

Mise en place de MOABI AP-HP & SeqOIA

Quelles leçons?

Alban LERMINE – Séminaire Bioinfo-diag – Mercredi 15 Mai 2024



Plateforme de service en bio-informatique dédiée aux activités de séquençage ADN haut-debit de l'Assistance Publique - Hôpitaux de Paris

18

Hôpitaux

49

Séquenceurs

53

Laboratoires

211

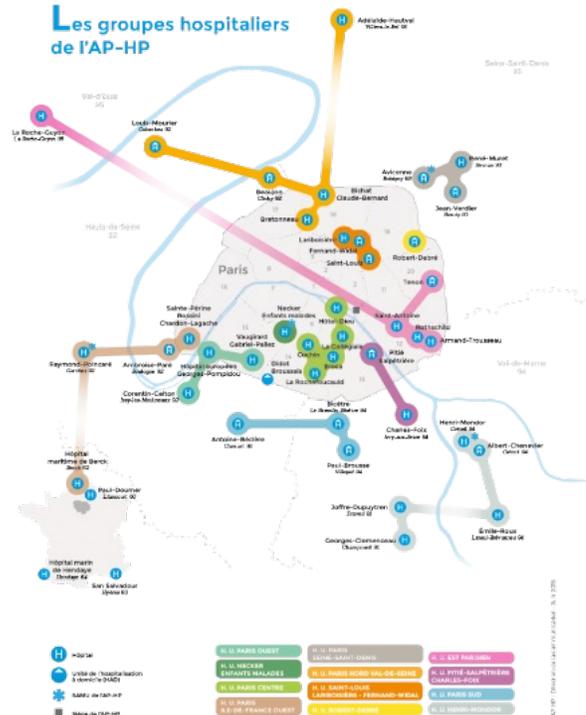
Panels de gènes

580

Utilisateurs

171K

Patients



600TB

Data

900

CPUs

15

Ingénieurs

Infrastructure & Équipe



<https://moabi-aphp.github.io/>



Laboratoire de Biologie Médicale du plan France Médecine Génomique 2025 – Intégration du génome complet dans le cadre du soin

80

Hôpitaux/CLCC

6

Séquenceurs

1

Laboratoire

2

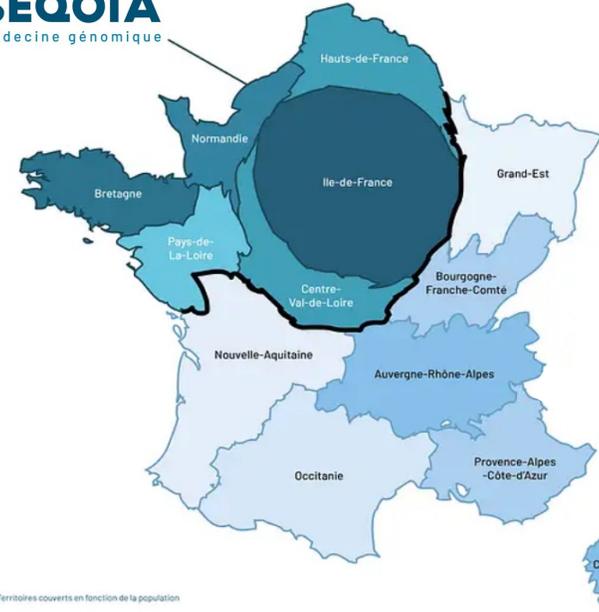
Pipelines

1759

Utilisateurs

41K

Patients



5,7PB

Data

2500

CPUs

18

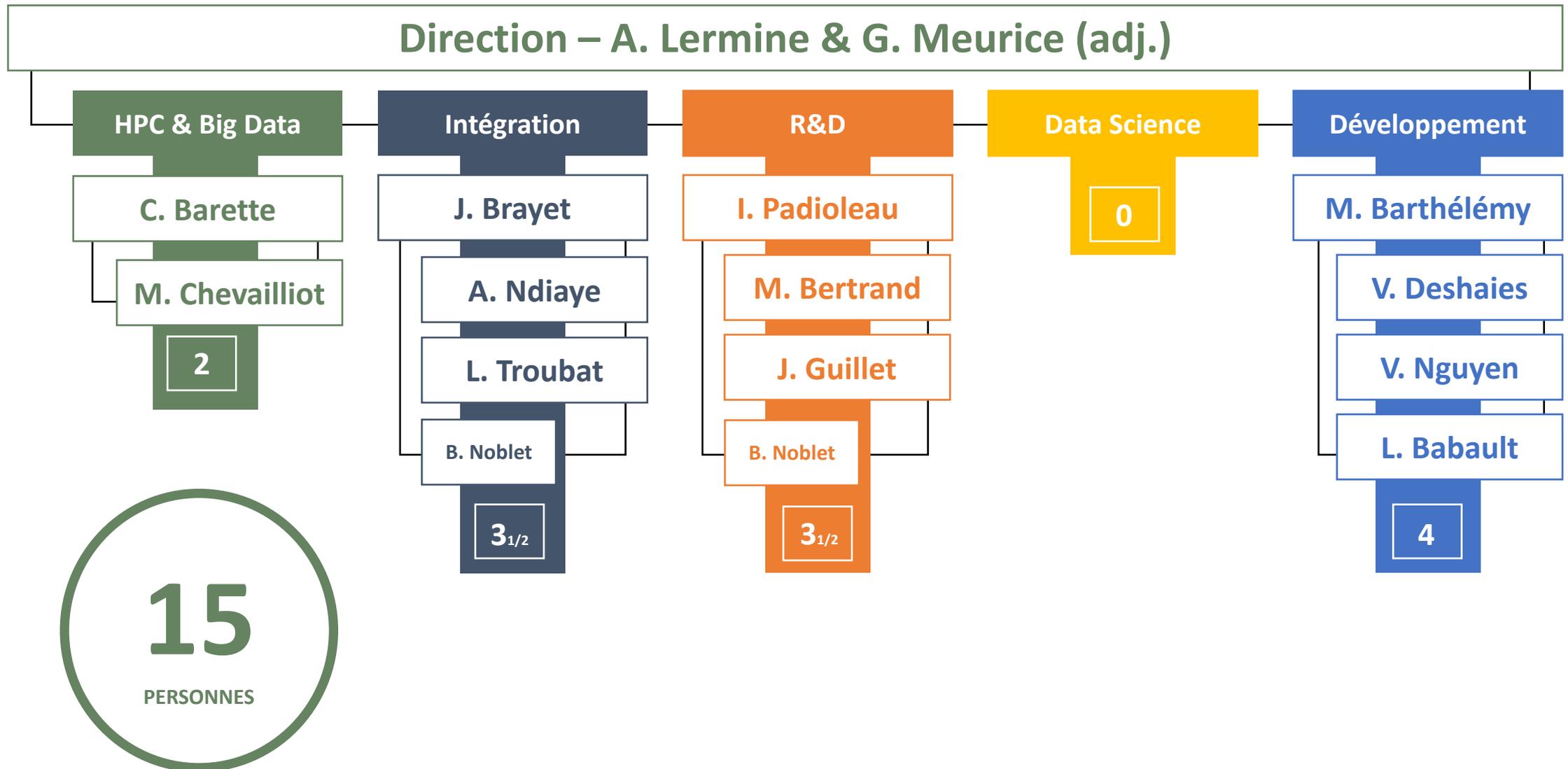
Ingénieurs

Infrastructure & Équipe

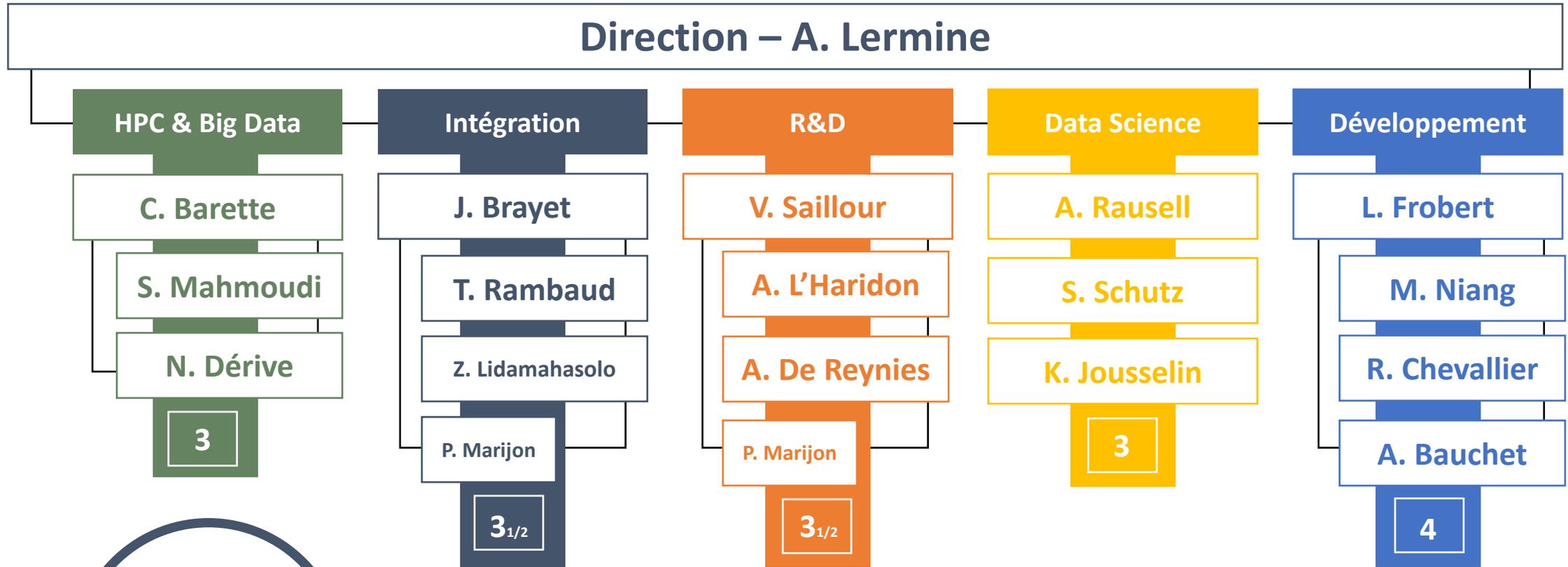


<https://laboratoire-seqoia.fr/>

MOABI AP-HP - Ressources humaines & organisation



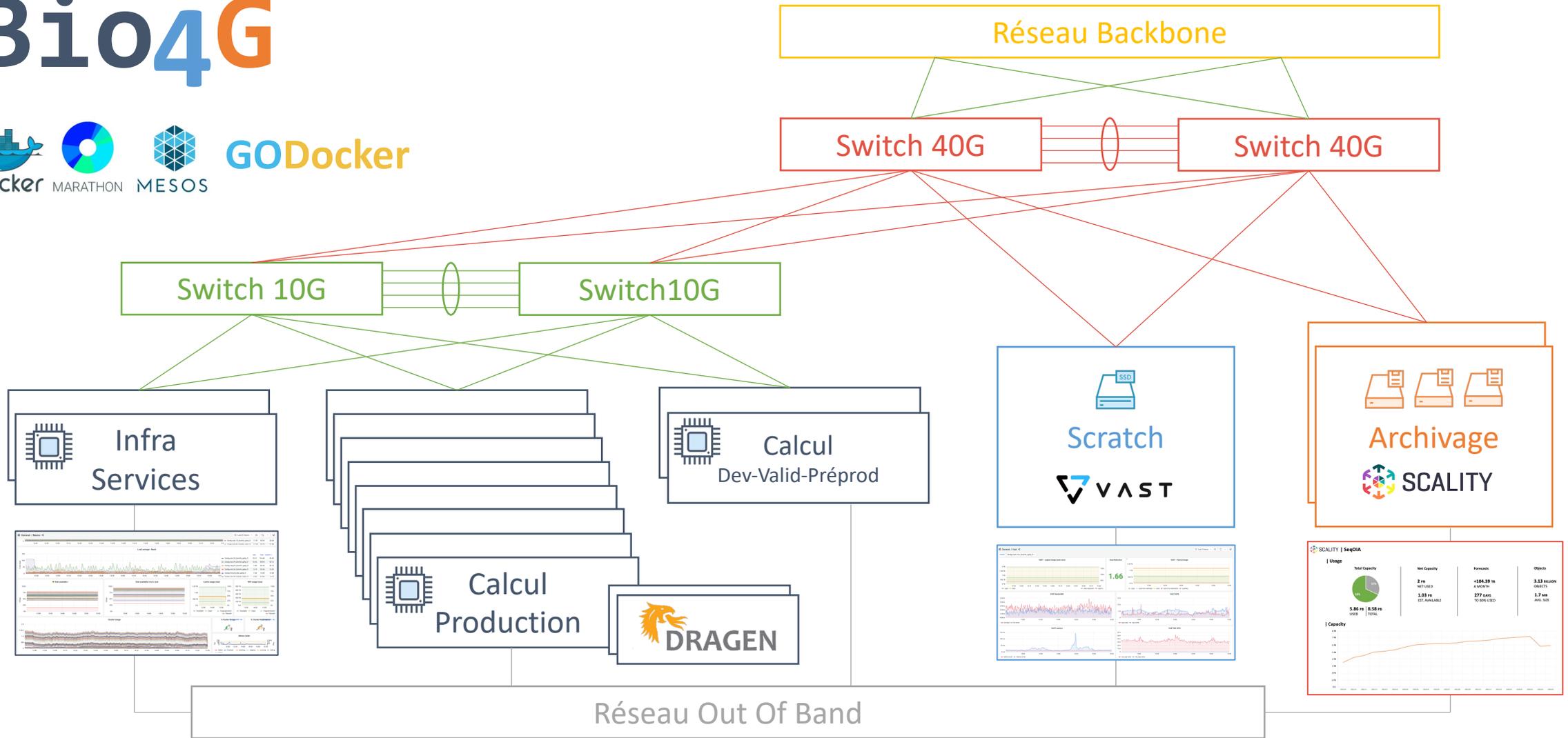
SeqOIA - Ressources humaines & organisation



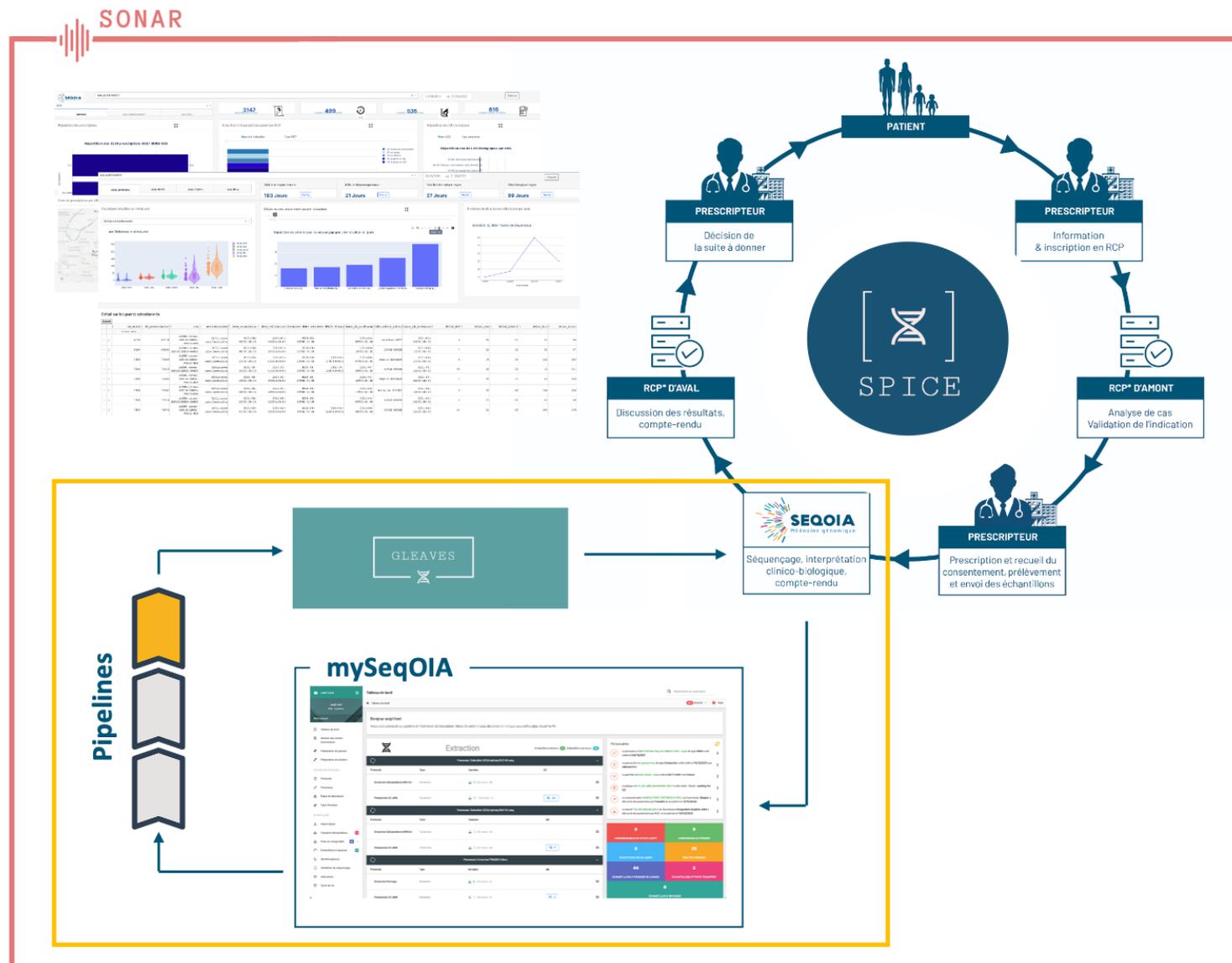
18
PERSONNES

Point clef 1 - Infrastructure Bio4G

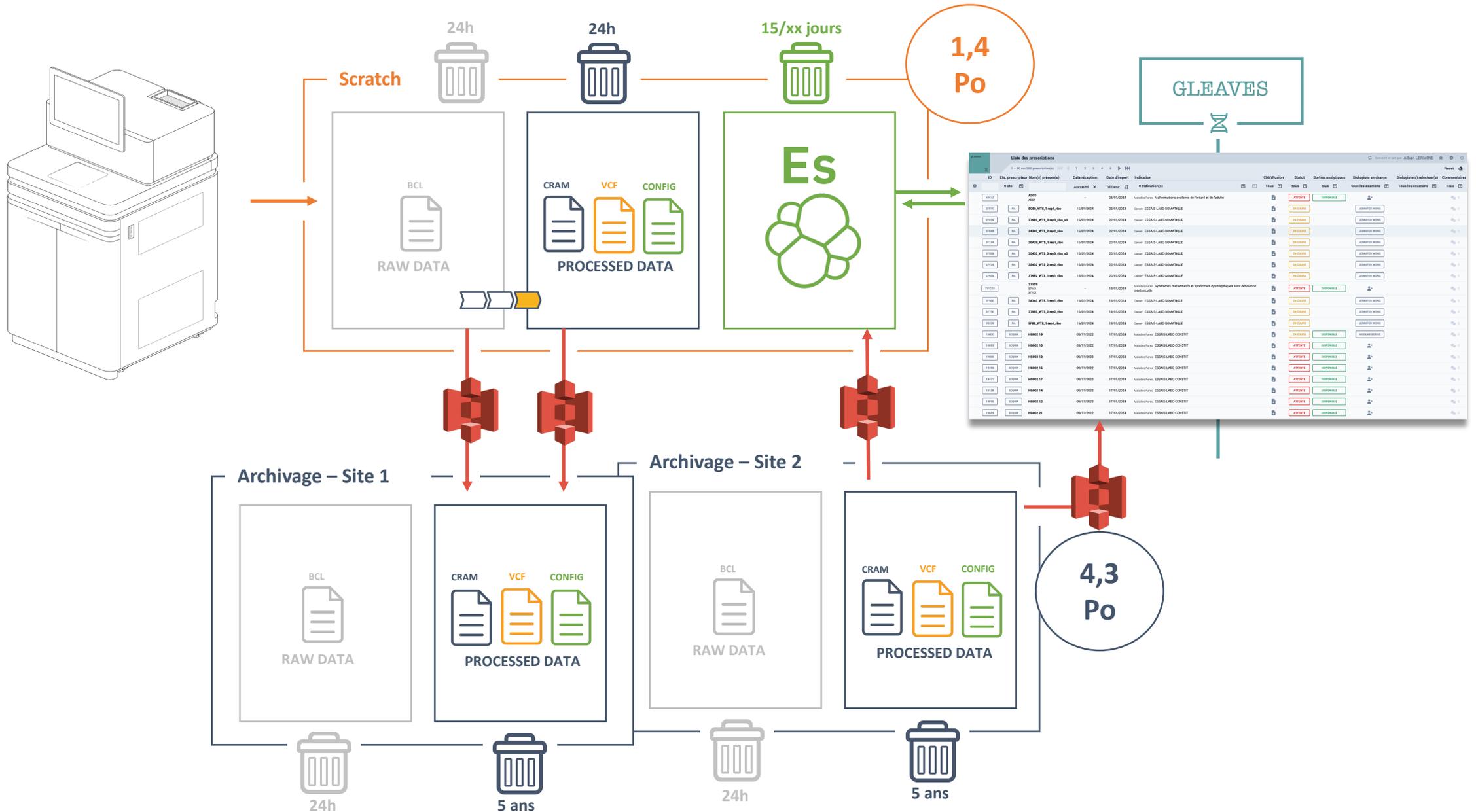
Bio4G



Point clef 2 - Cycle de vie de l'information



Point clef 4 - Cycle de vie des données



Liste des prescriptions

ID	Etq. prescripteur	Nom(s) patient(s)	Date réception	Date d'export	Indication	CHK/Validation	Statut	Séries analytiques	Biologie(s) en charge	Biologie(s) selecteur(s)	Commentaires
100001	ABCC	ABCC	20/05/2024	20/05/2024	Médecin Presc: Malformations oculaires de l'enfant et de l'adulte	OK	ATTENTE	RESPONSE 2			
100002	30448_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100003	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100004	34448_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100005	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100006	34448_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100007	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100008	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100009	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100010	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100011	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100012	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100013	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100014	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100015	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100016	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100017	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100018	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100019	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100020	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				
100021	37191_WTL3_1_rpt_Lib		15/05/2024	20/05/2024	Centre: ESSAIS-LABO-SOMATOLOGIE	OK	EN COURS				

Volumétrie en jeu - Scratch

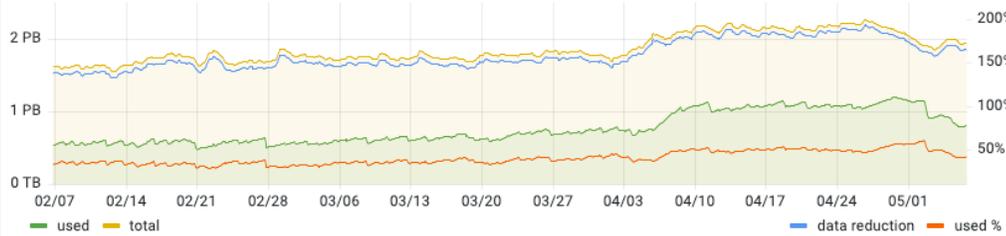
Scratch

General / Vast

Last 90 days

cluster bio4g-vast-vms_bioinfo_aphp_fr

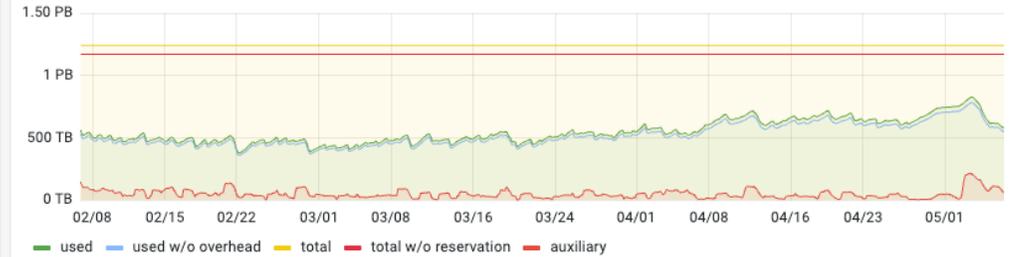
VAST - Logical Usage (seen size)



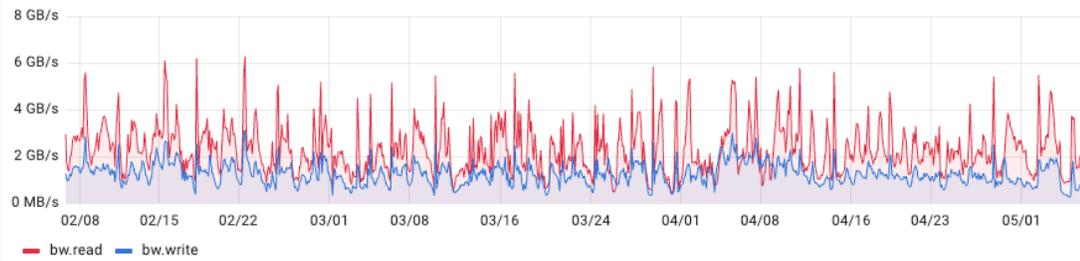
Data Reduction

1.66

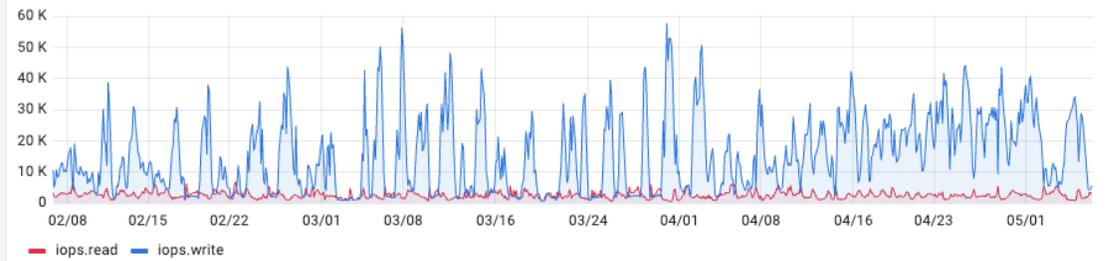
VAST - Physical Usage



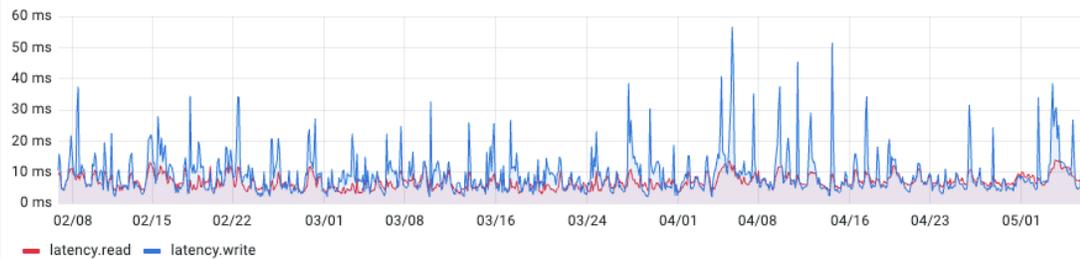
VAST Bandwidth



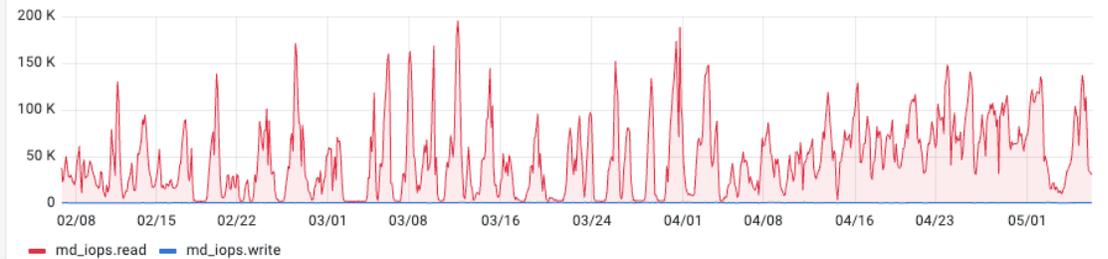
VAST IOPS



VAST Latency



VAST MD IOPS



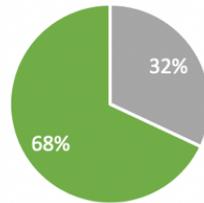
Volumétrie en jeu - Archivage

Archivage



| Usage

Total Capacity



5.86 PB USED | **8.58 PB** TOTAL

Net Capacity

2 PB
NET USED

1.03 PB
EST. AVAILABLE

Forecasts

+104.39 TB
A MONTH

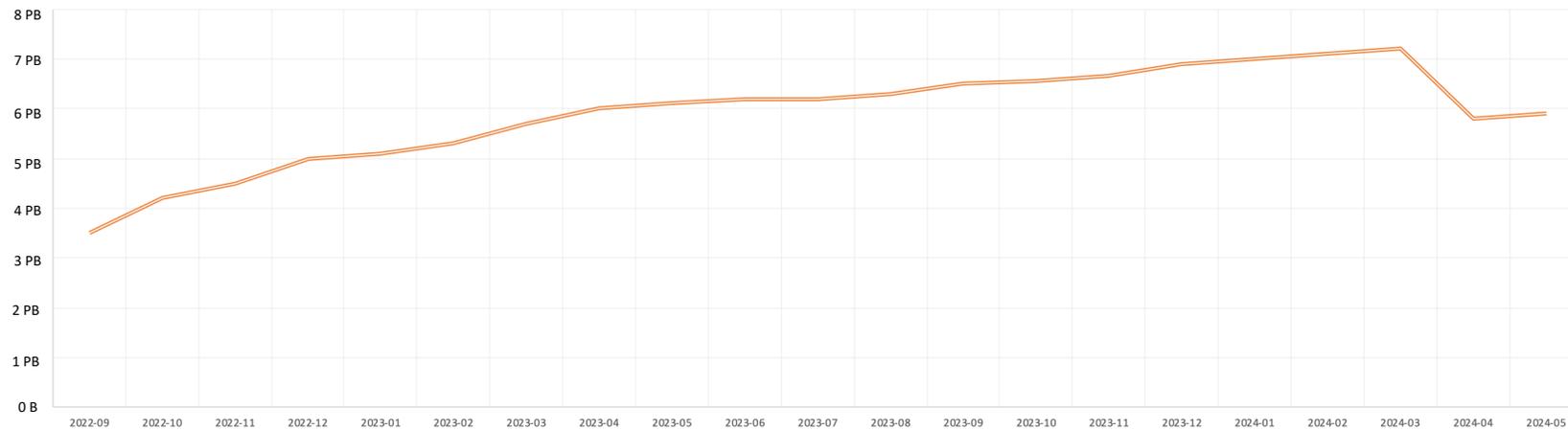
277 DAYS
TO 80% USED

Objects

3.13 BILLION
OBJECTS

1.7 MB
AVG. SIZE

| Capacity



Quelles leçons?

- 4 point clefs -

Infrastructure

Maitrise de l'Information

Maitrise du cycle de vie des données

Automatiser = Survivre

- Autres enseignements -

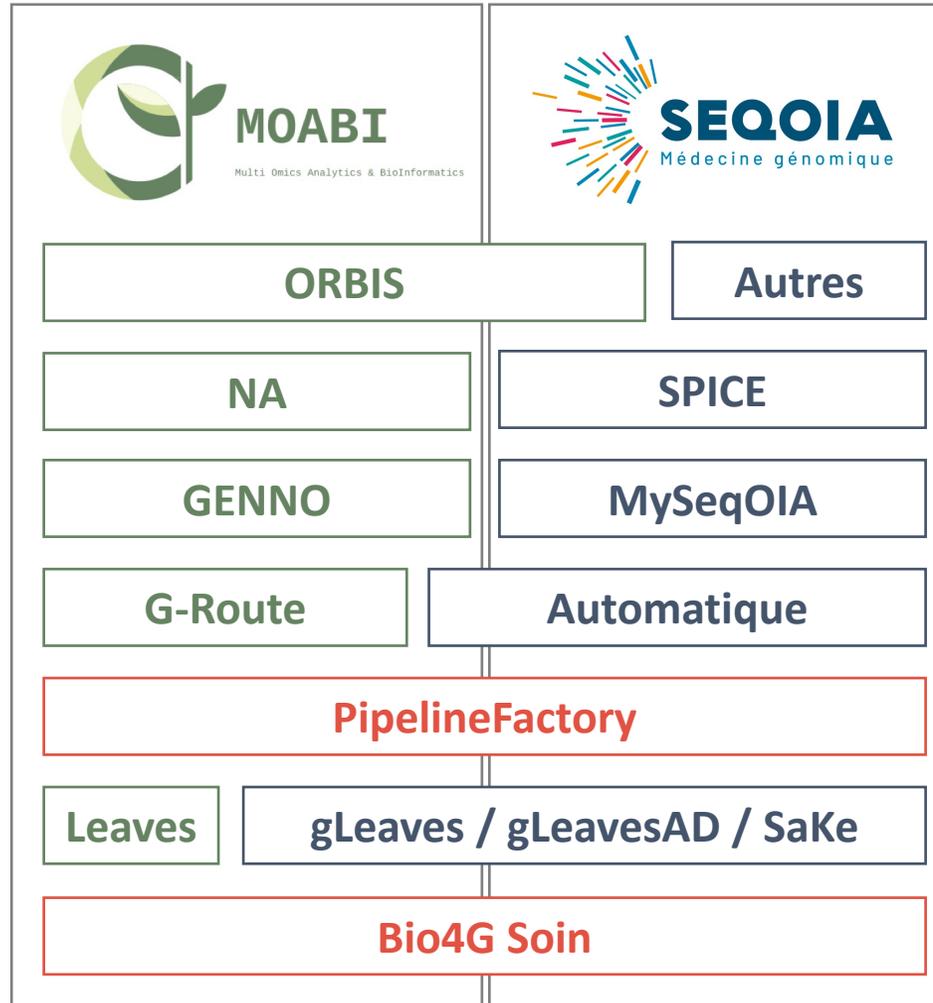
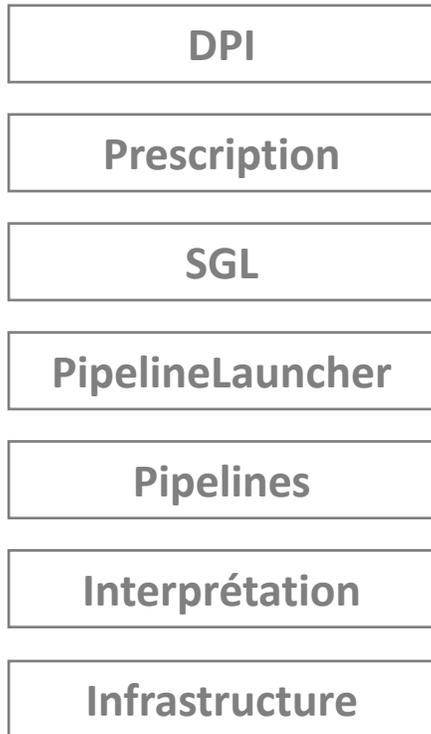
- Diversité infiniment plus complexe à gérer que la volumétrie
- L'internalisation des solutions permet:
 - La maitrise de l'information de bout en bout
 - La maitrise du cycle de vie des données
 - La liberté et la maitrise des évolutions
 - La jouissance des sources logicielles (cf. Distribution logicielle)
 - Un accès illimité à ses propres données (cf. Sequoia dAta laKE)
 - De faire des économies, c'est moins cher que la prestation SaaS!!!
 - 1 génome analysé/stocké 5 ans = 28€
- Mais l'internalisation c'est:
 - Un modèle d'investissement différent (RH +++) souvent redouté en milieu hospitalier
 - Recruter des profils pouvant être rares sur le marché à des salaires souvent inférieurs au marché
 - Rentable que si l'activité est suffisante



→ Automatisation à l'extrême

Systeme d'Information Soins – Aujourd'hui

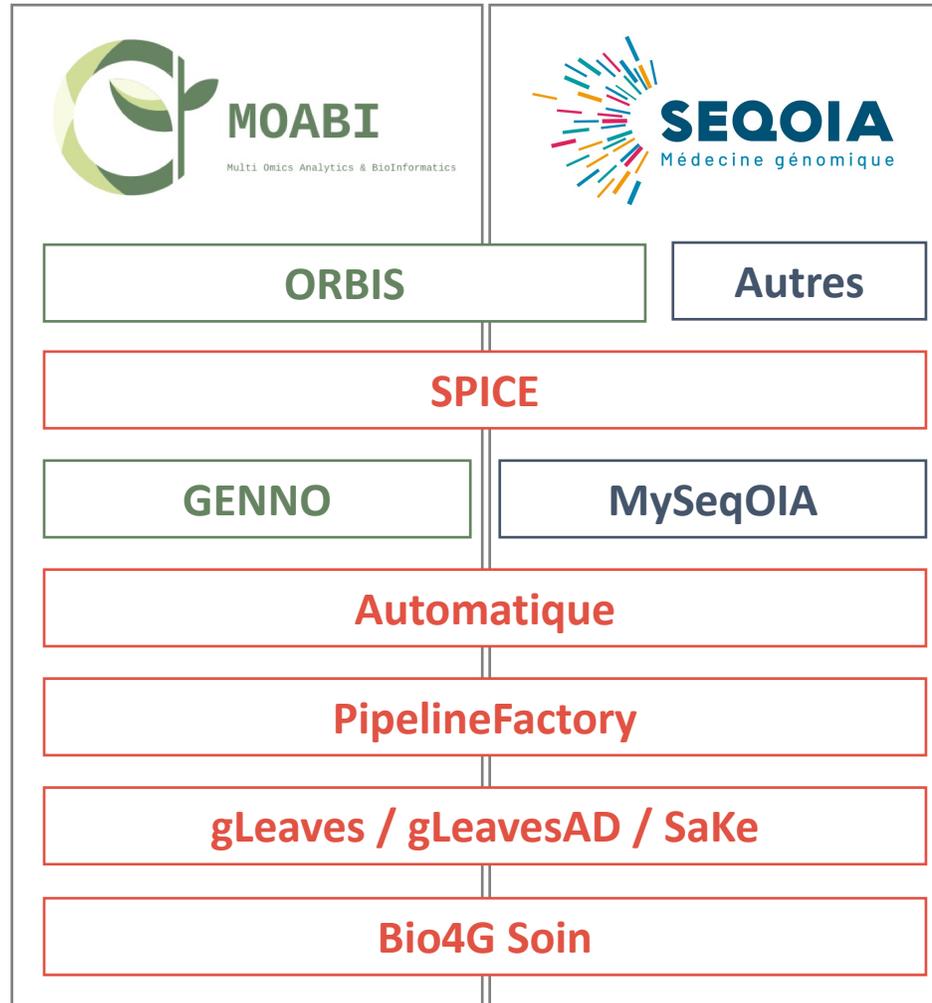
2024



Système d'Information Soins – Cible

Cible

DPI
Prescription
SGL
PipelineLauncher
Pipelines
Interprétation
Infrastructure



Distribution Logicielle

<https://gleaves.laboratoire-sequoia.fr>

<https://github.com/natir/variantplaner>

Solutions distribuées

GLEAVES



+



gLeavesAD



**Variant
-Planer**

 Freeware

<https://gleaves.laboratoire-sequoia.fr>

Documentation d'installation/exploitation

Images Docker + Annotations dynamiques

StackOverflow

 Open Source

<https://github.com/natir/variantplaner>

Documentation

Sources

Github

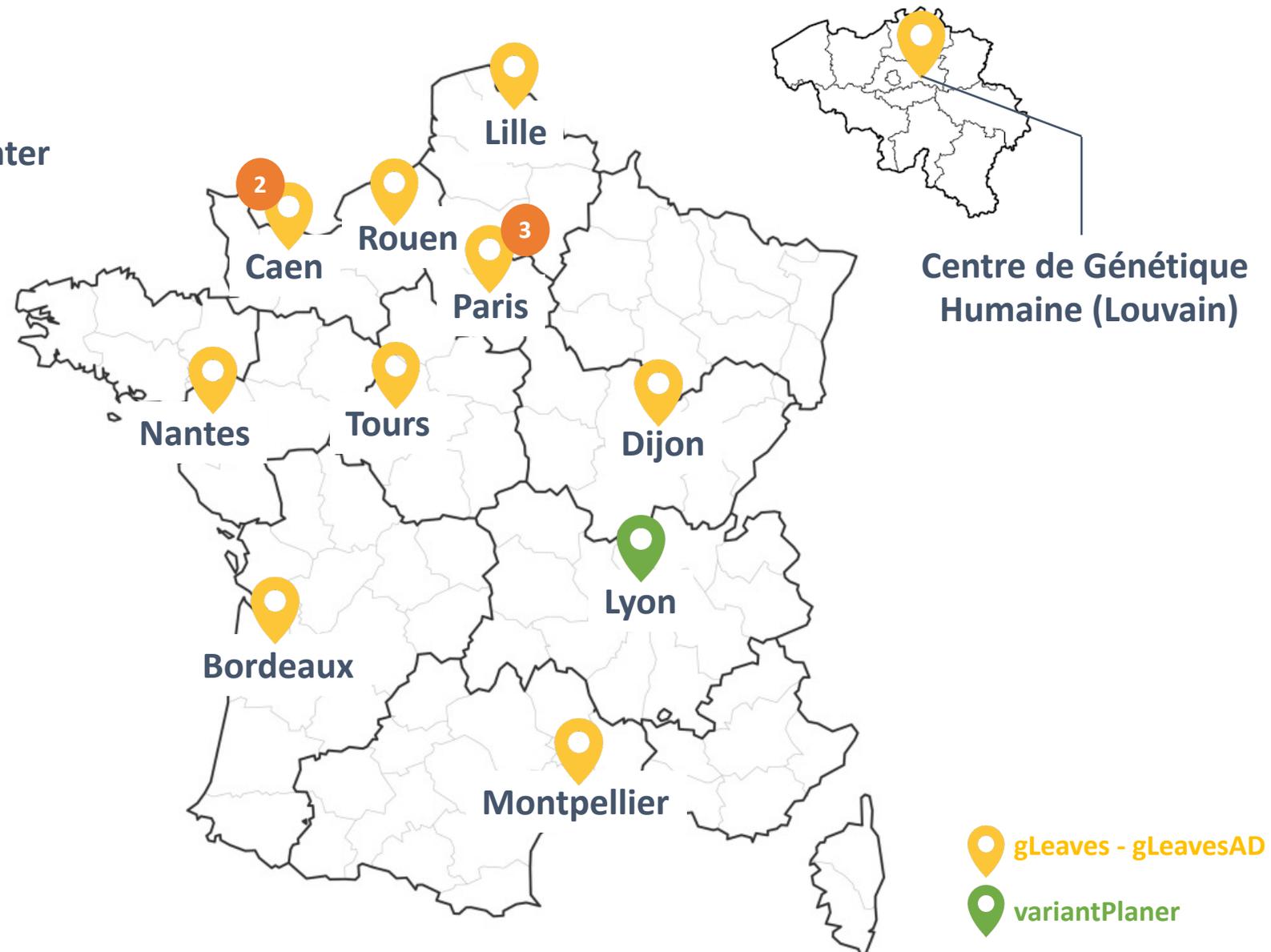
Distribution – Discussions en cours

Memory Sloan-Kettering Cancer Center



NEW YORK

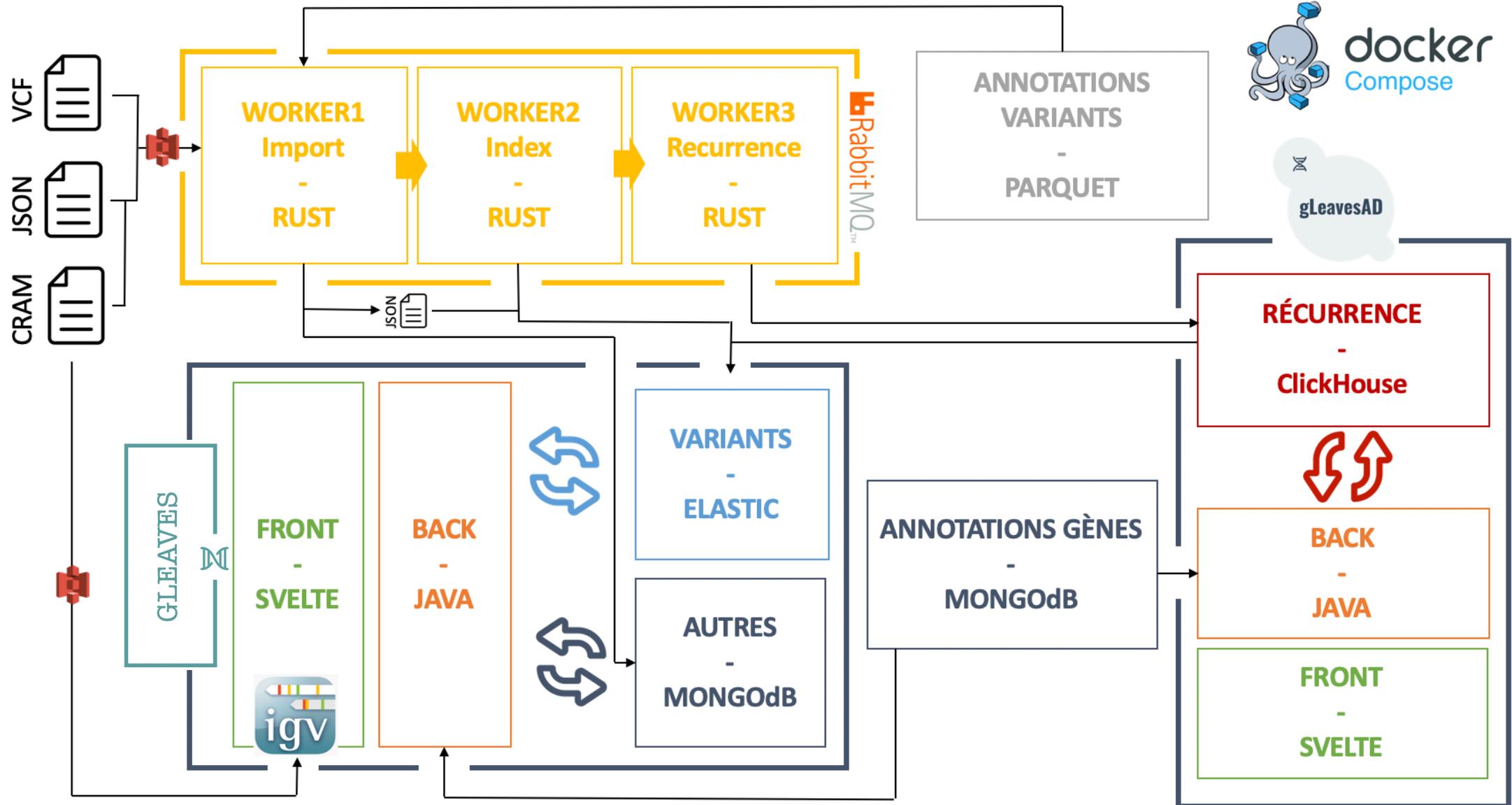
New York Genome Center



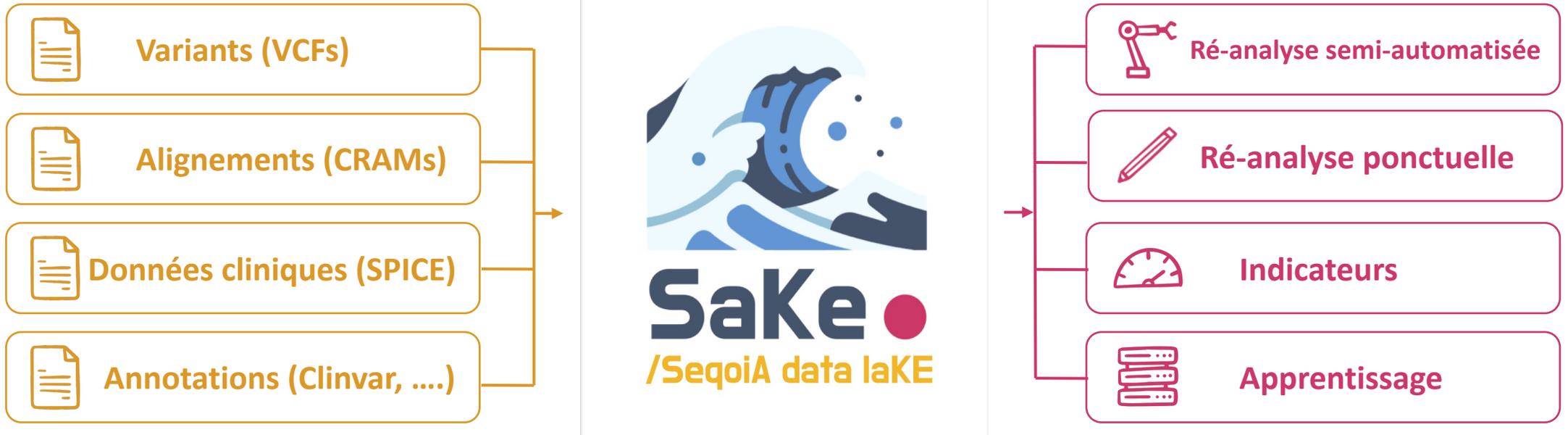
Centre de Génétique Humaine (Louvain)

-  gLeaves - gLeavesAD
-  variantPlanner

gLeaves & gLeavesAD - Architecture logicielle



SaKe – Sequoia dAta LaKE



Pierre MARIJON – VariantPlaner

Structuration des données

Nom/Prénom/Date de naissance
 Sexe
 Atteint/Non Atteint
 Pré-indication
 HPO
 Prescripteur/Etablissement
 ...

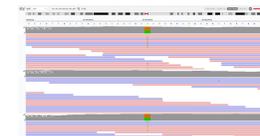
41K individus

700M variants

	7CD0	7CD1	7CD2	AB0	AB1	AB2	AB3	3280	3281	3282
1:12300213 A>T	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/1	0/1
1:381290839 G>CT	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2:123819238 AT>A	0/0	0/0	0/0	0/1	0/0	0/1	0/0	0/0	0/0	0/0
2:391283122 C>G	0/1	0/0	0/0	0/0	0/0	1/1	0/0	0/0	1/1	0/0
X:632169833 A>T	0/0	0/1	0/1	0/1	0/1	0/1	0/1	0/0	0/0	0/0
X:322169821 A>AT	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Chr/Position/ref/alt
 Gène
 Classe SeqOIA
 Fréquence SeqOIA
 Clinvar
 gnomAD

Génotype
 VAF
 Couverture
 Transmission
 Classe SeqOIA

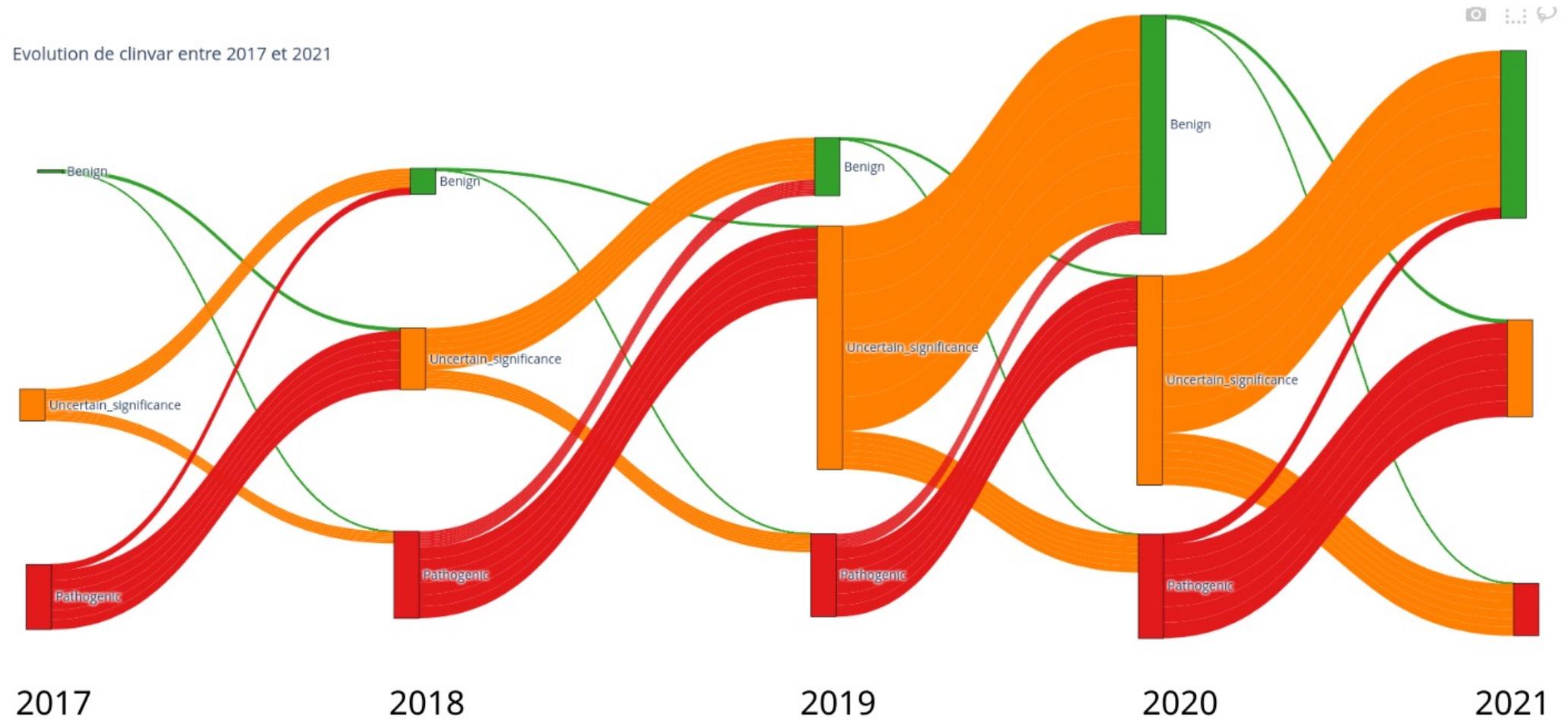


....

La ré-analyse périodique des examens non conclusifs produit un gain diagnostique annuel de l'ordre de 5 à 10 %

<https://pubmed.ncbi.nlm.nih.gov/26562225/>
<https://pubmed.ncbi.nlm.nih.gov/27711071/>
<https://pubmed.ncbi.nlm.nih.gov/27441994/>
<https://pubmed.ncbi.nlm.nih.gov/31231135/>
<https://pubmed.ncbi.nlm.nih.gov/33568551/>

Evolutions de Clinvar entre 2017 et 2021



L'interprétation perpétuelle



Ré-annotation (1x/mois)



Sélection automatisée (1x/mois)



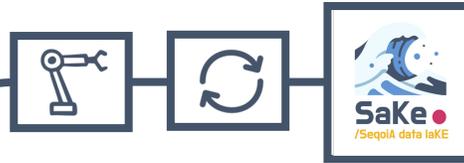
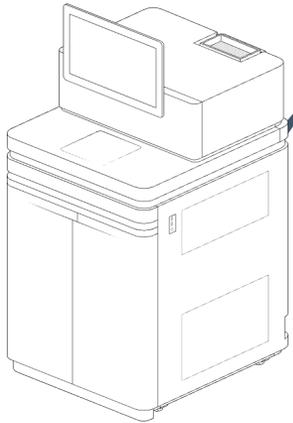
Biologiste



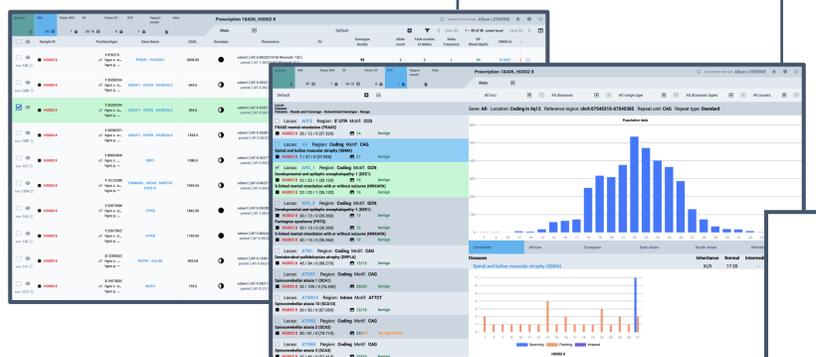
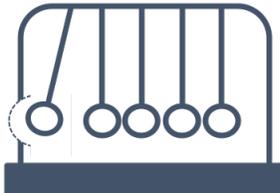
CR non-conclusif



CR concluant



```
{data_lake}
├─ germline/           // données de pré-indication constitutionnelles
├─ somatic/           // données de pré-indication somatiques
├─ samples/           // informations sur les dossiers
├─ annotations/       // base d'annotations des variants
└─ aggregations/      // données agrégées, pré-calculé
```



```
>
while $CR != concluant
do
  sake.requests
  gleaves.alert

  if $gleaves.alert = true
  then
    biologist.read($gleaves.alert)
  fi

  if $biologist.read($gleaves.alert) = patho || likely_patho
  then
    CR = concluant
  else
    CR = non-conclusif
  fi
done
```