



L'INSTITUT DE RECHERCHE TECHNOLOGIQUE BIOASTER AU CENTRE DE CALCUL DE L'IN2P3

Aspects opérationnels

Auteur NICOLAS SAPAY (NICOLAS.SAPAY@BIOASTER.COM)

Lieu CENTRE DE CALCUL DE L'IN2P3, LYON

Date 2 AVRIL 2013

LES PARTENAIRES FONDATEURS DE BIOASTER



FONDATION
DE COOPÉRATION SCIENTIFIQUE

INSTITUT DE RECHERCHE TECHNOLOGIQUE

Maladies infectieuses : bactéries, virus, champignons, parasites

Thérapies et vaccins

Diagnostic

Microbiote comme
indicateur et produit de santé



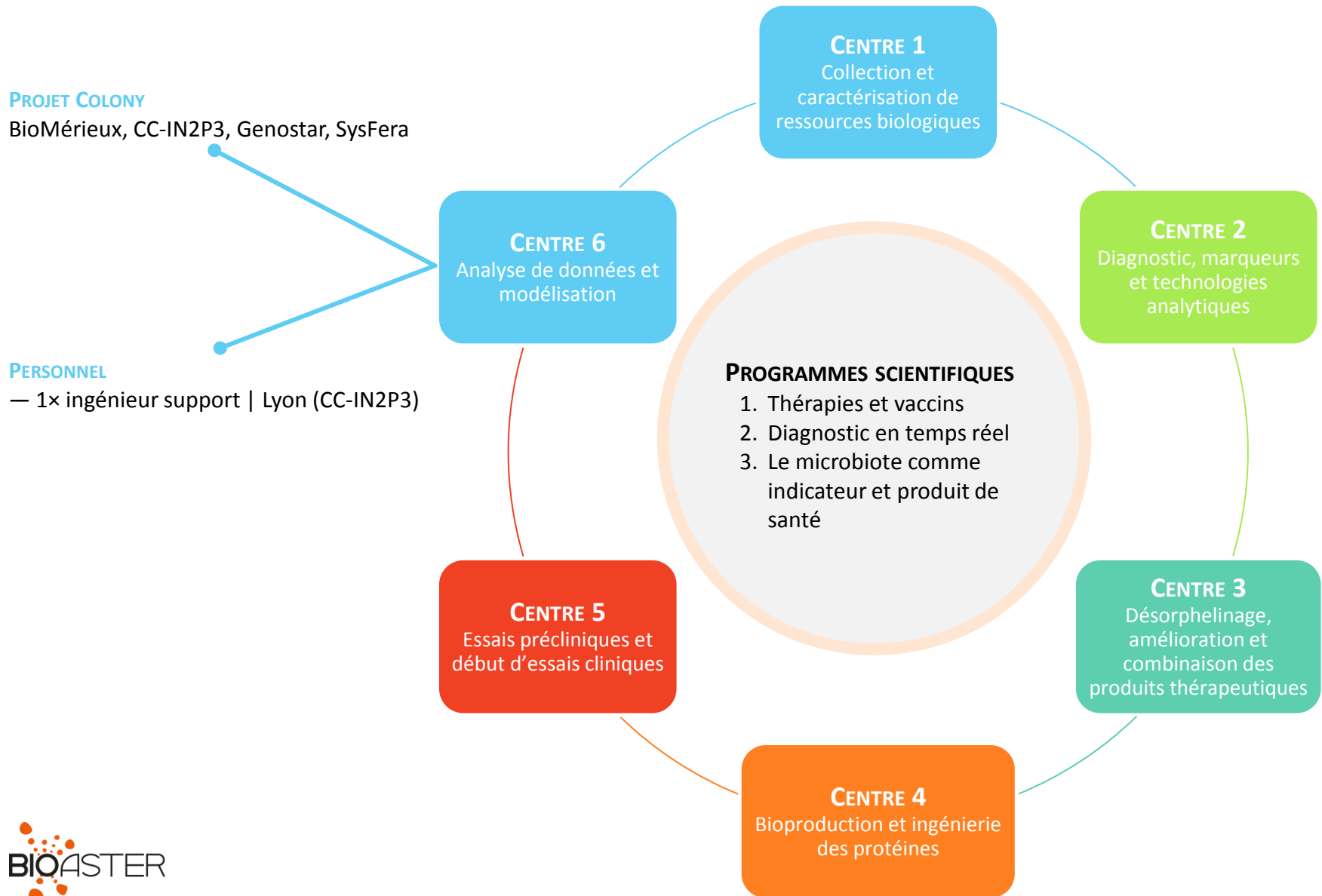
PROGRAMMES DE RECHERCHE ET CENTRES TECHNOLOGIQUES

PROJET COLONY

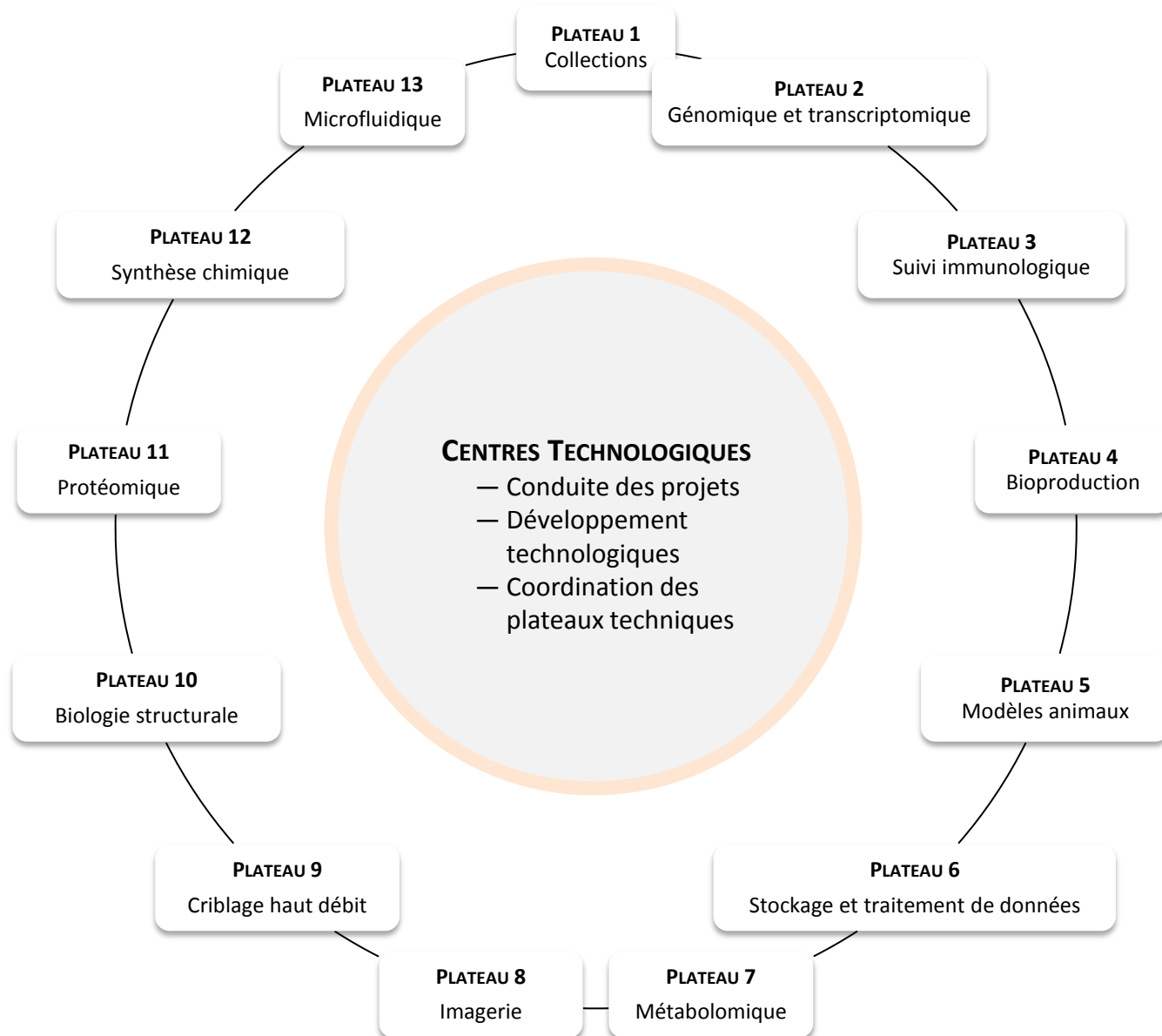
BioMérieux, CC-IN2P3, Genostar, SysFera

PERSONNEL

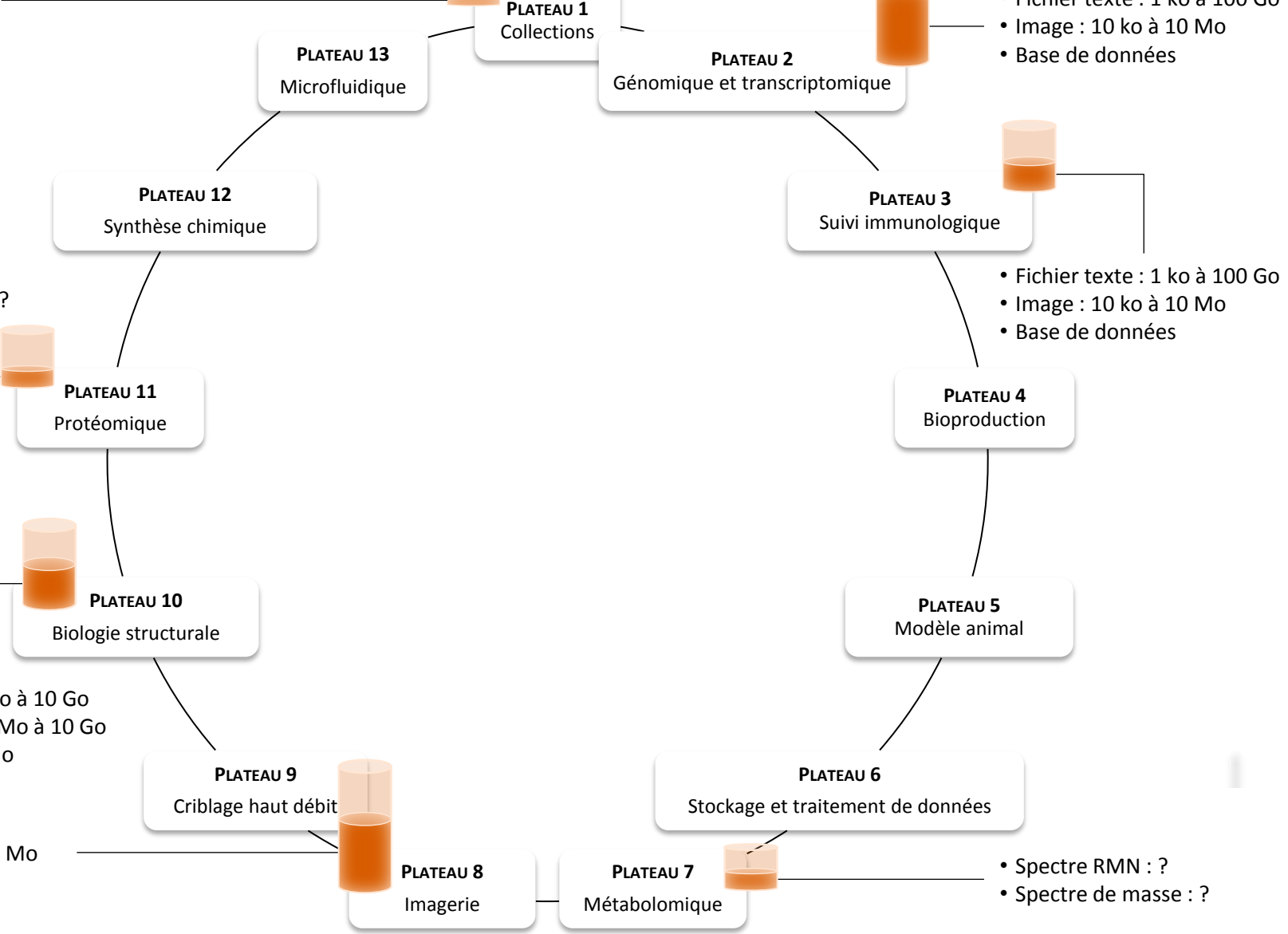
— 1x ingénieur support | Lyon (CC-IN2P3)



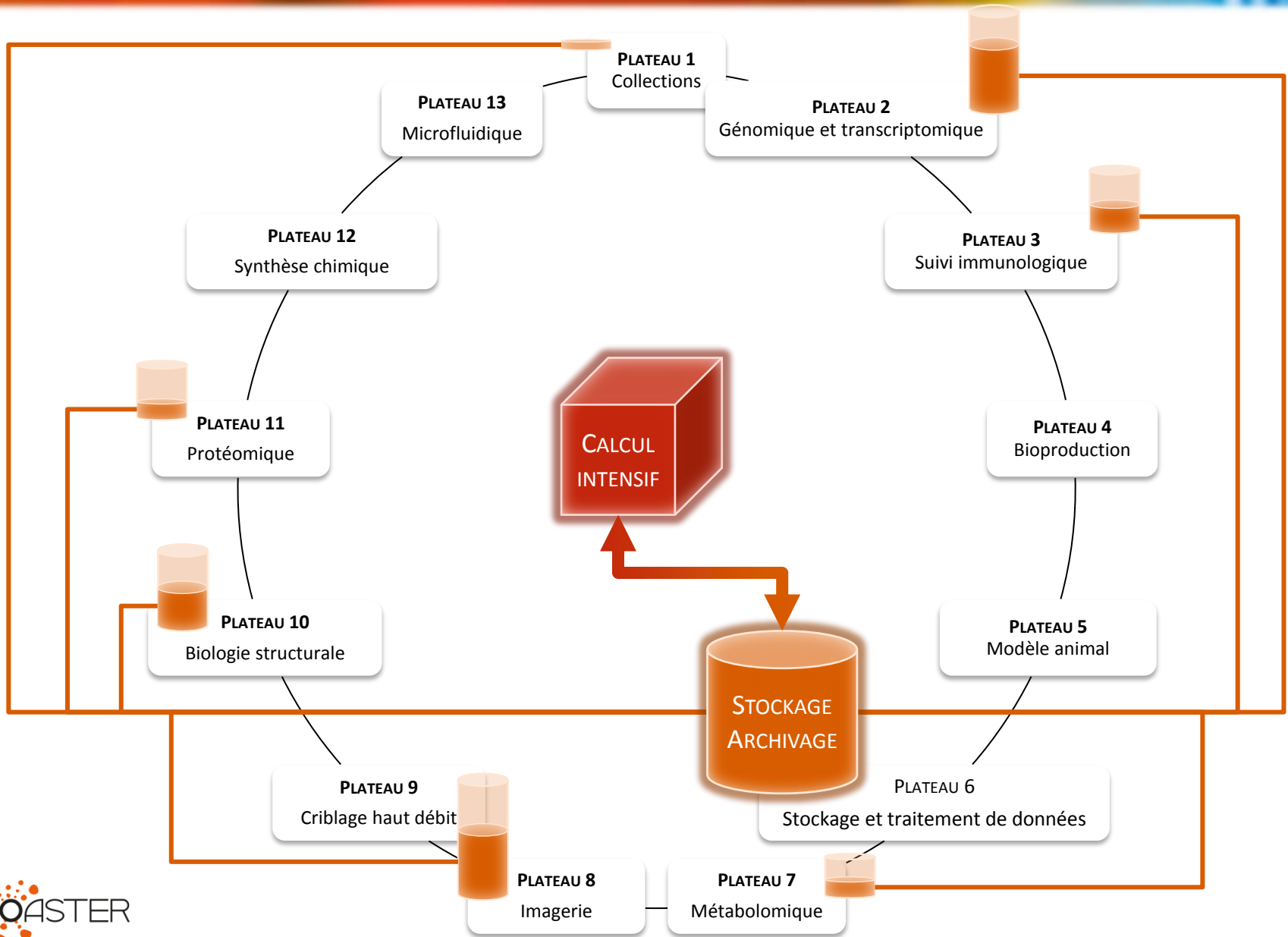
CENTRE TECHNOLOGIQUES ET PLATEAUX TECHNIQUES



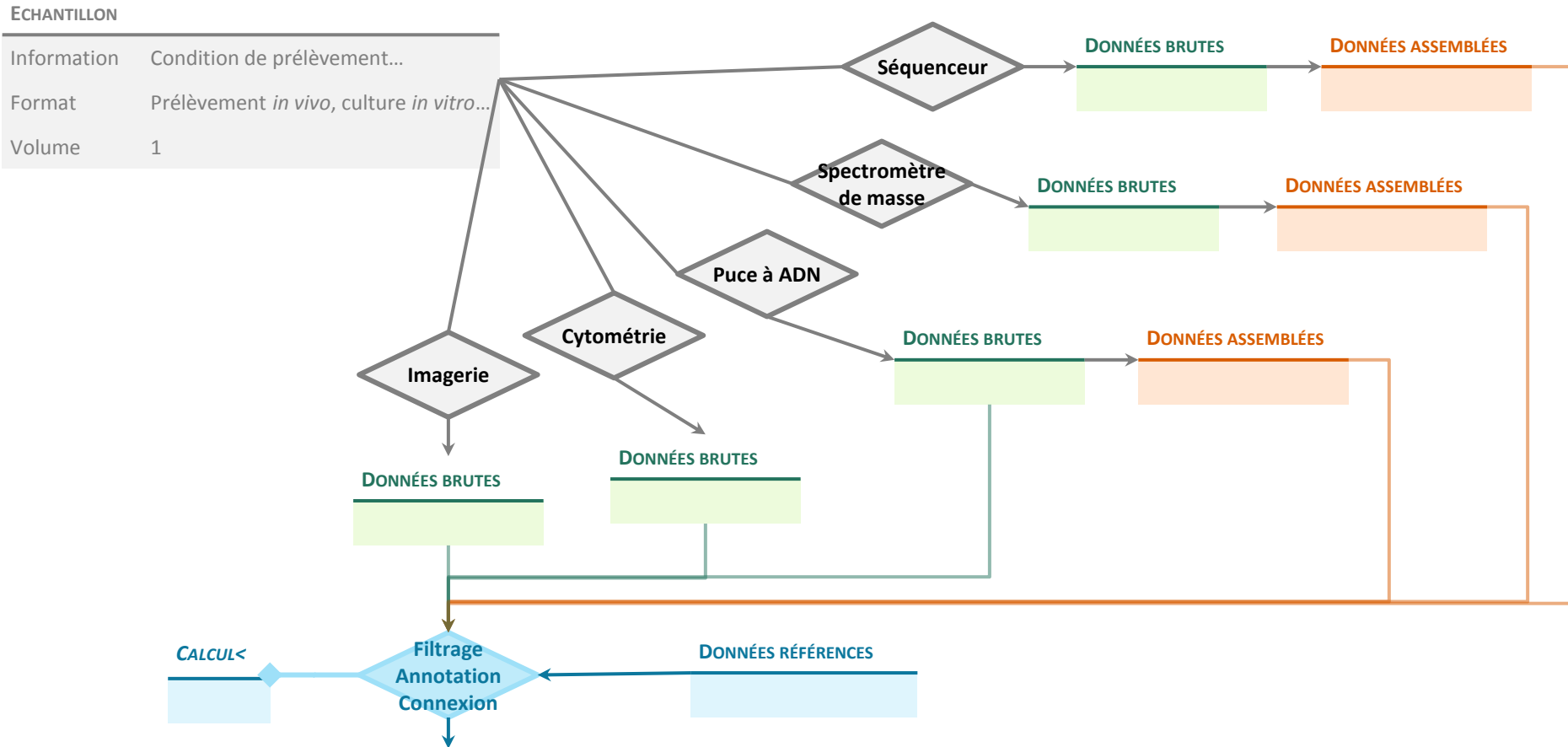
LES BESOINS EN STOCKAGES DES PLATEAUX TECHNIQUES



LES BESOINS EN STOCKAGES DES PLATEAUX TECHNIQUES



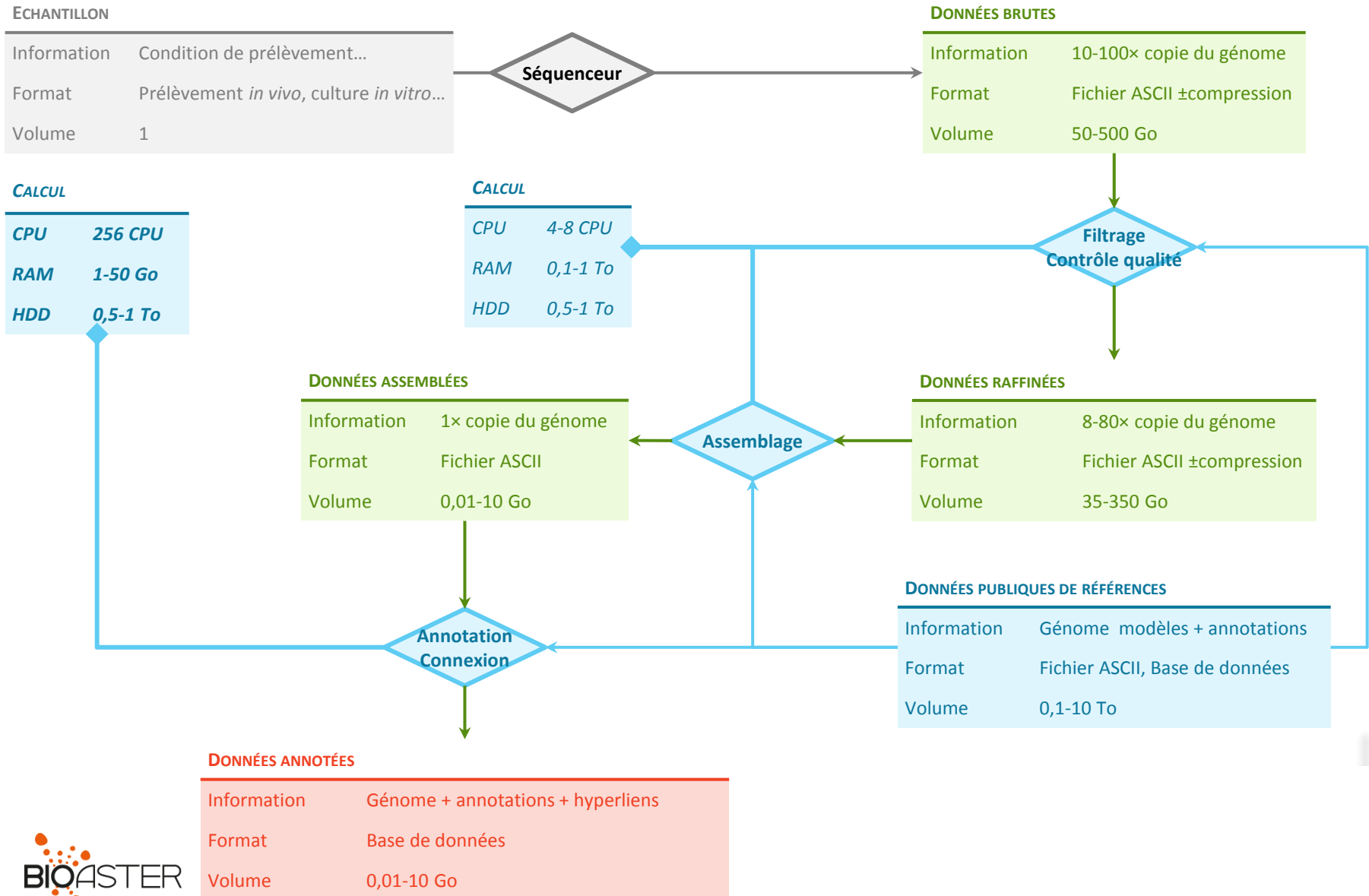
OBJECTIFS SCIENTIFIQUES ET TECHNOLOGIQUES



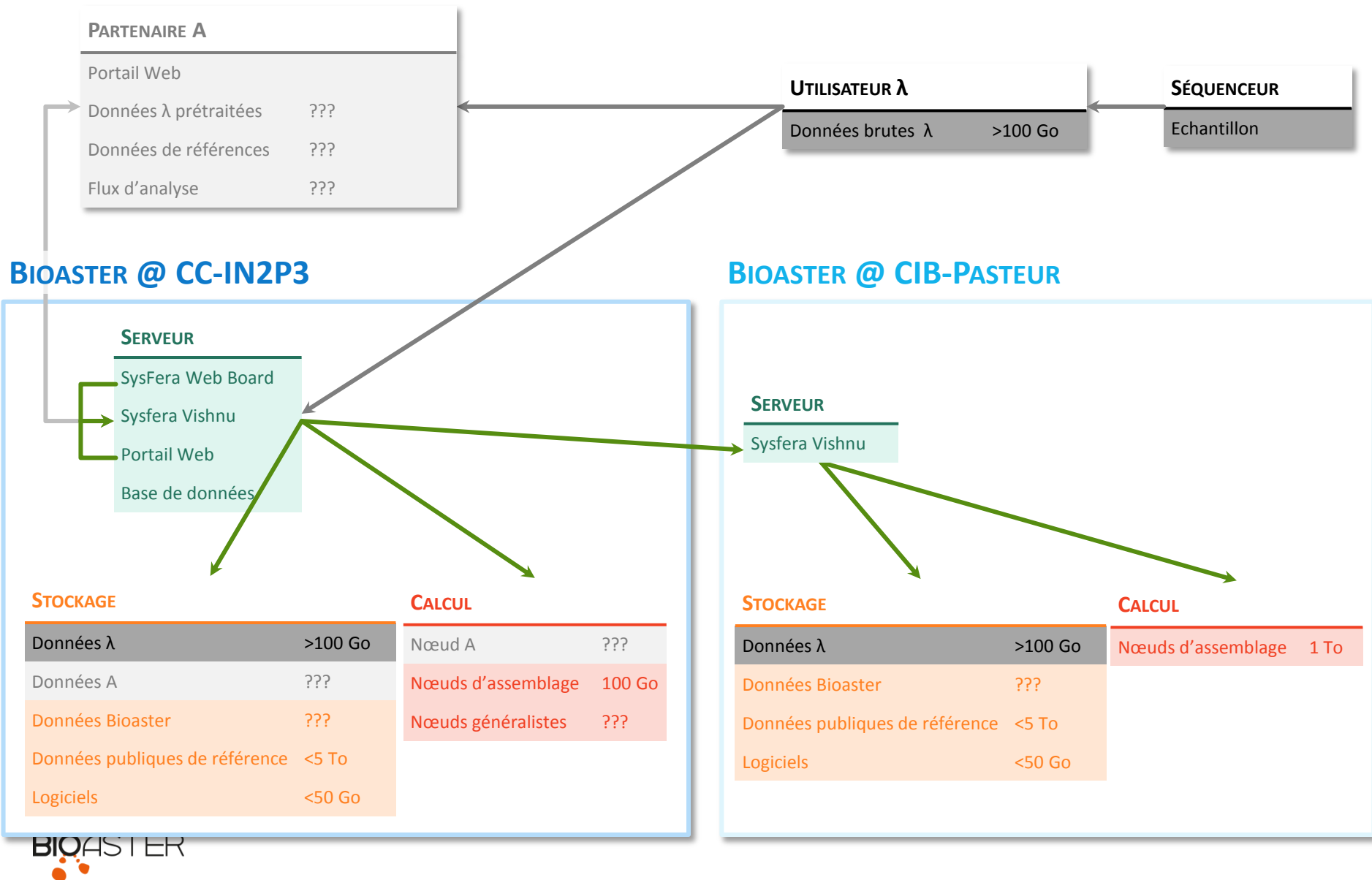
RÉSEAU D'INFORMATIONS

Information	Métadonnées + Connexion + Modèles
Format	Base de données
Volume	???

CAS DU SÉQUENÇAGE DE GÉNOMES



CAS DU SÉQUENÇAGE DE GÉNOMES



CAS DU SÉQUENÇAGE DE GÉNOMES

UTILISATEUR λ	LYON
Données brutes	>100 Go

UTILISATEUR μ	PARIS
Données brutes	>100 Go

Comment injecter les données ?

BIOASTER @ CC-IN2P3

BIOASTER @ CIB-PASTEUR

SYSFERA

Web Board
Serveur Vishnu

SYSFERA

Serveur Vishnu

Données traitées sur le site de stockage
⇒ Minimisation du transfert

STOCKAGE/ARCHIVAGE

Données brutes λ	>100 Go
Données assemblées λ	<5 Go
Données assemblées μ	<5 Go
Données publiques de référence	<5 To
Logiciels	<50 Go

CALCUL

Nœuds génome	100 Go
Nœuds généralistes	???

STOCKAGE

Données brutes μ	>100 Go
Données assemblées μ	<5 Go
Données publiques de référence	<5 To
Logiciels	<50 Go

CALCUL

Nœuds d'assemblage	1 To
--------------------	------



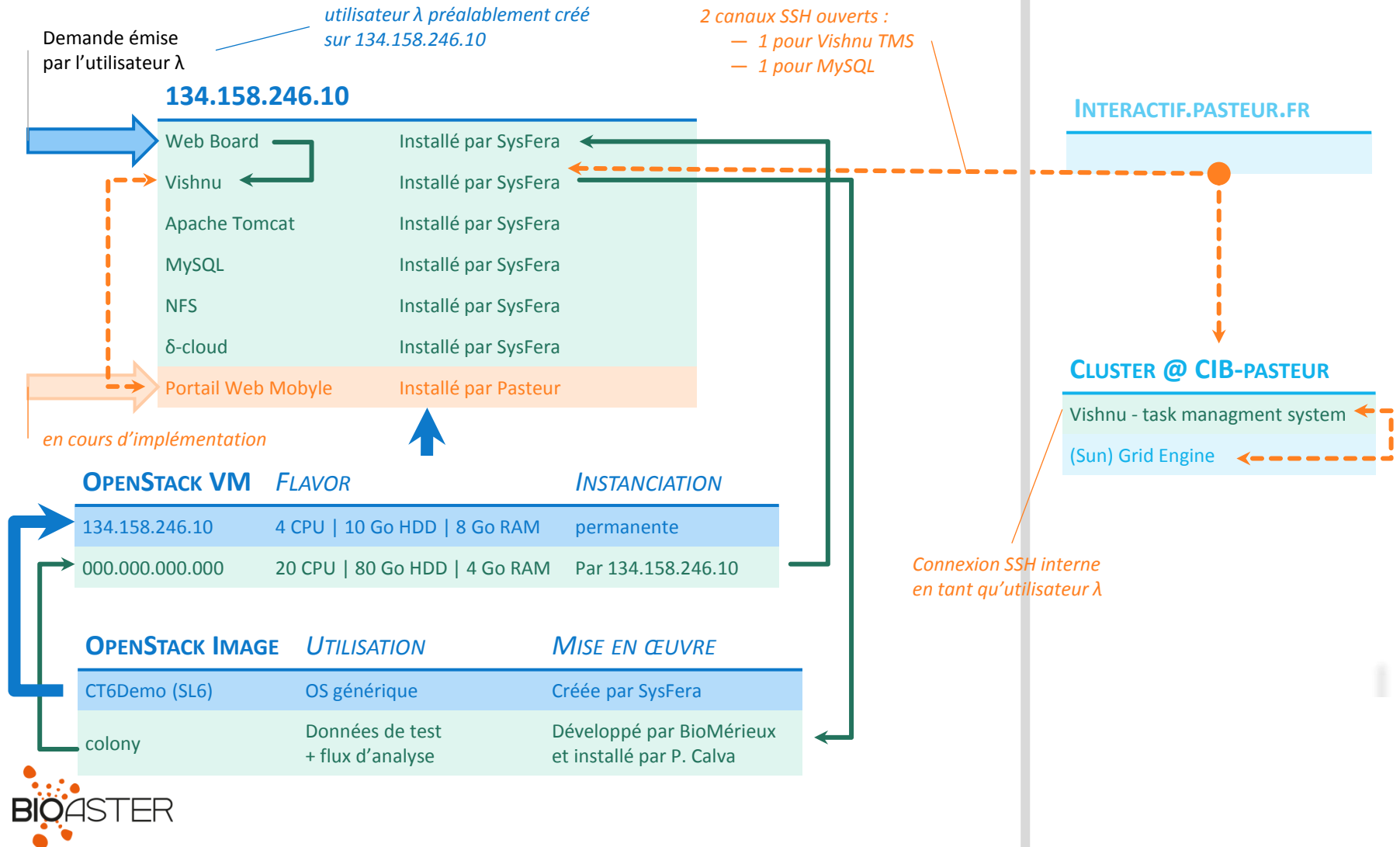
Comment synchroniser les références et les logiciels ?

(Stockage + Calcul) $\times 2$

ORGANISATION ACTUELLE DU PROTOTYPE DE PLATEAU TECHNIQUE 6

BIOASTER @ CC-IN2P3

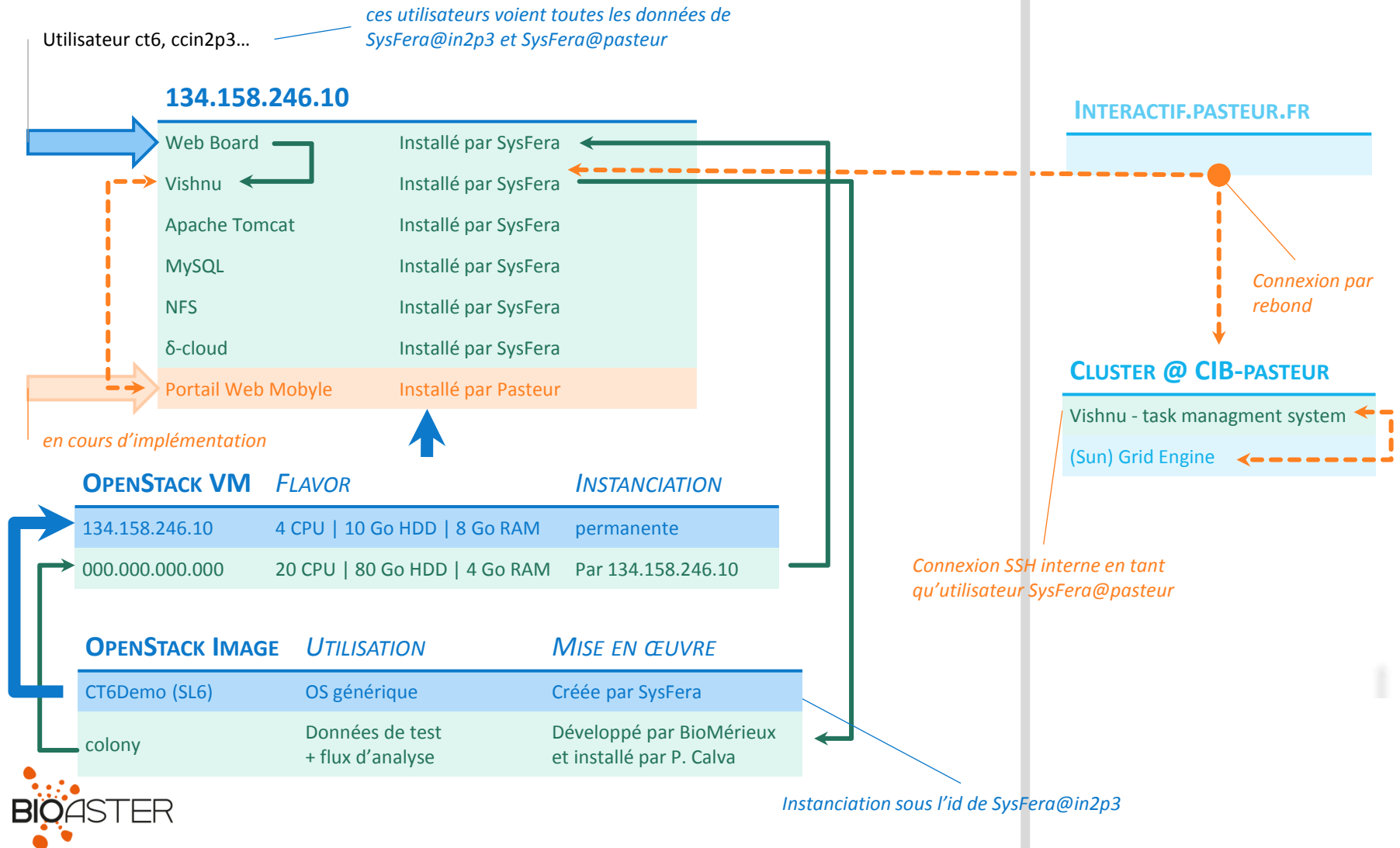
BIOASTER @ CIB-PASTEUR



ORGANISATION ACTUELLE DU PROTOTYPE DE PLATEAU TECHNIQUE 6

BIOASTER @ CC-IN2P3

BIOASTER @ CIB-PASTEUR



DÉROULEMENT IDÉAL DE LA SOUMISSION D'UNE TÂCHE

BIOASTER @ CC-IN2P3

BIOASTER @ CIB-PASTEUR

Tâche émise par l'utilisateur λ pour le projet « assemblage »

OPENSTACK VM	FLAVOR	INSTANCIATION
vishnu.in2p3.fr	4 CPU 10 Go HDD 4 Go RAM	permanente
000.000.000.000	4 CPU 500 Go HDD 1 To RAM	Par OGE
000.000.000.000	20 CPU 2 To HDD 20 Go RAM	Par OGE

Comment instancier à la demande ?

ORACLE GRID ENGINE

2x nœuds d'assemblage	4 cœurs 1 To RAM
8x nœuds généralistes	4 cœurs 10 Go RAM

Comment monter les espaces à la demande ?

STOCKAGE

Données brutes λ	>100 Go
Données assemblées λ	<5 Go
Données Bioaster	<100 Go
Données assemblées μ	<5 Go
Résultat d'analyses λ	<100 Go
Résultats d'analyses μ	<500 Go
Données publiques de référence	<5 To
Logiciels	<50 Go

OPENSTACK IMAGE UTILISATION

Assemblage	flux d'assemblage
Analyse 1	Logiciels d'analyse 1
Analyse 2	Logiciels d'analyse 2
Analyse n	Logiciels d'analyse n

Comment générer des images ?

Les espaces de stockage utilisateurs doivent être cloisonnés
⇒ Confidentialité des données

CONCLUSION

CONTRAINTES

- Cloisonnement des données entre projets
- *Exclusivité des ressources ?*
- *Organisation des utilisateurs et des groupes d'utilisateurs ?*

CHOIX TACTIQUES

- Analyse des données sur le lieu de production des données
- Guichet unique d'utilisation de gestion des droits et des ressources

DIFFICULTÉS

- Activité variable des plateaux techniques
- Estimation du volume de données
- Plusieurs utilisateurs sont attachés à la ligne de commande
- Plusieurs utilisateurs sont attachés à l'immédiateté
- Grande quantité de logiciels à gérer sur des durées de vies très courtes