



HAL
open science

Étude de faisabilité d'un service générique d'accueil et de diffusion des données simples : ambitions du service et scénarios de mise en œuvre

Collège Données De La Recherche

► To cite this version:

Collège Données De La Recherche. Étude de faisabilité d'un service générique d'accueil et de diffusion des données simples : ambitions du service et scénarios de mise en œuvre. [Rapport de recherche] Comité pour la science ouverte. 2021, 42 p. hal-03594196

HAL Id: hal-03594196

<https://hal-lara.archives-ouvertes.fr/hal-03594196v1>

Submitted on 2 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude de faisabilité d'un service générique
d'accueil et de diffusion des données simples

Synthèse de la phase 3 : ambitions du service et scénarios de mise en œuvre

Comité pour la Science Ouverte / MESRI

Éléments de contexte

Les données de la recherche

« Le terme « donnée de recherche » **semble [...] être une notion complexe, que ni les producteurs de données, ni les institutions qui en demandent l'ouverture ne parviennent à définir sous une forme élémentaire et unanimement reconnue.** » (1)

On peut retenir cette définition plus extensive, qui a le mérite d'englober les pratiques de nombreuses disciplines :

Le terme de données de la recherche désigne les données sous forme de **faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences** sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé. Les données peuvent être numériques, descriptives, visuelles [...]. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou média.

« *Research data means data in the form of **facts, observations, images, computer program results, recordings, measurements or experiences** on which an argument, theory, test or hypothesis, or another research output is based. Data may be numerical, descriptive, visual [...]. It may be raw, cleaned or processed, and may be held in any format or media.* » (2)

(1) Rebouillat, Violaine. « Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs ». Thèse de doctorat. Paris: Conservatoire national des arts et métiers - CNAM, 2019. <https://tel.archives-ouvertes.fr/tel-02447653>.

(2) ANDS, « What Is Research Data », 2017. https://www.ands.org.au/_data/assets/pdf_file/0006/731823/Whatis-research-data.pdf

Ouverture des données de la recherche



Évolutions récentes du cadre réglementaire et institutionnel :

- La **loi pour une République numérique** pose le principe de l'ouverture des données de la recherche publique (2016)
- Le programme européen **Horizon 2020** rend obligatoire l'ouverture des données de recherche quand cela est possible (2017)
- Le **Plan national pour la science ouverte** prévoit de rendre obligatoire l'ouverture des données de recherche issues de programmes financés par appels à projets sur fonds publics, lorsque cette ouverture est possible (2018)
- L'**Agence Nationale pour la Recherche** rend obligatoire le Plan de Gestion des Données pour les projets financés (2019)

Données ne rentrant pas dans le périmètre :

- Données issues des activités de recherche en santé impliquant des personnes humaines (Loi Jardé)
- Données sensibles (secret industriel, mettant en danger écosystème, individus...)
- Données personnelles non anonymisées ou pseudo-anonymisées

Diffusion des données de la recherche

Plateformes de publication

- Entrepôts **généralistes** : Zenodo, Figshare, Dryad ;
- Entrepôts **disciplinaires** : Nakala, portail des données marines ;
- Entrepôts **institutionnels** : Data INRAE, DataSuds (IRD) ;
- Portails et bases de données des **dispositifs d'observation** et **d'expérimentation** : Pôles de données des sciences de la Terre, de la biodiversité ;
- Entrepôts des **éditeurs** : Elsevier, Springer...

Données simples et risques associés

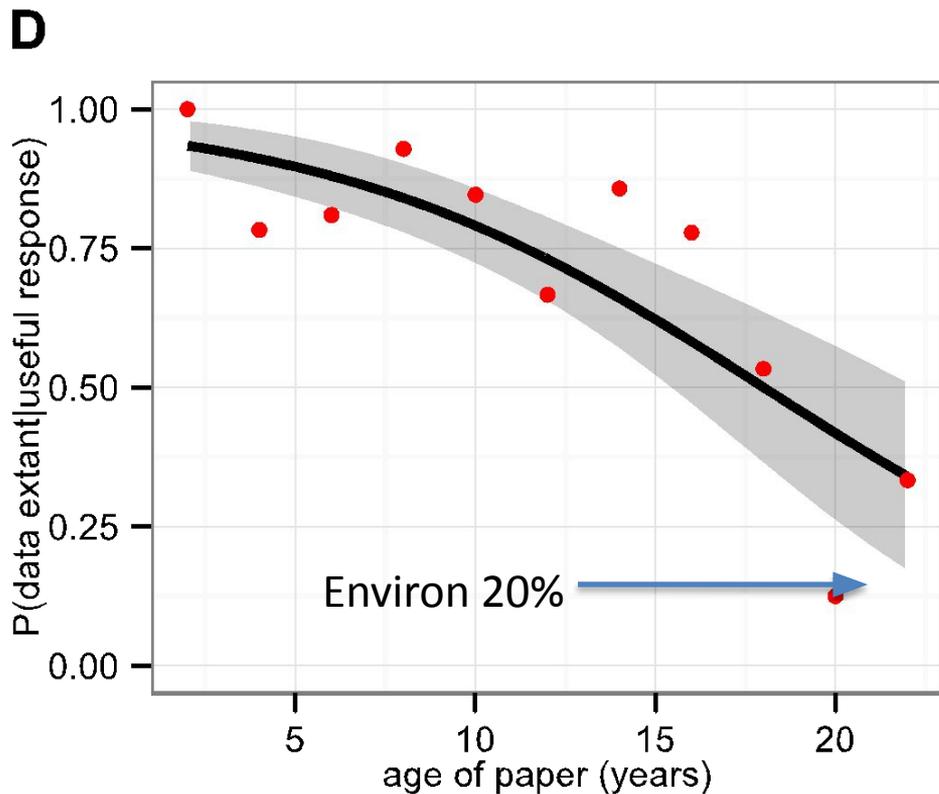
Exemples de données simples :

- **Supplementary material d'une publication**
- Mais également : **données de tableaux, données des graphiques** apparaissant dans une publication scientifique ;
- **Données de la longue traîne** qui ne font pas l'objet d'une préservation, d'une diffusion par l'intermédiaire de services en ligne dédiés.

Principaux risques associés :

- Perte des données
- Captation des données par les éditeurs de revue
- Recours à des entrepôts génériques sans curation, non modérés ou non certifiés
- Faible réutilisabilité de données déposées pas ou peu documentées

“La bibliothèque des données brûle”



Current Biology 24, 94–97, January 6, 2014 ©2014 Elsevier Ltd All rights reserved <http://dx.doi.org/10.1016/j.cub.2013.11.014>

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines,^{1,2,*} Arienne Y.K. Albert,³ Rose L. Andrew,¹ Florence Débarre,^{1,4} Dan G. Bock,¹ Michelle T. Franklin,^{1,5} Kimberly J. Gilbert,¹ Jean-Sébastien Moore,^{1,6} Sébastien Renaut,¹ and Diana J. Rennison¹

¹Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

²Molecular Ecology Editorial Office, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

³Women's Health Research Institute, 4500 Oak Street, Vancouver, BC V6H 3N1, Canada

⁴Centre for Ecology & Conservation Biosciences, University of Exeter, Cornwall Campus, Tremough, Penryn TR10 9EZ, UK

⁵Institute for Sustainable Horticulture, Kwantlen Polytechnic University, 12666 72nd Avenue, Surrey, BC V3W 2M8, Canada

⁶Department of Biology, Université Laval, 1030 Avenue de la Médecine, Laval, QC G1V 0A6, Canada

sets (23%) were confirmed as extant. [Table 1](#) provides a breakdown of the data by year.

We used logistic regression to formally investigate the relationships between the age of the paper and (1) the probability that at least one e-mail appeared to work (i.e., did not generate an error message), (2) the conditional probability of a response given that at least one e-mail appeared to work, (3) the conditional probability of getting a response that indicated the status of the data (data lost, data exist but unwilling to share, or data shared) given that a response was received, and, finally, (4) the conditional probability that the data were extant (either “shared” or “exists but unwilling to share”) given that an informative response was received.

There was a negative relationship between the age of the paper and the probability of finding at least one apparently working e-mail either in the paper or by searching online (odds ratio [OR] = 0.93 [0.90–0.96, 95% confidence interval

Des enjeux pour les données de la recherche



Garantir la **souveraineté** et la **pérennité** des données de recherche grâce aux infrastructures numériques de l'ESRI



Assurer la **visibilité** et l'**accessibilité** des données de la recherche française au-delà de l'ESRI



Assurer l'**accès à un dépôt** pour toutes les données de recherche



S'intégrer à la construction d'un **écosystème** des données de recherche, aux échelles **européenne** (EOSC) et internationale

Un service qui répond aux besoins des chercheurs

“ Si on veut **inciter** les gens à déposer leurs données, ça doit être **simple**. Les chercheurs n'ont **pas assez de temps** pour faire leur cœur de métier, on veut bien faire ce qui a un **vrai sens scientifique** mais on veut que ce soit simple ”

“ S'il y a un entrepôt de données, il faut [des personnes] qui accompagnent et dialoguent [...], qui fassent **tourner le moteur**. ”

“ Les multi appartenances [institutionnelles] sont infernales à gérer. Un entrepôt national **neutre** peut permettre de régler certains problèmes. ”

Un service à construire pour les chercheurs

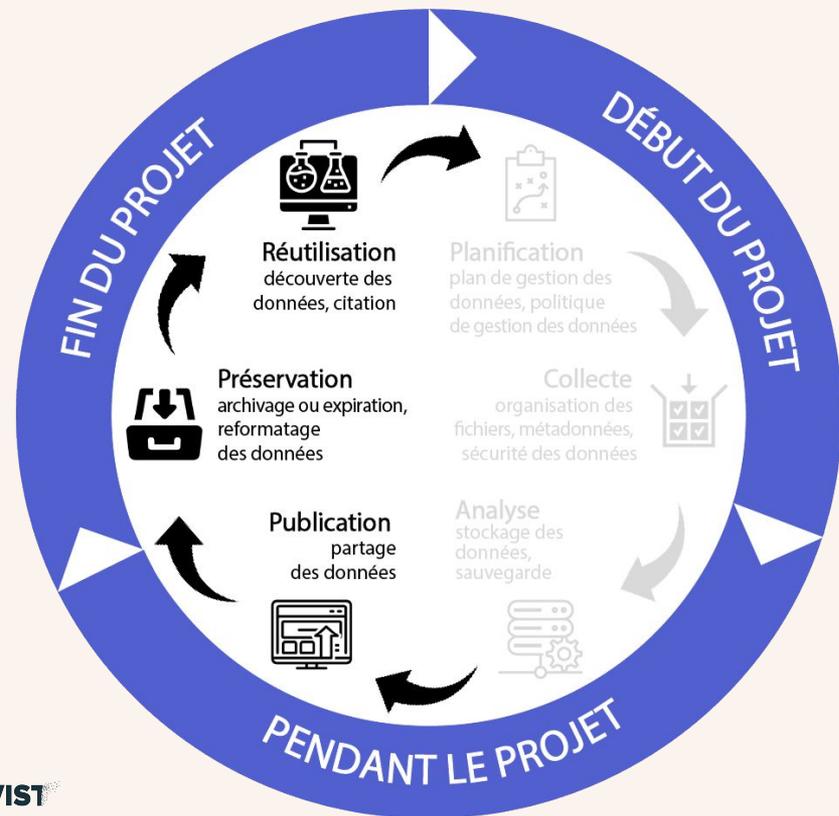
- Les chercheurs **producteurs de données** sont les premiers usagers d'un service d'accueil et de diffusion des données de la recherche. Nous décrivons le service en suivant leur parcours dans le dépôt d'un jeu de données pour **matérialiser les changements**, les services apportés et leurs impacts sur les établissements et les utilisateurs. Nous mettons l'accent sur l'accompagnement des déposants, selon **3 niveaux d'ambition d'accompagnement**.
- Nous adoptons ensuite le point de vue des **réutilisateurs de données**, pour décrire les fonctionnalités techniques d'un entrepôt de données généraliste selon **3 niveaux d'ambition technique**.
- Nous décrivons enfin, indépendamment des niveaux d'ambition, **3 scénarios** possibles pour la **mise en œuvre** d'un entrepôt de données généraliste, en adoptant le point de vue d'un établissement de la communauté ESRI.

Le service d'accueil et de diffusion des données dans le cycle de vie de la donnée

Publication et **préservation** pour une meilleure **réutilisation** des données.

Deux usages bien identifiés, qui seront décrits par deux parcours utilisateur :

- **Producteur de données** déposant ses données sur l'entrepôt ;
- **Réutilisateur** du point de vue des fonctionnalités qui lui sont proposées ;



Parcours producteur

Point de départ : Un chercheur doit déposer un jeu de données

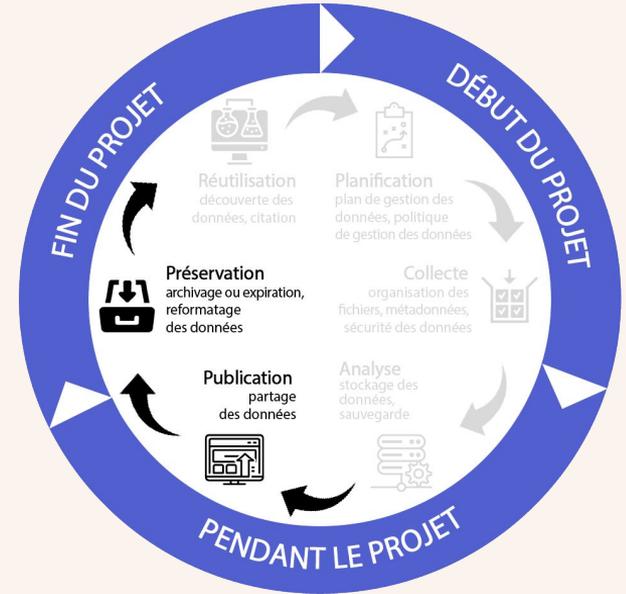
L'accompagnement des chercheurs : un facteur d'adhésion et de succès

- L'étude de recueil des besoins et des contraintes des usagers a fait ressortir l'importance, comme **facteur d'adhésion et de succès**, de l'**accompagnement** des chercheurs à la **gestion des données**.
- Le **dépôt des données** sur un entrepôt s'inscrit dans le cadre plus large de la gestion des données tout au long du cycle de vie d'un projet de recherche.

3 niveaux d'ambition dans l'accompagnement des producteurs de données pour le dépôt :

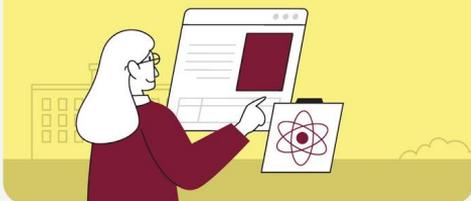
- * : assurer la possibilité de déposer, garantir la qualité minimale des dépôts ;
- ** : faciliter le dépôt, encourager l'amélioration des pratiques ;
- *** : proposer un accompagnement personnalisé à chaque étape du dépôt.

Les niveaux d'ambition ne sont pas monolithiques : chaque niveau d'ambition regroupe des éléments autonomes, qui peuvent être mis en œuvre indépendamment.



■ Producteur ■ Établissement ■ Opérateur national

JE DOIS PUBLIER MES DONNÉES ET OBTENIR UN DOI POUR FINALISER LA SOUMISSION DE MON ARTICLE.



JE DÉTERMINE LE OU LES ENTREPÔTS PERTINENTS POUR MES DONNÉES. JE PEUX BÉNÉFICIER DES RESSOURCES MISES EN LIGNE PAR LES SERVICES D'ACCOMPAGNEMENT.



JE PRÉPARE MES DONNÉES AU DÉPÔT. J'UTILISE LA DOCUMENTATION ET LES TUTORIELS MIS À DISPOSITION PAR LES SERVICES D'ACCOMPAGNEMENT. ILS PEUVENT RÉPONDRE À MES QUESTIONS.



JE DÉPOSE MES DONNÉES, JE BÉNÉFICIE D'UN SUPPORT SI JE RENCONTRE DES DIFFICULTÉS.



JE PEUX AMÉLIORER LA QUALITÉ DE MON DÉPÔT POUR FACILITER LA RÉUTILISATION DE MES DONNÉES (DONNÉES FAIR).

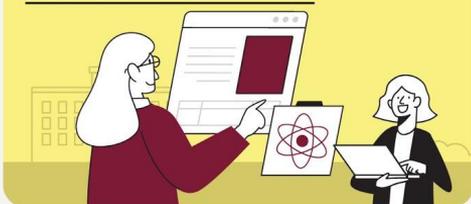


LES SERVICES D'ACCOMPAGNEMENT ASSURENT LA MODÉRATION ET LA CURATION DES DÉPÔTS, POUR GARANTIR LEUR CONFORMITÉ ET LEUR QUALITÉ, ET METTRE EN VALEUR LES DONNÉES.



■ Producteur ■ Établissement ■ Opérateur national

JE DOIS PUBLIER MES DONNÉES ET OBTENIR UN DOI POUR FINALISER LA SOUMISSION DE MON ARTICLE. LES SERVICES D'ACCOMPAGNEMENT M'AIDENT À PRÉPARER UN DÉPÔT CONFORME AU CADRE RÉGLEMENTAIRE.



JE DÉTERMINE LE OU LES ENTREPÔTS PERTINENTS POUR MES DONNÉES. JE PEUX BÉNÉFICIER DES CONSEILS DES SERVICES D'ACCOMPAGNEMENT.



JE PRÉPARE MES DONNÉES AU DÉPÔT. JE PEUX SUIVRE DES FORMATIONS EN LIGNE. LES SERVICES D'ACCOMPAGNEMENT PEUVENT RÉPONDRE À MES QUESTIONS.



JE DÉPOSE MES DONNÉES, JE BÉNÉFICIE D'UN SUPPORT SI JE RENCONTRE DES DIFFICULTÉS.



JE PEUX AMÉLIORER LA QUALITÉ DE MON DÉPÔT POUR FACILITER LA RÉUTILISATION DE MES DONNÉES (DONNÉES FAIR). JE PEUX SOLLICITER LES SERVICES D'ACCOMPAGNEMENT.

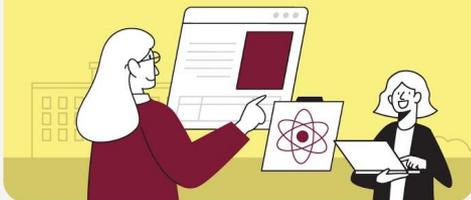


LES SERVICES D'ACCOMPAGNEMENT ASSURENT LA MODÉRATION ET LA CURATION DES DÉPÔTS, POUR GARANTIR LEUR CONFORMITÉ ET LEUR QUALITÉ, ET METTRE EN VALEUR LES DONNÉES.



■ Producteur ■ Établissement ■ Opérateur national

JE DOIS PUBLIER MES DONNÉES ET OBTENIR UN DOI POUR FINALISER LA SOUMISSION DE MON ARTICLE. LES SERVICES D'ACCOMPAGNEMENT M'AIDENT À PRÉPARER UN DÉPÔT CONFORME AU CADRE RÉGLEMENTAIRE.



JE DÉTERMINE LE OU LES ENTREPÔTS PERTINENTS POUR MES DONNÉES. LES SERVICES D'ACCOMPAGNEMENT ME CONSEILLENT ET VALIDENT LES ENTREPÔTS CHOISIS.



JE PRÉPARE MES DONNÉES AU DÉPÔT. JE PEUX SUIVRE DES FORMATIONS EN LIGNE. LES SERVICES D'ACCOMPAGNEMENT PEUVENT ME FOURNIR UNE ASSISTANCE PERSONNALISÉE.



JE DÉPOSE MES DONNÉES, JE BÉNÉFICIE D'UN SUPPORT SI JE RENCONTRE DES DIFFICULTÉS.



JE PEUX AMÉLIORER LA QUALITÉ DE MON DÉPÔT POUR FACILITER LA RÉUTILISATION DE MES DONNÉES (DONNÉES FAIR). JE PEUX SOLLICITER LES SERVICES D'ACCOMPAGNEMENT.



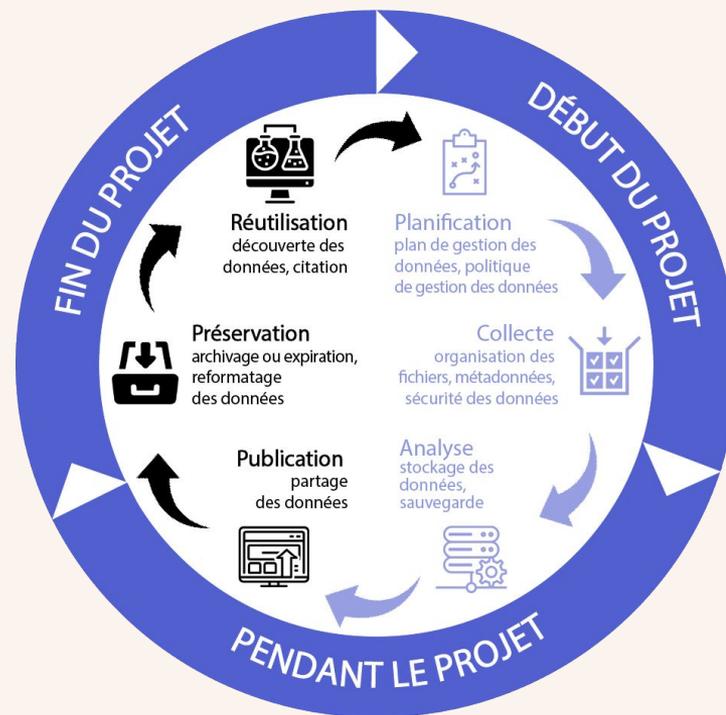
LES SERVICES D'ACCOMPAGNEMENT ASSURENT LA MODÉRATION ET LA CURATION DES DÉPÔTS, POUR GARANTIR LEUR CONFORMITÉ ET LEUR QUALITÉ, ET METTRE EN VALEUR LES DONNÉES.



Des besoins d'accompagnement tout au long du cycle de vie de la donnée

La description des niveaux d'ambition ci-dessus ne porte que sur deux étapes du cycle de vie de la donnée (publication et préservation).

Les besoins d'accompagnement sont plus larges et doivent à terme permettre de soutenir les chercheurs dès la rédaction du projet et du Plan de Gestion des Données.

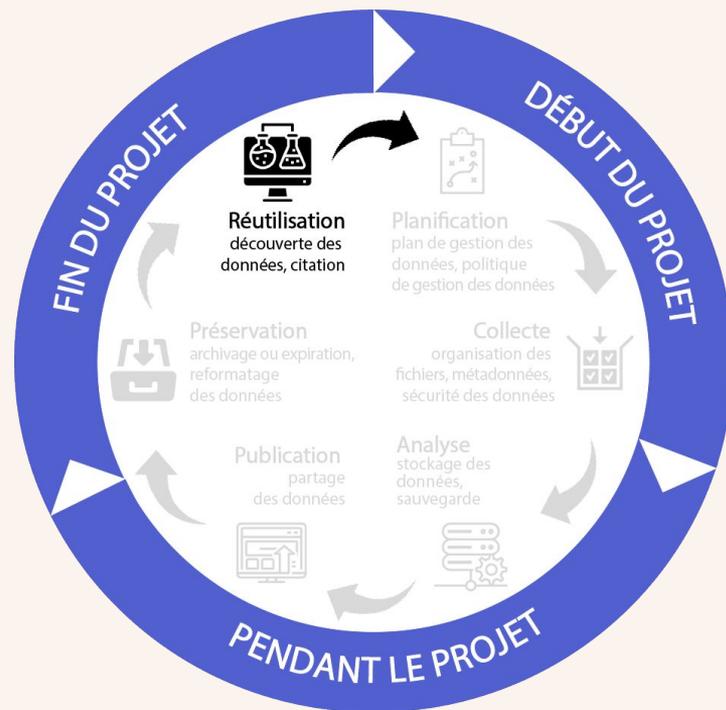


Parcours réutilisateur

Point de départ : Un chercheur souhaite identifier des jeux de données et évaluer rapidement leur adéquation à ses besoins.

Un entrepôt facile à trouver...

- L'étude de recueil des besoins et des contraintes des usagers a fait ressortir l'importance, comme **facteur d'adhésion et de succès**, de la capacité à fournir un point d'accès unique pour la découverte et le suivi des jeux de données de la recherche.
- L'étude a également mis en évidence les attentes fortes de la communauté quant au **niveau de qualité** des données déposées sur l'entrepôt.
- Le service générique n'est pas envisagé comme un silo supplémentaire mais s'inscrit dans un écosystème d'accès aux productions de la recherche en s'appuyant sur les services extérieurs à l'ESRI français.
- Les jeux de données publiés constituent un support à la discussion scientifique.



● Réutilisateur ● Entrepôt ● Autres services en ligne

A

JE CHERCHE DES JEUX DE DONNÉES SUR UN MOTEUR DE RECHERCHE. LES JEUX DE DONNÉES DE L'ENTREPÔT NATIONAL SONT INCLUS DANS LES RÉSULTATS. PAR UN LIEN, J'ARRIVE SUR LA PAGE DESCRIPTIVE.

- ★★ LE RÉFÉRENCIEMENT DES JEUX DE DONNÉES EST AMÉLIORÉ.
- ★★★ LES RESSOURCES LIÉES SONT AUSSI RÉFÉRENCÉES.



B

JE LIS UN ARTICLE SUR UNE REVUE EN LIGNE. JE CLIQUE SUR UN LIEN QUI M'AMÈNE À LA PAGE DESCRIPTIVE DU JEU DE DONNÉES SUR L'ENTREPÔT NATIONAL.

- ★★ LORS DU DÉPÔT, LES MÉTADONNÉES SONT ENRICHIES DES LIENS VERS LES PUBLICATIONS ASSOCIÉES AU JEU DE DONNÉES.



C

JE PARCOURS UNE COLLECTION ÉDITORIALISÉE DE JEUX DE DONNÉES SUR L'ENTREPÔT.



D

JE CHERCHE DES JEUX DE DONNÉES SUR L'ENTREPÔT. LES DONNÉES RÉPONDANT À DES CRITÈRES DE QUALITÉ SONT MISES EN ÉVIDENCE.

- ★★ LA FONCTIONNALLITÉ D'ASSISTANCE À LA SAISIE M'AIDE À FORMULER MA REQUÊTE.
- ★★★ LA RECHERCHE S'EFFECTUE SUR LES MÉTADONNÉES ET DONNÉES EN PLEIN TEXTE.



JE CONSULTE LA DESCRIPTION DÉTAILLÉE D'UN JEU DE DONNÉES SUR L'ENTREPÔT. JE VÉRIFIE SON ADÉQUATION À MON USAGE D'APRÈS LES MÉTADONNÉES ET LA LISTE DES RÉUTILISATIONS.



... pour des données faciles à réutiliser

3 niveaux d'ambition technique de l'entrepôt, pour la recherche et la réutilisation des données :

1. * : garantir une recherche et un téléchargement des données déposées dans l'entrepôt national avec la mise en valeur de leur niveau de qualité ;
2. ** : faciliter la découverte et l'évaluation de l'adéquation d'un jeu de données aux besoins du chercheur ;
3. *** : augmenter la capacité de réutilisation des données.

Les niveaux d'ambition ne sont pas monolithiques : chaque niveau d'ambition regroupe des éléments autonomes, qui peuvent être mis en œuvre indépendamment.

■ Réutilisateur ■ Entrepôt

JE CONSULTE LA DESCRIPTION DÉTAILLÉE D'UN JEU DE DONNÉES SUR L'ENTREPÔT. JE VÉRIFIE SON ADÉQUATION À MON USAGE D'APRÈS LES MÉTADONNÉES ET LA LISTE DES RÉUTILISATIONS.



JE RÉCUPÈRE LA CITATION DU JEU DE DONNÉES AU FORMAT QUI ME CONVIENT.



JE TÉLÉCHARGE LE JEU DE DONNÉES AU FORMAT DE DÉPÔT.



● Réutilisateur ● Entrepôt

JE CONSULTE LA DESCRIPTION DÉTAILLÉE D'UN JEU DE DONNÉES SUR L'ENTREPÔT. JE VÉRIFIE SON ADÉQUATION À MON USAGE D'APRÈS LES MÉTADONNÉES ET LA LISTE DES RÉUTILISATIONS. L'EXPLORATION VISUELLE DES DONNÉES ME LE CONFIRME.



JE RÉCUPÈRE LA CITATION DU JEU DE DONNÉES AU FORMAT QUI ME CONVIENT ET JE M'ABONNE AUX MISES À JOUR DU JEU DE DONNÉES.



JE TÉLÉCHARGE LE JEU DE DONNÉES DANS L'UN DES FORMATS PROPOSÉS.



JE PEUX POSER DES QUESTIONS SUR LE JEU DE DONNÉES ET SON UTILISATION ET PARTAGER MES COMMENTAIRES AVEC LA COMMUNAUTÉ.



■ Réutilisateur ■ Entrepôt

JE CONSULTE LA DESCRIPTION DÉTAILLÉE D'UN JEU DE DONNÉES SUR L'ENTREPÔT. JE VÉRIFIE SON ADÉQUATION À MON USAGE D'APRÈS LES MÉTADONNÉES ET LA LISTE DES RÉUTILISATIONS. L'EXPLORATION VISUELLE DES DONNÉES ME LE CONFIRME.



JE RÉCUPÈRE LA CITATION DU JEU DE DONNÉES AU FORMAT QUI ME CONVIENT ET JE M'ABONNE AUX MISES À JOUR DU JEU DE DONNÉES, AUX NOUVELLES CITATIONS ET RESSOURCES LIÉES.



JE TÉLÉCHARGE LE JEU DE DONNÉES DANS L'UN DES FORMATS PROPOSÉS. JE PEUX ANALYSER LES DONNÉES DIRECTEMENT DEPUIS UN AUTRE SERVICE EN LIGNE.



JE PEUX POSER DES QUESTIONS SUR LE JEU DE DONNÉES ET SON UTILISATION ET PARTAGER MES COMMENTAIRES AVEC LA COMMUNAUTÉ. JE PEUX PROPOSER DES AMÉLIORATIONS SUR LE JEU DE DONNÉES, AUX PRODUCTEURS ET À LA COMMUNAUTÉ DES RÉUTILISATEURS.



Scénarios de mise en œuvre du service de dépôt

Des conditions nécessaires à la pérennité du service

- une **communauté d'utilisateurs**, dont les retours permettront de guider les enrichissements fonctionnels, pour garantir une **qualité d'usage** ;
- une **communauté de développeurs**, dont les contributions garantiront la **pérennité** et la **dynamique** de la solution logicielle (open source) ;
- l'**interopérabilité** avec les autres dispositifs de publication, de diffusion et d'exploitation des publications, codes sources, données : intégration dans l'écosystème des données de la recherche ;
- un **opérateur national** qui maintient le service pour les établissements de l'ESRI, facilite la mutualisation et la coordination des moyens

Trois scénarios de mise en œuvre d'un service de dépôt

Point de départ : un établissement veut permettre à ses chercheurs de déposer leurs jeux de données.

- **mon entrepôt en 1 clic** : un entrepôt personnalisable pour mon établissement, en quelques minutes, sur une infrastructure mutualisée
- **mon entrepôt à la carte** : un entrepôt dédié pour mon établissement, sur une infrastructure mutualisée
- **mon entrepôt à déployer chez moi** : un entrepôt maîtrisé par mon établissement, sur un socle logiciel partagé installé sur mon infrastructure

Un seul de ces trois scénarios sera mis en œuvre.

SCÉNARIO 1 MON ENTREPÔT EN 1 CLIC

UN ENTREPÔT PERSONNALISABLE POUR MON ÉTABLISSEMENT,
EN QUELQUES MINUTES, SUR UNE INFRASTRUCTURE MUTUALISÉE

Administrateur Établissement Opérateur national

- + SIMPLE DANS LE CONTEXTE MULTI-TUTELLES
- + COÛT FAIBLE
- PERSONNALISATION LIMITÉE POUR MON ÉTABLISSEMENT
- ÉVOLUTION DU SERVICE DÉPENDANTE DE LA SYNCHRONISATION DES BESOINS

QUELLES SONT LES DÉMARCHES À FAIRE POUR QUE LES CHERCHEURS DE MON ÉTABLISSEMENT PUISSENT DÉPOSER SUR L'ENTREPÔT ?



L'OPÉRATEUR NATIONAL CRÉE UN ESPACE DE DÉPÔT STANDARD POUR MON ÉTABLISSEMENT.



JE GÈRE MON ESPACE.

- ✓ CRÉATION COLLECTIONS/SOUS COLLECTIONS
- ✓ CHARTE GRAPHIQUE
- ✓ RÔLES UTILISATEUR



L'OPÉRATEUR NATIONAL ASSURE LE STOCKAGE, LA SAUVEGARDE ET LA SÉCURITÉ DES DONNÉES.



AU FIL DU TEMPS, JE BÉNÉFICIE DES ÉVOLUTIONS FONCTIONNELLES DE LA SOLUTION LOGICIELLE.



Évaluation SWOT du scénario 1

<p style="text-align: center;">Forces</p> <ul style="list-style-type: none">• facilité de mise en œuvre et de sortie du service, sans infrastructure établissement• qualité de service garantie pour le stockage et la sécurité des données• très forte mutualisation des améliorations du service• rend facilement visible l'ensemble des données des chercheurs de l'établissement	<p style="text-align: center;">Faiblesses</p> <ul style="list-style-type: none">• pourrait ne pas répondre exactement aux besoins spécifiques de mon établissement• pourrait poser des contraintes pour connecter l'entrepôt à mon système d'information• faible valeur ajoutée par rapport aux entrepôts génériques existants
<p style="text-align: center;">Opportunités</p> <ul style="list-style-type: none">• construction d'un écosystème national• visibilité des données de la recherche au-delà de mon établissement	<p style="text-align: center;">Menaces</p> <ul style="list-style-type: none">• l'opérateur national pourrait imposer des procédures chronophages• éloignement par rapport au niveau local• les autres établissements continuent de créer leurs entrepôts

Un exemple à l'international : DataverseNO

DataverseNO est un entrepôt national générique, opéré à l'université arctique de Norvège (UiT The Arctic University of Norway). Chacun des 9 établissements partenaires bénéficie d'un affichage clair pour ses collections (16 au total). L'entrepôt est certifié CoreTrustSeal. Une équipe pluridisciplinaire de 9 personnes est dédiée à la gestion de l'entrepôt, la curation des données est décentralisée au sein des établissements.

Facteurs de réussite et d'adhésion, à conserver :

- la centralisation de l'infrastructure permet une facilité de mise en œuvre évidente pour les établissements,
- engagement auprès des établissements de la continuité de stockage et d'archivage des données et la migration vers une autre plateforme,
- qualité de service garantie : un entrepôt unique accélère la certification CoreTrustSeal
- la curation des données est très décentralisée dans les établissements

Points d'attention pour la transposition de ce modèle à l'ESRI français :

- le nombre d'établissements beaucoup plus important en France qu'en Norvège (9 institutions, 16 collections au total) implique d'adopter un modèle de gouvernance différent (en Norvège : 1 comité consultatif par collection, qui remonte les besoins de fonctionnalités ; création d'un comité de pilotage technique envisagée si l'ampleur croît)
- point d'attention majeur sur les RH : favoriser la montée en compétences, réflexion d'un réseau pour que le coût RH aille en diminuant, constituer un réseau d'experts locaux pour absorber la montée en volume anticipée

SCÉNARIO 2 MON ENTREPÔT À LA CARTE

UN ENTREPÔT DÉDIÉ POUR MON ÉTABLISSEMENT,
SUR UNE INFRASTRUCTURE MUTUALISÉE

Administrateur Établissement Opérateur national

- + FACILE À METTRE EN ŒUVRE
- + ADAPTABLE AUX PRINCIPAUX BESOINS DE MON ÉTABLISSEMENT
- GESTION ET USAGE COMPLEXES POUR LES PROJETS MULTI-TUTELLES
- CO-EXISTENCE D'ENTREPÔTS UTILISANT DES VERSIONS DIFFÉRENTES DE LA SOLUTION LOGICIELLE

JE VOUDRAIS UN ENTREPÔT
CLÉ EN MAIN POUR
MON ÉTABLISSEMENT.



L'OPÉRATEUR NATIONAL CRÉE UN
ENTREPÔT POUR MON ÉTABLISSEMENT
ET ME DONNE LES DROITS DE GESTION
ET D'ADMINISTRATION.



JE GÈRE MON ESPACE.

- ✓ CRÉATION COLLECTIONS/SOUS COLLECTIONS
- ✓ CHARTE GRAPHIQUE
- ✓ SCHEMA DE MÉTADONNÉES FLEXIBLE
- ✓ RÔLES UTILISATEUR



L'OPÉRATEUR NATIONAL ASSURE
LE STOCKAGE, LA SAUVEGARDE
ET LA SÉCURITÉ DES DONNÉES.



AU FIL DU TEMPS, JE CHOISIS DE BÉNÉFICIER
DES ÉVOLUTIONS FONCTIONNELLES
DE LA SOLUTION LOGICIELLE.



Évaluation SWOT du scénario 2

<p style="text-align: center;">Forces</p> <ul style="list-style-type: none">● compromis facilité/personnalisation● facilité de sortie du service● qualité de service garantie pour le stockage et la sécurité des données● autonomie de gestion de l'entrepôt● capacité d'adaptation aux besoins de mon établissement	<p style="text-align: center;">Faiblesses</p> <ul style="list-style-type: none">● pourrait ne pas correspondre tout à fait aux besoins de mon établissement● pourrait poser des contraintes pour connecter l'entrepôt à mon système d'information● nécessite des compétences diversifiées● difficulté potentielle de choix de l'entrepôt pour le dépôt
<p style="text-align: center;">Opportunités</p> <ul style="list-style-type: none">● construction d'un écosystème national● visibilité des données de la recherche au-delà de mon établissement	<p style="text-align: center;">Menaces</p> <ul style="list-style-type: none">● l'opérateur national pourrait imposer des procédures chronophages● éloignement par rapport au niveau local● les autres établissements continuent de créer leurs entrepôts● duplication des données entre entrepôts d'établissements

Un exemple à l'international : southafrica.figshare.com

L'éditeur commercial Figshare héberge et gère des entrepôts de données de recherche pour 22 universités d'Afrique du Sud.

Chaque université dispose de son portail propre, un portail commun southafrica.figshare.com fédère automatiquement l'ensemble des jeux de données publiés sur les portails des universités.

Un data librarian assure la modération des jeux de données publiés, et l'accompagnement des chercheurs.

Facteurs de réussite et d'adhésion, à conserver :

- Un socle technique homogène pour faire l'interconnexion avec les autres services et construire un écosystème ;
- Une évolution fonctionnelle plus facile à mettre en œuvre ;
- Une souplesse de gestion pour l'établissement

Points d'attention pour la transposition de ce modèle à l'ESRI français :

- **L'étude internationale réalisée par Hugo Catherine ne comporte aucun service national mis en œuvre sur ce modèle.**
- L'adoption de ce modèle par une entreprise établit sa faisabilité technique, mais ne valide pas sa soutenabilité économique sur le long terme. La stratégie commerciale d'un acteur économique peut être amenée à évoluer fortement.
- Les solutions commerciales comme figshare ont été exclues, car l'ESRI serait à la merci des changements de conditions commerciales d'un fournisseur.

SCÉNARIO 3 MON ENTREPÔT À DÉPLOYER CHEZ MOI

UN ENTREPÔT MAÎTRISÉ PAR MON ÉTABLISSEMENT, SUR UN SOCLE LOGICIEL PARTAGÉ INSTALLÉ SUR MON INFRASTRUCTURE

■ Administrateur ■ Établissement ■ Opérateur national

- ⊕ PERSONNALISATION FINE POUR MON ÉTABLISSEMENT
- ⊕ MAÎTRISE DES DONNÉES DE MON ÉTABLISSEMENT
- ⊖ COÛTS ÉLEVÉS DE PERSONNELS ET D'INFRASTRUCTURES À LA CHARGE DE MON ÉTABLISSEMENT
- ⊖ COUVERTURE DE TOUS LES ÉTABLISSEMENTS DE L'ESRI TRÈS DIFFICILE À ATTEINDRE

JE MONTE UNE ÉQUIPE DÉDIÉE QUI DÉFINIT LES BESOINS FONCTIONNELS ET TECHNIQUES DE L'ENTREPÔT.



L'OPÉRATEUR NATIONAL FOURNIT UNE SOLUTION LOGICIELLE SOCLE, MA DSI L'INSTALLE SUR SES SERVEURS.



JE GÈRE MON ESPACE.

- ✓ CRÉATION COLLECTIONS/SOUS COLLECTIONS
- ✓ CHARTE GRAPHIQUE
- ✓ SCHEMA DE MÉTADONNÉES FLEXIBLE
- ✓ RÔLES UTILISATEUR



MA DSI ASSURE LE STOCKAGE, LA SAUVEGARDE ET LA SÉCURITÉ DES DONNÉES.



JE CONTRACTUALISE AVEC L'OPÉRATEUR NATIONAL POUR FAIRE ÉVOLUER LES FONCTIONNALITÉS SUR LE SOCLE LOGICIEL, MA DSI FAIT LES MISES À JOUR.



Évaluation SWOT du scénario 3

<p style="text-align: center;">Forces</p> <ul style="list-style-type: none">• maîtrise du patrimoine de données de recherche de mon établissement• autonomie de gestion et d'administration de mon entrepôt• intégration de l'entrepôt dans le système d'information de mon établissement	<p style="text-align: center;">Faiblesses</p> <ul style="list-style-type: none">• nécessite des compétences diversifiées et une infrastructure dédiée• coût élevé pour mon établissement• difficulté potentielle de choix de l'entrepôt pour le dépôt• fragmentation des entrepôts• couverture de tous les établissements de l'ESRI très difficile à atteindre
<p style="text-align: center;">Opportunités</p> <ul style="list-style-type: none">• création d'une communauté technique entre établissements• développement de fonctionnalités avec des partenaires externes	<p style="text-align: center;">Menaces</p> <ul style="list-style-type: none">• perte de compatibilité entre les développements propres et les mises à jour du socle logiciel commun• duplication des mêmes données sur plusieurs entrepôts

Un exemple à l'international : DANS (NL)

Depuis 2005, **DANS** est un opérateur national qui fournit des services de publication (DataverseNL), d'archivage (EASY) et de découverte des résultats de la recherche (NARCIS). DANS réalise les développements logiciels mais chaque établissement a son déploiement Dataverse sur ses serveurs, reliés via DataverseNL et au méta-moteur NARCIS.

Facteurs de réussite et d'adhésion, à conserver :

- La mutualisation du travail de développement permet d'offrir des solutions prêtes à déployer dans chaque établissement.
- La création d'un écosystème facilite le lien entre publication, données, personnes (NARCIS).
- La construction de l'écosystème s'est faite étape par étape.

Points d'attention pour la transposition de ce modèle à l'ESRI français :

- La taille et la complexité institutionnelle des deux pays n'est pas comparable ;
- Besoins d'une volonté politique forte et d'un soutien financier continu.

DANS a un poids important dans les initiatives de construction d'infrastructure à l'échelle européenne (EOSC) + coalition RD-NL + European Social Science Data Archive, DARIAH, board de CoreTrustSeal. DANS est centre d'expertise reconnu sur les principes FAIR.

En bref ...

COMPARAISON FONCTIONNELLE DES SCÉNARIOS

DROITS ATTACHÉS AU PRODUCTEUR DE LA DONNÉE

SCÉNARIO 1
MON ENTREPÔT EN 1 CLIC

inchangés

SCÉNARIO 2
MON ENTREPÔT À LA CARTE

inchangés

SCÉNARIO 3
MON ENTREPÔT À DÉPLOYER CHEZ MOI

inchangés

MISE EN AVANT DE LA MARQUE ÉTABLISSEMENT



VISIBILITÉ NATIONALE ET INTERNATIONALE DES DONNÉES



PERSONNALISATION DU PROCESSUS DE MODÉRATION



VISIBILITÉ DE L'ENSEMBLE DES DONNÉES DU CHERCHEUR, Y COMPRIS CELLES DÉPOSÉES SUR DES SERVICES DE DONNÉES TIERS



FACILITÉ D'ÉTABLIR DES LIENS DEPUIS LES SERVICES NATIONAUX DE PUBLICATIONS ET CODES SOURCES VERS LES DONNÉES



INTÉGRATION AVEC DES ENVIRONNEMENTS NATIONAUX DE RÉUTILISATION DES DONNÉES (ANALYSE ET CALCUL)



COMPARAISON ÉCONOMIQUE DES SCÉNARIOS

MUTUALISATION DU
SOCLE TECHNOLOGIQUE

RESSOURCES
NATIONALES



RESSOURCES
ÉTABLISSEMENT



SOUTENABILITÉ
DU SERVICE



Les indicateurs ci-dessus expriment des ordres de grandeur des coûts induits par chacun des scénarios (€ = peu coûteux, €€€€€ = très coûteux).

Recommandations du Groupe de Travail

- **Accompagner** les chercheurs pour assurer un dépôt de **qualité élevée** ;
- Favoriser la **montée en compétences** et la mise en réseau des personnels d'appui à la recherche pour consolider cet accompagnement ;
- Constituer un **réseau d'experts locaux** pour absorber la montée en volume anticipée ;
- Réaliser **la modération** et la **curation au plus près des chercheurs**, avec la possibilité de **solliciter des experts** d'autres établissements pour répondre à des questions spécifiques.

*“ Pour chaque institution, stocker ses données de manière pérenne a un avantage mais la redondance et la **multiplicité** des diverses **initiatives** n'est pas écologique, et il faut réfléchir au fait qu'on n'est pas obligé de tout garder.”*

Recommandations du Groupe de Travail

- **Modérer** les dépôts et **certifier** l'entrepôt ;
- Assurer une **qualité de service** garantie par un opérateur national ;
- **Mutualiser** les développements de la **solution logicielle**.
- Inscrire et connecter le service aux **écosystèmes nationaux et internationaux**, en particulier celui de l'EOSC ;

*“ Certaines communautés sont perdues face à un type de données. L'idée serait d'imaginer un **aiguillage** et de leur proposer des solutions. L'institution locale ou nationale est là pour leur dire qu'en structurant leurs données de telle manière, ils pourront déposer dans tel entrepôt. ”*

*“ Il faut éviter de gaspiller les énergies et vouloir faire son truc à soi avec son logo. C'est une dépense d'énergie, d'impact, de tout parce que la concurrence n'est plus CNRS versus université mais recherche française versus Elsevier, Google etc. **S'il y a dispersion des forces, on aura les outils qu'ils nous donneront et pas les outils qu'on aura choisis.** ”*

Recommandations du Groupe de Travail

- Soutenir le service **financièrement et durablement** ;
- **Créer une communauté technique et fonctionnelle** à partir des compétences développées sur les entrepôts déjà déployés ;
- Construire un **service de référence** à l'échelle **européenne** ;

*“ Ce qui manque actuellement, c'est d'assurer la pérennité des services qui sont rendus. **Si je pars du laboratoire, ma plateforme de diffusion de données va disparaître** car il n'y aura personne pour la maintenir. ”*