

Le directeur

Paris, le 10/11/2020

A Madame Stéphanie Combes, directrice générale du GIP Plateforme de données de santé

Objet: Avis sur le projet Health Data Hub

Réf: - Courrier de saisine du 09 octobre 2020
- Annexe : services utilisés par la plateforme

Conformément au contrat de transformation du fonds pour la transformation de l'action publique (FTAP), vous m'avez saisi par courrier du 09 octobre 2020 pour avis concernant le projet « Health Data Hub » (HDH).

Je remercie tout d'abord le groupement d'intérêt public (GIP) pour la collaboration tant dans la préparation de ce dossier que dans son déroulement, dans un contexte de crise COVID d'une part et une période particulièrement chargée pour vous-même et vos équipes d'autre part.

Le projet HDH vise à mettre à disposition des environnements sécurisés où pourront être réalisés des projets d'intérêt public sur des données de santé pseudonymisées provenant de sources multiples. La plateforme HDH, nationale et centralisée, permet la réalisation des projets de recherche dans des conditions de sécurité et de traçabilité inégalées et sur un périmètre de données ciblées pour des utilisateurs habilités.

Si la crise sanitaire a freiné la mise en œuvre de la plateforme, des projets liés au COVID délivrent déjà leur valeur, notamment au travers de l'exploitation des données des bases OSCOUR de Santé Publique France, SNDS Fast Track de la CNAM ou encore SIVIC de la DGS, dont les réutilisateurs sont principalement publics (DREES, INSEE, université de Grenoble) avec une mise en place en cours pour la société privée Clinityx.

Le succès de la plateforme se traduit par les nombreuses (plus de 250) candidatures des deux appels à projets en janvier et décembre 2019. Le processus de sélection des projets a permis d'en retenir 36 dont la mise en œuvre s'étale sur l'année 2021. Plusieurs ouvertures

de projet sont prévues d'ici la fin du premier trimestre 2021, dont les réutilisateurs sont multiples (ICANS, EHESP, APHP, INSERM, Institut Curie, Centre Léon Berard de Lyon, CHU Limoges, DRCDC Occitanie, ARS IDF, CHU Bordeaux), avec également quelques structures privées (Implicity, E-Scopics ou encore l'ESPIC Fondation A. de Rothschild).

Vous avez pris soin de mandater un comité indépendant d'experts pour juger de la pertinence, de l'intérêt général, de l'adéquation des données demandées et de la méthodologie proposée de chacun des projets dans le cadre du processus de sélection. La valeur recherchée par ces projets n'est donc pas à discuter dans le cadre de la présente procédure. De plus, je considère que ces projets pris dans leur ensemble n'auraient pu se faire en dehors d'une telle plateforme au regard de la quantité de données à traiter et de la multiplicité des sources à manipuler dans les conditions de sécurité, de traçabilité et de délai nécessaires ne pouvant être atteintes par les projets individuellement.

Si l'ensemble des fonctionnalités de la plateforme prévues dans le dossier FTAP n'est pas encore complètement en place, je note que vous proposez un accompagnement personnalisé des projets à la fois pour l'appropriation des outils à disposition mais aussi pour permettre leur évolution en fonction des besoins projet. Par ailleurs, un comité scientifique et médical est en charge de recommander les prochains logiciels à installer sur la plateforme. Je note également la présence d'un Groupe de Travail sur les Logiciels visant à 1) recommander des logiciels à installer sur la plateforme du HDH, 2) proposer un « pipeline de certification » pour l'éligibilité des logiciels disponibles à la demande sur la plateforme, et 3) encourager le développement de logiciels et codes de recherche réutilisables pour permettre leur mise à disposition aux autres utilisateurs du HDH. Ces actions sont de nature à me rassurer concernant le bon déroulement des projets vis-à-vis de la valeur recherchée et d'une bonne gouvernance entre les outils utilisés et les développements faits dans le cadre des projets de recherche.

Après étude du dossier et des informations complémentaires échangées pendant la période d'instruction, je souhaite partager avec vous les points d'attention suivants nécessitant des mesures correctives et d'amélioration continue :

1. L'architecture en place est de qualité et à l'état de l'art, mais des éléments non redondés mettent en évidence l'existence de SPOF (single point of failure : points de défaillance uniques). Si des exigences de disponibilité réduites au lancement de la plateforme sont compréhensibles dans le cadre de projets de recherche scientifique et assumées par l'équipe du GIP, le délai de restauration / reconstruction de ces SPOF doit être améliorée.

La réplication des données sur 3 sites de région parisienne est de nature à accélérer la remise en service de la plateforme en cas d'incident grave sur le datacenter principal hébergeant les services nécessaires à son fonctionnement.

La plateforme répond aux exigences de sécurité de l'ANSSI, ce qui a permis ses homologations successives, dont la dernière a été prononcée le 28/10/2020 pour une durée de trois ans.

2. La réversibilité de la plateforme, bien que théoriquement évoquée dans les contrats des marchés support, doit faire l'objet d'une attention renforcée pour qu'elle puisse se dérouler dans des conditions financières et calendaires acceptables. L'étude de

réversibilité fournie est incomplète et ne permet pas d'établir un plan d'actions clair de nature à rassurer sur la capacité du HDH à être transféré vers un autre acteur du cloud offrant un panel de services similaires. Je note que la réversibilité est une préoccupation majeure du GIP, qu'une analyse complémentaire est en cours avec mes équipes, et que le GIP s'engage à mettre à jour cette étude annuellement.

3. Aucun processus d'optimisation des coûts permis par les infrastructures cloud n'a encore été mis en place. Sur les 11 M€ de coûts de fonctionnement annuels, 8,3 M€ sont provisionnés pour les projets (à date au nombre de 36 sur 2021) et 2 M€ pour le cœur de la plateforme (dont 400 k€ sur la volumétrie des données). Ces coûts sont cohérents et au niveau du marché pour un tel usage. Ils demandent toutefois à être optimisés conformément aux bonnes pratiques de consommation de services de cloud public.

Les coûts annuels de personnel (1,2 M€ T2) nécessaires à l'exploitation, MCO et MCS de la plateforme, appuyés par un recours aux renforts externes (en diminution de 36% à compter de 2023 pour se stabiliser à 1 M€) sont en ligne avec la maturité du projet.

4. Les engagements pris envers les citoyens sont un engagement fort du GIP visant à assurer de la transparence sur la plateforme et son usage : intérêt général, protection des données, respect des droits individuels, mais également informations sur les projets de recherche en cours ou à venir. Si la liste des projets est présente sur le site www.health-data-hub.fr, les informations disponibles sur les projets mériteraient d'être détaillées, afin que les réutilisateurs fassent eux aussi preuve de transparence concernant les projets qu'ils mènent en s'appuyant sur les données mises à leur disposition par la collectivité.

Au vu de ces éléments, les mesures suivantes devraient être mises en place dans les plus brefs délais, compte tenu de la croissance du nombre de projets que vous attendez pour les mois à venir :

1. Un plan de reprise d'activité (PRA) de la plateforme doit être réalisé et documenté, avec en plus de l'externalisation des sauvegardes, la préparation de matériels de rechange, et un travail d'architecture de résorption des SPOF.
2. En termes de sécurité, l'ANSSI souligne le travail important réalisé par le GIP pour améliorer la sécurité de ses plateformes, en particulier sur le stockage des tables de correspondances par opérateurs. En plus de maintenir l'analyse de risque à jour au regard des évolutions techniques et fonctionnelles apportées à la plateforme HDH, l'ANSSI suggère d'améliorer en continu la détection et la réponse à incident : de nombreux indicateurs ont été mis en place, mais le GIP doit améliorer la capacité de ses équipes de supervision à détecter et traiter les événements de sécurité.
3. Le plan de réversibilité doit être finalisé et tenu à jour : Une première analyse des services utilisés par la plateforme est fournie en annexe du présent avis mettant en lumière un premier niveau d'effort pour permettre l'intégration de chacun des services nécessaires au HDH. Cette étude devra être poursuivie et partagée dans l'établissement d'un cahier des charges en chiffrant le coût et la durée de substitution / transfert des services en fonction des cas d'usage, des exigences de performance et du niveau de criticité pour le HDH (fonctionnalité indispensable, fonctionnalité dont

l'étude d'une alternative peut être considérée, ou fonctionnalité de confort). Tout nouvel usage de services Azure Cloud non utilisés pour la version 1 devra être conditionné à son niveau de réversibilité et à la non remise en cause du cahier des charges des futurs services Cloud qui aura été établi.

Je tiens à préciser que l'analyse réalisée par nos équipes dans le cadre de cette procédure est de nature à me rassurer sur le fait que les choix réalisés à date par le GIP n'obèrent pas la transférabilité de la plateforme, moyennant des efforts de reconfiguration et redéveloppement de certains composants précisés en annexe. Entre autres, aucun service d'intelligence artificielle d'azure n'est utilisé sur la plateforme. Je vous demande de maintenir à jour cette l'étude de réversibilité en continuant l'analyse par service dans le but de capitaliser sur les investissements déjà réalisés, et d'évaluer la réintégration de certains services au gré de l'augmentation de leur utilisation. Mes équipes se tiennent à votre disposition dans ce cadre.

4. La mise en place d'une démarche d'optimisation des coûts d'usage cloud (« FinOps ») doit permettre de limiter les coûts de fonctionnement (achat de services) de la plateforme cloud au juste nécessaire. Cette optimisation permettra de réinvestir les budgets provisionnés pour les projets (et non dépensés) sur l'évolution de la plateforme et l'absorption de nouveaux projets.
5. Concernant les coûts de personnel, les activités de chaque grande catégorie de service offert (collecte de données, distribution de données, sécurité, création de plateforme projet) doivent faire l'objet de procédures écrites permettant d'en envisager leur automatisation selon des critères de sensibilité, de répétition, etc. Ces automatisations, lorsque pertinentes, permettront de gagner en fiabilité et en coût de personnel. Le plan de recrutement du GIP devra également tenir compte de la montée en charge.
6. Une présentation des projets sous l'angle du bénéfice (avéré ou recherché) avec une documentation claire et accessible par les citoyens mériterait d'être mise en place. Au-delà de l'aspect positif d'une telle communication sur la population, elle sera également un vecteur complémentaire d'attractivité des réutilisateurs privés.

Au vu des efforts engagés par le HDH pour conduire ce programme et de l'attente forte de l'écosystème de la recherche médicale, je considère que les éléments précités ne sont pas d'ordre à le remettre en cause. Afin de conserver et renforcer la dynamique engagée depuis le début du projet, **j'é mets un avis favorable sur le projet « HDH » sous réserve de la prise en compte des recommandations précédentes.**

Je souhaite à ce titre être informé de l'avancement des travaux et de l'application effective des recommandations.



Nadi BOU HANNA

Directeur interministériel du Numérique

Copie:

Etienne CHAMPION

Secrétaire Général des ministères chargés des affaires sociales

Fabrice LENGART

Directeur de la recherche, des études, de l'évaluation et des statistiques

Hélène BRISSET

Directrice du numérique

Laura LETOURNEAU

Déléguée ministérielle au numérique en santé