

Cahier des charges pour participer à un projet de filière sur les données d'accès précoce

Projet DANTE - Données d'Accès précoces Normalisées extraites par Traitement automatisé du langage en vue de leur Évaluation

Synthèse de ce cahier des charges

Synthèse exécutive

DANTE, pour « Données d'Accès précoces Normalisées extraites par Traitement automatisé du langage en vue de leur Évaluation », est un projet national porté par la Filière Intelligence Artificielle & Cancers avec AstraZeneca, Amgen et MSD. **Lauréat France 2030**, il s'inscrit dans la stratégie nationale d'intelligence artificielle et de données de santé. **Son ambition est simple : faire passer l'extraction automatisée de données cliniques par IA/NLP/TAL du stade de la promesse au stade de l'usage, sur un cas concret, utile et exigeant : les accès précoces.**

Les accès précoces permettent à des patients en impasse thérapeutique d'accéder plus tôt à des médicaments innovants. Ils créent aussi une responsabilité collective : documenter ce qui se passe en « vie réelle » après la mise à disposition de ces traitements. Aujourd'hui, les données de suivi restent trop souvent incomplètes, alors même qu'une grande partie des informations attendues existe déjà dans les dossiers patients informatisés. DANTE part de cette idée simple : lorsque la donnée existe déjà dans le DPI, il faut tester si l'IA peut aider à la retrouver, l'extraire et la structurer sans alourdir le quotidien des équipes de soin.

Le comité de sélection international de France 2030 a reconnu à DANTE un intérêt « fort et opérationnel », car le projet répond à un problème concret de qualité des données dans les programmes d'accès précoce et les données de vie réelle. Il a également souligné son **potentiel de généralisation**, grâce au recours à plusieurs partenaires technologiques, capables de tester les approches dans des conditions réelles. **La structuration du projet autour du RGPD, de la protection des patients, des lignes directrices de la CNIL et d'un comité de suivi dédié à la protection des données constitue également un point fort majeur.**

DANTE est un projet d'évaluation méthodologique : pour les mêmes patients, il compare les données issues de la saisie manuelle habituelle, les données extraites automatiquement par IA/NLP/TAL depuis les DPI, et une base de référence constituée par des ARC indépendants. DANTE n'est pas un classement de centres, de PME ou de réseaux, ni un benchmark commercial, ni une mise en concurrence de solutions. **Il vise à produire une photographie collective de la maturité de l'écosystème, partagée avec les autorités, afin de tirer l'ensemble des acteurs vers le haut.**

Le projet est ouvert aux centres de soins disposant d'une solution d'IA/NLP/TAL opérationnelle, aux centres travaillant avec une PME ou un prestataire spécialisé, ainsi qu'aux réseaux publics

ou privés et aux PME disposant de solutions déjà déployées ou activables rapidement. Les candidats devront être capables d'accéder aux sources nécessaires dans le DPI, d'extraire automatiquement un set prédéfini de 23 variables, de pseudonymiser les données, de les structurer dans un format exploitable et de les transmettre de façon sécurisée. Les données extraites par l'IA ne devront pas être corrigées manuellement avant transmission, car l'objectif est précisément d'évaluer la performance réelle de l'extraction automatisée.

Le modèle économique de défraiement et participation est volontairement simple et lisible : un droit d'usage de deux mille euros par dossier patient transmis conformément au protocole, dans le format attendu et dans des conditions de sécurité conformes. Cette rémunération vise à reconnaître le travail nécessaire pour mobiliser les sources, paramétrer ou activer la solution, extraire les données, les préparer, les pseudonymiser et les transmettre. Au-delà du projet, DANTE préfigure un modèle soutenable de production de données cliniques de qualité à partir des DPI.

L'intégration des partenaires se fera selon une logique de guichet ouvert : premier dossier complet, premier dossier instruit, dans la limite de l'enveloppe disponible et des objectifs scientifiques du protocole. Pour être recevable, la candidature devra comprendre deux éléments : une présentation concise du candidat et de sa capacité opérationnelle, ainsi qu'une lettre d'intention signée par les représentants habilités. Cette lettre (annexée) confirme que le candidat a pris connaissance du contrat de partenariat (disponible en ligne : [lien](#)) et en accepte le principe de signature, afin de permettre une contractualisation rapide et homogène des acteurs réellement prêts à participer.

Participer à DANTE, c'est contribuer à un projet situé à l'intersection de cinq attentes fortes : les centres souhaitent valoriser leurs données et leurs savoir-faire ; les industriels de santé ont besoin de données récentes, fines et fiables ; les autorités veulent mieux apprécier la performance en vie réelle ; l'écosystème doit pérenniser les accès précoces ; et les politiques publiques s'intéressent à la capacité de l'IA à documenter la pertinence et la qualité des soins.

Les partenaires seront associés à une démarche nationale, visible, utile et structurante, avec une restitution collective des enseignements et une contribution concrète à la méthode de probation des données extraites par IA.

Table des matières

1	Contexte et démarche de Filière	4
1.1	En synthèse	5
2	Rationnels réglementaire, médical, technologique et organisationnel	5
3	Définition des prestations attendues	6
4	Droits d'usage attachés à la collecte automatisée de données du projet	7
5	Modèle économique proposé.....	9
5.1	Principe général	9
5.2	Modalité d'intégration des partenaires : guichet ouvert « premier arrivé, premier servi »	9
5.3	Pérennité et anticipation du cadre européen de tarification	12
6	Interopérabilité, bras de référence et contrôle qualité	12
7	Sécurité des données et conformité au Règlement Général sur la Protection des Données	12
7.1	Comité de suivi dédié à la protection des données	13
7.2	La Filière Responsable de traitement.....	13
7.3	Séquence d'inclusion d'un patient à Dante	14
8	Dialogue autour des propositions de ce cahier des charges	14
9	Calendrier prévisionnel et vue d'ensemble de la prestation	14
9.1	Phase 1 – Publication échanges et ouverture du guichet.....	15
9.2	Phase 2 - Instruction au fil de l'eau et contractualisation	15
9.3	Phase 3 - Collecte automatisée par IA/NLP/TAL.....	15
9.4	Phase 4 - Constitution de la base de référence et analyse.....	15
10	Annexes	16
10.1	Définitions des acronymes utilisés, par ordre d'apparition.....	16
10.2	Set de variables à extraire	18
10.3	Modèle de lettre d'intention et d'acceptation de principe du contrat de partenariat DANTE	20

1 Contexte et démarche de Filière

La Filière Intelligence Artificielle & Cancers, ci-dessous « la Filière », est une association à but non lucratif issue du Conseil National de l'Industrie, présidé par le Premier Ministre.

Elle fédère des membres publics, dont l'Institut national du cancer, l'Agence d'Innovation en Santé et la Plateforme de Données de Santé, ainsi que des membres privés, dont l'Alliance pour la Recherche et l'Innovation des Industries de Santé, France Biotech et des industriels de santé internationaux engagés en oncologie.

Son ambition est de mener des projets préfigurateurs, c'est-à-dire des preuves de concept innovantes destinées à devenir des standards demain. Ces projets servent à tester, en conditions réelles, des méthodes nouvelles d'extraction, de structuration, de comparaison et de valorisation de données de santé, dans un cadre public-privé sécurisé et transparent.

Pour cela, la Filière publie des cahiers des charges qui décrivent un besoin, l'ouvrent à l'écosystème et permettent à des acteurs capables d'y répondre de se manifester. Le présent document vise à ouvrir, de manière simple et pédagogique, la participation au projet DANTE.

DANTE signifie : *Données d'Accès précoces Normalisées extraites par Traitement automatisé du langage en vue de leur Evaluation*. Il porte sur l'usage de solutions d'intelligence artificielle, notamment de NLP/TAL (les définitions des acronymes utilisés sont mentionnés en Annexe en page 16), pour extraire automatiquement des données utiles au suivi des médicaments en accès précoce à partir des dossiers patients informatisés.

Ce cahier des charges est public. Il a pour objet de permettre à deux grandes familles d'acteurs de candidater directement : **d'une part les centres de soins disposant d'une solution d'IA ou de NLP/TAL déjà opérationnelle ; d'autre part les PME spécialisées dans l'extraction automatisée de données cliniques, lorsqu'elles disposent d'une solution déployée ou activable dans un ou plusieurs centres.**

NB : ce cahier des charges a vocation à être mis à disposition selon les termes de la licence **Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions (CC BY-NC-SA)**. Cette licence autorise la réutilisation, la reproduction et l'adaptation du document, sous réserve de mentionner explicitement la source, de ne pas en faire un usage commercial et de diffuser toute œuvre dérivée selon une licence identique.

Les informations présentées dans ce cahier des charges sont exclusivement des informations publiques et non confidentielles. Elles ont pour seule finalité de décrire le cadre, les objectifs et les modalités du projet. Aucune information stratégique, économique ou technique propre à un acteur industriel n'y est incluse. Cette règle vise à garantir une démarche transparente, loyale et conforme au droit de la concurrence, notamment dans le secteur de la santé et de la recherche.

1.1 En synthèse

- DANTE compare deux façons de recueillir des données d'accès précoce : la saisie manuelle habituelle et l'extraction automatisée par IA/NLP/TAL,
- La comparaison se fait par rapport à une base de référence constituée par des ARC indépendants, à partir des dossiers patients informatisés. Le schéma d'ensemble est le suivant :

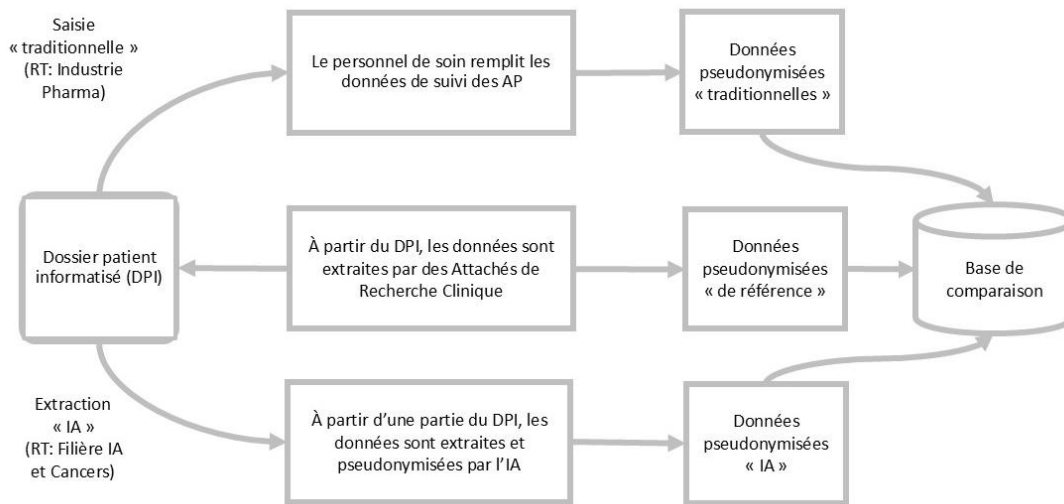


Figure 1 - Projet Dante : les flux des données

- Les candidats attendus sont les centres et les PME capables d'extraire automatiquement un set limité de variables (proposé en page 18), dans un cadre sécurisé et traçable.

2 Rationnels réglementaire, médical, technologique et organisationnel

Depuis la réforme de l'accès précoce aux médicaments, la Haute Autorité de santé encadre le dispositif d'accès précoce, en lien avec l'ANSM pour les situations préalables à l'autorisation de mise sur le marché. Les laboratoires qui demandent un accès précoce doivent soumettre un Protocole d'Utilisation Thérapeutique et de Recueil de Données, ou PUT-RD. Ce protocole prévoit des fiches à renseigner à différents moments du parcours : demande d'accès, instauration du traitement, visites de suivi et arrêt éventuel du traitement.

Ces données sont précieuses. Elles documentent souvent les premières utilisations en vie réelle de traitements innovants, dans l'indication de l'accès précoce. Elles peuvent donc éclairer les

conditions d'usage, la tolérance, l'efficacité observée et, plus largement, la compréhension du médicament en dehors des essais cliniques.

Pour autant, leur recueil est difficile. Il peut être perçu comme complexe ou redondant par les professionnels de santé, car une partie des informations demandées dans les PUT-RD est déjà présente dans le dossier patient informatisé. Les équipes sont par ailleurs confrontées à des plateformes de saisie multiples, à des ressources limitées et à des exigences variables selon les accès précoces. Ces contraintes expliquent en partie des niveaux de complétude inférieurs aux attentes des autorités.

DANTE part d'une idée simple : si une information existe déjà dans le dossier patient informatisé, alors une partie du recueil pourrait, à terme, être facilitée par des solutions d'IA capables de lire, repérer et structurer cette information. Les technologies de NLP/TAL, c'est-à-dire de traitement automatique du langage, sont particulièrement pertinentes lorsque l'information utile est contenue dans des documents non structurés : comptes rendus médicaux, lettres, notes d'évolution, prescriptions, comptes rendus d'hospitalisation ou éléments du suivi.

Le projet DANTE ne suppose pas que l'IA soit fiable par principe. Au contraire, il en organise une évaluation méthodique. Pour les mêmes patients, trois bases sont comparées : une base issue de la saisie manuelle habituelle du PUT-RD, une base issue de l'extraction automatisée par IA et une base de référence relue par des ARC indépendants à partir du dossier patient informatisé. C'est cette comparaison qui permettra d'objectiver l'exactitude, les erreurs, les omissions et la concordance des données.

Le rationnel est donc triple. Il est réglementaire, car les accès précoces exigent des données fiables et utiles à l'évaluation. Il est organisationnel, car il vise à réduire le fardeau de saisie pesant sur les professionnels. Il est technologique et stratégique, car il teste en conditions réelles la capacité de l'écosystème français à produire des données de vie réelle de qualité, récentes, traçables et comparables entre centres.

3 Définition des prestations attendues

La Filière sollicite des candidats capables d'assurer une collecte automatisée de données d'accès précoce par NLP/TAL, à partir des dossiers patients informatisés de centres participant à l'étude DANTE.

La prestation attendue vise des solutions déjà matures, déjà déployées ou activables rapidement en environnement hospitalier, capables de collecter et structurer un set prédéfini de variables dans des délais compatibles avec la conduite de l'étude.

Le candidat devra être capable de :

- Accéder aux sources nécessaires dans le dossier patient informatisé ;

- Extraire automatiquement le set minimal de variables défini en page 18 ;
- Pseudonymiser les données et utiliser le pseudonyme d'appariement défini pour DANTE ;
- Structurer ces variables dans un format exploitable (format CSV) par un Tiers de confiance analyste, mandaté par la Filière ;
- Transmettre les données de façon sécurisée ;
- Documenter les éventuelles limites connues ;
- Accepter les vérifications organisées par la Filière dans le périmètre strict de l'étude.

Une exigence est particulièrement importante : les données extraites par l'IA ne doivent pas être corrigées manuellement par le candidat avant transmission. Le but de DANTE est d'évaluer la performance réelle de l'extraction automatisée. Une correction humaine non tracée améliorerait artificiellement les résultats et rendrait la comparaison non interprétable.

La participation est ouverte :

- aux centres de soins, qu'ils soient CHU, CHRU, CH, CLCC, établissements privés ou autres établissements prescripteurs, dès lors qu'ils disposent d'une solution d'IA/NLP/TAL opérationnelle ;
- aux centres travaillant avec une PME ou un prestataire spécialisé ;
- aux réseaux, publics ou privés, ou aux PME disposant de solutions de NLP/TAL déployés en réseau, c'est-à-dire capables d'intervenir dans un ou plusieurs centres où leur solution est déjà installée, déployée ou activable rapidement ;

4 Droits d'usage attachés à la collecte automatisée de données du projet

Le présent cahier des charges a pour objet de financer une collecte automatisée de données dans le cadre du projet DANTE. Cette collecte comprend l'accès autorisé aux sources, l'extraction automatisée, la pseudonymisation, la structuration, la traçabilité et la transmission sécurisée des données nécessaires à l'étude.

Les données structurées transmises par les candidats serviront exclusivement aux finalités scientifiques et méthodologiques de DANTE : comparer la collecte manuelle habituelle, la collecte automatisée par IA et la base de référence constituée par la CRO centrale. **Elles ne pourront pas être réutilisées par la Filière pour d'autres finalités, notamment commerciales, de démonstration ou de prospection, sans base légale et autorisation spécifique.**

Par ailleurs, tous les candidats conservent l'entière propriété intellectuelle de leurs solutions, modèles, logiciels, savoir-faire, connecteurs et méthodes.

En revanche, les données traitées dans DANTE, leurs extractions pseudonymisées et les journaux d'audit associés seront utilisés, uniquement dans le cadre autorisé du projet, sous la responsabilité de traitement définie par la Filière et dans le respect des conventions signées.

Les données collectées dans le cadre du projet DANTE ont pour finalité exclusive de permettre l'évaluation scientifique et méthodologique de la capacité de l'écosystème à produire, par traitement automatisé du langage, des données d'accès précoces fiables, structurées et exploitables. Les résultats issus du projet seront produits et valorisés uniquement sous forme agrégée. Aucun résultat individualisé par centre, par établissement, par partenaire technologique ou par solution de NLP/TAL ne sera produit, communiqué ou publié.

Ce point constitue un principe fondamental du projet. DANTE n'a pas pour objet d'établir un classement, une comparaison commerciale ou un benchmark entre centres participants, ni entre solutions développées ou opérées par des PME. **L'objectif est de sonder collectivement la maturité de l'écosystème, d'identifier les conditions de réussite, les limites méthodologiques, les prérequis de qualité et les besoins de standardisation permettant, demain, de rendre ces approches plus robustes, plus reproductibles et plus utiles à l'ensemble des parties prenantes.**

En conséquence, les analyses et publications éventuelles porteront sur des tendances globales, des indicateurs agrégés, des enseignements méthodologiques et des recommandations à l'échelle de la filière. Elles ne devront pas permettre d'identifier, directement ou indirectement, la performance relative d'un centre, d'un réseau de centres, d'une PME ou d'une solution technologique particulière. Cette règle vise à garantir un cadre de confiance, à favoriser la participation la plus large possible et à préserver l'esprit partenarial du projet : apprendre collectivement plutôt que comparer les acteurs entre eux.

Les résultats agrégés de DANTE pourront être valorisés par la Filière, notamment dans des rapports, communications scientifiques, publications, échanges avec les autorités ou présentations à l'écosystème. Les modalités de publication et de communication seront encadrées par la gouvernance scientifique du projet.

Les prestations de data management global, de fusion des bases, de constitution de la base de référence, d'analyse statistique et de rédaction de rapports intermédiaires ou final ne sont pas l'objet principal de la prestation demandée au présent cahier des charges. Elles relèvent d'un Tiers de confiance mandaté par la Filière.

5 Modèle économique proposé

Le modèle économique proposé est volontairement simple. Il doit permettre de rémunérer un service réel : rendre possible, dans un centre donné, l'extraction automatisée d'un set limité de variables d'accès précoce, puis transmettre des données pseudonymisées, structurées, sécurisées et conformes au protocole.

L'enveloppe budgétaire prévisionnelle sera consommée au fil de l'eau. La Filière contractualisera les candidats éligibles dans l'ordre d'arrivée des dossiers complets et recevables, jusqu'à épuisement de cette enveloppe et/ou atteinte de l'objectif de recrutement de l'étude. L'objectif scientifique du protocole DANTE est d'atteindre 147 dossiers patients analysables, tous accès précoces confondus, avec une répartition équilibrée entre centres.

5.1 Principe général

La rémunération proposée est la suivante :

2000 euros (deux-mille euros) HT par droit d'usage d'un dossier patient transmis conformément au protocole, dans le format attendu et dans des conditions de sécurité conformes.

Cette rémunération inclut la mobilisation de ressources nécessaires à l'organisation des vérifications par la CRO centrale pour la constitution du bras de référence. Aucun frais complémentaire ne pourra être imputé à la Filière sans accord préalable et écrit.

5.2 Modalité d'intégration des partenaires : guichet ouvert « premier arrivé, premier servi »

Contrairement à un appel à projet ou à un dialogue compétitif « classique », le présent cahier des charges repose sur une logique de guichet ouvert.

Il n'y aura pas de classement général des candidats ni de sélection d'un ou plusieurs lauréats uniques. La Filière signera au fur et à mesure avec les candidats capables de réaliser la prestation demandée.

La règle est la suivante : premier dossier complet, premier instruit ; si le candidat satisfait aux exigences minimales et si le budget reste disponible, il pourra être contractualisé. La date de référence sera la date et l'heure de réception du dossier complet, et non la date d'envoi d'une manifestation d'intérêt incomplète.

Cette règle ne signifie pas que les exigences sont amoindries.

Pour être considéré comme complet, recevable et opérationnel, le dossier de candidature devra comprendre deux éléments ; (i) un dossier de candidature et (ii) une lettre d'intention signée par le ou les représentants habilités des parties candidates :

1. Le dossier de candidature doit permettre à la Filière de répondre à trois questions : le candidat est-il légitime pour intervenir dans le ou les centres proposés ? La solution est-elle opérationnelle ? Le candidat peut-il respecter le cadre de DANTE ?
 - a. Pour ce faire, un argumentaire de **10 pages MAXIMUM SVP** devra comprendre cinq éléments :
 - la présentation du candidat : centre de soin, réseau, PME ou candidature conjointe et le ou les centres de soins concernés ;
 - la présentation **succincte** de la solution d'IA/NLP/TAL : maturité, déploiements existants, sources accessibles, type de documents analysés, architecture générale ;
 - l'expérience en matière de gestion des règles du Règlement Général sur la Protection des Données ;
 - les références ou expériences comparables ;
 - le calendrier réaliste d'activation ;
 - b. Un candidat ne pourra être contractualisé que s'il démontre :
 - une solution d'IA/NLP/TAL mature et utilisable dans le ou les centres ;
 - une capacité réelle d'accès aux sources nécessaires, sous contrôle du centre et dans le respect du RGPD ;
 - une capacité à extraire automatiquement les variables attendues et un dispositif de pseudonymisation et de transmission sécurisée ;
 - une capacité à informer les patients concernés avec la **Note d'information** disponible **via ce [lien](#)¹** ;

¹ L'information du patient sera assurée au moyen de plusieurs outils complémentaires, comprenant notamment un site internet dédié au projet DANTE et une note d'information patient ; cette note devra être effectivement remise ou présentée au patient par le centre participant, qui ne devra s'engager dans le projet que s'il est en mesure de garantir que chaque patient concerné reçoit cette information dans des conditions permettant sa bonne compréhension et l'exercice effectif de ses droits. La note explique, dans

- une disponibilité compatible avec le calendrier de l'étude ;
 - l'acceptation du contrôle qualité et des vérifications prévues par la Filière.
2. La lettre d'intention, signée par le ou les représentants habilités des parties candidates, vise à **confirmer que le candidat, ou le groupement candidat lorsqu'un centre et une PME répondent ensemble, a pris connaissance du contrat de partenariat accessible en ligne via ce [lien](#) et en accepte le principe de signature.**
- a. Cette lettre est proposée en Annexe en page 19, elle est importante pour garantir la rapidité d'exécution du projet, en effet, dans une logique de recrutement au fil de l'eau, un dossier ne pourra être regardé comme pleinement opérationnel que si les parties concernées confirment, dès le dépôt de leur candidature, qu'elles sont prêtes à signer le contrat de partenariat proposé.
 - b. En pratique, selon la configuration de candidature, cette lettre devra être signée :
 - i. par le centre, lorsqu'il dispose en propre d'une solution d'IA/NLP/TAL opérationnelle ;
 - ii. par la PME, lorsqu'elle candidate en qualité de partenaire technologique disposant d'une solution en réseau ;
 - iii. ou conjointement par le centre, un réseaux ou la PME lorsque l'extraction repose sur une coopération opérationnelle entre eux.

Cette lettre permet d'attester que le candidat a compris le cadre juridique, opérationnel et financier du projet DANTE, et qu'il ne conditionne pas sa participation à une renégociation du contrat de partenariat.

Ce point reflète la logique même du projet : permettre une inclusion rapide des partenaires capables d'exécuter la prestation demandée, dans un cadre contractuel commun, homogène et sécurisé.

Le contrat prévoit précisément la possibilité pour des partenaires de rejoindre progressivement le dispositif, en signant les clauses applicables au projet DANTE.

La date de réception du dossier complet, au sens du présent cahier des charges, sera celle à laquelle la Filière aura reçu ces deux éléments demandés.

des termes accessibles, pourquoi les données sont utilisées, quelles données sont concernées, qui peut y accéder, comment la confidentialité est préservée, combien de temps les données sont conservées et comment les patients peuvent exercer leurs droits, notamment leur droit d'opposition.

5.3 Pérennité et anticipation du cadre européen de tarification

Au-delà du projet DANTE, l'enjeu est de faire émerger un modèle soutenable d'extraction de données cliniques de qualité à partir des dossiers patients informatisés. Les montants proposés visent donc aussi à préfigurer une rémunération « en run » des solutions mobilisées, afin de financer leur maintien en conditions opérationnelles, leur documentation, leur fiabilisation et leur diffusion progressive.

Cette logique anticipe le cadre européen d'accès secondaire aux données de santé et les besoins de tarification associés. **La Filière souhaite contribuer à un modèle prévisible, équitable et transparent, qui rémunère le travail réel des centres et des innovateurs tout en garantissant une qualité probante des données produites.**

6 Interopérabilité, bras de référence et contrôle qualité

Le projet DANTE n'a pas seulement pour objet de produire des données. Il vise à produire des données comparables. Les données seront structurées selon un dictionnaire commun de variables, avec un format standardisé qui sera précisé dans le plan de data management proposé par la CRO centrale.

Cette CRO centrale pourra procéder à des vérifications dans les dossiers patients informatisés, en lecture seule, afin de constituer la base de référence.

L'accès devra être limité dans le temps, journalisé et conforme aux procédures du centre. Le contrôle portera exclusivement sur le set de variables de l'étude nécessaires à l'audit de la collecte.

7 Sécurité des données et conformité au Règlement Général sur la Protection des Données

Parce que sans confiance des patients, tout le système n'a pas de sens, le projet DANTE est conduit dans **le plus strict respect du Règlement Général sur la Protection des Données**, de la loi Informatique et Libertés modifiée et des référentiels applicables aux données de santé et aux recherches n'impliquant pas la personne humaine.

Le cadre opérationnel et juridique du projet DANTE a été défini par la Filière IA & Cancers, en dialogue étroit avec la Commission Nationale de l'Informatique et des Libertés (CNIL), et documenté dans une analyse d'impact relative à la protection des données. Cette analyse d'impact, réalisée selon la méthodologie et avec l'outil PIA de la CNIL, décrit les finalités du traitement, les rôles et responsabilités des acteurs, les catégories de données traitées, le

cycle de vie des données, les mesures de sécurité, les modalités d'information des patients, les droits des personnes concernées, ainsi que l'analyse des risques et les mesures de maîtrise associées. **Elle a vocation à être tenue à jour pendant toute la durée du projet**, afin que la conformité au RGPD soit suivie comme un processus vivant et non comme une formalité accomplie une fois pour toutes. Elle est disponible **via ce [lien](#)**.

7.1 Comité de suivi dédié à la protection des données

Un comité de suivi dédié à la protection des données sera mis en place. Il associera la Filière, des volontaires parmi les acteurs opérationnels concernés et la CNIL.

Ce comité aura pour objet de suivre le projet du démarrage à sa clôture, de documenter les choix structurants, de vérifier le respect continu des principes de minimisation, de pseudonymisation, de cloisonnement, de sécurité, de traçabilité et de transparence, et d'adapter, si nécessaire, les mesures organisationnelles et techniques au fil de la vie du projet. L'objectif est de garantir que le RGPD accompagne effectivement l'exécution du projet de bout en bout.

7.2 La Filière Responsable de traitement

La Filière IA & Cancers agit en qualité de responsable de traitement pour la finalité propre du projet DANTE, à savoir la comparaison méthodologique entre les données collectées manuellement dans le cadre des accès précoces et les données extraites automatiquement par IA/NLP/TAL à partir des dossiers patients informatisés. Les centres participants, les réseaux ou PME mobilisant des solutions d'IA/NLP/TAL, ainsi que les prestataires mandatés pour les besoins du projet, interviennent en sous-traitant pour le compte de la Filière.

Le projet articule plusieurs cadres de conformité. Les données collectées manuellement dans le cadre habituel des accès précoces relèvent du référentiel applicable à ces dispositifs. Leur réutilisation à des fins de recherche, dans le cadre de DANTE, est conduite conformément à la méthodologie de référence MR-004 de la CNIL, dès lors qu'il s'agit d'un traitement de données de santé à finalité scientifique, sans intervention directe auprès des personnes, reposant sur une mission d'intérêt public et sur des garanties appropriées.

Lorsque certaines opérations, en particulier celles relatives à la constitution de la base de référence par la CRO centrale, nécessitent des modalités complémentaires au cadre de référence applicable, elles feront l'objet des formalités ou autorisations requises auprès de la CNIL avant leur mise en œuvre effective. Les traitements mis en œuvre dans le cadre du projet reposent sur un principe de cloisonnement strict.

7.3 Séquence d'inclusion d'un patient à Dante

Le schéma fonctionnel est le suivant :

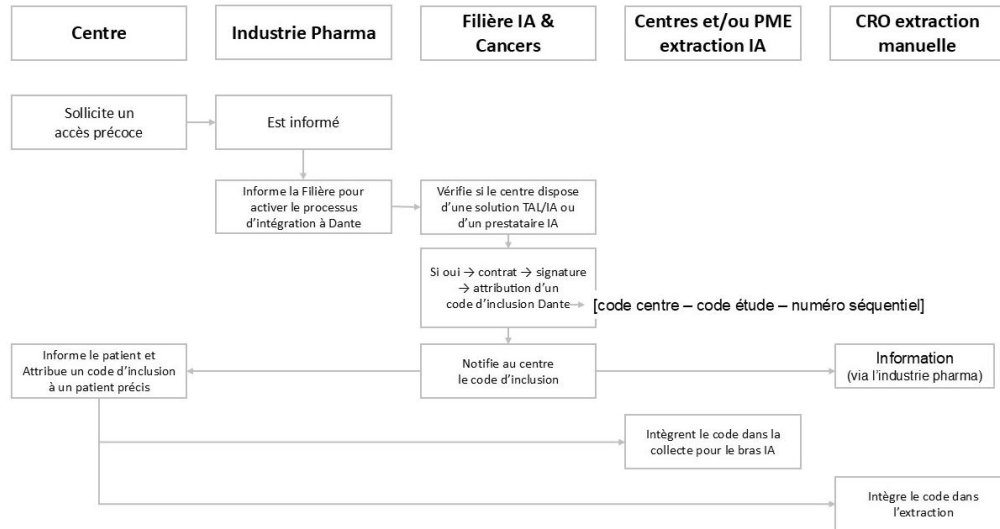


Figure 2 - Séquence d'inclusion (vue par acteurs)

8 Dialogue autour des propositions de ce cahier des charges

Parce que l'attractivité nationale est un des enjeux clef pour la Filière, celle-ci proposera des échanges dédiés afin de présenter le projet DANTE, le set minimal de variables, le modèle économique, la règle de guichet ouvert et les exigences de sécurité.

Le dialogue autour de ce cahier des charges n'a pas pour objet d'organiser une « compétition » entre candidats. Il vise à permettre aux acteurs de comprendre rapidement les attentes et de déposer un dossier complet. Les questions-réponses utiles à tous pourront être partagées publiquement, sous une forme anonymisée lorsque nécessaire.

9 Calendrier prévisionnel et vue d'ensemble de la prestation

Les dates précises seront ajustées selon la publication du cahier des charges, le calendrier des accès précoces participants et les contraintes opérationnelles des centres. La logique générale est la suivante :

9.1 Phase 1 – Publication échanges et ouverture du guichet

- Publication du cahier des charges public.
- Ouverture de l'adresse de dépôt des candidatures : **21 mai**
- Réponses aux questions utiles à tous les candidats via des rendez-vous en one-one, à [cette adresse](#).

9.2 Phase 2 - Instruction au fil de l'eau et contractualisation

- Réception des dossiers complets dans l'ordre d'arrivée.
- Vérification de l'éligibilité et de la capacité opérationnelle.
- Contractualisation des candidats éligibles, dans la limite du budget disponible.
- Attribution ou confirmation des codes d'appariement et des modalités de transmission.

9.3 Phase 3 - Collecte automatisée par IA/NLP/TAL

- Extraction du set de variables à partir des dossiers patients informatisés des patients inclus.
- Pseudonymisation, structuration et transmission sécurisée à la CRO centrale.

9.4 Phase 4 - Constitution de la base de référence et analyse

- Constitution de la base de référence par la CRO centrale, via des ARC indépendants, à partir des DPI. Fusion et chaînage des bases manuelle, IA et référence. Gel de base, analyse statistique et calcul des indicateurs d'exactitude, d'erreur, d'omission et de concordance.
 - Restitution des résultats, valorisation scientifique et bilan des enseignements pour l'écosystème.
-

10 Annexes

10.1 Définitions des acronymes utilisés, par ordre d'apparition

DANTE : Données d'Accès précoces Normalisées extraites par Traitement automatisé du langage en vue de leur Évaluation. Nom du projet faisant l'objet du présent cahier des charges.

NLP/TAL : Natural Language Processing / Traitement Automatisé du Langage. Ensemble de méthodes d'IA permettant d'analyser du texte libre, par exemple des comptes rendus médicaux, afin d'y repérer, extraire et structurer des informations utiles.

ARC : Attaché de Recherche Clinique. Professionnel chargé notamment de collecter, vérifier ou contrôler des données de recherche clinique ou de recherche sur données de santé. Dans DANTE, des ARC indépendants contribueront à constituer la base de référence à partir des dossiers patients informatisés.

HAS : Haute Autorité de santé. Autorité publique indépendante intervenant notamment dans l'évaluation des produits de santé et dans l'encadrement du dispositif d'accès précoce.

ANSM : Agence nationale de sécurité du médicament et des produits de santé. Autorité publique compétente notamment pour la sécurité sanitaire des médicaments et produits de santé, associée à l'encadrement de certains accès précoces.

AMM : Autorisation de Mise sur le Marché. Autorisation permettant la commercialisation d'un médicament dans une indication donnée, après évaluation de sa qualité, de sa sécurité et de son efficacité.

PUT-RD : Protocole d'Utilisation Thérapeutique et de Recueil de Données. Document encadrant l'utilisation d'un médicament en accès précoce et précisant les données à renseigner aux différentes étapes du parcours : demande d'accès, instauration, suivi et arrêt éventuel du traitement.

DPI : Dossier Patient Informatisé. Dossier numérique contenant les informations produites ou utilisées dans le cadre de la prise en charge du patient.

CRO : Contract Research Organization, ou société de recherche contractuelle. Dans DANTE, la CRO centrale est le tiers mandaté par la Filière pour fusionner, chaîner, contrôler et analyser les bases de données, ainsi que pour constituer la base de référence avec des ARC indépendants.

CHU : Centre Hospitalier Universitaire. Établissement public de santé associant activités de soins, d'enseignement et de recherche.

CHRU : Centre Hospitalier Régional Universitaire. Établissement public de santé de référence régionale, exerçant des missions de soins, d'enseignement et de recherche.

CH : Centre Hospitalier. Établissement public de santé assurant des missions de soins hospitaliers.

CLCC : Centre de Lutte Contre le Cancer. Établissement spécialisé dans la prise en charge, la recherche et l'innovation en cancérologie.

PIA : Privacy Impact Assessment. Outil et méthode d'analyse d'impact relative à la protection des données. Dans le présent document, le terme renvoie notamment à l'outil proposé par la CNIL pour documenter les risques et les mesures de maîtrise d'un traitement.

DPO : Data Protection Officer, ou Délégué à la Protection des Données. Personne chargée d'informer, conseiller et contrôler la conformité d'un organisme au RGPD.

MR-004 : Méthodologie de référence n°004 de la CNIL. Cadre de référence applicable à certaines recherches, études ou évaluations n'impliquant pas la personne humaine, portant sur des données de santé et réalisées dans des conditions définies par la CNIL.

HDS : Hébergement de Données de Santé. Certification applicable aux hébergeurs traitant des données de santé à caractère personnel, destinée à garantir un niveau renforcé de sécurité, de disponibilité et de confidentialité.

AIPD : Analyse d'Impact relative à la Protection des Données. Analyse permettant d'identifier les risques d'un traitement pour les droits et libertés des personnes concernées, puis de documenter les mesures prévues pour les maîtriser.

10.2 Set de variables à extraire

Le set ci-dessous reprend, sous forme pédagogique, les variables issues du modèle de PUT-RD de la HAS retenues pour la comparaison dans DANTE. La numérotation ci-dessous suit celle du protocole source.

NB : au-delà de cette présentation pédagogique ; un dictionnaire précis des variable comprenant leur format attendu (texte libre, nombre, valeurs possibles...), les nomenclature exactes des colonnes, etc. ; sera proposé par la CRO Centrale via la Filière aux participants.

Fiche PUT-RD	Catégorie de la fiche	N° et description de la variable	Valeurs attendues
NA	NA	1 Médicament en AP	Liste prédéfinie selon les laboratoires participant
Demande d'accès	Identification du patient	2 Date de naissance (MM/AAAA)	mm/aaaa
		3 Sexe	- Masculin - Féminin
Instauration du traitement	Conditions d'utilisation	4 Date de 1ère administration ou d'instauration du traitement	jj/mm/aaaa
	Posologie prescrite	5 Posologie 1 : dose	Numérique
		6 Posologie 2 : dose	Numérique
		7 Posologie 1 : voie d'administration	- Orale - Intraveineuse - Sous-cutanée
	8 Posologie 2 : voie d'administration	- Orale - Intraveineuse - Sous-cutanée	
Effet(s) indésirable(s) / Situation(s) particulière(s)	9 Y a-t-il eu apparition d'effet(s) indésirable(s) ou une situation particulière ?	- Oui - Non	

Fiche PUT-RD	Catégorie de la fiche	N° et description de la variable	Valeurs attendues
Suivi(s) du traitement (autant d'occurrences que d'administrations du traitement)	Suivi du traitement	10 Date de la visite de suivi	jj/mm/aaaa
		11 Posologie 1 : dose	Numérique
	Posologie prescrite	12 Posologie 2 : dose	Numérique
		14 Posologie 1 : voie d'administration	- Orale - Intraveineuse - Sous-cutanée
		15 Posologie 2 : voie d'administration	- Orale - Intraveineuse - Sous-cutanée
	Interruption/arrêt temporaire	16 Date d'interruption	jj/mm/aaaa
		17 Motif d'interruption	- Progression de la maladie - Effet indésirable - Souhait du patient - Autre (préciser)
		18 Date de reprise	jj/mm/aaaa
	Effet(s) indésirable(s) / Situation(s) particulière(s)	19 Y a-t-il eu apparition d'effet(s) indésirable(s) ou une situation particulière depuis la dernière visite?	Oui Non

Fiche PUT-RD	Catégorie de la fiche	N° et description de la variable	Valeurs attendues
Arrêt définitif du traitement	Arrêt définitif	20 Date de l'arrêt définitif de traitement	jj/mm/aaaa
	Raisons de l'arrêt de traitement	21 Raisons de l'arrêt de traitement	<ul style="list-style-type: none"> - Fin de traitement (définie dans le RCP) - Survenue d'un effet indésirable suspecté d'être lié au traitement - Progression de la maladie - Effet thérapeutique non satisfaisant - Décès - Souhait du patient d'interrompre le traitement - Patient perdu de vue - Ne remplit plus les critères d'éligibilité - Autre (préciser)
	Si décès	22 Date de décès	jj/mm/aaaa
	Si décès	23 Raison du décès	<ul style="list-style-type: none"> - Décès lié à un effet indésirable - Décès lié à la progression de la maladie - Autre raison (préciser)

10.3 Modèle de lettre d'intention et d'acceptation de principe du contrat de partenariat DANTE

Une fois signée, à envoyer s'il vous plaît scannée à l'adresse projet-dante@filiere-ia.fr pour décompte de la date, et en courrier simple à l'adresse ci-dessous

À l'attention de :
Filière Intelligence Artificielle & Cancers
PariSanté Campus
2-10 rue d'Oradour-sur-Glane
75015 Paris

Objet : Lettre d'intention — Candidature au projet DANTE et acceptation de principe du contrat de partenariat

Madame, Monsieur,

Nous soussignés :

[Nom de la structure candidate n°1],
[forme juridique], dont le siège est situé **[adresse]**,
représentée par **[nom, prénom, fonction]**, dûment habilité(e) aux fins des présentes,

et, le cas échéant,

[Nom de la structure candidate n°2],
[forme juridique], dont le siège est situé **[adresse]**,
représentée par **[nom, prénom, fonction]**, dûment habilité(e) aux fins des présentes,

déclarons par la présente notre intention de candidater au projet DANTE — « Données d'Accès précoces Normalisées extraites par Traitement automatisé du langage en vue de leur Évaluation » — dans le cadre du cahier des charges publié par la Filière Intelligence Artificielle & Cancers.

Nous confirmons avoir pris connaissance du cahier des charges du projet DANTE, de ses objectifs scientifiques, opérationnels, réglementaires et financiers, ainsi que du contrat de partenariat, disponible sur le site de la Filière.

Nous confirmons également accepter le principe de signature de ce contrat de partenariat. Nous comprenons que cette acceptation de principe constitue un prérequis de recevabilité opérationnelle de notre dossier de candidature. Elle a pour objet de permettre à la Filière d'instruire rapidement les candidatures dans une logique de recrutement au fil de l'eau, jusqu'à épuisement du budget disponible ou atteinte des objectifs de recrutement du projet.

Nous déclarons être informés que le contrat de partenariat encadre notamment :

- la collecte automatisée des données par IA/NLP/TAL à partir des dossiers patients informatisés ;
- la pseudonymisation des données collectées ;
- la transmission sécurisée des données pseudonymisées au tiers de confiance désigné par la Filière ;
- la limitation stricte des finalités de traitement au projet DANTE ;
- le respect du RGPD et des clauses contractuelles applicables entre responsable de traitement et sous-traitants ;
- les conditions financières applicables à la participation au projet.

Nous confirmons que notre candidature s'inscrit dans l'une des configurations suivantes :

- candidature d'un centre disposant en propre d'une solution d'IA/NLP/TAL opérationnelle ;
- candidature d'un réseau public ou privé ou d'une PME disposant d'une solution d'IA/NLP/TAL en réseau ;
- candidature conjointe d'un centre et d'une PME partenaire technologique ;

Nous comprenons que la présente lettre d'intention ne vaut pas encore sélection définitive ni contractualisation effective. Elle exprime toutefois notre engagement de bonne foi à poursuivre l'instruction de notre candidature et, en cas de validation par la Filière, à signer le contrat de partenariat à compter de l'envoi du lien de signature électronique.

Fait à **[lieu]**, le **[date]**.

Pour **[nom de la structure candidate n°1]**

Nom : **[à compléter]**

Fonction : **[à compléter]**

Signature :

Pour **[nom de la structure candidate n°2, le cas échéant]**

Nom : **[à compléter]**

Fonction : **[à compléter]**

Signature :