Échantillons et données : améliorer leur accessibilité et leur réutilisation au service de la recherche

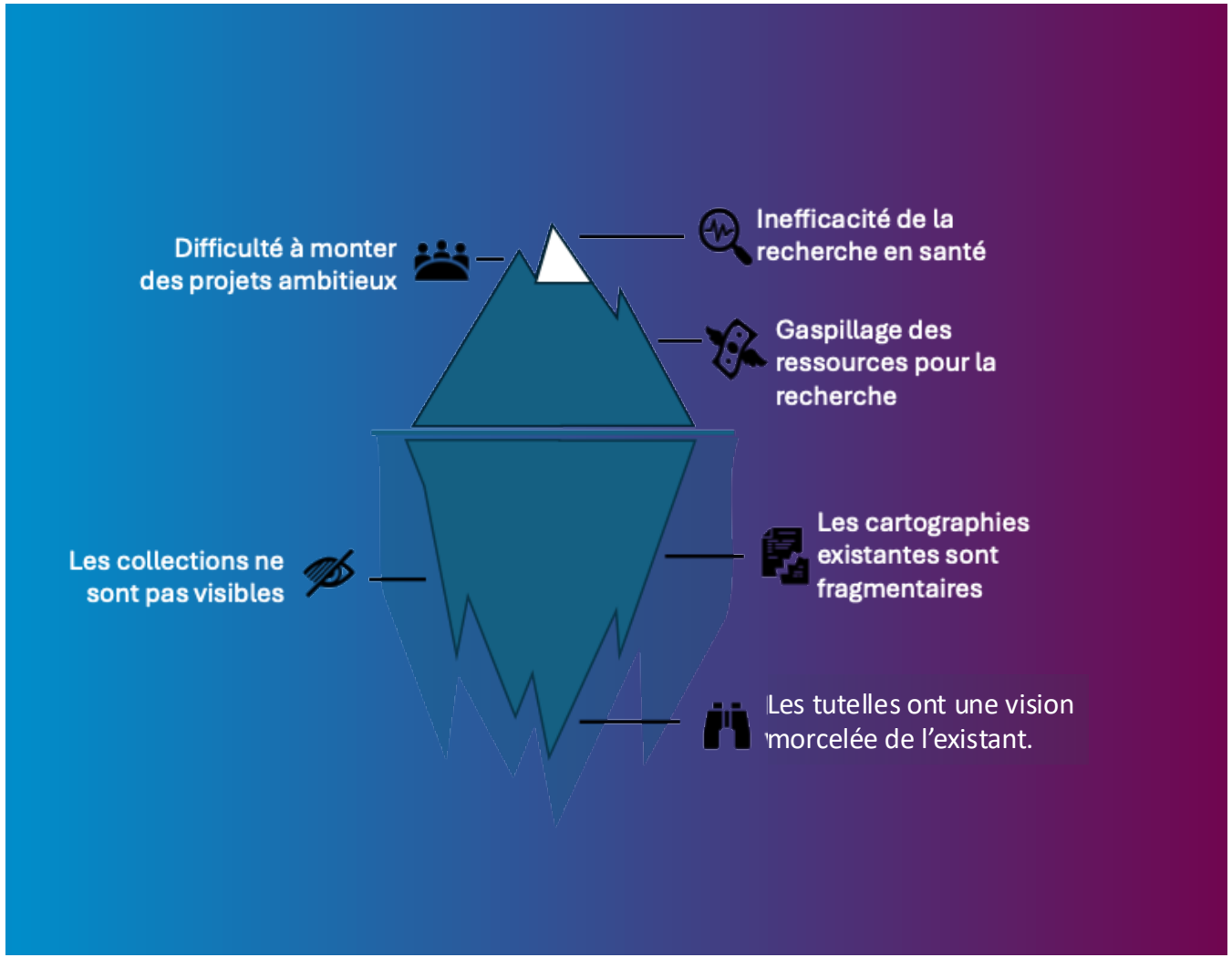
Work Package 2 : Cartographie & Database

Journée Nationale FrBioNet

Jeudi 24 septembre 2026

Dr. Séverine TABONE EGLINGER – Dr. Yves Edouard HERPE

VALORISER UN CAPITAL DE RESSOURCES BIOLOGIQUES DE LA FRANCE ENCORE PEU VISIBLE



Collections biologiques **essentielles au progrès** des sciences de la vie et de la médecine

Développement biologie à grande échelle, des approches « omique », de la médecine personnalisée et, plus récemment de l'intelligence artificielle

La France possède de nombreuses collections biologiques de haute valeur scientifique, mais :

- **Multitude de structures** distribuées sur tout le territoire

Il en résulte:

- **Hétérogénéité des pratiques**
- **Mauvaise visibilité** des ressources pour les chercheurs

Frein à l'utilisation de ces ressources pour les projets de recherche académiques ou industriels

VALORISER UN CAPITAL DE RESSOURCES BIOLOGIQUES DE LA FRANCE ENCORE PEU VISIBLE

PREMIÈRE MINISTRE
Liberté
Égalité
Fraternité
Secrétariat général pour l'investissement, en charge de France 2030

Le Secrétariat général pour l'Investissement

La Direction générale de la recherche et de l'innovation

La Direction de l'offre de Soins

Paris, le 31 MAI 2023

Madame la Déléguée Générale de France Universités,
Monsieur le Président Directeur Général de l'Inserm,
Monsieur le Président du CNCR,
Monsieur le Directeur Général du RESPIC,

Le plan France 2030 répond aux grands défis de notre temps à travers un plan d'investissement massif pour faire émerger les futurs champions technologiques de demain. Ce plan ambitionne de permettre à la France de retrouver le chemin de son indépendance environnementale, industrielle, technologique, sanitaire et culturelle et de prendre un temps d'avance dans ces secteurs stratégiques.

Annoncé par le Président de la République, le volet santé de France 2030 « Innovation Santé 2030 » prévoit un ensemble de mesures législatives et règlementaires ainsi que 7,5 milliards d'euros pour faire de la France la nation la plus innovante et souveraine en santé d'Europe. L'objectif est d'accompagner les transformations de notre système de santé, au bénéfice de tous les Français, en produisant les traitements innovants tout en assurant la sécurité d'approvisionnement des produits essentiels.

Les collections biologiques sont des ressources essentielles au progrès des sciences de la vie et de la médecine, en particulier avec le développement de la biologie à grande échelle, des approches « omiques » et de la médecine personnalisée.

Les échantillons biologiques, qu'ils soient collectés pour des études cliniques, épidémiologiques, ou dans le cadre du soin, sont stockés dans des biobanques, le plus souvent des centres de ressources biologiques (CRB) labellisés et certifiés, généralement adossés à des établissements de soins. Certaines collections importantes sont aussi également stockées en dehors de ces structures, dans les établissements de soin ou les structures de recherche.

Cette organisation caractérisée par l'existence d'une multitude de structures distribuées sur tout le territoire entraine, malgré la mise en place d'un système de management de la qualité dans la plupart d'entre elles, une certaine hétérogénéité dans les pratiques et les procédures, et une mauvaise visibilité des ressources pour les chercheurs.

Il en résulte une variabilité dans la qualité des échantillons et une complexité d'accès qui constituent des freins à l'utilisation de ces ressources pour les projets de recherche académiques ou industriels.

Dans le cadre de France 2030, le Gouvernement souhaite confier conjointement aux acteurs de la recherche et aux établissements de santé publics ou privés groupés au sein d'un consortium, une mission de structuration des biobanques visant à définir une politique harmonisée de gestion et d'accès aux ressources biologiques et aux données générées, dans le respect des principes « FAIR » (Findable, Accessible, Interoperable, Reproducible) et de la science ouverte soutenus par la France.

Les porteurs de projets seront a minima l'Inserm, le CNCR, le RESPIC et France Universités. Il est fortement attendu des porteurs de projets que la réponse soit construite de manière concertée avec une gouvernance permettant de s'assurer la représentation de l'ensemble des partenaires et structures (notamment Conférence des DG de CHU et Réseau national des CRB de CHU, Unicancer, Conférence des DG de CH, FEHAP, RESPIC, Service de santé des armées, industriels et acteurs du mode économique).

Hôtel de Cassini – 32, rue de Babylone – 75007 Paris
Tel : 01.42.75.64.32 – contact.sgpi@pm.gouv.fr

Cette structuration aura pour objectifs principaux :

- **de structurer le pilotage des biobanques au niveau national et représenter l'écosystème national à l'international ;**
- **d'améliorer la qualité de collections biologiques** afin d'augmenter leur potentiel de réutilisation, en harmonisant les pratiques des biobanques ;
- **de développer des capacités d'accompagnement à la création de nouvelles collections biologiques associées à des données clinico-biologiques, sociodémographiques, de phénotypage profond et/ou de génotypage, à fort potentiel de réutilisation** et d'augmenter, le cas échéant, les capacités de stockage des biobanques à cet effet ;
- **de simplifier l'accès pour les porteurs de projets de recherche publics ou privés aux collections biologiques** à fort potentiel de réutilisation et aux données associées.

Pour atteindre ces objectifs, cette structuration devra :

- **réaliser un bilan de la situation** en matière d'organisation des biobanques à l'échelle du territoire ;
- **dimensionner la réponse** à l'évolution des besoins des futures collections biologiques à fort potentiel de réutilisation des acteurs de la recherche académique ou industrielle ;
- **d'articuler avec les initiatives lancées**, en particulier avec l'action de la DGOS sur les biobanques

Nous vous prions d'agréer, Madame la Déléguée Générale, Messieurs les Présidents et Directeurs, l'expression de notre considération distinguée.

Claire GIRY
Directrice générale de la recherche et de l'innovation

Bruno BONNEL
Secrétaire général pour l'Investissement, en charge de France 2030

Marie DAUDE
Directrice générale de l'offre de soins

une mission de structuration des biobanques visant à définir une politique harmonisée de gestion et d'accès aux ressources biologiques et aux données générées, dans le respect des principes « FAIR » (Findable, Accessible, Interoperable, Reproducible) et de la science ouverte soutenus par la France.

LES PRINCIPES FAIR AU SERVICE DES BIO-COLLECTIONS

Un cadre commun pour mieux identifier, partager, connecter et valoriser les collections.



RENDRE LES COLLECTIONS « TROUVABLES »

- Catalogue national commun
- Métadonnées structurées et interrogeables
- Liens vers les études et données associées (ex: portail FreSH)



CLARIFIER LES CONDITIONS D'ACCÈS

- Règles explicites
- Processus harmonisés avec le w/WP4 (recommandations)
- Propositions de contrats types, démarches simplifiées w/WP4



FAIRE DIALOGUER LES SYSTÈMES

- Formats et standards communs w/WP3&5
- Terminologies et ontologies partagées w/WP3&5



PRÉPARER DE NOUVEAUX USAGES

- Données documentées et contextualisées
- Conditions de réutilisation claires
- Valorisation dans de futures collections

Identifier → Accéder → Connecter → Réutiliser

Dresser l’état des lieux des biocollections

L'ENQUÊTE DE CARTOGRAPHIE : UN DISPOSITIF NATIONAL ET ÉVOLUTIF

OBJECTIFS

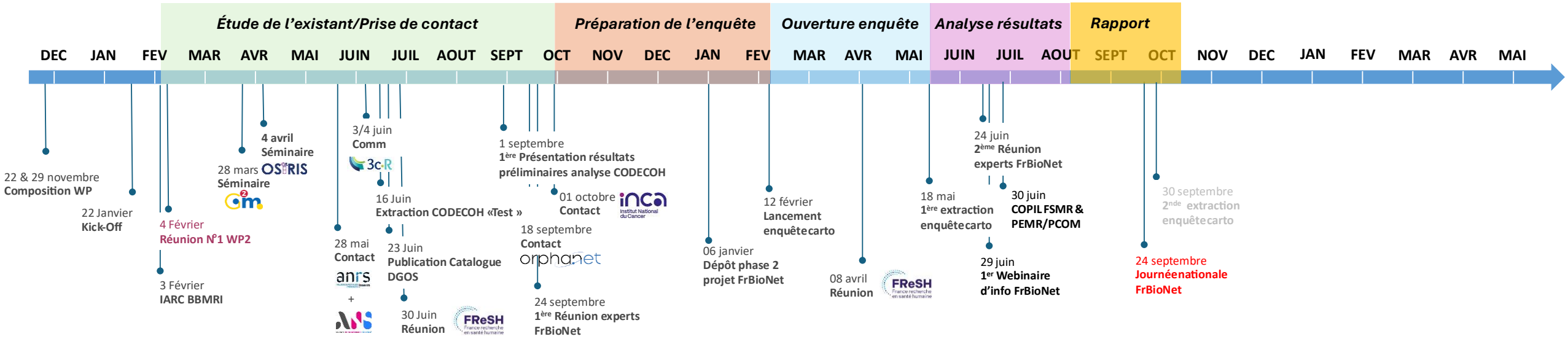
- **Cartographie des CRB et des collections valorisables**
- **Identification des données pour le catalogue (w/ WP5).**
- **Préconisations concernant les items minimaux à collecter pour les futures collections**

CE QUI A ÉTÉ FAIT

- Conception et déploiement de l'enquête nationale sur Démarches Numériques **(94 questions / 3 blocs Structurants)**
- Webinaires et productions de guides et documents d'aide au remplissage
- Collecte des données structurée en trois niveaux, avec saisie en ligne ou sur fichiers Excel (DGOS/ DGRINES Préremplissage pour 65 structures de la MSER03)
- Consolidation des données sous SQLite et lancement des analyses statistiques



L'ENQUÊTE DE CARTOGRAPHIE : UN DISPOSITIF NATIONAL ET ÉVOLUTIF



CARTOGRAPHIER LES BIOCOLLECTIONS EN FRANCE

Au 18 mai 2026.

170 **Dossiers déposés**

*Issues de **103 établissements** différents (SIRET) de **78 entités juridiques** différentes (SIREN).*

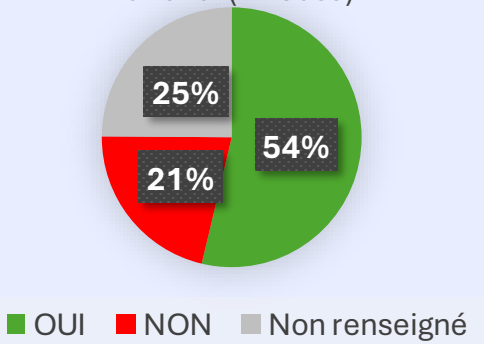
33 797 920 **Échantillons au total**

Toutes natures d'échantillons confondues

5 038 **Collections**

*Top 5 des champs **thématiques** les plus cités : Cancer, M. rares, M. inflammatoires, M. infectieuses, Neurologie.*

Publication dans le catalogue national (N=5038)

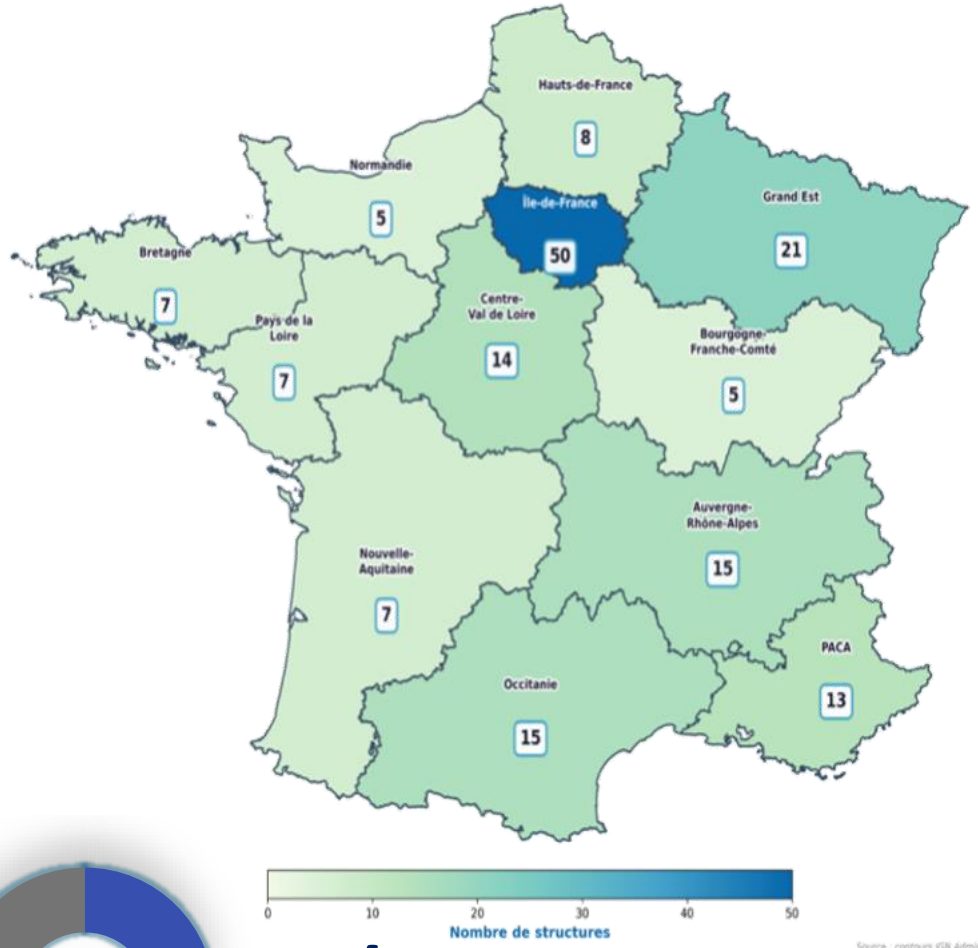


UN TRÈS FORT POTENTIEL

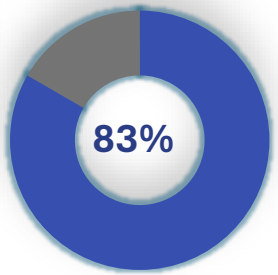
54 %

des collections sont ouvertes à la publication dans le futur catalogue national.

UNE COUVERTURE DÉJÀ NATIONALE *Métropôle et DROM)*







15/18 **Régions représentées**

La cartographie couvre 37 départements sur 101 (37 %) et 15 régions sur 18 (83 %). Régions non représentées : Corse, Guyane et Mayotte.

CERTIFICATIONS DES STRUCTURES QUI HÉBERGENT LES BIOCOLLECTIONS

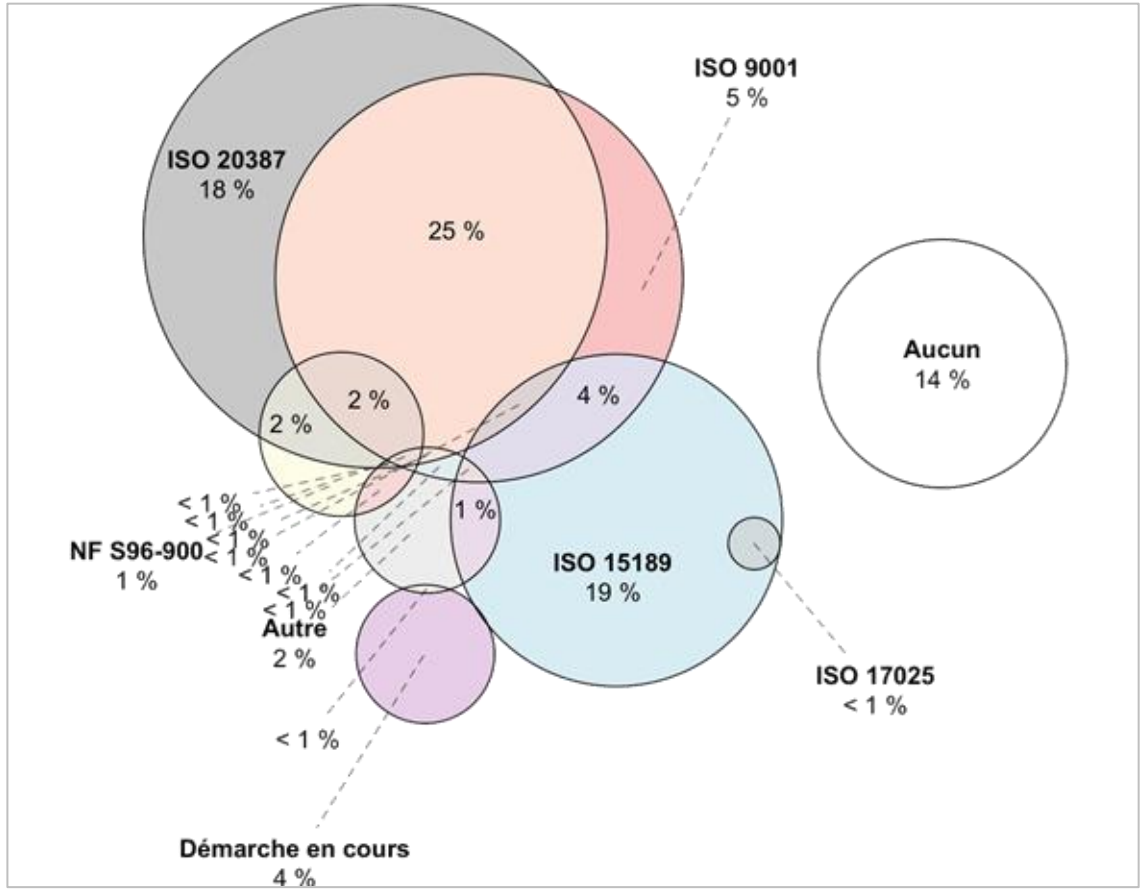


Fig. Pourcentage de structures par combinaison de certifications (n=170)

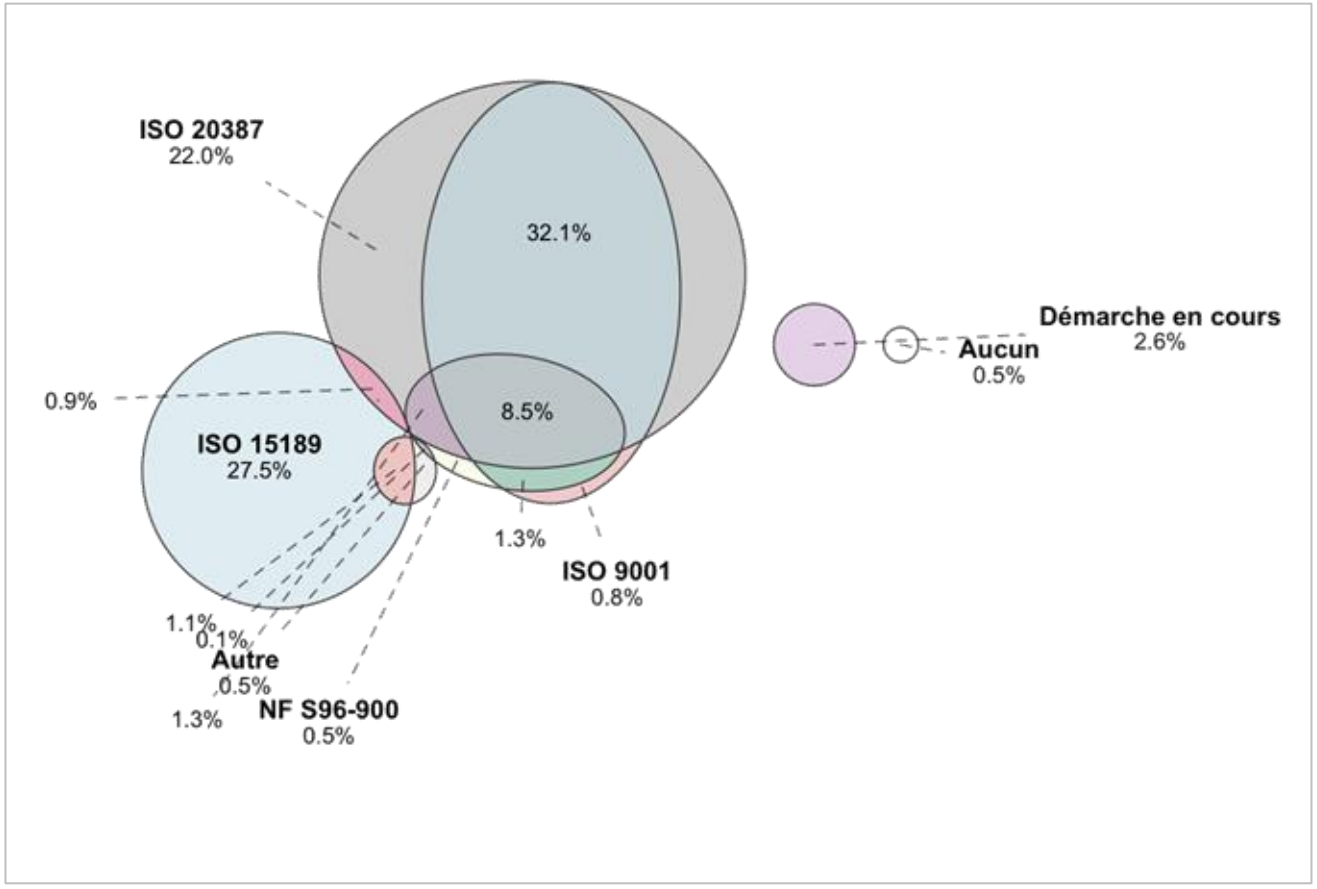


Fig. Pourcentage de collections par combinaison de certifications (N=5038 collections)

UN ECOSYSTEME D'OUTILS LOGICIELS TRES DIVERSIFIE

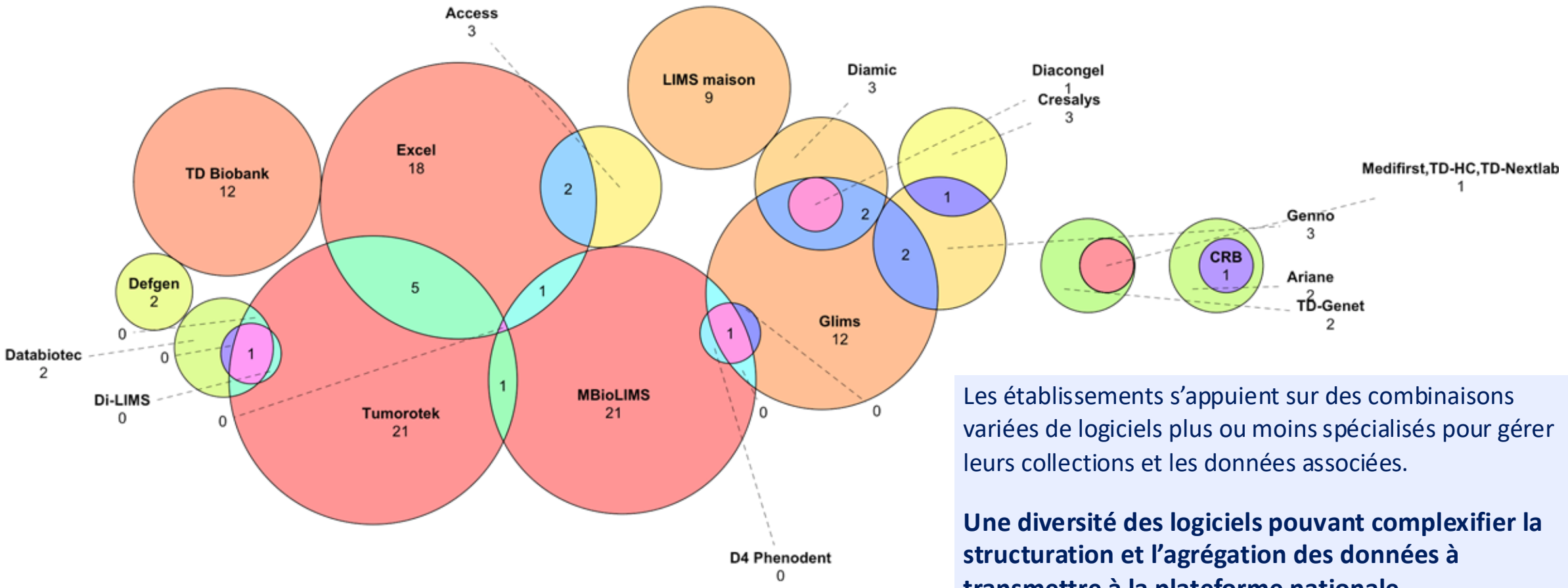


Fig. Combinaisons de systèmes informatiques utilisés comme LIMS par les établissements

Les établissements s'appuient sur des combinaisons variées de logiciels plus ou moins spécialisés pour gérer leurs collections et les données associées.

Une diversité des logiciels pouvant complexifier la structuration et l'agrégation des données à transmettre à la plateforme nationale.

CADRE DE COLLECTE DES ÉCHANTILLONS DES BIO-COLLECTIONS

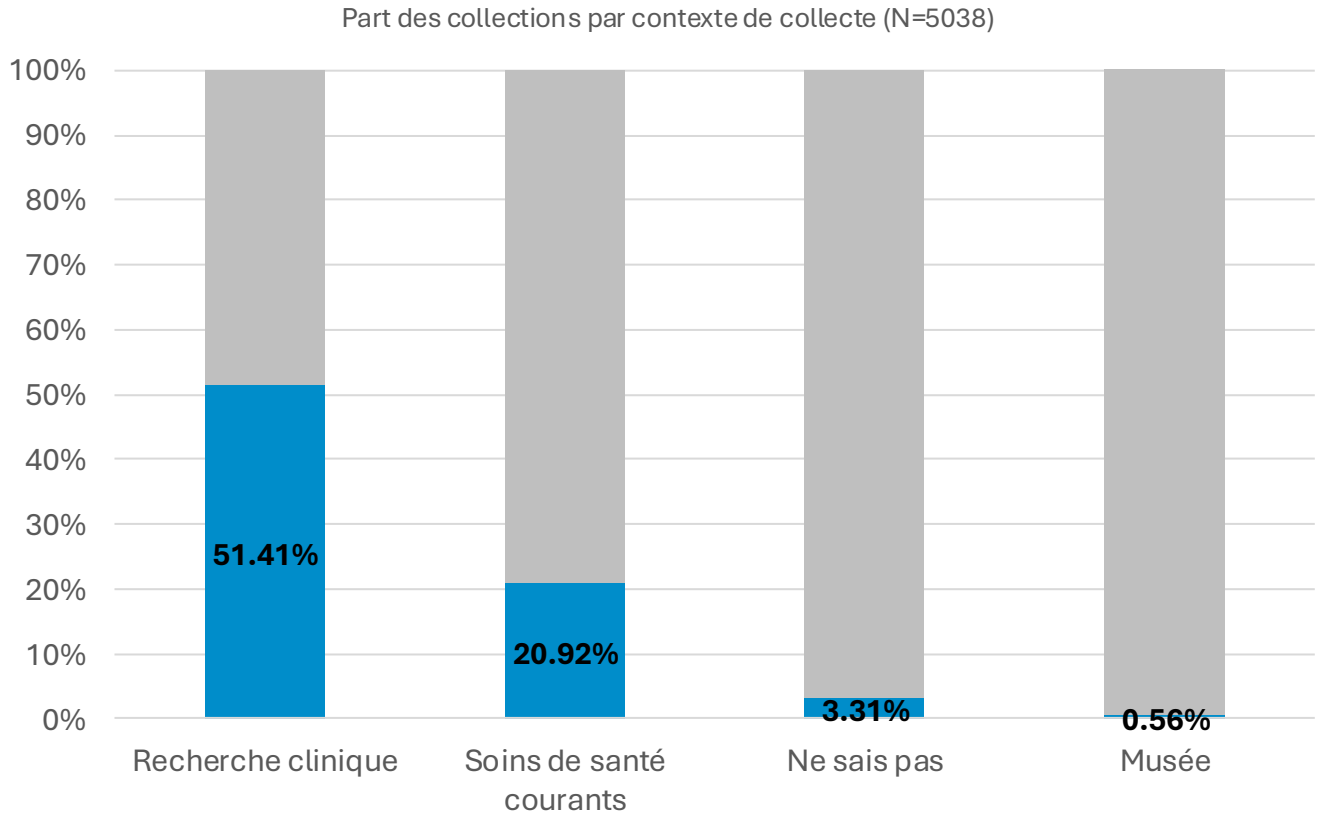


Fig. Répartition des bio-collections selon le contexte de collecte (N = 5 038)

51,4 %

des bio-collections recensées sont issues de la recherche clinique.

Les soins courants constituent le deuxième contexte de collecte, avec 20,9 % des collections.

PRIORITÉ POUR LA PHASE 2

Connecter les bio-collections aux grandes bases de données publiques de recherche clinique (FReSH, Clinicaltrials, etc.) afin d'enrichir les données associées.

Enjeu : rendre les collections plus documentées, interopérables et réutilisables.

DES DONNEES ASSOCIES ESSENTIELLEMENT CLINIQUES

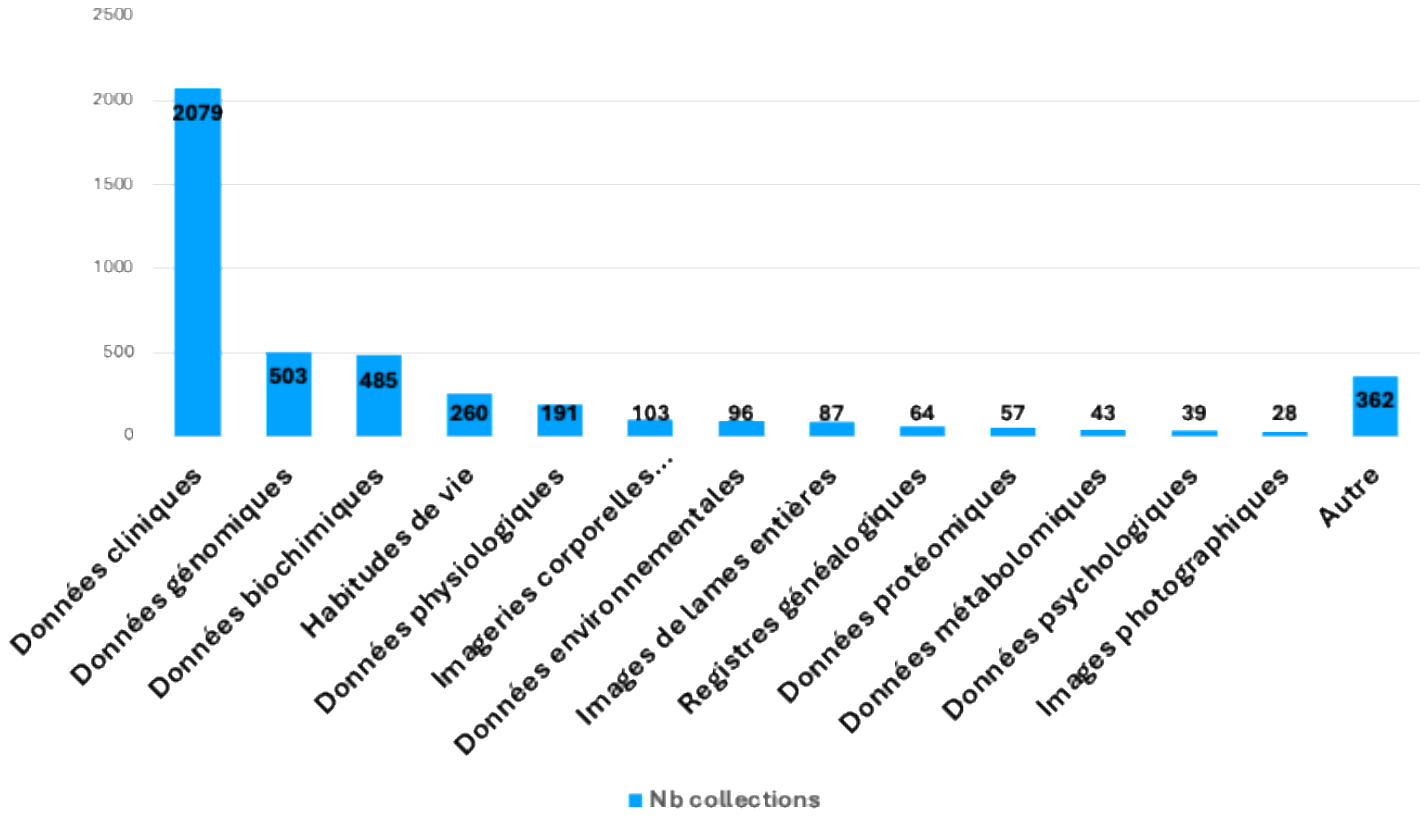


Fig. Nombre de collections selon les types de données associées

2 079

collections sont associées à des données cliniques, de très loin le type le plus représenté.

Les données génomiques (503 collections) et biochimiques (485) arrivent ensuite, à des niveaux proches.

La collecte de nouvelles données doit être guidée par les objectifs des études ; l'enjeu central de mieux structurer et rendre visibles les données disponibles pour faciliter leur identification et leur réutilisation. **Cela implique un travail d'interface étroit et structurant avec les EDS**, afin de mieux caractériser les données accessibles et de renforcer le potentiel de valorisation des bio-collections.

LES ONTOLOGIES : UN LEVIER D'INTEROPÉRABILITÉ ENCORE PEU ADOPTÉ

Trois référentiels structurent le codage des pathologies et l'annotation des échantillons.

ICD-11

PATHOLOGIES

Classification internationale des maladies, dans sa version la plus récente.

ADICAP

ANATOMIE PATHOLOGIQUE

Thésaurus dédié à l'annotation des échantillons d'anatomie et de cytologie pathologiques.

ORDO

MALADIES RARES

Ontologie Orphanet utilisée pour décrire et coder les maladies rares.

ADOPTION ENCORE MARGINALE

Le codage ontologique reste peu répandu. Il concerne surtout les structures financées par la **DGOS / DGRINES qui a impulsé la dynamique en 2024** avec le déploiement d'un 1er catalogue des biocollections des établissements de santé.

POSITIONNEMENT À ADOPTER POUR FRBIONET

1- Capitaliser sur la dynamique engagée
2- Accompagner les structures dans l'adoption de référentiels communs,
3- Associer les éditeurs de systèmes informatiques à la démarche afin de construire avec eux les conditions de leur intégration et de leur utilisation.

Les catalogues existants restent encore très perfectibles

L'enquête de cartographie révèle un double enjeu : élargir la couverture et renforcer les fonctionnalités.

63 %

des structures ne disposent d'aucun catalogue

37 % disposent d'un catalogue

Une part importante des collections reste donc difficile à repérer et à consulter en ligne.

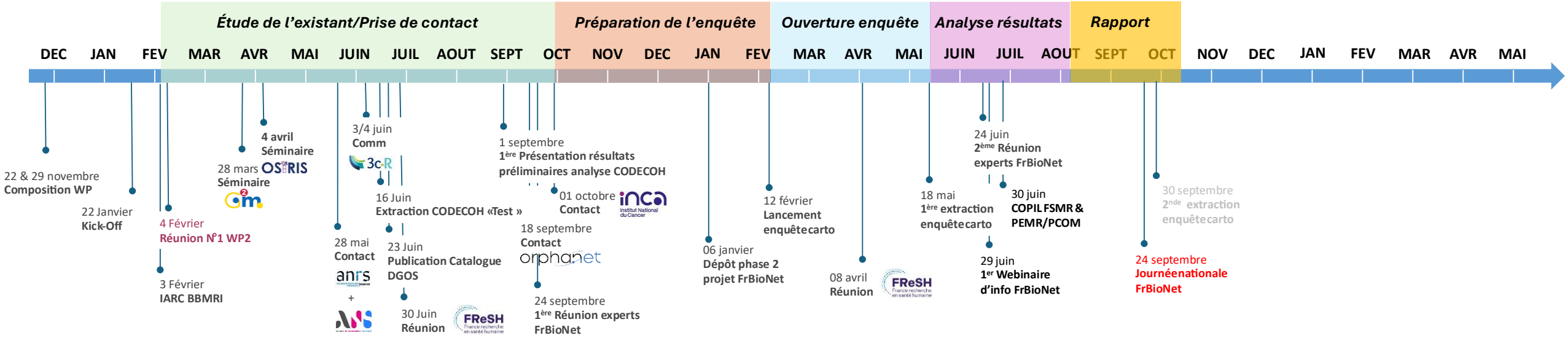
Parmi les catalogues existants…

87 %	sont à accès entièrement public
19 %	permettent des recherches et des filtres (requêtables)
7,9 %	proposent une section privée
6,3 %	sont mis à jour automatiquement

Pourcentages calculés parmi les structures disposant d'un catalogue.

PRIORITÉ POUR LE CATALOGUE NATIONAL **Rendre les collections visibles, interrogeables et régulièrement actualisées.**

L'ENQUÊTE DE CARTOGRAPHIE : UN DISPOSITIF NATIONAL ET ÉVOLUTIF

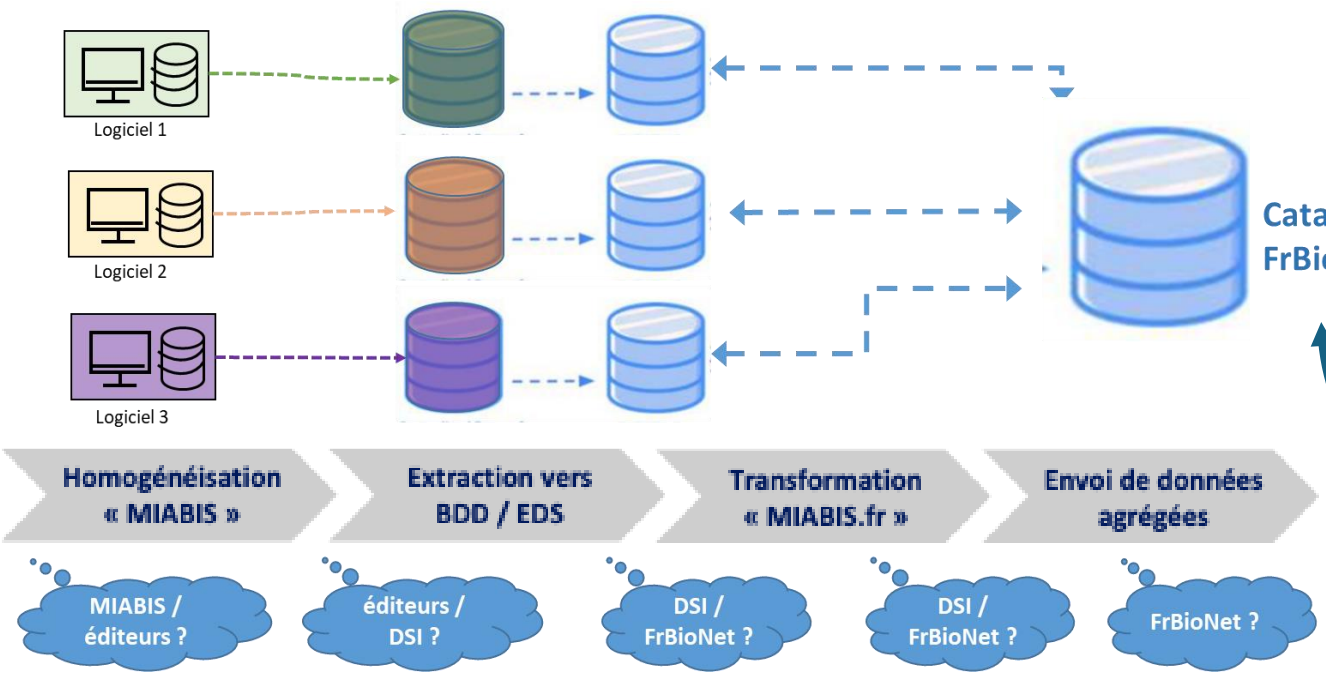


CE QUI EST À VENIR

- **La plateforme reste ouverte (+11 nouveaux dossiers déposés)**
- Améliorer le questionnaire, notamment l'outil de collecte, au fil des retours de la communauté
- Produire et publier le rapport de synthèse complet sur la page FrBioNet.

Perspectives de la phase II :
vers un catalogue national interconnecté

VERS UN CATALOGUE NATIONAL INTERCONNECTÉ



Catalogue FrBioNet

BBMRI - national nodes



Rendre visible au niveau européen / international

Grande **hétérogénéité** entre biobanques dans la manière de structurer les données, de les nommer et de les documenter.

Necessité de les mettre en correspondance et de faire communiquer les systèmes **= interopérabilité**.

FR IT4CT French IT sytems for clinical research tools

FReSH France recherche en santé humaine

Enrichir avec les données cliniques et/ou générées dans les laboratoires

DEFINIR UN MODÈLE COMMUN DE DONNÉES

OBJECTIFS

Définir un **socle commun de données pour les biobanques**, permettant d'harmoniser les pratiques, d'améliorer l'interopérabilité des collections et faciliter leur valorisation et leur réutilisation à l'échelle nationale.

MIABIS

Minimum Information About Biobank data Sharing · BBMRI-ERIC

- Standard européen de référence pour le partage des données de biobanques
- Recommandé à l'échelle de l'UE et déjà déployé (BBMRI Directory)
- Cadre commun et minimal → interopérabilité avec l'Europe et recommandé dans l'ISO 29387

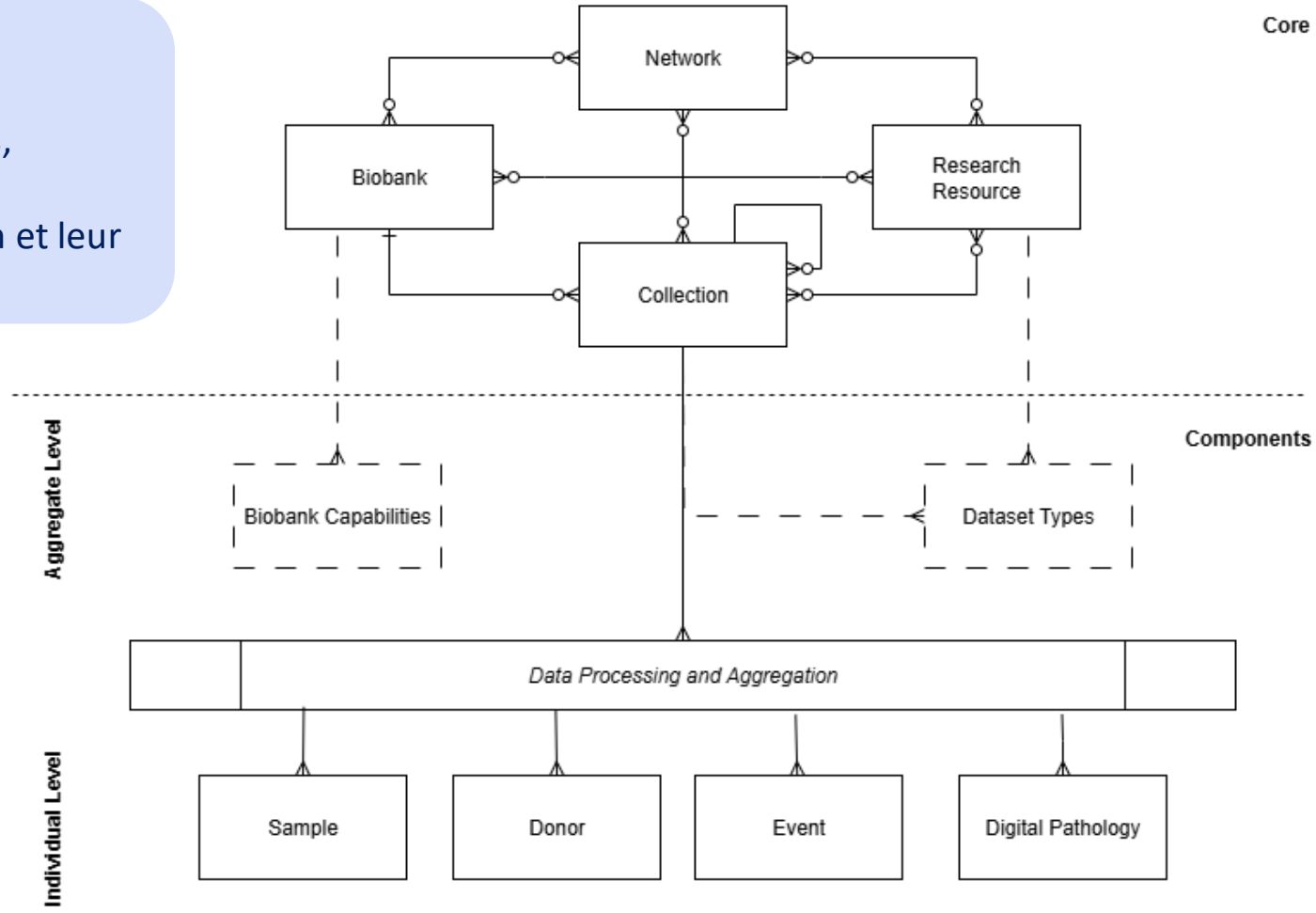


Fig. Modèle conceptuel de données du MIABIS core v3 (Eklund et al., 2024)

SELECTIONNER LES ITEMS MINIMAUX

Le data-set MIABIS qui comprend 32 attributs n'est pas exhaustif → à compléter

CONSULTATION D'AUTRES RÉFÉRENTIELS

- **SPREC (ISBER)** = pré-analytique
- **BRISQ =** descriptif experimental
- **IARC =** techniques de biobanking

CONSULTATION D'AUTRES CATALOGUES

- **BBMRI Directory =** échelle européenne
- **Swiss Biobanking Platform =** exemple d'harmonisation nationale.
- **RD-Connect =** maladies rares

CE QUI A ÉTÉ FAIT

- Consolidation de listes de données de la littérature
- Évaluation de la présence et du codage de chaque item au niveau local / chaque expert
- Sélection des items minimaux pour une v1

CE QUI EST À VENIR

- Modéliser les entités métiers et leurs relations
- Choisir et accorder les référentiels de codage (dictionnaire de données)
- Définir la gouvernance (validation, versionning, gestion des évolutions, etc.)
- Déployer sur les projets structurants le modèle retenu

CHAMPS RETENUS POUR LA V1 DU CATALOGUE

Liste des 26 champs retenus par le groupe d'experts pour la V1 du catalogue :

Biobanque

ID
Identifiant unique de la biobanque

Acronyme
Sigle ou acronyme officiel de la biobanque

Nom
Nom complet de la biobanque

URL
Adresse internet officielle de la biobanque

Établissement
Personne morale responsable de la biobanque

Pays
Pays où se situe la biobanque

Infos de contact
Informations permettant de contacter la biobanque

Description
Description synthétique de la biobanque

Services (3 champs)
Services fournis par la biobanque

Certifications qualité
Norme(s) de management de la qualité appliquée(s) à la biobanque

Collection

ID
Identifiant unique de la collection

Nom
Nom de la collection

Description
Description synthétique de la collection

Origine échantillons
La source de prélèvement des échantillons

Nature d'échantillons
Type d'échantillon collecté ou stocké

Cadre de collecte
Cadre ou circonstances de collecte des échantillons

Numéro NCT
Numéro d'enregistrement ClinicalTrials.gov

Nombre de participants (2 champs)
Nombre total de participants inclus dans la collection ou étude

Investigateur principal
Investigateur principal de la collection

Nombre total d'échantillons
Nombre total d'échantillons de la collection

Échantillon

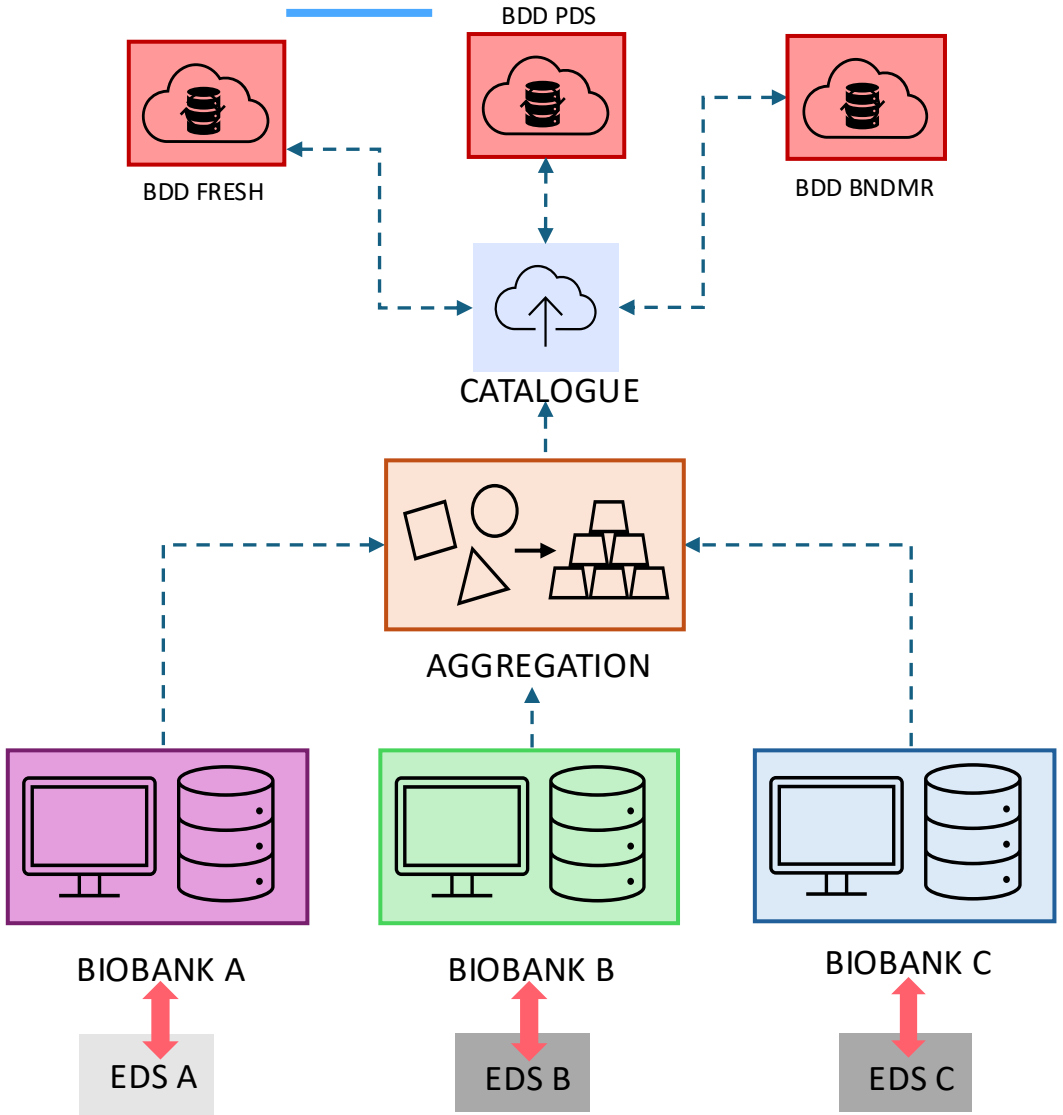
Pathologie
Maladie, affection ou pathologie associée au participant ou à l'échantillon

Température de conservation
Température de stockage long terme

Site anatomique
Site anatomique d'où est issu l'échantillon

La liste étendue comporte 208 items au total, l'objectif étant d'implémenter de manière incrémentale certains de ces items au fil du temps.

ÉTABLIR DES PREUVES DE CONCEPT EN S'APPUYANT SUR LES PROJETS STRUCTURANTS



Tester deux modes complémentaires d'alimentation de la plateforme

Étape 1 « MODE CENTRALISÉ »

Les données sources seront mises en correspondance avec les champs du catalogue. Puis les ETL seront générés depuis les structures afin d'extraire, harmoniser et agréger les données avant leur transfert vers la plateforme.

Étape 2 « MODE CONNECTÉ »

Des connecteurs automatiseront ensuite les échanges avec les bases des centres et les initiatives partenaires, notamment FRESH, BNDMR, PDS etc. afin d'alimenter directement la plateforme.

OBJECTIF Valider la faisabilité, puis préparer le passage à l'échelle nationale.

LA TRANSVERSALITÉ AVEC LES AUTRES WP POUR RENFORCER LES ACTIONS

WP2 • ACTIONS TRANSVERSALES

WP2 × WP3 × WP6

Harmoniser les pratiques et la qualité des données

LIGNES DIRECTRICES
DÉPLOIEMENT DES STANDARDS

- Promouvoir l'utilisation d'ontologie commune et de standards qualité (SPREC)
- Diffusion de guidelines pour la gestion des données.
- Renforcer la traçabilité (lecteurs 2D interfacés avec les LIMS).

WP2 × WP5

Transformer les standards en outils opérationnels

OUTILS INFORMATIQUES
INTEROPERABILITÉ et CATALOGUE

- Accompagner les CRB pour rendre leur données plus interopérables.
- Concevoir les outils informatiques nécessaires à l'intégration de leur données dans le catalogue.

WP2 × WP4

Harmoniser les termes et simplifier les échanges de données

OUTILS
GLOSSAIRE, NOTICE INFOBOX, CODECOH

- Mise en cohérence des travaux du WP4 avec l'utilisation des standards (glossaire, cadre de la collecte et consentement)
- Réflexion avec les instances (MESR, DGOS/DGRINES) à une meilleure interopérabilité avec leurs outils informatiques.

WP2 : quelques chiffres



24 membres experts

De tous horizons : scientifiques, Data, IT, interopérabilité, responsables de biobanques, chefs de projets, etc.

→ INSERM, CHU, CLCC, Unicancer, DGOS/DGRINES, FReSH, Institut Pasteur, PSCC, CNCR, ESPIC, etc.



23 réunions

Depuis juin 2025



Work package leaders : 2 responsables de réseaux

Séverine TABONE-EGLINGER *(Centre Léon Bérard, Animatrice du réseau des CRB d'UNICANCER) et* ***Yves-Edouard HERPE*** *(CHU Amiens-Picardie, Coordinateur du réseau national des CRB de CHU)*

WP2 : nos experts

Merci à tous ☺





Karine CHANTREL-GROUSSARD	Institut Thématique Génétique Génomique et Bioinformatique - Inserm
Paul HOFMAN	Université Côte d'Azur, CHU de Nice et IHU RespirERA
Edouard TUAILLON	CHU de Montpellier, UMR INSERM 1058
Emeline DUPONT	CRB de Touraine - CHRU de Tours
Sophie BROUTIN	UNICANCER CRB Gustave Roussy
Mikaël MAZIGHI	RESPIC - hôpital Rothschild
Bruno COUTARD	Aix Marseille Université
Odette MARIANI	UNICANCER Institut Curie
Stanie GAETE	CRB CHU de la Guadeloupe
Véronique ROBERT	UNICANCER
Julien HUYARD	UNICANCER
Clément ROCHER	CRB du CHU de Bordeaux
Théo BLOQUET	Biobanque de Picardie
Caroline JAMOT	CHU Clermont Ferrand
Moussa EL ASSLANI	CRB-TBR CHU Toulouse
Vincent LETEXIER	Centre Léon Bérard
Rémy ARTUS	Institut Pasteur
Blandine RIMBAULT	Inserm ANRS MIE
Lesya BAUDOIN	INSERM/FReSH
Cécile GIRARD	CHU Nantes
Hélène BLASCO	CHU Tours
Jérôme LEMONNIER	Unicancer
Frédéric VÉLY	PSCC
Zhanna SANTYBAYEVA	DGOS

WP2 : perspectives d'évolution

Intéressé.e par les travaux du WP2 cartographie & les bases de données ?

Vous souhaitez contribuer ?



Appel à participation au lancement de la phase 2

Merci pour votre attention