



# INSU: Quelques Eléments sur Science Ouverte et Données de la Recherche

Rencontres RESIF: 15 Novembre 2021

A. Chambodut, M. Gerin, Nathalie Pothier, J.P. Vilotte

# EOSC: Europe et France

**Europe:** EOSC steering board: L. Crouzet, V. Beckmann (MESRI/DGRI)

- *Trusted, federated, multi-provider environment across borders and scientific disciplines enabling Open Science and FAIR data management practices together with Data policies and rights.*

## EOSC Association (AISBL) : 29 Juillet 2020

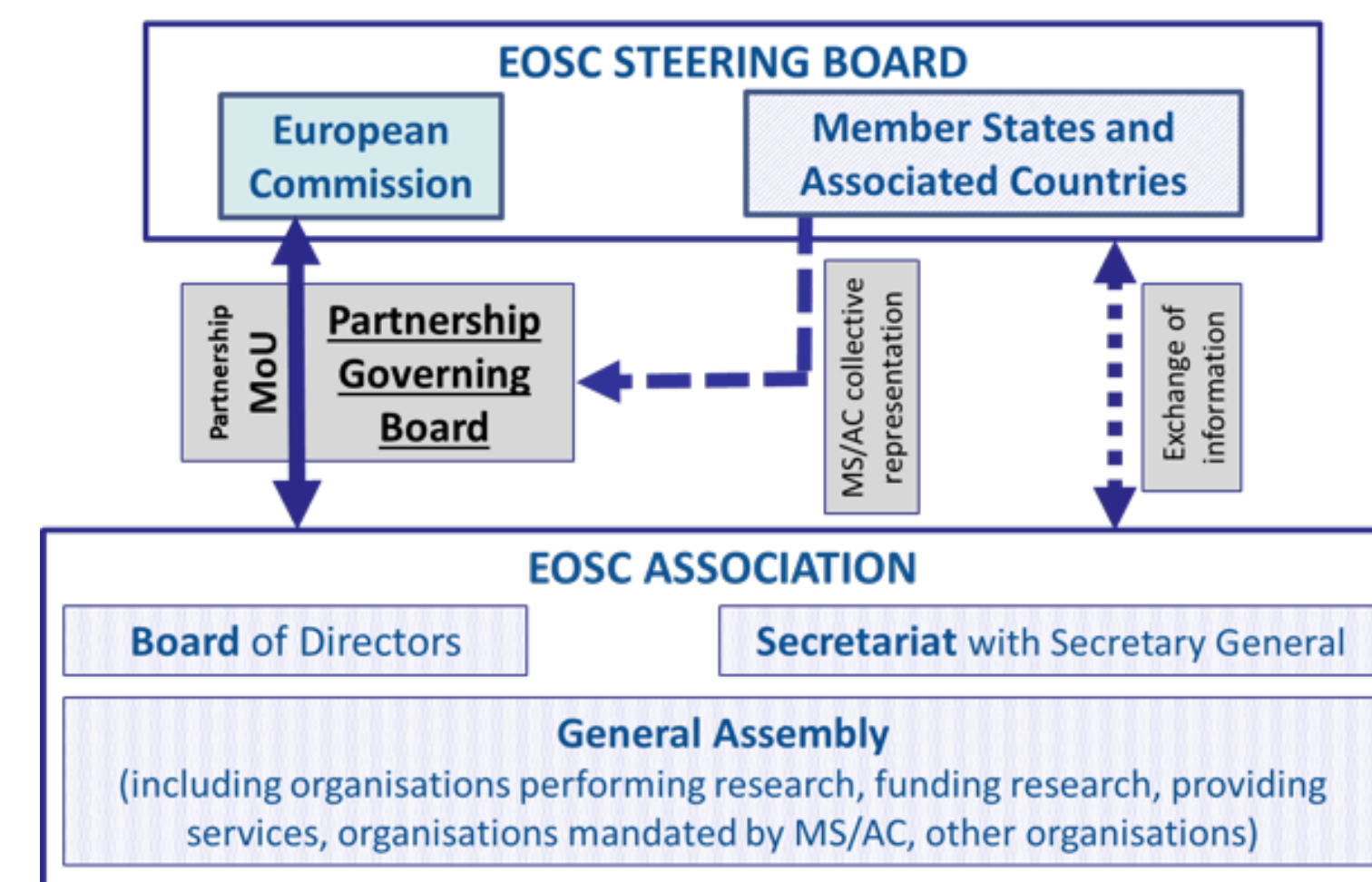
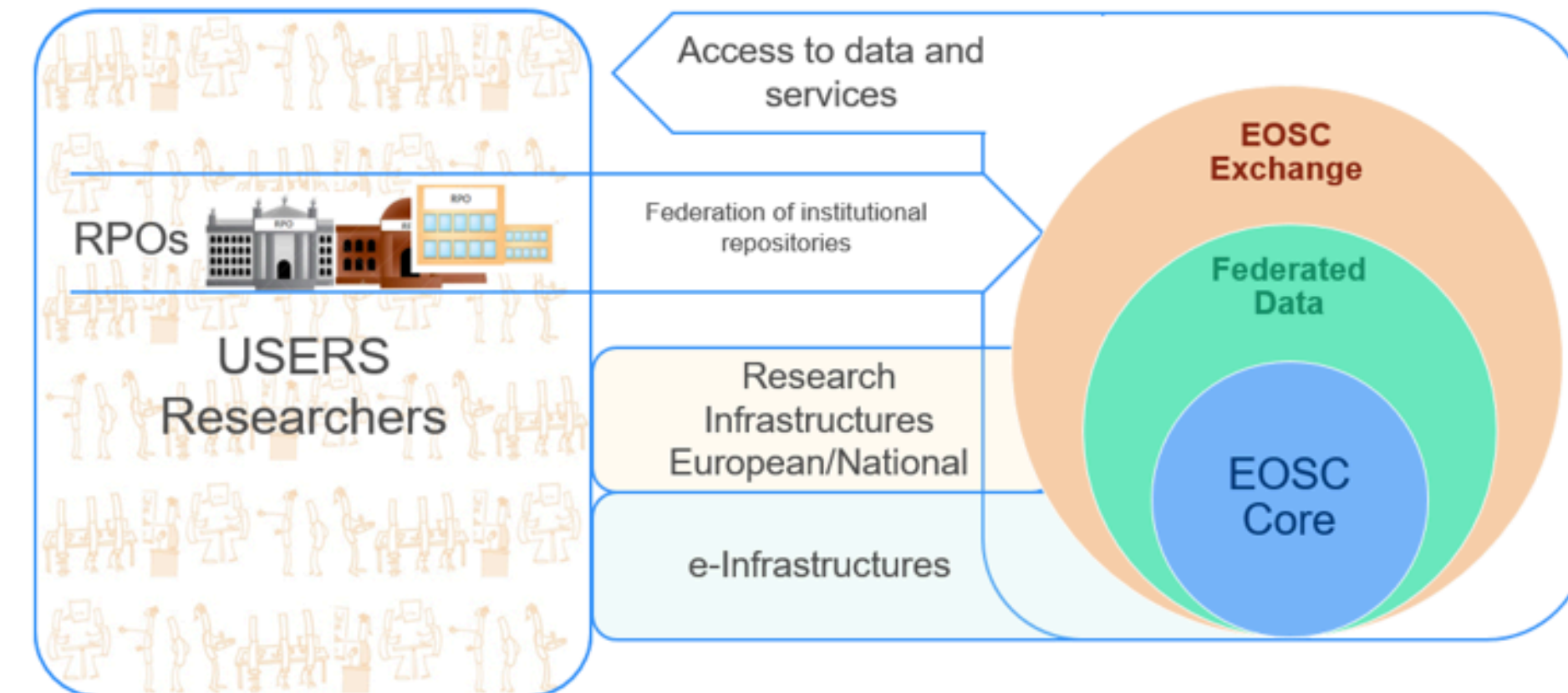
- **Board of Directors** (S. Dumouchel, CNRS) & **General assembly:**
  - ▶ **15 Membres français** : CEA, CINES, CNRS, CPU, GENCI, IFREMER, INRAE, INRIA\*, INSERM, Obs. Paris, RENATER, Université de Bordeaux, Université Paris, Université Paris-Saclay, Université Strasbourg
  - ▶ **8 Observateurs français** : CGE, Couperin, ANR, IRD, Univ. de Lorraine, Univ. de Nantes, Univ. Grenoble Alpes, Univ. Paris 1 Panthéon-Sorbonne

## • EOSC advisory groups - EOSC Task Forces

- ▶ **Implementation of EOSC** : PID policy and implementation, Researcher engagement and adoption, Rules of participation (RoP) compliance monitoring
- ▶ **Metadata and data quality** : FAIR metrics and data quality, Semantic interoperability
- ▶ **Research careers and curricula** : Data stewardship curricula and career paths; Research careers and recognition and credit, Up-skilling countries to engage EOSC
- ▶ **Technical challenges on EOSC** : AAI architecture, Infrastructure for quality research software, Technical interoperability and services
- ▶ **Sustaining EOSC** : Defining funding models, Long-term data preservation

## ➔ Strategic Research Innovation Agenda (SRIA); Multi-Annual Roadmap

(co-programmation EOSC et INFRA calls Horizon Europe)



**France** (V. Beckmann, DGRI)

**DGRI** : Structurer paysage français de l'EOSC

**Collège EOSC France (COSO et COSIN) :**

- Représentants d'organismes de l'ESR
- CNRS: DDOR - Denis Veynante, Sylvie Rousset

# Contexte national & CNRS: Données & Science Ouverte

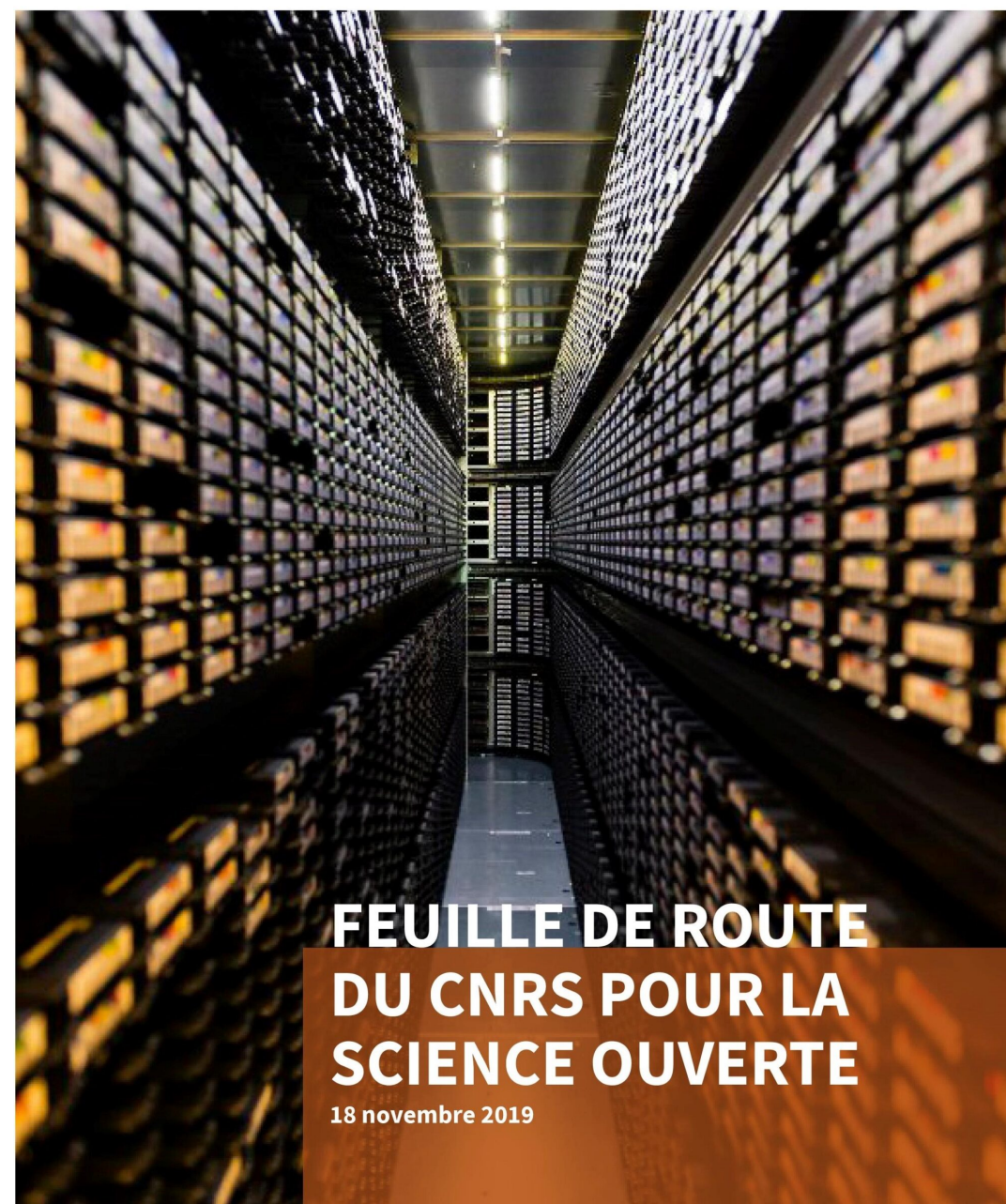
## Deuxième plan national pour la science ouverte (2021-2024) :

*Généraliser l'accès ouvert aux publications, Structurer, partager et ouvrir les données de la recherche, Ouvrir et promouvoir les codes sources produits par la recherche, Transformer les pratiques pour faire de la science ouverte le principe par défaut*

- Comité pour la science ouverte (COSO); Fond National pour la Science Ouverte (FNSO), Comité Stratégique Infrastructures Numériques (COSIN)

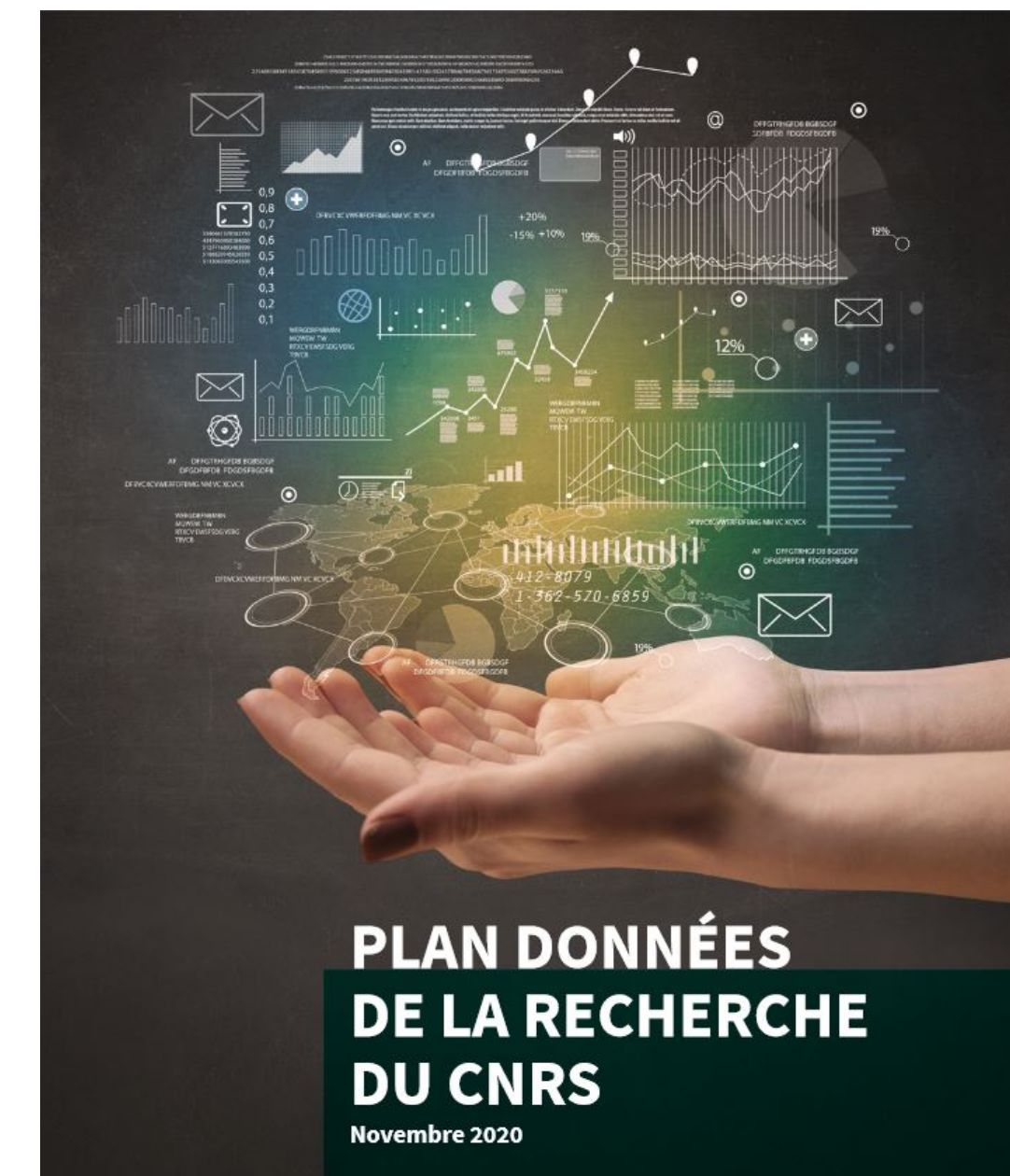
## Feuille de route Science Ouverte du CNRS

Consulter le PDF (version française)



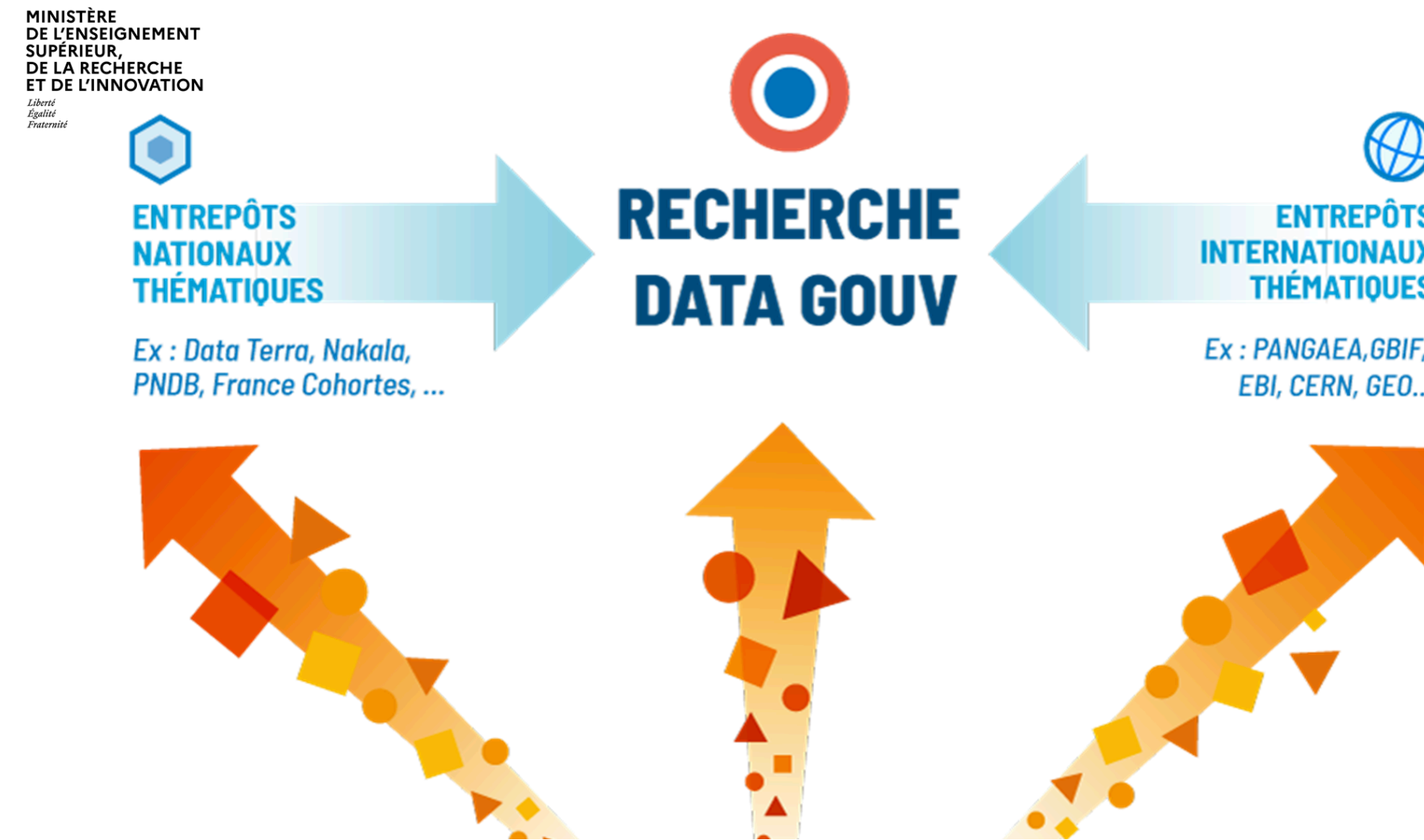
## Plan Données de la recherche" du CNRS

Consulter le PDF (version française)



# Plateforme *Fédérée* Données de la Recherche (Data Gouv)

## Modules (5)



## Accompagnement



## Subsidiarité, Confiance, Mutualisation, Visibilité

- **Principe de subsidiarité** avec entrepôts disciplinaires existants (nationaux, internationaux), modération des dépôts et accompagnement des chercheurs
- **Une plateforme modérée** avec démarche certification
- **Mutualisation** des moyens (logiciels et infrastructures)
- **Visibilité données et provenance**: établissements, structures/infrastructures de recherche, SNOs, chercheurs
- **Principe de non cession de la propriété intellectuelle** sur les données et logiciels
- **Visibilité internationale et contribution à EOSC**

# Data Gouv : projet entrepôt et catalogue de données

Une équipe projet dédiée qui mobilise des compétences INRAE, des partenaires ESR (dont CNRS) et des prestataires: principe de subsidiarité, catalogues inter-opérables et moissonables avec entrepôts thématiques

## PILOTAGE DU PROJET (WP1, INRAE)

Comité projet régulier  
avec tous les responsables de WP

## PORTAIL RECHERCHE DATA GOUV (WP2, PARTENAIRE ESR)

Offrir une entrée ergonomique et éditorialisée  
aux interfaces de dépôt/catalogue. Ateliers avec les usagers

## ENTREPÔT ET CATALOGUE (WP3, INRAE)

Offrir une qualité de service aux usagers compatible  
avec l'état de l'art (gestion et partage de la donnée,  
technologie). Ateliers avec les usagers

## FORMATION, ACCOMPAGNEMENT, COMMUNICATION (WP5, PARTENAIRE ESR)

Garantir la meilleure  
appropriation et utilisation  
des services par les usagers

## DISPOSITIF CIBLE (WP6, INRAE)

Définir les éléments clés  
(organisation, compétences,  
modèle économique)  
indispensables pour  
la pérennisation  
de l'entrepôt/catalogue  
de la plateforme nationale  
de données

## INFRASTRUCTURE APPLICATIVE ET TECHNIQUE (WP4, INRAE)

Concevoir et déployer une architecture résiliente  
et adaptable aux évolutions (volumétrie et usages)



### ENTREPÔT

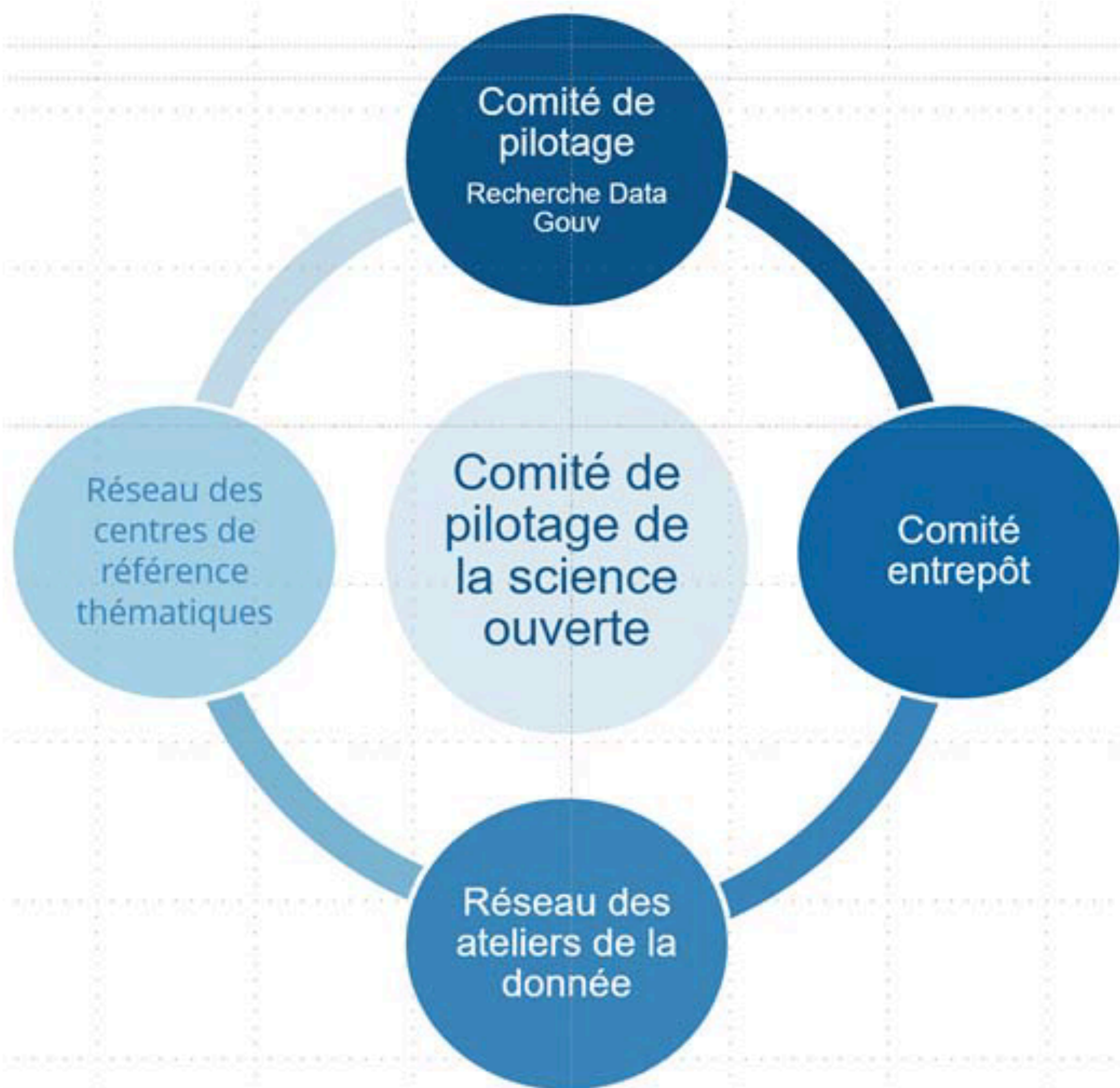
Interface Web  
de dépôt de ses données  
par le chercheur  
+ espace de modération



### CATALOGUE

Repérage et moissonnage  
des données  
des entrepôts externes

# Recherche Data Gouv : gouvernance version beta



## Réseau des ateliers de la donnée

**Rôle :** Préfiguration du réseau des ateliers de la donnée par les premiers lauréats des ateliers

- Préfigure et installe les services des ateliers de la donnée
- Contribue à la création du réseau des ateliers de la donnée
- Contribue l'expression des besoins et aux tests de l'entrepôt et du catalogue
- Rend compte au Comité de pilotage Recherche Data Gouv

**Composition :**

- MESRI (Présidence), préfigureurs des ateliers de la donnée

**Fréquence :** mensuelle

## Réseau des centres de référence thématiques

**Rôle :** porte les référentiels thématiques et disciplinaires

- Conçoit et porte les référentiels de pratiques thématiques/disciplinaires
- Contribue à l'arborescence thématiques/disciplinaires de données de l'entrepôt
- Porte le catalogue de référence des entrepôts spécialisés nationaux et internationaux
- Soutient l'articulation de la plateforme avec les dispositifs thématiques spécialisés
- Soutient l'inscription dans l'écosystème international

**Composition :**

- MESRI, centres de référence thématiques (p.ex. Instituts du CNRS)

**Fréquence :** bimestrielle

# Data Gouv : ateliers de la donnée

## Ateliers de la donnée (politique de sites)

- Point d'entrée en proximité des chercheurs sur toute nature de besoin relatif à la donnée
- Regroupent une palette de compétences inter-disciplinaires et des ressources issues de différents établissements sur un ou plusieurs sites (**INSU: OSUs, SNOs, dans le respect de leurs missions**)
- Apportent un *premier niveau d'expertise et de services*
- S'appuient sur l'expertise des services juridiques, partenariales, SI de proximité
- S'appuient sur des expertises scientifiques nationales (**INSU: OSUs, SNOs, centres de référence thématiques en respectant leurs missions**)
- Animés en réseau pour mutualiser leurs pratiques, formations... (**INSU: synergie/coordination avec Data Terra, GAIA Data**)

## Appel à manifestation d'intérêt

- Labellisation initiatives existantes et les nouvelles initiatives
- Accompagnement de la réflexion des établissements pendant les phases de maturation et de préfiguration de leurs propositions
- Structuration au plan territorial le réseau des ateliers de la donnée dans une logique de partenariat entre établissements sur un ou plusieurs sites (**INSU: synergie/coordination avec Data Terra, Gaia Data ...**)

## Modalités

- 3 Appels par an (- 3 millions d'euros), soutien financier pour le démarrage/renforcement des initiatives 50 à 200 k€ / projets
- Suivi des projets labellisés à 18 et 36 mois
- **1ère phase:** Appels à intentions incluant une *lettre d'engagement* des différents partenaires
  - **22 novembre 2021:** ouverture soumission des intentions; **20 Décembre 2021:** clôture soumission des intentions
- **Journée de rencontre entre les projets :** 6 janvier 2022
  - Interactions/échanges entre projets, convergence des intentions dans l'optique d'un lancement de l'appels à projets dédié
- **2ème phase :** Appel à projets dédié
  - **10 janvier 2022:** ouverture soumission projets; **23 février 2022:** clôture soumission des projets
- **Première liste d'ateliers labellisés :** **fin avril 2022:** annonce première liste d'ateliers labellisés

**Assurer visibilité des OSUs et SNOs (dans le respect de leurs missions) dans les réponses à l'AMI en lien avec Data Terra et GAIA Data**

# Contexte national: Data Centres Régionaux

Réflexion engagée 2015 : un data centre labellisé par région (avec des exceptions !!)

- **Critères** : *niveau technique (engagement de fermeture de salles existantes); connexion réseau; engagement d'hébergements; engagement des acteurs ESRI partenaires; gouvernance et soutien des présidences d'établissements*
- **Labellisation** : comité présidé par la DGRI (DGRI, régions, DINU, CPU, CDEFI, DNE)

## Data centres régionaux labellisés et opérationnels :

- **Bretagne** : préfiguration du Data centre Eskemm-data hébergé par B<>com, opérée par le GIP Numérique de Bretagne
- **Bourgogne Franche Comté** : Data centre « DC-UBFC » à Dijon, opéré par l'Université Bourgogne Franche Comté
- **Centre-Val de Loire** : Data centre hébergé sur le site Orléans Grand Campus, opéré par le BRGM
- **Grand Est** : Data centre de Strasbourg, opéré par l'Université de Strasbourg
- **Hauts-de-France** : « Datacentre-HdF » à Villeneuve D'Ascq (cité des sciences), opéré par l'Université de Lille
- **Normandie** : Data centre CDR-1 à Saint-Etienne-du-Rouvray, opéré par le CRIANN
- **Pays de la Loire** : Data centre hébergé et opéré par l'Université de Nantes
- **Provence Alpes Côte d'Azur** : Data centre « DC-Sud » à Marseille, opéré par l'Université Aix-Marseille
- **Occitanie** : Espace Clément Ader à Toulouse, opéré par l'Université Fédérale Toulouse Midi Pyrénées

## Data centres régionaux labellisés en construction :

- **Auvergne-Rhône-Alpes** : futur Data centre GR3@Grenoble
- **Bretagne** : futur Data centre régional Eskemm-data sur le Campus de Beaulieu à Rennes
- **Grand Est** : futur Data centre situé à Nancy bâtiment Apollo du CNRS Vandoeuvre les Nancy
- **Nouvelle-Aquitaine**: futur Data centre régional, projet Campus@datacenter portée par l'Université de Bordeaux
- **Normandie** : futur Data centre régional CDR-2 opéré par le CRIANN qui remplacera l'actuel CDR1
- **Pays de Loire** : futur Data centre régional dans le projet « Dacas », Universités Nantes, Angers et le Mans; en remplacement du DC Nantes



# Data Centres: suivi des opérations

## DGRI :

- Financement (3,6 M€) : études amont et quelques opérations d'amorçage de projets
- Demandes de financement CPER 2021-2027 pour la plupart des Initiatives
- Soutien financier sur la durée: nécessaire pour accompagner les trajectoires de consolidation
- Fléchage de crédits: levier incitatif fort pour le succès de cette transformation.

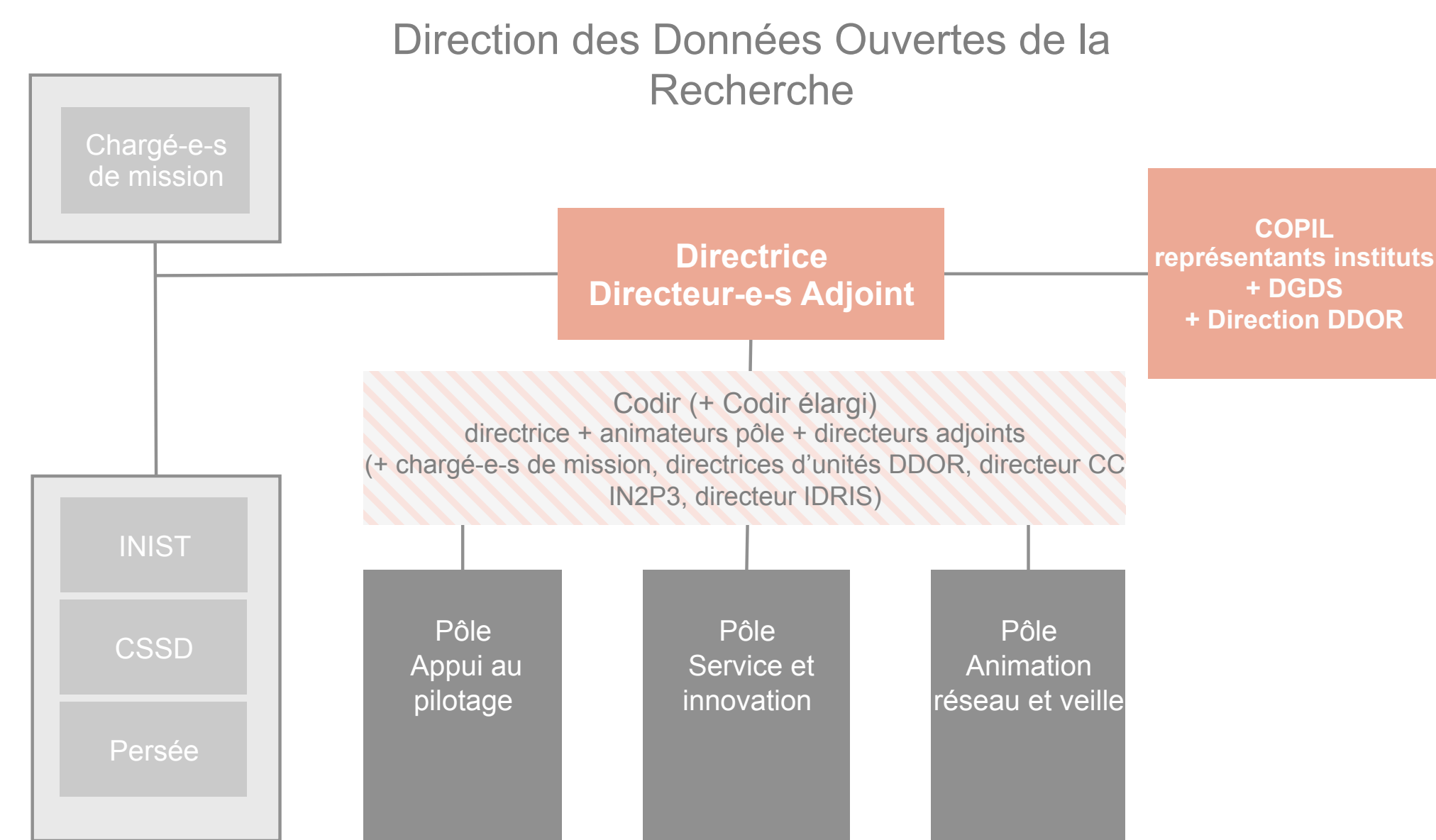
## Groupe de travail "data centres labellisés" : en cours de constitution, animé par le MESRI

- Suivi des projets labellisés, coordination des prochaines étapes
- Fédération de l'offre de services entre data centres régionaux et nationaux (CINES, CC-IN2P3, IDRIS et TGCC)

## CNRS-INSU :

- Inquiétude sur la composition du groupe de travail (CNRS: articulation DDOR et Instituts)
- Articulation avec les Infrastructures de Recherche (Data Terra, Gaia Data, CDS)
- Articulation au niveau régional avec les OSUs et SNOs (hébergements, politique d'accès, services)
- Stratégie transversale au sein de l'INSU

# Contexte CNRS: DDOR



## Direction des Données Ouvertes de la Recherche (DDOR) : décembre 2020

- **Direction** : Sylvie Rousset & Denis Veynante
  - ✓ *Plan de données de la recherche et feuille de route Science Ouverte du CNRS*
- **CoPil** : un représentant par Institut (INSU: J.-P. Vilotte)
- **Missions** : données FAIR & services de données, publications, infrastructures/e-infrastructures numériques, calcul intensif (HPC, HPDA, AI), Science Ouverte
- **EOSC** : DDOR représente le CNRS dans le collège EOSC France

## Premiers éléments d'organisation :

- **GTs** : données de la recherche (Aude Chambodut, J.P. Vilotte), publications (Nathalie Pothier), infrastructures numériques (Maryvonne Gérin Laslier, J.P. Vilotte), EOSC (J.P. Vilotte, Aude Chambodut, Maryvonne Gerin Laslier, Baptiste Cecconi, Chiara Ferrari, Mark Allen, Christelle Pierkot)

## CNRS - INSU:

- **Coordination INSU**: données, calcul, publications, infrastructures numériques et de recherche, science ouverte
  - J-P. Vilotte, Maryvonne Gerin Laslier, Aude Chambodut, Nathalie Pothier
- ➔ **Implication DAS**: liens avec communautés, infrastructures de recherche, projets européens

# CNRS-INSU: actions en cours

## EOSC :

- Articulation GT EOSC (DDOR), Collège EOSC France
- Renforcement liens avec Infrastructures de Recherche
- Coordination participation INSU: EOSC Task Forces, appels EOSC & INFRA Horizon

## Exascale France :

- GT Exascale INSU: DAS et représentants CS : A&A, OA, ST
- Articulation avec les infrastructures recherche:
- Stratégie de soutien INSU aux applications exascale stratégiques

## CNRS - DDOR:

- CoPiL & GTs DDOR (données de la Recherche, Publications, infrastructures numériques, Science Ouverte)
- Implication DDAS :
  - Validation et modération informations Cat OPIDoR
  - Harmoniser ontologies entre les domaines INSU
  - Liens entre Infrastructures, OSUs, SNOs

## GT IRs et Science Ouverte :

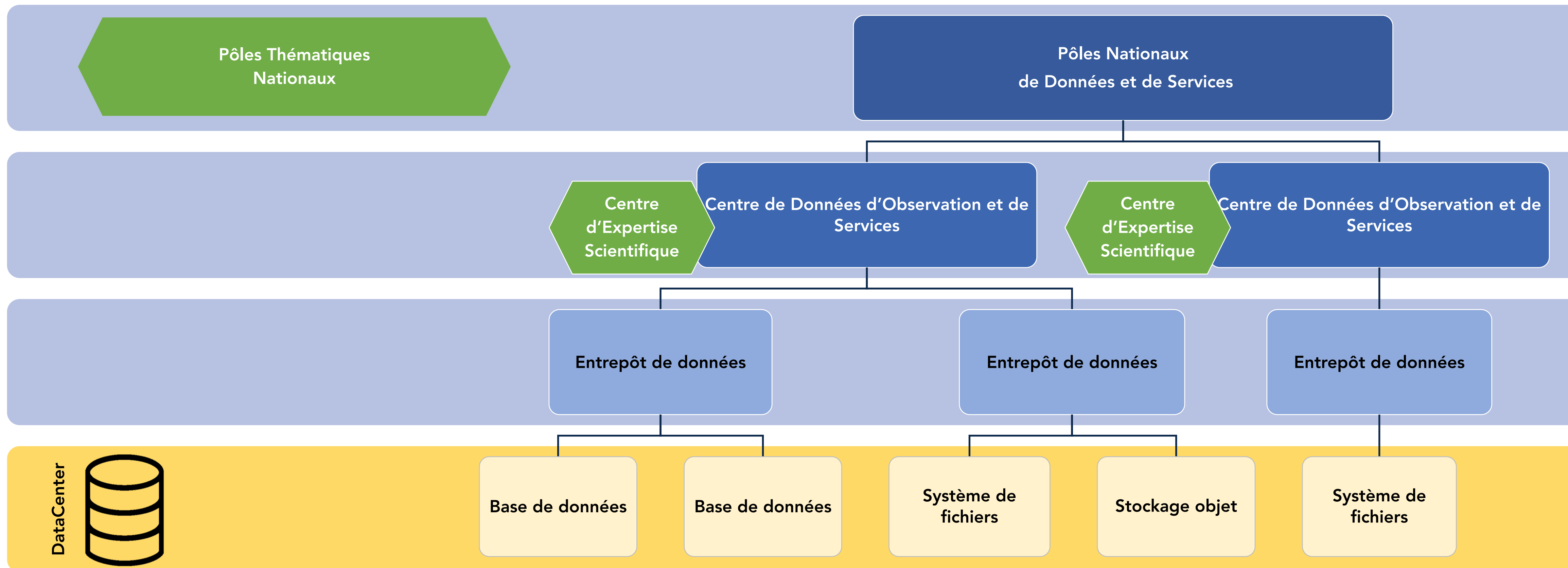
- Ateliers transverses Infrastructures: Données FAIR & multi-source; Science Ouverte, nouvelles expertises
- Analyse matrices d'auto-évaluation Infrastructures
- Co-identifier axes de collaborations transverses
- Contribution stratégie INSU
- Actions collaboratives de recherche multi-disciplinaire

## Science Ouverte :

- Réseau de correspondants dans chaque OSU
- Coordination membres INSU des collèges COSO
- GT transverse INSU: DOI & Licenses
- Préparation école thématique INSU 2022
- Politique et évolution du HAL - INSU

# E-Infrastructure: données - pôles

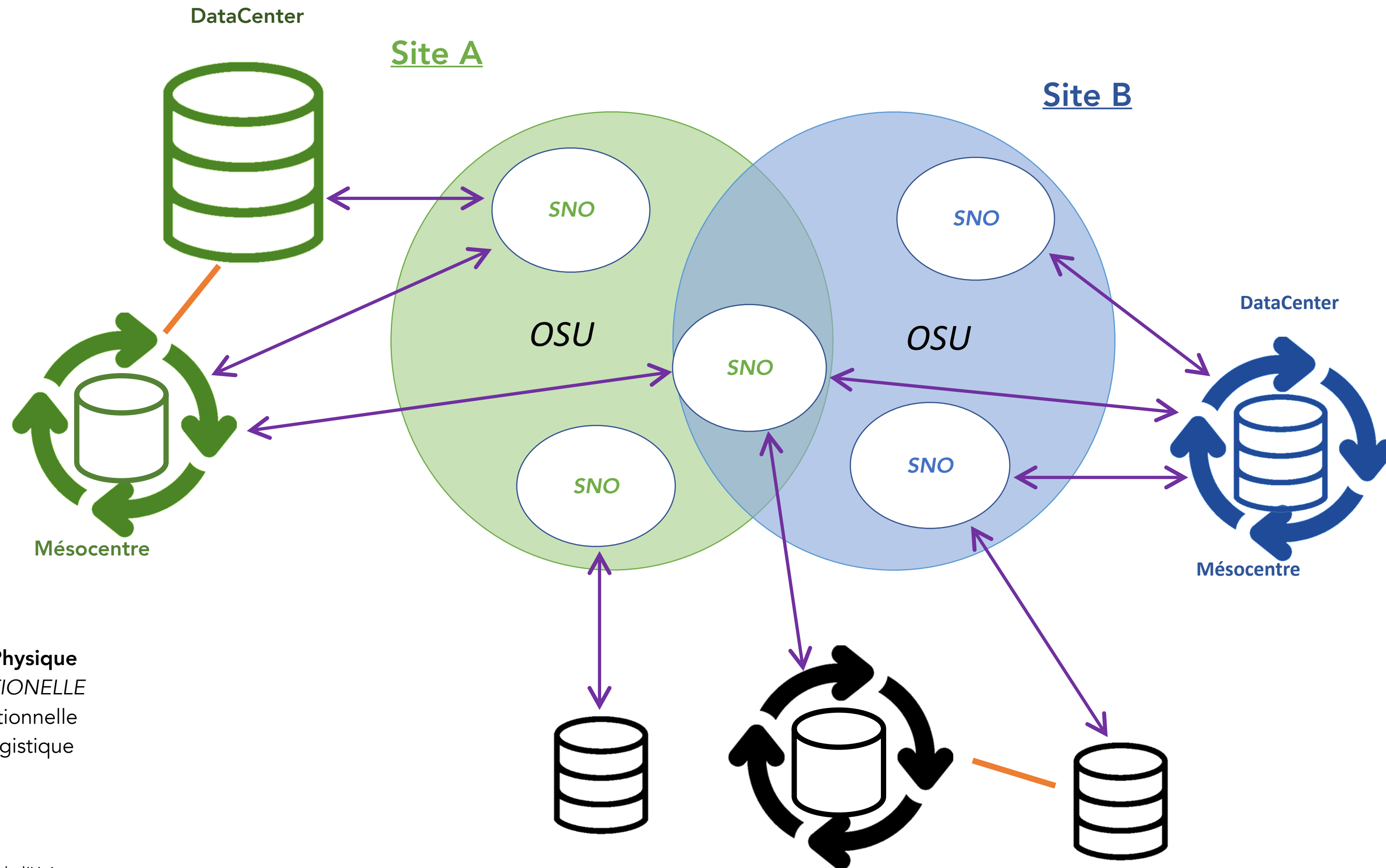
*Élément en discussion*



opérationnel / coordination / physique

# Organisation: vision OSU

*Elément en discussion*



**Infrastructure Physique**  
ENTITE FONCTIONNELLE  
interaction fonctionnelle  
lien réseau et logistique



OSU - Observatoire des Sciences de l'Univers  
SNO - Service National d'Observations  
ESR - Enseignement Supérieur et Recherche

Organismes de l'ESR, Europe, International

# Organisation: vision data repository

*Élément en discussion*

## Éléments en provenance

### Instances

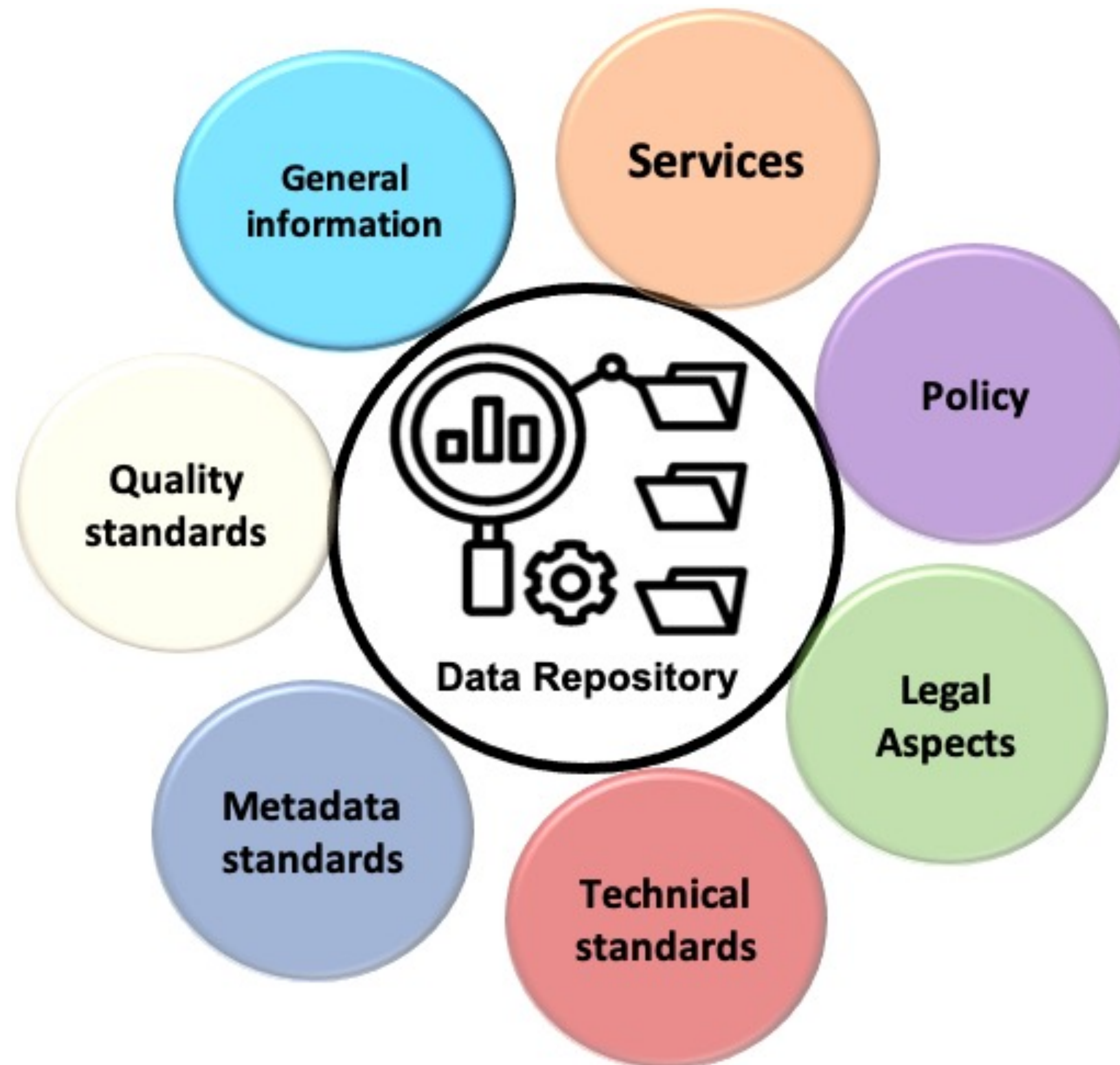
- *Policy*
- *Legal aspects, e.g.: Licence, IPR*

### Communauté scientifique

- Standards données/métadonnée
- Référentiel données/métadonnées

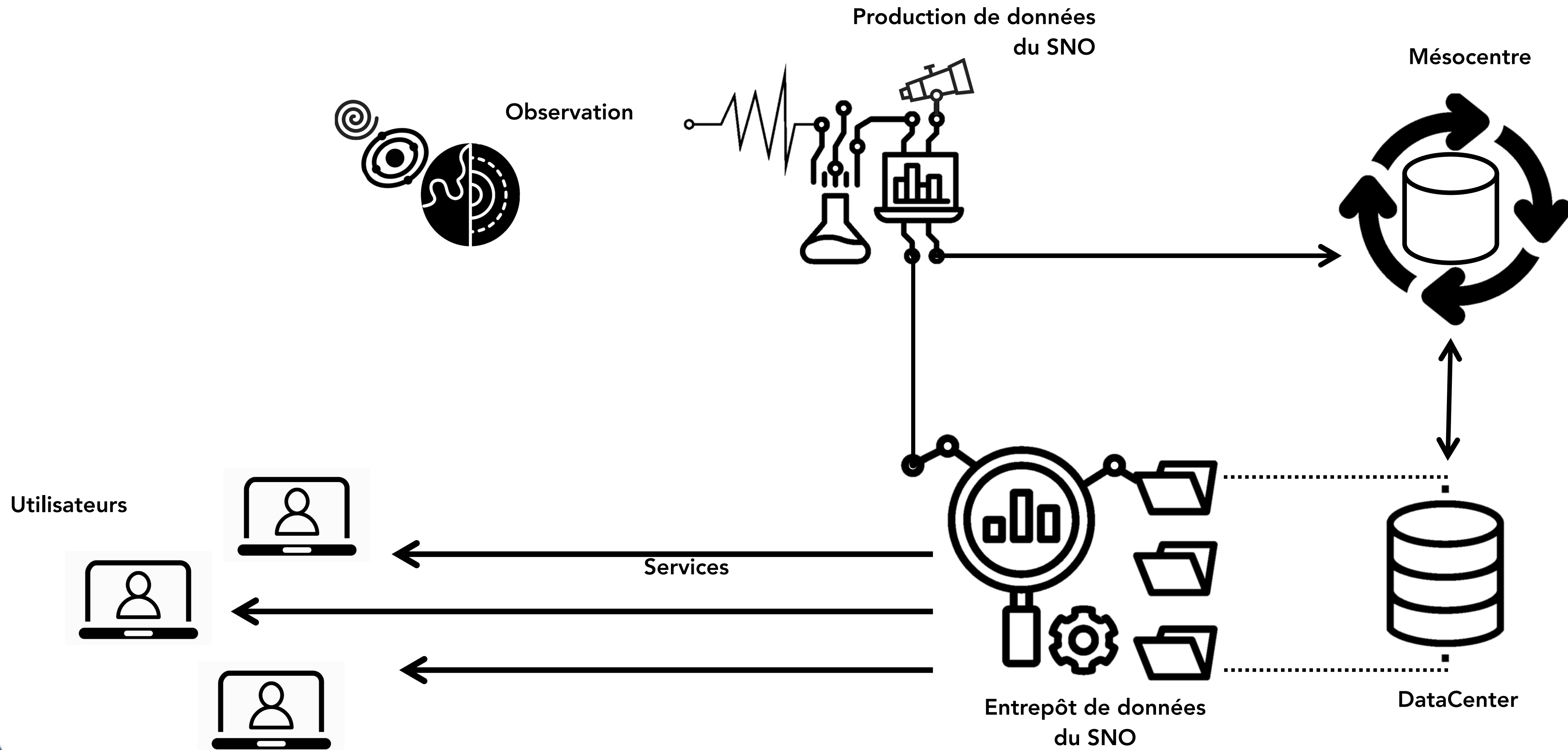
### Entrepôt

- *Services*
- *Data stewardship*



# Organisation: vision SNO

*Elément en discussion*



# Organisation: vision modélisation/simulation numérique

*Elément en discussion*

