

Note à l'attention du Comité Stratégique du Système National des Données de Santé

Objet : bilan d'étape et perspectives du groupe de travail "Financement" concernant la tarification de mise à disposition des données de santé

L'entrée en application du *Data Governance Act* européen en ce mois de septembre 2023 redéfinit le cadre légal permettant aux acteurs publics de percevoir des redevances en contrepartie de la mise à disposition de données de santé à des utilisateurs disposant des autorisations nécessaires. La mise en place d'un modèle commun de tarification pour la mise à disposition des données de santé à l'échelle nationale apparaît aujourd'hui comme une priorité pour l'ensemble de l'écosystème qui doit s'adapter à cette nouvelle réglementation. Garantissant transparence et égalité de traitement envers les porteurs de projets ainsi qu'une juste rémunération des organismes qui coordonnent la collecte des données de santé, le modèle commun permettra de faciliter la réutilisation des données de santé en y accélérant l'accès. Dans le but d'identifier les caractéristiques principales de ce futur modèle de tarification commun, le groupe de travail "Financement" a organisé au cours des six derniers mois un premier recueil des retours d'expérience et attentes de l'écosystème des données de santé. A cette occasion, plusieurs organismes dépositaires de données de santé ont partagé leurs outils internes de calcul de redevances. Présentant des caractéristiques complémentaires, les outils élaborés au sein de l'AP-HP et de l'Ouest Data Hub ont été sélectionnés pour une première phase de test à l'issue de laquelle le groupe de travail a proposé une harmonisation de ces outils afin qu'ils permettent d'aboutir à des résultats cohérents et comparables. Une nouvelle phase de test de ces outils harmonisés apparaît désormais nécessaire pour finaliser le cadrage et la définition du modèle commun de valorisation économique du patrimoine national des données de santé.

I. Définir un modèle commun pour la valorisation économique des données de santé est une priorité pour assurer la pérennisation du patrimoine national des données de santé

Lancé en juin 2022 par le comité stratégique des données de santé, le groupe de travail "Financement" a pour objectifs d'établir un état des lieux des mécanismes de financement des bases de données de santé et d'effectuer des propositions dans le but de pérenniser le financement de ces bases d'intérêt stratégique afin de consolider le patrimoine français des données de santé. Au premier trimestre 2023, le groupe a rendu un premier travail d'estimation des coûts de mise en place et de maintien d'entrepôts de données de santé (EDS) au sein des établissements de santé afin de faciliter la réutilisation secondaire de données cliniques, encore peu accessibles aujourd'hui. Ce premier travail a notamment permis d'établir un premier état des lieux des sources de financement des EDS hospitaliers et de constater qu'à l'heure actuelle, les redevances perçues dans le cadre d'activités tarifées ne représentent qu'une part très minoritaire du budget total des EDS. En effet, la plupart des acteurs travaillent aujourd'hui à la structuration et au déploiement opérationnel des EDS et peu d'entre eux ont atteint un niveau de maturité suffisant pour permettre un

accompagnement en routine de projets externes facturables. Par ailleurs, pour les quelques établissements disposant d'une offre de service, l'analyse juridique effectuée dans le cadre de ces premiers travaux du groupe de travail a démontré que les redevances perçues ne pourront en aucun cas couvrir l'intégralité du budget de fonctionnement d'un EDS. En effet, l'entrée en vigueur du *Data Governance Act* (DGA) européen en juin 2022 proscrit la possibilité pour les dépositaires de données de santé publics de générer un bénéfice grâce aux activités de mise à disposition de ces données qui doivent être facturées sur la base des coûts marginaux engendrés. Cette nouvelle réglementation européenne, pleinement applicable à partir de septembre 2023, écarte ainsi la possibilité pour les dépositaires de bases de données de santé de s'autofinancer grâce aux fonds versés par des acteurs économiques.

Dans ce contexte, le groupe de travail a souhaité poursuivre les travaux lancés en 2022 en approfondissant les réflexions relatives à la mise en place d'un modèle de tarification commun pour la mise à disposition de données de santé à l'échelle nationale. En effet, le déploiement à travers l'ensemble du territoire d'un modèle commun présente un double intérêt :

- Ce modèle permettra dans un premier temps d'identifier et de standardiser les tâches usuelles de mise à disposition de données de santé et d'harmoniser les tarifs associés ce qui permettra, à terme, d'accélérer la production de devis et la contractualisation avec les porteurs de projets disposant des autorisations nécessaires. La mise en place d'un modèle national public garantira par ailleurs une meilleure transparence vis-à-vis des porteurs de projets afin d'empêcher des traitements arbitraires.
- Il a par ailleurs pour ambition d'identifier l'ensemble minimal de tâches nécessaires à l'extraction et à la mise à disposition des données nécessaires à la conduite du projet déposé. Chaque organisme dépositaire de données de santé sera alors libre de construire, en sus de ce prérequis minimal, sa propre offre de service lui permettant de valoriser son savoir-faire et ses compétences. Le périmètre de tarification commune pourra ainsi être amené à évoluer avec le retour des établissements suite à sa mise en pratique.

II. Un travail mené sur six mois pour comprendre les attentes de l'écosystème et analyser les méthodes de valorisation déployées actuellement à l'échelle locale

Le groupe de travail, qui, similairement à la première phase des travaux, associe représentants institutionnels et représentants de divers organismes dépositaires de données de santé, a organisé le travail du dernier semestre autour de quatre objectifs principaux :

1. Recueillir les retours d'expériences et attentes de l'écosystème des données de santé en auditionnant dépositaires de données comme utilisateurs ;
2. Cadrer, selon les conclusions de cette première phase exploratoire, les caractéristiques principales d'un modèle de tarification commun ;
3. Répertoire les modèles de tarification existants et analyser les différences d'approche ;
4. Tester opérationnellement les modèles existants sur des cas d'usage concrets afin d'évaluer la réponse aux besoins exprimés par l'écosystème.

1. Conclusions de la phase de consultation de l'écosystème des données de santé

L'hiver 2023 a permis au groupe de travail de mener dix auditions, dont le détail est donné en **annexe 1**. Entrepôts, bases de données de santé, cohortes mais également acteurs économiques ont été entendus afin de dresser un premier bilan de la tarification de mise à disposition des données de santé en France. Cette phase de consultation a notamment permis de souligner les difficultés éprouvées par chaque partie dans la mise à disposition et l'accès aux données :

- **Des contractualisations et des négociations aux durées jugées excessives :** exercice complexe et chronophage pour les équipes l'estimation des coûts totaux engendrés par un projet nécessite souvent une expérience et un recul dont peu d'organismes dépositaires disposent aujourd'hui. L'absence d'un modèle ou d'un outil de tarification standardisé suppose d'itérer à plusieurs reprises au sein des équipes de l'organisme dépositaire ainsi qu'avec les équipes du porteur de projet pour stabiliser les modalités de cette tarification et pour choisir le cadre contractuel qui correspond à la demande (par exemple prestation versus collaboration). Plusieurs utilisateurs ont ainsi mentionné une attente supérieure à un an entre le dépôt de leur demande et l'accès aux données afin de s'accorder sur les contrats et sur les montants des redevances.
- **Une hétérogénéité territoriale.** L'absence d'un modèle de tarification unifié à l'échelle nationale provoque une forte variabilité de coûts d'accès pour les utilisateurs entre les différents dépositaires. Selon les utilisateurs interrogés, les devis fournis pour un même projet peuvent être multipliés par dix en fonction du dépositaire sollicité¹. Il est toutefois à noter que le groupe de travail n'a, à ce jour, pas identifié de modèle de tarification unifié à l'échelle nationale dans d'autres pays.
- **Une connaissance des données disponibles et une harmonisation des modalités de tarification insuffisantes.** L'ensemble des utilisateurs auditionnés soulignent par ailleurs une absence d'informations publiques concernant le contenu des bases de données françaises et les tarifs applicables pour leur mise à disposition auprès d'acteurs ayant obtenu les autorisations nécessaires.

Globalement, les acteurs auditionnés soulignent à travers ces constats le manque de compétitivité des acteurs français de la donnée de santé qui pratiqueraient par ailleurs des tarifs parmi les plus élevés à l'international pour des données comparables.

Cette consultation de l'écosystème des données de santé ne s'est également pas limitée aux frontières françaises puisque le groupe de travail a eu l'opportunité d'auditionner Findata, autorité en charge des autorisations d'accès et de la mise à disposition des données du système de santé finlandais. Findata a déployé en 2019 un modèle de tarification de mise à disposition des données de la population finlandaise basé sur une logique de coûts marginaux en accord avec le DGA. Trois ans plus tard, les recettes de la tarification s'élèvent à environ 380k€ par an pour Findata et 1,3M€ pour l'ensemble des dépositaires de données de santé finlandais². L'autorité Finlandaise estime ainsi avoir atteint un plafond quant au niveau de revenu pouvant être issu de la tarification de son activité de centralisation et de mise à disposition des données³.

2. Premières conclusions relatives au cadrage d'un modèle de tarification commun

¹ Observation faite par l'entreprise Owkin et communiquée au groupe de travail lors de leur audition par ce dernier le 17 mars 2023.

² Source : <https://findata.fi/en/about-findata/annual-report-2022/#2022-in-figures>

³ Demeurent toutefois en-dehors du champ de FINDATA les autorisations d'accès accordées par les détenteurs de données publics directement, lorsqu'un seul détenteur est concerné.

Les entretiens réalisés ont permis au groupe de travail de conclure sur la nécessité d'appuyer opérationnellement les dépositaires de données de santé dans l'exercice d'estimation des coûts d'accompagnement d'un projet en fournissant, à court terme, un outillage adapté. L'accès à un tel outil permettra aux dépositaires de données de santé d'accélérer la fourniture d'un devis aux utilisateurs, phase clé et préalable de la contractualisation entre les parties. Selon les attentes de l'écosystème, cet outil devra :

- Permettre d'identifier les principaux postes de coûts liés à la mise à disposition de données à un porteur de projet afin de générer une tarification sur la base de ces coûts en accord avec le DGA européen ;
- Être adaptable aux différents types de bases de données de santé ;
- Permettre l'estimation des coûts d'une variété de cas d'usages selon différents niveaux de complexité ;
- Être simple d'utilisation afin de garantir une prise en main aisée pour l'ensemble des acteurs, quel que soit leur niveau de maturité.

L'élaboration d'un tel outil commun semble une première étape fondamentale dans la constitution d'un modèle commun garantissant à terme une tarification juste, rapide, transparente et non-arbitraire, conforme à la loi et plébiscitée par les utilisateurs.

3. Analyse des outils existants

Élément clé de la viabilité économique des bases de données de santé, plusieurs organismes dépositaires ont travaillé à la conception d'outils internes d'évaluation des coûts d'un projet afin de standardiser leur processus de tarification. Le groupe de travail tient ainsi à remercier sincèrement les équipes de l'AP-HP (EDS et BNDMR) et de l'Ouest Data Hub porté par le GCS HUGO pour le partage de leurs outils respectifs au groupe de travail. Fondés sur une liste exhaustive des tâches de mise à disposition, ces grilles diffèrent dans leur approche de remplissage :

- Les outils de l'AP-HP proposent à l'utilisateur d'associer à chaque tâche un ou plusieurs profils et d'estimer *ad-hoc* la charge pour chaque profil dans la réalisation de cette tâche. Cette charge est ensuite multipliée par un coût journalier moyen, fonction du salaire attribué à chaque profil.
- L'outil développé par le Ouest Data Hub quant à lui associe à chaque tâche un profil et une charge de manière standardisée. Un ensemble de variables permet de qualifier la complexité du projet (par exemple : nombre de centres de données impliqués, chaînage avec d'autres bases, etc) et génère des scores de majoration de la charge standard estimée pour chaque tâche.

Chaque grille prévoit par ailleurs une majoration ou une minoration du coût final, fonction de plusieurs facteurs : profil du porteur de projet, intérêt scientifique du projet ou encore intérêt du projet pour l'entrepôt. Aucun de ces coefficients n'a été fixé à l'heure actuelle dans l'attente des réflexions menées à l'échelle nationale.

4. Réponse de l'existant aux besoins exprimés par les membres

Afin d'évaluer l'adéquation des outils proposés avec les attentes exprimées par les membres du groupe de travail, une phase de test des grilles partagées par l'AP-HP et par le Ouest Data Hub a été lancée à l'été 2023. Pendant un mois, cinq établissements membres du groupe de travail ont testé les deux grilles sur un ensemble de cas d'usage concrets. Cette phase de test, dont une description détaillée est disponible en **annexe 2**, a permis de confirmer l'intérêt pour chaque dépositaire de disposer d'un outil de modélisation des coûts

d'un projet. Les deux outils ont été appréciés par les membres qui ont toutefois souligné des écarts dans leurs contextes d'utilisation respectifs :

- L'outil proposé par l'AP-HP ne proposant pas de modélisation *a priori* de la charge et des profils mobilisés est plus flexible dans l'estimation des coûts. Chaque utilisateur peut adapter le chiffrage aux spécificités du projet considéré. Son remplissage nécessite recul et expérience.
- L'outil proposé par le GCS HUGO quant à lui est particulièrement adapté au chiffrage de projets mobilisant plusieurs centres de données. La modélisation *a priori* de la charge a été appréciée par plusieurs établissements car elle permet d'aboutir rapidement à une première estimation des coûts totaux pour des projets sur lesquels les établissements disposent de peu de recul. La complexité de la grille la rend toutefois complexe d'utilisation pour les établissements moins matures et il est à noter que l'estimation initiale obtenue requiert une revue et validation par les experts du Ouest Data Hub afin de s'assurer de sa bonne adéquation aux exigences du projet considéré.

III. A court terme, le groupe de travail propose une harmonisation des méthodes de calcul des redevances

La phase de test réalisée par le groupe a permis de conclure quant à l'utilité des deux outils qui, dans leur différence d'approche dans le chiffrage des coûts d'un projet, permettent a priori de répondre à des besoins complémentaires des établissements. Le groupe de travail a ainsi statué sur l'intérêt de poursuivre l'exploration de deux outils communs, fondés respectivement sur les modèles proposés par l'AP-HP et l'Ouest Data Hub, et harmonisés afin d'aboutir à des résultats cohérents et comparables. L'été 2023 a ainsi permis au groupe de travail de s'accorder sur des listes de tâches et de profils communs nécessaires à la mise à disposition de données et à l'accompagnement d'un projet de recherche. Le groupe a également harmonisé les coûts associés à chaque profil en reprenant les hypothèses formulées au second semestre 2022 lors de l'exercice d'estimation des coûts d'un EDS.

Concernant les facteurs de minoration et de majoration du coût total de mise à disposition, le groupe ne souhaite pas statuer à ce stade sur l'utilisation de tels facteurs. Il semble toutefois important de porter à l'attention du comité stratégique, l'impossibilité, dans le cadre de l'application du DGA, de majorer le coût total de la mise à disposition. En effet, une telle majoration rentrerait directement en contradiction avec le principe de tarification au coût marginal. Toutefois, si la majoration des coûts est exclue, le DGA stipule également que les organismes du secteur public doivent favoriser les utilisations non commerciales, les PME et les startups de la société civile et peuvent leur appliquer des tarifs réduits ou la gratuité.

IV. Un travail à prolonger sur six mois pour cadrer et définir le modèle commun de valorisation économique du patrimoine national des données de santé

Les travaux menés au cours des six derniers mois ont permis de fournir un premier état des lieux de la tarification de la mise à disposition des données de santé en France et d'établir un premier constat sur la nécessité d'établir un modèle de tarification unifié à l'échelle nationale. Permettant une rémunération des organismes organisant la collecte et le partage des données de santé et ayant vocation à assurer un traitement qui soit jugé transparent et équitable par l'ensemble des utilisateurs, ce modèle doit faciliter l'accès aux données de santé pour les porteurs de projets.

Afin de consolider ces premières observations et d'obtenir des métriques plus précises sur la réalité de la tarification entre dépositaires de données et utilisateurs, le groupe de travail souhaite lancer à l'automne 2023 une phase de test prolongée. Cette phase de test permettrait le déploiement des deux outils d'estimation des coûts d'un projet au sein des organismes partenaires afin qu'ils puissent être utilisés en conditions réelles. Lancée sur six mois, cette phase permettra d'obtenir, entre autres, des mesures concrètes concernant :

- Le nombre et le type de projets et d'études déposés auprès des organismes partenaires ;
- Le nombre de devis réalisés et les montants de ces devis par type de projet ;
- Le temps consacré à l'estimation des coûts et à la génération des devis par les équipes des organismes partenaires.

L'organisation de points d'étapes au cours des six prochains mois permettra de faire remonter régulièrement les constats et éventuelles difficultés rencontrées par les organismes partenaires dans l'application des outils et la mise en place de redevances. Ces réunions de revoyure permettront par ailleurs de faire évoluer les outils en continu en fonction des retours transmis par les organismes. Suite à cette phase de test, le groupe de travail pourra préciser le périmètre minimal des activités sujettes à une tarification homogène sur le territoire national ainsi que le principe tarifaire retenu pour les activités (avec ou sans pré-modélisation de la charge notamment), et pourra se pencher sur le déploiement d'un modèle de tarification universel et national, concourant à l'harmonisation des coûts d'accès aux données de santé essentielle dans la reconnaissance et la valorisation du travail fourni pour la constitution du patrimoine des données de santé.

Annexe 1 - Organisation du groupe de travail

Le groupe de travail s'est réuni à 10 reprises dans le cadre de séances d'une heure et demie entre les mois de février et juillet 2023 afin d'approfondir les réflexions relatives à la mise en place d'un modèle de tarification commun pour la mise à disposition de données de santé à l'échelle nationale et proposer deux outils uniques et harmonisés de calcul des redevances. Il est piloté par Jean-François Lefebvre (CHU de Toulouse) et est composé de représentants hospitaliers et institutionnels aux profils variés. Ci-dessous la liste des membres du groupe :

Liste des membres

Prénom NOM	Fonction	Organisation
Pauline Beaufrère	Responsable du GCS NOVA	GCS NOVA
Leila Ben-Janette	Directrice administrative et financière	France Cohortes
Léonardo Bonilla	Juriste en charge des projets de partenariats sur les données	CHU Lille / GCS G4
Jean-Christophe Calvo	Directeur des Systèmes d'Information	CHRU Nancy
Cécile Chevassus-Clément	Chargée de mission EDS	GCS G4
Stéphanie Combes	Directrice	HDH
Brigitte Courtois	Directrice de la recherche et de l'innovation	CHU Lille / GCS G4 / Commission recherche et innovation de la conférence des directeurs généraux de CHU
Marc Cuggia	Coordinateur médical de la plateforme Ouest Data Hub (ODH)	CHU Rennes / GCS HUGO
Quentin Demanet	Adjoint au directeur de la Recherche Clinique et de l'Innovation	AP-HP
Stéphanie Dioszeghy	Directrice de la Recherche et de l'Innovation	CHU Brest / Commission recherche et innovation de la conférence des DG de CHU

Carole Dufouil	Chargée de mission	DGRI
Caroline Dunoyer	Responsable Unité des Données de Santé	CHU Montpellier
Vanina Duwoye	Directrice des Affaires Financières	CHU Montpellier / sous commission MERRI de la commission des DAF de CHU
Florence Favrel-Feuillade	Directrice Générale	CHRU Brest / Commission recherche et innovation de la conférence des DG de CHU
Fanny Gaudin	Déléguée Générale	GCS HUGO
Charlotte Geay	Directrice de projet, Entrepôt de données Include	CHU Lille / GCS G4
Isabelle Gentil	Directrice de Projet	DNS
Claude Gissot	Directeur de projet AMDAC	DREES
Valentine Guiton	Cheffe de projet plateforme Ouest Data Hub	GCS HUGO
Anne Gysembergh-Houal	Responsable Partenariats R&D DRCI	AP-HP
Housseyni Holla	Directeur Général	ATIH
Vianney Jouhet	Responsable de l'EDS du CHU de Bordeaux	CHU Bordeaux / GCS NOVA
Jean-François Lefebvre	Directeur général	CHU de Toulouse
Anne-Laure Martin	Directrice Data et Partenariats	Unicancer
Annaëlle Paris	Chargée de mission	DGE
Christine Pichon	Directrice des systèmes d'information	CHU Rennes
Grégoire Rangé	Coordinateur médical	France PCI

Pauline Renaud	Cheffe du service Data	ATIH
Grégoire Rey	Directeur	France Cohortes
Mathieu Robain	Conseiller stratégique, Direction des Data	Unicancer
Arnaud Sandrin	Directeur opérationnel de la BNDMR	AP-HP
Hadrien Scheibert	Directeur des Affaires Financières	AP-HP/ sous commission MERRI de la commission des DAF de CHU
Ludivine Sellem	Chargée de mission accès aux données	ATIH
Pascaline Villié	Responsable des opérations du Laboratoire commun AP-HP - Inria Daniel Bernoulli	AP-HP
Vincent Vuiblet	Directeur de l'Institut d'Intelligence Artificielle en Santé	CHU Reims
-	-	DGOS

Les travaux se sont agencés autour de 3 phases principales :

1. Consultation de l'écosystème afin de réaliser un état des lieux des méthodes de valorisation déployées actuellement à l'échelle locale et mieux comprendre les attentes des acteurs
2. Test de deux outils de calcul des redevances partagés par l'AP-HP et le GCS HUGO par cinq établissements membres du groupe
3. Construction de deux outils de calcul des redevances harmonisés sur la base des attentes exprimées par l'écosystème et des retours de la phase de test

Auditions

L'état des lieux initial et le recueil des attentes de l'écosystème ont mené le groupe de travail à conduire un ensemble d'auditions avec des organismes dépositaires de données de santé comme des utilisateurs des données de santé :

- 3 février 2023 : Audition de la Banque Nationale de Données Maladies Rares
 - Personne auditionnée :
 - Arnaud Sandrin, Directeur opérationnel de la BNDMR
- 3 février 2023 : Audition de la cohorte Memento
 - Personne auditionnée :

- Carole Dufouil, Directrice de recherche INSERM et co-investigatrice principale de la cohorte Memento
- 17 février 2023 : Audition du Groupement de Coopération Sanitaire des Hôpitaux Universitaires du Grand Ouest
 - Personnes auditionnées :
 - Pr Marc Cuggia, Coordinateur médical de la plateforme Ouest Data Hub (ODH)
 - Fanny Gaudin, Déléguée générale du GCS HUGO
 - Valentine Guiton, Cheffe de projet de la plateforme ODH
- 17 février 2023 : Audition de l'Assistance Publique - Hôpitaux de Paris
 - Personnes auditionnées :
 - Anne Gysebergh-Houal, Responsable Partenariats R&D DRCI de l'AP-HP
 - Pascaline Villié, Responsable des opérations du Laboratoire commun AP-HP - Inria Daniel Bernoulli
- 10 mars 2023 : Audition de l'Agence Technique de l'information sur l'Hospitalisation
 - Personnes auditionnées :
 - Pauline Renaud, Cheffe de service Demande, Traitement et Analyse des Données (DATA) de l'ATIH
 - Ludivine Sellem, Chargée de mission accès aux données à l'ATIH
- 17 mars 2023 : Audition de l'entreprise Owkin
 - Personnes auditionnées :
 - Agathe Arlotti, Vice-Présidente en charge des partenariats chez Owkin
 - Yedidia Levy-Zauberman, Senior Vice-President Public Affairs & Impact chez Owkin
- 24 mars 2023 : Audition d'INSERM Transfert
 - Personne auditionnée :
 - Pascale Augé, Présidente du directoire d'INSERM Transfert
- 24 mars 2023 : Audition d'Unicancer
 - Personne auditionnée :
 - Mathieu Robain, Conseiller stratégique à la Direction des Data d'Unicancer
- 31 mars 2023 : Audition de l'Association Française des CROs (*Contract Research Organization*)
 - Personnes auditionnées :
 - Stéphane Bouée, Secrétaire Général de l'AFCROs
 - Magali Lemaitre, Membre du Comité Directeur de l'AFCROs
- 31 mars 2023 : Audition des Entreprises du Médicament
 - Personnes auditionnées :
 - Thomas Borel, Directeur des Affaires Scientifiques du Leem
 - Nathalie Manaud, Directrice innovation du Leem
- 9 juin 2023 : Audition de Findata
 - Personnes auditionnées :
 - Johanna Seppänen, Directrice générale de Findata
 - Mervi Siltanen, Directrice des services Data de Findata

Annexe 2 - Description de la phase de test des outils existants

La phase de consultation de l'écosystème réalisée au cours de l'hiver 2022 a permis à plusieurs membres du groupe de présenter les réflexions initiées et travaux lancés au sein de leurs établissements autour de la tarification de la mise à disposition de données de santé. Le GCS HUGO et l'AP-HP ont ainsi pu présenter au groupe de travail les travaux en cours sur l'élaboration d'outils de calcul des coûts d'accompagnement d'un projet mobilisant les données de leurs EDS respectifs.

L'AP-HP et le GCS ayant par la suite accepté de partager ces outils internes au groupe de travail, celui-ci a ainsi souhaité lancer une phase de test de ces outils auprès de différents membres. L'objectif poursuivi au cours des divers ateliers menés au mois de juin, était d'identifier, en confrontant la grille à des environnements différents de celui de l'AP-HP ou du GCS HUGO, des items de chiffrage complémentaires ou des points de difficulté rencontrés par les autres établissements afin d'affiner le cadrage d'un outil commun.

Description des outils testés

Description de l'outil de calcul du Ouest Data Hub

- L'outil élaboré par les équipes du Ouest Data Hub consiste en un tableur organisé autour de 11 onglets :
 - 5 onglets permettent de lister les tâches de mise à disposition et d'accompagnement d'un projet mobilisant les données des EDS du GCS HUGO et l'utilisation de la plateforme technologique du Ouest Data Hub. Ces tâches sont réparties en 5 lots.
 - 1 onglet permettant de qualifier le projet et les besoins d'accompagnement (par exemple : flux de données mobilisés ou besoins en matière de chaînage de données)
 - 1 onglet permettant de présenter le diagramme de Gantt du projet
 - 3 onglets de restitution du chiffrage permettant de consolider les estimations de chiffrage au sein d'un devis et d'offrir une vision de répartition des coûts par centre de données et par ressource humaine mobilisée.
- Utilisation de l'outil :
 - L'utilisateur n'a qu'un seul onglet à remplir : l'onglet de qualification du projet.
 - L'ensemble des autres onglets se mettent à jour automatiquement grâce à un calcul de la charge fonction des caractéristiques projet pour chaque ligne d'activité.
 - L'utilisateur peut également négativer manuellement certaines lignes de coût (par exemple : supprimer les coûts d'utilisation de la plateforme technologique si le porteur de projet ne souhaite pas l'utiliser).
 - Pour chaque ligne d'activité, sont affichés : la ressource humaine mobilisée, le coût journalier de la ressource, la charge estimée en jour.homme, le coût total de la tâche et le centre d'imputation du coût.

Description de l'outil de calcul de l'EDS de l'AP-HP

- L'outil développé par l'AP-HP a pour objectif d'appuyer les équipes de l'EDS dans l'élaboration de devis à destination de porteurs de projets.
- L'outil, formalisé au sein d'un tableur, est fondé sur le socle de compétences nécessaire à l'accompagnement des projets :

- L'ensemble des ressources humaines (par exemple : data engineer, data scientist, chef de projet, etc) susceptibles d'intervenir dans l'accompagnement d'un projet ont été répertoriés.
- Pour chaque ressource, un coût journalier moyen a été calculé à partir des grilles salariales de l'AP-HP.
- Utilisation de l'outil :
 - Les profils du socle de compétences en colonne et les tâches d'accompagnement du projet en ligne forment une matrice.
 - L'utilisateur remplit la matrice avec une estimation de la charge (en jour.homme) pour chaque ressource pour l'ensemble des tâches pertinentes dans le cadre du projet accompagné. La charge est estimée par les équipes de l'AP-HP à partir de leur connaissance du projet accompagné.
 - La matrice de charge est convertie en matrice de coûts humains grâce aux coûts journaliers moyens.
 - Un tableau dédié permet d'estimer les coûts d'infrastructure en cas d'utilisation de la plateforme technologique de l'AP-HP.

Cas d'usage testés

Chaque membre du groupe de travail ayant exprimé le souhait de participer à cette phase de test s'est prêté à l'exercice d'utilisation de chaque outil partagé sur un à deux cas d'usage concrets.

Le groupe a rapidement exprimé le besoin de définir un cas d'usage commun afin d'évaluer la robustesse de chaque outil aux changements d'environnement d'utilisation. Chaque membre a par ailleurs eu la possibilité de proposer un cas d'usage spécifique à son environnement afin d'illustrer la variété de projets déposés.

Cas d'usage commun

De manière générale, les projets centrés sur la donnée peuvent être décomposés en 4 étages de complexité :

- Cibler une population d'intérêt au sein d'une base ou d'un EDS (*cohorting*)
- Extraire les données liées à la population d'intérêt, créer un *datamart* et éventuellement exporter les données vers un tiers
- Chaîner les données extraites avec une autre base
- Exploiter les données, par exemple pour créer un modèle de statistiques ou d'intelligence artificielle

Dans la définition du projet commun, le groupe de travail a proposé de qualifier un projet de *cohorting* aboutissant à l'exportation d'un *datamart* pour plus de simplicité et afin que le projet puisse être pertinent pour l'ensemble des membres du GT. Le CHU de Reims a, sur la base de ce cadrage, proposé un protocole pour le projet commun visant à :

- Cibler les patients atteints de diabète de type 2, sans insuffisance rénale chronique et pris en charge en cardiologie pour décompensation cardiaque aiguë ;
- Construire un datamart à partir des données cliniques, de comorbidités, de biologie médicale, de traitement et de parcours.

Conclusions de la phase de test

- 5 membres du groupe de travail - AP-HP, GCS G4, GCS HUGO, CHU de Montpellier, CHU de Reims - ont participé à la phase de test des grilles. Au-delà du cas d'usage commun, cinq cas d'usage complémentaire ont été proposés par les membres.
- Concernant les résultats du chiffrage du projet commun, la phase de test a permis de constater :
 - Une variabilité du chiffrage entre établissements en utilisant le même outil : facteur 1 à 10 constaté lié à l'interprétation du projet et au contexte de l'établissement
 - Une variabilité du chiffrage (facteur 1 à 10) fourni par les deux outils au sein d'un même établissement
 - Que le remplissage a principalement mobilisé les équipes des EDS (responsables, chefs de projet)
- Concernant les résultats du chiffrage des projets complémentaires , la phase de test a permis de constater :
 - Comme pour le projet commun, lorsque les deux grilles sont testées, elles produisent un résultat différent
 - Les 2 grilles n'ont pas toujours été testées, chaque établissement s'étant de manière générale approprié un outil plus facilement que l'autre (répartition égale)
- La variabilité des chiffrages provient principalement :
 - De difficultés dans l'appropriation et la prise en main des grilles qui peuvent conduire à des erreurs de chiffrage.
 - D'un recoupement imparfait des listes de tâches identifiées
 - D'écarts importants dans les coûts journaliers moyens présentés dans les deux outils
 - Des hypothèses prises par les établissements testeurs pour le cas d'usage commun (par exemple pour les modalités de réalisation de la requête pour la constitution de la cohorte selon le degré de précision souhaité, les conditions techniques de l'export des données, etc.)
- Chaque membre ayant participé à la phase de test a toutefois souligné l'utilité de disposer d'un outil d'estimation des coûts d'accompagnement d'un projet qui permet de standardiser les approches dans l'élaboration des devis.

Suite à cette phase de test, le groupe de travail a ainsi décidé de conserver les deux outils de calcul et de procéder à une harmonisation des ressources, des coûts journaliers et des tâches afin de garantir une plus grande robustesse des outils.