons également suggéré des formations dédiées pour les *data managers*. Je pense que le besoin est important en France. La présence de personnes qui, sans être médecins, comprennent ce qu'ils traitent et savent le traiter pourrait, au sein des GHT, transformer l'exploitation de la donnée et permettre une vision plus stratégique de l'information médicale.

Je vais maintenant vous montrer le fonctionnement de l'accès sécurisé.

*Une démonstration de l'accès au serveur est proposée par le Dr Michel LEHMANN.*

Je travaille vraiment comme si j'étais sur mon propre ordinateur. Dans les tables, l'ATIH a fait un travail très intéressant de labellisation des variables.



Chacun structure ses programmes comme il le souhaite. Je vous ai proposé un exemple sur le développement de la récupération accélérée après une chirurgie orthopédique d'un membre inférieur. J'ai une table de format qui permet de récupérer des données et qui va me donner des libellés de GHM comprenant le mot «Sprothèse». J'utilise beaucoup les formats, sous SAS, qui remplacent assez efficacement les jointures SQL, et qui apportent une plus grande souplesse pour réaliser certaines tâches. Je crée ensuite une table, l'équivalent d'une feuille Excel, où je définis des variables et leur affecte certaines propriétés. J'interroge ensuite la base MCO de 2018, qui comprend une vingtaine de millions de données. J'obtiens la réponse en quelques secondes.

Ensuite, avec SAS, nous pouvons utiliser différentes méthodes pour sélectionner des données.

Je pourrais vous montrer un programme un peu plus développé, sur lequel nous pouvons lire plusieurs années de suite très facilement.

## V. Utilité pour la recherche

### Antoine DUCLOS

Nous souhaitons remercier l'ATIH pour leur invitation. Avec Cécile Payet et Stéphanie Polazzi, nous représentons le service des données de santé des Hospices civils de Lyon. La vocation de ce service est d'exploiter les données disponibles aussi bien sur les entrepôts de données locaux que sur les entrepôts de données nationaux. Nous avons testé le CASD, l'autre système d'exploitation de la donnée PMSI. Nous travaillons également sur la plateforme ATIH, la plateforme EGB, la plateforme SNDS, mais aussi sur nos propres plateformes. La plateforme de l'ATIH fonctionne bien. L'ATIH a fait un travail remarquable dans un contexte réglementaire qui a beaucoup évolué. La loi a été très mouvante ces derniers temps, et

### 1. Conception d'études

Le CHU travaille sur ces données PMSI depuis 20 à 30 ans, d'abord sur des données locales, puis progressivement sur des données régionales, puis nationales. Aujourd'hui, dès lors que nous avons une granularité suffisante disponible dans ces bases, nous travaillons sur ces données nationales.

Nous travaillons aussi bien sur des études observationnelles qu'interventionnelles. Nous pouvons être amenés à développer des schémas expérimentaux.

De nombreuses méthodes permettent d'accéder aux données. Le système de glisser/déposer est également très utilisé. Nous pouvons facilement ajouter des filtres pour affiner la sélection.

Lorsqu'on ne connaît pas bien SAS, un temps d'acclimatation et d'apprentissage est nécessaire. De nombreuses personnes sont cependant capables rapidement de l'utiliser de façon rationnelle et efficiente.

Je ne vais pas entrer plus en détail dans la démonstration, je préfère garder du temps pour répondre à vos questions ou vos remarques. Je vous remercie pour votre attention.

### Damien MURE

De nombreuses questions ont été posées suite à la première intervention, nous pouvons passer la parole aux représentants des HCL.

l'ATIH a réussi à mettre à notre disposition un dispositif qui fonctionne dans un temps record.

Nous allons procéder en trois temps. Je vais commencer par vous présenter notre historique et le type d'études que nous menons sur les données de la plateforme ATIH.

Stéphanie Polazzi vous expliquera ensuite de quelle manière elle prépare ces données, nous prendrons l'exemple de deux projets en particulier.

Enfin, Cécile Payet vous présentera la manière dont elle analyse les données préparées par Stéphanie.

Nous utilisons également les données pour des travaux méthodologiques. Nous nous interrogeons beaucoup sur la qualité des données. Nous sommes amenés à utiliser différents standards pour voir si les données que nous retrouvons dans les bases correspondent à la réalité du patient. À ce titre, nous avons des conventions spécifiques, notamment avec l'ATIH, pour voir par exemple si nous pouvons mettre à disposition des ratios normalisés de mortalité hospitalière pour les établissements.

Historiquement, nous sommes amenés à faire beaucoup de choses avec l'ATIH.

Pour n'importe quelle étude que nous menons, la première étape – celle qui nous prend le plus de temps – est la formulation précise d'une hypothèse ou d'une question de recherche.

Avant de faire le moindre travail sur les données, nous protocolisons. Les protocoles, en fonction des études, peuvent être plus ou moins longs.

Les éléments clés d'un protocole sont généralement les suivants :

- titre de l'étude ;
- investigateurs avec accès aux données ;
- rationnel avec justification du caractère d'intérêt public de l'étude ;
- objectif(s) ;
- méthodologie : schéma d'étude, sources de données, population étudiée, critères d'évaluation, variables d'intérêt, analyses ;
- calendrier prévisionnel ;
- références bibliographiques.

Si nous n'avons pas tous ces éléments, nous ne touchons pas aux données. La MR-005 peut paraître compliquée, mais elle est très simple. Pour nous, il est plus simple aujourd'hui de faire une étude sur les données de l'ATIH en national que sur les données de nos propres entrepôts. Toutes les procédures sont en place et la plateforme fonctionne bien, tout peut

donc aller très vite. Encore une fois, nous ne faisons pas forcément des protocoles de 50 pages.

Les étapes de la MR— 005 sont les suivantes :

- déposer un protocole à l'INDS ;
- remplir la Déclaration publique d'intérêt ;
- rédiger une Fiche Résumé INDS ;
- fixer une date prévisionnelle de fin d'étude ;
- restituer les résultats à la fin de l'étude à l'INDS.


À ce jour, seules quatre études sont déclarées en MR-005 auprès de l'INDS en France, dont deux portées par les HCL. Nous devons faire ces déclarations. En cas de dérapages, les patients pourraient nous reprocher notre manque de transparence.

L'une des deux études menées par les HCL est l'étude Bariatrique, dont l'objectif est de comparer le risque de reprise chirurgicale entre différents types d'interventions (sleeve, by-pass et anneau). Cette étude est donc déclarée et affichée sur le site de l'INDS.

La deuxième étude, l'étude Elipse, a été démarrée avant la mise en place de la méthodologie de référence. Nous sommes donc passés par le circuit classique de la CNIL et elle n'est pas affichée sur le site de l'INDS. L'objectif de cette étude est de voir si les patients opérés dans des hôpitaux réduisant significativement leur durée de séjour au cours du temps étaient exposés à un risque plus élevé de réhospitalisation pour événement indésirable grave après l'opération.

## 2. Préparation des données

### Stéphanie POLAZZI

Toutes les personnes qui ont accès au serveur au sein d'un établissement ont accès au même espace commun, où elles peuvent partager des données. Nous nous servons de cet espace pour partager tout ce qui concerne le projet et avoir un suivi entre tous les acteurs du projet.  **FIGURE 3**

Notre espace commun et le dossier projet comprend les éléments suivants :

- les programmes de création de la base et des analyses ;
- la base de données ;
- le dictionnaire des variables ;
- le flowchart ;
- le rapport de data-management et d'analyse.

Nous avons donc un dossier par projet.

En parallèle, nous pouvons construire des tables à partir des tables initiales que nous utilisons assez régulièrement. Etant donné que nous travaillons la

plupart du temps en chirurgie, nous avons besoin d'horodatage des actes. Nous reconstruisons donc une table d'acte où l'horodatage est corrigé. Nous nous sommes en effet rendu compte, en faisant des contrôles de cohérence, que certains horodatages étaient un peu aberrants. Nous reconstruisons donc une table que nous pouvons stocker facilement dans l'espace commun. Nous utilisons les horodatages pour calculer nos *outcomes* à partir de dates, que ce soit pour le projet Elipse ou le projet Bariatrique.

Nous construisons également d'autres tables comme les urgences, une table des décès, etc.

Pour ces études, nous utilisons moins de la moitié, voire un tiers des tables qui existent pour les données MCO. Nous utilisons un historique de données assez important pour réaliser nos études (neuf ans pour l'étude Elipse et six ans pour l'étude Bariatrique).

FIGURE 3



Je vous ai également mentionné les tables complémentaires que nous avons utilisées pour chacune des deux études. Pour Elipse, par exemple, nous souhaitons avoir le nombre de lits de MCO, de manière à ajuster les facteurs confondants. Nous nous sommes également servis des données de la SAE. Pour l'étude Bariatrique, nous souhaitons récupérer toutes les consommations de soins de l'année précédente, nous nous sommes donc servis de tous les DMI en valorisation.

Les nomenclatures sont très nombreuses, nous en découvrons quasi quotidiennement. Nous utilisons

de nombreux codes de correspondance entre le code GEO et le code COMMUNE. Pour l'étude Bariatrique, étant donné que nous avons besoin de faire des classifications de la CCAM, des codes CIM-10 et des dispositifs médicaux, nous avons découvert toute la hiérarchie qui existe.

Au final, nous téléchargeons de nombreuses données. Nous utilisons des données médico-sociales pour le patient, le revenu médian des ménages de la commune, et les données ENC ainsi que la classification ATC.

### 3. Analyse des données

#### Cécile PAYET

Je vais vous présenter rapidement la manière dont nous analysons les données. Deux logiciels sont disponibles pour réaliser les analyses statistiques : SAS Enterprise Guide et R Studio. Nous utilisons également Microsoft Office 2013, et notamment Word et Excel. Le fait que tous ces outils soient disponibles sur le serveur nous permet d'automatiser nos rapports et de mettre directement les résultats dans les rapports.

Je vais vous parler plus précisément de l'utilisation de SAS avec l'exemple de l'étude Bariatrique, puis de l'utilisation de R Studio avec l'étude Elipse. Nous utilisons beaucoup plus SAS, mais R Studio nous permet d'accéder à beaucoup plus de nouveautés.

Pour l'étude Bariatrique, nous souhaitons donc comparer le risque de reprise chirurgicale entre différents types d'interventions (sleeve, by-pass et anneau).

Dans cette étude, nous avons apparié un patient opéré d'une sleeve avec un patient opéré d'un by-

pass, ainsi qu'un patient opéré d'une sleeve avec un patient opéré d'un anneau au sein de chaque établissement, ce qui nous permettait de tenir compte des caractéristiques des patients. Nous avons souhaité également tenir compte des facteurs organisationnels des établissements, c'est la raison pour laquelle nous avons apparié les patients au sein de chaque établissement.

Pour réaliser ces appariements, nous avons utilisé le score de propension, qui nous permet d'avoir une valeur par patient, qui tient compte de toutes ces caractéristiques. Nous avons utilisé une procédure disponible dans SAS.

Dans un deuxième temps, nous avons voulu appairer les patients en prenant en compte plusieurs éléments, en nous servant de toute l'information disponible dans les bases. Nous sommes allés regarder les codes des actes, des diagnostics, des DMI, des molécules onéreuses, etc. Une méthodologie est

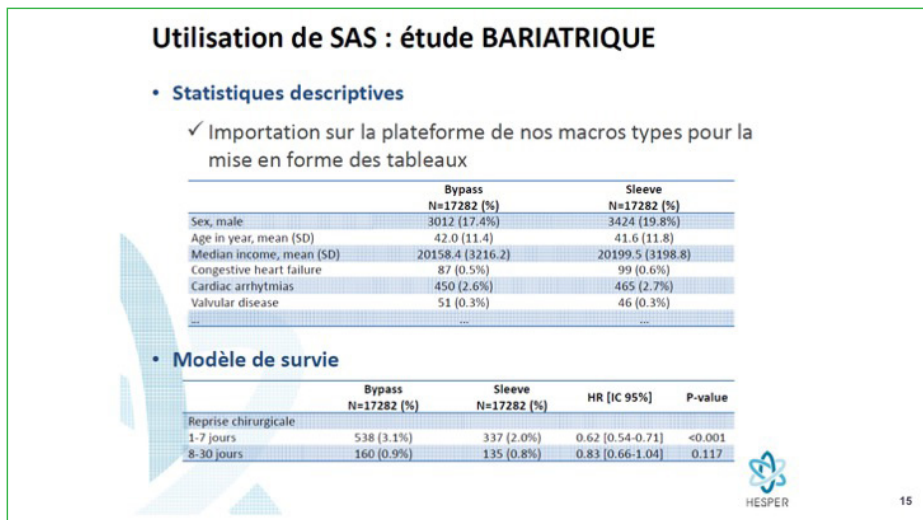
disponible pour ce faire, il s'agit des scores de pro-  
pension à haute dimension. Ces scores ne sont cepen-  
dant pas disponibles sur SAS, nous avons donc  
dû importer ces macros et le fichier associé, écrit en  
Java. Nous avons tout de même demandé l'accord à  
l'ATIH pour nous assurer que ce fichier était compa-  
tible avec la plateforme.

Nous utilisons donc SAS avec ses procédures clas-  
siques, mais nous utilisons également parfois des  
macros disponibles par ailleurs, que nous impor-  
tons et qui nous permettent de réaliser nos analyses

statistiques, comme si nous étions en local sur notre  
ordinateur.

Ensuite, nous réalisons nos rapports statistiques en  
effectuant les statistiques descriptives ou les réalisa-  
tions que nous avons à faire. Auparavant, nous avions  
des macros types qui nous permettaient d'automati-  
ser nos rapports, de mettre en forme des tableaux,  
des graphiques, etc. Nous avons importé toutes ces  
macros sur le serveur, et nous les utilisons en routine.  
Nous avons pu réaliser des analyses simples, comme  
des plus complexes, et, jusqu'à présent, nous avons  
toujours pu faire tourner les modèles. ➔ **FIGURE 4**

FIGURE 4



Nous sommes donc assez satisfaits du fonctionne-  
ment sur le volet analyse.

Concernant l'utilisation de R Studio, nous pouvons  
prendre l'exemple de l'étude Elipse, pour laquelle  
nous souhaitons étudier si les patients opérés dans  
des hôpitaux réduisant significativement leur durée  
de séjour au cours du temps étaient exposés à un  
risque plus élevé de réhospitalisation pour événe-  
ment indésirable grave après l'opération.

L'une des premières étapes était de définir des  
groupes d'hôpitaux selon l'évolution de leur durée  
de séjour. Nous avons donc identifié trois groupes :  
les hôpitaux qui ont une diminution majeure de leur  
durée de séjour, ceux qui ont une diminution modé-  
rée et ceux qui n'ont pas de diminution. Nous sou-  
haitons utiliser une méthode K-means, qui était plus  
facile à réaliser sur R Studio.

Sur le serveur de l'ATIH, le R disponible est le R de  
base. Nous ne pouvons donc pas importer des  
packages comme nous le ferions en local. Nous  
sommes obligés de passer par le support, qui installe  
le package demandé. Lorsque le package est instal-  
lé sur la plateforme CRAN, il n'y a pas de problème

pour les utiliser. En revanche, la situation semble plus  
complexe sur la plateforme Github.

Je peux à présent répondre à vos questions.

**De la salle**

Quels sont les critères qui vous amènent à déclarer  
un traitement réalisé sur les bases nationales ?

**Antoine DUCLOS**

Lorsque nous avons un projet constitué suite à une  
demande qui nous est faite, nous ouvrons un projet.  
L'important est de déclarer la finalité, l'intention du  
projet. Nous ne sommes pas obligés de détailler  
toutes les variables que nous allons utiliser. Le fait  
qu'un projet soit déclaré sur la plateforme fait qu'il est  
transparent pour tout le monde, à commencer par le  
patient.

**De la salle**

Les packages R installés par l'ATIH sont-ils dispo-  
nibles pour tous les utilisateurs, ou chaque utilisateur  
doit-il demander ses packages ?

**Antoine DUCLOS**

Ils sont installés pour chaque utilisateur, cependant une fois validés, l'accès est facilité pour les autres utilisateurs.

**Damien MURE**

Le répertoire qui contient les packages est commun à tous les utilisateurs de la plateforme. C'est la raison pour laquelle nous ne souhaitons pas que les utilisateurs puissent installer des packages.

Seuls les packages de la plateforme CRAN sont officiellement reconnus et supportés par les personnes de R Studio. Sur Github, aucun contrôle n'est réalisé. Nous pouvons nous permettre d'installer les packages contrôlés et de les rendre disponibles à tous.

**De la salle**

J'avais une deuxième question concernant l'antériorité des données. Etant donné que l'antériorité fonctionne avec une période glissante, comment faites-vous pour les projets qui demandent une plus grande antériorité ?

**Antoine DUCLOS**

Je crois que cette antériorité glissante est inscrite dans les textes. Il est en effet dommage de ne pas laisser les données vieillir.

**De la salle**

Si nous passons par un circuit classique INDS/CEREES/CNIL, est-il possible de charger davantage de tables sur les serveurs ?

**Antoine DUCLOS**

À priori, oui.

**De la salle**

Les packages présents sur CRAN ne sont pas supportés par R Studio. Ils ne sont pas validés, ils sont simplement diffusés par CRAN, qui s'assure simplement que les personnes qui postent le package répondent à des critères de qualité en termes de documentation de fonction. Les packages ne sont pas lus par les personnes de R Studio. Ils répondent simplement à une charte qualité.

Si vous vous limitez aux plateformes CRAN, cela signifie que vous limitez les packages développés en interne. Le fait que nous ne puissions pas utiliser les packages développés en interne sur la plateforme me paraît problématique. Techniquement, nous travaillons sur des outils que nous ne pouvons pas reproduire sur la plateforme pour des raisons d'ouverture publique.

**Damien MURE**

Ces questions n'ont pas été posées au début du projet. Cependant nous sommes ouverts pour étudier ces problématiques. Concernant les modules qui pourraient être ajoutés à la plateforme, nous devons nous poser la question de ce que font ces programmes. Des possibilités existent, nous devons travailler ensemble. Pour le moment, les modules qui sont installés sont accessibles par tout le monde.

Nous devons clore la séance de questions/réponses, je vais devoir conclure.

Merci pour les présentations que nous avons eues. J'espère que vous avez mieux compris les possibilités qu'offrait la plateforme d'accès sécurisé et que ces présentations vous ont donné envie d'y aller.

Je vous rappelle que la plateforme est gratuite pour les établissements, elle est simple d'utilisation. Notre responsable qualité mène des enquêtes pour l'ATIH, et 95 % des personnes interrogées sont satisfaites ou très satisfaites de l'utilisation de la plateforme sur l'année 2017. Nous vous avons également transmis le lien qui vous permet d'accéder au site de l'ATIH, sur la page qui vous explique la démarche à suivre pour avoir accès à la plateforme.



**ATIH**

117, bd Marius Vivier Merle  
69329 Lyon cedex 03  
Tél. : 04 37 91 33 10  
Fax: 04 37 91 33 67

**Pour plus d'informations:**  
**[www.atih.sante.fr](http://www.atih.sante.fr)**