cad Collecteur Analyseur de Données
La santé du futur s'inscrit dans notre ADN

Collecteur Analyseur de Données

Projets de recherche

Mise en œuvre

Fonctionnement des bulles sécurisées

31 mars 2025 - bioinfodiag

FRANCE 2030

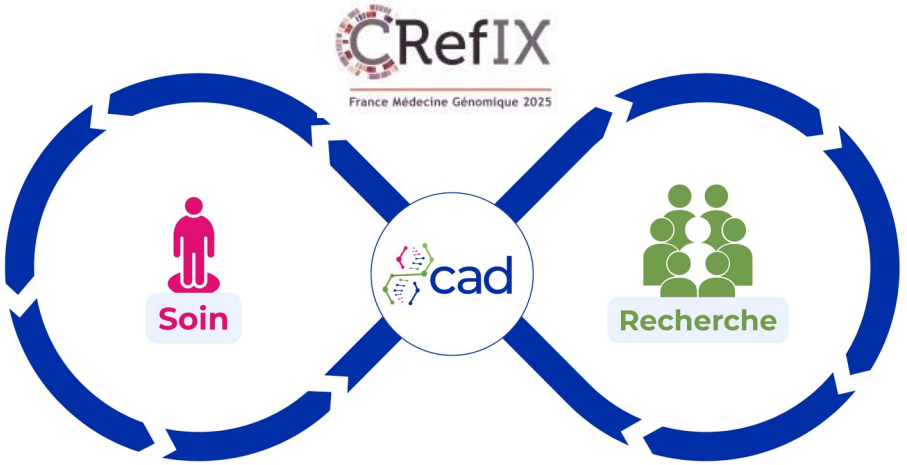
Infrastructures du plan

SeqOIA, Auragen, CRefIX, CAD

Le Plan France Médecine Génomique (PFMG2025) a pour objectif d'implémenter la Médecine Génomique en France dans un continuum soin-recherche. Il s'appuie sur 3 piliers principaux :

Un réseau de laboratoires de séquençage génomique (LBM-FMG pour *Laboratoire de Biologie Médicale France Médecine Génomique)*

Un Centre de Référence, d'Innovation, d'expertise et de transfert (CRefIX)



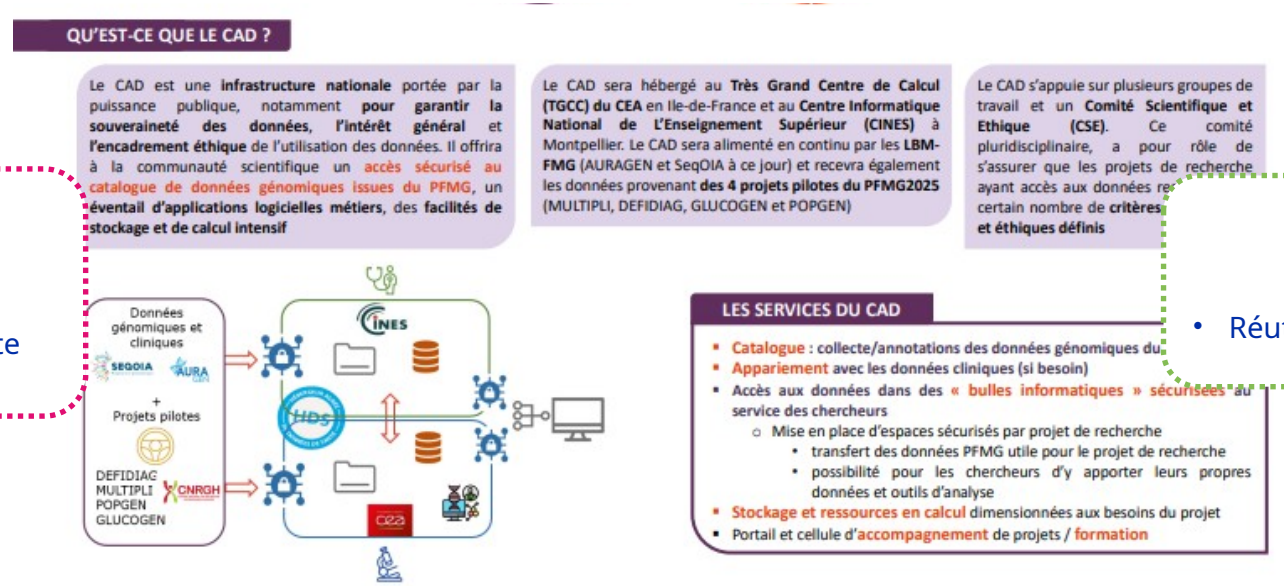
Un centre national de Collecte et d'Analyse des Données (CAD)

Collecteur Analyseur de Données
Quelques dates

- Appel à soumission de projets en 2022
- Recrutements des équipes courant 2024
- Autorisation EDS - juin 2024
- Données des LBM / infrastructure / premiers projets ouverts en cours

CAD <> Soins

- Ré-analyse des données dans la poursuite de la démarche diagnostique



CAD <> Recherche

- Réutilisation des données pour la recherche

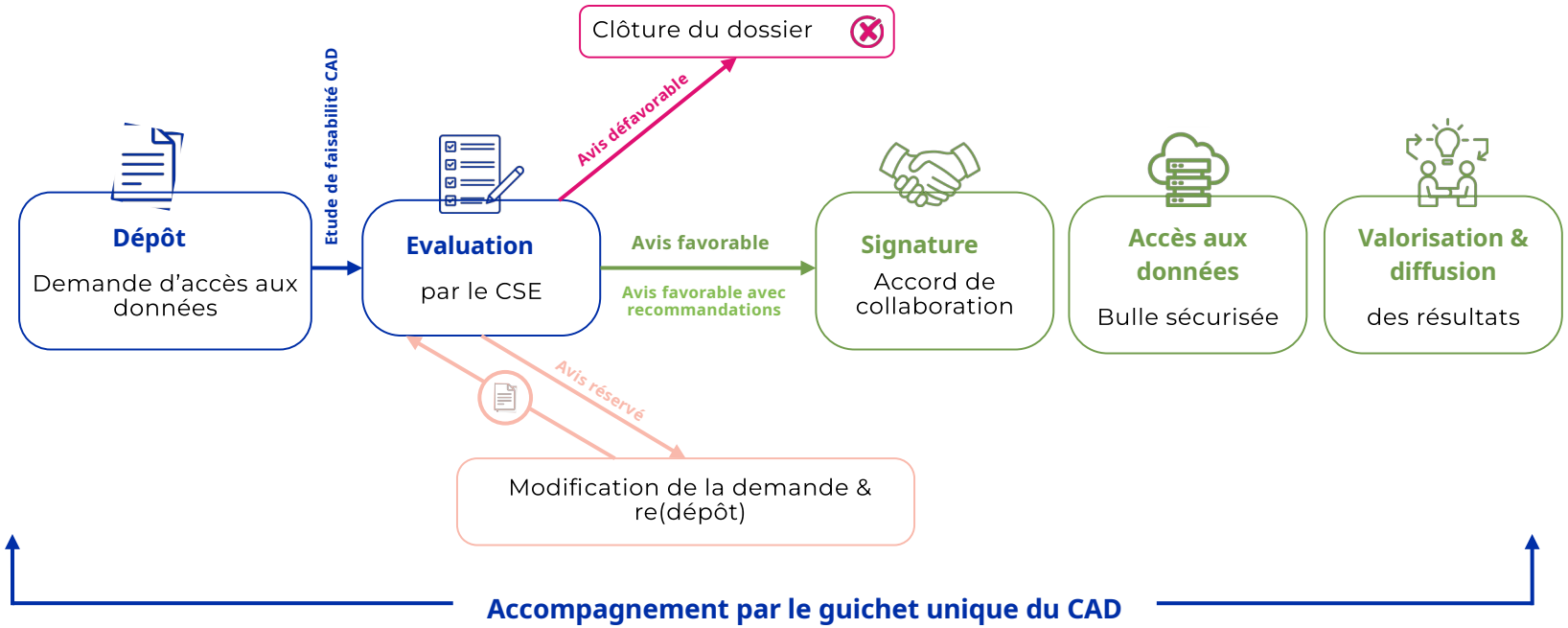
Projets de recherche
Cellule d'accompagnement des projets

Projets de recherche

- Conseil Scientifique et Ethique
- Données des Laboratoires de Biologie Médicale
- Bulle sécurisée - outils / données

Vers les "nouveaux diags"

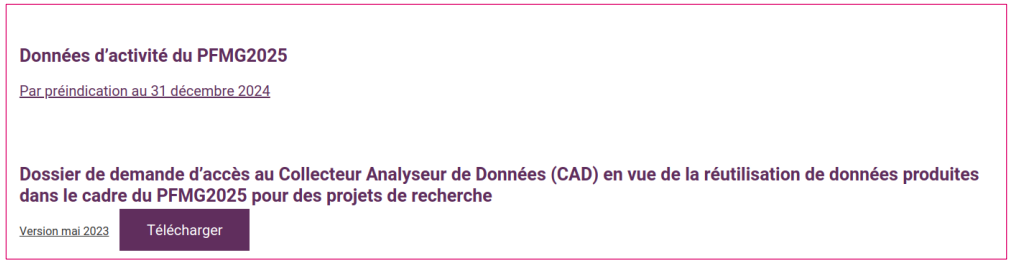
Accompagnement de toutes les étapes



Comment déposer une étude?

https://pfmg2025.aviesan.fr/acces-aux-donnees/

- Données générées par les LBM ?
- Demande d'accès : etude@genomecad.fr



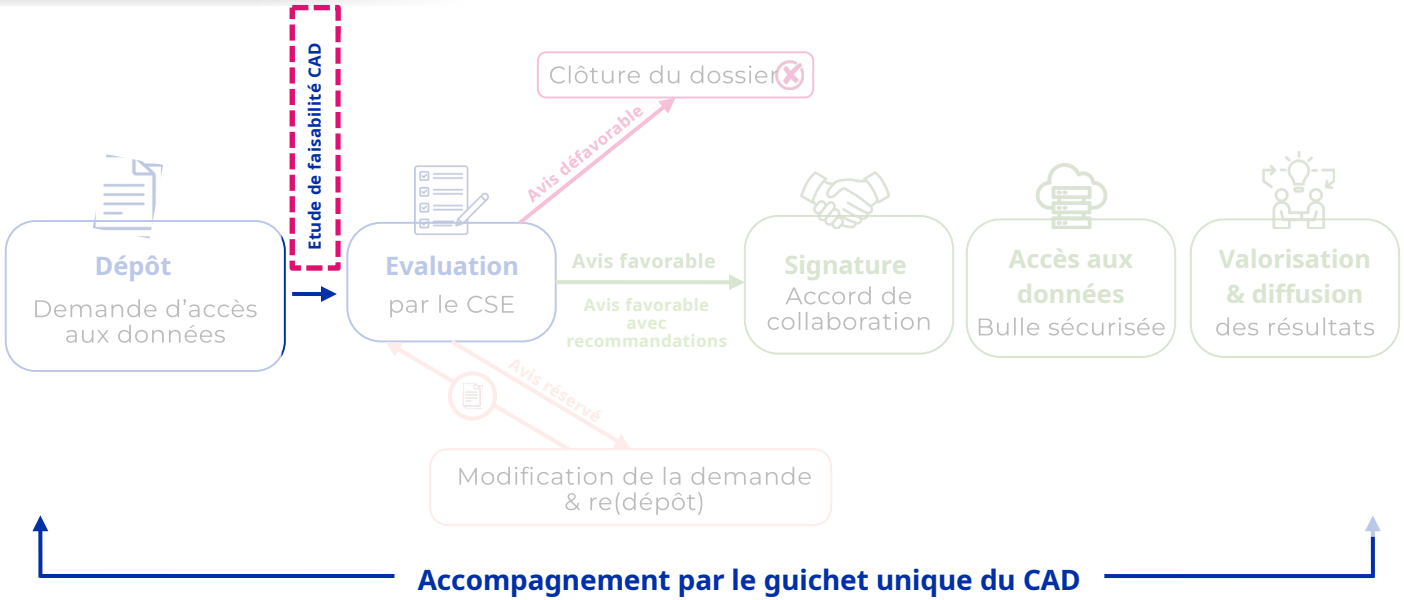
Données d'activité du PFMG2025 au 31 décembre 2024

	Nb de prescriptions validées en RCP-FMG	Nb de dossiers complets	Nb de dossiers clos
Angiodèmes bradykiniques héréditaires	3	3	2
Anomalies du développement, syndromes malformatifs et syndromes dysmorphiques sans déficience intellectuelle	9 098	8 123	6 040
Anomalies sévères de la différenciation sexuelle d'origine gonadique et hypothalamo-hypophysaire	186	141	96
Aplasies et hypoplasies médullaires	65	59	52

Mise en place "faisabilité CAD" : pourquoi?

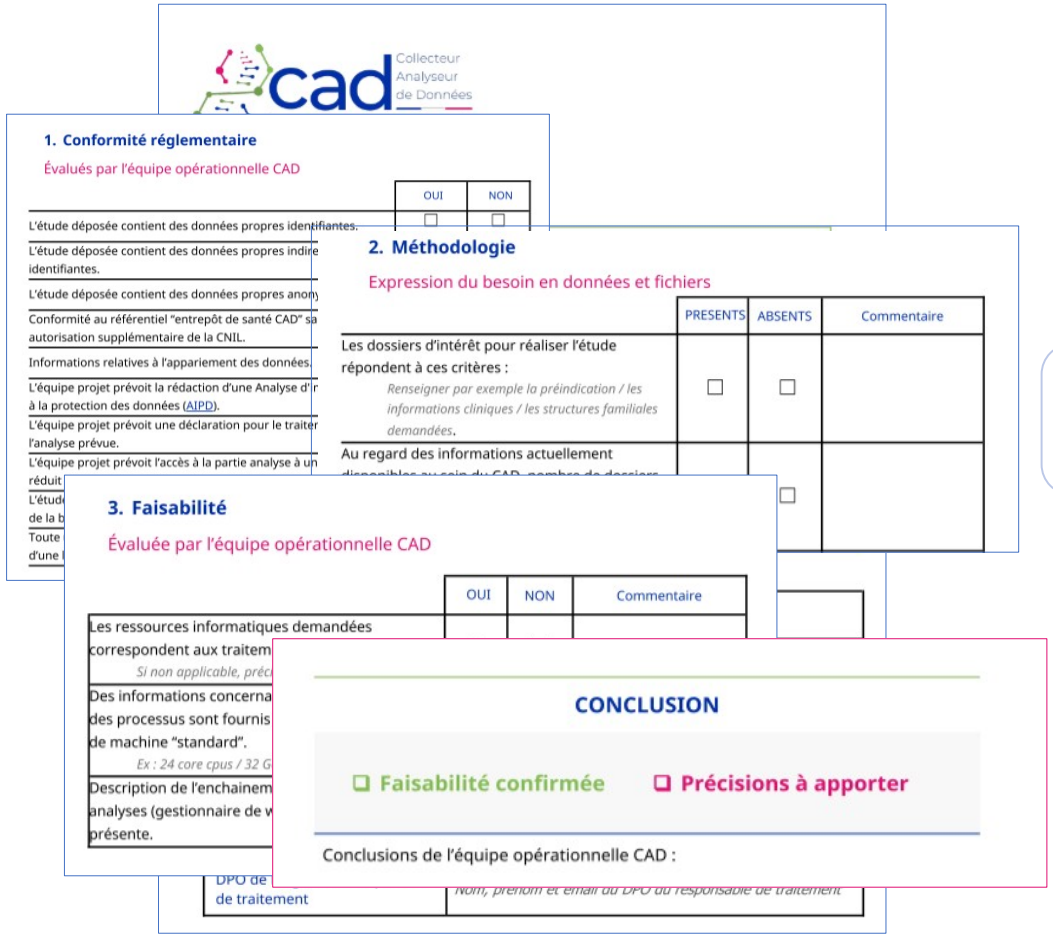
Avant mise en place d'une "faisabilité CAD"

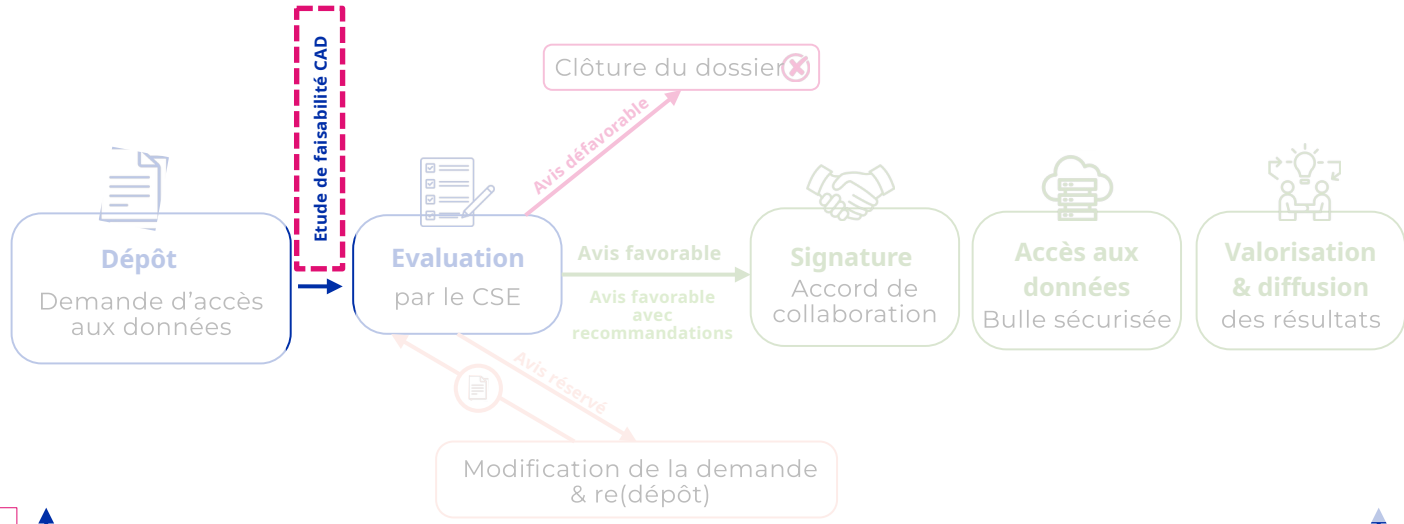
- Conformité réglementaire ?
- Est ce que les données demandées sont générées ? Disponibles au sein du CAD ?
- Ce qui est décrit dans le projet, est-ce faisable au CAD ? Outils … infra … ?



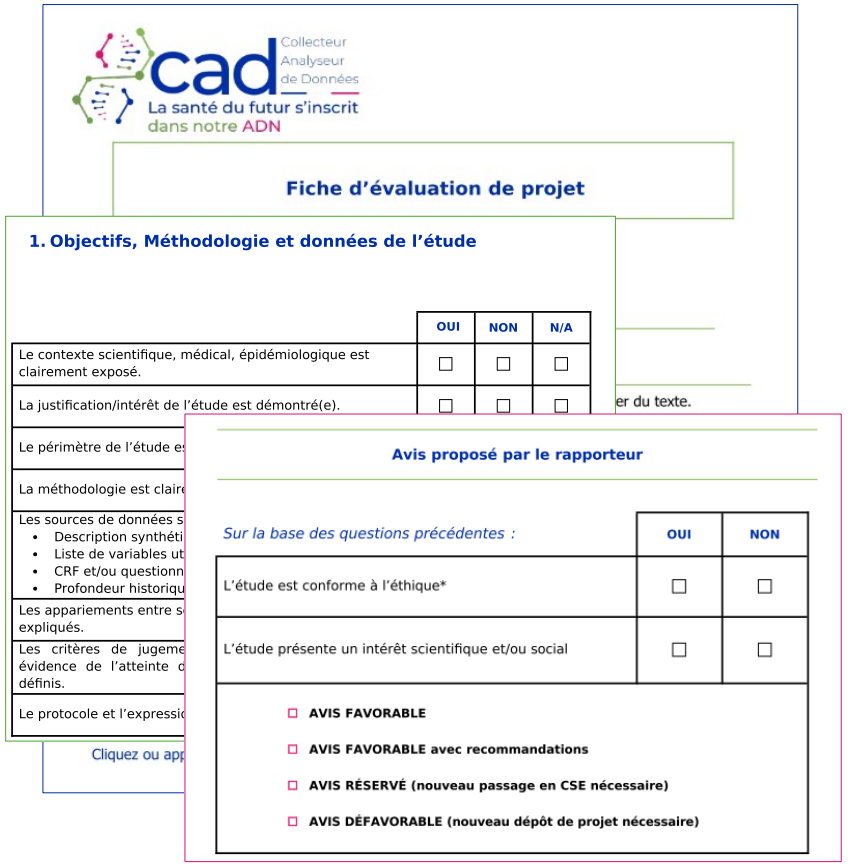
- Proposition d'accompagnement Faisabilité CAD
- Structurer pour limiter les projets qui stagnent
- Accompagner CSE

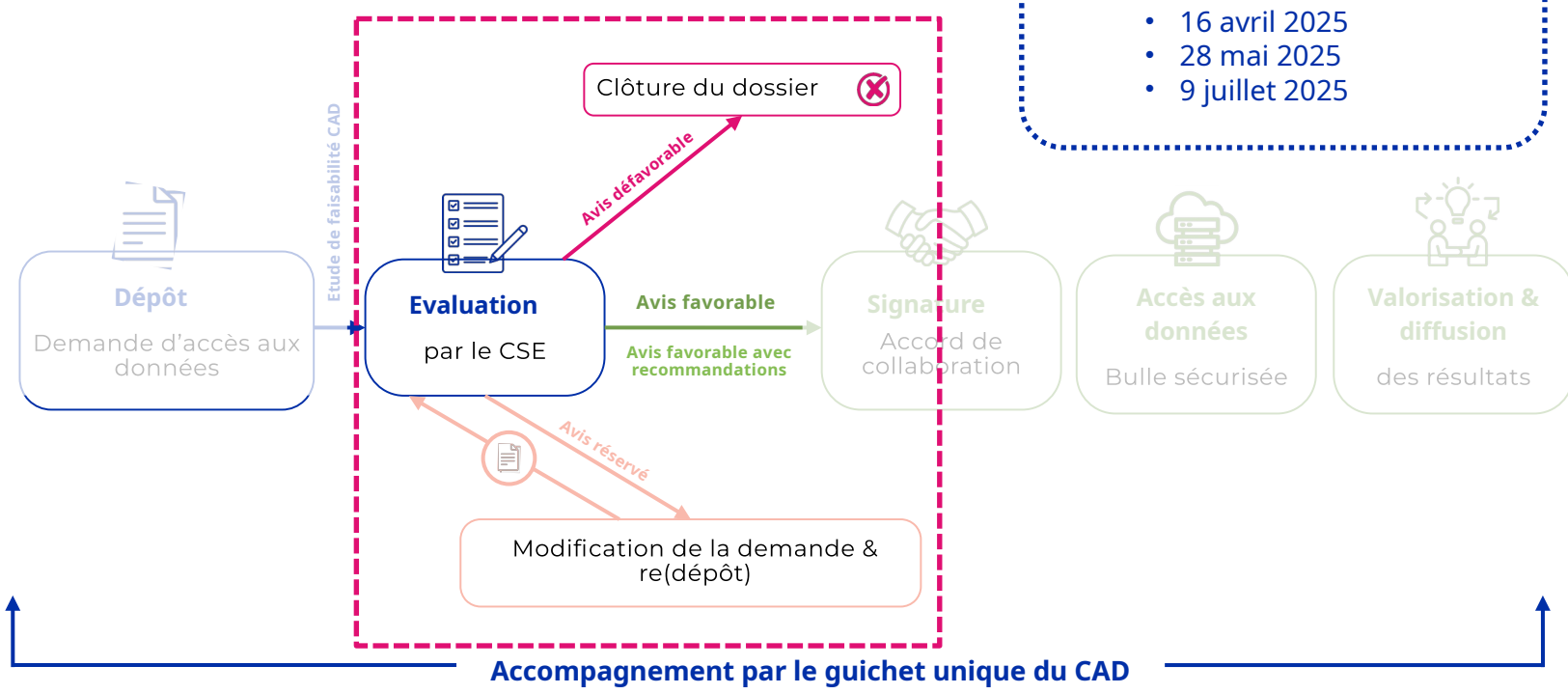
Faisabilité CAD : retours aux membres CSE & porteurs des projets





Evaluation par le Conseil Scientifique et Ethique - CSE





Dates réunion CSE

- 22 janvier 2025
- 5 mars 2025
- 16 avril 2025
- 28 mai 2025
- 9 juillet 2025

2 rapporteurs par projets discuté, discussion avec tout le CSE

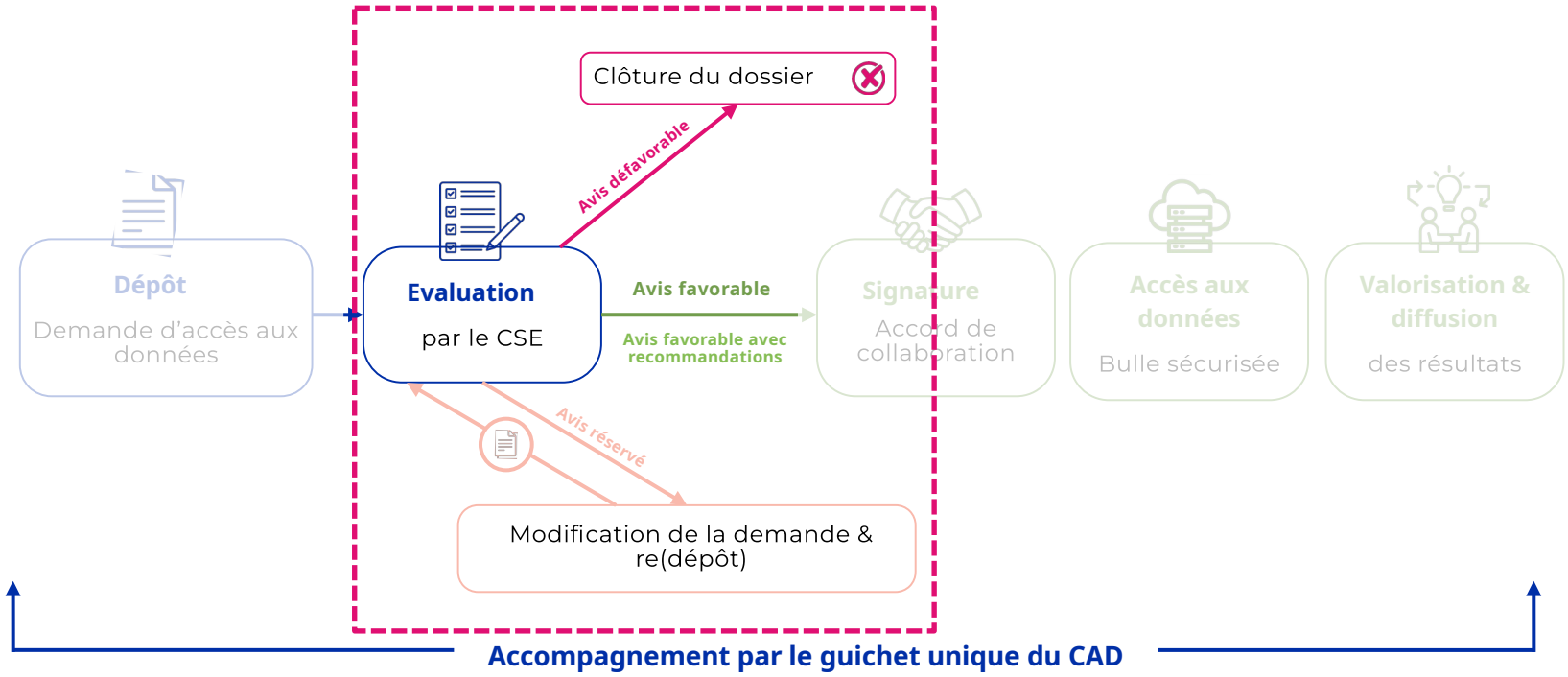
Les CSE de 2022 à 2024 : 17 projets favorables

Évaluation CSE	2022	2023	2024
Défavorable	-	1	-
Favorable avec recommandations			
Favorable	3	8	6
Réservé	5	13	8

- Faisabilité CAD = atouts CSE
- Action pour mise en oeuvre des bulles

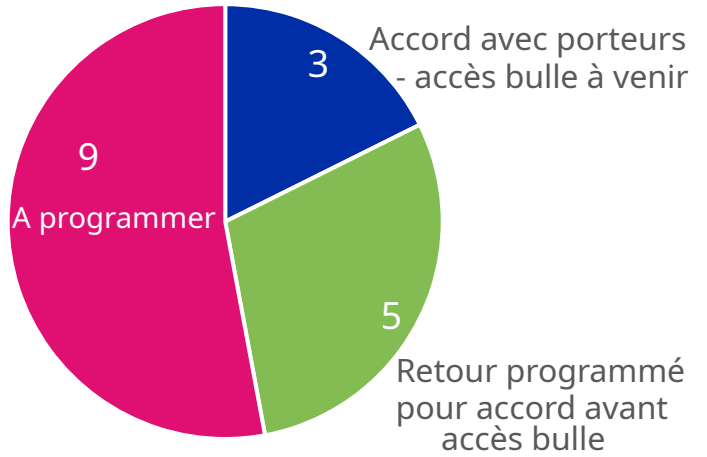
https://pfmg2025.aviesan.fr/projets-de-recherche-2/

Faisabilité CAD en pratique



- Faisabilité pour acter les outils / données à mettre en accès dans les bulles sécurisées
- **2025 : Mise en place des accès aux bulles**

Faisabilité CAD



Transfert des données des LBM

Soin & transfert des données
- de prescription
- de séquençage (dossiers interprétés)

- Qualité / population de référence
- Catalogue d'outils, annotations
- Besoins spécifiques du projet
- Pseudonymisation

Projet X

- Données laboratoires pseudonymisées
  - Critères cliniques
- Données propres
- Outils du catalogue / propres au projet
- Puissance / stockage demandé
- Jeu de référence asymptomatique

Projet Y

- Données laboratoires pseudonymisées
  - Critères cliniques
- Données propres
- Outils du catalogue / propres au projet
- Puissance / stockage demandé
- Jeu de référence asymptomatique

Validations internes avant mise en œuvre projets

Réglementaire

Conforme “Entrepot Santé CAD"
Contractualisation
Analyse d’impact relative à la protection des données
Déclaration Health Data Hub
Habilitation membres du projet

Formalisation / procédures en cours de rédaction

Logiciels / bulle

Faisabilité CAD : précision du besoin logiciels
Fiche technique bulle sécurisée
Configuration espace logiciels
Ressources de calcul et stockage
Données de référence dans la bulle
Bulle sécurisée HDS accessible

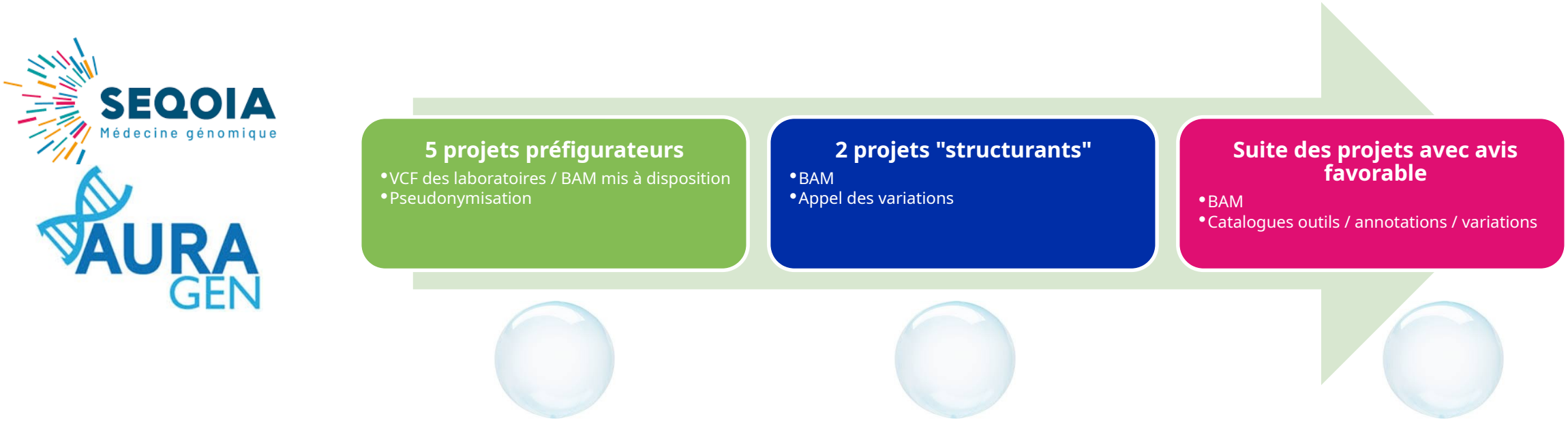
Données

Faisabilité CAD - précision du besoin des porteurs de projet
Génération des données sur les plateformes
Transfert des données PFMG dans la bulle
Données propres attendues dans la bulle
Vérification des données transférées dans la bulle

Ressources

Bioinformaticien
Espace collaboratif CAD

Trois grandes étapes de mise en œuvre
Projets avec données des LBM



Projets "structurants"

Disponibilité des équipes pour rendre la bulle fonctionnelle et permettre au projet de se réaliser

SEQOIA Médecine génomique

AURA GEN

doublons
...

Pseudonymisation

Bulle sécurisée

Données LBM

Outils spécifiques

Population de référence Défidiag
- SNV
- CNV
- SV

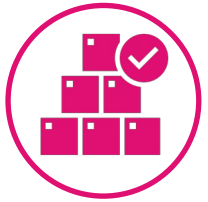
Besoin d'ajouter des outils ? Des données ?
Ex :
- Visualisation des données
- Dictionnaire HPO

Catalogue CAD
- Outils / annotations
- Variations

Mise en place d'un **"welcome pack"** défini en fonction des besoins pointés par projets structurants

Sera incrémenté avec le temps

Catalogues qui vont s'incrémenter



Dictionnaire des termes HPO
Outil de visualisation IGV
Python

Annotations structurales et fonctionnelles

- Annotations structurales - gencode / RefSeq
- Annotations géniques - HGNC / OMIM
- Données génomiques - TCGA
- Annotations cliniques et phénotypiques - Orphanet / HPO
- Annotations populationnelles - 1000 genomes / gnomAD exome / genome
- Annotations pathogénicité - COSMIC / ClinVar
- Annotations d'impact - CADD / dbNSFP
- … en cours!

Outils d'appel des variations

- samtools 1.19
- snpeff 5.2
- snpsift 5.2
- finsurf a2c989c
- bedtools 2.31.1
- bedops 2.4.41
- R 4.3.2
- epacts 3.4.2
- spip 2.1
- vep 111
- … en cours

Portail CAD : objectif transparence et requêtage des données présentes dans l'EDS / des catalogues

	Nb de prescriptions validées en RCP-FMG	Nb de dossiers complets	Nb de dossiers clos
Angiodèmes bradykiniques héréditaires		3	2
Anomalies du développement, syndromes malformatifs et syndromes dysmorph		8 123	6 040
Anomalies sévères de la différenciation sexuelle d'origine gonadique et hypotha		141	96
Aplasies et hypoplasies médullaires	65	59	52

Aujourd'hui :
Données fournies par le PFMG
+ faisabilité CAD

A construire :

Informations requêtable : dossiers accessibles pour projet

- Préindication?
- Informations cliniques ?
- Structures familiales
- … à définir en fonction des besoins projets

- Catalogue "variations" ?
  SNV CNV SV …

Quels outils sont installés ?

- Variations du nombre de copies ?
  Détail
- Fusions ?
- ..

Quelles annotations peuvent être utilisées ?

- Annotations structurales ?
  Détail
- Annotations fonctionnelles ?

Projet favorable CSE – dans la bulle ?

Outils / ressources mises à disposition **"Welcome pack"**

Données populationnelles disponibles au CAD (DEFIDIAG, ...)

Outils du catalogue qui répondent aux questions du projet

Outils spécifiques validés "faisabilité CAD"…….

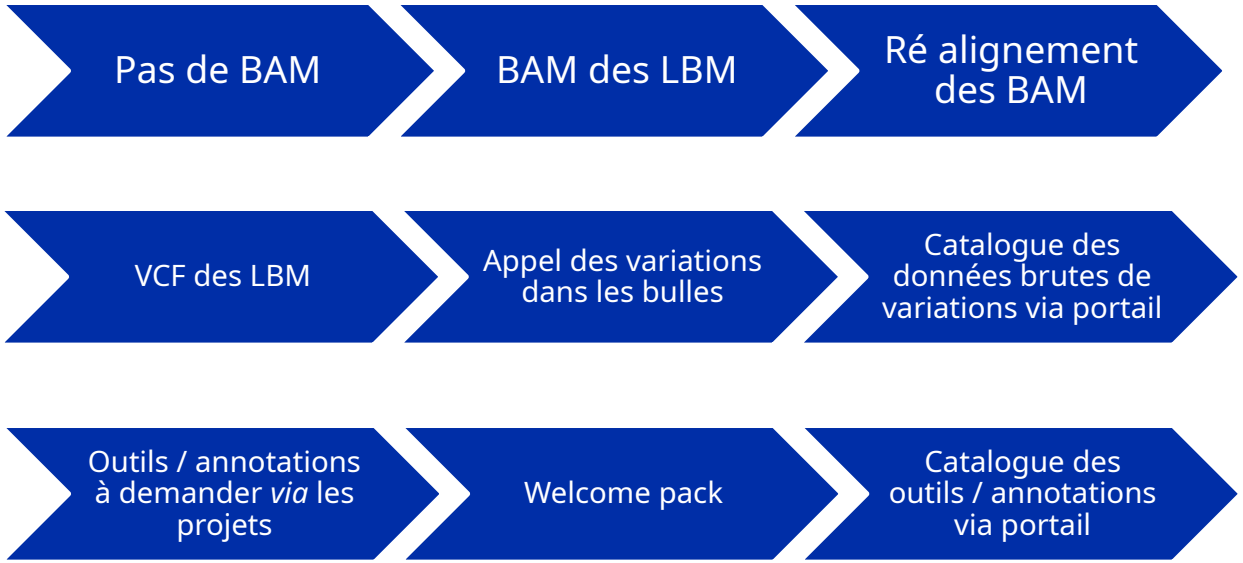
Ressources d'annotation du catalogue

Annotations spécifiques validés "faisabilité CAD"…….

Données des LBM – informations cliniques

Données propres des projets - déclaré dans le projet

Fonctionnement actuel, et objectifs



Autonomie des porteurs à requêter les catalogues outils / annotations / variations

Processus automatisés

Calendrier 2024 - 2026

cad Collecteur Analyseur de Données
La santé du futur s'inscrit dans notre ADN

	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12
	2024					2025												2026											
Projet MR VCF (BAM attendu) **Données LBM**																													
Projet MR VCF (BAM attendu) **Données LBM**																													
Projet cancer BAM **Données LBM**																													
Projet MR BAM + VCF **Données Défidiag**																													
Projet MR BAM + VCF **Données Défidiag**																													
Projet MR VCF (sans BAM) **Données LBM**																													
Projet MR VCF (sans BAM) **Données LBM**																													
Projet MR structurant **Données LBM avec cohorte**																													
Projet cancer structurant **Données LBM avec cohorte**																													
Suite des projets cancer /																													

Projets préfigurateurs

Projets pilotes

Projets "structurants"

Suite des

Inserm 2 projets maladies rares
1 projet cancer

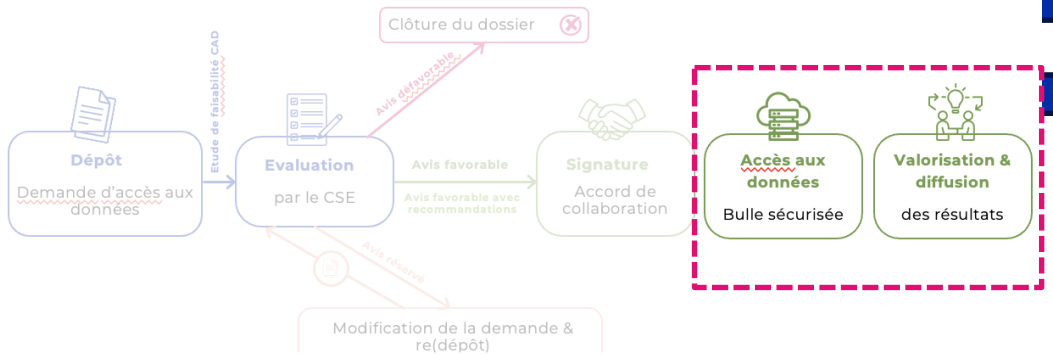
cea CNRGH 2 projets maladies rares

Inserm 2 projets maladies rares

1 projet maladies rares
1 projet cancer

Accès aux données & valorisation

	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12
	2024					2025												2026											
Projet MR VCF (BAM attendu) **Données LBM**																													
Projet MR VCF (BAM attendu) **Données LBM**																													
Projet cancer BAM **Données LBM**																													
Projet MR BAM + VCF **Données Défidiag**																													
Projet MR BAM + VCF **Données Défidiag**																													
Projet MR VCF (sans BAM) **Données LBM**																													
Projet MR VCF (sans BAM) **Données LBM**																													
Projet MR structurant **Données LBM avec cohorte**																													
Projet cancer structurant **Données LBM avec cohorte**																													
Suite des projets cancer /																													



Accompagnement par le guichet unique du CAD

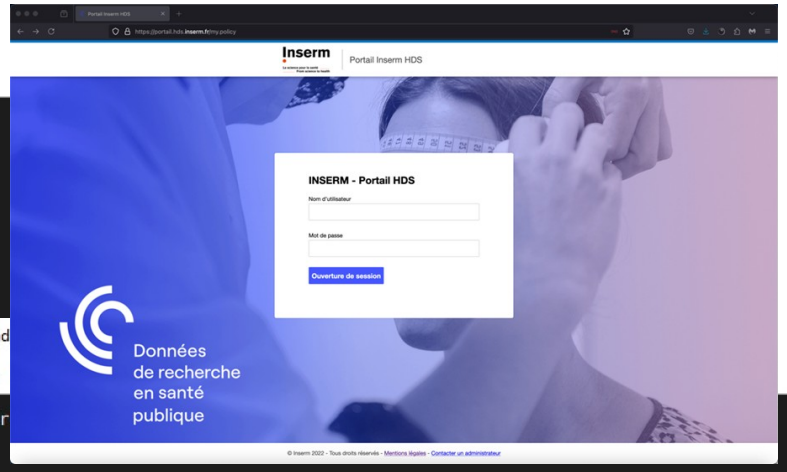
Bulle sécurisée

cad Collecteur Analyseur de Données
La santé du futur s'inscrit dans notre ADN

8. dc5vmscad-[redacted].hds.inserm.fr (test_cad)

Re-attach | Fullscreen | Stay on top | Duplicate | Hide toolbar

```
[test_cad@dc5vmscad-[redacted] ~]$ cad_containers_inventory
Inventory of tools:
vep
samtools
snpeff
spip
r[redacted]
bedtools
bedops
snpsift
finsurf
epacts
[test_cad@dc5vmscad-[redacted] ~]$ bedtools --version
bedtools v2.31.1
[test_cad@dc5vmscad-[redacted] ~]$ bedtools --help
bedtools is a powerful toolset for genome arithmetic.

Version:   v2.31.1
About:     developed in the quinlanlab.org and by many contributors worldwide.
Docs:      http://bedtools.readthedocs.io/
Code:      https://github.com/arq5x/bedtools2
Mail:      https://groups.google.com/forum/#!forum/bedtools-discuss

Usage:     bedtools <subcommand> [options]

The bedtools sub-commands include:

[ Genome arithmetic ]
    intersect     Find overlapping intervals in various ways.
    window        Find overlapping intervals within a window around an interval.
    closest       Find the closest, potentially non-overlapping interval.
    coverage      Compute the coverage over defined intervals.
    map           Apply a function to a column for each overlapping interval.
    genomecov     Compute the coverage over an entire genome.
    merge         Combine overlapping/nearby intervals into a single interval.
    cluster       Cluster (but don't merge) overlapping/nearby intervals.
    complement    Extract intervals _not_ represented by an interval file.
    shift         Adjust the position of intervals.
    subtract      Remove intervals based on overlaps b/w two files.
    slop          Adjust the size of intervals.
    flank         Create new intervals from the flanks of existing intervals.
    sort          Order the intervals in a file.
    random        Generate random intervals in a genome.
    shuffle       Randomly redistribute intervals in a genome.
    sample        Sample random records from file using reservoir sampling.
    spacing       Report the gap lengths between intervals in a file.
    annotate      Annotate coverage of features from multiple files.

[ Multi-way file comparisons ]
    multiinter    Identifies common intervals among multiple interval files.
    unionbedg     Combines coverage intervals from multiple BEDGRAPH files.

[ Paired-end manipulation ]
    pairtobed     Find pairs that overlap intervals in various ways.
    pairtopair    Find pairs that overlap other pairs in various ways.

[ Format conversion ]
    bamtobed      Convert BAM alignments to BED (& other) formats.
    bedtobam      Convert intervals to BAM records.
    bamtofastq    Convert BAM records to FASTQ records.
```



8. dc5vmscad-[redacted].hds.inserm.fr (test_cad

Re-attach | Fullscreen | Stay on top | Close

```
[test_cad@dc5vmscad-[redacted] ~]$ tr
/data
├── cad
│   ├── [redacted]71_MR
│   │   ├── [redacted]86.father.vcf.gz
│   │   ├── [redacted]86.father.vcf.tbi
│   │   ├── [redacted]71.index.vcf.gz
│   │   ├── [redacted]71.index.vcf.tbi
│   │   ├── [redacted]41.mother.vcf.gz
│   │   ├── [redacted]41.mother.vcf.tbi
│   │   └── metadata.json
│   ├── pop
│   │   ├── [redacted]20.vcf.gz
│   │   ├── [redacted]20.vcf.tbi
│   │   ├── [redacted]78.vcf.gz
│   │   └── [redacted]78.vcf.tbi
│   └── refs
│       ├── CADD
│       │   └── v1.6
│       │       ├── CADD_v1.6.sha256CADD_v1.6.sha256
│       │       ├── whole_genome_SNVs.tsv.gz.tbiwhole_genome_SNVs.tsv.gz.tbi
│       │       └── whole_genome_SNVs.tsv.gzwhole_genome_SNVs.tsv.gz
│       └── dbNSFP
│           └── 4.2a
│               ├── dbNSFP4.2a_grch38.gzdbNSFP4.2a_grch38.gz
│               ├── dbNSFP4.2a_grch38.gz.tbidbNSFP4.2a_grch38.gz.tbi
│               └── dbNSFP_4.2a.sha256dbNSFP_4.2a.sha256
└── external

9 directories, 17 files
[test_cad@dc5vmscad-[redacted] ~]$ tree /working/
/working/
└── test

0 directories, 1 file
[test_cad@dc5vmscad-[redacted] ~]$
```

Pour résumer

Participation active de la bioinfo, bel atout!

Structuration de la mise en oeuvre des projets en 2025

- Infrastructure
- Premières données des dossiers LBM
- Mise en place de la qualité
- Mise à disposition de bulles sécurisées en cours
- Catalogue des outils / des données
- "Welcome pack" à définir (et incrémenter)
- Calendrier des livraisons

Perspectives: les projets "au fil de l'eau" avec catalogue requêtable

Merci !

Cellule accompagnement des projets

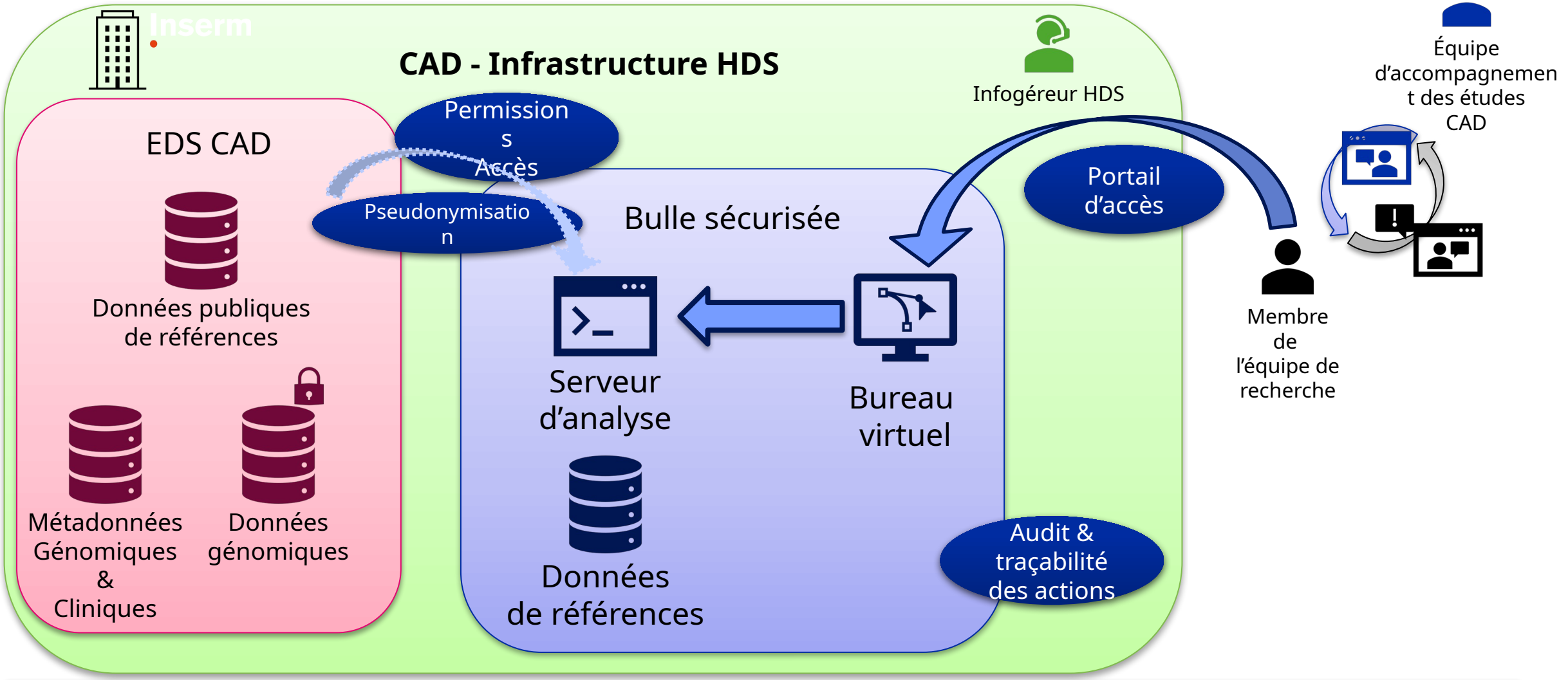
Adrien Josso Rigonato
Flavian Rique
Syrine Bouallegue
Cécile Meslier
Virginie Bernard

cad Collecteur Analyseur de Données

La santé du futur s'inscrit dans notre ADN

Contact : Virginie Bernard

Environnement pour la recherche



Moins de projets en 2024 ?
Anticipation de l'intérêt CADMOS

Évaluation CSE	2022	2023	2024
Défavorable	-	1	-
Favorable avec recommandations			
Favorable	3	8	6
Réservé	5	13	8

- Projets avec focus sur gène d'intérêt
- Variations d'intérêt avec clinique spécifique? >> CADMOS
- Ex. RNU4-2 & biais