

Données de la recherche : Réflexions partagées par la communauté toulousaine

Nathalie Aussenac-Gilles, Pascal Gaillard, Michelle Sibilla
*Comité de Réflexion pour le partage et la valorisation des données de la
recherche et la coordination de la Science Ouverte*

A L'ATTENTION DU RESEAU CAPITOU

Plan de la présentation

- I. Des données à leur ouverture
- II. Les enjeux scientifiques
- III. Réflexions et Actions du CéSO
vers le partage et l'ouverture des résultats de la
recherche
- IV. Conclusions et perspectives

PARTIE I – Des données à leur ouverture

- Quelles données sont concernées ?
- Vers l'ouverture
- Nécessité d'un accompagnement tout au long du cycle de vie des données
- Besoins d'environnement(s) support
- Conclusion

Quelles sont les données concernées ?

Définition de l'OCDE des **données de la recherche** financée sur fonds publics ⁽¹⁾ :

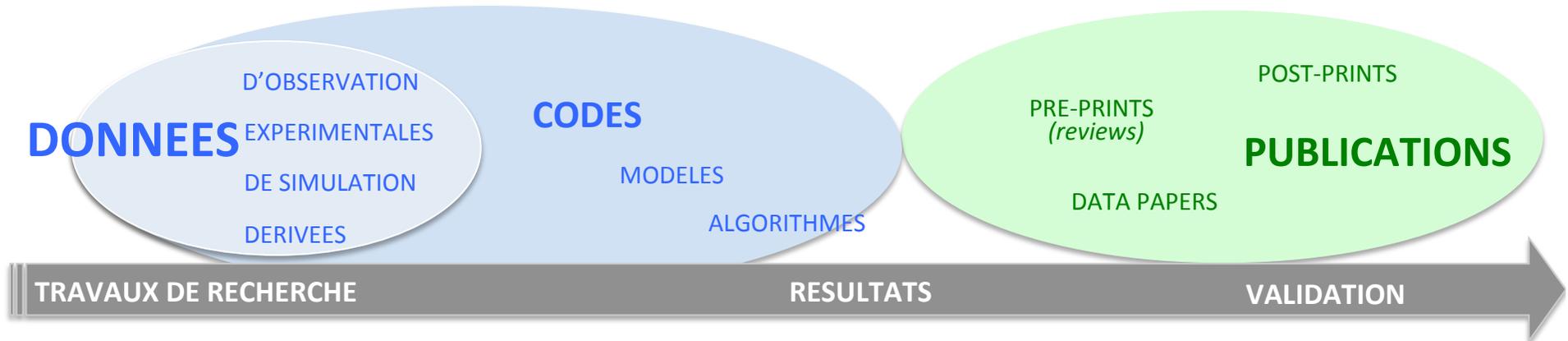
« Les «**données de la recherche** » sont définies comme des enregistrements factuels (chiffres, textes, images, sons) utilisés comme **sources principales** pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour **valider** les résultats de la recherche. »

(1) : Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics
<http://www.oecd.org/fr/science/inno/38500823.pdf>

Quelles sont les données concernées ?

Définition de l'OCDE des données de la recherche financée sur fonds publics ⁽¹⁾ :

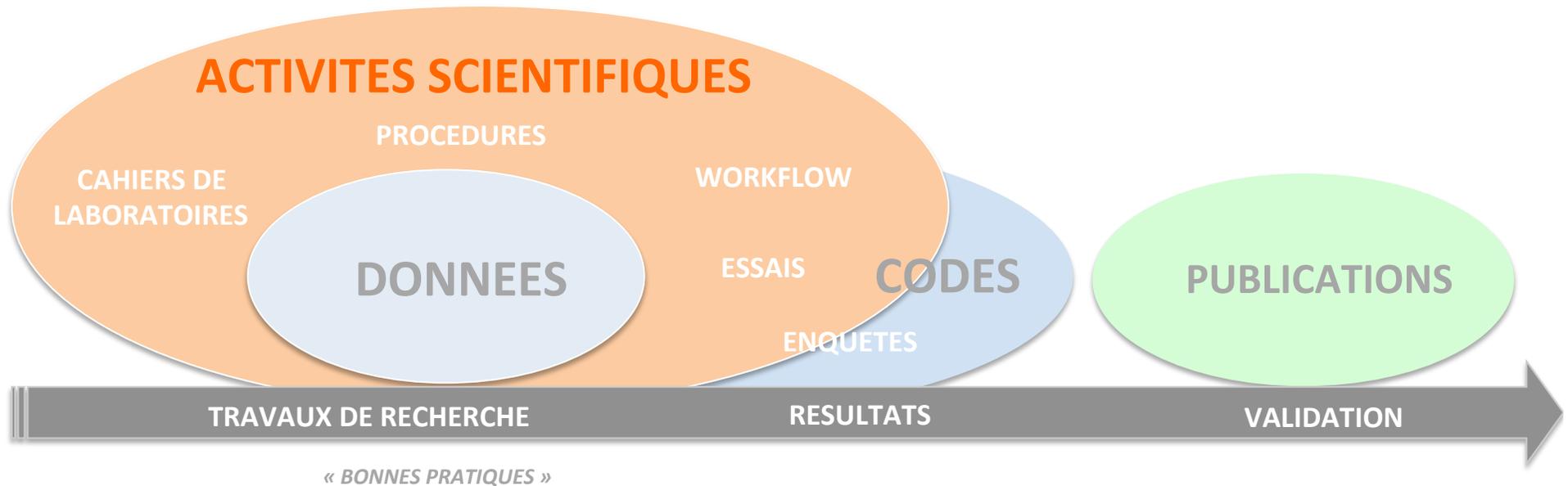
« Les «données de la recherche» sont définies comme des enregistrements factuels (chiffres, textes, images, sons) utilisés comme *sources principales* pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour *valider* les résultats de la recherche. »



Quelles sont les données concernées ?

Autres « données » partageables :

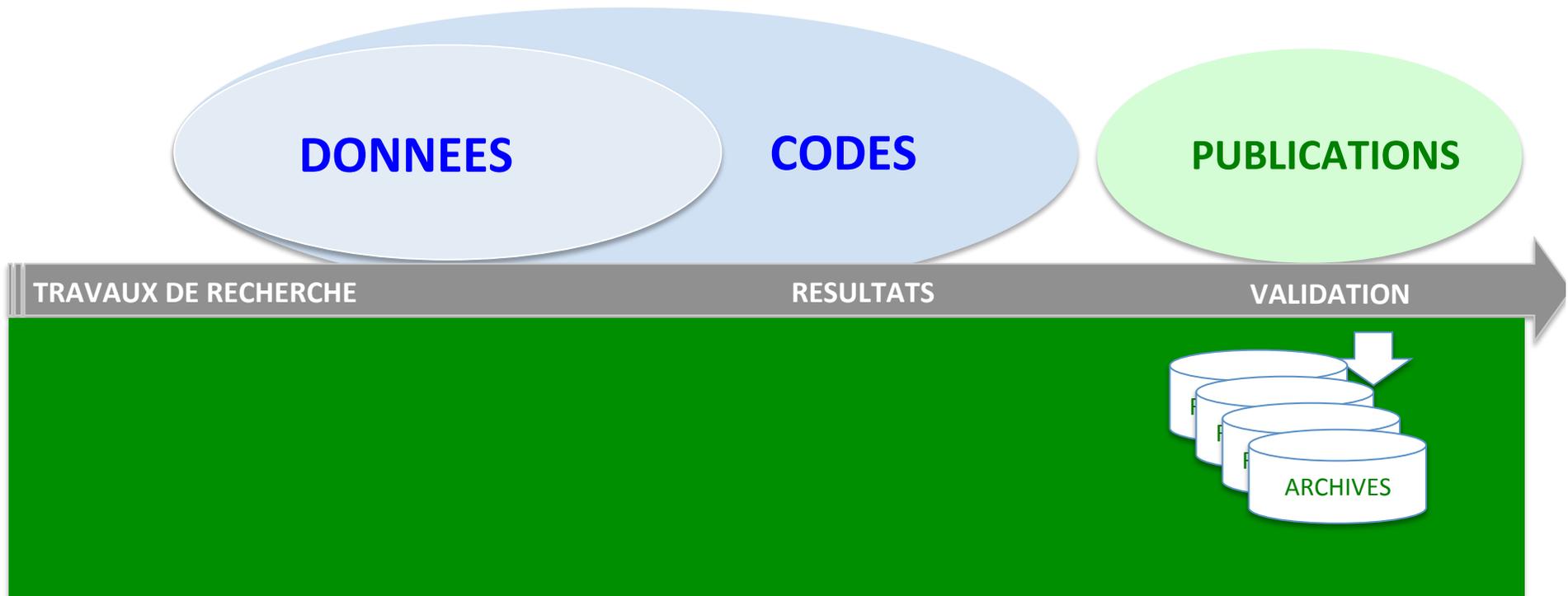
Les « données » liées aux activités scientifiques



Vers l'Ouverture



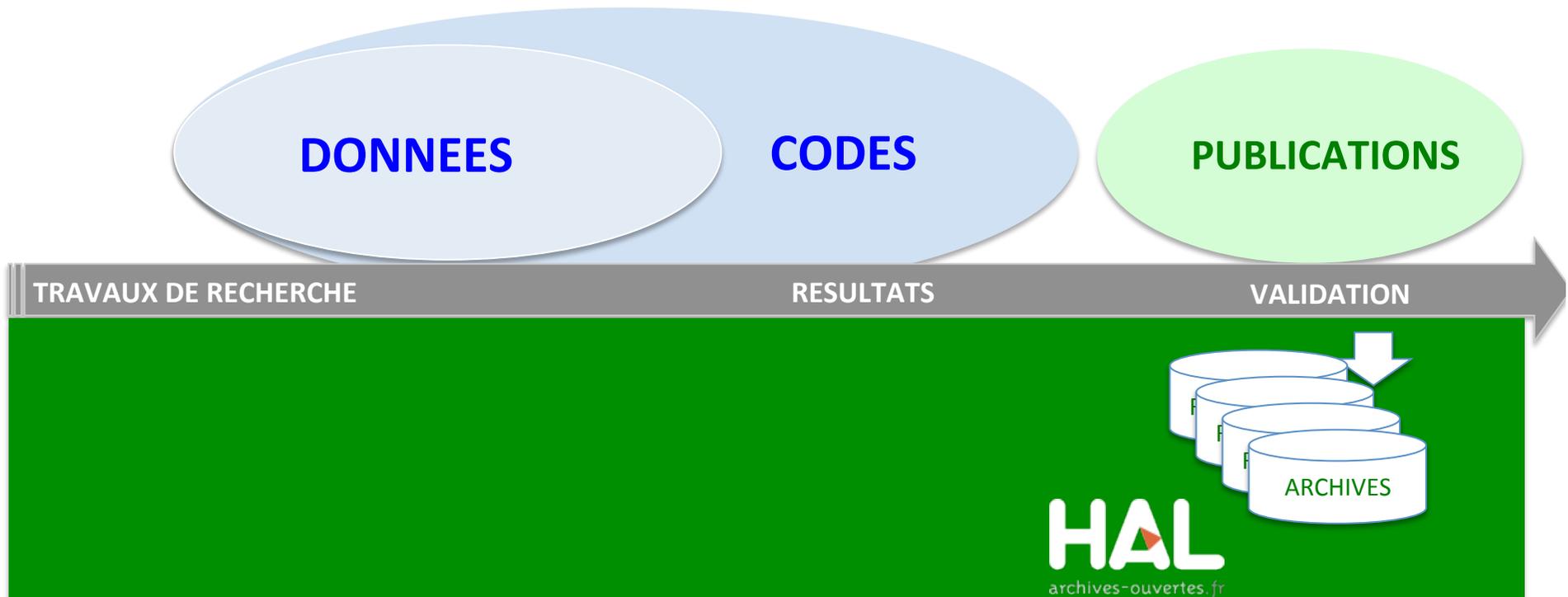
PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS



Vers l'Ouverture



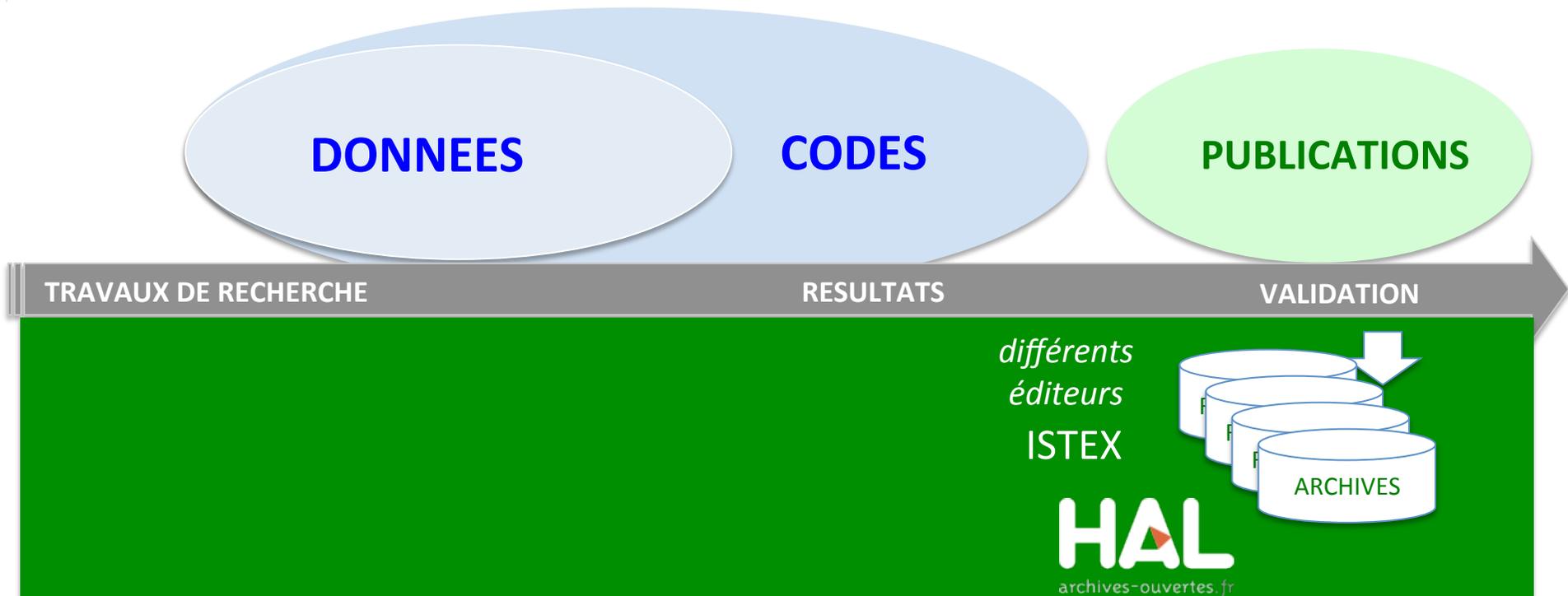
PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS



Vers l'Ouverture



PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS

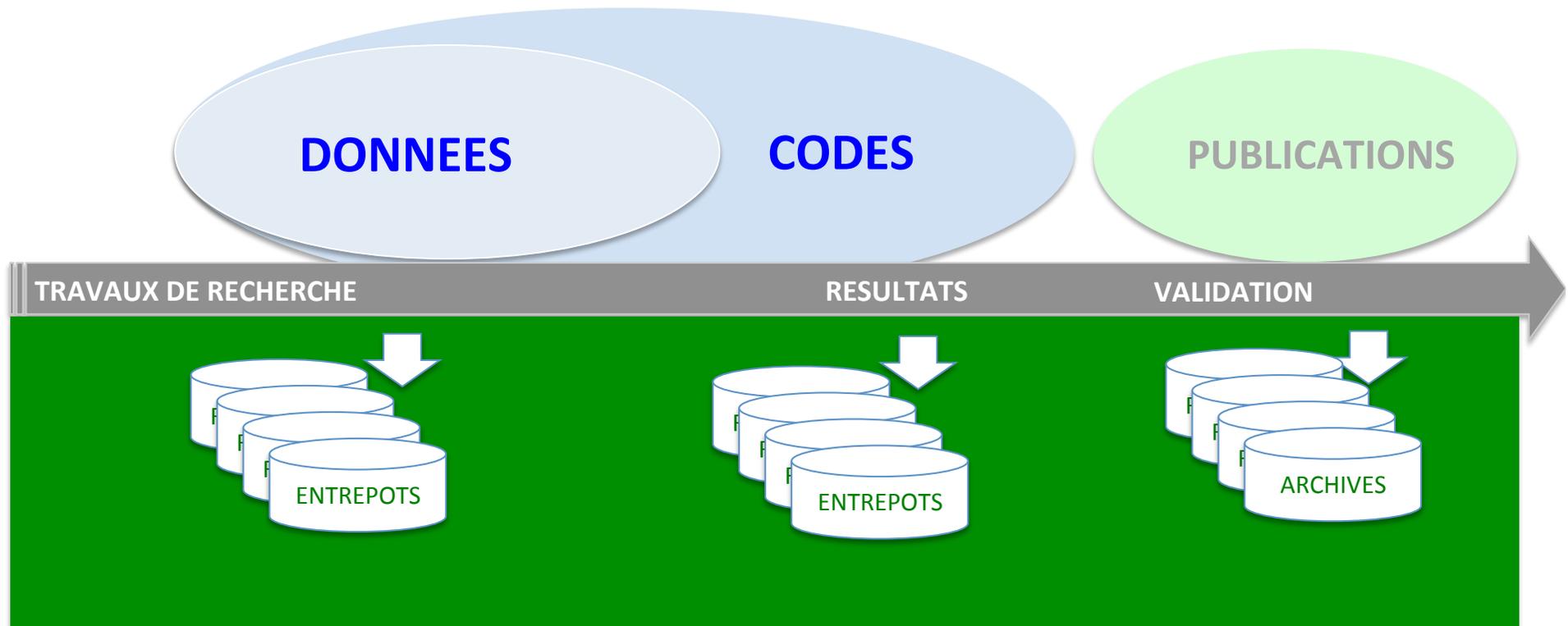


Vers l'Ouverture



PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS

DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE

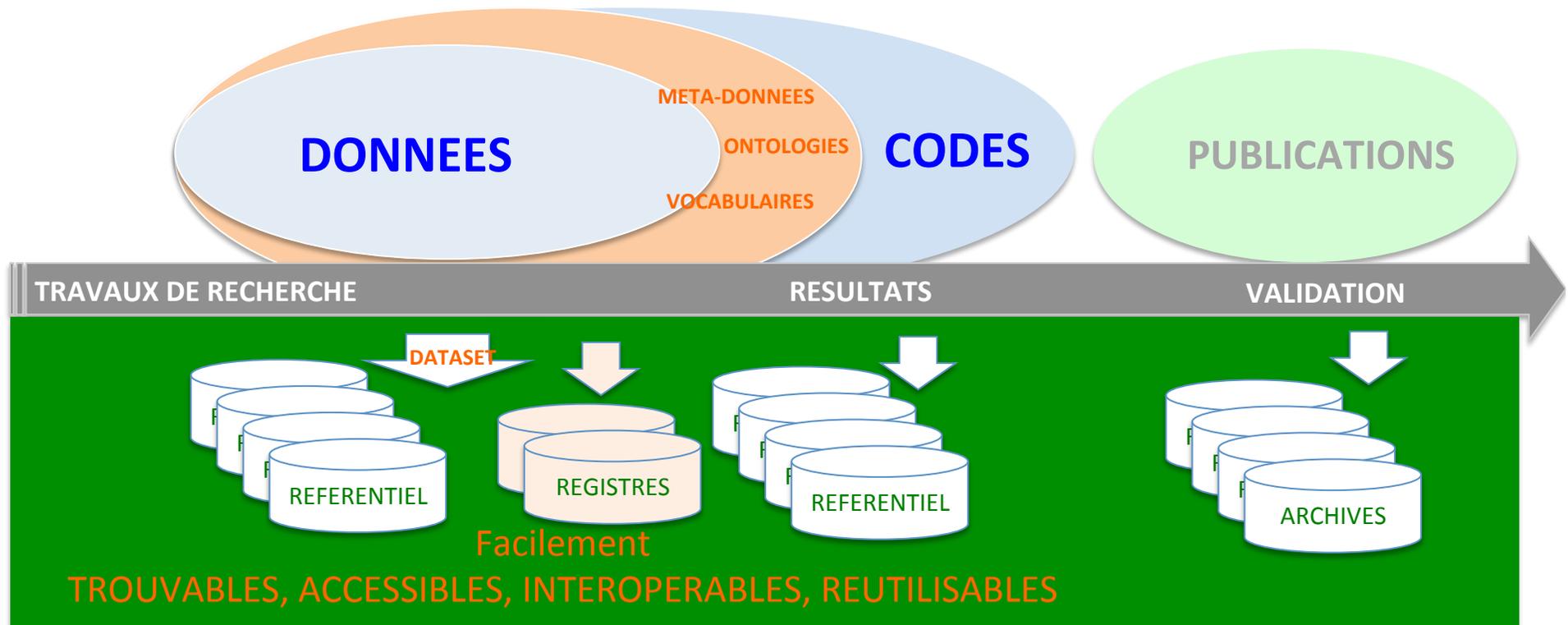


Vers l'Ouverture



PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS

DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE

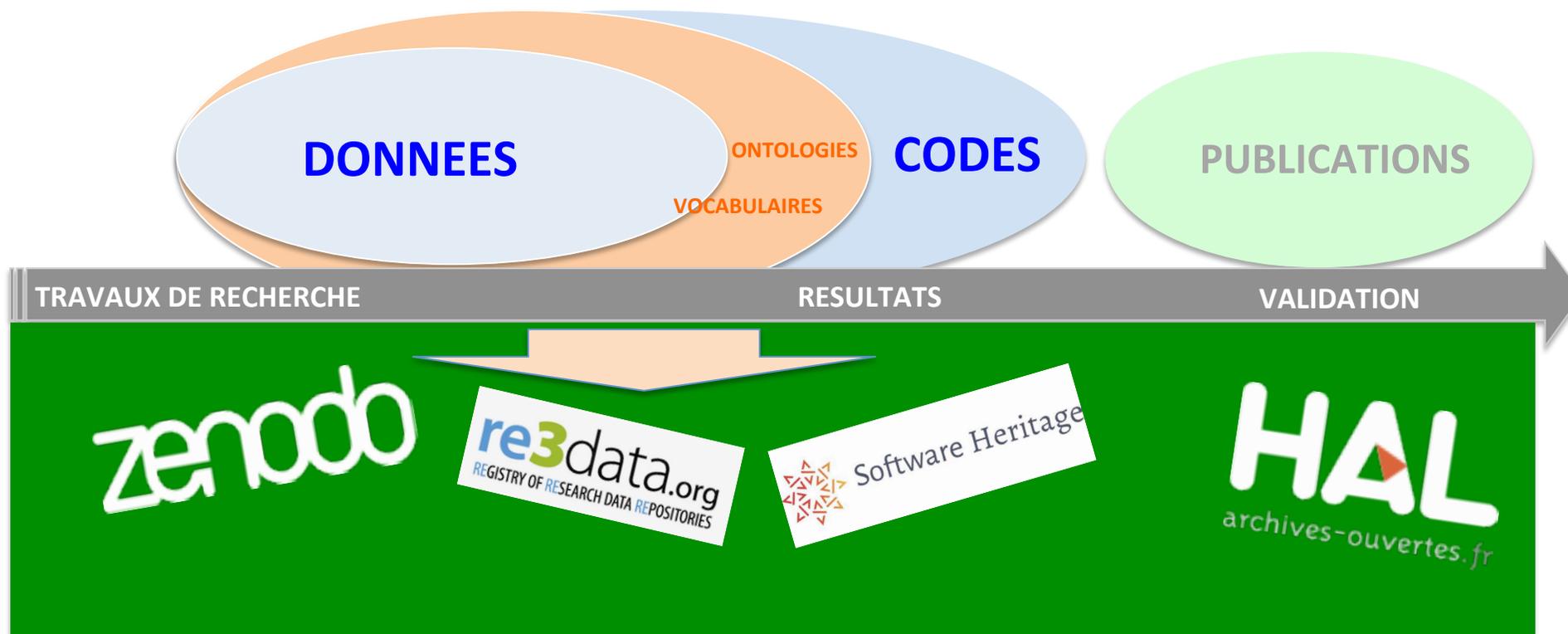


Vers des entrepôts généraux de référence



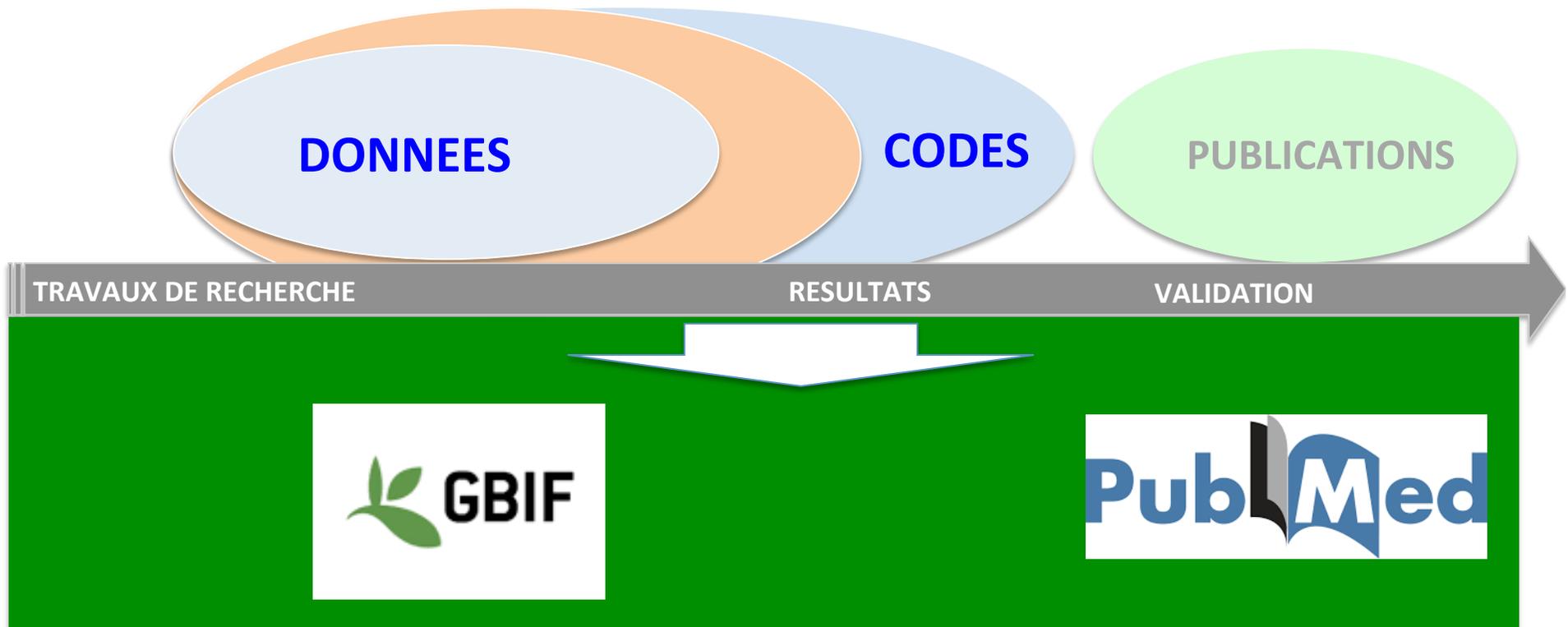
PREMIER AXE : GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS

DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE



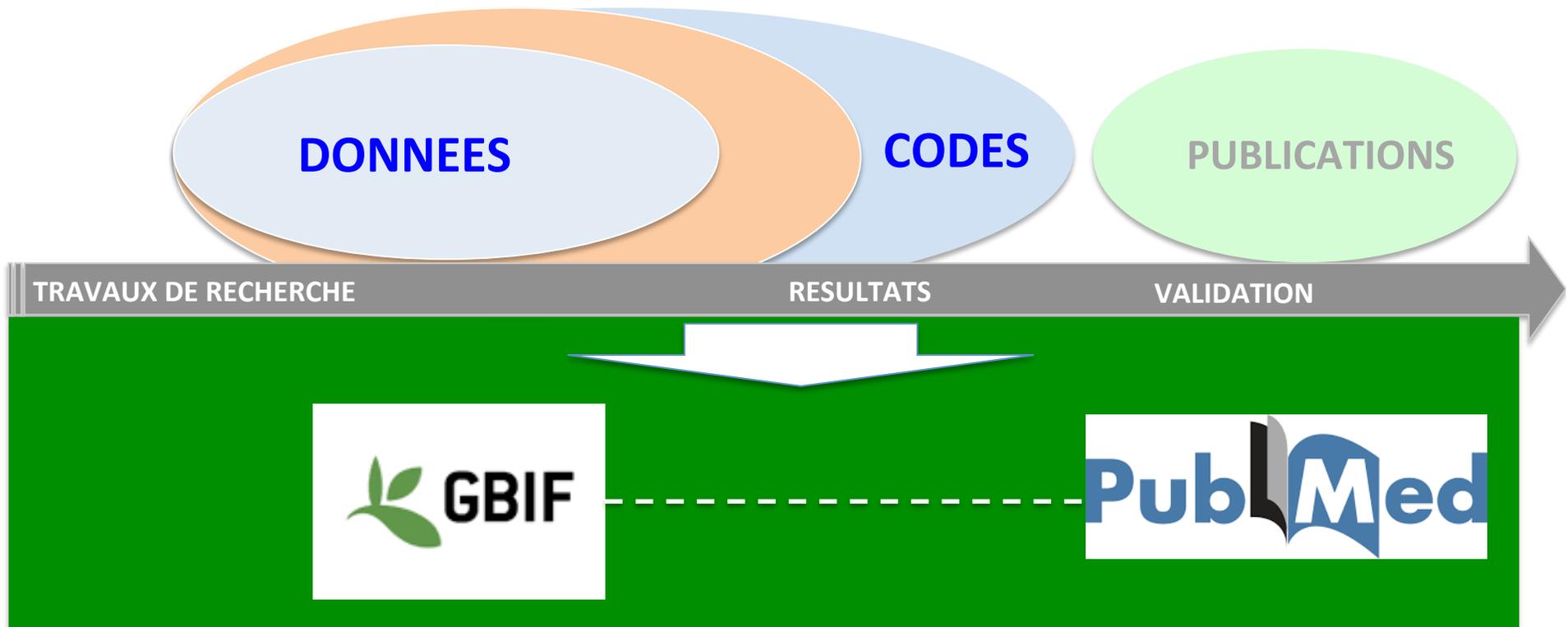
Vers des entrepôts spécifiques de domaine adoptés par des communautés

Exemple pour le domaine de la Biodiversité



Vers des entrepôts spécifiques de domaine adoptés par des communautés

Exemple pour le domaine de la Biodiversité



Les questions fusent ...

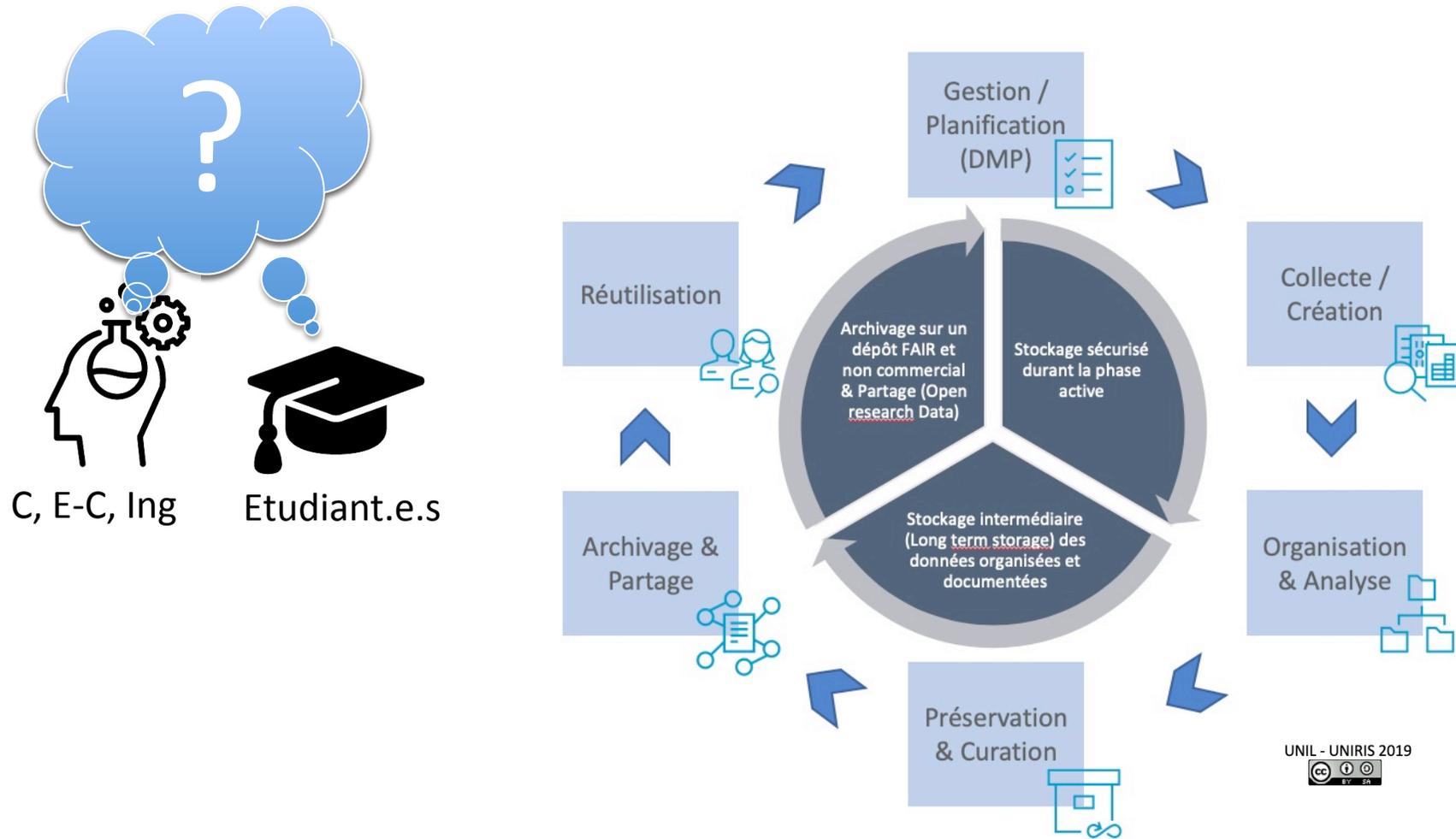


C, E-C, Ing

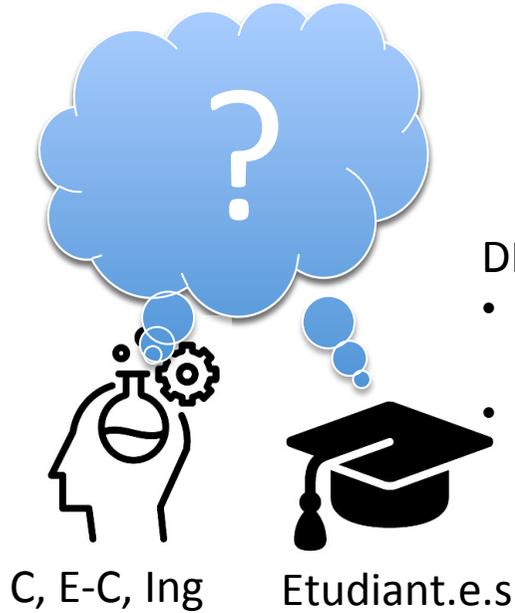


Etudiant.e.s

... Nécessité d'une **gestion** tout au long du **cycle de vie des données**

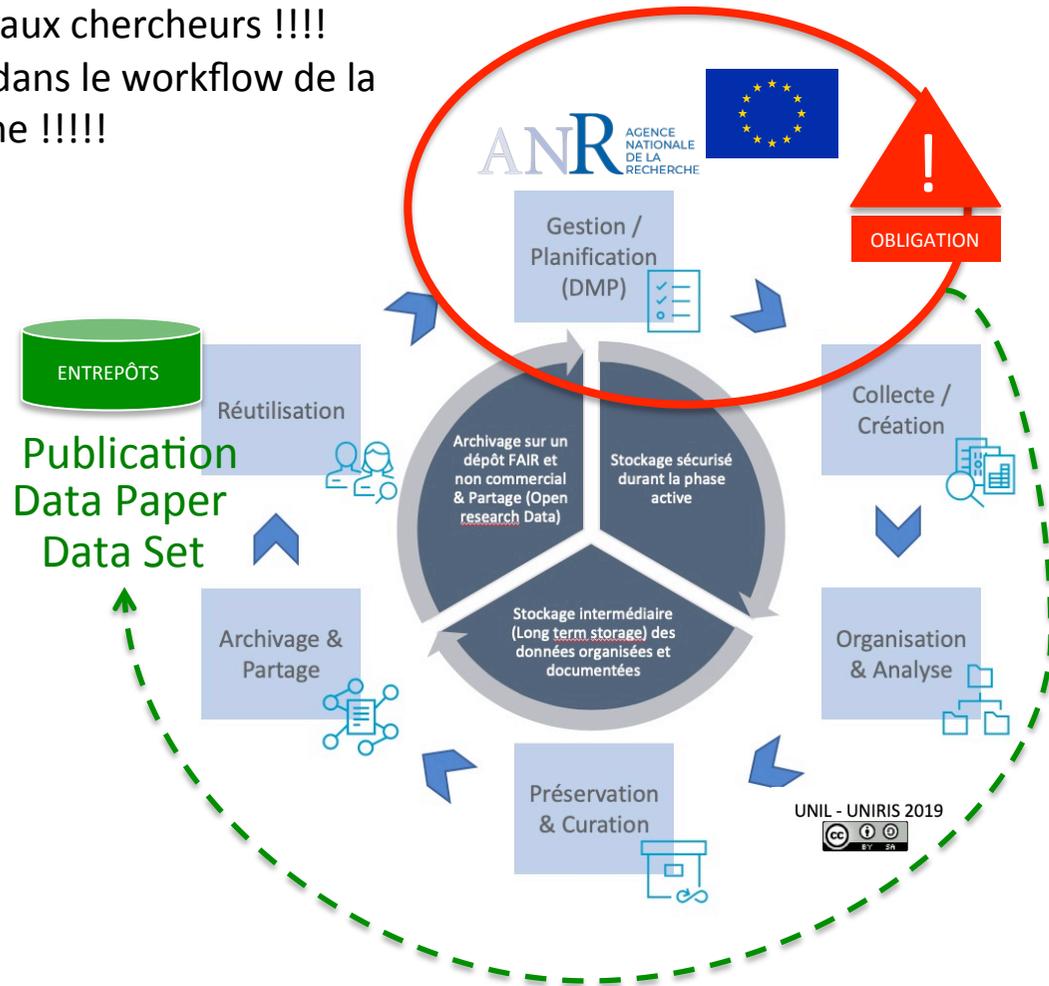


Un point de départ : le PGD/DMP



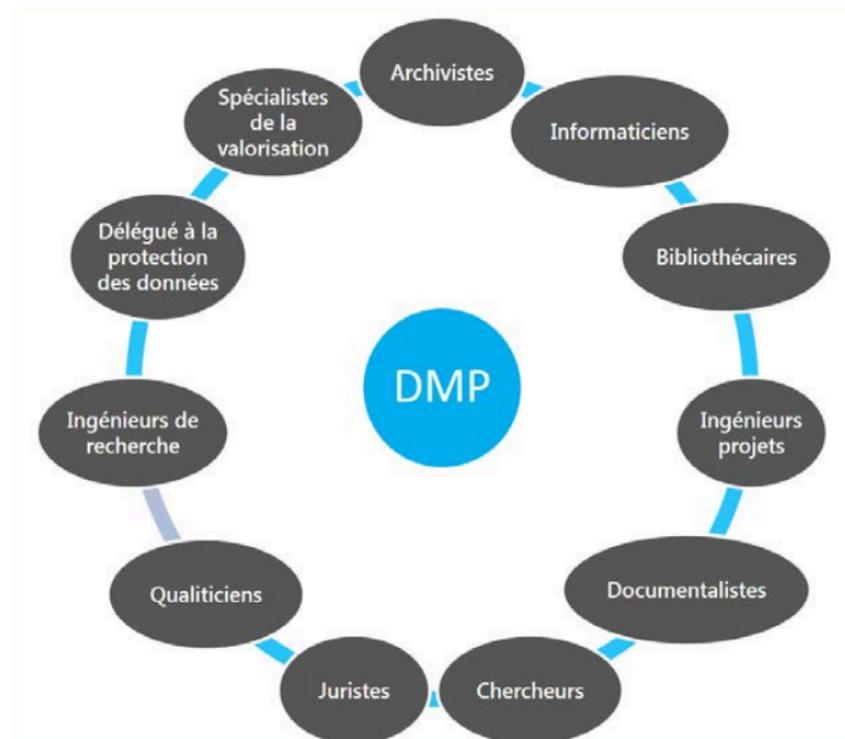
DMP utile/dynamique ...efficient

- doit s'inscrire dans un écosystème support aux chercheurs !!!!
- intégré dans le workflow de la recherche !!!!!



- Le DMP promeut les bonnes pratiques ...
- il est l'inventaire de tous les produits de la recherche

Nécessité d'un accompagnement tout au long du cycle de vie des données



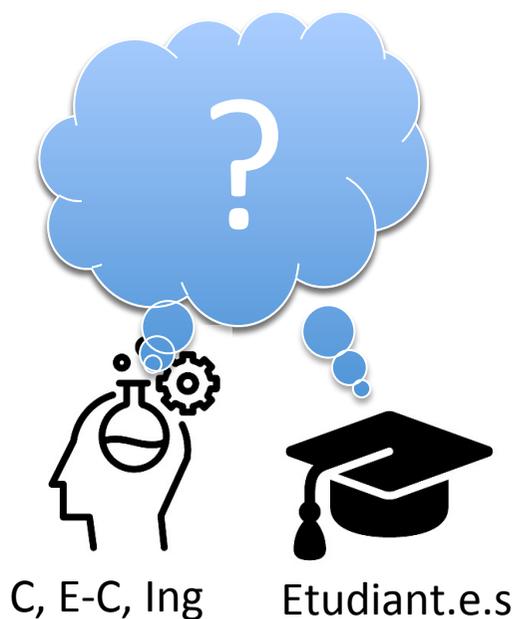
Source :

[Le Data Management Plan \(DMP ou Plan de Gestion de Données \(PGD\) et OPIDoR, Collège de France – Direction des réseaux et partenaires documentaires – Département de bibliométrie et des ressources en ligne](#)

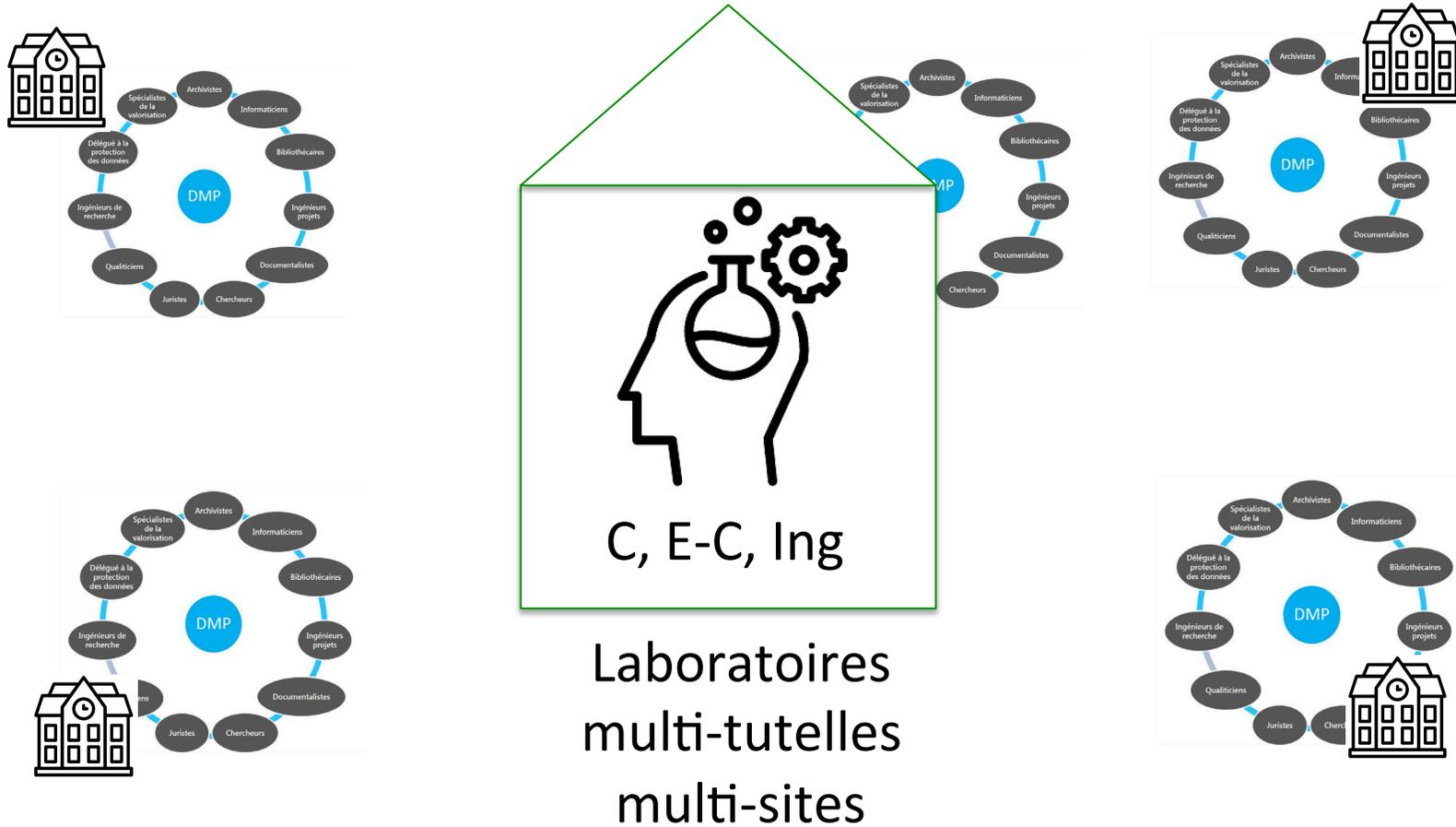


Etablissement

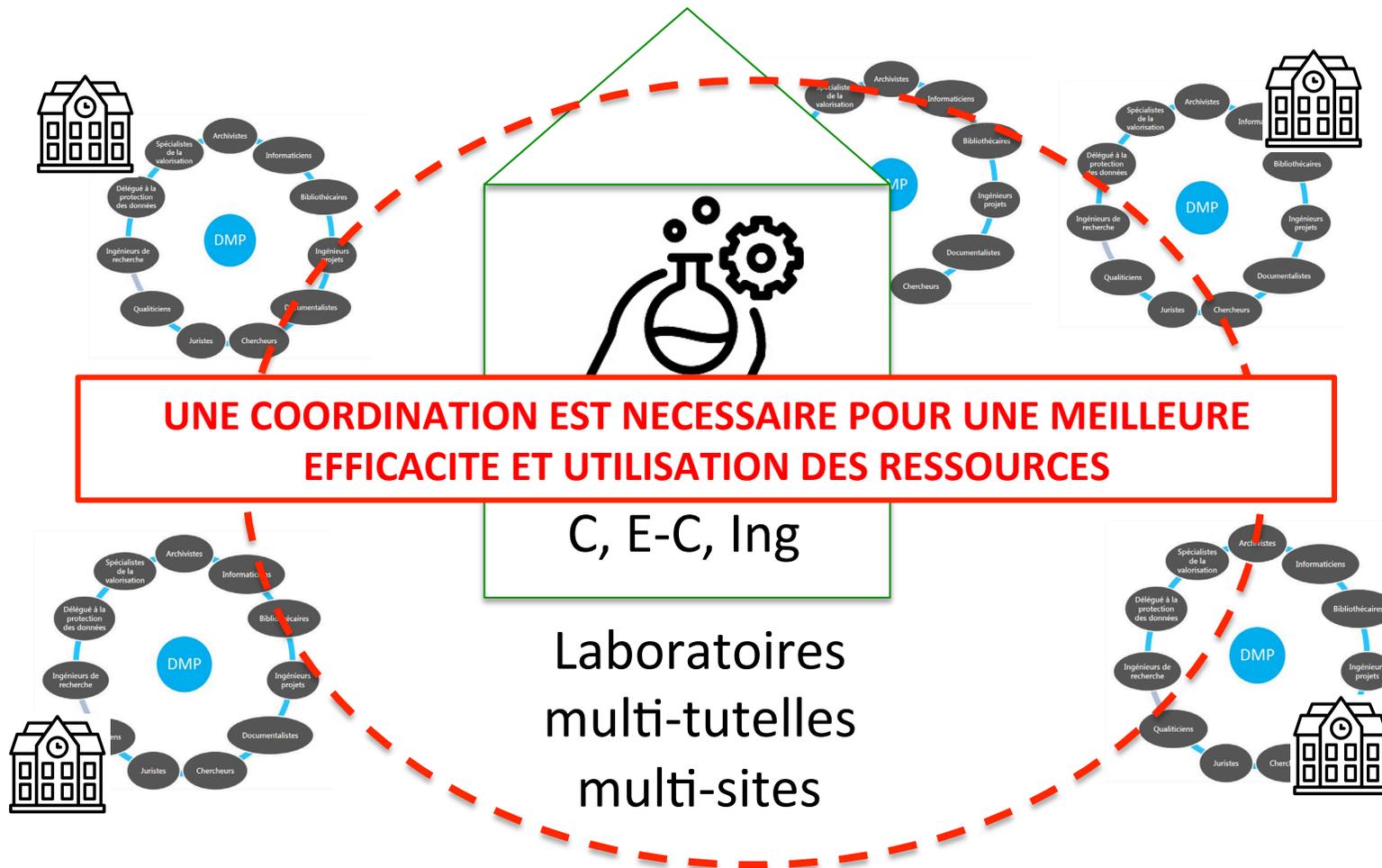
Nécessité d'un accompagnement tout au long du cycle de vie des données



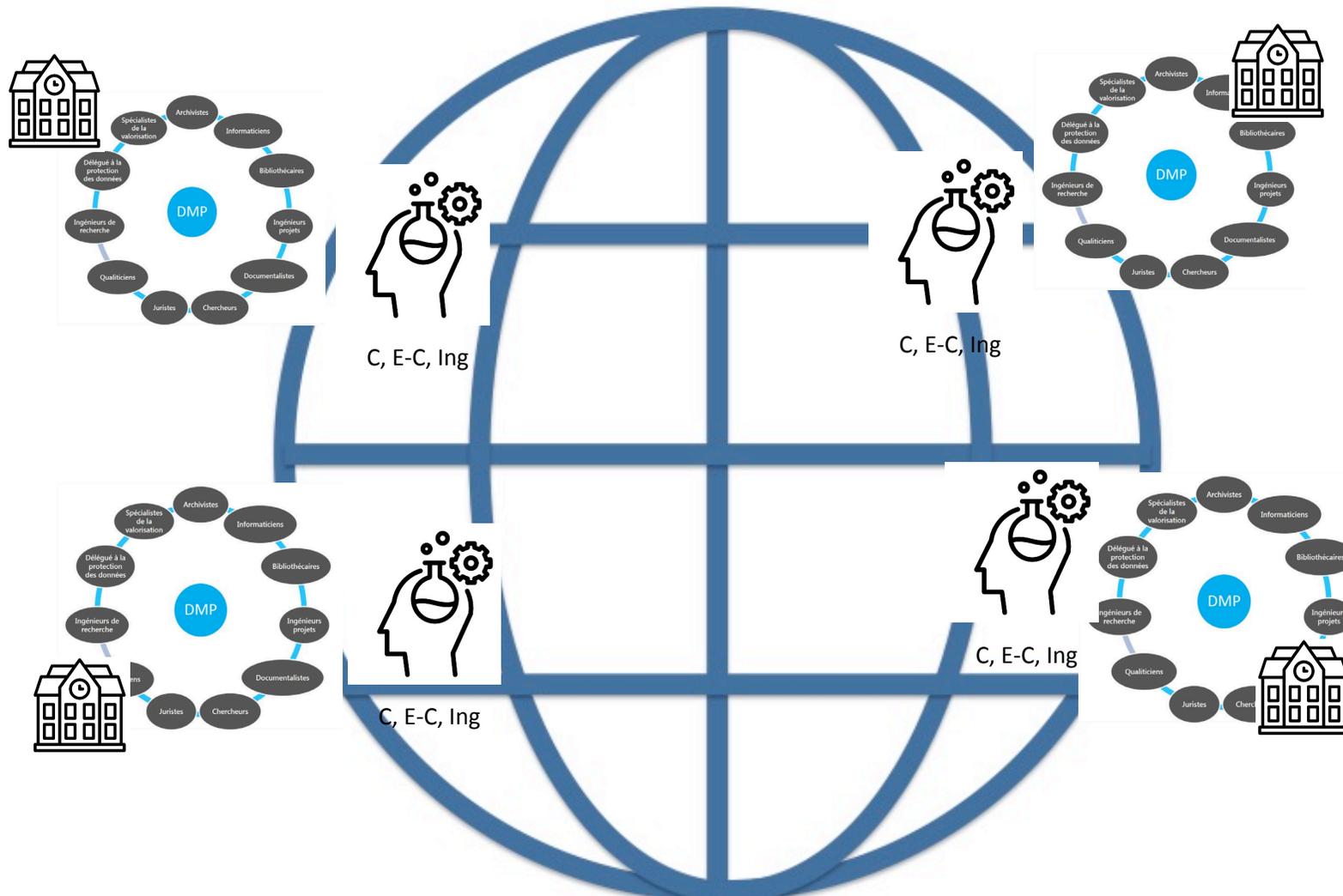
Situation de gestion des données de la recherche (laboratoire)



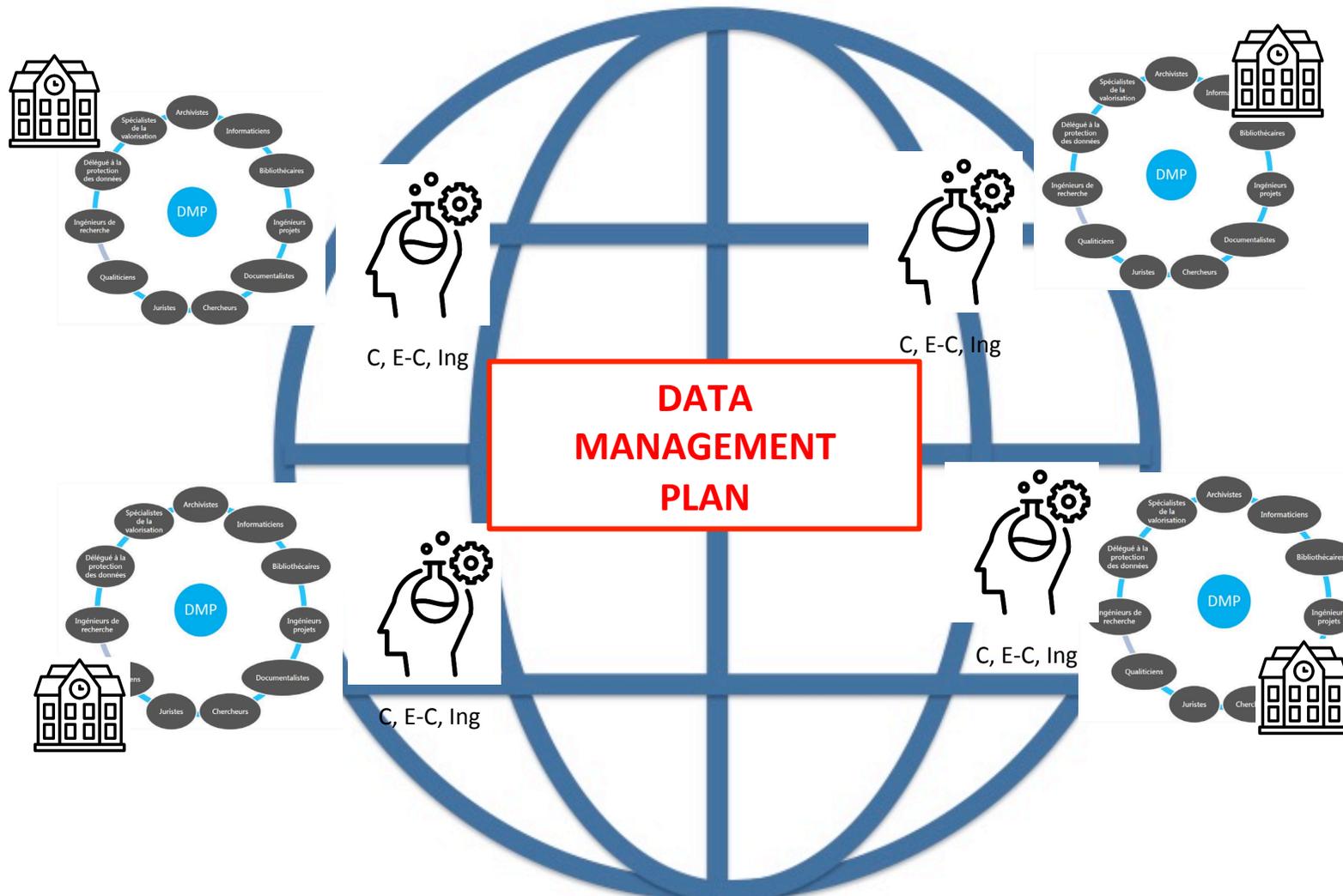
Situation de gestion des données de la recherche (laboratoire)



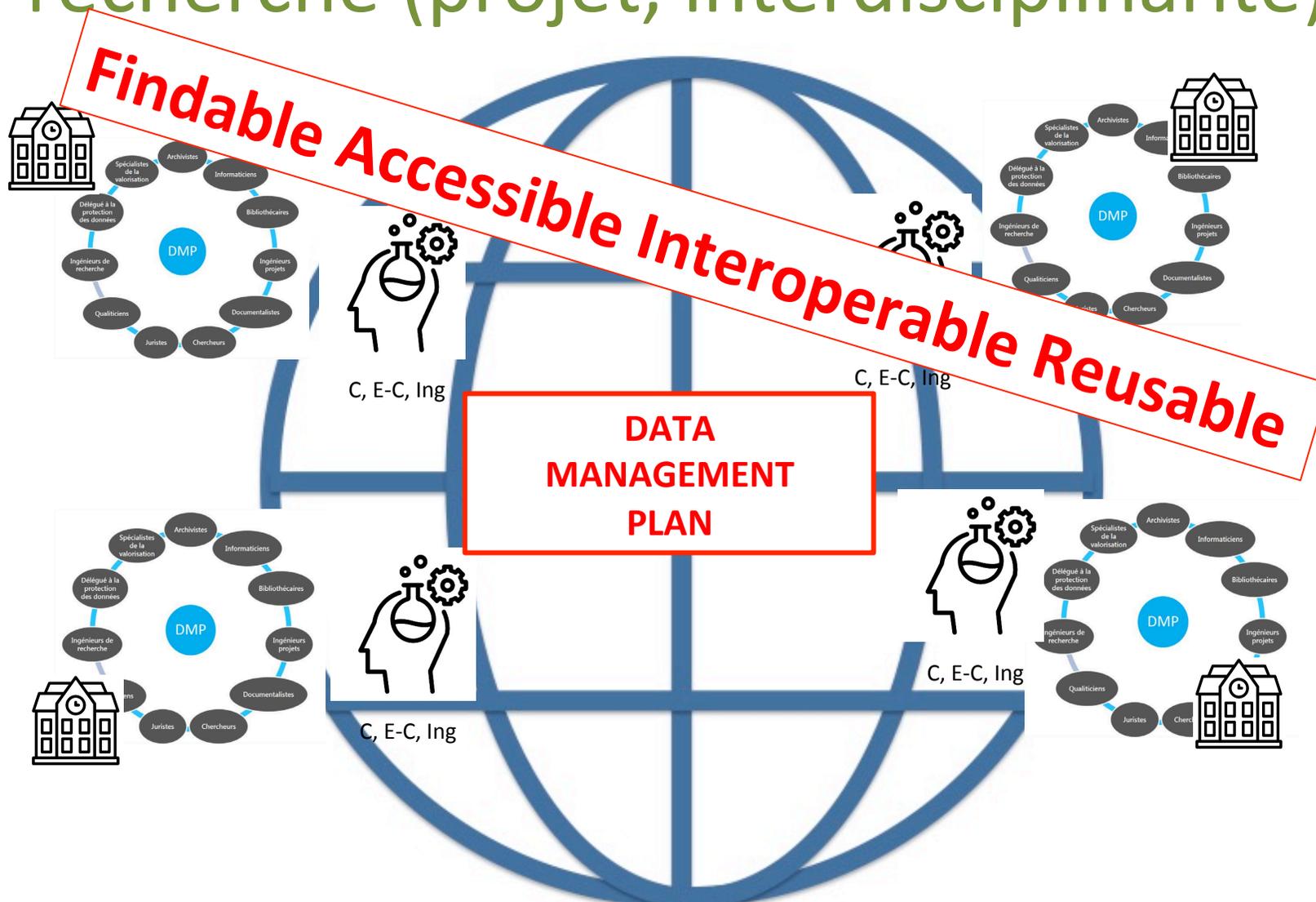
Situation de gestion des données de la recherche (projet)



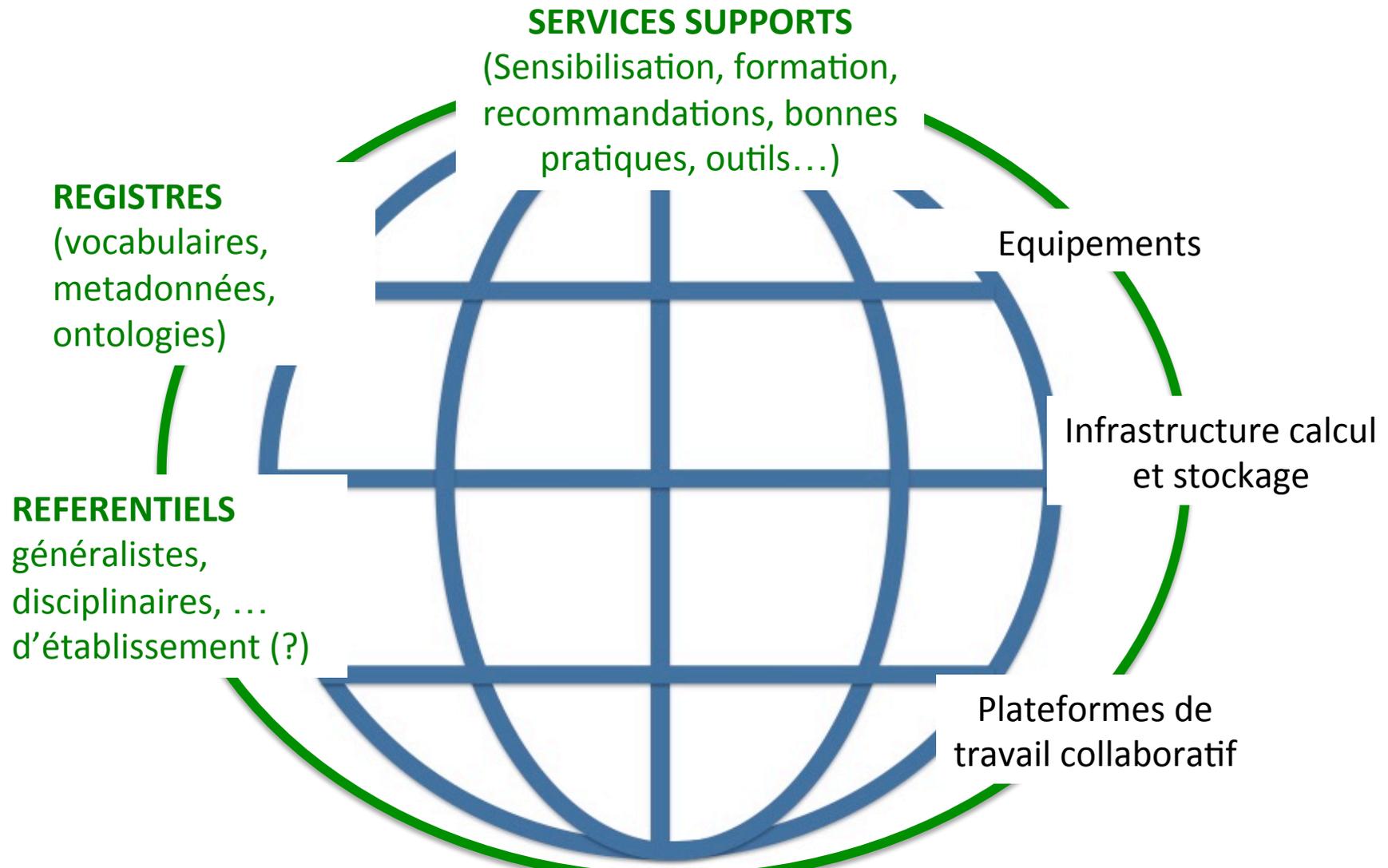
Situation de gestion des données de la recherche (projet, interdisciplinarité)



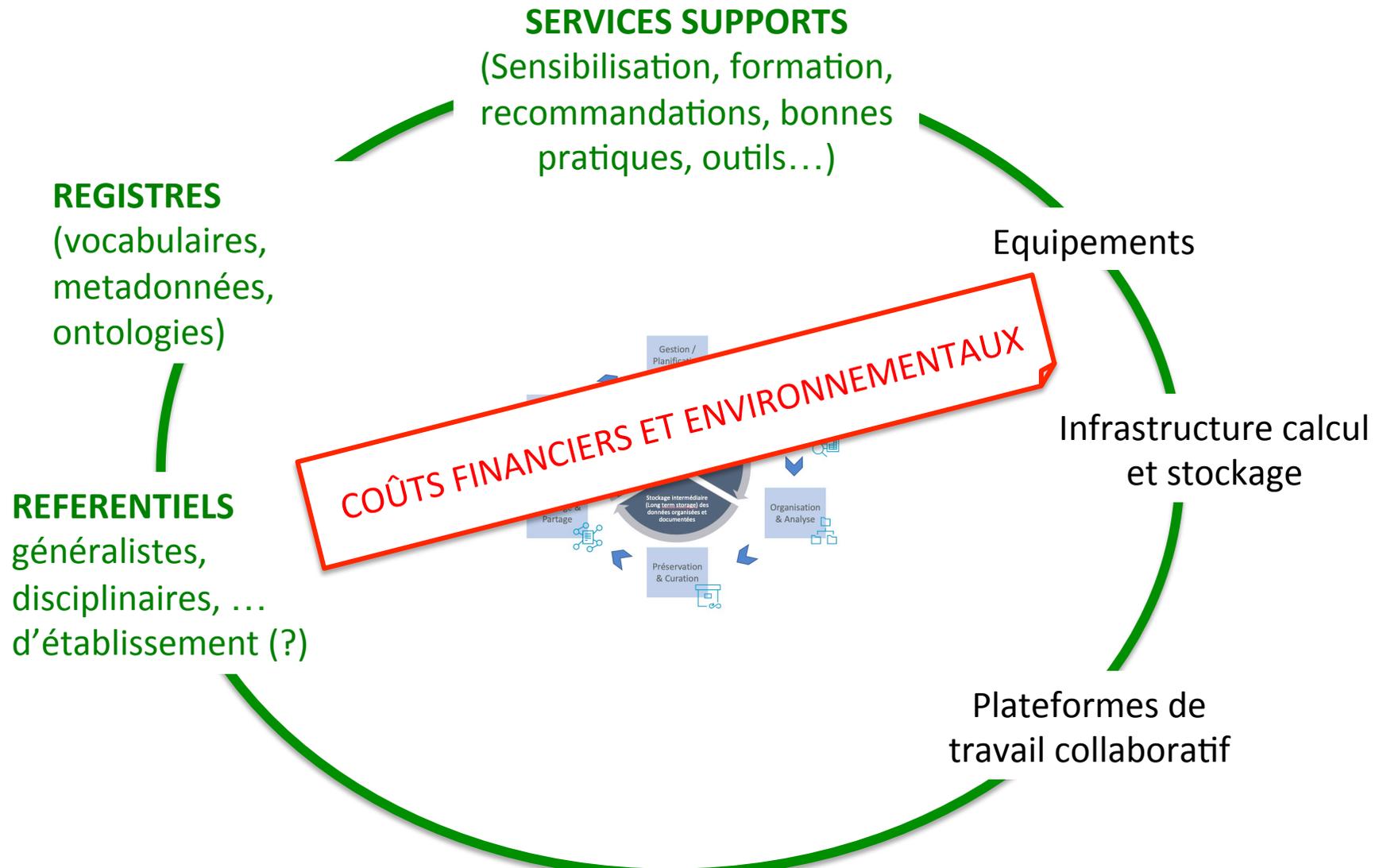
Situation de gestion des données de la recherche (projet, interdisciplinarité)



Besoins d'environnement(s) support... de confiance, de qualité, pérenne, ...**et interconnectés**



Besoins d'environnement(s) support



Conclusion

- Il faut BEAUCOUP de moyens : compétences, accompagnements, supports, infrastructures
- Chaque communauté scientifique est plus ou moins avancée et organisée
- Même disparité au niveau des établissements

Comment gagner en efficacité ? en qualité ?
POUR TOUTES LES COMMUNAUTÉS SCIENTIFIQUES
tout en réduisant nos coûts ?

Partie II

Les enjeux scientifiques et des établissements

Enjeux pour



C, E-C, Ing



Etudiant.e.s



Etablissements, organismes

- + Aller vers de la Data Driven Science
- + Disposer d'un écosystème
- + Initier de nouvelles recherches (inter)disciplinaires
- + Valoriser les résultats
- + ...

- + Enjeu de visibilité des recherches (pôles, interdisciplinarité)
- + Offrir des services de qualité, adaptés aux disciplines/communautés
- + Approche responsable des points de vue éthique et environnement
- + ...

Enjeux scientifiques

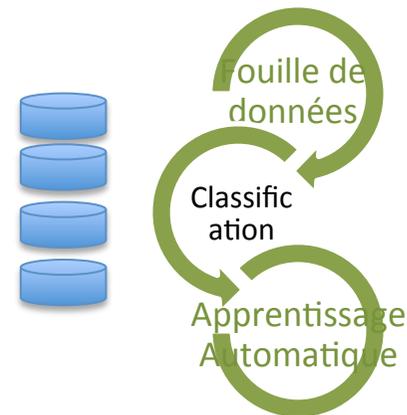
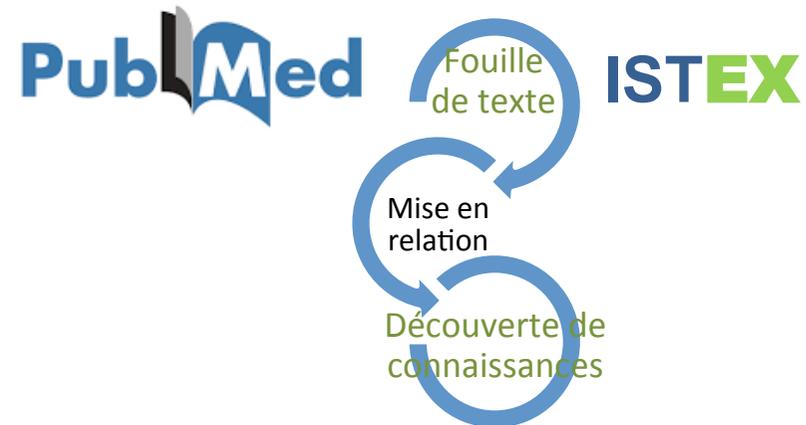


C, E-C, Ing



Etudiant.e.s

- + Aller vers de la Data Driven Science
 - + Changement des les pratiques des chercheurs
- + Disposer d'un écosystème
- + Initier de nouvelles recherches (inter)disciplinaires
- + Valoriser les résultats
- + ...



- Recherche de régularités passées pour interpréter de nouvelles données
- Alternatives à la modélisation mathématique ou à la simulation
- Anticiper des comportements, des choix ...

Enjeux scientifiques

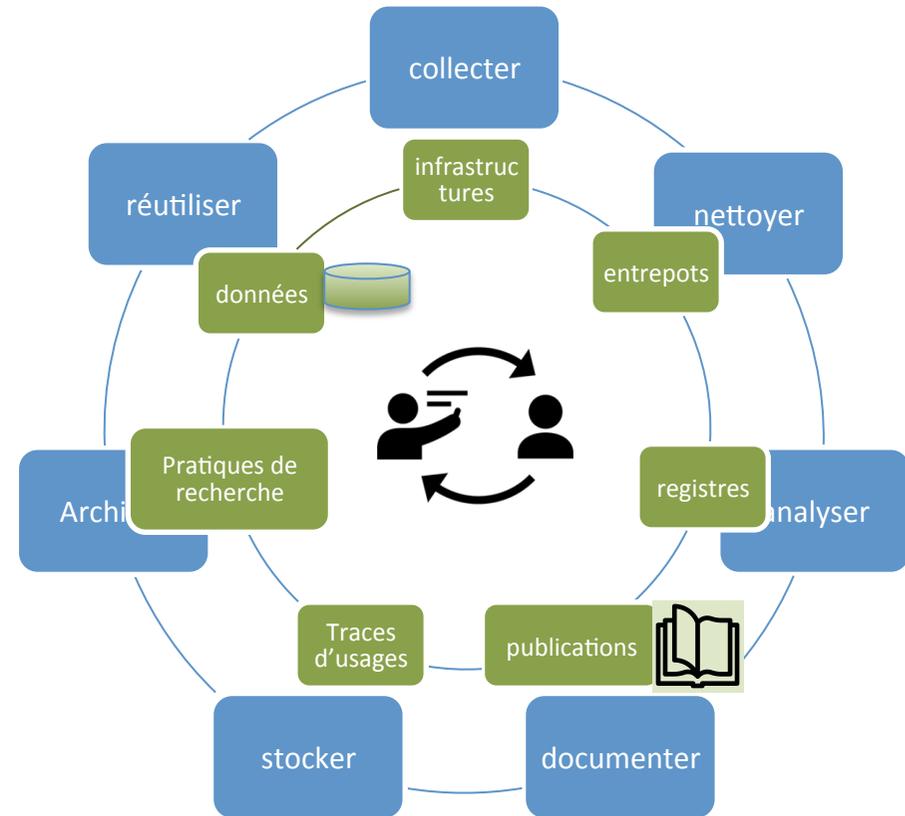


C, E-C, Ing



Etudiant.e.s

- + Aller vers de la Data Driven Science
- + Disposer d'un écosystème
 - + Partager des données, des expériences et en trouver beaucoup d'autres
 - + Mieux connaître l'état des lieux et les initiatives
- + Initier de nouvelles recherches (inter)disciplinaires
- + Valoriser les résultats
- + ...



Écosystème par domaine :



Ecosystème de site :



Enjeux scientifiques pour

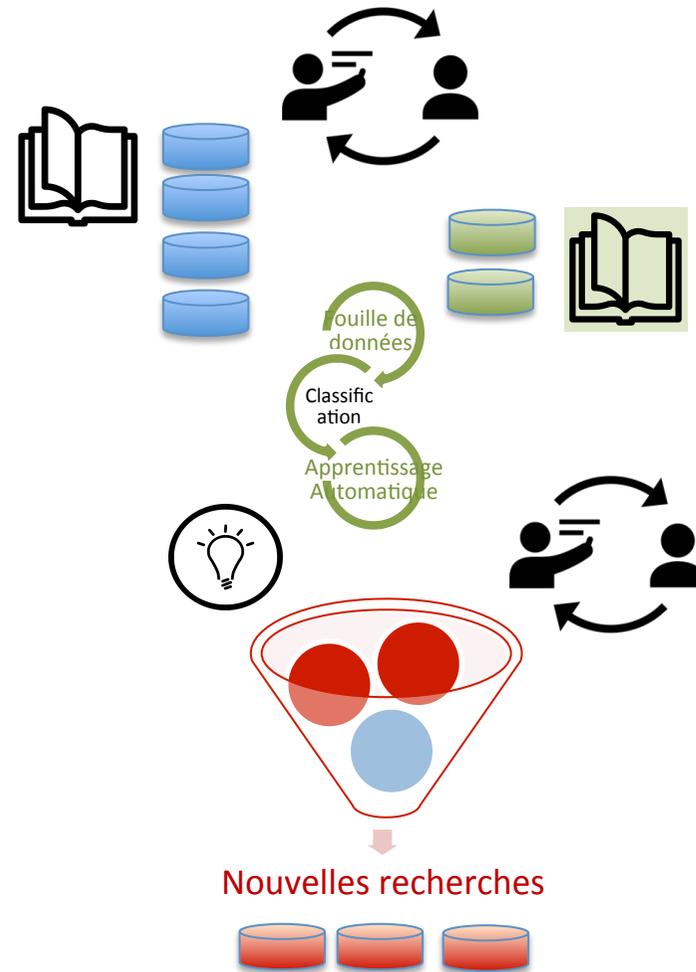


C, E-C, Ing



Etudiant.e.s

- + Aller vers de la Data Driven Science
- + Disposer d'un écosystème
- + Initier de nouvelles recherches (inter)disciplinaires
 - + Se connaître (jusqu'à l'international)
 - + Etre efficace
- + Valoriser les résultats
- + ...



Enjeux scientifiques



C, E-C, Ing



Etudiant.e.s

- + Aller vers de la Data Driven Science
- + Disposer d'un écosystème
- + Initier de nouvelles recherches (inter)disciplinaires
- + Valoriser les résultats
 - + Porter une visibilité disciplinaire/communautaire
 - + Aller vers une reconnaissance (acteurs de la SO, évaluation)
- + ...

Rendre visibles et accessibles ses données et ses publications c'est

- Amplifier leur visibilité
- Circulation des métadonnées des portails vers des registres sans intervention du chercheur



- Pouvoir reproduire les expériences des autres chercheurs
- Mieux comparer les contributions
- Gage de qualité des recherches

Enjeux scientifiques

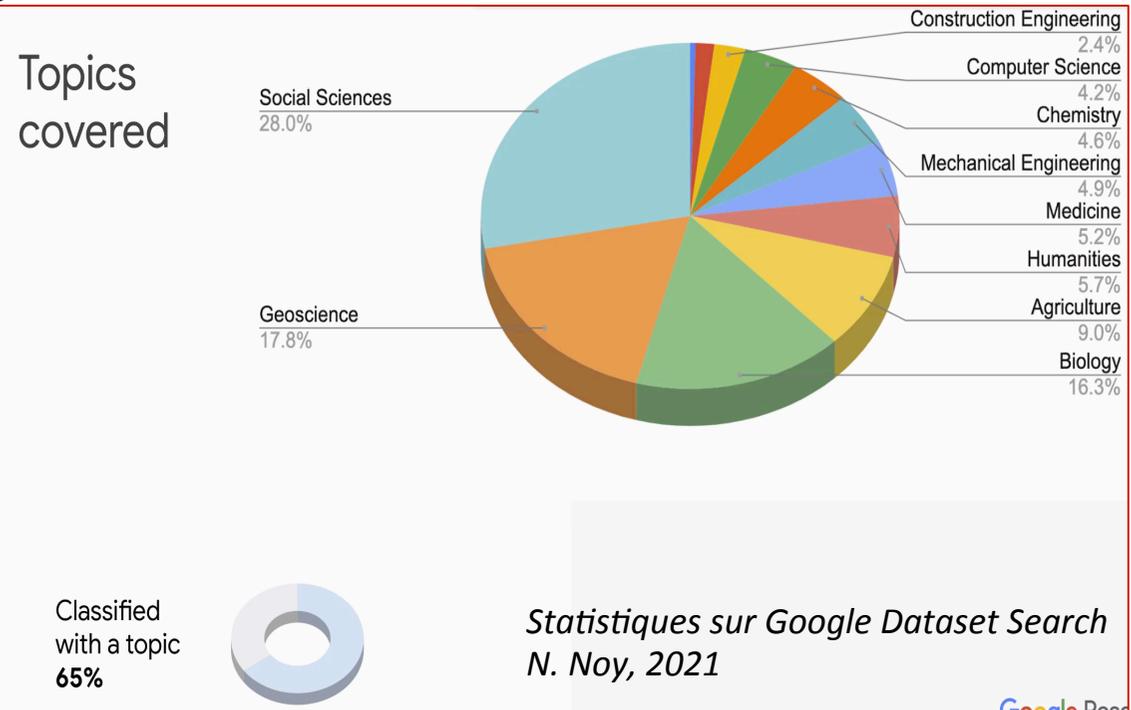


C, E-C, Ing



Etudiant.e.s

- + Aller vers de la Data Driven Science
- + Disposer d'un écosystème
- + Initier de nouvelles recherches (inter)disciplinaires
- + Valoriser les résultats
 - + Porter une visibilité disciplinaire/collaborative
 - + Aller vers une reconnaissance (acte d'évaluation)
- + ...



Enjeux pour



C, E-C, Ing



Etudiant.e.s

- + Porter une visibilité disciplinaire/communautaire simple
- + Partager des données, des expériences et en trouver beaucoup d'autres
- + Aller vers une reconnaissance (SO, évaluation)
- + Se connaître (jusqu'à l'international)
- + Mieux connaître l'état des lieux et les initiatives
- + Initier de nouvelles recherches interdisciplinaires
- + Disposer d'un écosystème
- + Etre efficacité
- + ...



Etablissements, organismes

- + Enjeu de visibilité des recherches (pôles, interdisciplinarité)
 - + Image positive, cohérente et coordonnée
 - + Réduire les disparités entre communautés scientifiques (locales et nationales)
- + Offrir des services de qualité, adaptés aux disciplines/communautés
 - + Faciliter les activités (individuelles, projets, communautés) des chercheurs
 - + Rationaliser des moyens, des services, des compétences
 - + Réduire les coûts
- + Approche responsable des points de vue éthique et environnement
- + ...

Partie III

Réflexions et actions du CéSO vers le partage et l'ouverture



- Le CéSO et ses missions
- Organisation de la réflexion et de la coordination
- Premières initiatives, actions et incitations

Le CéSO



*Accélérons l'adoption
de recommandations,
de standards, d'outils
en partageant nos
besoins, nos pratiques,
nos compétences dans
un cadre coordonné.*

- Validé par le Conseil des Membres de l'UFTMiP en Juin 2020
- Lancé en Septembre 2020 dans un continuum d'actions engagées par certains établissements et disciplines

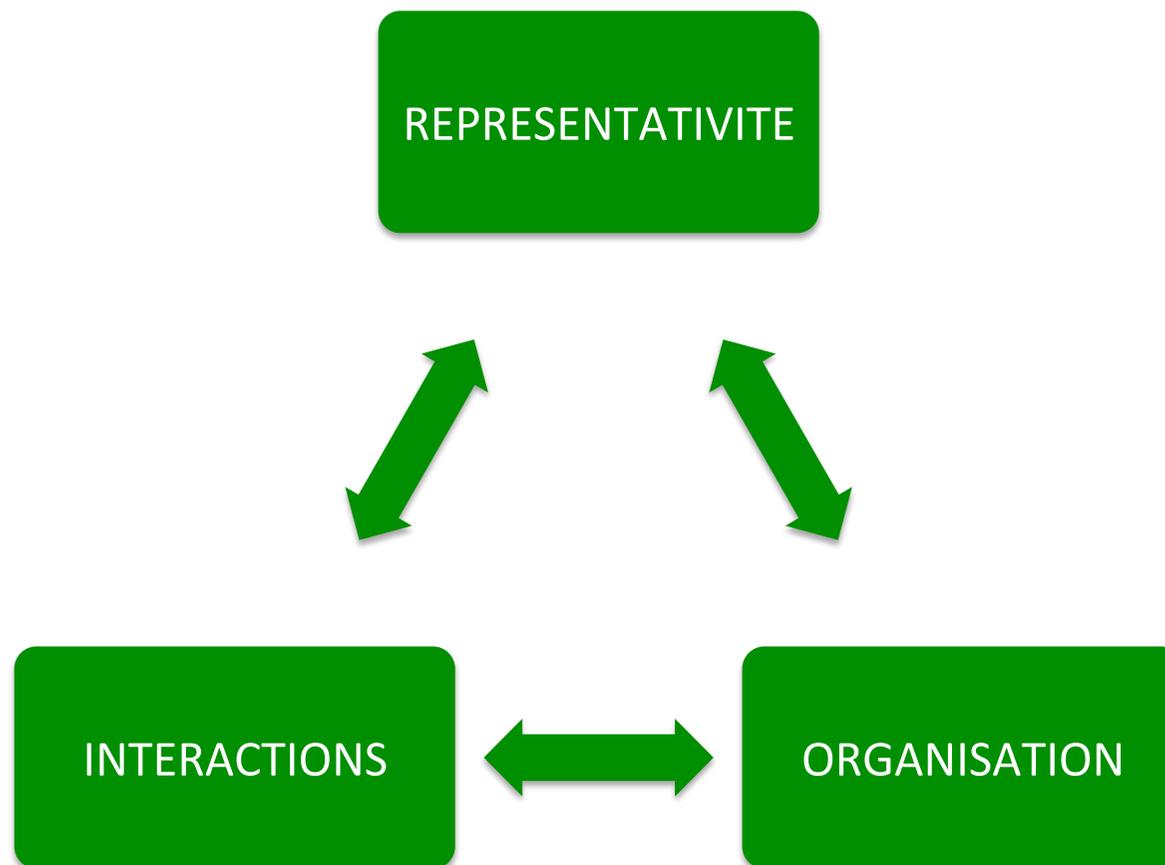
Les missions du CéSO

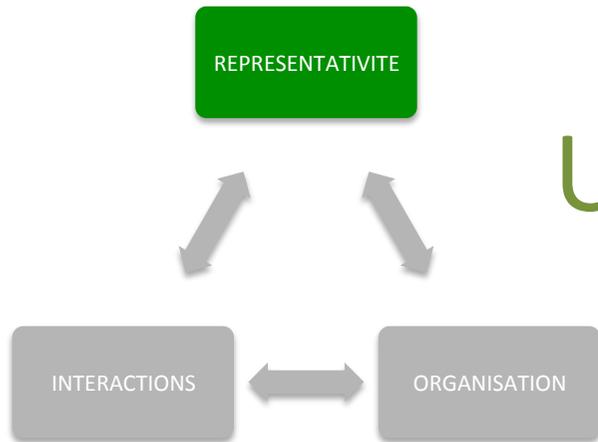
- **COORDONNER** une campagne de sensibilisation/formation à la Science Ouverte avec le réseau des bibliothèques
- **PROPOSER** une « politique de site » de gestion des données de la recherche (PGD) et ses déclinaisons (pôles/établissement/ laboratoires) avec l'aide des différents groupes de travail (existants ou à définir)
- **IDENTIFIER** des référentiels de données conformes aux préconisations nationales et européennes (FAIRisation), liés à des archives ouvertes (données de la recherche, publication, ...)
- **PRECONISER** une urbanisation des services autour de ces entrepôts de données
- **COORDONNER, SUSCITER, STIMULER** des réponses aux appels à projets nationaux et européens pour assurer la présence du site sur les AAP les plus importants ... en collaboration avec l'alliance académique



<https://datanoos.univ-toulouse.fr/>

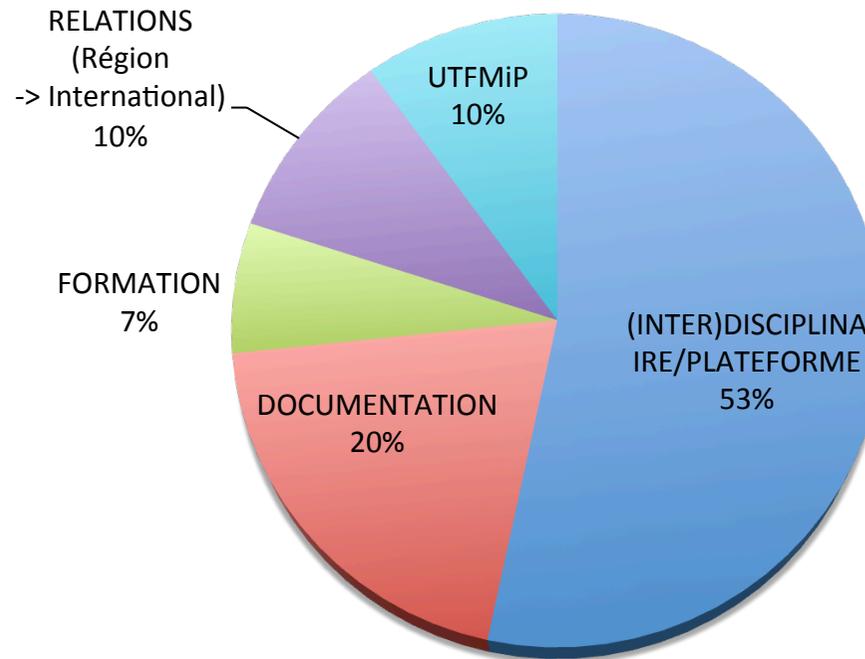
Organisation de la réflexion et de la coordination

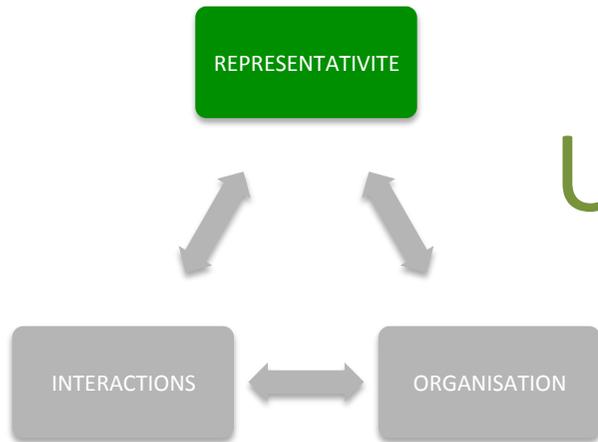




Une représentativité large

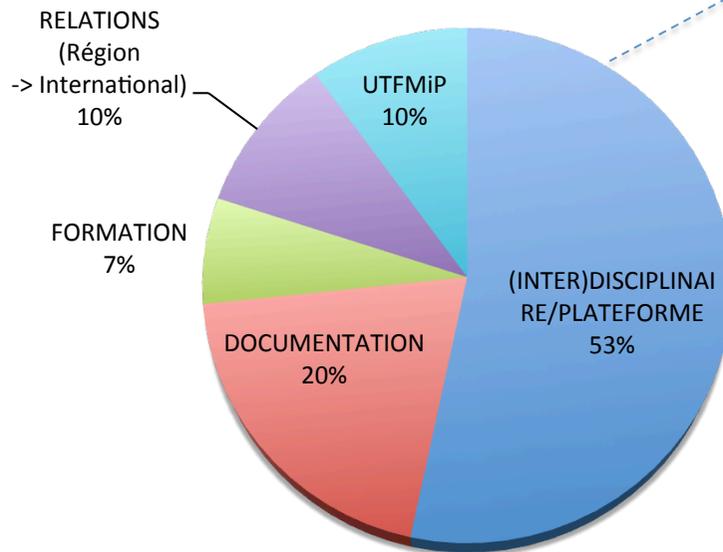
Composition du CéSO
(30 membres)





Une représentativité large

Composition du CéSO
(30 membres)



PÔLES	STRUCTURES	PLATEFORMES
H-SHS	MSHS-T	PUD-T PROGEDO
DSPEG	TSE Lab Exp	
UPEE	OMP	DATA TERRA
BABS	IFERISS, INSERM	GENOTOUL
MST2I	Fédération FERMAT	CALMIP
SDM	Institut de Chimie	
Transversal	ANITI	Alliance dataNooS

REPRESENTATIVITE



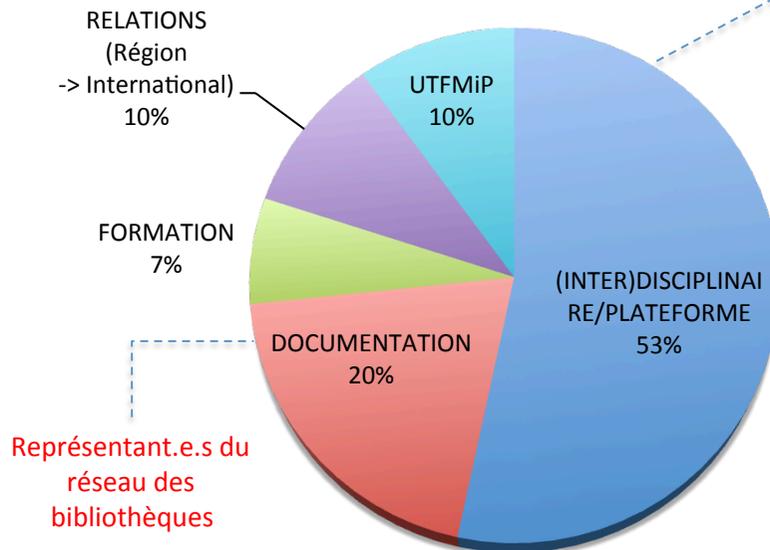
INTERACTIONS

ORGANISATION

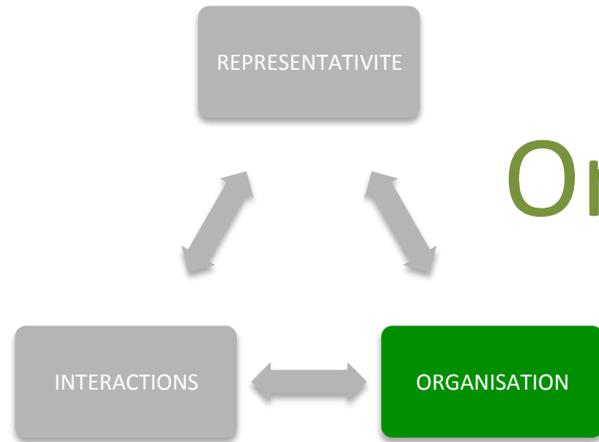
Une représentativité large

Représentant.e.s du
réseau CAPITOUL

Composition du CéSO (30 membres)



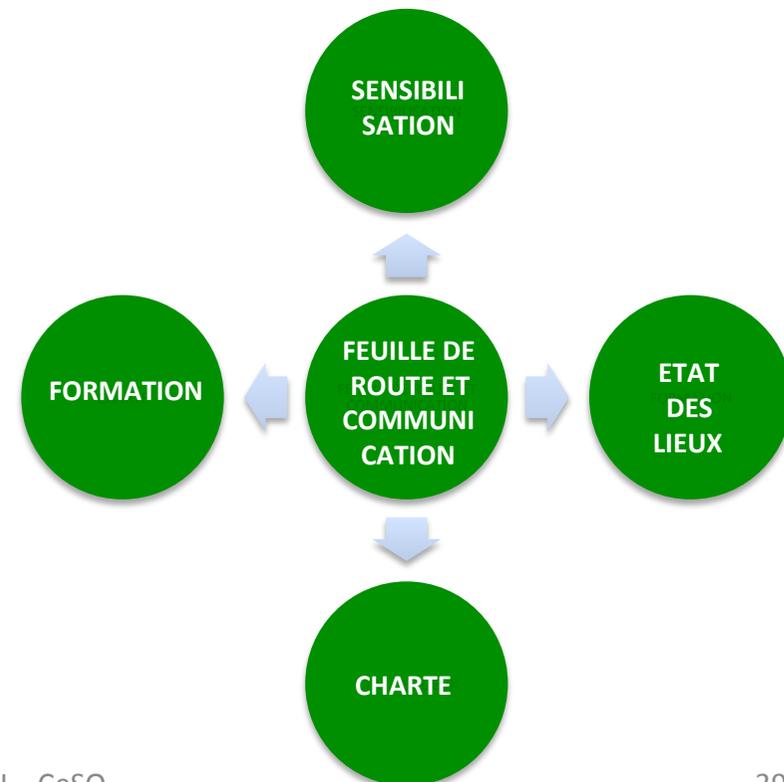
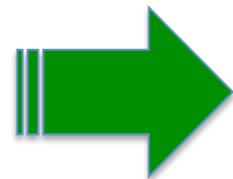
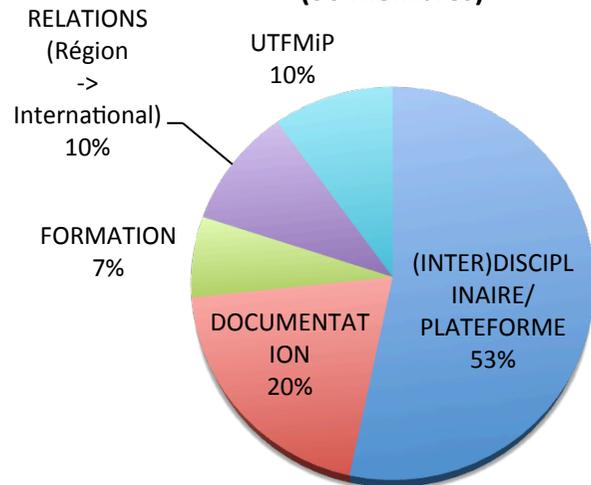
PÔLES	STRUCTURES	PLATEFORMES
H-SHS	MSHS-T	PUD-T PROGEDO
DSPEG	TSE Lab Exp	
UPEE	OMP	DATA TERRA
BABS	IFERISS, INSERM	GENOTOUL
MST2I	Fédération FERMAT	CALMIP
SDM	Institut de Chimie	
Transversal	ANITI	Alliance dataNooS

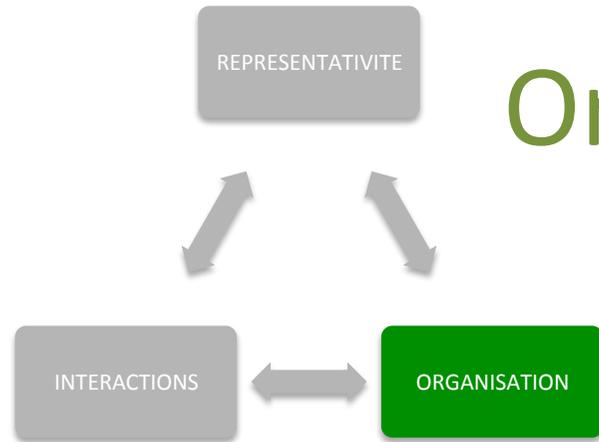


Organisation participative

Groupes de travail

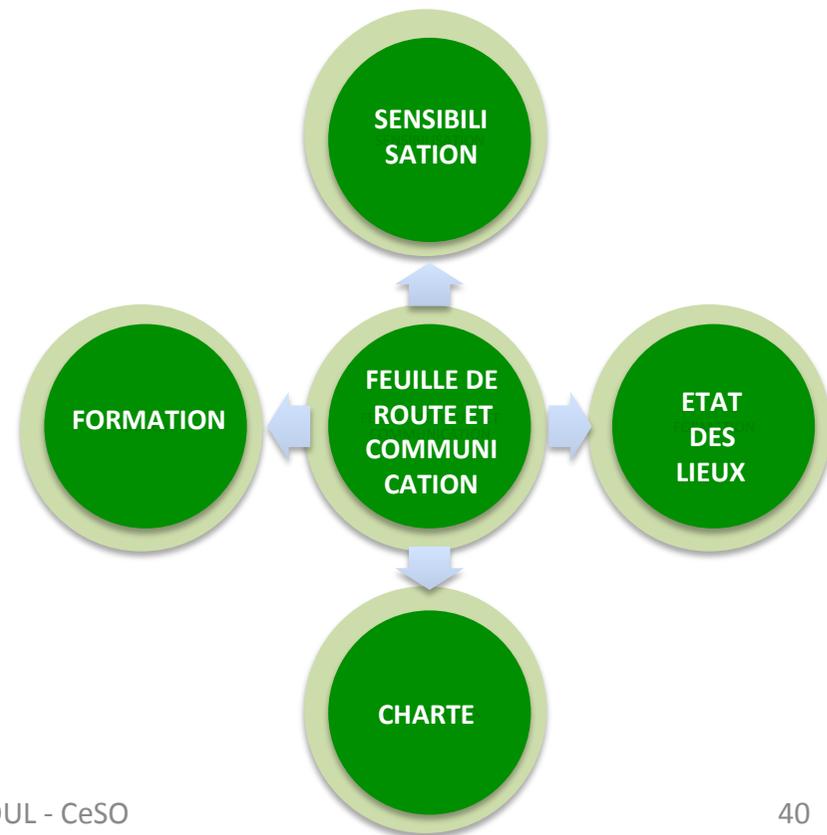
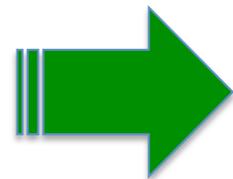
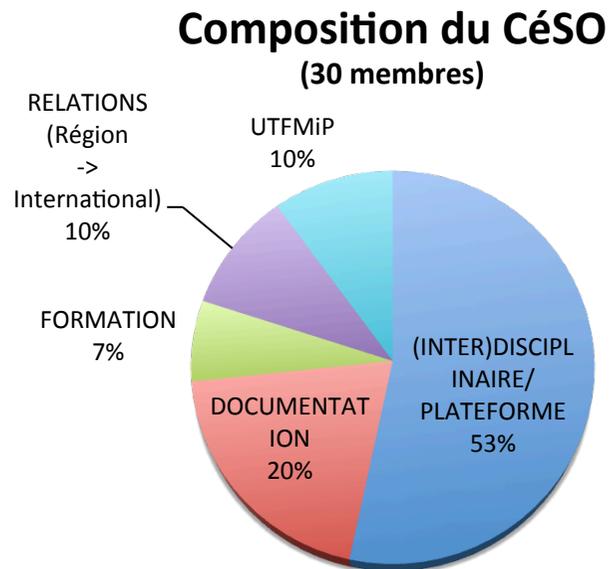
**Composition du CéSO
(30 membres)**





Organisation participative et élargie

Groupes de travail



REPRESENTATIVITE



Interactions multi-échelles

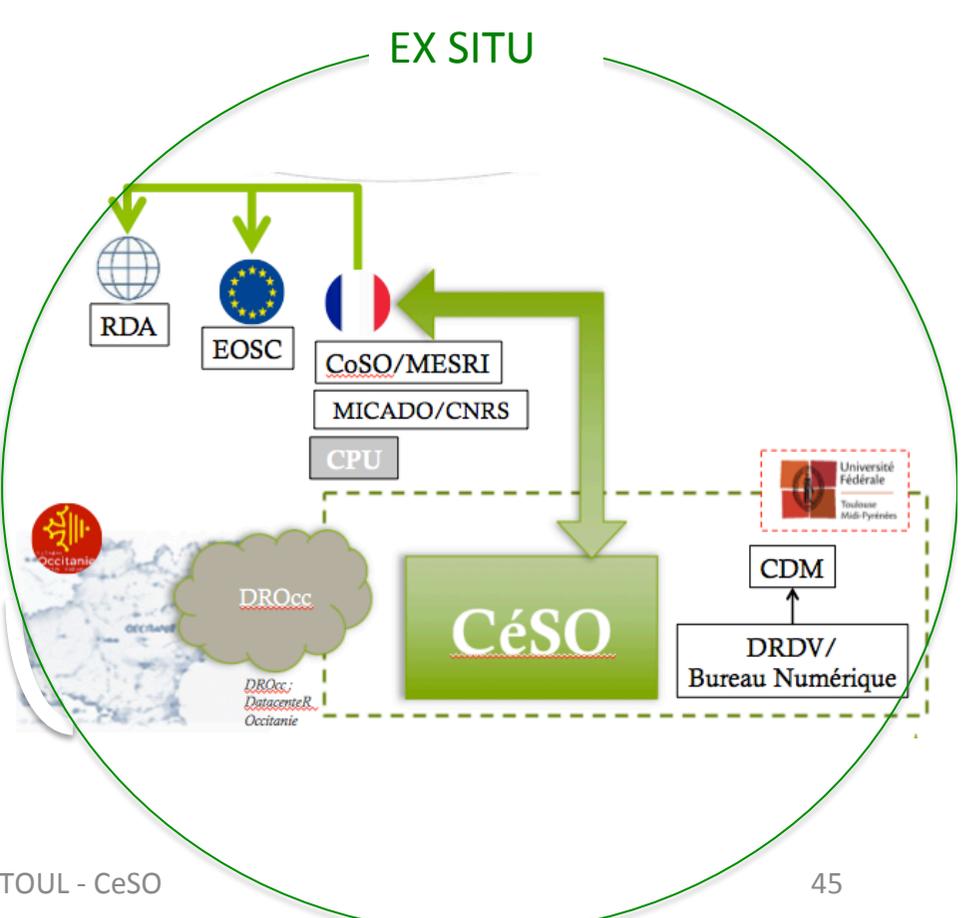
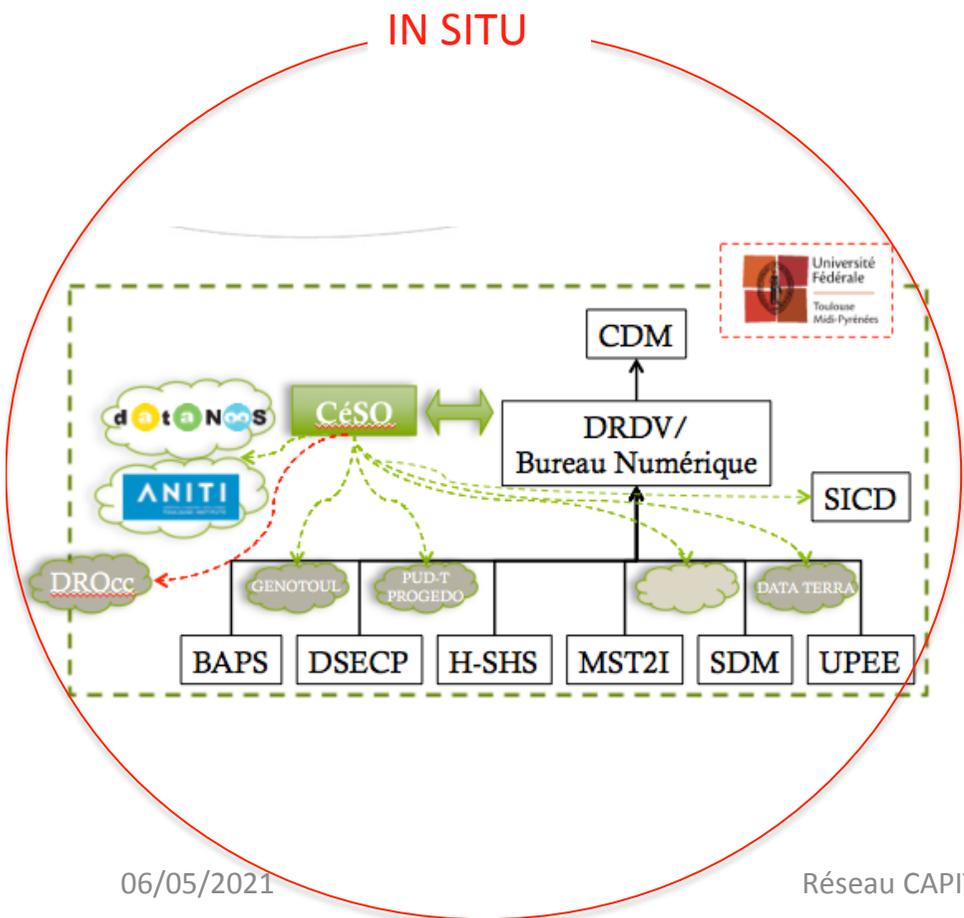
INTERACTIONS

ORGANISATION



IN SITU

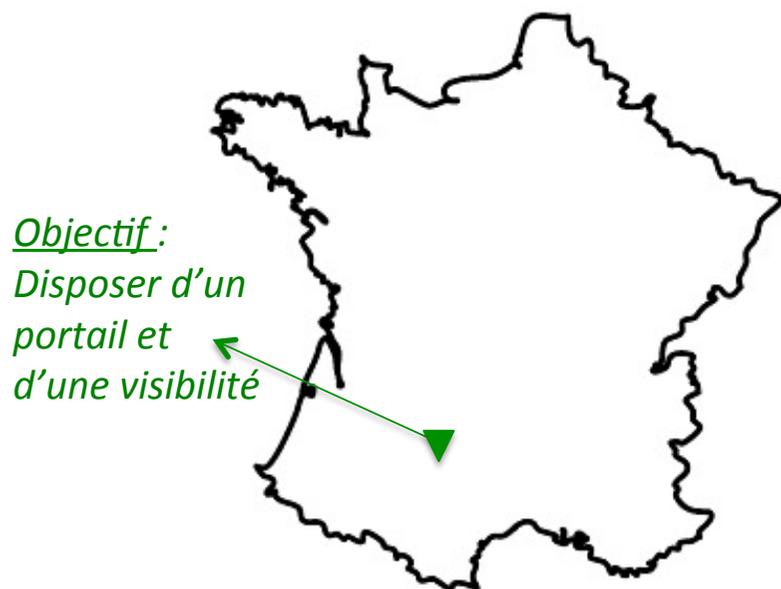
EX SITU



Premières initiatives coordonnées

« 2021 : Année de la valorisation des données de la recherche toulousaine et la Science Ouverte »

UNE PRIORITE :
INITIER UN ETAT DES LIEUX
SUR LES DONNEES DE LA RECHERCHE



UNE URGENCE :
(IN)FORMER SUR LA CONSTRUCTION
DES DMP



Objectif: Garantir la qualité des données et la gestion de leur cycle de vie

Actions en cours et Incitations

- Événements déjà planifiés en 2021 (Love Data Week, OA Week, ...) : ouverture à d'autres établissements
- Adaptation de formations et événements à de nouvelles thématiques disciplinaires
- Catalogue partagé des formations existantes
- Recueil des besoins auprès des chercheurs et partage de bonnes pratiques so
- Etat des lieux : cartographie des rôles/référents/ correspondants, des données, des infrastructures, ...

Echanges locaux

Relations directes institutionnelles

- Les établissements du sites (Présidents, VP Recherche, VP Numérique, Pôles, Stratégie scientifique partagée)

Autres relations (à développer) :

- Référents Science Ouverte, HAL et Données par établissement
- Réseau CAPITOUL : Rôle à jouer de relais des recommandations, besoins et supports techniques ?
- CéSO = Lieu d'échange et de mise en réseau coordonnée

Echanges locaux et (inter)nationaux

Relations directes institutionnelles

- Les établissements du sites (Présidents, VP Recherche, VP Numérique, Pôles, Stratégique scientifique partagée)

Relations (par l'intermédiaires des membres locaux du CéSO) avec

- la mission « Calcul, données ouvertes » du CNRS
- la conférence des Présidents d'Universités (CPU)
- les collèges du Comité National pour la SO (CoSO) : collège Publication, contact Collège Données de la Recherche)
- certains groupes de la Research Data Alliance (RDA) (inter)nationale
- des communautés nationales bien avancés (SHS, Environnement, Science du Vivant) et d'autres à co-construire

Partie IV

Conclusion et Perspectives

Conclusion

Sensibiliser les personnels, établissements et l'UFTMiP de la nécessité de se coordonner pour le partage et l'ouverture des données et des pratiques dans un esprit :

- trans-, inter- et disciplinaire
- inclusif
- éthique
- d'équilibre

Perspectives

Axes de travail de collaboration CéSO / CAPITOUL

- Effort constant de communication, sensibilisation, formation
- Mise en réseau des compétences pour soutenir l'ouverture de la Science
(référents SO/RD/HAL, DPO, Data stewardship, ...)
- Appel à retour d'expériences
 - « Success Stories » pour convaincre et valoriser
 - « Failed Stories » pour orienter la réflexion et les actions prioritaires à mener
- Promouvoir le partage/l'ouverture de données et de pratiques

Parlons-en ou Contactez-nous !

aussenac@irit.fr sibilla@irit.fr et pascal.gaillard@univ-tlse2.fr



<https://datanoos.univ-toulouse.fr/>

- dataNoosphère :
 - un écosystème autour des données
 - Faire de la science ouverte un sujet de recherche
- Agora : Un groupe d'animation depuis 2018
 - Faire émerger des besoins
 - Identifier l'existant
 - Favoriser des échanges
- Un groupe de travail et d'expérimentation
 - Cas d'usage :
 - Données environnementales
 - Données en aéronautique
 - Prototypes de composants d'une plateforme pour la science ouverte