

Programme annuel de travail du CépiDc pour 2026

Mars 2026

Document de travail du CépiDc N°07

Table des matières

1) Assurer la production dans un environnement évolutif	3
a) Assurer la production et la diffusion des données dans le calendrier fixé par Eurostat	3
b) Adaptations aux évolutions.....	3
2) Sécuriser et rationaliser la production.....	4
a) Automatisation et industrialisation du codage automatique (IA + Iris).....	4
b) Professionnaliser le suivi de collecte/production	4
c) Documentation.....	5
d) Systématiser le binôme	5
e) Sécuriser le SI.....	5
3) Améliorer la qualité.....	6
a) Qualité de la collecte	6
b) Qualité du codage	7
c) Qualité de la diffusion	7
4) Promouvoir, diffuser, collaborer	7
a) Collaboration de recherche.....	7
b) Publications et colloques.....	8
c) Collaborations internationales	8
5) Annexe - Liste des publications 2025	9

1) Assurer la production dans un environnement évolutif

a) Assurer la production et la diffusion des données dans le calendrier fixé par Eurostat

Le premier semestre 2026 sera consacré à la validation des données relatives aux décès survenus en 2024, dont le codage s'achèvera fin février, et à leur diffusion (Eurostat, OMS, SNDS, open data). La sortie des publications annuelles est prévue pour le 23 juin 2026, soit une quinzaine de jours plus tôt que les données du millésime précédent. Nouveauté de l'année 2026 : deux publications sortiront au format Etudes et Résultats de la Drees, de manière à isoler le bilan épidémiologique de l'analyse anglée, et à accélérer le processus de relecture.

Concernant les données relatives aux décès survenus en 2025, leur codage a commencé fin janvier et leur validation débutera au 2^{ème} semestre. Outre le maintien en conditions opérationnelles du batch automatique Iris réalisé en continu, un certain nombre d'évolutions sont prévues pour faciliter et sécuriser la production des données :

- une évolution sera demandée à la TMA, de manière à pouvoir reconstruire et stocker en base de données (tables Iris) le texte déclaré dans les certificats dans sa forme standardisée : les spécifications seront livrées à la fin du 1^{er} semestre pour des développements et une recette au 3^{ème} trimestre ;
- des améliorations seront également apportées dans les traitements post-batch pour mieux prendre en compte les clusters et les blocages de certaines causes de décès, notamment concernant les chutes ;
- si le temps le permet, le code de certains programmes sera revu de manière à être plus efficace.

Afin de préparer la production des données 2026 et suivantes, le dictionnaire en entrée d'Iris sera mis à jour. Pour ne pas produire du bruit dans les évolutions qui ne serait lié qu'à des artefacts venant de changements annuels de dictionnaire, décision a en effet été prise de ne plus le modifier à un rythme annuel, mais tous les 3 à 5 ans. Il n'a ainsi pas été mis à jour pour la production des données 2025.

Parmi les modifications qui seront mises en œuvre en 2026 :

- intégration des codes liés aux maladies rares (collaboration Orphanet) ;
- intégration des choix de codes ;
- meilleure prise en compte des sepsis ;
- améliorations en continu suite aux retours consolidés de l'équipe de codage ;
- travail sur la standardisation pour préparer le passage à la CIM-11.

b) Adaptations aux évolutions

L'ouverture de la certification aux infirmier.e.s a été généralisée le 26 avril 2025. Les décès survenus fin 2025 sont donc les premiers à avoir pu faire l'objet d'une certification par un.e professionnel.le de santé qui ne soit pas médecin. L'impact de cette ouverture sur la nature des déclarations et la manière de remplir le certificat et ses conséquences sur le codage sera étudiée au 2^{ème} trimestre.

La ministre de la santé a demandé à l'été 2025 une évolution du certificat de décès, pour être en capacité de suivre par ce biais la mise en œuvre de la loi sur la fin de vie, qui devrait être votée courant 2026. Cette demande a été le catalyseur pour lancer des travaux de mise en cohérence du modèle de

certificat avec le modèle international. Suite à plusieurs réunions d'un groupe de travail dédié co-piloté avec la DGS, un projet de certificat a été mis au point. Il sera testé dans la cadre de tests cognitifs (S1) auprès de professionnel.le.s certificateurs.trices et validé définitivement d'ici la fin du premier semestre. Le 2^{ème} semestre sera consacré à l'adaptation du système d'information à ce nouveau modèle : prise en compte par Iris, par le modèle IA, évolutions de la base de données, de CertDc, etc.

Si la date de bascule à la CIM-11 n'est pas encore fixée, un gros travail collectif sera mené en 2026 pour être en capacité de doubler-coder manuellement dans les deux versions de la CIM en 2027 au moins sur un échantillon de certificats, pour les données 2026. Il s'agira en particulier :

- de mettre à disposition un environnement de codage Iris en CIM-11 (T1) puis de tester le batch automatique Iris en CIM-11 (T2-T3) ;
- de construire un mapping CIM10-CIM11- (T2-T3) ;
- de déterminer la stratégie à mettre oeuvre pour le codage par l'IA (T3-T4) ;
- de réaliser des comparaisons structurées CIM-10-CIM11 sur des échantillons EDP (S2) ;
- de former l'équipe de codage à la CIM-11 ;
- d'anticiper l'impact du passage à la CIM-11 sur les séries longues de mortalité.

2) Sécuriser et rationaliser la production

Désormais stabilisée dans ces principes, la production de la statistique sur les causes médicales de décès doit maintenant être sécurisée et rationalisée, de manière à disposer d'une chaîne de production plus facilement pilotable.

a) Automatisation et industrialisation du codage automatique (IA + Iris)

La stratégie de codage mise en place à partir du traitement des données 2021 est désormais éprouvée, mais sa mise en œuvre nécessite beaucoup d'interventions manuelles qui fragilisent le processus. Aussi, 2026 sera assez largement consacrée à l'automatisation et à l'industrialisation du codage automatique (IA et IRIS) :

- sécurisation et industrialisation complète de la chaîne IA ;
- automatisation de l'import en base de données des sorties de la chaîne IA (S1) ;
- amélioration du jeu de test, avec intégration du codage Gold et prise en compte des résultats du stage réalisé à l'été 2025 ;
- écriture d'une procédure de retour arrière en production, en cas de régression d'un modèle (T2) ;
- automatisation du lancement du batch Iris depuis l'environnement self de production grâce à la mise en place d'une tâche planifiée.

b) Professionnaliser le suivi de collecte/production

Dans la même logique de sécurisation, un des objectifs de 2026 est d'industrialiser le suivi et le contrôle de toute la chaîne de production, de la collecte à la validation des résultats, réalisés jusqu'ici de manière artisanale. Plusieurs axes de travail sont prévus :

- automatiser l'alimentation du tableau de bord de suivi de la production (en y ajoutant des indicateurs de suivi de la collecte des certificats et les indicateurs de suivi du batch automatique Iris) ;
- automatiser la production des sorties visant à alimenter le rapport de production annuel, à assurer l'exhaustivité de la collecte des certificats papier, etc. ;

- construire des indicateurs de performance et de validation des modèles IA ;
- standardiser les sorties de la chaîne IA pour en faciliter la validation ;
- formaliser une procédure de validation scientifique des modèles IA avant mise en production (jeu test, critères, comité de validation) ;
- concevoir des indicateurs de performance du batch automatique Iris, en fonction de la complexité du certificat.

Une étude sera également menée pour déterminer s'il est possible de réduire le recours à la reprise manuelle des certificats électroniques par le prestataire de saisie afin de rationaliser les coûts associés.

c) Documentation

Afin toujours de sécuriser la production, mais aussi de faciliter les passations et les montées en compétences des nouvelles recrues comme de faciliter l'exploitation des données mises à disposition, l'ensemble des programmes sera mis à disposition et versionné sur un Git (privé pour les programmes de production, public pour ceux permettant de reproduire les tableaux publiés). Le stockage sur le serveur de fichiers interne sera également réorganisé avec le même objectif.

De la même manière, une recherche sera menée pour déterminer la meilleure méthode d'archivage des modèles IA et des jeux de données d'entraînement.

La documentation (modes opératoires, fiches de processus, documentation technique) sera systématisée.

d) Systématiser le binôme

Avec la stabilisation de la production et du périmètre de l'équipe, organiser un binôme systématique de toutes les missions est un autre axe de sécurisation de la production courante. Aussi, le codage en binôme sera instauré comme modalité de travail régulière ; le compagnonnage sera favorisé afin d'assurer une meilleure appropriation des règles de codage et le partage des compétences et des expertises sur l'ensemble des thématiques de codage sera encouragé. Une logique identique sera adoptée sur les sujets liés au suivi de collecte et au sein du pôle automatisation.

e) Sécuriser le SI

Le système d'information du CépiDc a été identifié comme étant un des principaux points de fragilité de la production des causes médicales de décès : 2026 sera l'année de sa refonte, avec un premier projet de cadrage qui aura lieu au premier semestre et un second de développement qui sera lancé au second semestre.

En parallèle, d'autres travaux en lien avec le SI seront également engagés :

- une étude de faisabilité de bascule sur le cloud Inserm pour la réalisation de toutes les missions IA sera réalisée au 1^{er} semestre. Si une telle bascule est possible, cela permettrait de disposer d'une solution pérenne (fin de la garantie des machines IA actuelles en 2027), fiable (sauvegardes automatiques régulières) et managée ;
- une autre étude de faisabilité sera engagée (T2/T3) pour déterminer s'il est possible de formater l'ensemble des log applicatives, de les centraliser dans Splunk et quels en seraient les coûts/avantages ;

- au premier semestre également, l'active directory (AD) spécifique CépiDc sera abandonné au profit d'une bascule générale dans l'AD national Inserm. Cela permettra de répondre aux exigences de sécurité informatique posées par l'ANSI et devrait limiter considérablement les problèmes quotidiens liés à l'utilisation jusqu'ici de trois AD différents ;
- une évolution du flux de la veille sanitaire sera étudiée pour l'enrichir des prédictions IA non finalisées. Ce dernier chantier sera à coordonner avec le projet de refonte globale du SI.

Enfin, les travaux IA seront mis en conformité, lorsque ce n'est pas déjà le cas, avec les réglementations IA européenne et Inserm.

3) Améliorer la qualité

a) Qualité de la collecte

La qualité de la collecte, en particulier avoir une bonne couverture, est un élément essentiel de la qualité finale de la statistique sur les causes médicales de décès. Deux axes de travail thématiques ont été identifiés pour 2026 :

- améliorer la remontée des volets médicaux complémentaires (VMC) qui concernent tous les décès pour lesquels une autopsie a été demandée. Pour cela, les modalités actuelles de remontée des VMC seront analysées et les freins à leur utilisation par les certificateurs seront identifiés. Des solutions alternatives de remontée d'information (Medle ou flux de données spécifique à l'image de ce qui est fait avec l'IML de Paris) seront investiguées.
- accroître la qualité de la statistique de mortalité infantile, par la fourniture d'un état des lieux sur l'utilisation des VMC lors de décès d'enfants de moins d'un ou deux ans, la mise en place d'un appariement avec les données de l'OMIN et du MIN afin de mettre en évidence les données manquantes ou différentes ou encore par la collaboration avec l'équipe Snoopi de la DREES pour étudier finement les causes de décès de ces enfants.

Le renforcement de l'usage de la certification électronique et la formation des certificateurs à la manière de « bien » remplir un certificat sont également deux angles d'attaque pour améliorer la qualité de la collecte. Aussi, le CépiDc continuera avec la DGS à promouvoir l'utilisation de la certification électronique (en identifiant les trous de collecte électronique pour cibler cette communication) et reprendra le travail lancé pour introduire un outil de traitement automatique du langage dans l'application CertDc, de manière à lever d'éventuelles ambiguïtés dans la saisie ou demander des précisions (S1). En parallèle, le CépiDc lancera des actions à destination des certificateurs directement : mise en place d'un dispositif permettant de recontacter les certificateurs dans des cas précis comme certains décès sensibles, communication/formation au remplissage à destination des certificateurs remplissant un grand nombre de certificats (SMURistes par exemple). Les supports d'aide et de formation (vidéos, infographies, consignes) seront revus, pour les améliorer et les rendre plus percutants.

b) Qualité du codage

En symétrie des actions lancées sur la collecte, un travail similaire sera mené sur la codification, avec un focus sur le codage des VMC, en développant une expertise dédiée et en renforçant l'harmonisation des pratiques de codage. Les supports de formation interne seront enrichis et l'équipe de codage sera progressivement formée à l'utilisation des logiciels du Pack Office et sensibilisée aux traitements statistiques par le biais d'ateliers R.

Dans le but d'améliorer la qualité du codage également, une étude de faisabilité d'intégration dans les modèles IA des intervalles d'apparition des différentes étapes du processus morbides présents sur les certificats papier sera réalisée. Si la mise en œuvre se révèle possible, cela permettra de mieux séparer les maladies des séquelles et d'harmoniser les traitements entre certificats électronique et papier.

c) Qualité de la diffusion

Afin de rendre le site de mise à disposition des données en *open data* plus utilisable, il sera enrichi de différents éléments ou fonctionnalités (courbes, cartes, interrogations par API). La possibilité de mettre à disposition les tableaux de données labellisées sur data.gouv.fr sera étudiée.

Les programmes associés aux différentes publications seront mis à disposition des utilisateurs par le biais d'un dépôt git public (T3).

Une note méthodologique sur le calcul des grandes tendances sera rédigée et le mode de calcul de ces dernières sera consolidé grâce à une collaboration avec le DMS de l'Insee.

Un travail sera réalisé en collaboration avec l'ONISR de manière à être en capacité d'expliquer les écarts constatés entre les séries produites par chacune des unités, sur le champ des accidents de la route.

4) Promouvoir, diffuser, collaborer

En continu, le CépiDc répond aux demandes qui lui sont faites et réalise des produits de diffusion sur mesure pour répondre aux différents besoins.

a) Collaboration de recherche

Différentes collaborations de recherche seront poursuivies ou entamées en 2026 :

- collaboration avec des équipes de recherche sur les modèles IA CIM-11 (LISN/ANS) ;
- collaboration à l'enquête fin de vie de l'Ined, avec notamment le tirage de l'échantillon ;
- mise en qualité de la variable donnant le lieu de décès (Insee) ;
- collaboration avec les équipes en charge de la veille sanitaire pour permettre la transmission des prédictions issues des modèles IA ;
- poursuite des travaux sur le suivi décès par chutes des personnes âgées (SPF, ATIH) ;
- poursuite de la collaboration avec l'Ined sur les décès multi-cause ;
- collaboration avec l'Insee pour l'actualisation des projections de population à horizon 2070 ;
- étude sur la sous-estimation des suicides avec l'accueil en visite académique d'une doctorante et une collaboration avec l'ONS (automne 2026).

b) Publications et colloques

Outre les publications annuelles désormais habituelles (BEH, Etudes et résultats), un article traitant de la stratégie d'échantillonnage pour le codage manuel sera rédigé et présenté au colloque francophone international sur les sondages (mai 2026). Une soumission a également été acceptée, pour la conférence qualité de 2026, sur le thème du suivi de la qualité et de la pertinence de la stratégie de codage mise en place, par rapport à une campagne « classique » (juin 2026). Les travaux sur la mortalité infantile seront aussi valorisés sous la forme d'une publication, dans un format à déterminer.

Le CépiDc essaiera également de réactiver le groupe de travail avec les data-scientists de l'AP-HP. A défaut, l'idée serait *a minima* d'organiser un séminaire commun (T2).

c) Collaborations internationales

En 2026, le CépiDc poursuivra son implication dans les différents groupes de travail OMS et Iris. Des travaux internationaux plus ponctuels seront également engagés :

- collaboration avec l'OMS pour intégrer des règles de standardisation dans l'API (T2) ;
- collaboration européenne (Italie, Allemagne, Hongrie, Angleterre) pour améliorer le batch automatique Iris, au niveau des postes de la shortlist Eurostat ;
- consolidation des échanges techniques autour d'Iris et étude de l'intégration dans le logiciel d'un modèle d'IA pour un usage international.

5) Annexe- Liste des publications 2025

- Adnot P., Vegas N., Coutier L., Gras Leguen C., Giuseppi A., Aubineau Y., *et al.*, Benefit/risk balance of prone positioning as first line treatment for upper airway obstruction, *in* Robin sequence. *Eur J Pediatr*
- Godet F., Costemalle V., Aubineau Y., Fouillet A., Coudin E., Causes de décès en France en 2023 : des disparités territoriales ; Etudes et Résultats, n°1342
- Coudin E., Fouillet A., De l'importance du volet médical du certificat de décès pour la connaissance des causes de décès en France, *La Presse Médicale Formation.*,
- Fouillet A., Aubineau Y., Godet F., Costemalle V., Coudin E., Grandes causes de mortalité en France en 2023 et tendances récentes, *Bull Épidémiol Hebd*, 13:218–43.
- Godet F., Aubineau Y., Anonymisation des données diffusées en opendata par le CépiDc, 15^{ème} édition des Journées de méthodologie statistique de l'Insee (JMS 2025)
- Boulet P. *et al*, Rapport de production - Année de décès 2023, Document de travail du CépiDc n°6/2025
- Godet F., Définitions des indicateurs sur les causes de décès utilisés sur le site du CépiDc, Fiche méthodologique du CépiDc
- Godet F., Taux standardisé de mortalité – un indicateur permettant les comparaisons dans le temps et l'espace, Fiche méthodologique du CépiDc
- Ruptures de série dans le codage des causes de décès, Fiche méthodologique du CépiDc
- Robert A., Zambetta E., Razakamanana N., Coudin E., Identifier et classer les causes de décès en automatisant le traitement des certificats en langage naturel, 15^{ème} édition des Journées de méthodologie statistique de l'Insee (JMS 2025)
- Boulet P., Coudin E., Godet F., Robert A., La statistique sur les causes de décès, 15^{ème} édition des Journées de méthodologie statistique de l'Insee (JMS 2025)
- Razakamanana D., Robert A., Zambetta E., Coudin E., Aligning French Medical Dictionary with ICD-11 using a Hybrid Approach: RAG, Mapping Tables and CodeFusion, WHO-FIC, poster #327, Octobre 2025
- Coudin E., Hebbache Z., Cuenot M., *et al.*, French WHO-FIC Collaborating - Center Annual Report 2024-2025, WHO-FIC, poster #314, Octobre 2025
- O'Brien K., Loorbach-van Zutphen W., Heuser F., Quayle G., Coudin E., Causes of death coding in Iris with ICD-11 testing results, WHO-FIC, poster #200, Octobre 2025