

**PiREN  
Seine**

Les ateliers  
du  
**PIREN-Seine**

PIREN-SEINE BASÉE  
B-T4A1c-2103-STOP  
Vb = 150 ml

PIREN-SEINE BASÉE  
B-T4A1a-2103-STOP  
Vb = ml

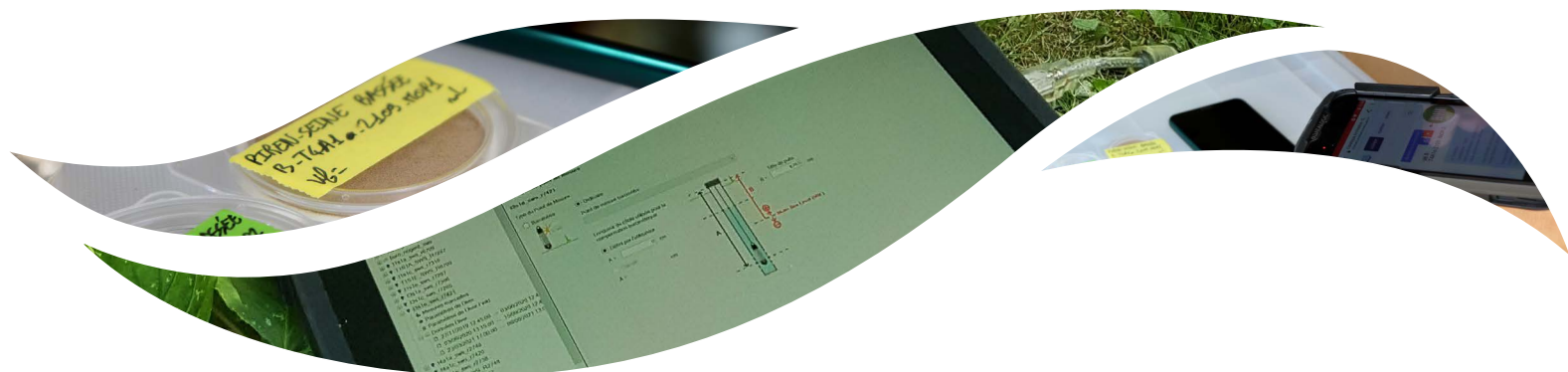
Retour à la rec...

LM B-  
T4A1c-2109- N

# TERRAIN  
Localisation : T4A1  
Date prélèvement : 07/05  
Heure prélèvement : hh:mm  
Date filtration : j/mm/aaaa  
Heure filtration : hh:mm  
Température (°C) :  
Conductivité (µS/cm) :  
Oxygène (mg/l) :  
pH :  
Observation :

Rendre  
**accessibles** les  
**données** du  
**PIREN-Seine**

# Les ateliers du PIREN-Seine



## Rendre **accessibles** les **données** du **PIREN-Seine**

Le 09 avril 2021



# SOMMAIRE

|  |    |
|--|----|
| ● Introduction   | 3  |
| ● Organismes   | 3  |
| ● Participants   | 3  |
| ● Enjeux et objectifs de l'atelier                                       | 4  |
| ● Notions clés et liens utiles   | 5  |
| ● Présentation et synthèse des échanges autour du Géocatalogue           | 8  |
| ● Synthèse des discussions   | 10 |
| #1 - <i>Quelles sont les habitudes et les besoins des chercheurs ?</i>   |    |
| #2 - <i>Quels jeux de données peut-on co-construire ensemble ?</i>       |    |
| #3 - <i>Quid de la propriété des données co-construites ?</i>            |    |
| ● Retour d'expérience de la Zone Atelier Armorique par Françoise Le Moal | 11 |
| ● Conclusion   | 12 |



## Introduction

Le PIREN-Seine rassemble des équipes de recherche et des partenaires opérationnels sur un territoire commun : le bassin de la Seine. Depuis plus de 30 ans, ces fructueux échanges ont généré d'importants volumes de données qui ont contribué à l'émergence de connaissances nouvelles. Aujourd'hui, le PIREN-Seine s'inscrit dans l'impulsion nationale et internationale en faveur de l'ouverture des données de la Recherche. Dans sa phase 8 (2019-2023), le PIREN-Seine s'engage à identifier, cataloguer, diffuser et valoriser les données ainsi que les métadonnées associées. Dans le paysage en constante évolution de l'Open data, le dialogue entre les chercheurs et les partenaires opérationnels est essentiel pour identifier les besoins ou les verrous et co-construire des solutions communes adaptées aux enjeux de la Science Ouverte.

L'atelier a été organisé en 4 parties :

1. Présentation des notions clefs et des liens utiles relatifs à la Science Ouverte.
2. Présentation et discussion autour du Geocatalogue PIREN-Seine-ZA Seine.
3. Discussion autour des pratiques et des échanges de données envisageables.
4. Présentation du retour d'expérience du Geocatalogue ZA Armorique.

## Organisateurs

David Eschbach (METIS) et Marie Silvestre (FIRE), avec le concours d'Alexandre Deloménie (Arceau-IDF) et de Laurence Lestel (METIS).

Contact : [data@piren-seine.fr](mailto:data@piren-seine.fr)

## Participants

L'atelier a réuni 37 participants, dont près d'un tiers était représenté par les partenaires du PIREN-Seine. Les partenaires scientifiques représentés par le GIP Seine Aval et OPUR sont également présents, ainsi que le BRGM et la Zone Atelier Armorique. Des chercheurs des 5 axes du PIREN-Seine sont présents.

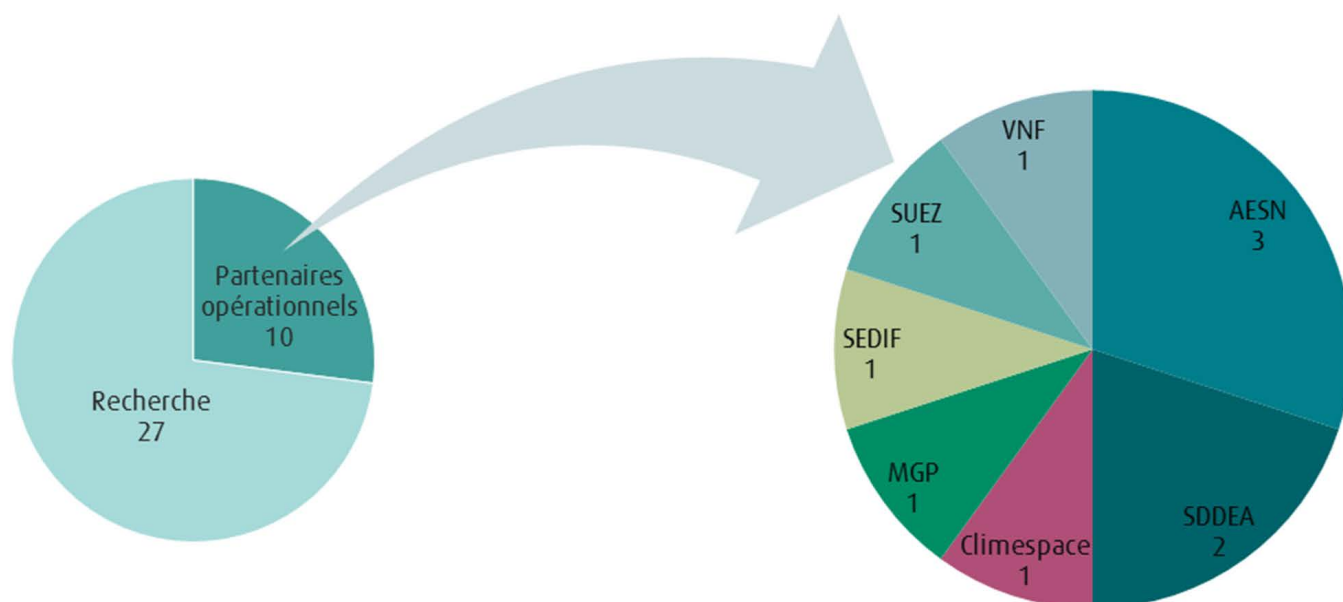


Figure 1 – Gauche : répartition des participants entre chercheurs et partenaires opérationnels ; Droite : répartition des partenaires opérationnels ayant participé à l'atelier en fonction de leur l'organisme d'attachement.

### Les acronymes

**AESN** : Agence de l'eau Seine-Normandie

**SDDEA** : Régie du syndicat mixte de l'eau de l'assainissement collectif, de l'assainissement non collectif, des milieux aquatiques

**MGP** : Métropole du Grand-Paris

**SEDIF** : Syndicats des eaux d'Île-de-France

**VNF** : Voies navigables de France

## Enjeux et objectifs de l'atelier

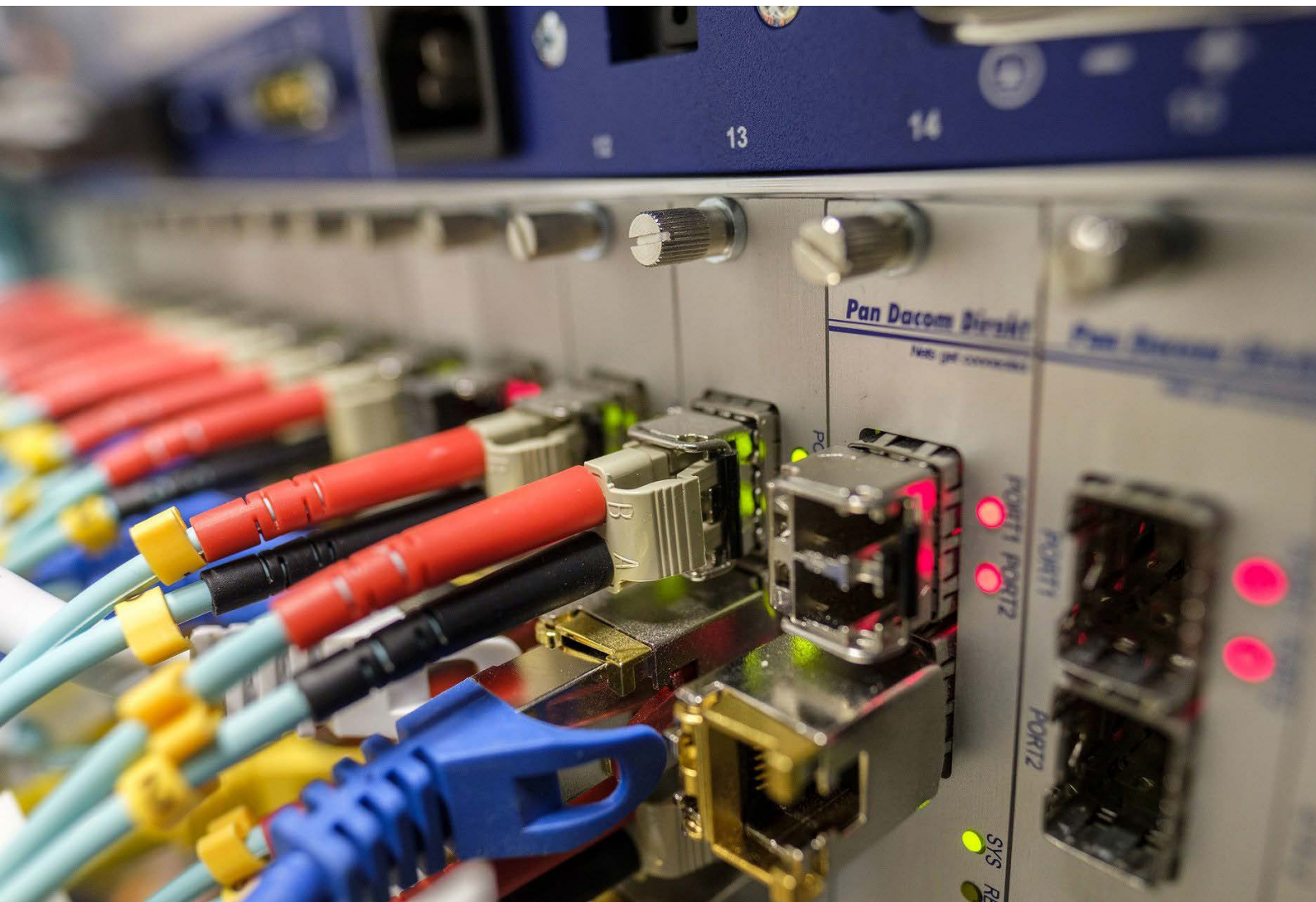
Cette première rencontre s'inscrit dans la volonté du PIREN-Seine de s'engager dans une action d'ouverture de la science. Saisissant l'opportunité que représente la mutation numérique, le PIREN-Seine s'engage à diffuser ses résultats, ses méthodes et les produits de ses recherches. Intégré à l'axe 6 de la phase 8, le bloc 6.2 « Des données partagées et accessibles » vise à accompagner les chercheurs du PIREN-Seine dans cette action. Outre les ateliers de co-réflexion annuels permettant de stimuler les échanges entre chercheurs et partenaires, de nombreux événements sont organisés tout au long de la phase 8 : *metadata-party*, collecte des données patrimoniales, formations...

Dans le cadre de ce premier atelier, il s'agit d'une part de **partager les pratiques de gestion de données** entre les équipes de recherche et les partenaires, et d'autre part **d'identifier les jeux de données existants** pouvant être mis en commun et/ou construits collectivement.

Marqueur d'une nouvelle action en faveur de la Science Ouverte, cet atelier symbolise également l'émergence de nouvelles pratiques autour de l'utilisation et la gestion des données de la recherche.

Quatre défis prioritaires sont identifiés :

- créer une base de « métadonnées » dédiée aux productions du PIREN-Seine ;
- collecter, sauvegarder et archiver les données patrimoniales ;
- créer des projets de données autour de recherches collectives (e.g. Bassée, axe Seine, etc.) ;
- mettre en place une base de données collective incluant tous les acteurs opérationnels et scientifiques.



## Notions clés et liens utiles

La Science Ouverte est la diffusion sans entrave des publications et des données de la recherche ([Ministère de l'enseignement supérieur, de la recherche et de l'innovation](#)). Elle s'appuie notamment sur un ensemble de réglementations européennes déclinées à l'échelle nationale telle que la Directive Inspire ([Directive 2007/2/CE](#)). En France, la Loi pour une République numérique ([Loi 2016-1321](#) du 7 octobre 2016) a permis d'impulser une dynamique d'ouverture des données et des connaissances. En 2018, le Plan National pour la Science Ouverte ([PNSo](#)) a permis d'inscrire la recherche française dans le mouvement mondial d'ouverture des données et de la transparence de l'action publique. Le PNSo a récemment fait l'objet de révisions qui couvrent la période 2021-2024 ([PNSo2](#)).

La déclinaison locale de la réglementation en faveur de la Science Ouverte doit être détaillée dans un **plan de gestion des données** rédigé à l'échelle d'un projet, d'une institution ou d'un programme de recherche (Fig. 2). Il permet de définir précisément les principes de gestion des données ainsi que les ressources qui devront être allouées (moyens humains et matériels) à chaque étape du cycle de vie de la donnée (Fig. 5).





|   | Cadre   | Facilitateur                                   | Outil  |
|---|---|--|--|
| <p>Niveau européen</p>  |  | Reiso  | Guide pratique pour l'harmonisation internationale |
|                        | Loi Lemaire<br>PNSo   | CoSo<br>INIST-CNRS<br>SIST<br>DoraNum<br>[...] | OPIDOR<br>Guides des bonnes pratiques              |
| <p>Niveau local</p>    | DMP<br>Convention   | Data manager<br>Tuto<br>Formations             | Géocatalogue<br>TGIR HumaNum<br>HAL<br>[...]       |

Figure 2 - Le contexte réglementaire de l'Open Data à différentes échelles.

**Le guide des bonnes pratiques sur la gestion des données** (téléchargeable [ici](#)) est une condition préalable essentielle à la découverte de nouvelles connaissances et un support incontournable pour la gestion des données. Des principes de bonnes gestions des données sont régulièrement diffusés, notamment dans la communauté des gestionnaires de données mais aussi sur des sites institutionnels comme [Doranum](#) qui regroupe et synthétise de nombreuses informations concernant les données de la recherche et de l'apprentissage numérique (Fig. 3).

The screenshot shows the Doranum website interface. At the top, there is a navigation bar with the Doranum logo on the left and menu items: ACCUEIL, RESSOURCES, ACTUALITES, A PROPOS, CONTACT. On the right of the navigation bar are social media icons for Twitter, YouTube, LinkedIn, and GitHub, along with a search icon and a globe icon. Below the navigation bar, the text 'PARCOUREZ LES RESSOURCES CLASSÉES PAR THÉMATIQUES' is displayed. The main content area is a grid of nine resource cards, each with a title, a brief description, and an icon:

- ENJEUX & BÉNÉFICES**: Pourquoi partager les données ? Qu'est-ce que l'Open Science ? (Icon: Atom)
- ASPECTS JURIDIQUES, ÉTHIQUES, INTÉGRITÉ SCIENTIFIQUE**: Que puis-je partager, réutiliser ? Quelles pratiques devrais-je respecter ? (Icon: Scales)
- PLAN DE GESTION DE DONNÉES**: Pourquoi et comment rédiger un plan de gestion des données ? (Icon: Document)
- MÉTADONNÉES**: Comment décrire les données ? (Icon: ID card)
- IDENTIFIANTS PÉRENNES**: Comment associer durablement des données à son auteur ? (Icon: ID card)
- DÉPÔT & ENTREPÔTS**: Comment et où déposer mes données ? (Icon: Warehouse)
- STOCKAGE & ARCHIVAGE**: Quelles données conserver à long terme et comment ? (Icon: Clouds)
- DATA PAPERS & DATA JOURNALS**: Comment publier mes données comme un article scientifique ? (Icon: Document)
- ACCÈS & VISUALISATION**: Où et comment extraire et visualiser les données qui m'intéressent ? (Icon: Eye)

Figure 3 - DoraNum regroupe de nombreuses ressources destinées à accompagner les chercheurs dans la mise à disposition de leurs jeux de données ([doranum.fr](#)).



**Les principes FAIR** sont des lignes directrices permettant de Faciliter le repérage, l'Accessibilité, l'Interopérabilité et la Réutilisation des ressources numériques. Suivre ces principes vise à faciliter la reproductibilité et le croisement des données originelles (Fig. 4).

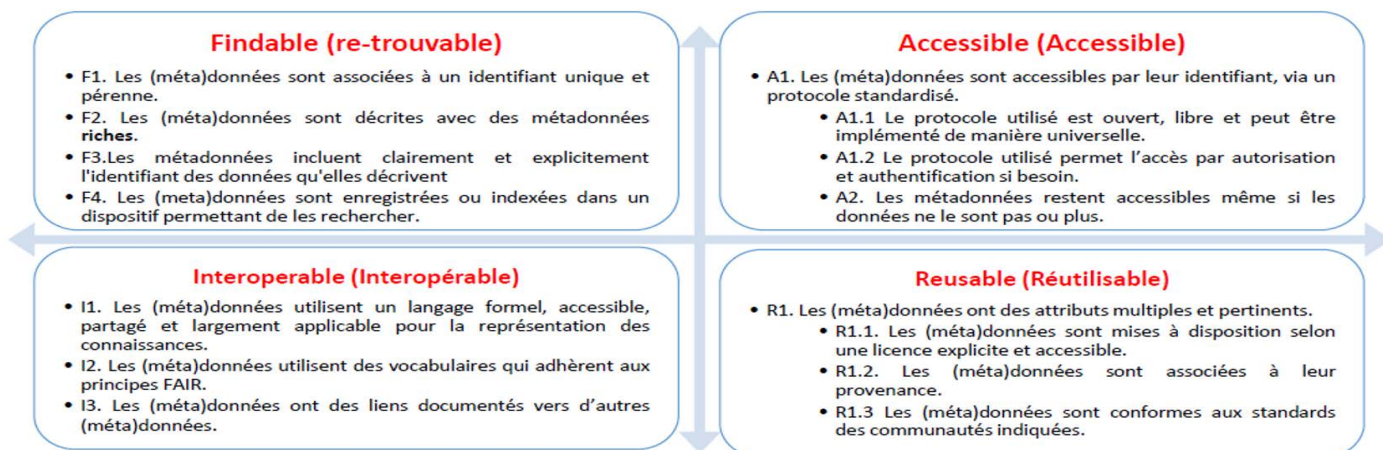


Figure 4 - Le modèle FAIR selon Wilkinson et al. (2016)<sup>1</sup>

**Le cycle de vie de la donnée** (Fig. 5) décrit la manière dont les données sont gérées depuis le processus de création jusqu'à leur réutilisation, en passant par leur publication.

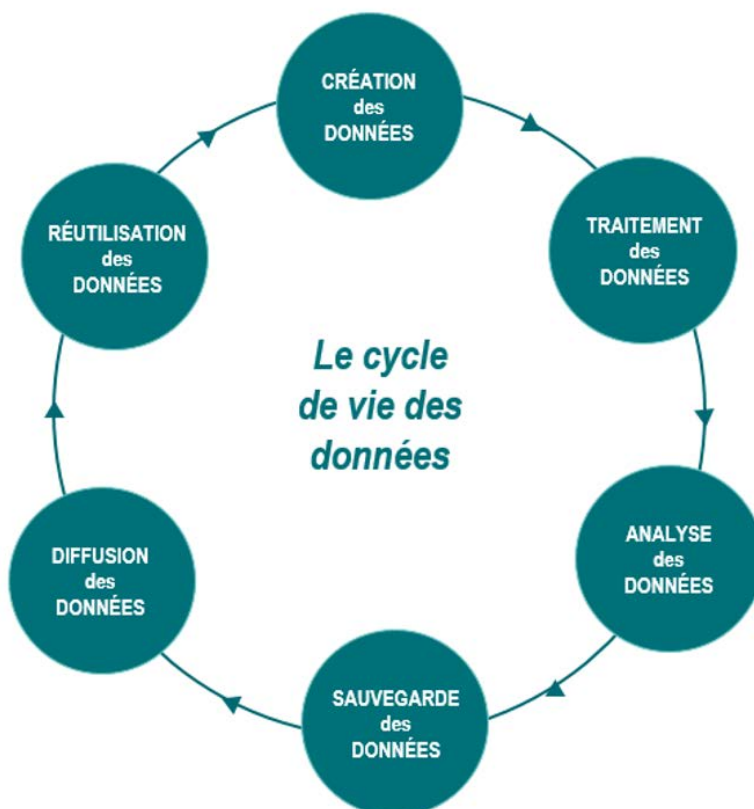


Figure 5 - Le cycle de vie des données, adapté de UK Data Archive.

<sup>1</sup>Wilkinson, M. D. et al., 2016. *The FAIR Guiding Principles for scientific data management and stewardship*. Sci. Data 3:160018 doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)



# Présentation et synthèse des échanges autour du Géocatalogue

**Un géocatalogue PIREN-Seine-Zone Atelier Seine (ZAS)** - La solution retenue est une Infrastructure de Données Spatiales (IDS). On parle plus couramment de géocatalogue. Ce dernier permet de réunir, de visualiser et de croiser des données spatialisées mais il peut également être utilisé pour compiler, organiser et gérer tous types de données.

Le géocatalogue permet de répondre à 3 objectifs principaux :

1. Rassembler les données.
2. Diffuser les données.
3. Valoriser les données.

Une solution de catalogage est en cours de déploiement au sein du PIREN-Seine et de la ZA Seine (Fig. 6). Il s'appuie sur l'IDS GeoOrchestra sous licence libre et est opérationnel depuis septembre 2021 à l'adresse [data.za-seine.fr](https://data.za-seine.fr).



Figure 6 – Présentation schématique de l'outil retenu par le PIREN-Seine et la ZA Seine. L'IDS se compose d'un géocatalogue (GeoNetwork), d'un serveur permettant de gérer les données spatiales (GeoServer) et d'un visualiseur (MapStore2). L'ensemble est mis en musique par GeoOrchestra.

**Mode de fonctionnement** - Les jeux de données sources sont identifiés à partir des résultats publiés et décrits dans des fiches de métadonnées. Des fiches-types ont été construites pour faciliter le travail de saisie et homogénéiser les informations descriptives tout en se conformant au modèle européen INSPIRE et répondre aux principes FAIR. Les fiches de métadonnées associées à la donnée seront centralisées sur le géocatalogue. Les droits d'accès et les licences associées à chaque jeu de données sont à définir au cas par cas par leur propriétaire.

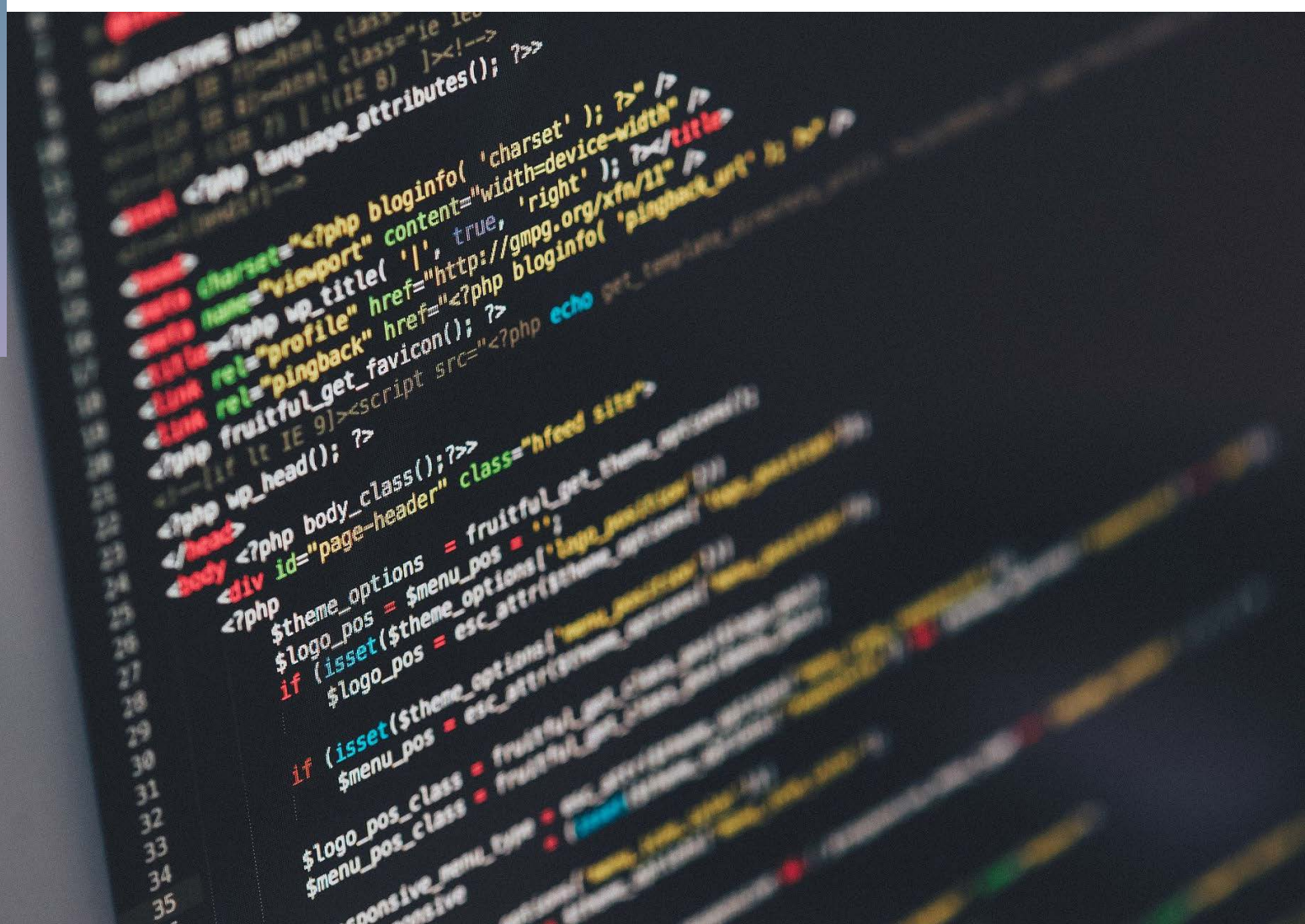
Des DOI (**Digital Object Identifier**) peuvent être créés sur les fiches de métadonnées et sur les jeux de données géographiques afin de mettre en place des chemins d'accès pérennes.

L'interface de visualisation cartographique actuelle (Mapfishapp) sera remplacée par une interface plus ergonomique (MapStore 2) en début d'année 2022.

**Synthèse des échanges autour de la présentation du Géocatalogue** – Les échanges ont notamment été guidés par des questions d'ordre techniques et opérationnelles. Les points abordés ont été les suivants :

- DOIser les jeux de données est une condition essentielle qui contribue à la partie « Accessible » du principe FAIR ;
- Nécessité d'identifier les données prioritaires à échanger entre partenaires et chercheurs ;
- Définir une procédure pour garantir la qualité du catalogue de métadonnées et le respect des principes FAIR. Pour le moment il n'y a pas d'outil pour mesurer la qualité des informations renseignées et le degré de FAIRness, mais simplement un contrôleur binaire qui valide les conditions INSPIRE (est-ce que le champ est renseigné ou non) ;
- Renseignements concernant la diversité des thésaurus disponibles et les modalités d'accès au géocatalogue (inscription, connexion, etc.).

Ces échanges ont mis en évidence l'importance d'élucider les zones d'ombres persistantes, notamment sur le plan technique. Il apparaît que les habitudes autour de la gestion des données sont plutôt lacunaires, voire inexistantes mais qu'il existe un vif intérêt des chercheurs et des partenaires pour améliorer les processus de mise à disposition de leurs données. L'ensemble des participants s'accorde à dire que la solution proposée facilitera la valorisation et la diffusion des données de la recherche.



# Synthèse des discussions concernant les pratiques et les échanges de données envisageables

Un sondage a été réalisé lors de la préparation de l'atelier. Trois questions-clefs ont été identifiées au moment de l'analyse des résultats. La première interroge les habitudes et les besoins des chercheurs face au processus d'édition des métadonnées (#1), la seconde concerne la gestion et la co-construction des jeux de données (#2), la dernière porte sur la propriété des données co-construites (#3).

## #1 – Quelles sont les habitudes et les besoins des chercheurs dans le processus d'édition des métadonnées ?

De nombreux chercheurs éditent déjà des métadonnées dans des bases de données personnelles, mais peu d'entre eux poussent leurs données dans des géocatalogues ou des entrepôts de données. Le retour d'expérience d'autres organismes déjà bien avancés sur cette question (e.g. INRAE) montre qu'il est important de bien structurer la démarche en amont, d'où l'intérêt de mettre en place un plan de gestion des données le plus tôt possible (enjeu prioritaire du premier semestre 2022).

Recommandations générales : avant le versement dans le géocatalogue, toutes les données doivent au préalable faire l'objet d'un tri, d'un nettoyage, voire d'une validation. Des outils permettant le tri et la validation des données existent déjà chez certains partenaires, c'est notamment le cas du GIP Seine Aval.

Mutualiser l'utilisation du géocatalogue et utiliser autant que possible les outils déjà existants pour faciliter la compilation et l'intégration des métadonnées dans le géocatalogue (e.g. Geoflow : outil en cours de développement à INRAE permettant de renseigner automatiquement les métadonnées de plusieurs catalogues depuis un tableur).

Par ailleurs, le géocatalogue PIREN-Seine-ZA Seine pourrait devenir un portail permettant de faire le lien vers d'autres catalogues existants.

## #2 - Quels jeux de données peut-on co-construire ensemble ? Comment gérer ces données ?

Bien que de nombreuses bases de données existent sur des sujets proches de ceux du PIREN-Seine, la récupération des données reste complexe, notamment lorsque la compilation des jeux de données répond à une structuration interne spécifique à chaque organisme. Pour faciliter les transferts et les échanges des données préexistantes, le choix des mots-clés et des thésaurus est donc essentiel. La question de la mutualisation et de l'interconnexion des jeux de données est difficile, notamment sur le plan juridique mais aussi sur le plan opérationnel (simulation en temps réel par exemple). L'optimisation de la gestion des données ainsi qu'un moissonnage systématique permettrait de faciliter les échanges et la mise en commun des jeux de données préexistants.

## #3 - Quid de la propriété des données co-construites ?

Les conventions de financement des équipes de recherche du PIREN-Seine prévoient que si plusieurs équipes participent à la création de jeux de données, leur propriété est partagée entre les différentes équipes ayant participé au projet. La propriété est donc copartagée à hauteur de la contribution des équipes. Il en va de même pour les données co-construites avec des partenaires opérationnels. Des protocoles d'accord sont généralement mis en place et doivent préciser la propriété des données et les licences associées. Il est recommandé de se faire accompagner par les services juridiques des établissements concernés pour la rédaction de ce document.

À noter : ces aspects font partie des éléments à intégrer dans le plan de gestion de données. Celui-ci doit être rédigé en amont de la création des données et son contenu peut évoluer tout au long du cycle de vie des données.

# Retour d'expérience de la Zone Atelier Armorique par Françoise Le Moal

Les données de la Zone Atelier Armorique (ZAAr) sont collectées depuis 1993 et la démarche s'inscrit dans le cadre réglementaire de la Science Ouverte. Pour cataloguer ses données dans un format standardisé, la ZAAr s'appuie sur l'infrastructure de données géographiques [OSURIS](#) (fig. 7), mise en place par l'Observatoire des Sciences de l'Université de Rennes (OSUR). Le développement de cet outil fait l'objet de prestations externes ponctuelles, notamment pour certaines maintenances et mises à jour. Les technologies utilisées sont proches de celles du géocatalogue PIREN-Seine-ZA Seine.

Aucun agent n'est dédié spécialement (à 100%) à l'administration, la maintenance et l'animation de l'outil OSURIS. Des % d'ETP des agents des unités constituant l'OSUR y sont dévolus. Au démarrage du projet, une dizaine d'agents ont participé à la mise en place du serveur, au déploiement, puis à l'administration et à l'animation. Actuellement, seuls 2 ou 3 agents participent encore officiellement au bon fonctionnement de l'outil, toutes tâches réunies, c'est-à-dire, moins de 0,3 ETP.

Concernant la gestion des données de la ZAAr, aucune personne n'est dédiée spécifiquement à cette tâche. Un agent est « référent » données/métadonnées ZAAr pour OSURIS à hauteur de 5% de son temps. Chaque unité participant à la ZAAr doit désigner un « référent » données pour chaque projet.

Le catalogue est essentiellement utilisé pour l'édition des métadonnées (fonction Geonetwork). Avec 358 fiches de métadonnées publiées, le bilan est positif. De nombreuses cartes ont été produites et diffusées, des *Metadata party* sont organisées. Le bilan est plus contrasté quant au nombre de jeux de données accessibles qui reste faible et à l'impossibilité de connaître les statistiques d'utilisation. Même si la demande des utilisateurs reste forte, que les référents sont bien identifiés et que la communauté est bien mobilisée, la collecte des données se heurte encore souvent à la réticence des chercheurs. Il manque notamment du personnel disponible pour maintenir et développer l'outil. La question du stockage à long terme se pose également.



Figure 7 : Capture d'écran du site de l'infrastructure de données géographiques OSURIS.

## Conclusion

Ce premier atelier autour des données – métadonnées a permis de soulever de nombreuses questions concernant les pratiques de gestion des données. Il a clairement mis en lumière le besoin de formation (fig. 8) des équipes de recherche et des partenaires sur les principes FAIR et le rôle structurant de déploiement d'un plan de gestion de données adapté aux périmètres du PIREN-Seine et de la ZA Seine. Des enjeux importants ont été identifiés, notamment en ce qui concerne la propriété de la donnée et sa diffusion, mais aussi la difficulté de mettre en commun des données d'origine très différente. La question de la donnée est aujourd'hui centrale dans le travail des chercheurs, mais d'importants efforts restent à fournir pour optimiser les moyens de collecte, de bancarisation et d'échanges des données, notamment entre les différents partenaires. Ce travail ambitieux nécessite des moyens, à la fois financier et humain, comme nous l'a rappelé Françoise Le Moal dans son retour d'expérience. Les outils existent et certains sont en cours de déploiement au sein du PIREN-Seine et de la ZA Seine, à nous (chercheurs, partenaires scientifiques et opérationnels) d'identifier les solutions les mieux adaptées à nos besoins et de construire ensemble le plan de gestion qui permettra de structurer nos relations avec les données sur le long terme.

À l'issue de l'atelier, un questionnaire de retour a été envoyé aux participants. Il a permis de mettre en évidence une demande prioritaire pour la mise en ligne de jeux de données issus de publications ou de suivis temporels longs et un intérêt moindre pour les données brutes. Par ailleurs, des besoins spécifiques ont été identifiés concernant notamment la rédaction d'un plan de gestion de données, le choix et l'utilisation des licences ainsi que l'accompagnement dans la saisie de métadonnées.

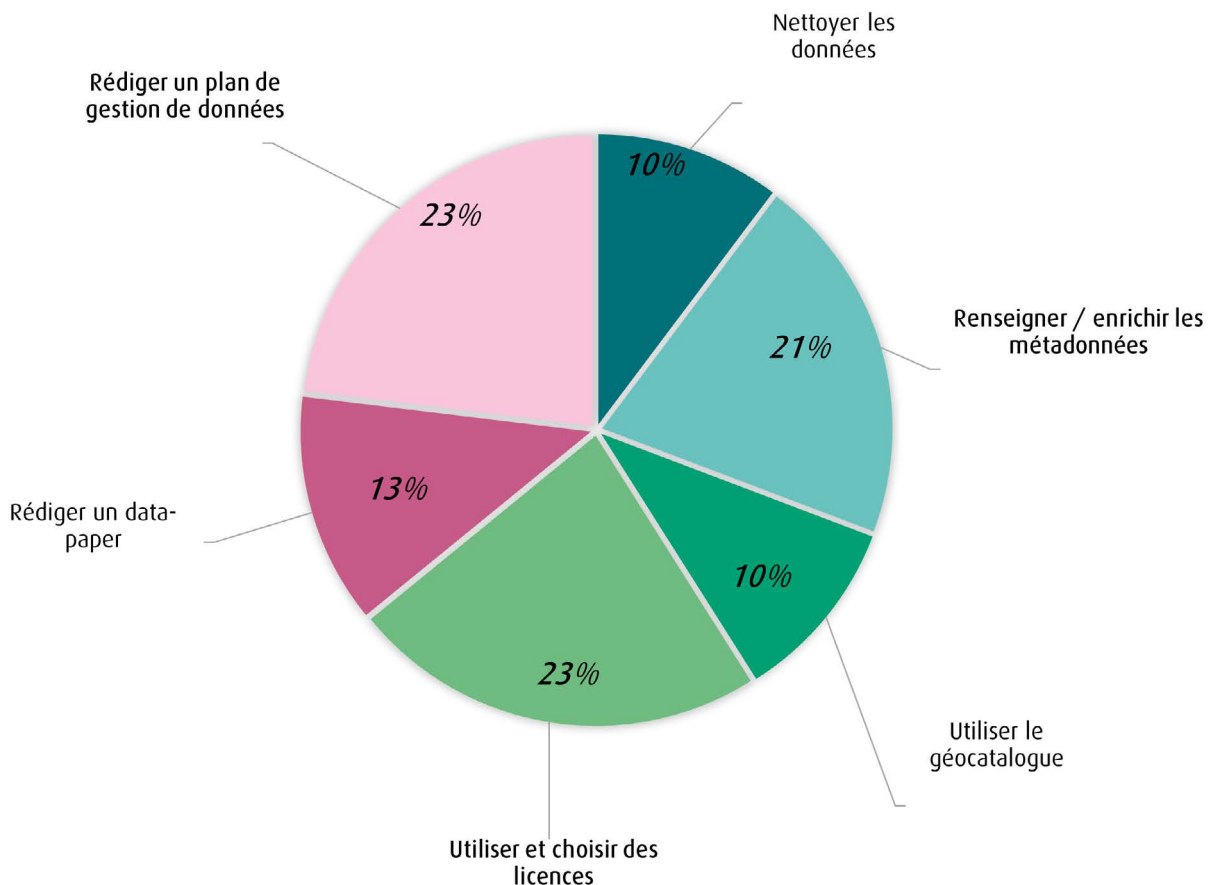


Figure 8 : Les besoins en formation des équipes de recherche et des partenaires sur les principes FAIR issus du questionnaire post-atelier.

Le **PIREN-Seine** est un programme de recherche interdisciplinaire en environnement dont l'objectif est de développer une vision d'ensemble du fonctionnement du bassin versant de la Seine et de la société humaine qui l'investit, pour permettre une meilleure gestion qualitative et quantitative de la ressource en eau.

La cellule transfert des connaissances a pour but de valoriser les productions de savoirs scientifiques issues des recherches du **PIREN-Seine**, et de favoriser la mise à disposition de ces informations à un large public, des professionnels de la gestion de l'eau aux élus en passant par les usagers. Animée depuis octobre 2016 par l'association ARCEAU IdF, cette cellule répond à une forte volonté de la part des chercheurs de participer au transfert des savoirs scientifiques et techniques vers la société civile. Elle est ainsi chargée de la rédaction et de l'édition de fiches thématiques, de la mise en ligne de contenus scientifiques adaptés à la fois aux professionnels et au grand public, et de la mise en place d'ateliers de co-réflexion du programme.

Les ateliers de co-réflexion du **PIREN-Seine** ont pour objectif de réunir chercheurs et opérationnels autour de sujets thématiques, afin de faciliter les échanges entre les différents acteurs de l'environnement de la Seine. Ces rencontres permettent d'établir ou de renforcer des bases de travail communes, pour que les recherches menées par les scientifiques du **PIREN-Seine** soient construites en cohésion avec les actions effectuées sur le terrain par les partenaires et les prises de décisions des pouvoirs publics.



*Les partenaires opérationnels de la phase 8 du PIREN-Seine*



*Les partenaires scientifiques de la phase 8 du PIREN-Seine*

