

Compte rendu du congrès

Jeudi 6 novembre 2025



360 RWE

CONSTRUIRE UNE VISION PARTAGÉE AUTOUR
DES DONNÉES DE VIE RÉELLE : DES BÉNÉFICES
INDIVIDUELS AUX ENJEUX COLLECTIFS

UN ÉVÉNEMENT ORGANISÉ PAR



360 RWE

CONSTRUIRE UNE VISION PARTAGÉE AUTOUR
DES DONNÉES DE VIE RÉELLE : DES BÉNÉFICES
INDIVIDUELS AUX ENJEUX COLLECTIFS

SAVE THE DATE

Rendez-vous pour la 4^e édition
Le jeudi 12 novembre 2026
Au Verso Victoire, Paris

www.rwe360congress.com

Un remerciement à nos sponsors pour leur soutien lors de la 3^e édition
et la parution du compte rendu du congrès



abbvie



Appartient au groupe



PEGASE
HEALTHCARE
Accompagnateur du changement

Contact : d.verza@pegase-healthcare.com



Sommaire

Compte rendu rédigé par Thomas Rohban, médecin rédacteur indépendant.



OUVERTURE

Fabrice ANDRÉ (Gustave Roussy, Villejuif)

PAGE 5



KEYNOTE

DÉCRYPTER LA STRATÉGIE INTERMINISTÉRIELLE EN DONNÉES DE SANTÉ 2025 – 2028 ET SON INTÉGRATION VIS-À-VIS DU RÉGLEMENT EUROPÉEN

Modérateurs : **Fabrice BARLESI** (Gustave Roussy, Villejuif) et **Caroline GERMAIN** (AP-HP)

Intervenant : **Aymeric PERCHANT** (DNS)

PAGE 8



PITCH & TALK

EDS HOSPITALIERS : QUELS RETOURS D'EXPÉRIENCES EN 2025 ?

Modératrices : **Marie DE TAYRAC** (CHU de Rennes) et **Céleste LEBBÉ** (Hôpital Saint-Louis, Paris)

Intervenants : **Caroline GERMAIN** (AP-HP), **Flavien GILLES** (Institut Curie, Paris), **Claire MORGAND** (INCa)

PAGE 13



TABLE RONDE

PLACE ACTUELLE ET FUTURE DU RWD-RWE DANS L'ÉVALUATION DES MÉDICAMENTS ?

Modératrice : **Anne-Laure MARTIN** (Unicancer)

Intervenants : **Pablo BERLANGA** (Gustave Roussy, Villejuif), **Isabelle BORGET** (Gustave Roussy, Villejuif), **Pierre COCHAT** (HAS et Commission de la Transparence), **Pierre-Olivier FARENQ** (ANSM), **Ariane GALAUP PACI** (LEEM)

PAGE 17



TABLE RONDE

BEST PRACTICE SHARING : RWE SANS FRONTIÈRES !

Modérateur : **Laurent PEYRIN-BIROULET** (CHRU de Nancy)

Intervenants : **Emmanuel BACRY** (Health Data Hub), **Nora BENHABILES** (CEA), **Pierre-Henri BERTOYE** (Unicancer, CNRIPH, DGS), **Gilles PAUBERT** (Cegedim Health Data), **Pierre-Alexandre SQUARA** (Pfizer)

PAGE 22

Sommaire



TABLE RONDE

DONNÉES DE SANTÉ EN DEHORS DE L'HÔPITAL : EXEMPLE DE LA MÉDECINE DE VILLE

Modérateurs : **Olivier GUÉRIN** (CHU de Nice) et **Catherine RIOUFOL** (Hospices Civils de Lyon)
Intervenants : **Cécile BADIOLA LAGARDÈRE** (HEVA),
Mehdi DJELAMANI (Stane),
Olivier SAINT-LARY (CNGE, P4DP)

PAGE 26



CONTROVERSE

COHORTE OBSERVATIONNELLE, REGISTRE, EDS, BASE DE DONNÉES : QUEL EST LE BON CHOIX ?

Modérateurs : **Marc BOURLIÈRE** (Hôpital Saint Joseph, Marseille) et **Jean-Marc PHELIP** (CHU de Saint-Étienne)
Intervenants : **Anne-Sophie JANNOT** (BNDMR),
Nicolas PALEIRON (HIA Sainte-Anne, Toulon), **Grégoire REY** (UMS France Cohortes), **Xavier TROUSSARD** (RRHMBN)

PAGE 30



FLASH

QUID DE LA TRANSPORTABILITÉ DES DONNÉES : EXEMPLE

Modérateurs : **Christos CHOUAÏD** (CHI de Créteil) et **David PÉROL** (Centre Léon Bérard, Lyon)
Intervenants : **Blythe ADAMSON** (Flatiron Health),
Alison ANTOINE (Roche), **Mathieu ROBAIN** (Unicancer)

PAGE 34



TABLE RONDE

IMPACT DE L'IA SUR LA PERSONNALISATION DES TRAITEMENTS ET LE PARCOURS DE SOINS

Modérateurs : **Marco FIORINI** (FIAC)
Intervenants : **Philippe-Jean BOUSQUET** (CAD),
Franck MOUTHON (INSERM), **Pascal PUJOL** (SFMPP)

PAGE 38



TAKE-HOME MESSAGES & CONCLUSION

David PÉROL (Centre Léon Bérard, Lyon)

PAGE 42



Photos : © Eric Durand – Maquette : Offset 5



OUVERTURE



Dans son propos d'ouverture de la journée 360 RWE, 3^e édition, Fabrice André situe les *Real-World Data* (RWD) et la *Real-World Evidence* (RWE) comme un champ scientifique arrivé à un point de maturité. Longtemps abordées avec scepticisme par une communauté formée aux essais cliniques prospectifs et contrôlés, ces données issues de la pratique réelle sont désormais omniprésentes. Toutefois, leur potentiel reste largement sous-exploité en raison de verrous conceptuels, méthodologiques et organisationnels qui freinent leur impact réel en oncologie.

Le premier point fondamental concerne l'absence de terminologie partagée. L'hétérogénéité du langage médical, observable jusque dans la rédaction des comptes rendus cliniques, empêche la construction de bases de données exploitables à grande échelle. Cette instabilité sémantique constitue un obstacle majeur à l'usage de méthodes algorithmiques et d'intelligence artificielle (IA). Les initiatives

portées par les sociétés savantes, notamment l'ESMO, et les projets d'ontologie menés à Gustave Roussy visent à structurer des définitions consensuelles. Cette démarche apparaît indispensable pour permettre l'interopérabilité des données et la reproductibilité des analyses, tout en dépassant une dépendance excessive aux régulateurs dans la définition des concepts cliniques.

Le deuxième point concerne les modalités de collecte des données, en particulier l'émergence de la saisie automatisée. Les bases de données constituées à Gustave Roussy, regroupant plusieurs centaines de milliers de patients, servent de terrain d'expérimentation pour des outils d'IA capables d'extraire et structurer l'information à partir des dossiers médicaux. Les résultats montrent des performances globalement satisfaisantes mais inégales selon les types de variables, avec des taux d'erreurs non aléatoires, concentrés sur certains *clusters* comme les dates ou les événements complexes. Cette situation

pose la question, encore largement débattue, du seuil d'erreur acceptable pour qualifier une base de données comme utilisable à des fins scientifiques. L'absence de cadre explicite définissant ces seuils limite la confiance dans les bases existantes et freine leur intégration dans des projets de recherche de grande ampleur.

Le troisième point porte sur le *reporting* scientifique des études utilisant des RWD. Il existe une hétérogénéité majeure des méthodologies et des pratiques de publication, souvent masquée par l'usage générique du terme *Real-World Data*. Contrairement à d'autres domaines, comme la recherche sur les biomarqueurs, qui ont bénéficié de critères de standardisation tels que REMARK, les études RWE ne disposent pas encore de cadres largement adoptés. Cette absence de normalisation complique l'interprétation des résultats, empêche les comparaisons entre études et limite la possibilité de réaliser des méta-analyses fiables. Des initiatives existent au niveau des sociétés savantes, mais leur diffusion reste insuffisante au regard de la diversité et de l'ampleur des acteurs impliqués dans ce domaine.

Quant à l'accès aux données, malgré l'existence d'infrastructures performantes, les procédures d'autorisation multiples, les délais de validation et la lourdeur administrative découragent les porteurs de projets, en particulier les cliniciens investis dans la pratique quotidienne. À terme, la pertinence des investissements réalisés dans les infrastructures de données sera évaluée sur la base de leur capacité à générer des avancées médicales tangibles.

Parmi les cas d'usage, la constitution de bras de contrôle synthétiques et l'utilisation de contrôles historiques soulèvent des enjeux méthodologiques majeurs. En oncologie, les définitions des sous-types tumoraux, notamment fondées sur les biomarqueurs, évoluent rapidement. Les changements successifs dans la définition du statut HER2 illustrent le risque de comparer des populations qui ne sont pas équivalentes d'une période à l'autre. Cette instabilité compromet la validité des comparaisons rétrospectives et impose une vigilance accrue dans l'interprétation des résultats issus des RWD.

Les maladies rares constituent en revanche un champ d'application particulièrement prometteur pour les RWD. De même, les approches basées sur la similarité entre patients et la création de patients virtuels représentent une évolution importante de la prise de décision médicale. Cette transition vers une décision fondée sur la similarité, plutôt que sur



des résultats moyens issus d'essais de phase III, pose des questions majeures de validation scientifique et de régulation.

Les RWD peuvent également jouer un rôle dans la recherche exploratoire et la découverte scientifique. Ces approches nécessitent toutefois l'implication directe de chercheurs et une réflexion explicite sur les objectifs scientifiques poursuivis. Enfin, l'intégration des PRO (*patient-reported outcomes*) apparaît comme un enjeu incontournable. Les données issues des dossiers médicaux sous-estiment largement les toxicités et l'expérience quotidienne des patients. La collecte prospective, facilitée par les outils de suivi numérique à distance, est indispensable pour compléter les RWD traditionnelles et offrir une évaluation plus complète des bénéfices et des risques des traitements.

En conclusion, le champ des RWD doit désormais se structurer autour de cadres conceptuels solides, incluant l'ontologie, la qualité des données, la standardisation du *reporting* et l'accessibilité à l'exploitation scientifique. Le développement de nouveaux cas d'usage et la simplification des parcours d'accès aux données conditionneront la capacité des RWD à générer un impact réel et mesurable sur la prise en charge des patients en oncologie.

Fabrice ANDRÉ

(Directeur de la recherche –
Gustave Roussy, Villejuif)

Nous innovons chaque jour
pour proposer **des avancées qui
changent la vie des patients.**

Découvrez-en plus sur [Pfizer.fr](https://www.pfizer.fr)





KEYNOTE

Décrypter la stratégie interministérielle en données de santé 2025 – 2028 et son intégration vis-à-vis du règlement européen

Modérateurs : **Fabrice BARLESI** (Directeur général – Gustave Roussy, Villejuif) et **Caroline GERMAIN** (Directrice du pôle innovation et données – AP-HP)

INTERVENANT

Aymeric PERCHANT (Coordinateur national de la stratégie d'accélération digitale en santé – Délégation au numérique en santé (DNS))

INTRODUCTION



Fabrice BARLESI
(Gustave Roussy, Villejuif)



Caroline GERMAIN
(AP-HP)

La session consacrée à la stratégie interministérielle des données de santé 2025-2028 et son articulation avec le règlement européen sur l'Espace européen des données de santé réunissait Fabrice Barlesi, directeur général de Gustave Roussy, Caroline Germain, directrice du pôle innovation et données de l'AP-HP, et Aymeric Perchant, coordinateur national de la stratégie d'accélération digitale en santé à la Délégation au numérique en santé (DNS). L'objectif était de clarifier la vision française sur les données de santé et l'intelligence artificielle (IA), de préciser la place du futur règlement européen et de dégager les principaux chantiers opérationnels pour les années à venir.

D'emblée, les modérateurs ont rappelé la complexité du cadre dans lequel s'inscrit cette stratégie. Il s'agit à la fois d'un environnement réglementaire exigeant, d'un champ scientifique en évolution rapide et d'un enjeu organisationnel majeur pour les établissements et les équipes de recherche. Les attentes sont fortes : il faut rendre les données plus accessibles et plus utiles, sans affaiblir les garanties de sécurité et de protection de la vie privée. Dans



ce contexte, le règlement européen créant l'Espace européen des données de santé, adopté en 2024, joue un rôle structurant. Il prévoit une mise en œuvre graduelle sur environ trois à quatre ans, avec la possibilité pour un projet démarrant dans un État membre d'accéder, via un mécanisme unique, à des données provenant de plusieurs autres pays. La France doit donc non seulement adapter ses propres dispositifs, mais aussi s'assurer que ceux-ci soient compatibles avec ce futur environnement européen harmonisé.

DONNÉES DE SANTÉ – STRATÉGIE FRANCE



Aymeric PERCHANT
(DNS)

Aymeric Perchant a rappelé que la stratégie française a été élaborée sur la base d'un volume inhabituel de rapports et de contributions produits en 2023-2024 par la Cour des comptes, des missions gouvernementales, des fédérations professionnelles et de nombreux *think tanks*. Deux grands volets se dégagent : un volet « données de santé », déjà arbitré, qui définit les axes d'organisation, d'accès et de gouvernance des données ; et un volet

« Intelligence artificielle (IA) en santé », centré sur les usages et non sur la technologie, en phase de finalisation après une consultation publique ayant recueilli près de quatre-vingt-dix contributions. L'orientation générale est claire : il s'agit de fédérer l'ensemble des acteurs, incluant les établissements de santé, les chercheurs, industriels, et les autorités publiques, autour de cas d'usage concrets, plutôt que d'imposer d'en haut des solutions techniques déconnectées du terrain.

La question du positionnement de la France par rapport aux autres pays européens a été largement discutée. Tous les États membres n'avancent pas à la même vitesse et les régimes nationaux d'accès aux données sont très hétérogènes. Alors que certains pays ne disposent pas d'équivalents aux méthodologies de référence françaises permettant un accès par simple déclaration à la CNIL, d'autres ont des dispositifs plus centralisés ou plus libéraux. Dans ce paysage, la France dispose d'atouts scientifiques importants en matière de données, de mathématiques et d'IA, mais reste en retrait sur certains aspects technologiques et d'infrastructures souveraines. La question de la souveraineté ne peut pas être éludée : il serait difficilement acceptable de fonder les grandes décisions de santé publique sur des données extra-européennes, par exemple chinoises, sans maîtrise sur leur collecte ni sur leurs biais. Le règlement européen introduit d'ailleurs un

changement de paradigme fort : la donnée en elle-même n'a pas de valeur économique propre ; seule compte la valeur créée par ses usages. L'accès aux données devra se faire à un coût marginal, et les motifs de refus seront limités, par exemple à la protection de la propriété intellectuelle ou du secret des affaires. Cette approche, très différente de la logique de valorisation de la donnée prévalant en Amérique du Nord ou en Asie, doit permettre de développer la recherche et l'innovation sur des données françaises et européennes, au service du soin et de la santé publique.

SIMPLIFICATION DE L'ACCÈS ET RÔLE DES CLINIENS

Les échanges ont également permis de préciser les attentes des cliniciens vis-à-vis des usages secondaires des données de santé. Les études de vie réelle apparaissent essentielles pour comparer des stratégies thérapeutiques dans des contextes où aucun standard ne fait consensus, notamment en neuro-oncologie. Dans ces situations, les essais randomisés ne répondent qu'à une partie des questions, tandis que des études de population à l'échelle nationale sont nécessaires pour analyser des parcours de soins hétérogènes, évaluer l'efficacité en conditions réelles et mettre en évidence d'éventuelles disparités territoriales. Une vigilance particulière est toutefois requise face au risque de standardisation excessive, notamment en cancérologie, où les définitions de biomarqueurs, les classifications histologiques et les pratiques cliniques évoluent rapidement et diffèrent selon les spécialités. L'implication étroite des cliniciens dans la définition, la collecte et l'analyse des données est indispensable, d'autant que certaines informations clés, comme les données histologiques, ne sont pas présentes dans le SNDS, ce qui limite certaines analyses. Dans ce contexte, un appui sur les sociétés savantes apparaît pertinent pour définir, par spécialité, un socle minimal de données à collecter de manière prospective, possiblement articulé à des registres nationaux, notamment en cancérologie.

Cette approche s'inscrit dans une vision selon laquelle les cliniciens doivent être pleinement au centre de ces dispositifs, non comme simples parties prenantes, mais comme acteurs moteurs. La définition des variables pertinentes ne peut relever uniquement des administrations centrales ; elle doit partir des cas d'usage et des besoins du terrain, en particulier pour les usages secondaires des données. L'un des axes explicites de la stratégie consiste ainsi à travailler avec les communautés de soins et de recherche afin d'identifier, par spécialité, les données manquantes, de clarifier les objectifs poursuivis, puis de traiter de manière cohérente les enjeux d'interopérabilité, de collecte et, le cas échéant, de création de registres

dédiés. Une gouvernance spécifique est en cours de structuration pour garantir l'association systématique des expertises cliniques et scientifiques aux décisions structurantes.

Enfin, les limites actuelles du SNDS en matière de données véritablement cliniques ont été largement soulignées, en particulier pour la biologie, l'imagerie ou les examens réalisés en ville. Cette insuffisance explique le maintien, par de nombreuses équipes, d'observatoires et de registres spécifiques, faute de pouvoir s'appuyer sur les bases nationales. Le SNDS reste en effet centré sur des données médico-administratives, autorisant essentiellement des inférences cliniques. Le cadre réglementaire européen introduit une ambition plus large, en posant le principe d'un accès, à terme, à l'ensemble des données de santé pour des usages secondaires. Une partie importante de ces données cliniques est déjà présente dans le Dossier médical partagé (DMP) et « Mon espace santé », alimentés notamment par le programme Ségur du numérique, avec des comptes rendus médicaux, d'imagerie et, à terme, des documents cliniques de plus en plus structurés. Les données d'anatomopathologie et certaines données génomiques demeurent toutefois difficiles à intégrer, malgré les investissements consentis pour l'équipement des laboratoires ; des obstacles persistent en matière de conservation à long terme, de volumes de données et de formats.

Le calendrier européen prévoit qu'en mars 2029 les données du DMP devront être accessibles pour les usages de recherche autorisés. D'ici là, il faudra construire la trajectoire technique, juridique et organisationnelle rendant cet accès possible. Aujourd'hui, le droit français limite encore l'utilisation des données de « Mon espace santé » à des finalités de soin, pour protéger la confiance des citoyens et encourager l'alimentation de ces dossiers. Des évolutions législatives sont donc attendues, avec une attention particulière à la préservation de la confiance et au respect du droit d'opposition. L'*opt-out* prévu pour « Mon espace santé » a été conçu en cohérence avec celui du règlement européen sur l'Espace des données de santé : l'idée est de massifier les données disponibles tout en garantissant à chaque patient la possibilité de s'opposer à l'usage secondaire de ses données. À terme, les États membres qui ne respecteraient pas ces obligations d'accessibilité des données pourraient être exposés à des sanctions européennes.

QUALIFICATION DES DONNÉES

Les échanges ont également fait ressortir les enjeux liés à la qualité et à la qualification des données. Les niveaux d'exigence ne peuvent être uniformes pour l'ensemble des usages : les besoins diffèrent selon qu'il s'agisse du soin individuel, de la description

épidémiologique, de la constitution de registres ou de travaux de recherche exploratoires générant des hypothèses. Les attentes en matière de complétude, de précision et de structuration des données doivent donc être adaptées à chaque finalité. Or, cette logique de qualification différenciée des jeux de données demeure encore insuffisamment intégrée dans la réflexion relative au futur répertoire national des ensembles de données.

Dans ce contexte, des travaux sont engagés à l'échelle européenne pour élaborer un dispositif de labellisation de la qualité des données, fondé sur une qualification plus fine et plus explicite des niveaux de fiabilité et d'usage. Par ailleurs, l'ouverture accrue des données de santé s'accompagnera d'exigences renforcées de transparence, tant sur les autorisations d'accès accordées que sur les résultats des projets conduits à partir de ces données. Cette transparence permettra une évaluation *a posteriori* de la pertinence et de la robustesse des jeux de données mobilisés pour différents types de travaux.

Un point de tension majeur demeure toutefois entre le temps nécessaire à la labellisation, à la documentation et à la qualification rigoureuse des ensembles de données, et le besoin de disposer de cycles d'exploitation courts ainsi que de données récentes, en particulier pour les projets reposant sur l'IA. Une multiplication excessive des procédures de certification ferait courir le risque de ralentir les usages et de réduire la valeur opérationnelle des données. L'enjeu réside ainsi dans la recherche d'un équilibre permettant de garantir un socle minimal de qualité et de documentation, tout en assurant un accès suffisamment fluide et rapide pour soutenir la recherche et l'innovation.

La place de « Mon espace santé » et du DMP a également suscité de nombreuses interrogations, tant du point de vue des professionnels de santé que de celui des patients. Les questions ont porté sur la nature des données effectivement appelées à alimenter le DMP, sur le caractère obligatoire ou facultatif de cette alimentation, ainsi que sur les éventuelles incitations ou sanctions prévues en cas de non-contribution par les professionnels. Des incertitudes subsistent également quant aux données rendues accessibles à des fins de recherche et aux modalités d'information et de consentement des patients. Sur le plan réglementaire, les dispositions relatives aux obligations et aux mécanismes incitatifs ou contraignants relèvent du cadre législatif et, en particulier, des lois de financement de la sécurité sociale, dont les arbitrages demeurent en cours. S'agissant des usages de recherche, le principe retenu repose sur un droit d'opposition binaire : chaque citoyen peut accepter ou refuser l'utilisation de ses données de santé à des fins de recherche, sans possibilité de modulation projet

par projet ou pathologie par pathologie. Le cadre réglementaire n'impose pas une information individuelle détaillée sur chaque utilisation des données ; en revanche, les autorisations accordées et les grandes catégories de projets concernés feront l'objet d'une publication, garantissant un niveau de transparence collectif.

La mise en œuvre d'une traçabilité exhaustive permettant d'identifier, pour chaque individu, l'ensemble des projets ayant mobilisé ses données apparaît difficilement soutenable sur les plans technique et organisationnel. En parallèle, la possibilité de réidentifier un patient doit néanmoins être préservée lorsqu'une recherche met en évidence une information cliniquement significative le concernant. Cette exigence suppose une articulation maîtrisée entre les projets de recherche, les données pseudonymisées et l'intervention d'un professionnel de santé responsable de l'information et de l'accompagnement du patient. Elle renvoie à des enjeux majeurs de bioéthique et de gouvernance, qui devront être intégrés aux évolutions réglementaires et législatives à venir.

CRÉATION D'UN RÉPERTOIRE DE DONNÉES

Un autre pilier de la stratégie, largement commenté, est la création d'un répertoire national des ensembles de données. Exigé par le règlement européen et attendu pour 2029, ce répertoire ne vise pas à centraliser les données mais à référencer l'ensemble des jeux disponibles, leurs métadonnées, leurs niveaux de qualité et leurs conditions d'accès, via un point d'entrée unique. Le défi est important, car de nombreux entrepôts (AP-HP, Inserm, centres hospitaliers, centres de lutte contre le cancer, *clusters*, *Health Data Hub*, etc.) disposent déjà de leurs propres catalogues, reposant sur des formats parfois incompatibles. Les travaux actuels visent à harmoniser les schémas de description et les procédures de déclaration, sans imposer de redondances aux équipes qui gèrent ces entrepôts. Le futur Organisme responsable de l'accès aux données (ORAD) devrait s'appuyer sur l'évolution de la Plateforme des données de santé nationale, en lien étroit avec la CNIL. Un groupe de travail spécifique, rattaché au Conseil du numérique en santé, sera chargé de coordonner ces actions et d'éviter les doublons techniques ou financiers.

Ce répertoire constitue aussi un levier pour structurer un modèle économique durable autour des données de santé. Il ne s'agit pas de monétiser directement les données, mais de valoriser les services rendus par les détenteurs : accompagnement méthodologique, mise à disposition d'environnements sécurisés d'analyse, expertise clinique et statistique, etc. Ces services permettront aux entrepôts de faire vivre et d'enrichir leurs données, en complément des

financements publics et des financements européens. Ce modèle économique n'était pas encore stabilisé et constituait l'un des grands chantiers à venir, notamment pour trouver le bon équilibre entre financements publics, contributions privées et incitations adaptées.

En conclusion, Aymeric Perchant a souligné trois conditions majeures pour que la France soit bien positionnée dans l'utilisation des données de santé. La première concerne la gouvernance et l'efficacité des organismes responsables de l'accès aux données et des détenteurs de confiance : il s'agit de s'appuyer sur les structures existantes qui ont déjà démontré leur capacité à avancer, plutôt que de repartir de zéro dans une logique trop centralisatrice. La deuxième est précisément la construction d'un modèle économique soutenable dans le nouveau paradigme où la donnée en tant que telle n'est plus monétisée, ce qui impose de repenser les équilibres entre financement public, financement privé et valorisation des services. La troisième condition est la simplification, qu'il s'agisse des référentiels de sécurité, des méthodologies de référence ou des procédures d'autorisation, avec une confiance affirmée dans les dispositifs éthiques et réglementaires déjà en place dans les établissements et les entrepôts de données.

10 POINTS CLES

1. Stratégie française en deux volets : données de santé (gouvernance, accès) et IA en santé centrée sur les usages.
2. Le règlement européen EHDS comme cadre structurant avec un accès transfrontalier harmonisé aux données.
3. Changement de paradigme : la donnée n'a pas de valeur économique propre, seule la valeur des usages compte.
4. La France dispose d'atouts scientifiques forts mais reste fragile sur les infrastructures et la souveraineté technologique.
5. Simplifier l'accès aux données est une priorité pour accélérer recherche, innovation et études en vie réelle.
6. Les cliniciens sont au cœur des usages secondaires et doivent piloter la définition des données pertinentes.
7. Le SNDS reste limité à des données médico-administratives, insuffisantes pour certaines analyses cliniques.
8. « Mon espace santé » et le DMP deviendront centraux pour la recherche à l'horizon 2029, sous conditions réglementaires.
9. La qualité et la qualification des données doivent être adaptées aux finalités, avec une future labellisation européenne.
10. Création d'un répertoire national des données pour référencer les jeux existants et structurer un modèle économique durable.



EDS hospitaliers : quels retours d'expériences en 2025 ?

Modérateurs : **Marie DE TAYRAC** (Responsable de l'unité de bio-informatique médicale – CHU de Rennes) et **Céleste LEBBÉ** (Responsable du centre d'oncodermatologie – Hôpital Saint-Louis, Paris)

INTERVENANTS

Caroline GERMAIN (Directrice du pôle innovation et données – AP-HP) ; **Flavien GILLES** (Responsable de la Data Factory au sein de la direction des données – Institut Curie, Paris) ; **Claire MORGAND** (Directrice de la direction de l'observation des sciences des données et de l'évaluation – Institut national du cancer (INCa)).

INTRODUCTION

Les entrepôts de données de santé (EDS) sont devenus, en moins de dix ans, des infrastructures essentielles pour l'hôpital et la recherche. L'essor des méthodes d'intelligence artificielle depuis 2017 a catalysé ce mouvement, en donnant une valeur nouvelle à des volumes de données jusqu'ici peu exploitables. En cancérologie, cette dynamique s'inscrit dans un paysage déjà structuré par les registres, les cohortes et les bases nationales médico-administratives, et s'articule désormais autour de deux chantiers majeurs : le futur registre national des cancers et l'espace européen des données de santé.

Cet article croise les points de vue de trois types d'acteurs : un important EDS hospitalier généraliste (AP-HP), un centre de lutte contre le cancer doté d'une *Data Factory* (Institut Curie) et l'INCa, dont l'entrepôt préfigure le registre national. Les échanges mettent en évidence les convergences sur quatre enjeux : clarifier les finalités des EDS, organiser la qualité et l'interopérabilité, assurer la soutenabilité et construire une complémentarité opérationnelle avec registres et cohortes, notamment pour la recherche clinique.

ARCHITECTURES ET FINALITÉS DES ENTREPÔTS EN 2025



Caroline GERMAIN
(AP-HP)

L'EDS de l'AP-HP, autorisé par la CNIL en 2017, agrège les données issues des logiciels métiers de 38 hôpitaux, couvrant environ 19 millions de patients et près d'un milliard d'examen de biologie, des images, médicaments et bientôt un socle conséquent de données génomiques, de maladies rares et de physiologie. Une centaine de professionnels aux profils variés (médecins, ingénieurs, mathématiciens, juristes, *data scientists*...) travaillent à la fois sur l'architecture de l'entrepôt et sur le développement d'outils d'exploration et de modèles d'intelligence artificielle (IA). Après une phase de croissance rapide, l'enjeu est de stabiliser une offre de services industrialisée, capable de s'intégrer à l'espace européen des données de santé, autour de quatre priorités : enrichissement des données, fiabilité, adaptation des capacités de calcul et modèle économique soutenable.

À l'Institut Curie, la direction des données est organisée en deux pôles : la *Data Factory*, qui gère les flux et la structuration, et une équipe « Conseil et services » qui accompagne les projets et diffuse une culture de la donnée auprès des cliniciens. L'EDS, récemment autorisé, est encore en phase de mise en conformité. La priorité est de connecter l'ensemble des logiciels métiers, puis de transformer ces données hétérogènes en « fiches tumeur » standardisées, utiles pour la recherche. Les comptes rendus restent un gisement d'information majeur, pour lequel les techniques d'IA appliquées au langage clinique sont mobilisées. Un axe récent consiste à



utiliser l'entrepôt pour le pilotage opérationnel du soin (parcours, délais, volumes d'activité), ce qui renforce le dialogue avec les services et améliore la qualité de la donnée à la source.



Claire MORGAND
(INCa)

L'INCa, enfin, n'a pas de mission de soins et se situe d'emblée dans une logique de recherche, principalement épidémiologique. Son entrepôt, construit dès 2010 autour de la « cohorte cancer », agrège les cas prévalents et incidents de cancer issus du SNDS (environ 12 millions de personnes), enrichis par des appariements indirects avec les registres de cancers (taux d'appariement élevé) et, à terme, par des données biologiques (anapath, génétique, moléculaire). Cet entrepôt, autorisé par la CNIL en 2019, constitue la base technique du futur registre national des cancers instauré par la loi du 30 juin 2025.

QUALITÉ, STANDARDISATION ET INTEROPÉRABILITÉ



Flavien GILLES
(Institut Curie, Paris)

Un consensus de ces échanges se dégage sur la nécessité des standards pour tout projet de fédération d'EDS. Dans le domaine de l'oncologie, des modèles comme OSIRIS-RWD se développent, tandis qu'OMOP et FHIR s'imposent comme références internationales. Cependant, il est rappelé qu'un modèle de données n'est pas synonyme de qualité : on peut remplir correctement un standard avec des données erronées, incomplètes ou mal contextualisées. L'AP-HP a ainsi expérimenté plusieurs modèles en parallèle avant de recentrer son effort sur une « base centrale » FHIR, considérée comme suffisamment générique pour documenter à la fois la sémantique et l'interopérabilité technique. Des couches supplémentaires (OMOP, OSIRIS) sont appliquées au cas par cas, selon les projets. À l'Institut

Curie, la logique est similaire : la standardisation est pensée à partir des usages (fiches tumeur, fédérations d'entrepôts) plutôt qu'abstraitemment.

Le contenu des modèles reste le point critique. Dans les EDS hospitaliers, une grande partie de l'information est encore portée par les comptes rendus ou par des formulaires très hétérogènes. L'exemple d'ORBIS à l'AP-HP (22 modules, 15 000 formulaires) illustre la complexité d'un système dans lequel chaque service a développé ses propres habitudes de saisie. La qualification de ces données suppose un travail patient de rétro-ingénierie, qui ne peut se faire qu'en associant étroitement médecins et ingénieurs.

Les éditeurs de logiciels métiers occupent ici une place stratégique. Sans structuration minimale « en entrée », tout doit être reconstruit « en sortie », ce qui renchérit fortement le coût des EDS. Or l'évolution des logiciels pour intégrer des modèles communs, par exemple en anatomopathologie, représente un investissement important, particulièrement en ville où sont réalisés la majorité des examens d'anapath. La définition de modèles nationaux ou européens robustes, accompagnés d'outils et de financements adaptés, apparaît comme un levier majeur pour améliorer la qualité en amont.

SOUTENABILITÉ, GOUVERNANCE ET MAILLAGE

Les trois structures insistent sur les questions de soutenabilité. Les coûts initiaux de mise en place des EDS sont élevés, puis décroissent, mais les besoins en ressources humaines spécialisées demeurent importants. L'INCa fonctionne avec une petite équipe (22 personnes) aux profils très qualifiés, ce qui limite la vitesse de montée en charge. À l'AP-HP ou à l'institut Curie, la compétition pour attirer et retenir des compétences rares (data, IA, cybersécurité) est forte, dans un contexte de contraintes budgétaires.

La gouvernance progresse vers plus de structuration. À l'INCa, la préparation du registre national conduit à formaliser des comités d'utilisateurs et de producteurs de données, de façon à organiser les retours d'expérience sur la qualité : les équipes de recherche signalent les incohérences, l'entrepôt ajuste (correction, retrait de variables) et documente. Des dispositifs analogues émergent dans les EDS hospitaliers, via des réseaux de *data scientists* et des comités associant la direction des systèmes d'information (DSI), cliniciens et chercheurs.

Au-delà des établissements, un maillage s'organise à l'échelle nationale. Les CHU se coordonnent au sein d'une commission « data » et du projet « Fait Data » pour mutualiser bonnes pratiques, infrastructures et modèles de services. Les dispositifs nationaux maladies rares (BaMaRa, Banque nationale de données maladies rares, plateformes AURAGEN et SeqOIA)

illustrent un autre type de maillage, plus thématique, mais fortement connecté aux réseaux européens. L'espace européen des données de santé, enfin, impose d'anticiper des standards d'interopérabilité et des identifiants communs, dans une perspective de compétitivité internationale.

EDS, REGISTRES ET COHORTES : UN ÉCOSYSTÈME À ARTICULER

Les échanges ont mis en lumière les tensions mais aussi la complémentarité entre EDS, registres et cohortes. Du côté des registres de cancers, trois difficultés majeures sont rappelées : une granularité variable selon qu'ils couvrent tous les cancers ou une localisation spécifique ; des délais importants de disponibilité des données (jusqu'à quatre ans), liés au recueil manuel et aux incertitudes réglementaires post-RGPD ; et l'absence historique de base légale explicite pour leur activité, qui a conduit certains établissements à restreindre les transmissions.

La loi sur le registre national des cancers apporte une réponse en donnant un fondement législatif à ces activités et en organisant un système d'information commun, piloté par l'INCa et FRANCIM. L'idée est de combiner une remontée automatisée à l'échelle nationale avec un recueil local ciblé pour compléter et valider la donnée. Les registres resteront les « gold standards » de qualité et d'exhaustivité sur leurs territoires, tandis que l'entrepôt de l'INCa offrira une couverture plus large, même si elle ne sera jamais totalement exhaustive. Les cohortes prospectives (MELBASE, CANTO, ESME, etc.) apportent, elles, une donnée de grande qualité, recueillie dans des cahiers d'observation structurés pour des objectifs scientifiques précis. Leur limite tient à la taille des effectifs et à la représentativité. À l'inverse, les EDS offrent des données plus massives et plus proches de la « vraie vie », mais de qualité variable.

D'où l'idée, largement partagée, de ne pas opposer ces dispositifs. Les EDS peuvent servir à explorer des signaux faibles, décrire des parcours, estimer des effectifs, préparer des protocoles ou alimenter partiellement des cahiers d'observations électroniques (eCRF). Les registres et cohortes fournissent des références pour valider les algorithmes issus des EDS et mener des études exigeant une forte précision. Certains travaux à venir viseront à confronter des modèles développés dans les EDS à des données de cohorte pour en mesurer la robustesse. L'articulation entre EDS, registres et cohortes est devenue essentielle pour structurer la donnée en cancérologie. Chaque dispositif apporte une valeur spécifique, mais aussi des limites qu'il faut combiner intelligemment plutôt que confronter.

Les registres de cancers rencontrent trois défis majeurs : une granularité hétérogène selon leur périmètre (généraux *versus* spécialisés), des délais importants

liés à un recueil encore largement manuel, et l'absence historique de fondement légal qui avait freiné certaines transmissions. Le registre national des cancers apporte une réponse en sécurisant juridiquement l'activité et en organisant un système d'information unifié. L'objectif est d'associer une remontée automatisée nationale à un recueil local ciblé, les registres locaux demeurant les « gold standards » de qualité, tandis que l'entrepôt de l'INCa assure une couverture plus large, sans viser l'exhaustivité.

LES EDS SONT-ILS PRÊTS POUR LA RECHERCHE CLINIQUE ?

Une question clé concerne la capacité réelle des EDS à soutenir aujourd'hui la recherche clinique comparative, notamment pour évaluer l'efficacité des traitements. Les exigences méthodologiques dans ce domaine sont élevées : définition précise des variables, temporalité fine, complétude quasi totale et critères de jugement complexes. Or, dans la plupart des EDS, une part substantielle de l'information reste enfouie dans les comptes rendus, avec des pratiques de saisie hétérogènes. Les outils d'IA de structuration progressent mais demeurent dépendants du contexte local et ne garantissent pas encore une qualité suffisante. Ainsi, les EDS ne peuvent pas, à ce stade, se substituer aux dispositifs conçus spécifiquement pour la recherche clinique (essais, cohortes structurées). Ils apportent toutefois une valeur importante : génération et priorisation d'hypothèses, analyses descriptives à large échelle, documentation des parcours de soins et des inégalités, études de faisabilité et contribution partielle au pré-remplissage d'eCRF lorsque la donnée est bien structurée.

Pour renforcer leur rôle dans la recherche clinique, plusieurs leviers sont identifiés : développement de modèles de données par pathologie, plus proches des CRF ; amélioration et évaluation systématique des algorithmes de structuration sur des jeux de vérité issus de registres et de cohortes ; meilleure structuration des informations à la source ; et mise en place de protocoles hybrides combinant EDS, registres et cohortes afin de tirer parti des forces de chacun.

En conclusion, les retours d'expérience de 2025 montrent que les EDS ont franchi un cap : ils ne sont plus seulement des projets techniques, mais des infrastructures structurantes pour l'hôpital, la recherche et la régulation en cancérologie. Leur potentiel ne se mesure pas uniquement à la masse de données collectées, mais à la façon dont ils s'insèrent dans un écosystème plus large et dont ils répondent à des cas d'usage bien définis.

Quatre messages se dégagent. D'abord, les finalités doivent être explicites : pilotage, recherche épidémiologique, appui à la décision clinique, préparation d'essais, etc., car c'est l'usage qui doit guider

les choix de modèles et d'investissements. Ensuite, la qualité et l'interopérabilité des données doivent être pensées sur toute la chaîne, des logiciels métiers aux projets de fédération européenne, en associant étroitement éditeurs, cliniciens et ingénieurs. Troisièmement, la soutenabilité – financière, technique et humaine – impose des choix collectifs, des mutualisations et des modèles économiques clairs. Enfin, les EDS n'ont de sens que dans une complémentarité assumée avec les registres et les cohortes, en particulier pour la recherche clinique. L'enjeu des prochaines années sera de transformer cette mosaïque en un système cohérent, gouverné, soutenable et utile, au service de la connaissance, de l'innovation et, *in fine*, de la prise en charge des patients atteints de cancer.

10 POINTS CLÉS

1. Les EDS sont devenus des infrastructures stratégiques pour l'hôpital et la recherche, portées par l'essor de l'IA et leur articulation avec le registre national des cancers et l'espace européen des données de santé.
2. Les finalités des EDS se différencient selon les acteurs, combinant recherche, innovation et pilotage du soin pour les hôpitaux, et usages épidémiologiques et réglementaires pour l'INCa.
3. Les architectures des EDS entrent dans une phase de stabilisation après une expansion rapide, avec un focus sur l'industrialisation des services, la fiabilité des données et l'intégration européenne.
4. La standardisation des données est nécessaire mais insuffisante, car les standards (FHIR, OMOP, OSIRIS) ne garantissent pas à eux seuls la qualité clinique des données.
5. Les comptes rendus cliniques constituent à la fois le principal gisement d'informations et la principale difficulté, nécessitant un important travail de structuration appuyé par l'IA.
6. La structuration des données à la source est identifiée comme un levier critique pour limiter les surcoûts des EDS et améliorer durablement la qualité.
7. La soutenabilité humaine et financière des EDS reste fragile, du fait des coûts élevés et de la tension sur les compétences spécialisées en données et IA.
8. Une gouvernance plus structurée et collaborative se met en place, associant cliniciens, DSI, chercheurs et réseaux nationaux pour améliorer la qualité et la cohérence des données.
9. Les EDS, les registres et les cohortes doivent être articulés de manière complémentaire, chacun apportant des forces spécifiques et répondant à des limites différentes.
10. Les EDS ne se substituent pas encore à la recherche clinique comparative, mais renforcent la génération d'hypothèses, l'analyse des parcours et la préparation des essais.



TABLE RONDE

Place actuelle et future du RWD-RWE dans l'évaluation des médicaments ?

Modératrice : **Anne-Laure MARTIN** (Directrice des Data et Partenariats – Unicancer)

INTERVENANTS

Pablo BERLANGA (Responsable de la cohorte SACHA – Gustave Roussy, Villejuif) ; **Isabelle BORGET** (PU-PH en économie de la santé – Gustave Roussy, Villejuif) ; **Pierre COCHAT** (Membre du Collège de la HAS et Président de la Commission de la Transparence) ; **Pierre-Olivier FARENQ** (Directeur Europe et Innovation – Agence nationale de sécurité du médicament et des produits de santé (ANSM)) ; **Ariane GALAUP PACI** (Directrice de la recherche clinique – Les entreprises du médicament (LEEM)).

INTRODUCTION



Anne-Laure MARTIN
(Unicancer)

Depuis une dizaine d'années, la France a profondément structuré l'accès précoce aux innovations thérapeutiques et la réutilisation des données de vie réelle. Les dispositifs d'accès dérogatoire (accès précoce, accès compassionnel, cadres de prescription compassionnelle) offrent aux patients des traitements innovants avant leur remboursement en droit commun, tout en générant des données d'utilisation, d'efficacité et de sécurité.

Pourtant, malgré un cadre perçu comme attractif au niveau international, le potentiel des RWD et de RWE reste encore sous-exploité dans l'évaluation des médicaments. Les débats réunissant représentants de l'ANSM, de la HAS, du LEEM, d'Unicancer et des acteurs académiques montrent à la fois l'enthousiasme, les limites actuelles et les pistes

très concrètes de progrès. Les échanges se sont articulés autour de trois thématiques : le bilan des dispositifs d'accès précoce et les limites du recueil actuel ; la place de la RWE dans les méthodologies d'évaluation ; les initiatives structurantes, dont les projets méthodologiques nationaux et un exemple de registre pédiatrique.

L'ACCÈS PRÉCOCE : UN PARI RÉUSSI POUR LES PATIENTS, ENCORE FRAGILE POUR LA DONNÉE



Pablo BERLANGA
(Gustave Roussy, Villejuif)

L'accès précoce, outil essentiel pour permettre à des patients en impasse thérapeutique de bénéficier rapidement de traitements innovants, constitue un facteur d'attractivité important pour l'écosystème français. De manière plus large, il s'inscrit dans un ensemble de dispositifs dérogatoires créés pour soutenir l'innovation à l'interface entre recherche clinique et soins courants. Lorsqu'une molécule entre dans un dispositif d'accès précoce, les données disponibles sont encore insuffisantes pour un passage direct en droit commun. Le dispositif repose donc sur une logique de transition : l'accès est accordé sur la base d'un bénéfice clinique présumé, tandis que la phase d'accès précoce doit permettre d'enrichir la connaissance grâce à un recueil organisé de données en vie réelle.

Ce recueil repose principalement sur les protocoles d'utilisation thérapeutique et de recueil de données (PUT-RD), conçus comme un garde-fou accompagnant le plan de développement du médicament. Avec plusieurs années de recul, le bilan de ces PUT-RD apparaît mitigé : exhaustivité insuffisante, qualité variable selon les centres, grande hétérogénéité entre produits et plateformes de recueil. Les informations réunies sont



utiles pour documenter la tolérance, suivre l'usage ou comprendre les situations où un accès précoce fonctionne moins bien que prévu. En revanche, elles ont très rarement apporté des éléments susceptibles d'influencer la valeur du service médical rendu (SMR) ou de l'amélioration du service médical rendu (ASMR) lors du passage en droit commun.

Du côté de la sécurité, les agences rappellent que les données de vie réelle existent depuis longtemps, notamment via les réseaux de pharmacovigilance. Le paysage actuel est cependant marqué par une fragmentation des outils et une surcharge de saisies : un même prescripteur peut devoir renseigner successivement le dossier médical, un registre de pathologie, puis une plateforme dédiée au dispositif d'accès précoce. Cette multiplicité des entrées entraîne une perte d'information et une fatigue opérationnelle qui nuisent à la qualité du recueil.

VERS UN RECUEIL DE DONNÉES PLUS CIBLÉ, STANDARDISÉ ET SOUTENABLE



Isabelle BORGET
(Gustave Roussy, Villejuif)

L'un des messages les plus récurrents concerne la nécessité de définir avec précision ce qui doit être recueilli, et pour quelle finalité. Certains PUT-RD se sont alourdis au fil du temps, cherchant à répondre simultanément à des objectifs d'efficacité, de sécurité, de qualité de vie, d'équité d'accès ou encore d'aide au passage en droit commun. Ce cumul rend les cahiers de recueil trop complexes, alors même

que les équipes hospitalières doivent assumer cette charge sans ressources dédiées.

Face à ce constat, plusieurs pistes d'amélioration convergentes ont été évoquées. La première est d'établir un socle commun de variables minimales, identique pour tous les PUT-RD, recentré sur les informations réellement utiles aux autorités. Des modules additionnels pourraient être ajoutés au cas par cas, uniquement lorsqu'une incertitude spécifique l'exige. La deuxième piste est de mobiliser davantage les données hospitalières existantes : les entrepôts de données, les logiciels de chimiothérapie et les systèmes de pharmacie ou encore le SNDS contiennent déjà de nombreuses informations structurées. À terme, des extractions automatisées permettraient d'éviter la ressaisie. La troisième piste vise à standardiser les outils de recueil ; la coexistence de multiples plateformes, chacune développée par un industriel, étant un frein majeur. L'idée d'une plateforme nationale unique, garantissant l'interopérabilité et l'homogénéité des données, fait consensus. Et la dernière piste serait de valoriser et professionnaliser l'activité de recueil, étant donné que sans reconnaissance financière ou organisationnelle, la qualité restera limitée. L'idée de flécher une partie des financements hospitaliers existants vers cette activité est régulièrement avancée.

Ainsi, il ne faut pas attendre des PUT-RD ce qu'ils ne peuvent pas fournir. Même optimisés, ils ne remplaceront pas les études post-inscription conçues spécifiquement pour répondre à une question clinique ou méthodologique précise. Leur rôle se situe davantage dans la description de l'usage réel, le repérage des signaux de sécurité et l'éclairage du fonctionnement du dispositif.

PLACE DE LA RWE DANS LES MÉTHODOLOGIES D'ÉVALUATION : CONDITIONS, LIMITES ET ATTENTES



Pierre COCHAT
(HAS)

Au-delà des accès précoces, la place de la RWE dans l'évaluation des médicaments dépend de la qualité méthodologique des études observationnelles. L'accompagnement proposé par les autorités pour les projets d'innovation, en particulier via des guichets scientifiques dédiés, vise à encourager une réflexion précoce sur le rôle des données de vie réelle. L'idée est d'anticiper les choix méthodologiques, de définir les sources mobilisables et d'identifier les scénarios où la RWE peut réellement compléter ou renforcer l'évidence issue des essais cliniques. Du côté de la HAS, l'essai clinique randomisé reste le standard de référence chaque fois qu'il est faisable, y compris dans nombre de maladies rares (en dehors des situations d'ultra-rareté où des études monobras peuvent être acceptées).

Les comparaisons indirectes et les bras synthétiques sont recevables, mais leur niveau de preuve est, par nature, inférieur ; ils ne peuvent donc être valorisés que lorsqu'ils respectent une exigence méthodologique élevée. Au niveau européen, des initiatives comme le projet DARWIN, piloté par l'EMA, permettent de mutualiser des bases qualifiées provenant de plusieurs pays, dont la France. Ces infrastructures facilitent déjà certains travaux sur l'usage et pourraient, à terme, supporter davantage d'analyses d'efficacité et de sécurité.

Les autorités nationales soulignent toutefois que la montée en puissance de la RWE ne doit pas conduire à des attentes irréalistes. Les données de vie réelle ne peuvent être pleinement utiles que si elles sont anticipées dès le plan de développement et conçues avec un degré de rigueur comparable à celui des essais cliniques. Les travaux méthodologiques récents, notamment ceux publiés dans le cadre des ateliers de Giens, proposent des critères d'acceptabilité qui constituent aujourd'hui une référence (Cucherat *et al.*, 2025). Pour la démonstration d'efficacité à partir de données observationnelles, le consensus recommande d'élaborer un protocole rigoureux avec une question scientifique clairement formulée ; d'adopter une logique d'inférence causale explicite (définition d'estimands, stratégie de gestion des données manquantes, choix précoce de la population d'analyse) ; d'anticiper les biais propres au contexte observationnel (confusion, sélection, changements de pratique) et de prévoir des méthodes d'ajustement adaptées (score de

propension, appariement, pondération, etc.) ; de maîtriser le risque alpha, souvent mal contrôlé dans les études observationnelles multipliant les analyses exploratoires. Enfin, une transparence totale dans l'analyse et le *reporting* est de règle.

Du côté des industriels, les analyses réalisées par le LEEM confirment que l'essai randomisé reste la voie royale : la quasi-totalité des essais contrôlés, en ouvert ou en double aveugle, sont intégrés par la HAS dans ses évaluations. Lorsque des essais non randomisés sont bien justifiés, ils peuvent également être pris en compte dans une proportion significative. En revanche, les études reposant sur des méthodologies alternatives au sens strict de la doctrine (émulations d'essais, bras synthétiques, comparaisons indirectes complexes) ne sont aujourd'hui acceptées que de façon marginale. En effet, les agences constatent encore que peu de dossiers reposant sur ces méthodologies sont de bonne qualité. Il en résulte un cercle vicieux : la HAS souhaiterait disposer de davantage de dossiers bien construits pour affiner sa doctrine, tandis que les industriels hésitent à investir dans des approches qu'ils perçoivent comme peu valorisées. D'où la demande récurrente des acteurs d'un « mode d'emploi » pragmatique, précisant dans quels contextes les RWE peuvent réellement peser sur l'évaluation et sous quelles conditions méthodologiques minimales.

REGISTRES STRUCTURES ET PROJETS MÉTHODOLOGIQUES : EXEMPLES ET PERSPECTIVES



Ariane GALAUP PACI
(LEEM)

Dans cet environnement en transition, la cohorte pédiatrique SACHA constitue un exemple de RWE dont la qualité méthodologique permet une utilisation effective par les autorités. SACHA répond à une problématique bien définie : documenter l'utilisation et la tolérance des médicaments innovants administrés aux enfants atteints de cancer en dehors des essais cliniques (hors-AMM, accès compassionnel, accès précoce). Le registre repose sur un recueil prospectif harmonisé, conduit dans l'ensemble des centres d'oncologie pédiatrique à l'échelle nationale, avec un contrôle centralisé de la cohérence des données et un volet de pharmacovigilance active. Ce type de cohorte montre que la production d'une donnée exploitable repose avant tout sur une gouvernance solide, des objectifs clairs et une implication forte des équipes cliniques.

L'originalité du registre SACHA tient à son point d'entrée centré sur le médicament, et non uniquement



sur la pathologie. Cette structuration rend les données immédiatement mobilisables dans les évaluations relatives à des produits donnés. En quelques années, plus d'un millier de prescriptions ont été intégrées, avec des taux de couverture élevés. La cohorte est désormais explicitement reconnue par la HAS et l'ANSM comme l'une des rares sources de données françaises répondant aux standards attendus pour une RWE exploitable. Son essaimage international (Royaume-Uni, Italie) témoigne de la pertinence du modèle et de la possibilité de transposer ce type de registre dans d'autres domaines, sous réserve d'une gouvernance forte, d'un cahier de données ciblé et d'un partenariat étroit avec les autorités.

Parallèlement, un appel à projets national dédié aux nouvelles méthodologies soutient actuellement des travaux portant sur l'émulation d'essais cliniques et l'usage de données artificielles. Les projets retenus bénéficieront d'un accompagnement structuré des différentes autorités. Les résultats attendus à l'horizon 2028 devraient permettre d'identifier, sur la base de cas d'usage concrets, les situations dans lesquelles la RWE apporte une réelle plus-value et les limites à ne pas dépasser.

En conclusion, les discussions mettent en lumière un paysage contrasté. D'un côté, la France dispose d'atouts considérables : dispositifs d'accès précoce performants, bases de données de santé étendues, réseaux hospitaliers structurés et dynamique méthodologique réelle. De l'autre, l'utilisation opérationnelle de la RWE dans l'évaluation des médicaments reste limitée, en raison d'une préparation insuffisante, d'outils hétérogènes et d'un recueil encore trop lourd pour les équipes de terrain.

Trois orientations se dégagent clairement : anticiper, standardiser, renforcer la rigueur. Le défi des prochaines années sera de transformer un potentiel conceptuel en une pratique robuste, régulière et utile à la décision publique.

10 POINTS CLÉS

1. La France dispose d'un cadre d'accès précoce attractif, mais la valeur des données générées reste encore peu exploitée dans l'évaluation des médicaments.
2. Les PUT-RD apportent des informations utiles sur l'usage et la tolérance, mais influencent rarement le SMR ou l'ASMR lors du passage en droit commun.
3. La charge de recueil des données est aujourd'hui excessive et fragmentée, entraînant ressaisies multiples, fatigue des équipes et perte de qualité.
4. Un recentrage sur un socle minimal de données réellement utiles est indispensable, complété par des modules ciblés selon les incertitudes.
5. Les données déjà disponibles dans les EDS hospitaliers, le SNDS et les logiciels métiers doivent être prioritairement exploitées via des extractions automatisées.
6. La standardisation des outils de recueil constitue un levier majeur, avec un consensus autour de l'idée d'une plateforme nationale unique.
7. La RWE ne se substitue pas à l'essai clinique randomisé, qui reste le standard dès qu'il est faisable, mais peut le compléter dans des contextes précis.
8. L'acceptabilité de la RWE repose sur une forte exigence méthodologique : protocole rigoureux, inférence causale explicite, maîtrise des biais et transparence.
9. Un cercle vicieux persiste entre les attentes des agences et l'investissement des industriels, faute de doctrine opérationnelle suffisamment stabilisée.
10. Des registres structurés comme la cohorte SACHA démontrent qu'une RWE bien conçue peut être reconnue et réellement utilisée par les autorités.

L'INFORMATION AU CŒUR DE VOS PRATIQUES

My ONCO'NEWS

AIRES THÉRAPEUTIQUES :



Gastro-
intestinal



Génito-
urinaire



Séno-
gynéco



Thoracique



Médecine
de précision



DÉCOUVREZ DÈS MAINTENANT
VOTRE FIL D'ACTUALITÉS
EN ONCOLOGIE

www.tribunek-onco.fr/myonconews/

NOS FORMATS



**FIGURES
COMMENTÉES**



**SÉLECTION
BIBLIO**



**INTERVIEWS
DÉBATS**



QUIZ



ÉMISSION



**SOUS L'ÉGIDE
D'UN COMITÉ
SCIENTIFIQUE**



**Marine
GROSS-GOUPIL**
Bordeaux



**Emmanuelle
SAMALIN SCALZI**
Montpellier



**Elsa
KALBACHER**
Clermont-Ferrand



**Isabelle
SOUBEYRAN**
Bordeaux



**Delphine
LOIRAT**
Paris



**Aurélie
SWALDUZ**
Lyon

Édité par



KEPHREN PUBLISHING est amenée à traiter vos données personnelles. Vous attestez avoir pris connaissance de notre charte de protection des données personnelles – data privacy <https://www.pegase-healthcare.com/data-privacy/> et consentez au traitement de vos données personnelles par les sociétés du groupe Pegase Healthcare, conformément à ladite charte. Nous vous précisons que vous disposez d'un droit d'accès, de rectifications, de suppression et de portabilité des informations vous concernant ainsi que d'un droit d'opposition ainsi que d'un droit de limitation du traitement que vous pouvez exercer en nous contactant à l'adresse suivante : dataprivacy@pegase-healthcare.com. Conformément à la loi « Informatique et libertés » du 6 janvier 1978 modifiée par la loi n°2018-493 du 20 juin 2018 relative à la protection des données personnelles, vous disposez d'un droit d'accès, de rectification, de suppression, de limitation au traitement de vos données, de définir des directives sur le sort des données vous concernant, de portabilité de vos données vers un prestataire tiers. Vous pouvez en outre vous opposer à tout moment au traitement de vos données personnelles. Pour ce faire, vous pouvez faire valoir vos droits en adressant un mail à : dataprivacy@pegase-healthcare.com et en joignant une pièce d'identité. KEPHREN PUBLISHING, société par actions simplifiée au capital de 10.000 euros, sise 86-90 Rue du Dôme – 92100 Boulogne-Billancourt, immatriculée au registre du commerce et des sociétés de Nanterre sous le numéro 750 538 084.



TABLE RONDE

Best practice sharing : RWE sans frontières !

Modérateur : Laurent PEYRIN-BIROULET (Gastro-entérologue – CHRU de Nancy)

INTERVENANTS

Emmanuel BACRY (Représentant du Health Data Hub – European data innovation board) ; **Nora BENHABILES** (Directrice adjointe DRF en charge de l'innovation et de la valorisation – CEA) ; **Pierre-Henri BERTOYE** (Ex-directeur de la prospective réglementaire, des référentiels et de la formation (Unicancer), ex-président (Commission Nationale des Recherches Impliquant la Personne Humaine – CNRIPH) et ex-chargé de mission (Direction Générale de la Santé – DGS)) ; **Gilles PAUBERT** (Directeur de l'international – Cegedim Health Data) ; **Pierre-Alexandre SQUARA** (Scientifique RWE en hématologie – Pfizer).

taires. L'enjeu central, rappelé d'emblée, est de dépasser une approche strictement nationale afin de comparer, reproduire et valoriser les résultats issus de différents systèmes de santé, et d'inscrire le RWE dans une logique européenne, voire globale. Après une introduction des intervenants issus des milieux académique, institutionnel, industriel et des plateformes de données, les échanges se sont structurés autour de quatre axes : la cartographie des initiatives européennes, le rôle structurant des plateformes et de l'Espace européen des données de santé (EHDS), les applications du RWE pour la R&D et les essais cliniques, et enfin les freins et leviers conditionnant un *leadership* européen durable.

CARTOGRAPHIE EUROPÉENNE DES PROJETS EN DONNÉES DE SANTÉ

INTRODUCTION



Laurent PEYRIN-BIROULET
(CHRU de Nancy)

La session « *Best practice sharing* : RWE sans frontières » a été consacrée aux conditions nécessaires à un usage véritablement transnational des données de vie réelle (RWE) en santé. Elle s'est appuyée sur les dynamiques européennes actuelles, le rôle croissant des plateformes de données, des cas d'usage concrets pour la recherche et le développement pharmaceutiques, ainsi que sur l'évolution des cadres réglemen-



Emmanuel BACRY
(Health Data Hub)

Il est primordial de cartographier de manière systématique les projets européens faisant usage de données de santé, en particulier dans le champ de la santé digitale. L'hétérogénéité des systèmes de santé européens, incluant plus de 27 systèmes, sans véritable marché unique, se traduit par un foisonnement de projets, de standards et de méthodes, rarement mis en regard de façon globale. Les financements européens, dispersés entre différents *clusters* (programmes de recherche fondamentale, impacts sociétaux, partenariats publics-privés, initiatives ciblant les



pays moins innovants), ont laissé une large liberté aux scientifiques pour proposer leurs propres approches, standards et modes d'utilisation des données. Cette liberté, propice à l'innovation, rend en retour difficile l'établissement d'un état de l'art consolidé.

Le travail engagé vise à recenser les projets financés entre 2020 et 2025 en analysant plusieurs dimensions : origine géographique et type d'acteurs impliqués (hôpitaux, universités, industriels, PME), catégories de données utilisées (données cliniques, cohortes, registres, données numériques), standards et modèles de données mobilisés, méthodes analytiques et *endpoints*, ainsi que les liens avec l'industrie et les perspectives de valorisation. L'analyse porte également sur les impacts des cadres réglementaires nationaux sur l'accès au marché et la diffusion des innovations. Cette démarche, fondée sur l'exploitation structurée des bases de données européennes et conduite en partenariat entre académiques et industriels, permet d'identifier à la fois des modèles de réussite scientifique et économique, et des zones de fragilité : projets restés au stade de preuve de concept, absence de brevets ou de modèles économiques, ou innovations ne parvenant pas jusqu'au patient. La question de la saturation des espaces de propriété intellectuelle par des acteurs non européens est également intégrée, car elle conditionne les capacités futures d'exploitation.

L'objectif est de limiter la multiplication de projets financés sans débouchés opérationnels, en favorisant la réutilisation de ressources existantes, la constitution de nouveaux consortiums fondés sur des briques déjà dérisquées (standards, méthodes, infrastructures), et la visibilité des meilleures innovations européennes. Cette cartographie constitue ainsi les bases d'un observatoire partagé de l'innovation en données de santé.

PLATEFORMES NATIONALES ET ESPACE EUROPÉEN DES DONNÉES DE SANTÉ



Nora BENVHABILES
(CEA)

Les plateformes nationales d'accès aux données (ORAD/HDAB) sont appelées à jouer un rôle central dans l'écosystème européen des données de santé : point d'entrée pour les demandes d'accès, accompagnement méthodologique et juridique des porteurs de projets, standardisation et exposition des métadonnées. En France, ces missions sont largement assurées par la plateforme des données de santé, en lien étroit avec l'autorité de protection des

données et les stratégies nationales en intelligence artificielle (IA). Un enjeu structurant réside dans la constitution de catalogues nationaux de bases de données de santé considérées comme stratégiques. Ces bases, documentées par des métadonnées standardisées, sont accessibles dans des environnements sécurisés et s'inscrivent dans une perspective européenne de fédération des métadonnées. L'objectif est que chaque pays expose de manière harmonisée l'ensemble de ses ressources disponibles via un portail européen unique.

Dans le cadre des projets pilotes de l'EHDS, un réseau de plateformes nationales a expérimenté la fédération des métadonnées à l'aide d'un standard commun (HealthDCAT) et la mise en place de formulaires d'accès harmonisés, au prix d'un important travail juridique et organisationnel. Les cas d'usage transfrontaliers testés ont mis en évidence plusieurs goulots d'étranglement, la rapidité d'accès aux données apparaissant comme le frein majeur, en particulier pour les projets complexes impliquant plusieurs pays ou des chaînages de données sophistiqués. L'introduction de délais contraints pour l'instruction des demandes, assortis de mécanismes incitatifs, constitue l'un des apports structurants attendus de l'EHDS.

APPORTS DU RWE À LA R&D PHARMACEUTIQUE

Plusieurs présentations ont illustré le potentiel des données de vie réelle pour soutenir et accélérer la recherche et le développement pharmaceutiques. L'un des exemples portait sur la constitution d'une base européenne de données de médecine de ville couvrant plusieurs pays et reposant sur des systèmes d'information hétérogènes. Structurées de manière fine et longitudinales, ces données permettent d'explorer de nouvelles indications thérapeutiques ou d'identifier des sous-groupes de patients d'intérêt pour le développement clinique.

La base présentée regroupe plusieurs dizaines de millions de dossiers médicaux enrichis par une meilleure structuration des informations cliniques (symptômes, diagnostics, prescriptions, biologie, vaccinations) et par un effort particulier de liaison entre diagnostics et prescriptions. Cette structuration a nécessité une évolution des logiciels métiers, initialement conçus pour la facturation, vers une logique davantage médicale. L'exploitation de ces données repose sur des méthodes d'IA développées en collaboration avec des équipes académiques, permettant un travail par similarité de phénotypes, notamment en immunologie et en oncologie. Cette approche ouvre la voie à une caractérisation plus fine des populations, avec des perspectives d'intégration future de données génomiques.

Les résultats suggèrent des gains substantiels, tant en termes d'identification de nouvelles indications que de réduction des délais de développement. Ces approches ont déjà contribué à renforcer la crédibilité de dossiers de financement portés par des biotechs européennes, illustrant l'impact concret du RWE sur la compétitivité de l'écosystème.

RWE ET ESSAIS CLINIQUES : RECRUTEMENT ET ATTRACTIVITÉ INTERNATIONALE



Pierre-Henri BERTOYE
(Unicancer, CNRIPH, DGS)

L'utilisation des données hospitalières pour optimiser le recrutement des essais cliniques a constitué un autre axe majeur des discussions. L'estimation imprécise des populations éligibles reste l'une des causes principales de retards, de surcoûts, voire d'échecs d'essais. Les données nécessaires sont pourtant présentes dans les systèmes hospitaliers, et les progrès de l'IA et du traitement du langage naturel permettent désormais d'identifier plus finement les patients correspondant aux critères d'inclusion et d'anticiper les capacités de recrutement par centre.

En pratique, la mise en œuvre de ces approches reste entravée, notamment en France, par des verrous organisationnels : délais importants de contractualisation, coûts élevés d'accès aux données, et processus décisionnels internes parfois peu réactifs. Ces contraintes affectent l'attractivité européenne pour les études observationnelles, au bénéfice de pays comme les États-Unis, où des acteurs privés ont structuré des réseaux centralisés offrant rapidité, prévisibilité et maîtrise de la chaîne d'analyse. Des exemples européens historiques, tels que l'étude I-CARE en maladie de Crohn, rappellent toutefois qu'un *leadership* européen est possible, à condition d'investir dans des phases préparatoires longues, une coordination forte et une gouvernance claire du partage des données.

CADRE RÉGLEMENTAIRE ET DÉFIS MÉTHODOLOGIQUES



Gilles PAUBERT
(Cegedim Health Data)

Les essais dits « hybrides », combinant données issues d'essais cliniques et données de vie réelle, soulèvent des enjeux réglementaires spécifiques. En

France, un travail de clarification interinstitutionnelle a permis de réduire le risque de double instruction et de blocage, même si des adaptations restent nécessaires au niveau européen. Des guides méthodologiques nationaux encadrent désormais l'usage de bras contrôles externes basés sur des données de vie réelle, en insistant sur la licéité de la collecte, l'information des patients, la robustesse du design et la qualité des données.

L'hétérogénéité des cadres juridiques et techniques entre États membres demeure néanmoins un obstacle majeur à l'exploitation transnationale. Les divergences dans les définitions de l'anonymisation et de la pseudonymisation, ainsi que les différences d'organisation des systèmes de données, contribuent à cette complexité, malgré un cadre français souvent perçu comme relativement structuré.

PERSPECTIVES ET CONDITIONS DU LEADERSHIP EUROPÉEN



Pierre-Alexandre SQUARA
(Pfizer)

La table ronde finale a insisté sur la nécessité d'une approche collective pour asseoir un leadership européen en matière de RWE. L'innovation existe, mais elle reste dispersée, insuffisamment visible et parfois sous-exploitée. La cartographie systématique des projets, la mutualisation des acquis, la valorisation des résultats existants et le renforcement des partenariats public-privé constituent des leviers essentiels.

Les conditions de réussite apparaissent clairement identifiées : amélioration et harmonisation de l'accès aux données, modèles économiques soutenables,

gouvernance renforcée au niveau local, cadre réglementaire stabilisé et investissements durables dans la qualité et l'interopérabilité des données. Si ces conditions sont réunies, l'Europe dispose d'atouts majeurs, notamment une diversité des systèmes de santé, une richesse des cohortes, et une expertise méthodologique. Les capacités de l'Europe en IA et la tradition de régulation protectrice mais évolutive permettront de faire du RWE non seulement un outil au service de ses patients, mais aussi un domaine d'excellence reconnu au niveau mondial.

10 POINTS CLÉS

1. Le RWE doit dépasser une approche strictement nationale pour permettre la comparaison et la reproductibilité des résultats à l'échelle européenne.
2. L'écosystème européen des données de santé est riche mais fragmenté, du fait de la diversité des systèmes, standards et projets.
3. La cartographie structurée des projets européens est indispensable pour identifier les meilleures pratiques et éviter les redondances.
4. Les plateformes nationales d'accès aux données jouent un rôle central dans l'EHDS comme points d'entrée et d'accompagnement des projets.
5. La fédération européenne des métadonnées est réalisable mais nécessite un fort investissement juridique et organisationnel.
6. La rapidité d'accès aux données reste le principal frein aux projets transfrontaliers en RWE.
7. Le RWE constitue un levier majeur pour accélérer et sécuriser la R&D pharmaceutique.
8. Les données de vie réelle peuvent améliorer l'attractivité et le recrutement des essais cliniques en Europe.
9. Les cadres réglementaires et méthodologiques demeurent hétérogènes entre États membres.
10. Le *leadership* européen en RWE repose sur une action collective, durable et coordonnée.



TABLE RONDE

Données de santé en dehors de l'hôpital : exemple de la médecine de ville

Modérateurs : **Olivier GUÉRIN** (Chef du pôle gériatrie – CHU de Nice) et **Catherine RIOUFOL** (Chef de service Pharmacie du groupement hospitalier sud – Hospices Civils de Lyon)

INTERVENANTS

Cécile BADIOLA LAGARDÈRE (Directrice générale – HEVA) ; **Mehdi DJELAMANI** (Co-fondateur et Président – Stane) ; **Olivier SAINT-LARY** (Représentant du Collège National des Généralistes Enseignants (CNGE) – Platform For Data in Primary Care (P4DP)).

INTRODUCTION



Olivier GUÉRIN
(CHU de Nice)



Catherine RIOUFOL
(Hospices Civils de Lyon)

Le fil conducteur de la session « Données de santé en dehors de l'hôpital » porte sur la valorisation des données produites hors de l'hôpital, dans un contexte où environ 80 % des actes médicaux sont réalisés en ambulatoire. Les objectifs poursuivis concernent l'amélioration des soins, le renforcement de la prévention et le développement de la recherche clinique.

Le cadre conceptuel repose sur la distinction entre données de santé primaires et secondaires. Les données primaires correspondent aux informations recueillies au cours de la pratique quotidienne : anamnèse, examen clinique, prescriptions, résultats d'examens ou comptes rendus. Elles sont saisies dans les logiciels métiers des professionnels de santé et utilisées en premier lieu pour la prise en charge des patients. Elles deviennent secondaires lorsque ces informations sont extraites, anonymisées ou pseudonymisées, transformées puis intégrées dans des entrepôts de données de santé pour des usages distincts du soin direct, tels que la recherche clinique interventionnelle ou observationnelle, l'épidémiologie, les études médico-économiques ou l'entraînement de modèles d'intelligence artificielle.

Le SNDS est identifié comme un socle central d'observation des parcours de soins, en raison de sa couverture quasi exhaustive et de sa capacité à documenter la continuité ville-hôpital. Son caractère essentiellement médico-administratif limite toutefois sa capacité à répondre à des questions cliniques fines. L'enjeu majeur réside donc dans l'enrichissement de ce socle par des données cliniques détaillées produites en ville, issues des logiciels de médecine générale, des spécialistes ambulatoires, des laboratoires de biologie et des pharmacies. Cet enrichissement doit permettre d'évaluer la qualité des parcours de soins, l'efficacité des stratégies de dépistage et de prévention, ainsi que les résultats de santé en vie réelle, et de renforcer la robustesse des analyses médico-économiques.



FREINS À LA VALORISATION DES DONNÉES DE VILLE



Cécile BADIOLA LAGARDÈRE
(HEVA)

La recherche en soins primaires est présentée comme un domaine historiquement en retard en France par rapport à la recherche hospitalière. Ce retard s'explique notamment par une structuration tardive de la filière universitaire de médecine générale et par une culture de la recherche encore peu diffusée parmi les praticiens de ville. Une dynamique de rattrapage est cependant engagée, avec la montée en puissance de la médecine générale universitaire, la mobilisation de milliers de maîtres de stage, la participation croissante d'autres professions de santé de première ligne (sages-femmes, pharmaciens, kinésithérapeutes) et le développement de plateformes de données dédiées aux soins primaires. L'objectif est de disposer de bases de données cliniques de proximité, interopérables avec les entrepôts hospitaliers et le SNDS, afin de mieux caractériser les trajectoires de soins réelles et de soutenir des projets de recherche à fort impact.

L'analyse des freins à la valorisation des données de ville met en évidence trois types de verrous.

Le verrou technique est lié à la fragmentation des systèmes d'information et au faible degré de structuration de la donnée utile. Les logiciels métiers sont nombreux, les modèles de données hétérogènes et une part importante de l'information clinique est saisie en texte libre. La constitution d'entrepôts de qualité nécessite donc des chaînes d'interopérabilité capables de collecter automatiquement les données à partir des logiciels, d'en extraire les concepts cliniques pertinents, notamment grâce au traitement automatique du langage, de les mapper sur des standards partagés et de les organiser dans des formats exploitables, dans le respect des exigences réglementaires. Les standards et obligations issus du Ségur du numérique en santé constituent un premier cadre de convergence, mais des développements technologiques complémentaires demeurent nécessaires.

Le verrou financier concerne la soutenabilité des investissements requis pour la création, l'industrialisation et la maintenance des infrastructures de données. Les projets d'entrepôts, de plateformes d'analyse et de chaînage entre sources distinctes (SNDS, données cliniques de ville, biologie, imagerie, etc.) requièrent des investissements importants dans la durée. Une articulation entre financements publics, programmes nationaux, dispositifs tels que France 2030 ou soutien à des plateformes d'intérêt général, et investissements privés apparaît indispensable. La viabilité des modèles économiques

suppose également que la valeur créée par l'usage des données soit partagée de manière transparente avec les producteurs de données, en particulier les professionnels de santé de ville.

Le verrou culturel est lié à l'absence de tradition de recherche en médecine de ville en France. Une proportion importante de praticiens a été formée sans exposition significative à la recherche clinique ni à l'intérêt des données pour l'amélioration des pratiques. La diffusion d'une culture de la recherche passe par la formation initiale, notamment par la réalisation de thèses reposant sur des bases de données structurées, ainsi que par la formation continue, les congrès professionnels et le partage de retours d'expérience autour de projets concrets. La mise à disposition de tableaux de bord, d'indicateurs de pratique, d'outils d'aide à la décision et la participation à des essais cliniques contribuent à démontrer l'intérêt direct de la recherche pour le suivi des patients et favorisent l'adhésion des praticiens.

INITIATIVES DE VALORISATION DES DONNÉES DE VILLE



Mehdi DJELAMANI
(Stane)

La comparaison avec d'autres pays met en évidence le retard français et offre des repères utiles. Des bases telles que le CPRD au Royaume-Uni ou les réseaux de données de soins primaires aux Pays-Bas et dans les pays nordiques illustrent les bénéfices d'une culture de recherche implantée de longue date en médecine générale. Elles ont produit un volume important de publications et permis une forte réactivité en situation de crise, notamment lors de la pandémie de Covid-19. Ces expériences soulignent également les risques de dépendance technologique lorsque la structuration ou l'exploitation des données repose sur des acteurs extra-européens, renforçant la nécessité de consolider une souveraineté numérique européenne et française.

Plusieurs initiatives françaises illustrent des approches complémentaires de structuration et de valorisation des données de ville. Une plateforme nationale de données de soins primaires vise à constituer un entrepôt clinique alimenté automatiquement par de nombreux logiciels de médecine générale, en s'appuyant sur les exigences de standardisation issues du Ségur du numérique en santé. À terme, cette plateforme doit rassembler plusieurs millions de patients suivis par des milliers de médecins. Trois cas d'usage prioritaires sont identifiés : la restitution de tableaux de bord dynamiques aux

médecins pour analyser leur patientèle et améliorer leurs pratiques, la contribution à la veille épidémiologique en lien avec des réseaux existants, et la facilitation de l'inclusion dans les essais cliniques par le repérage de patients éligibles.

Un réseau de recherche clinique en médecine de ville, adossé à un entrepôt de données autorisé par la CNIL, illustre une autre stratégie. Ce réseau regroupe plusieurs milliers de professionnels de santé de ville et se situe à l'interface entre santé numérique et recherche clinique. Un « système d'exploitation de la donnée » assure la collecte automatisée, la pseudonymisation et la structuration des informations issues des logiciels métiers, malgré les hétérogénéités de saisie. Un portail permet aux promoteurs d'essais cliniques de déposer leurs protocoles ; des algorithmes identifient ensuite les patients répondant aux critères d'éligibilité et restituent aux médecins des listes de patients potentiels via des tableaux de bord dédiés. Parallèlement, un assistant numérique intégré au cabinet facilite la réalisation de l'acte de recherche au plus près de l'acte de soin. L'entrepôt permet également de mettre à disposition, dans des environnements sécurisés, des cohortes destinées aux études observationnelles et à l'entraînement de modèles d'intelligence artificielle (IA) en santé.

Une structure de valorisation de la donnée, filiale d'un grand groupe de services postaux, constitue un troisième exemple. Initialement spécialisée dans l'exploitation du SNDS, elle a progressivement intégré des compétences en biostatistique, en recherche clinique et en ingénierie de données, ainsi que des capacités d'hébergement certifiées. Ses activités couvrent les études observationnelles classiques et des projets à forte valeur ajoutée reposant sur le chaînage de données du SNDS avec des données cliniques complémentaires, notamment des données de biologie médicale issues de grands réseaux privés. Des autorisations spécifiques de la CNIL permettent ce chaînage dans des environnements sécurisés temporaires, comme l'illustre une étude récente sur la prévalence de l'hépatite B fondée sur la sérologie couplée aux données du SNDS.

Les données sont traitées dans le cadre du régime de non-opposition, avec une obligation d'information des patients sur les usages envisagés. Des espaces sécurisés temporaires (« bulles ») sont mis en place à l'échelle des projets : les données y sont agrégées, anonymisées ou pseudonymisées, accessibles via des outils d'analyse, mais non exportables sous forme brute. À la clôture des projets, les bulles sont fermées et les données détruites, afin de limiter les risques de réutilisation non contrôlée et de préserver la confiance des citoyens.

LEVIERS D'EXPLOITATION DES DONNÉES EN SOINS PRIMAIRES



Olivier SAINT-LARY
(CNGE, P4DP)

La mobilisation effective des professionnels de santé de ville repose sur plusieurs leviers complémentaires. La simplification technique vise à réduire au minimum les actions spécifiques nécessaires pour contribuer aux entrepôts, en s'appuyant sur les flux existants des logiciels métiers. La formation et l'accompagnement permettent de clarifier les enjeux scientifiques, réglementaires et pratiques, et de soutenir la mise en œuvre des outils. Des modèles économiques sont développés afin de redistribuer une partie de la valeur issue des projets aux producteurs de données, en complément des bénéfices scientifiques et cliniques. Cette redistribution est d'autant plus importante que l'activité de recherche s'ajoute à une charge de soin déjà élevée en médecine de ville.

Les échanges ont également mis en évidence l'intérêt d'intégrer d'autres professions de santé de ville à ces dispositifs. Les pharmacies d'officine sont identifiées comme des sources majeures de données sur les traitements, la polymédication, les comédications et les tests rapides d'orientation diagnostique. L'existence de systèmes de centralisation des données pharmaceutiques et d'archives numériques anciennes crée un potentiel important pour la recherche et la surveillance, sous réserve de résoudre les enjeux d'interopérabilité, de gouvernance et d'articulation avec les données de médecine générale et de biologie. L'extension des entrepôts à l'ensemble des professionnels de santé de proximité, y compris dans des structures pluriprofessionnelles, apparaît ainsi comme une évolution logique.

L'usage des données de ville pour le diagnostic, notamment des maladies rares, constitue un autre axe de développement. Ces pathologies, nombreuses mais individuellement peu fréquentes, rendent difficile pour un médecin généraliste la maîtrise de l'ensemble des tableaux cliniques. Le développement d'outils d'aide au diagnostic, fondés sur des combinaisons de signes cliniques issus des données de ville et orientant vers les centres de référence, est envisagé. La stratégie repose sur l'utilisation des entrepôts de données secondaires pour entraîner et valider des modèles d'IA, puis sur leur intégration progressive dans les logiciels métiers pour une utilisation en temps réel. Des travaux comparables sont déjà conduits pour des pathologies

fréquentes telles que la BPCO, en particulier pour la prédiction des exacerbations et l'identification des hospitalisations évitables.

Dans l'ensemble, la session met en évidence que les données de santé produites en dehors de l'hôpital constituent un levier stratégique majeur pour la transformation du système de santé. Leur exploitation structurée permet de mieux documenter les parcours de soins en conditions réelles, d'optimiser la prévention, d'améliorer l'évaluation des stratégies thérapeutiques, de renforcer la souveraineté numérique et de rapprocher la recherche de la pratique quotidienne. La réalisation de ce potentiel suppose la levée simultanée des verrous techniques, financiers et culturels, la mise en place d'infrastructures interopérables, sécurisées et souveraines. Elle repose également sur l'émergence de modèles de coopération équilibrés entre acteurs publics et privés, en lien étroit avec les professionnels de santé de ville et les patients.

10 POINTS CLÉS

1. Les données de santé produites en ville sont stratégiques, car près de 80 % des actes médicaux sont réalisés hors de l'hôpital.
2. La distinction entre données primaires et secondaires structure les usages entre soin, recherche, épidémiologie, médico-économie et intelligence artificielle.
3. Le SNDS constitue un socle central pour l'analyse des parcours ville-hôpital, mais reste limité pour les analyses cliniques fines.
4. L'enrichissement du SNDS par des données cliniques détaillées issues de la médecine de ville est indispensable.
5. La valorisation des données de ville se heurte à trois verrous majeurs : techniques, financiers et culturels.
6. La structuration et l'interopérabilité des données reposent sur les standards, le Ségur du numérique et des technologies complémentaires.
7. Des initiatives françaises démontrent la faisabilité de plateformes et de réseaux dédiés aux soins primaires.
8. L'automatisation du recueil depuis les logiciels métiers est essentielle pour limiter la charge des professionnels de santé.
9. L'adhésion des professionnels de ville repose sur la formation, la restitution d'indicateurs utiles et le partage de la valeur créée.
10. Les données de ville constituent un levier majeur pour améliorer prévention, diagnostique, recherche et souveraineté numérique.



CONTROVERSE

Cohorte observationnelle, registre, EDS, base de données : quel est le bon choix ?

Modérateurs : **Marc BOURLIÈRE** (Chef de service hépato-gastroentérologie – Hôpital Saint Joseph, Marseille) et **Jean-Marc PHELIP** (Gastro-entérologue et hépatologue – CHU de Saint-Etienne)

INTERVENANTS

Anne-Sophie JANNOT (Directrice médicale – Banque Nationale de Données Maladies Rares (BNDMR) ; **Nicolas PALEIRON** (Chef du service pneumologie – HIA Sainte-Anne, Toulon) ; **Grégoire REY** (Directeur – UMS France Cohortes) ; **Xavier TROUSSARD** (Président – Registre régional des hémopathies malignes de Basse-Normandie (RRHMBN)).

REGISTRES ET COHORTES : SIMILARITÉS ET DIFFÉRENCES



Nicolas PALEIRON
(HIA Sainte-Anne, Toulon)

INTRODUCTION



Marc BOURLIÈRE
(Hôpital Saint Joseph, Marseille)



Jean-Marc PHELIP
(CHU de Saint-Étienne)

L'objectif de cette session n'était pas de confronter des modèles concurrents, mais de clarifier les définitions, d'explicitier les forces et limites de chaque dispositif et de montrer en quoi registres, cohortes, entrepôts et bases nationales sont aujourd'hui profondément complémentaires pour la recherche et la veille sanitaire.

Les registres de cancers ont été présentés comme des dispositifs fondés sur des données personnelles, ce qui impose un respect strict du cadre réglementaire, en particulier du RGPD et de la loi Informatique et libertés. Leur caractéristique majeure est l'exhaustivité sur un territoire défini : tous les cas incidents doivent être recensés, quels que soient l'âge, le lieu de prise en charge ou le type de structure (hôpital public, clinique privée, laboratoire, etc.). Cette exhaustivité repose sur un travail de terrain. Les équipes de registre se déplacent dans les structures de soins, vérifient les dossiers, contrôlent la concordance des informations et s'assurent que les patients résident bien dans la zone de couverture. Cette démarche produit des données de très haute qualité, contrôlées, vérifiées, disponibles pour les professionnels de santé comme pour les autorités sanitaires. Les registres permettent ainsi un suivi temporel sur des dizaines d'années, la surveillance de l'incidence, l'étude des tendances spatiales et temporelles des cancers et la réalisation d'études de recherche populationnelle. Leur limite principale réside dans le délai de disponibilité des données. La nécessité de vérifier chaque cas conduit générale-



ment à un recul d'environ deux ans entre la survenue du cancer et sa mise à disposition dans les bases nationales. Sur le plan économique, les registres bénéficient d'un financement national, notamment par l'Institut national du cancer, qui couvre environ 60 % des besoins. Le reste doit être obtenu chaque année par des financements complémentaires, principalement pour les ressources humaines, ce qui fragilise la pérennité de certains registres pourtant essentiels à la surveillance épidémiologique.

Une cohorte est définie comme un ensemble de patients partageant une pathologie ou une situation clinique, inclus selon un protocole défini, puis suivis de façon longitudinale dans le temps. Dans l'exemple d'une grande cohorte française de patients atteints d'hépatites virales, plus de 25 000 patients ont été inclus dans un réseau de centres spécialisés, au moment où de nouveaux antiviraux révolutionnaient la prise en charge de l'hépatite C. La cohorte a permis d'évaluer l'efficacité réelle de ces traitements dans la « vraie vie », de démontrer la réduction des complications, l'amélioration du pronostic et la diminution du risque de carcinome hépatocellulaire. Le chaînage avec le Système national des données de santé (SNDS) s'est révélé déterminant pour mettre en évidence des effets indésirables inattendus, en particulier un surrisque de troubles du rythme cardiaque après l'arrêt des antiviraux, signal qui n'avait pas émergé des essais cliniques. Ce couplage a également permis d'évaluer les pratiques médicales, par exemple la qualité de la surveillance radiologique des cirrhoses, et de montrer l'écart entre les recommandations et la réalité observée. Sans chaînage au SNDS, les cohortes restent limitées aux variables prévues

dans le protocole initial, et ne permettent pas de détecter facilement des événements non anticipés. Le lien avec des bases de remboursement et d'hospitalisation ouvre au contraire la possibilité d'explorer des issues inattendues et d'analyser les parcours de soins à grande échelle.

L'expérience d'un groupe académique en cancérologie thoracique illustre le rôle des EDS. Selon la définition de la CNIL, il s'agit d'une base qui agrège des données provenant de sources multiples, structurée en vue de leur réutilisation à des fins de recherche. La mise en place d'un entrepôt nécessite le respect d'un référentiel CNIL spécifique, avec une procédure d'autorisation lourde, mobilisant juristes, informaticiens et cliniciens. Une fois l'autorisation obtenue, avec un délai d'environ un an et demi dans l'exemple discuté, l'outil devient en revanche très souple d'utilisation. Ainsi, les projets de recherche peuvent être multipliés sans redéposer à chaque fois un dossier complet, dès lors qu'ils entrent dans la finalité déclarée. Deux grandes configurations ont été distinguées. Les entrepôts fondés sur des données de recherche, qui agrègent des essais cliniques et des études observationnelles, disposent de données extrêmement robustes, monitorées ou contrôlées, portant aussi bien sur la survie, l'imagerie, la génétique tumorale que les données cliniques. Ils permettent de constituer des cohortes témoins, d'évaluer des biomarqueurs pronostiques ou de développer des modèles prédictifs sophistiqués, notamment par intelligence artificielle (IA). Les entrepôts hospitaliers de « vraie vie » intègrent quant à eux l'ensemble des données produites dans un établissement, avec une couverture beaucoup plus large mais une hétérogénéité et des lacunes plus importantes. Le modèle économique de ces entrepôts repose sur le

coût du stockage et surtout sur celui de l'ouverture d'espaces sécurisés de travail. La stratégie la plus efficace consiste à bien préparer les projets, ouvrir des « bulles » analytiques pour une durée limitée, réaliser les analyses puis fermer les accès, afin de maîtriser les coûts. Le financement est en pratique projet par projet, sur appels d'offres, avec des budgets globalement bien inférieurs au coût d'un essai clinique, même si les projets impliquant de gros volumes d'imagerie ou de génomique restent plus onéreux.

CADRES ORGANISATIONNELS SPÉCIFIQUES



Anne-Sophie JANNOT
(BNDMR)

La Banque Nationale de Données Maladies Rares (BNDMR) s'inscrit dans un cadre organisationnel spécifique. Une maladie rare est définie en Europe par une prévalence inférieure à 1 sur 2 000, ce qui signifie qu'en France une cohorte de 15 000 patients peut encore correspondre à une maladie rare. On estime à plusieurs milliers le nombre de maladies rares distinctes. Les cancers rares et les maladies infectieuses, bien que parfois peu fréquents, ne sont pas inclus dans le périmètre des plans nationaux maladies rares car ils relèvent d'autres dispositifs de financement. Les différents plans nationaux français ont labellisé environ 2 300 centres experts structurés en 23 filières. La BNDMR s'appuie sur un système d'information déployé dans ces centres, qui collecte de façon homogène un set minimal de données pour chaque maladie rare lors de chaque venue du patient. La collecte conditionne le financement des centres, ce qui garantit une très bonne exhaustivité. Les données ainsi recueillies permettent à la fois de calculer incidences et prévalences, comme un registre, et de suivre les trajectoires de soins dans le temps, comme une cohorte. L'entrepôt associé, mono-source, héberge des données dé-identifiées et adossées à des procédures d'accès très encadrées, avec mise à disposition de « bulles » sécurisées et grilles budgétaires définies. Le chaînage au SNDS est possible pour certaines études, et de nouveaux développements visent le suivi de l'utilisation des traitements, notamment les accès précoces, à partir d'un set de données élaboré en collaboration avec la HAS. Sur le plan européen, l'organisation française a inspiré les réseaux de référence, avec une structuration en filières, un set minimal commun et une harmonisation des formats de données via le modèle OMOP, notamment pour permettre la participation à des initiatives comme DARWIN de l'Agence européenne du médicament. La BNDMR illustre ainsi un dispositif hybride, à la fois registre national structurant et entrepôt de données de soin réutilisables pour la recherche.

FRANCE COHORTES



Grégoire REY
(UMS France Cohortes)

L'infrastructure France Cohortes répond à un autre besoin : la pérennisation et la mise en synergie des grandes cohortes nationales, souvent financées initialement dans le cadre du programme d'investissements d'avenir. Pilotée par l'Inserm et financée par le ministère de la Recherche et l'ANR, France Cohortes vise d'abord à proposer un hébergement mutualisé au niveau d'exigence requis pour le chaînage au SNDS, en particulier l'homologation au référentiel de sécurité. De nombreuses cohortes n'ont pas la capacité de se hisser seules à ce niveau de conformité technique et réglementaire. L'infrastructure fournit également un accompagnement méthodologique et réglementaire pour la rédaction des analyses d'impact, la gestion des risques et la mise en conformité avec le RGPD. Le troisième axe de travail concerne la science ouverte : documentation détaillée des données, catalogues, métadonnées, description des biais et des limites de recueil, afin de rendre les cohortes trouvables, accessibles, interopérables et réutilisables. L'accès à France Cohortes est conditionné à une évaluation scientifique, à travers les campagnes de l'ANR, les stratégies nationales de recherche ou le conseil scientifique de l'infrastructure, qui prévoit de lancer des appels à projets dédiés pour sélectionner les cohortes à héberger en priorité. À l'horizon 2026-2030, France Cohortes a également pour mission de faciliter la constitution de consortiums nationaux et européens. Les enjeux sont importants, car il n'existe pas encore de référentiel de sécurité harmonisé à l'échelle de l'Union, au-delà du RGPD, et le nouveau règlement sur l'espace européen des données de santé ne résout pas entièrement cette question. L'infrastructure se positionne donc comme un acteur clé pour organiser des analyses couplées et des projets multicohortes dans un cadre sécurisé.

MODÈLES ÉCONOMIQUES ET INTEROPÉRABILITÉ



Xavier TROUSSARD
(RRHMBN)

Plusieurs questions transversales sont revenues au cours des échanges, en particulier les modèles économiques et la valorisation des bases de données. Que ce soit pour les registres, les entrepôts ou les grandes cohortes, le financement est dominé par les coûts de personnel pour le recueil et la qualité des

données, et par les coûts techniques liés au stockage sécurisé et aux environnements de calcul. Au-delà des financements récurrents (INCa, ANRS, ANR, DGOS...), les équipes doivent développer des grilles budgétaires claires et attractives pour les projets académiques ou partenariaux, afin de favoriser l'utilisation effective des bases constituées. Les intervenants ont insisté sur l'enjeu de démocratiser l'accès à ces ressources, en particulier pour les jeunes chercheurs. Il a été rappelé qu'il n'est pas nécessaire d'être expert en programmation pour mener des projets simples sur un registre ou un entrepôt, dès lors qu'un accompagnement méthodologique et statistique est disponible. L'un des risques serait de voir des infrastructures coûteuses mais sous-utilisées, faute d'appropriation par la communauté clinique.

L'interopérabilité et l'IA constituent un autre champ de réflexion. Les règles actuelles imposent que les données ne sortent pas des entrepôts, ce qui complique les analyses inter-EDS. Des approches de type apprentissage fédéré sont en développement : un algorithme est envoyé successivement sur plusieurs entrepôts, s'entraîne localement, puis les modèles résultants sont agrégés sans que les données brutes ne circulent. Ces approches restent toutefois à consolider et à comparer à la solution, plus simple scientifiquement mais plus lourde juridiquement, qui consiste à constituer une base centralisée unique. Pour que ces stratégies fonctionnent, un important travail de standardisation est nécessaire, notamment à travers des modèles de données communs comme OMOP, ainsi que la mise à disposition de catalogues de métadonnées permettant aux chercheurs d'identifier précisément les variables disponibles, les cohortes mobilisables et les procédures d'accès.

En conclusion, les registres, cohortes, entrepôts de données de santé, base nationales thématiques et infrastructures de cohorte ne sont pas des dispositifs concurrents mais des briques complémentaires d'un même écosystème de données de santé. Les registres apportent l'exhaustivité populationnelle et la profondeur historique, les cohortes cliniques offrent un recueil riche et prospectif, les entrepôts permettent la réutilisation flexible de données de soin ou de recherche, les grandes infrastructures nationales et européennes organisent la mutualisation, la standardisation et l'ouverture contrôlée des données. L'ensemble est structuré par un cadre réglementaire exigeant, qui protège les personnes mais impose des compétences nouvelles aux équipes. Dans ce contexte en évolution rapide, la priorité est double : assurer la pérennité et la qualité de ces dispositifs par des modèles économiques stables, et développer une véritable culture de la donnée parmi les cliniciens, en particulier les plus jeunes, afin que ces ressources puissent pleinement contribuer à l'amélioration des connaissances et de la prise en charge des patients.



10 POINTS CLÉS

1. Les dispositifs de données ne sont pas concurrents mais complémentaires, chacun répondant à des objectifs spécifiques de recherche et de veille sanitaire.
2. Les registres se distinguent par leur exhaustivité populationnelle et la très haute qualité des données, au prix de délais de disponibilité plus longs.
3. Les cohortes offrent un suivi longitudinal fin et prospectif, mais restent limitées aux variables prévues par le protocole et à la taille des populations incluses.
4. Le chaînage des cohortes avec le SNDS augmente fortement leur valeur en permettant l'analyse des parcours de soins et la détection d'événements inattendus.
5. Les entrepôts de données de santé (EDS) facilitent la réutilisation flexible des données une fois le cadre réglementaire mis en place.
6. Deux grands types d'EDS coexistent : ceux fondés sur des données de recherche très robustes et ceux de vraie vie, plus larges mais plus hétérogènes.
7. Les modèles économiques des dispositifs reposent principalement sur les coûts humains liés au recueil, à la qualité et à la sécurisation des données.
8. La BNDMR illustre un modèle hybride combinant registre national et entrepôt de données, avec une forte exhaustivité et des usages de recherche encadrés.
9. France Cohortes joue un rôle structurant pour la pérennisation, la mutualisation et l'ouverture contrôlée des grandes cohortes nationales.
10. Le choix du dispositif le plus adapté dépend avant tout de la question scientifique posée et nécessite souvent une combinaison de plusieurs outils.



FLASH

Quid de la transportabilité des données : exemple

Modérateurs : **Christos CHOUAÏD** (Pneumologue – CHI de Créteil) et **David PÉROL** (Directeur de la recherche clinique et de l'innovation – Centre Léon Bérard, Lyon)

INTERVENANTS

Blythe ADAMSON (Directrice de la recherche sur les résultats de santé et génération de données – Flatiron Health) ; **Alison ANTOINE** (Spécialiste en biostatistique et en science des données médicales – Roche) ; **Mathieu ROBAIN** (Directeur scientifique – Direction des Data – Unicancer).

(complétude, standardisation, traçabilité) avec une compréhension des différences de pratiques cliniques, d'accès aux traitements et d'organisation des parcours de soins. L'enjeu est d'évaluer, pour une question donnée, dans quelle mesure des résultats observés dans un pays peuvent être appliqués de façon valide à un autre.

FLATIRON HEALTH ET ESME

INTRODUCTION



David PÉROL
(Centre Léon Bérard, Lyon)

L'utilisation des données de vie réelle en oncologie est devenue un complément indispensable aux essais cliniques pour décrire les pratiques, comparer l'efficacité des traitements et éclairer les décisions réglementaires et médico-économiques. Leur emploi au-delà du contexte national pose toutefois la question de la transportabilité des résultats entre systèmes de soins, cadres réglementaires et populations, surtout lorsque les bases locales sont limitées en effectifs ou en recul. Cette session Flash illustre comment articuler la qualité intrinsèque des entrepôts



Blythe ADAMSON
(Flatiron Health)

Flatiron Health exploite des données issues des dossiers médicaux électroniques de réseaux d'oncologie dans plusieurs pays, principalement aux États-Unis, mais aussi au Royaume-Uni, en Allemagne et au Japon. Les bases sont construites à partir d'un modèle de données commun, avec des variables standards pour l'ensemble des cancers et des variables spécifiques par localisation. La production de données repose sur une combinaison d'abstraction humaine experte, de traitements automatisés, de méthodes de traitement du langage naturel et, plus récemment, de grands modèles de langage, avec des validations en partenariat avec agences réglementaires et équipes académiques. Un grand nombre de publications ont été générées à partir de la base de données Flatiron.



ESME est un entrepôt de données de santé thématique français, multicentrique, dédié aux cancers avancés. Les données sont extraites des dossiers médicaux d'environ 39 établissements selon les cohortes, via un travail manuel de techniciens d'études cliniques formés et encadrés, selon un dictionnaire commun d'environ 200 variables. L'entrepôt comprend plus de 110 000 patients, dont plus de 60 000 atteints de cancer bronchique, avec un recul médian important et une forte granularité des informations cliniques et biologiques, notamment les biomarqueurs. Les données sont mises à jour par gels annuels et ont déjà donné lieu à de nombreuses publications, y compris dans le cadre de collaborations internationales.

APPROCHE MÉTHODOLOGIQUE DE LA TRANSPORTABILITÉ

L'une des premières étapes méthodologiques a consisté à comparer les populations sources, de Flatiron et ESME, afin d'identifier les convergences et divergences initiales. Des différences démographiques ont été observées, notamment en termes d'âge et de distribution par sexe, avec une population française globalement plus jeune et comportant une proportion plus élevée d'hommes. En revanche, les principales caractéristiques cliniques pronostiques, telles que le statut de performance, l'histologie tumorale, le statut tabagique ou le

stade au diagnostic, apparaissaient globalement comparables entre les deux bases. La majorité des patients dans les deux pays était diagnostiquée au stade métastatique d'emblée, ce qui constitue un point essentiel pour l'interprétation des analyses de survie.

L'analyse de l'évolution temporelle des prescriptions de traitements de première ligne entre 2015 et 2023 met en évidence des tendances similaires dans les deux pays, notamment la diminution progressive des chimiothérapies à base de sels de platine au profit d'une utilisation croissante de l'immunothérapie, seule ou en association avec une chimiothérapie. Les thérapies ciblées, utilisées en monothérapie chez des sous-groupes sélectionnés de patients, montrent une stabilité relative de leur prescription au cours du temps. Néanmoins, un décalage temporel notable dans l'accès aux innovations thérapeutiques est observé, avec une introduction plus précoce des immunothérapies dans le contexte américain, en moyenne deux ans avant leur disponibilité et leur remboursement en France. Ce décalage, bien documenté au niveau international entre autorités réglementaires, constitue un facteur déterminant dans l'interprétation des différences observées entre populations.

L'étude des trajectoires thérapeutiques de la première à la troisième ligne met en évidence une

structure globale très similaire : en première ligne, la chimiothérapie à base de platine reste le schéma le plus fréquent, suivie par l'immunothérapie associée à la chimiothérapie, l'immunothérapie seule puis les thérapies ciblées. Les proportions varient, mais l'ordre de fréquence et l'architecture des parcours sont comparables, suggérant une convergence des standards de prise en charge malgré des organisations de soins différentes.

Le cœur de l'analyse repose sur l'évaluation de la transportabilité des résultats de survie globale. L'objectif était de déterminer si les *outcomes* observés dans une cohorte américaine pouvaient être appliqués, après ajustement, à une population française comparable. Cette démarche repose sur le principe de validité externe et suppose la disponibilité de facteurs pronostiques communs permettant un ajustement robuste. Les analyses ont été conduites dans plusieurs sous-groupes définis par le type de traitement de première ligne, correspondant aux stratégies les plus fréquemment utilisées dans les deux pays. Les méthodes d'ajustement employées reposaient principalement sur des approches de pondération par score de propension, permettant de rééquilibrer la population américaine sur les caractéristiques de base de la population française. Les variables utilisées incluaient notamment l'âge, le sexe, le statut de performance, l'expression de PD-L1 et d'autres facteurs cliniques majeurs. Cette étape est essentielle pour limiter les biais liés aux différences de structure des populations et permettre une comparaison plus juste des *outcomes*.

RÉSULTATS



Alison ANTOINE
(Roche)

Dans le groupe traité par chimiothérapie à base de sels de platine en première ligne, les courbes de survie globale étaient très proches entre les deux pays, avant et après ajustement. La quasi-superposition des courbes suggère une transportabilité élevée dans ce contexte où les pratiques sont standardisées et les lignes ultérieures moins dépendantes de l'accès aux innovations. Chez les patients recevant une immunothérapie seule en première ligne, les courbes de survie brute divergeaient nettement entre États-Unis et France. Après pondération, la courbe américaine se rapprochait de la courbe française, montrant que les différences de distribution des facteurs pronostiques expliquent une partie de l'écart, mais un différentiel persistant suggère un rôle de déterminants non mesurés ou contextuels.

Des résultats comparables ont été observés avec les inhibiteurs de l'EGFR en première ligne : l'ajustement réduit la différence de médiane de survie sans l'abolir, et les courbes s'écartent à distance. La transportabilité apparaît ici partielle, probablement influencée par la gestion des lignes suivantes, la disponibilité de nouvelles molécules ciblées et les stratégies de suivi. Dans le groupe traité par immunothérapie associée à une chimiothérapie, les courbes américaines et françaises demeuraient distinctes malgré l'ajustement. Cette divergence illustre les limites de la transportabilité lorsque interviennent des facteurs plus complexes : différences d'organisation des soins, séquences thérapeutiques, critères d'éligibilité aux traitements et préférences de fin de vie. Elle rappelle qu'un ajustement statistique, même sophistiqué, ne compense pas l'absence de variables explicatives clés ni les spécificités structurelles des systèmes de santé.

LIMITES DES MÉTHODES DE TRANSPORTABILITÉ



Mathieu ROBAIN
(Unicancer)

Plusieurs limites doivent être prises en compte. L'harmonisation de sources distinctes repose sur des algorithmes de dérivation des lignes de traitement proches mais non identiques, ce qui peut générer des biais. L'exclusion de patients avec données manquantes sur certaines covariables d'ajustement introduit un risque de biais de sélection ; des analyses de sensibilité avec imputation sont prévues pour tester la robustesse des résultats. Par ailleurs, un biais résiduel est probable en raison de variables non disponibles ou non harmonisées telles que statut socio-économique, ethnicité, type et nombre de métastases. Ces facteurs peuvent influencer la survie et limiter l'interprétation causale des écarts observés. Les différences de calendrier d'accès aux innovations constituent un autre déterminant majeur : des analyses restreintes aux périodes où les options thérapeutiques étaient effectivement disponibles dans les deux pays montrent des résultats globalement stables, mais soulignent la nécessité d'un cadrage temporel précis pour chaque question de recherche.

PERSPECTIVES

Du point de vue réglementaire et décisionnel, ces travaux montrent que la transportabilité des données de vie réelle doit être appréciée au cas par cas, en fonction du contexte clinique, du traite-

ment étudié et de l'effet attendu. Un écart de survie d'un à trois mois peut être acceptable ou non selon l'ampleur de l'effet recherché et le niveau d'incertitude toléré. L'adéquation entre question de recherche, différence cliniquement pertinente et capacité d'ajustement des bases est centrale. Ils illustrent également l'intérêt des collaborations internationales structurées pour tester empiriquement la validité externe des données. Des bases volumineuses, bien caractérisées et de haute qualité, comme Flatiron et ESME, permettent de développer et d'évaluer des méthodologies d'ajustement avancées et de renforcer la crédibilité des données de vie réelle auprès des autorités, des payeurs et de la communauté scientifique.

En conclusion, l'étude présentée démontre que la transportabilité des *outcomes* issus de données de vie réelle est possible dans certaines configurations, notamment lorsque les pratiques sont homogènes et les déterminants majeurs bien capturés. Elle met également en évidence ses limites dans des contextes plus complexes, en particulier pour les traitements innovants dont l'utilisation est étroitement liée aux cadres réglementaires et aux organisations de soins. Ces enseignements plaident pour une approche méthodologique rigoureuse, transparente et pragmatique de la transportabilité, intégrée dès la conception des études, et pour le développement continu de standards internationaux en matière de données de vie réelle.

10 POINTS CLÉS

1. La transportabilité des données de vie réelle constitue un enjeu central pour l'utilisation internationale de la RWE en oncologie.
2. La qualité intrinsèque des bases (complétude, standardisation, traçabilité, recul) conditionne toute analyse de transportabilité.
3. Flatiron et ESME reposent sur des modèles de production différents mais robustes et scientifiquement validés.
4. La comparaison initiale des populations est indispensable pour identifier convergences et divergences démographiques et cliniques.
5. Les pratiques thérapeutiques évoluent de manière globalement similaire entre États-Unis et France, avec un décalage d'accès aux innovations.
6. Les trajectoires de soins de la première à la troisième ligne présentent une architecture comparable entre les deux pays.
7. La transportabilité est élevée pour les traitements standardisés comme la chimiothérapie à base de platine.
8. Elle devient partielle pour les traitements innovants, en particulier l'immunothérapie et les thérapies ciblées.
9. Les méthodes d'ajustement statistique réduisent les biais mais ne compensent pas les différences structurelles entre systèmes de soins.
10. La transportabilité doit être évaluée au cas par cas selon la question clinique, le traitement étudié et le niveau d'incertitude acceptable.



TABLE RONDE

Impact de l'IA sur la personnalisation des traitements et le parcours de soins

Modérateur : **Marco FIORINI** (Directeur général, Filière Intelligence Artificielle & Cancers (FIAC))

INTERVENANTS

Philippe-Jean BOUSQUET (Collecteur analyseur de données – Central data analyzer for Genomics (CAD)) ; **Franck MOUTHON** (Directeur exécutif de l'agence de programmes pour la recherche en santé – INSERM) ; **Pascal PUJOL** (Président – Société française de médecine prédictive et personnalisée (SFMPP)).

puvant sur les contributions de spécialistes de la donnée génomique, de la recherche en santé et de l'oncologie de précision.

L'IA n'est plus un concept prospectif, mais un outil déjà opérationnel à plusieurs étapes du parcours patient. Elle ouvre des perspectives majeures, mais pose également des défis techniques, économiques, éthiques et sociétaux qui conditionneront son déploiement équitable et sécurisé au sein du système de santé.

UNE INTÉGRATION DÉJÀ EFFECTIVE DE L'IA DANS LES DÉCISIONS DE SOINS

INTRODUCTION



Marco FIORINI
(FIAC)

L'intelligence artificielle (IA) occupe désormais une place centrale dans l'évolution de la médecine personnalisée. Elle transforme la manière d'analyser les données, d'organiser les parcours de soins et de prendre des décisions cliniques. Cette table ronde, modérée par Marco Fiorini, a mis en lumière les principaux enjeux, en s'ap-



Franck MOUTHON
(INSERM)

L'IA est désormais présente dans différents domaines de la médecine, souvent de façon discrète mais structurante. En imagerie médicale, par exemple, des algorithmes permettent depuis longtemps de détecter et caractériser des lésions, d'automatiser le contourage ou de guider l'analyse diagnostique. Ces outils, intégrés depuis des années dans les *workflows* cliniques, illustrent la capacité de l'IA à devenir un instrument de travail accepté et maîtrisé.



En oncologie de précision, l'IA joue un rôle essentiel dans l'interprétation des données génomiques, l'évaluation des biomarqueurs, la prédiction de la réponse aux traitements et l'orientation vers des thérapies ciblées ou des essais cliniques. Les réunions de concertation pluridisciplinaires s'appuient de plus en plus sur des outils capables d'agréger des informations complexes et d'améliorer la pertinence des décisions thérapeutiques.

Certaines applications ont déjà démontré leur impact quantifiable. Dans le domaine de la transplantation, par exemple, des modèles issus de l'IA permettent d'améliorer significativement l'appariement entre donneurs et receveurs, augmentant la précision du pronostic de rejet et diminuant les complications à moyen terme. Ces avancées illustrent l'apport direct de l'analyse de données massives à la décision clinique.

LA DONNÉE GÉNOMIQUE AU CŒUR DE LA PERSONNALISATION



Pascal PUJOL
(SFMPP)

La médecine prédictive a connu une transformation notable ces dernières années, reposant sur deux évolutions majeures : l'essor des connaissances génétiques et le développement d'approches multimodales intégrées. L'augmentation du nombre de gènes identifiés, l'élargissement des panels de séquençage et l'avènement des scores de risque polygénique ont permis une estimation plus fine du risque individuel.

Toutefois, la génétique n'explique qu'une partie du déterminisme du risque. L'enjeu actuel consiste à associer à ces données d'autres informations : environnement, habitudes de vie, antécédents, imagerie, protéomique, épigénétique. L'IA offre précisément la capacité d'exploiter simultanément ces sources hétérogènes. Ces approches permettent de personnaliser non seulement les stratégies de prévention intensifiée pour les individus à haut risque, mais aussi, lorsque c'est pertinent, de réduire le recours à certains examens pour les personnes à risque inférieur à la moyenne. Le modèle évolue ainsi d'un dépistage uniforme vers un continuum de niveaux de surveillance individualisés.

Le séquençage génomique massif génère des volumes de données considérables, impossibles à interpréter sans technologies d'IA. Un génome humain comporte trois milliards de bases ; entre deux individus, on recense plusieurs millions de

variations. Isoler celles qui expliquent une maladie rare ou une différence de réponse à un traitement exige des méthodes d'analyse automatisées puissantes. La collecte à large échelle de génomes, initiée dans plusieurs programmes nationaux, vise à réduire l'errance diagnostique, affiner les classifications de maladies rares, mieux comprendre la variabilité interindividuelle en cancérologie et favoriser l'émergence des jumeaux numériques. Ces derniers permettraient de simuler différentes trajectoires thérapeutiques pour éclairer le choix du traitement optimal.

L'un des défis majeurs réside dans l'articulation des données génétiques avec les données de vie réelle, condition essentielle pour progresser vers une véritable personnalisation des prises en charge. Chez des patients en situation d'impasse thérapeutique, par exemple, l'IA peut faciliter l'identification d'essais cliniques adaptés en confrontant automatiquement les données cliniques et moléculaires d'un patient aux bases d'études en cours, tout en vérifiant les critères d'inclusion et l'état réel des inclusions. Cette automatisation contribue à réduire les délais, à éviter des déplacements inutiles et à améliorer l'équité d'accès à l'innovation.

PRIORITÉS NATIONALES COHÉRENTES AVEC DES OBJECTIFS DE SANTÉ



Philippe-Jean BOUSQUET
(CAD)

La structuration nationale de la recherche en IA appliquée à la santé nécessite une priorisation claire. Face à un champ extrêmement large, les stratégies de financement doivent être alignées sur des objectifs explicites de santé publique : amélioration de l'espérance de vie en bonne santé, réduction des maladies chroniques évitables, montée en puissance de la prévention, diminution des pertes de chance. La démarche d'analyse et d'anticipation stratégique repose sur l'examen des publications, des brevets, des financements préexistants, des projets internationaux et des capacités nationales. Il s'agit d'identifier les domaines où la France dispose d'avantages compétitifs à amplifier, mais aussi les zones à risque où l'absence d'investissement entraînerait une dépendance ou une vulnérabilité technologique.

La prévention représente un domaine prioritaire, notamment en raison du retard national. L'IA y jouera un rôle essentiel pour adapter des modèles internationaux à la population française, exploiter au mieux les données existantes et concevoir des outils de prédiction robustes permettant d'ajuster les politiques publiques.

Plusieurs limites freinent encore un déploiement optimal de l'IA dans le système de santé. Sur le plan psychologique et culturel, l'innovation technologique peut susciter des craintes parmi les professionnels, liées à l'idée de perte de contrôle ou à la complexité perçue des modèles. Or certaines IA surpassent déjà l'humain dans des tâches spécifiques. Le défi consiste à considérer l'IA comme un soutien au raisonnement clinique, et non comme une menace. À ces obstacles s'ajoutent des limites économiques et environnementales. Former des modèles performants est coûteux et énergivore. Le secteur de la santé doit définir des modèles économiques stables pour financer, intégrer et maintenir les solutions utiles, tout en sélectionnant celles dont l'impact clinique est démontré. Enfin, des enjeux sociétaux et éthiques persistent, marqués par un manque de clarté dans la relation entre données de santé et citoyen, l'absence d'un « pacte numérique » partagé et la circulation de fausses informations alimentant la défiance.

La protection de la vie privée et la souveraineté des données doivent être garanties, mais aussi explicitées de manière lisible. La confiance constitue le socle de toute stratégie de données. Les infrastructures nationales ont fait le choix d'un hébergement souverain, limitant strictement la circulation des données en dehors du cadre européen. Cette approche garantit une sécurité renforcée mais exige une pédagogie claire sur les modalités d'usage. La transparence des usages, l'explication systématique des finalités et la mise en avant des bénéfices cliniques constituent des leviers essentiels pour favoriser l'adhésion. Les exemples tirés des cohortes nationales montrent que, bien informés, les citoyens participent volontiers à la recherche, conscients que les données qu'ils fournissent bénéficient à l'ensemble de la collectivité.

LES COMPÉTENCES NÉCESSAIRES POUR UNE IA DE CONFIANCE

Le développement d'une IA fiable et utile nécessite des compétences diversifiées. Les experts en données sont indispensables, mais difficiles à recruter et à fidéliser dans le secteur de la santé face à la concurrence d'autres industries. En parallèle, la formation des professionnels de santé doit évoluer pour intégrer une compréhension des principes, limites et biais des modèles d'IA.

L'enjeu n'est pas de transformer les soignants en ingénieurs, mais de leur permettre de dialoguer avec les spécialistes de l'IA, de comprendre les logiques des outils qu'ils utilisent et de garder une maîtrise clinique éclairée de la décision.



10 POINTS CLÉS

1. L'IA est déjà intégrée dans le parcours de soins, notamment en imagerie médicale, en oncologie de précision et dans les réunions de concertation pluridisciplinaires.
2. L'IA renforce la qualité et la pertinence des décisions cliniques en facilitant l'analyse conjointe de données cliniques, biologiques et génomiques pour orienter les choix thérapeutiques et l'accès aux essais cliniques.
3. La donnée génomique constitue un pilier central de la médecine personnalisée et nécessite des outils d'IA pour être interprétée à grande échelle.
4. La personnalisation des soins repose sur des approches multimodales combinant génétique, environnement, modes de vie, imagerie et données cliniques.
5. L'IA transforme les stratégies de prévention et de dépistage en permettant des niveaux de surveillance adaptés au risque individuel.
6. L'articulation entre données génomiques et données de vie réelle est essentielle pour une personnalisation efficace des traitements et des parcours de soins.
7. Les priorités nationales en IA doivent être alignées sur des objectifs clairs de santé publique, en particulier la prévention et la réduction des pertes de chance.
8. Des freins culturels, économiques et environnementaux limitent encore le déploiement à grande échelle des solutions d'IA en santé.
9. La confiance des citoyens repose sur la souveraineté des données, la protection de la vie privée et la transparence des usages.
10. Le succès de l'IA en santé dépend des compétences humaines, associant experts en données et professionnels de santé formés à un usage éclairé des outils.



Take-home messages & conclusion



David PÉROL

(Directeur de la recherche clinique et de l'innovation – Centre Léon Bérard, Lyon)

Pour conclure cette journée scientifique particulièrement riche et stimulante, David Pérol a tenu à remercier l'ensemble du comité d'organisation, les membres du comité scientifique, les participants, les partenaires académiques et industriels ainsi que le public, pour leur participation active et soutenue tout au long de la journée.

À l'issue de ces échanges, plusieurs questions initialement soulevées ont pu trouver des éléments de réponse. Tout d'abord, les travaux présentés ont permis de clarifier des notions fondamentales de terminologie et de sémantique, en distinguant précisément les concepts de cohortes en données de santé, de bases de données populationnelles et de registres, dont les structurations, les usages et les finalités diffèrent profondément. Il est apparu clairement que la donnée, et *a fortiori* son volume, ne constitue pas une finalité en soi, mais la qualité, la qualification et l'adéquation des données à des objectifs clairement définis demeurent essentielles.

Par ailleurs, la journée a mis en évidence des avancées réglementaires notables. Les échanges avec les représentants des instances de régulation ont souligné une écoute favorable à la valorisation de dossiers reposant sur des données observationnelles, à condition qu'ils s'inscrivent dans le respect de recommandations nationales et internationales et

qu'ils mobilisent des méthodes robustes et efficaces. De nombreux exemples ont illustré la manière dont ces approches méthodologiques permettent aujourd'hui de répondre à des questions cliniques insuffisamment couvertes par les essais cliniques randomisés.

Les présentations ont mis en évidence l'intérêt des comparaisons indirectes entre essais cliniques, en particulier pour les cancers rares ou très rares et, plus généralement, pour les pathologies à faible incidence, y compris pour l'analyse de sous-groupes ou de critères intermédiaires. Ces travaux soulignent la nécessité de renforcer les collaborations internationales. À cet égard, les résultats présentés sur la transportabilité des données entre les États-Unis et la France illustrent la faisabilité de recherches comparatives de qualité et ouvrent des perspectives prometteuses en matière d'interopérabilité des données entre pays, susceptibles d'intéresser également les autorités de régulation.

Enfin, les discussions ont porté de manière transversale sur les données de vie réelle et observationnelles en médecine de ville, les enjeux de financement, la place actuelle et future de l'intelligence artificielle dans ces dispositifs, ainsi que sur les exigences en matière de *reporting* et de valorisation scientifique.

La qualité des présentations et des échanges a été unanimement saluée. Face à l'intérêt croissant pour ces thématiques et à la richesse des questions encore à explorer, un rendez-vous est donné pour la prochaine édition, qui se tiendra le **12 novembre 2026**, consacrée à nouveau aux données de vie réelle.