



CAHIER DES CLAUSES TECHNIQUES PARTICULIERES (CCTP)

REFONTE DU SYSTEME D'INFORMATION DU CENTRE D'EPIDMIOLOGIE SUR LES CAUSES MEDICALES DE DECES

En date du 02/07/2026

Table des matières

1. PREAMBULE.....	4
1.1. Présentation du CépiDc	4
1.2. Objet du document.....	4
1.3. Contexte du projet.....	5
1.3.1. Un SI actuellement ingouvernable	5
1.3.2. Une dette technique conséquente	6
1.3.3. Travaux de cadrage réalisés au premier semestre 2026	6
1.4. Présentation du projet.....	7
1.4.1. Objectifs pour le CépiDc	7
1.4.2. Les utilisateurs du futur SI	7
1.4.3. Les points d'attention	8
2. DESCRIPTION MACRO DES BESOINS FONCTIONNELS	9
2.1. Une architecture en couches successives	9
2.2. Un modèle de données organisé autour l'événement métier	9
2.3. Un socle technologique homogène hébergé à l'Inserm.....	9
2.4. Séparation du contenu métier et des données échangées.....	10
2.5. Principe d'opérabilité métier.....	10
2.6. Périmètre du SI cible et hypothèses de dimensionnement	10
2.7. Couverture fonctionnelle du progiciel.....	12
3. DESCRIPTION DES PRESTATIONS ATTENDUES	14
3.1. Liste et description des unités d'œuvres utilisées	14
3.2. Exigences méthodologiques	14
Le système d'information objet du présent marché est un système de production statistique reposant sur des traitements intensifs de données (<i>data processing</i>). À ce titre, les candidats devront démontrer leur maîtrise des méthodes, architectures et bonnes pratiques propres à ce type de système d'information.	14
3.3. Prestations de base	15
3.3.1. Vue d'ensemble du projet.....	15
3.3.2. Phase 1 - Fondations du système d'information.....	15
3.3.3. Phase 2 - Traitements métier et mise en qualité	16
3.3.4. Phase 3 – Figeage, diffusion et mise en production	18
3.4. Prestations de fonctionnalités supplémentaires à la demande.....	19
3.4.1. Reprise de l'historique des données	19
3.4.2. Intégration de données tierces	20

3.4.3.	Tierce maintenance applicative	20
3.4.4.	Conduite du changement et accompagnement à la transformation organisationnelle.....	20
3.5.	Planning prévisionnel	21
4.	ORGANISATION ET GOUVERNANCE DU PROJET	22
4.1.	Composition de l'équipe projet	22
4.2.	Présence sur site.....	22
4.3.	Gouvernance du projet.....	23
4.3.1.	Comité de pilotage (COFIL)	23
4.3.2.	Comité projet (COPROJ)	23
4.3.3.	Ateliers de travail	23
4.4.	Reporting et indicateurs	23
4.5.	Gestion documentaire	24
5.	LIVRABLES ATTENDUS	25
5.1.	Livrables de conception fonctionnelle	25
5.2.	Livrables de conception technique.....	25
5.3.	Livrables relatifs aux traitements statistiques et aux flux de données	25
5.4.	Livrables de configuration du progiciel	26
5.5.	Livrables de tests et de recette	26
5.6.	Livrables de sécurité et de conformité	26
5.7.	Livrables de déploiement	27
5.8.	Livrables d'exploitation et de maintenance	27
5.9.	Livrables de formation et d'accompagnement	27
5.10.	Livrables relatifs à la gouvernance des données	28
5.11.	Réversibilité.....	28
5.12.	Livrables par palier	28
6.	ANNEXES.....	29
6.1.	Documentation sur l'existant	29
6.2.	Travaux de cadrage réalisés au S1 2026.....	29
6.3.	Critères d'évaluation	29
6.4.	Cadre de réponse technique	30

1. Préambule

1.1. Présentation du CépiDc

Le CépiDc, unité de service de l'Inserm, est l'autorité nationale responsable de la production et de la diffusion des statistiques sur les causes de décès en France construites à partir des informations renseignées par les médecins et infirmières certificateurs sur les volets médicaux des certificats de décès. À ce titre, il collecte, code, consolide et diffuse chaque année les données de mortalité qui alimentent en temps réel la veille sanitaire (Santé publique France), mais aussi le système statistique européen (Eurostat), le Système National des Données de Santé (SNDS), l'Organisation Mondiale de la Santé (OMS) et la recherche épidémiologique par la mise à disposition annuelle d'une base de données statistique. La production de ces données est encadrée par le règlement européen n°1338/2008, labellisée par l'Autorité de la Statistique Publique et soumise au Code des bonnes pratiques de la statistique européenne (règlement UE n°223/2009).

Le CépiDc est une structure d'une vingtaine de personnes dont l'activité se résume schématiquement de la façon suivante :

- Accueil et mise à disposition des informations pour alimenter la veille sanitaire
- Production d'une base de données annuelle (intégration, contrôle, traitement des volets médicaux des certificats de décès, mise en cohérence avec les décès déclarés à l'Insee...)
- Analyses statistiques de données
- Diffusion de résultats (*via* des publications, vers des partenaires réguliers, en open-data, au SNDS...).

1.2. Objet du document

Le présent appel d'offre a pour objet la refonte du système d'information (SI) du CépiDc pour la production des données relatives aux causes de décès.

Le système d'information actuel du CépiDc n'est en effet plus suffisamment adapté au processus métier. Le CépiDc a besoin d'un système d'information robuste et optimisé avec les outils et technologies actuels, tout en respectant la sobriété (sans développement d'outils démesurés pour la tâche demandée). Ce nouveau SI devra être maintenable, évolutif avec une réelle modularité, apte à supporter la charge tout en garantissant un haut niveau de sécurité et de disponibilité. Cette refonte a aussi pour objectif d'optimiser la gestion des données au sein du système d'information CépiDc et d'améliorer l'industrialisation et l'automatisation des chaînes de traitement. Elle doit s'inscrire dans l'approche habituelle d'une chaîne de production statistique comme ce qui est réalisé pour des traitements homologues de statistique publique.

Les données étant uniques, sensibles (les causes médicales de décès sont des données de santé au sens du RGPD) et couvertes par le secret professionnel, leur traitement doit garantir l'intégrité des informations et leur inviolabilité. Par ailleurs, l'organisation du SI devra être conforme au schéma directeur de l'Inserm.

Les prestations attendues sont décrites dans les différentes parties du document. Les principaux objectifs sont d'obtenir un :

- système d'information restructuré et rationalisé : le modèle de données sera revu, les technologies seront unifiées et les traitements seront rationalisés ;

- système d'information orchestrable : l'intégration et l'orchestration des pipelines de données sera prise en charge par le SI ;
- système d'information documenté, maintenable et évolutif : l'évolutivité et la paramétrabilité du système en faciliteront la maintenance ;
- dispositif qui doit rendre possible, au moins pour la majorité des tâches, le télétravail.

Les prestations décrites dans le présent appel d'offre seront soumises à une obligation de résultat.

1.3. Contexte du projet

Afin de produire la base médicale des causes de décès, le système d'information du CépiDc permet :

- la réception de flux de données externes, essentiellement
 - o le contenu des certificats médicaux (texte libre et variables statistiques), et les images correspondantes, avec des flux différents selon l'origine (électronique ou papier) du certificat de décès,
 - o des flux externes provenant de l'Insee pour mise en cohérence avec les décès déclarés à l'état civil,
- l'intégration, le contrôle, et la correction des données par une IHM spécifique,
- des traitements individuels par des agents grâce à des applications spécifiques (IHM qui permet aux agents spécialisés de coder les certificats, logiciel métier international à conserver),
- des traitements de masse par batchs automatisés,
- des traitements de masse par batchs spécifiques (définis et opérés par le métier), modèles d'IA.

La mise à disposition de données au fil de l'eau se fait par flux de données automatisés vers des partenaires réguliers, notamment Santé Publique France. Le SI doit également permettre la mise à disposition ponctuelle durant la phase de production de bases de données *ad hoc*.

A l'issue du processus de production, des traitements statistiques sont réalisés pour constituer les produits de diffusion dont une diffusion de statistiques fines en open-data. Le système d'information de production est étroitement articulé avec le système d'information de diffusion des résultats, avec lequel il entretient de fortes interactions fonctionnelles et techniques. Les équipes en charge du SI de diffusion sont ainsi des parties prenantes du SI de production et participent à la définition des interfaces et des données mises à disposition. Toutefois, la refonte du système de diffusion, des traitements spécifiques associés à la diffusion des résultats ainsi que des outils de valorisation et de publication des données ne fait pas partie du présent marché, qui se limite au périmètre de la production de la base médicale des causes de décès.

1.3.1. Un SI actuellement ingouvernable

Le système d'information de production du CépiDc s'est construit progressivement depuis plus de vingt ans, par accumulation de briques techniques répondant à des besoins successifs. Ainsi, il est aujourd'hui composé de 4 bases de données, plus de 1 000 variables, 25 tables au sein desquelles 4 identifiants coexistent et une centaine de traitements sont réalisés sans que leur enchaînement ne soit optimisé. Cette sédimentation a produit un SI dont la complexité est aujourd'hui un frein majeur à l'évolution du dispositif.

D'autre part, le processus de production des données a évolué (au niveau de la collecte par l'utilisation d'une plateforme électronique de certification de décès ; au niveau du traitement avec l'intégration

de nouvelles méthodes de codage des causes de décès impliquant des réseaux de neurones ; au niveau de l'organisation et du pilotage...).

Le SI de production s'est adapté, au fil de l'eau, à ces différentes évolutions structurantes. Il doit désormais être repensé dans sa globalité afin de le rendre harmonisé, compréhensible et gouvernable.

1.3.2. Une dette technique conséquente

Il existe seulement treize flux fonctionnels vers et depuis le SI du CépiDc mais le système d'information s'appuie sur une multitude de mécanismes techniques accumulés au fil du temps : traitements batch, macros et applications périphériques se sont superposés jusqu'à former près d'une centaine de chaînes de traitement interdépendantes. L'enjeu principal de la refonte est précisément de supprimer cette complexité héritée en consolidant l'ensemble dans quelques pipelines unifiés, cohérents et facilement exploitables.

1.3.3. Travaux de cadrage réalisés au premier semestre 2026

Afin de préparer la présente consultation, le CépiDc a mené, au cours du premier semestre 2026, une phase de cadrage du projet de refonte du système d'information de production.

Cette phase a permis de définir les principes directeurs, les orientations fonctionnelles et techniques, ainsi que la trajectoire cible du futur système d'information. Elle a également conduit à l'élaboration de plusieurs livrables portant notamment sur :

- la gouvernance des données ;
- les exigences fonctionnelles et techniques ;
- la cartographie des flux et des interfaces ;
- les hypothèses de volumétrie et de dimensionnement ;
- les principes d'infrastructure, de sécurité et d'exploitation ;
- la trajectoire de migration et les modalités de reprise de l'historique.

Ces documents sont fournis en annexe du présent CCTP.

La prestation de cadrage général du projet ayant déjà été réalisée, il n'est pas attendu des candidats qu'ils réalisent ou chiffrent une nouvelle phase de cadrage de même nature. Les offres devront s'appuyer sur les éléments fournis afin de proposer les prestations de conception détaillée, de réalisation, de migration, de recette, de déploiement et d'accompagnement au changement attendues dans le cadre du présent marché.

Le titulaire retenu devra néanmoins prévoir une phase d'appropriation et de consolidation des travaux existants afin de préparer la conception détaillée de la solution. Cette phase ne devra pas conduire à remettre en cause les orientations structurantes validées lors du cadrage, sauf accord préalable.

1.4. Présentation du projet

1.4.1. Objectifs pour le CépiDc

Le CépiDc souhaite mettre en place une plateforme de gestion et d'orchestration des données permettant d'améliorer la performance globale de l'organisation, d'optimiser les processus métiers, de renforcer la fiabilité des reportings et d'accélérer la prise de décision à tous les niveaux.

Le CépiDc a besoin d'un système d'information industrialisé, supprimant les traitements manuels actuellement effectués. Pour ce faire, une solution de type progiciel devra être proposée.

Elle devra permettre :

- d'automatiser l'intégration et la consolidation des données provenant de systèmes hétérogènes,
- de garantir la qualité, la cohérence et la traçabilité des données de bout en bout,
- de réduire les lancements de traitements et les corrections manuels, et les risques d'erreurs liés aux échanges de fichiers.
- de permettre la mise en œuvre de règles explicites de gouvernance des données et de contrôles qualité automatisés.
- d'offrir une plateforme évolutive permettant au service de traiter les modifications d'ampleur modérée sans faire appel à une maintenance (évolution de modalités de variables, changement de nomenclature, modification des règles de traitement, etc.),
- de sécuriser les flux de données et garantir leur disponibilité pour l'ensemble des utilisateurs autorisés.
- de mettre à disposition des indicateurs paramétrables, actualisés et partagés nativement au sein de l'outil afin d'en faciliter le pilotage à la fois stratégique et opérationnel.
- de faciliter la prise de décision grâce à des données consolidées, fiables et accessibles en temps réel.

1.4.2. Les utilisateurs du futur SI

Le système d'information du CépiDc a plusieurs types d'utilisateurs avec des besoins différents.

- Le pilote de campagne annuelle : il est responsable de l'organisation, du suivi et de la coordination de l'ensemble des opérations permettant de produire les données statistiques. Il garantit que les données sont produites dans les délais, avec le niveau de qualité attendu, et qu'elles sont exploitables. Les indicateurs de suivi des opérations seront disponibles dans le futur SI et consultable par le pilote de campagne.
- Les responsables techniques du codage des causes de décès (codage par intelligence artificielle, codage manuel, codage par système expert) : le responsable du codage IA est en charge de l'entraînement des modèles, de la production d'un nouveau modèle, du versionning des modèles et participe à la validation de ces modèles. Le futur SI devra permettre la supervision des travaux sur le codage par IA. Le responsable du codage manuel pourra suivre grâce au nouveau SI l'avancée du codage manuel et consulter les indicateurs de codage. Le responsable du codage par système expert Iris pourra suivre les indicateurs de codage *via* ce système.

- Le responsable de la diffusion : il pilote la mise à disposition des données et des résultats statistiques auprès des utilisateurs. Il garantit la qualité, la cohérence (annuelle et pluri-annuelle) et la conformité des publications aux règles de confidentialité et de protection des données. Il coordonne les calendriers de diffusion et les publications. Il contribue à la valorisation des données et veille à rendre les informations statistiques accessibles, compréhensibles et exploitables par les différents publics.
- L'administrateur d'application : il assure le bon fonctionnement des applications utilisées pour la collecte et le traitement des données statistiques. Il gère les paramétrages, les droits d'accès des utilisateurs et le suivi des incidents et alerte le cas échéant. Il veille à la disponibilité et à la performance du SI. Il coordonne les mises à jour et les évolutions techniques avec les équipes métiers et informatiques. Enfin, il apporte un support aux utilisateurs et contribue à l'amélioration continue des outils de production.

1.4.3. Les points d'attention

Le futur système d'informations devra résoudre la majorité des difficultés rencontrées à l'utilisation du SI actuel du CépiDc. En particulier :

- Les échanges entre applications ou SI tiers devront reposer sur des contrats d'interface clairement définis et normalisés.
- Chaque enregistrement devra être associé à un identifiant unique garantissant son unicité et sa cohérence dans l'ensemble du système.
- Les informations administratives et les données médicales seront dissociées afin de mieux structurer les traitements et de réduire les temps de traitement.
- Un défaut de qualité sur une donnée ne devra plus entraîner le blocage des flux, permettant ainsi la continuité des traitements tout en isolant les anomalies.
- Le codage par système-expert Iris sera maintenu de manière stable, i le futur SI doit prévoir l'adaptation des traitements à ce système. L'ensemble des opérations et transformations devra faire l'objet d'une traçabilité afin de garantir le suivi et l'auditabilité des données.
- Les référentiels (géographiques, de transcodage, de nomenclatures...) devront être gérés par version et systématiquement associés à une campagne donnée pour assurer la cohérence et la comparabilité des traitements dans le temps.

2. Description macro des besoins fonctionnels

2.1. Une architecture en couches successives

Le système d'information cible s'appuiera sur une chaîne de traitement centrale unique, découpée en trois espaces successifs, chacun disposant d'une responsabilité précise et de règles de passage de l'un à l'autre explicitement définies.

- La première couche, dédiée à la réception, conservera la totalité des données d'origine en entrée du SI, sans modification ni suppression. Elle constituera la zone de référence du SI, permettant de rejouer l'ensemble des traitements à tout moment en cas de besoin.
- La deuxième couche correspondra à l'espace de transformation et de valorisation des données : contrôles qualité, enrichissements, production du codage et gestion de l'exhaustivité en mobilisant les données de l'Insee y sont réalisés.
- La troisième couche sera consacrée à la mise à disposition des données consolidées. Elle hébergera les bases statistiques dans leur état final, accessibles en lecture seule et destinées aux différents utilisateurs tels que la chaîne de diffusion qui alimentera le SNDS, Eurostat etc.

Le franchissement d'une couche à l'autre sera daté et auditable rendant la traçabilité consultable et exploitable par les équipes.

2.2. Un modèle de données organisé autour l'événement métier

Le modèle actuel est structuré autour du certificat de décès qui constitue l'objet principal de collecte. Le modèle cible devra adopter une logique différente : le certificat deviendra un support de collecte, tandis que les éléments centraux du système seront désormais l'identité de la personne décédée et l'événement de décès lui-même.

Cette évolution permettra notamment de supprimer la séparation historique entre les circuits papier et électronique qui traverse aujourd'hui l'ensemble du modèle. Par ailleurs, les multiples identifiants actuellement associés au certificat seront remplacés par un identifiant unique et pérenne, généré directement par le SI. Cet identifiant restera stable durant tout le cycle de vie de la donnée au sein du SI. Les anciens identifiants seront conservés comme informations de rapprochement afin de garantir la traçabilité historique.

2.3. Un socle technologique homogène hébergé à l'Inserm

Le CépiDc a besoin d'un système d'information industrialisé, supprimant les traitements manuels actuellement effectués et simplifiant le pilotage du système d'information. L'orientation retenue repose sur un progiciel intégrant l'ensemble des fonctions techniques du SI : intégration des flux, orchestration des traitements, supervision, gestion de la qualité, traçabilité et administration des référentiels et des campagnes (traitement des décès par année civile).

Le progiciel proposé doit être déjà installé sur le marché, la solution doit être éprouvée avec de solides références. Ce progiciel doit permettre de centraliser, fiabiliser et exploiter efficacement les données issues de multiples sources. La solution proposée devra intégrer nativement, sans développements spécifiques, les fonctions suivantes : orchestration des traitements et gestion des dépendances entre pipelines, gestion versionnée des référentiels (nomenclatures, règles de contrôle, tables de

transcodage), traçabilité des opérations et auditabilité des transformations, tableau de bord de pilotage paramétrable par un administrateur métier. Ces fonctions doivent être accessibles et configurables par des profils non-développeurs, sans intervention du prestataire. Elle devra aussi offrir une administration simplifiée et évolutive afin de réduire les coûts d'exploitation et la dépendance aux ressources techniques. Les traitements métier aujourd'hui dispersés sur des postes individuels pourront être versionnés, centralisés au sein de cette plateforme et exécutés par différents utilisateurs dans un environnement maîtrisé et sécurisé.

L'hébergement est actuellement prévu sur le datacenter virtualisé de l'Inserm. Mais la solution devra être portable vers d'autres environnements (cloud par exemple) de manière à garantir sa pérennité.

L'authentification des utilisateurs repose sur l'annuaire national Inserm *via* un mécanisme d'authentification unique.

Enfin, le moteur de codage IRIS, placé sous gouvernance internationale, est maintenu sans modification : l'architecture du SI est conçue pour s'intégrer autour d'IRIS et non pour en transformer le fonctionnement. Les entrées/sorties d'IRIS dans le SI sont normalisées et actuellement versionnées en lot. Chaque lot soumis sera tracé avec la version du moteur utilisée et les paramètres d'exécution. L'interface devra être conçue pour absorber les nouvelles versions du moteur sans rupture des formats d'échange, et permettre la comparaison de résultats de codage entre versions sur un même lot.

2.4. Séparation du contenu métier et des données échangées

Le cœur du SI cible produit et consomme un format unique, indépendamment du protocole utilisé pour échanger avec chaque partenaire.

Chaque connecteur partenaire est responsable de la conversion entre ce format interne de référence et les protocoles spécifiques d'échange. Chaque interface (entrante ou sortante) fait ainsi l'objet d'un contrat d'interface formalisé.

2.5. Principe d'opérabilité métier

Le système d'information cible devra être conçu pour être majoritairement opéré par les équipes métier du CépiDc. À ce titre, les utilisateurs habilités devront pouvoir assurer de manière autonome les activités de supervision fonctionnelle, de pilotage des traitements, de lancement et reprise des traitements, d'analyse des anomalies, de consultation de la traçabilité, de suivi de la qualité des données, de gestion des référentiels métier et de contrôle des flux, sans recours systématique à une expertise technique ou à l'intervention du titulaire. La solution devra privilégier les mécanismes de paramétrage, les interfaces de pilotage et les tableaux de bord accessibles aux utilisateurs métier, afin de limiter les dépendances vis-à-vis des équipes techniques et de favoriser l'autonomie opérationnelle du CépiDc.

2.6. Périmètre du SI cible et hypothèses de dimensionnement

Le système d'information cible a vocation à couvrir l'ensemble du processus de production statistique du CépiDc, depuis la réception des certificats de décès et des données partenaires jusqu'à la constitution de la base statistique figée et historisée. Il devra constituer une plateforme unique permettant l'intégration, le contrôle, l'enrichissement, le codage, la consolidation, et le pilotage des

données de mortalité, tout en assurant une traçabilité complète des traitements et une gouvernance unifiée de la donnée.

Par contre, le périmètre fonctionnel du socle ici décrit ne couvre pas :

- Le codage par système expert Iris qui fait l'objet d'une gestion et d'une production à part – hors refonte. Le futur SI doit prévoir l'adaptation des traitements et des référentiels à ce système ;
- Les moteurs des modèles d'intelligence artificielle : seuls les bases d'apprentissage et les résultats prédits par ces modèles sont intégrés (et versionnés) dans le périmètre ;
- Le SI diffusion qui s'adaptera à l'évolution du SI de production.

Pour le dimensionnement de la solution, les candidats devront prendre en compte les hypothèses suivantes :

Indicateur	Valeur de référence
Décès traités annuellement	Environ 650 000 (effectif appelé à augmenter sensiblement avec l'arrivée à des âges de mortalité élevée des générations du babyboom)
Certificats intégrés quotidiennement	Environ 1 800
Utilisateurs nominatifs	Environ 30
Utilisateurs simultanés en période de pointe	Environ 10
Nombre de contrôles	Environ 50
Flux critiques fonctionnant en continu	24h/24 et 7j/7 <u>Durée d'indisponibilité tolérée : 4h</u>
Synchronisation annuelle Insee	Actuellement traitement lourd de plusieurs heures
Nombre actuel de traitements techniques	Environ 100
Nombre de campagnes gérées simultanément	Minimum 3 campagnes en parallèle

Le titulaire devra proposer une architecture capable d'absorber ces volumes tout en intégrant :

- la gestion de plusieurs référentiels versionnés et millésimés ;
- la conservation des métadonnées de traçabilité et des journaux d'exécution ;
- l'exécution simultanée des traitements d'intégration, de contrôle, de codage, de synchronisation, d'anonymisation et de diffusion ;
- l'intégration future de nouveaux flux, partenaires, référentiels ou traitements sans remise en cause de l'architecture cible.

Les hypothèses de dimensionnement proposées par les candidats devront couvrir les besoins de stockage, de calcul, de reprise après incident, d'archivage, de montée en charge et de disponibilité. Les candidats devront expliciter leurs hypothèses de volumétrie, les marges de croissance retenues ainsi que les mécanismes permettant d'assurer les niveaux de service attendus par le CépiDc.

2.7. Couverture fonctionnelle du progiciel

Le progiciel attendu est de nature self serve data processing. Le futur S.I. sera opéré directement par le métier. La nature des fonctions attendues du progiciel sont les suivantes :

Data production

- Fonctions d'orchestration permettant de définir des chaînes de traitement (pipelines) avec séquençement, scheduling, gestion des dépendances entre étapes, reprise sur erreur et rejeu sélectif d'une étape sans retraitement complet. Ce moteur est configurable par un administrateur fonctionnel sans écriture de code.
- Les nœuds d'exécution sont technologiquement neutres. Le progiciel peut piloter des traitements écrits en Python, R, SQL ou tout autre langage, exécutés sur des environnements distincts. Le soumissionnaire précise pour chaque environnement si la prise en charge est native ou nécessite un développement d'adaptateur. Des bibliothèques de traitements prêtes à l'emploi peuvent être un plus (support de fonctions standards de data processing. NB : dans le cadre du présent appel d'offre, il n'y a pas besoin de faire appel à des fonctions avancées de type data science).
- Le progiciel permet le lancement planifié, manuel ou conditionnel (déclenchement sur événement ou sur disponibilité d'une donnée) des traitements. Les règles de déclenchement sont paramétrables sans code.
- Le progiciel dispose de fonctions de supervision : vue consolidée en temps réel de l'état de chaque traitement actif (en cours, en attente, en erreur, terminé) avec identification du composant défaillant et accès aux journaux d'exécution depuis l'interface, génération d'alertes automatiques paramétrables par type d'anomalie, par seuil (ex. : taux de rejet > x%, délai de traitement > n heures) et par destinataire.
- Dans l'idéal il offre au métier un cockpit permettant le suivi simultané des performances, de la qualité, des statuts d'exécution des traitements, des anomalies et suivant les différents modes de traitement (automatisé, manuel) sur une même interface.

Data management

- Le progiciel intègre un module d'administration des référentiels (tables de codes, nomenclatures, règles de transcodage) permettant le chargement, le versionnage automatique et l'activation d'une version via une interface IHM dédiée, sans intervention sur la base de données ni développement. Chaque chargement validé d'une nouvelle référence est tracé (métadonnées – horodatage, version, statut...) constituant la version de référence. Les traitements consomment la version de référence au moment de leur exécution. La version utilisée est tracée dans les résultats produits, permettant de reconstituer rétrospectivement l'ensemble des éléments ayant contribué à un traitement.
- Le progiciel permet la gestion de règles de qualité de données, versionnées et paramétrables en termes de mode passant ou non, activation/désactivation de règles, ajout, modification des seuils. Le résultat de l'exécution des règles qualités est directement visualisable par traitement exécuté (le traitement a réussi, a échoué, à lever un problème qualité mais a continué).
- Le progiciel permet la visualisation du data lineage d'un jeu de données.

Data intégration et stockage

- Le progiciel permet via une logique de connecteurs de s'intégrer avec des sources externes de données (NB : dans le cas du présent appel d'offre, les sources sont de nature fichiers structurés, à plat, XML, excel, CSV voire API – pas de progiciel spécifique à intégrer).
- Le progiciel s'interface avec les bases de données standards du marché. L'open source est privilégié.
- Le progiciel permet des fonctions de data catalogage.

Data publication

- Le progiciel dispose d'un portail de mise à disposition des données permettant aux utilisateurs habilités de rechercher, filtrer et télécharger les jeux de données produits dans plusieurs formats (au minimum tableur et CSV), sans intervention technique. L'accès est sécurisé selon les habilitations.
- Chaque jeu de données publié dans le portail est identifiable, par l'ajout de métadonnées : producteur, date de production, millésime, niveau de contrôle qualité appliqué et millésime des référentiels utilisés.

Self data

- Rappel des fonctions self (exécution par le métier). L'ensemble des opérations courantes du progiciel (lancement de campagnes de traitements, gestion des référentiels, paramétrage des règles de contrôle - qualité, supervision des traitements, publication dans le portail) peut être réalisé par des profils métier via l'interface, sans accès au code source ni intervention d'une maîtrise d'œuvre.

Data sécurité

- Le progiciel s'intègre à un annuaire d'entreprise (LDAP/AD) pour la gestion centralisée des identités et des habilitations. Les droits d'accès sont administrables par périmètre fonctionnel (pipeline, ensemble de traitements de données, référentiel, données et jeu de données).
- Le progiciel doit permettre la mise en œuvre des règles de confidentialité relatives au RGPD (gestion de données sensibles).

Data framework

- L'architecture du progiciel doit s'inscrire dans une logique de type framework (structure d'accueil), facilitant la construction du futur S.I. et son évolution dans le temps.
- Le progiciel doit faciliter : la prise en compte de nouvelles sources de données, le déploiement et l'évolution de pipelines de données – ajout-évolution de traitements de données, l'ajout et l'évolution de règles de qualité de données, l'exposition de données à de nouveaux consommateurs, l'interfaçage avec des fonctions de pilotage avancée (type business intelligence) – pour faciliter la production de macro-contrôles.
- Le tout dans un environnement tracé (métadonnées automatiquement produites).

3. Description des prestations attendues

3.1. Liste et description des unités d'œuvres utilisées

Le présent marché s'inscrit dans une démarche de pilotage par la valeur visant à maximiser les bénéfices métier apportés par la solution tout en respectant l'enveloppe budgétaire allouée au projet. Le soumissionnaire devra néanmoins décomposer dans sa réponse le coût total de chacune des phases du projet en sous-items fonctionnel cohérents.

Les prestations supplémentaires à la demande sont optionnelles et font l'objet d'un chiffrage distinct. Elles ne seront activées que par décision formelle du comité de pilotage du projet.

Afin de garantir la transparence des coûts et de faciliter les arbitrages éventuels en cours d'exécution, le soumissionnaire devra présenter dans son offre une décomposition détaillée des prestations sous la forme d'unités d'œuvre cohérentes correspondant à des fonctionnalités, processus métier, interfaces, flux, composants techniques, ...

Pour chaque unité d'œuvre, le soumissionnaire précisera *a minima* :

- le périmètre couvert ;
- les hypothèses de réalisation retenues ;
- les charges estimées par profil ;
- les dépendances éventuelles ;
- les livrables associés ;
- les critères de validation ;
- la contribution de l'unité d'œuvre aux objectifs métier du projet.

Le soumissionnaire identifiera notamment les éléments indispensables à la mise en service et à l'atteinte des objectifs fondamentaux du projet, les éléments pouvant faire l'objet d'un déploiement progressif ou différé, les options ou variantes susceptibles d'être mobilisées en fonction des contraintes budgétaires ou des priorités métier, les conséquences de ces arbitrages sur les fonctionnalités, les performances ou les délais de réalisation. La décomposition proposée devra permettre au CépiDc de disposer, tout au long du projet, d'une visibilité suffisante sur la répartition des coûts, les charges associées et la valeur attendue des différents incréments, afin de faciliter les arbitrages et les réévaluations de priorités dans le respect de l'enveloppe budgétaire du projet.

3.2. Exigences méthodologiques

Le système d'information objet du présent marché est un système de production statistique reposant sur des traitements intensifs de données (*data processing*). À ce titre, les candidats devront démontrer leur maîtrise des méthodes, architectures et bonnes pratiques propres à ce type de système d'information.

Dans leur offre, les soumissionnaires décriront la méthodologie qu'ils proposent pour concevoir, mettre en œuvre et maintenir un système de production de données répondant aux exigences de robustesse, de qualité, de traçabilité et d'évolutivité attendues par le CépiDc.

Ils présenteront notamment les compétences et retours d'expérience des profils mobilisés sur des projets comparables, ainsi que les méthodes, standards et outils qu'ils mettent en œuvre pour garantir la qualité de réalisation.

La méthodologie proposée devra notamment couvrir les aspects suivants :

- la conception de pipelines de données robustes, idempotents, traçables et réutilisables, s'appuyant sur des standards de développement permettant de limiter l'hétérogénéité des traitements et de faciliter leur maintenance et leur évolution ;
- la conception et l'orchestration des traitements, notamment la gestion des dépendances, des reprises sur incident, des relances et de la supervision des exécutions ;
- la définition d'une architecture de données adaptée aux besoins du CépiDc (notamment architecture de type médaillon ou équivalente), ainsi que les principes de modélisation des données retenus ;
- la conception des interfaces d'intégration et de diffusion des données, assurant la réception des flux entrants, leur traitement et la mise à disposition des données auprès des systèmes consommateurs ;
- l'implémentation des fonctions de gestion des données (*data management*), incluant notamment la qualité des données, la gestion des référentiels, la gestion des métadonnées, la traçabilité (*data lineage*), la gestion des habilitations et la gouvernance des données ;
- la conception et la mise en œuvre des mécanismes de migration, de reprise et de transformation des données.

Les candidats expliciteront la manière dont ces principes méthodologiques seront appliqués dans le cadre du projet afin de produire une base statistique fiable, pérenne, maintenable et évolutive, répondant aux exigences fonctionnelles et techniques du présent CCTP.

3.3. Prestations de base

3.3.1. Vue d'ensemble du projet

La réalisation du projet est organisée en trois phases successives et complémentaires, correspondant aux trois couches définies *supra*. Chaque phase apporte un élargissement maîtrisé du périmètre fonctionnel, livrant de la valeur de façon incrémentale et permettant au CépiDc de procéder à des recettes à l'issue de chacune d'elles.

Phase 1 — Fondations : mise en place de la plateforme technique et des mécanismes d'accueil et de contrôle de premier niveau des données.

Phase 2 — Traitements métier : mise en qualité, enrichissement, codage des données et alimentation en continu de la veille sanitaire.

Phase 3 — Diffusion et mise en production : figeage des données, mise à disposition des produits aux destinataires, et décommissionnement du SI actuel.

3.3.2. Phase 1 - Fondations du système d'information

3.3.2.1. Objectifs

La première phase pose les bases techniques et conceptuelles du nouveau SI. Elle poursuit deux objectifs simultanés, traités en parallèle : d'une part, la conception détaillée du SI cible ; d'autre part, la mise en œuvre du socle technique permettant l'accueil et le contrôle de premier niveau des données.

À l'issue de cette phase, le SI sera en mesure de réceptionner et d'intégrer en flux continu les certificats de décès papier et électroniques. Les données reçues pourront être qualifiées au fil de l'eau, sans que les contrôles de qualification ne soient bloquants pour le processus d'intégration.

3.3.2.2. Axe 1 - Conception détaillée

La phase de conception détaillée a pour finalité la production du modèle physique de données opérationnel. Elle doit être conduite indépendamment de la mise à disposition de l'infrastructure. Les livrables attendus sont les suivants :

- Modèle physique de données complet, couvrant l'ensemble des entités du SI cible.
- Mapping détaillé variable par variable entre le SI actuel et le nouveau SI, permettant une traçabilité complète des transformations de données.
- Spécification exhaustive des règles de contrôle de qualification, reprenant et étendant les règles existantes pour garantir une couverture complète.
- Formalisation des règles de pilotage des processus, couvrant la supervision, le reporting, la traçabilité et le recueil des données.
- Élaboration des contrats d'interface avec l'ensemble des partenaires émetteurs et destinataires de données.

3.3.2.3. Axe 2 - Réalisation technique

La phase de réalisation technique consiste à déployer la plateforme et à industrialiser les fonctions d'intégration de données. Elle comprend les travaux suivants :

- Mise à disposition, installation et paramétrage du progiciel sur l'infrastructure informatique du CépiDc, dans trois environnements distincts : développement, recette et production.
- Création du modèle physique de données dans le SI cible.
- Mise en place de la chaîne d'intégration des données sources, avec contrôle de premier niveau à réception.
- Anonymisation des données collectées sous forme de texte libre.
- Chargement et gestion des référentiels annuels (millésimés) via un module dédié.

3.3.2.4. Résultats attendus en fin de phase 1

À l'issue de cette première phase :

- Les certificats de décès sont intégrés en continu dans la zone de réception du nouveau SI.
- Un identifiant unique est attribué à chaque certificat dès sa réception.
- Les contrôles de qualification sont opérationnels et produisent automatiquement des rapports (signalements qualifiants et bloquants).
- La traçabilité complète des données est assurée depuis leur réception.
- Les référentiels millésimés sont chargés et actifs.

3.3.3. Phase 2 - Traitements métier et mise en qualité

3.3.3.1. Objectifs

La deuxième phase constitue le cœur fonctionnel du SI. Elle regroupe l'essentiel des traitements métier réalisés sur les données par le CépiDc, organisés autour de cinq piliers : la correction des données administratives, le codage des causes médicales de décès, la synchronisation avec les données de l'Insee, la transmission à la veille sanitaire, et l'orchestration globale de la chaîne de production. À l'issue de cette phase, la chaîne de traitement sera intégralement opérationnelle et pourra être pilotée de manière centralisée.

3.3.3.2. Pilier 1 - Correction des données administratives

Les certificats de décès comportent deux catégories de données : des données médicales et des données administratives. Jusqu'à présent traitées conjointement dans le SI actuel, ces deux dimensions doivent être dissociées dans la nouvelle architecture afin que l'avancement global du traitement ne soit pas conditionné par la levée de contrôles portant sur les données administratives. Un circuit de traitement spécifique devra être mis en place pour les données administratives. Ce circuit intégrera les règles de contrôle existantes, enrichies si nécessaire. L'IHM actuellement utilisée pour contrôler et corriger ces données administratives sera évaluée et, le cas échéant, remplacée par une solution plus adaptée.

3.3.3.3. Pilier 2 - Codage des causes médicales de décès

Le système de codage des causes médicales de décès mis en œuvre par le CépiDc repose sur trois modes complémentaires :

- Le codage automatique par traitement batch, réalisé *via* le logiciel métier international IRIS,
- Le codage assisté par intelligence artificielle (IA),
- Le codage manuel, lui aussi assisté par IRIS *via* une interface dédiée.

Le nouveau SI devra s'interfacer avec IRIS et avec les modèles d'IA, ces deux composantes restant inchangées. Il devra permettre de stocker les données produites en sortie de codage quel que soit le mode de codage dont il est issu et, dans la mesure du possible, de conserver les journaux d'exécution (logs) des modèles d'IA à des fins de traçabilité et d'audit.

3.3.3.4. Pilier 3 - Synchronisation avec les données Insee

La synchronisation des données du CépiDc avec les données d'état civil de l'Insee constitue un enjeu majeur de cohérence entre les deux institutions sur leur champ commun. Cette étape du processus de production vise à :

- Apparier les enregistrements CépiDc avec les données Insee,
- Enrichir les données CépiDc avec les informations complémentaires mises à disposition par l'Insee,
- Contrôler l'exhaustivité des données de part et d'autre.

Les éléments suivants seront mis en place dans le cadre de cette phase :

- Un espace d'accueil dédié aux données Insee, avec contrôles à réception,
- Une traçabilité complète des choix d'appariement dans la zone de traitement (fonctionnalité non disponible dans le SI actuel),
- Un mécanisme d'enrichissement des données CépiDc à partir des données Insee,
- Une procédure de gestion des certificats manquants côté CépiDc.

3.3.3.5. Pilier 4 - Alimentation de la veille sanitaire

La transmission régulière de données à la veille sanitaire est l'une des missions centrales du CépiDc. Les flux de données devront être adaptés au nouveau modèle de données cible et maintenus en continu. Ce pilier constitue l'un des livrables prioritaires de la phase 2.

3.3.3.6. Pilier 5 - Orchestration de la chaîne de production

La phase 2 inclut la livraison du système complet d'orchestration de la chaîne de production. Celui-ci comprend :

- L'articulation des différents pipelines de traitement et la gestion de leurs dépendances.
- Un système centralisé de suivi et d'alertes automatiques en cas d'anomalie ou de rupture de flux.
- Un tableau de bord de pilotage de la campagne, et de suivi de la production présentant des indicateurs clés avec des fréquences de calcul et de mise à jour paramétrables.
- La capacité à générer des photographies intermédiaires des données de diffusion, à des fins d'analyse et d'éventuelles rétro-actions.

3.3.3.7. Résultats attendus en fin de phase 2

À l'issue de cette deuxième phase :

- La chaîne de production est intégralement opérationnelle pour tous les traitements.
- Le SI est capable d'intégrer, de contrôler et de corriger les données administratives et médicales.
- Les résultats du système de codage sont intégrés et traçables.
- La synchronisation avec les données de l'Insee est fonctionnelle.
- Les flux vers la veille sanitaire sont opérationnels.
- L'orchestration, le pilotage de la chaîne et le suivi de l'ensemble des traitements sont automatisés, assurés au travers d'une plateforme unique et centralisée permettant de superviser l'ensemble de la chaîne de production. Cette plateforme offre une vision consolidée de l'état d'avancement des traitements, des flux de données, des contrôles réalisés, des incidents détectés et des actions correctives engagées. Elle permet également l'automatisation des processus, le suivi des performances, la traçabilité des opérations et la production d'indicateurs de pilotage nécessaires au fonctionnement et à l'amélioration continue du système.

3.3.4. Phase 3 – Figeage, diffusion et mise en production

3.3.4.1. Objectifs

La troisième et dernière phase a pour objet la mise en place du processus de figeage des données et la mise à disposition des produits finis aux différents destinataires. Elle conduit à la mise en production complète du nouveau SI et au décommissionnement de l'ancien.

3.3.4.2. Processus de figeage des données

Le processus de figeage est conçu pour garantir la qualité, la traçabilité et la reproductibilité des données diffusées. Il comprend les éléments suivants :

- Estampillage de la base finale (producteur, niveau de contrôle, périmètre, conditions de production, etc.).
- Documentation complète et association de métadonnées standardisées.
- Procédures de contrôle orientées diffusion, incluant la vérification de l'anonymisation des textes en clair.

3.3.4.3. Campagne de production en double

Afin de valider le nouveau SI de bout en bout, une campagne de production en parallèle avec l'ancien SI sera conduite durant cette phase. La comparaison des bases produites par les deux systèmes constituera le critère de décision formel (go / no-go) pour la mise en production définitive.

3.3.4.4. Résultats attendus de la phase 3

Le processus de figeage est opérationnel et les données sont disponibles pour les destinataires. Le nouveau SI est validé en production et l'ancien SI peut être décommissionné.

3.4. Prestations de fonctionnalités supplémentaires à la demande

3.4.1. Reprise de l'historique des données

La reprise de l'historique constitue un chantier distinct et qui pourra être déclenché à la demande. Elle consiste à migrer vers le nouveau modèle de données les données de production accumulées sur plusieurs décennies d'exploitation du SI actuel.

Les enjeux principaux de cette reprise sont structurants pour la valeur à long terme du SI. Il s'agit d'une part d'assurer la continuité des séries statistiques, pour répondre au besoin de séries longues des utilisateurs et d'autre part d'assurer une comparabilité temporelle des données de mortalité, sans laquelle toute analyse épidémiologique perd de son sens. Par ailleurs, la reprise du linéage des données lors de cette phase permettra d'avoir la capacité d'expliquer le parcours de la donnée.

Différents écueils sont néanmoins à prendre en compte :

- La profondeur historique des données est grande. Aussi le format des données est hétérogène (changement de nomenclature, introduction du certificat électronique, etc.), la structure des bases a évolué au fil du temps.
- Un travail de réconciliation des 4 identifiants coexistant dans le SI actuel devra être mené.
- Les référentiels en particulier géographiques ne sont pas millésimés. La qualité des données diminue plus on remonte dans le temps.

Si cette option est activée, plusieurs scénarios correspondant à des profondeurs temporelles différentes devront être étudiés par le prestataire, sur la base d'une analyse coût/bénéfice intégrée à la réponse à l'appel d'offres. La profondeur minimale est estimée à cinq ans.

Quel que soit le scénario retenu, la reprise des données historiques devra suivre les étapes suivantes :

- Extraction des données historiques par année de décès, avec conservation des identifiants d'origine.
- Génération d'un identifiant rétrospectif au format cible pour chaque enregistrement.
- Constitution d'une table de correspondance entre le nouvel identifiant et les anciens identifiants.
- Transcodification des variables
- Reconstitution des référentiels millésimés pour chaque année concernée
- Chargement dans la zone de diffusion du nouveau SI, en état figé rétrospectif, en lecture seule.

La procédure de reprise sera validée statistiquement par comparaison des indicateurs clés (nombre de décès, principales causes, répartition par âge et sexe) entre l'ancien et le nouveau SI, année par année.

3.4.2. Intégration de données tierces

La prise en charge et l'intégration de données tierces issues de systèmes d'information externes dont la structure, le format ou les règles de gestion diffèrent des données principales traitées par le SI est déclenchée à la demande. À titre d'exemple, cette prestation pourra concerner l'intégration de données produites par les instituts médico-légaux ou tout autre fournisseur de données. La solution devra permettre la réception, le stockage, la transformation, le rapprochement et l'harmonisation de ces données afin de les rendre exploitables au sein du système cible. Elle devra également offrir des mécanismes de contrôle de qualité, de détection des incohérences et de traçabilité des traitements réalisés. Le titulaire décrira les capacités de la solution à gérer des formats hétérogènes (fichiers plats, XML, Excel, bases de données, API, etc.). L'intégration de ces données tierces devra être réalisée sans impact sur les traitements standards du système et garantir la cohérence globale des données exploitées.

3.4.3. Tierce maintenance applicative

Une fois le nouveau SI mis en production, il devra régulièrement être maintenu. En particulier, une adaptation forte sera à prévoir, même si anticipée dès la conception, pour le changement de nomenclature de codage et le passage à la version 11 de la classification internationale des maladies.

3.4.4. Conduite du changement et accompagnement à la transformation organisationnelle

La refonte du système d'information de production du CépiDc est susceptible d'entraîner des évolutions significatives des modes de fonctionnement, des responsabilités, des processus opérationnels et des activités réalisées par les équipes métier.

A la demande, le titulaire proposera une prestation d'accompagnement à la conduite du changement et à l'analyse des impacts organisationnels induits par la mise en œuvre du futur système d'information.

Cette prestation devra permettre d'identifier et de formaliser l'organisation cible permettant de tirer pleinement parti des capacités offertes par le nouveau système d'information. Le titulaire proposera une démarche permettant de définir :

- les rôles et responsabilités des différents acteurs de la chaîne de production statistique ;
- les modalités de pilotage et de supervision de la production ;
- l'organisation des activités de contrôle et de gestion des anomalies ;
- les modalités de gestion et de pilotage des évolutions ;
- les interactions entre les équipes métier, les équipes techniques et les partenaires externes ;
- les compétences clés à développer ou à renforcer pour assurer l'exploitation pérenne de la solution.

Les livrables attendus pourront comprendre :

- une description détaillée de l'organisation cible ;
- une cartographie des processus cibles ;
- une matrice des rôles et responsabilités (RACI) ;
- une identification des impacts sur les compétences et les métiers ;
- une trajectoire de transformation permettant de passer progressivement de l'organisation actuelle à l'organisation cible ;
- un plan d'accompagnement au changement et de montée en compétences des équipes.

Cette prestation fera l'objet d'un chiffrage distinct dans l'offre financière et sera présentée comme une prestation à la demande pouvant être activée à la demande par émission de bon de commande.

3.5. Planning prévisionnel

Une stratégie de bascule en big bang ferait courir d'importants risques sur la réussite du projet et la capacité à produire du CépiDc. Aussi, une démarche incrémentale devra être mise en place, de manière à aboutir à un double run sur la campagne de 2027 : cette stratégie permettra de sécuriser à la fois le projet et la production courante du CépiDc.

Sous l'hypothèse d'un démarrage de la prestation liée au présent marché à l'automne 2026, le double run sur la campagne 2027 (données des décès survenus en 2026) pourrait débuter au printemps 2027 pour se terminer à l'été 2028, avec le figeage des données de la campagne 2027, cette opération servant de juge de paix pour valider le nouveau SI. En suivant ce calendrier, le SI courant pourrait être décommissionné début 2029, pour le lancement du traitement des données 2028.

4. Organisation et gouvernance du projet

Le titulaire devra mettre en œuvre une organisation de projet permettant d'assurer le pilotage, le suivi, la coordination et la bonne exécution de l'ensemble des prestations prévues au marché. Cette organisation devra garantir une collaboration étroite avec les équipes de l'Inserm tout au long du projet, depuis la phase de cadrage jusqu'à la mise en production et la période de garantie.

Le titulaire s'engage à mettre en place les moyens humains, techniques et organisationnels nécessaires au respect des objectifs du projet, des délais contractuels et des exigences de qualité définies dans le présent CCTP.

4.1. Composition de l'équipe projet

Le titulaire devra constituer une équipe projet dédiée disposant des compétences nécessaires à la réalisation des prestations. Le titulaire présentera dans son offre :

- L'organisation de l'équipe projet ;
- Les rôles et responsabilités de chaque intervenant ;
- Les taux d'affectation prévisionnels ;
- Les modalités de remplacement des ressources en cas d'absence ou de départ.

Tout remplacement d'une ressource clé devra être préalablement validé par le CépiDc.

Dans le cadre de son offre, le soumissionnaire devra présenter son analyse des contributions attendues de la maîtrise d'ouvrage et des équipes du CépiDc tout au long du projet.

Cette analyse devra permettre d'identifier les sollicitations nécessaires des équipes métier, fonctionnelles et techniques afin d'assurer la bonne réalisation du projet, ainsi que les hypothèses retenues pour l'élaboration de l'offre.

Le soumissionnaire présentera, sous forme de tableau, une estimation des charges de la maîtrise d'ouvrage exprimées en jours/homme par profil et par phase du projet.

À titre indicatif, les profils suivants pourront être distingués :

- Direction de projet ;
- Product Owner / Responsable métier ;
- Experts métier ;
- Référents de production ;
- Équipe DSI ;
- Référents sécurité et conformité.

Le soumissionnaire précisera également l'ensemble des fournitures, informations, accès, environnements, jeux de données, référentiels, ressources humaines ou moyens techniques qu'il estime nécessaires pour mener à bien les prestations.

Toute hypothèse structurante concernant la disponibilité des équipes du CépiDc ou de la DSI Inserm devra être explicitement mentionnée dans l'offre. À défaut, le titulaire sera réputé avoir intégré dans son offre les niveaux de contribution raisonnablement nécessaires à la réalisation des prestations décrites au présent CCTP.

4.2. Présence sur site

Afin de favoriser les échanges avec les équipes métiers et techniques, le titulaire devra assurer une présence régulière sur site. *A minima*, le chef de projet ainsi que les ressources techniques concernées par les développements en cours devront être présents sur site une journée par semaine.

Le CépiDc pourra demander une fréquence supérieure lors des phases critiques du projet, notamment :

- Phase de cadrage ;
- Ateliers de conception ;
- Recette ;
- Préparation de la mise en production ;
- Accompagnement au démarrage

Certaines interventions pourront être réalisées à distance pour les activités ne nécessitant pas une présence physique, sous réserve de validation préalable.

4.3. Gouvernance du projet

La gouvernance du projet reposera sur plusieurs instances de suivi.

4.3.1. Comité de pilotage (COPIL)

Le comité de pilotage aura pour mission de valider les orientations stratégiques, de suivre l'avancement global du projet, d'arbitrer les sujets majeurs, de valider les livrables structurants, d'identifier et traiter les risques.

Il se réunira environ une fois par mois. Le titulaire préparera les supports de présentation et les comptes rendus.

4.3.2. Comité projet (COPROJ)

Le comité projet assurera le suivi opérationnel du projet. Il permettra notamment le suivi du planning, le suivi des actions, le suivi des risques et points bloquants, le suivi des développements, le suivi des recettes.

Le comité projet se réunira une fois par semaine.

4.3.3. Ateliers de travail

Le titulaire animera les ateliers nécessaires à la réalisation du projet :

- Ateliers de recueil des besoins ;
- Ateliers de conception fonctionnelle ;
- Ateliers de conception technique ;
- Ateliers de validation ;
- Ateliers de préparation des recettes.

Chaque atelier donnera lieu à un compte rendu validé par les participants.

4.4. Reporting et indicateurs

Le titulaire produira un reporting régulier comprenant notamment :

- L'état d'avancement du projet ;
- Les jalons réalisés ;
- Les actions en cours ;
- Les points d'attention ;
- Les risques identifiés ;

- Les indicateurs de qualité ;
- Les indicateurs de charge.

Un tableau de bord consolidé sera présenté à chaque comité de pilotage.

4.5. Gestion documentaire

Le titulaire assurera la mise à disposition et la maintenance de l'ensemble de la documentation projet. L'ensemble de la documentation produite dans le cadre du marché deviendra la propriété du CépiDc.

5. Livrables attendus

L'ensemble des livrables devra être fourni sous format électronique modifiable. Les livrables devront être maintenus à jour tout au long du projet et refléter en permanence l'état réel de la solution mise en œuvre.

Le titulaire veillera à produire une documentation suffisamment détaillée pour permettre au CépiDc d'assurer l'exploitation fonctionnelle de la solution, de limiter sa dépendance au titulaire et de faciliter les opérations de maintenance, d'évolution et de réversibilité.

5.1. Livrables de conception fonctionnelle

Le titulaire produira les documents décrivant les besoins métiers et les fonctionnalités de la solution :

- Dossier de conception fonctionnelle générale ;
- Dossier de conception fonctionnelle détaillée ;
- Description des processus métiers ;
- Cartographie des flux de données ;
- Description des règles de gestion ;
- Description des contrôles de qualité et de cohérence ;
- Description des règles d'orchestration des traitements ;
- Description des processus de gestion des campagnes ;
- Modélisation des référentiels ;
- Description des rôles et habilitations fonctionnelles ;
- Maquettes et prototypes des interfaces.

5.2. Livrables de conception technique

Le titulaire produira les documents techniques permettant l'exploitation et la maintenance de la solution :

- Dossier d'architecture technique ;
- Dossier d'architecture applicative ;
- Schémas d'infrastructure ;
- Description détaillée des échanges de données et contrats d'interfaces ;
- Modèle de données ;
- Dictionnaire des données ;
- Stratégie de sauvegarde et de restauration ;
- Stratégie de continuité d'activité ;
- Stratégie de reprise après incident ;
- Documentation relative à la sécurité.

5.3. Livrables relatifs aux traitements statistiques et aux flux de données

Le titulaire fournira les documents décrivant les traitements statistiques métier, ainsi que le contenu des flux de données :

- Documentation de l'interfaçage avec les traitements de codage (logiciel métier IRIS, codage IA)
- Description de l'intégration des données Insee et des modalités/mécanismes d'appariement et de traçabilité ;
- Description du processus de gestion des certificats manquants ;
- Documentation des corrections / traitements des données administratives
- Description du processus d'anonymisation et de ses modalités
- Description de tous les flux de données, entrant et sortant (contenu, format)

- Documentation du séquençement de la chaîne de production : dépendances entre pipelines, remontée d'alerte, suivi centralisé
- Documentation des indicateurs de suivi de production.

5.4. Livrables de configuration du progiciel

La solution reposant sur un progiciel, le titulaire fournira l'ensemble des éléments permettant son administration, son paramétrage et son évolution.

Les livrables attendus comprennent notamment :

- Dossier de paramétrage fonctionnel ;
- Dossier de paramétrage technique ;
- Documentation des workflows ;
- Documentation des pipelines ;
- Documentation des référentiels ;
- Documentation des nomenclatures ;
- Documentation des règles métier ;
- Documentation des règles de contrôle ;
- Documentation des tableaux de bord ;
- Documentation des alertes ;
- Documentation des profils utilisateurs ;
- Documentation des habilitations ;
- Journal des paramétrages réalisés ;
- Inventaire des personnalisations ;
- Inventaire des extensions réalisées ;
- Documentation des éventuels développements spécifiques.

Pour chaque fonctionnalité majeure, le titulaire indiquera si celle-ci repose :

- sur une fonctionnalité standard du progiciel ;
- sur un paramétrage spécifique ;
- sur un développement spécifique.

5.5. Livrables de tests et de recette

Le titulaire réalisera les tests relevant de sa responsabilité, notamment les tests unitaires, d'intégration, de bout en bout (end-to-end), système, de non-régression, de performance et, le cas échéant, de sécurité.

Il livrera au CépiDc

- un environnement de test, de manière à ce que le CépiDc puisse également réaliser des tests.
- une documentation fonctionnelle de chaque version livrée, décrivant en différence les apports par rapport à la version précédente.
- Une procédure (*via* un outil de ticketing ou pas) pour faire remonter les anomalies et suivre leur traitement
- Un plan de tests (fonctionnels, techniques, de performance et de non-régression) avec les cahiers de recette, jeux de données de test établis avec le CépiDc

Le titulaire assistera le CépiDc pendant toute la phase de recette jusqu'à la validation des livrables et à la levée des réserves éventuelles.

5.6. Livrables de sécurité et de conformité

Le titulaire fournira :

- Dossier d'architecture de sécurité ;
- Cartographie des habilitations ;
- Matrice des droits d'accès ;
- Documentation des mécanismes d'authentification ;
- Documentation d'intégration au SSO Inserm ;
- Documentation des mécanismes de chiffrement ;
- Documentation des journaux de sécurité ;
- Documentation des mécanismes d'audit ;
- Documentation des sauvegardes ;
- Documentation de restauration ;
- Analyse des risques de sécurité ;
- Éléments nécessaires aux démarches d'homologation ;
- Documentation relative au RGPD ;
- Registre des traitements concernés ;
- Documentation de traçabilité des accès ;
- Documentation des mécanismes de conservation et d'archivage des données.

5.7. Livrables de déploiement

Le titulaire produira le :

- Plan de déploiement ;
- Plan de migration ;
- Stratégie de reprise des données ;
- Procédures de mise en production (avec stratégie de bascule, plan de double run, procédure de retour-arrière, critères de go/no-go) ;
- Procès-verbal de mise en production ;
- Bilan de déploiement.
- Bilan du double-run et comparaison ancien SI/nouveau SI.

5.8. Livrables d'exploitation et de maintenance

Le titulaire fournira :

- Manuel d'administration ;
- Manuel d'exploitation ;
- Manuel utilisateur ;
- Guide des procédures courantes (gestion de campagnes, de référentiels, lancement et reprise des traitements, gestion des incidents, gestion des anomalies...) ;
- Documentation de supervision ;
- Documentation des mécanismes de sauvegarde ;
- Documentation des procédures de maintenance ;
- Base de connaissances des incidents et solutions.
- Documentation des mises à jour du progiciel.

5.9. Livrables de formation et d'accompagnement

Le titulaire produira :

- Supports de formation administrateurs ;
- Supports de formation utilisateurs ;
- Guides pratiques et de prise en main avec des tutoriels ;

- Exercices et cas pratiques

5.10. Livrables relatifs à la gouvernance des données

Le titulaire produira :

- Documentation des campagnes de production ;
- Catalogue des métadonnées ;
- Documentation des référentiels ;
- Documentation des nomenclatures ;
- Documentation des règles et indicateurs de qualité ;
- Documentation du linéage des données et de leur transformation ;
- Documentation du versionning (des modèles IA utilisés, des versions du moteur IRIS...)

5.11. Réversibilité

En fin de marché ou en cas de changement de titulaire, le prestataire devra fournir :

- Documentation consolidée et mise à jour ;
- Inventaire des composants ;
- Inventaire des flux ;
- Inventaire des développements spécifiques ;
- Export des données et métadonnées ;
- Procédures de reprise par un tiers ;
- Rapport de transfert de connaissances.

5.12. Livrables par palier

À l'issue de chaque palier du projet, le titulaire produira un dossier de validation comprenant :

- les fonctionnalités et paramétrages réalisés ;
- les interfaces réalisées ;
- les écarts constatés
- les résultats des tests et anomalies résiduelles ;
- les actions restant à mener ;
- les critères d'entrée du palier suivant
- la décision de validation du palier.

À minima :

- **Phase 1** : validation de la zone de réception, des mécanismes d'intégration et du socle de données ;
- **Phase 2** : validation des traitements métier, des mécanismes de contrôle, du codage, de la synchronisation Insee et de l'orchestration ;
- **Phase 3** : validation du figeage, du double-run, de la bascule, de la mise en production et du décommissionnement du SI historique.

6. Annexes

6.1. Documentation sur l'existant

Une documentation de l'existant est transmise aux candidats au sein du dossier de consultation. Les documents transmis sont les suivants :

- Présentation générale :
 - o Dossier de demande de re-labellisation des séries issues de la statistique sur les causes de décès - mars 2025
 - o Les statistiques sur les causes médicales de décès de A à Z – août 2022
 - o Courrier des statistiques N°12 - décembre 2024
 - o Présentation de la production du CépiDc – 2024
 - o RACI – cycle de vie de la donnée au CépiDc - avril 2026
- Documentation méthodologique :
 - o Document de travail du CépiDc sur le codage des causes de décès de 2018 et 2019 en CIM10 (approche combinant deep learning, système expert et codage manuel ciblé) - septembre 2023
 - o Fiche méthodologique : Définitions des indicateurs sur les causes de décès utilisés sur le site du CépiDc - janvier 2025
-
- Documentation relative au système d'information existant
 - o Analyse de l'existant - juillet 2024
 - o Analyse de l'existant : cadre général du SI du CépiDc – février 2026
 - o Présentation de la BDD du CépiDc – 2024 (y compris dictionnaire de variables, inventaire des applications et traitements intervenant dans le processus de production des données du CépiDc)
 - o Liste non exhaustive des contrôles existants actuellement dans le SI – avril 2026
- Flux, échanges de données et synchronisation avec les partenaires
 - o Présentation de l'opération de synchronisation avec l'Insee - 2024
 - o Document de travail sur l'opération de synchronisation – septembre 2025
 - o Guide du référent en établissement de santé (paramétrer CertDc pour son utilisation en établissement) - mai 2020

6.2. Travaux de cadrage réalisés au S1 2026

Afin de préparer la présente consultation, le CépiDc a mené, au cours du premier semestre 2026, une phase de cadrage du projet de refonte du système d'information de production.

Cette phase a conduit à l'élaboration des livrables suivants :

- Document 1 – Architecture, données et gouvernance
- Document 2 – Démarche projet et trajectoire
- Document 3 – Infrastructure et exigences techniques
- Document 4 – Synthèse exécutive

6.3. Critères d'évaluation

Les offres remises par les candidats feront l'objet d'une analyse comparative au regard des critères de jugement définis dans les documents de la consultation.

L'évaluation portera notamment sur :

- la compréhension des enjeux métier, fonctionnels et techniques du projet ;
- la pertinence de la solution proposée au regard des objectifs du CépiDc ;
- l'adéquation du progiciel proposé aux besoins exprimés dans le présent CCTP ;
- la couverture fonctionnelle de la solution et les modalités de paramétrage ;
- la qualité de l'architecture proposée ;
- les capacités d'intégration, de supervision, de traçabilité et de gouvernance des données ;
- les modalités de conduite du projet, de pilotage et de gestion des risques ;
- les moyens humains mobilisés et l'expérience des intervenants ;
- la qualité des prestations d'accompagnement, de formation et de transfert de compétences ;
- les garanties apportées en matière de sécurité, d'exploitation, de maintenabilité et d'évolutivité ;
- la cohérence des charges, du planning et des hypothèses de réalisation ;
- le coût global de la solution sur la durée du projet.

Les modalités détaillées de notation, les pondérations associées ainsi que les éventuels sous-critères figurent dans les documents de consultation et dans la grille d'évaluation des offres.

Le CépiDc se réserve la possibilité de demander aux candidats toute précision complémentaire qu'il jugerait nécessaire à l'analyse de leur offre.

Une attention particulière sera portée à la capacité du candidat à maximiser la couverture des besoins par les fonctionnalités standard du progiciel et à limiter le recours aux développements spécifiques.

6.4. Cadre de réponse technique

Le mémoire technique sera complété par la grille de conformité fonctionnelle et technique et par le cadre de réponse financier, qui constituent les supports de réponse aux exigences du présent CCTP.

