



HAL
open science

Une décennie de compétences collaboratives dans le cadre des 10 ans de Dialogu'IST

Sylvie Grésillaud, Fabien Borget, Magali Damoiseaux, Justine Fabre, Isabelle
Gras, Joanna Janik, Marie Roger-Chantin, Claire Tignolet

► **To cite this version:**

Sylvie Grésillaud, Fabien Borget, Magali Damoiseaux, Justine Fabre, Isabelle Gras, et al.. Une décennie de compétences collaboratives dans le cadre des 10 ans de Dialogu'IST. 2024. hal-04586604

HAL Id: hal-04586604

<https://hal.science/hal-04586604v1>

Submitted on 24 May 2024

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Dialogu

IST

Information
Scientifique et
Technique

10
ANS !



Une décennie de compétences collaboratives dans le cadre des 10 ans de Dialogu'IST

30 novembre 2023

Dialogu'IST est une action de Renatis, réseau national des professionnels de l'information du CNRS.



Le 14^e atelier Dialogu'IST porte sur les compétences collaboratives nécessaires pour structurer, partager et ouvrir les données de la recherche. Il a lieu le 30 novembre, date anniversaire pour célébrer ses 10 ans d'existence et s'est déroulé exceptionnellement en présentiel, à Marseille, son lieu de naissance. L'objectif de cet atelier est d'apporter un éclairage à la fois rétrospectif et actuel sur les modalités de travail, les rôles des différentes parties prenantes, ainsi que les métiers complémentaires et indispensables pour valoriser l'ensemble de la production scientifique dans l'esprit de la science ouverte.

La journée est notamment rythmée par une table ronde sur le thème "Qu'est-ce que la Science ouverte a changé dans votre quotidien ?", par des anecdotes et des retours d'expérience autour de collaborations fructueuses. Celles-ci représentent de nouvelles opportunités qui offrent de nouveaux espaces d'échanges et de collaborations au niveau national et international.

Programme et supports de présentation sur <https://renatis.cnrs.fr/a-noter-les-10-ans-de-dialoguist-le-30-novembre-2023>

Introduction de la journée par Martina Knoop (CNRS MITI) et Fabien Borget (chargé de mission Science Ouverte, AMU)

En 8 ans, depuis la mise en œuvre de la loi pour une République numérique (2016), des progrès énormes ont été faits en matière de science ouverte, notamment sur la question de l'accessibilité des publications et de la protection et droits des auteurs. Dans 5 ans, nous ne parlerons peut-être plus de science ouverte car ce sera le principe par défaut (en étant optimiste). Après le travail de sensibilisation qui a largement été fait, la démarche est axée sur la mise en place de services. Aujourd'hui, l'enjeu consiste à aller au-delà de la publication, afin notamment d'en libérer et partager les fondements (les données) et les outils (les protocoles ou le code-source).

Étude sur les métiers et activités de la donnée (MESR) – Isabelle Blanc

Cette étude menée début 2023 s'articule autour des **données, algorithmes et codes sources**. Les « données de la recherche » sont décrites comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique (OCDE 2007) et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. On englobe aujourd'hui sous le terme « écosystème » l'ensemble des outils, actions et paramètres nécessaires pour retrouver et réutiliser ces données (gain de temps et réduction de l'impact numérique).

Une réflexion est actuellement en cours sur **les activités et métiers** nécessaires au bon fonctionnement de cet écosystème. Un aspect important est la question de **la formation** : quelles sont les formations diplômantes ? Quelle est la place de l'ingénierie autour de l'IA ? Dans la filière IST, le besoin d'accélérer la formation se fait ressentir. Les problématiques d'intégrité et de déontologie sont également à prendre en considération. Il est important de développer le diplôme universitaire et la formation continue tout en s'appuyant sur DoraNum et les Urfist (feuille de route du parcours pédagogique pour la formation continue).

L'étude souligne la **diversité des métiers sur la donnée**, par exemple intendant de la donnée, data steward (mais quel périmètre ?), data librarian, data curator, profil à distinguer du metadata curator (plus proche du data librarian). En effet ces métiers interviennent sur différentes phases du cycle de vie de la donnée. Les métiers de la BAP E (Informatique, Statistiques et Calcul scientifique) semblent historiquement et par nature les plus appropriés pour prendre soin des données. Mais de plus en plus de professionnels de l'information scientifique et technique, qui relèvent plutôt de la BAP F, sont confrontés à ces enjeux et développent de nouvelles compétences ou s'efforcent de le faire. La complémentarité entre métiers est nécessaire pour traiter correctement la donnée, d'où qu'ils proviennent (le labo, une infrastructure de recherche, une BU...). Dans la sphère publique, l'INSEE, où les gestionnaires des BDD maîtrisent les enjeux et sont formés aux enquêtes, démontre sa force.

Pour l'ESR, les enjeux se situent dans la **gestion, la préservation, la circulation et l'ouverture des données**. Il importe aussi de développer et de reconnaître ces compétences côté enseignants-chercheurs, personnels d'appui et administratif, ce qui passera par l'identification, le soutien et l'amplification des formations existantes.

L'objectif est de concevoir des **données de recherche réutilisables**, mais se pose la question de la qualité et de l'accès à ces données. Quelles sont les données existantes (dont les premiers bénéficiaires sont les chercheurs et les collègues du laboratoire) ? Le travail de FAIRisation des données est à faire le plus souvent dans le cadre de l'unité, par ceux qui traitent les données, au plus près de l'origine des données. Ensuite, la question de la **diffusion et du croisement des données** est primordiale.

Comment accompagner les chercheurs ? Il s'agit de les sensibiliser à la donnée pour la gérer, la conserver (détruire aussi ce qui est nécessaire) et de leur faire comprendre l'importance des métadonnées. Se met actuellement en place un écosystème qui repose sur des ateliers de la donnée,

des centres de ressources, des centres de référence thématique et sur la mise en réseau d'établissements via ces ateliers. La labellisation de ces structures sur le territoire est devenue un sujet majeur aux yeux des gouvernances, ainsi que pour les personnels concernés, qui ont un besoin croissant de partage et de coopération afin de mettre au point de nouvelles pratiques. Des réseaux se structurent également en marge de cet écosystème mais grâce à sa dynamique il s'opère une autoformation et une montée en compétences, qui font converger des profils jusqu'à dessiner de nouveaux métiers. Mais attention cependant, l'impulsion vient aussi des grands éditeurs commerciaux, qui sont les premiers à avoir saisi l'importance des données.

La journée s'est poursuivie avec une table-ronde sur le thème « Qu'est-ce que la science ouverte a changé dans votre quotidien depuis votre intervention à Dialogu'IST ? »,

réunissant 4 participants qui sont intervenus à l'un des précédents ateliers : **Pascal Aventurier** (atelier #4), **Pierre-Yves Arnoud** (atelier #5), **Magali Moysan** (atelier #9) ainsi que la Directrice Adjointe Scientifique Calcul Intensif et Données (IN2P3) **Sabine Crépe-Renaudin**. Tous ont évolué depuis leur intervention et ont totalement intégré la science ouverte à leurs missions. Il ressort des échanges l'importance du savoir-être, de l'écoute et du savoir travailler en co-construction. Plus particulièrement, il est nécessaire de savoir traduire les besoins des chercheurs, de participer à des groupes de travail, d'être prospectifs sur de nouveaux métiers et multicompetents de façon à promouvoir les données au même titre que les publications scientifiques pour de bonnes réutilisations. Les métiers liés au catalogage ou aux référentiels sont toujours d'actualité. Il y faut ajouter ceux liés à l'archivage à long terme tout en tenant compte des domaines scientifiques concernés. L'objectif du collectif est de faire une science plus éthique et plus transparente.

La seconde partie de la journée était consacrée à des **retours d'expériences de collaborations réussies** :

- DoreMITI, l'atelier Données, groupe de travail inter-réseaux de la MITI,
- Un exemple de collaboration autour de la donnée territoriale : la plateforme connaissance du territoire en région Provence-Alpes-Côte d'Azur
- Un exemple de collaboration à l'échelle européenne, Skills4EOSC ou comment faire progresser les compétences en science ouverte en harmonisant le paysage actuel de la formation dans un écosystème européen commun et fiable.

Ces témoignages ont permis de dialoguer autour de plusieurs questions.

Comment travailler ensemble en fonction des méthodes suivies ? Le même niveau de langage pour une approche commune.

Le point départ de notre action est la **science ouverte (rôle politique)**, il est important qu'elle soit comprise et intégrée par les décideurs mais aussi sur le terrain et voir cela comme un aspect stratégique.

Pour pouvoir dialoguer il faut s'accorder sur le **vocabulaire employé**. Pour travailler ensemble il faut avoir du savoir-être et être à l'écoute des besoins des équipes scientifiques (rôle des ingénieurs de projet).

La gestion de la donnée nécessite un **dialogue avec d'autres métiers/compétences** (la qualité de la donnée avec la description des métadonnées par exemple, avec l'appui des archivistes), il faut savoir traduire les demandes des enseignants-chercheurs, travailler au sein de groupes de travail, mener une veille sur les compétences à développer, **être prospectif sur les nouveaux métiers**.

Il est à noter un exemple d'initiative commune : le guide des bonnes pratiques sur la gestion des données de la recherche (<https://gt-atelier-donnees.miti.cnrs.fr/guide>) et la carte des actions métiers (<https://gt-atelier-donnees.miti.cnrs.fr/download/GTInterreseaux-CartoSyntheseV6-optimise.pdf>) réalisés par DoreMITI, l'atelier Données, groupe de travail inter-réseaux de la MITI. On peut également citer le dispositif de coordination, d'animation et de mutualisation des données et études en PACA, qui est un exemple de stratégie régionale de la connaissance au service des territoires (<https://connaissance-territoire.maregionsud.fr/la-plateforme/les-donnees-ouvertes-regionales>). La donnée doit être travaillée de manière **interdisciplinaire**.

Problématiques communes rencontrées

La question de la **réutilisation des données** est cruciale (via des indicateurs qualitatifs ?). L'enjeu de demain, c'est la réutilisation via des algorithmes d'IA et/ou par des processus d'apprentissage profond (deep / machine learning) ; il faut dès lors uniformiser les référentiels des données. **L'exploitation et la réexploitation** posent la question de la **souveraineté** et de la **propriété des données** dans un cadre transnational, la question de la gouvernance et le rapport entre les données et la machine. Cette question est soulevée par différents acteurs, notamment la région Sud qui travaille sur la donnée territoriale. Il faut croiser les données et les expériences et mutualiser cela pour l'avenir. L'idée étant de fédérer les territoires grâce à un réseau et des moyens mutualisés, l'accompagnement à l'ingénierie et mettre en place des solutions pour la réutilisation des données ouvertes. L'objectif est de valoriser et de mutualiser nos connaissances, ce qui pose également la question de la **conservation à long terme** (plus de 5-10 ans) et d'une science plus éthique et plus transparente.

Les échanges ont mis en évidence la **problématique du recrutement** et de la formation. Il est difficile de recruter des personnes qui maîtrisent toute la chaîne de l'IST (catalogage, bibliométrie, publications, etc.) tout en ayant des compétences en matière de données.

L'exemple d'EOSC à l'échelle européenne permet de mettre en évidence trois lacunes :

- Le manque **d'expertise en science ouverte et en science des données** : il faut créer une expertise reconnue
- Le manque de définition claire **des expertises métiers et de parcours professionnels** correspondants : création de « MVS » Minimum Viable Skillsets
- Fragmentation dans les **ressources de formation** : l'objectif est l'harmonisation des formations, la production de supports FAIR et durables mais aussi de s'assurer de la qualité et de la certification des formations professionnelles (<https://zenodo.org/records/8305482>).

Comment s'emparer de la question des données pour la recherche (différence d'échelle / vision transversale de la gestion des données)

Pour pouvoir travailler sur cette question il faut faire le lien entre les données et les thématiques ainsi que **travailler de manière interdisciplinaire**. On note des différences d'échelle (mises en évidence par les intervenant.es au cours de la journée) : locale (unité/laboratoire/région/réseaux/ateliers) et les politiques gouvernementales nationale et internationale. On note l'importance des réseaux d'acteurs qui effectuent cette mise en réseau conjointement avec des actions de coordination.

Plus spécifiquement, le groupe de travail inter-réseaux de la MITI rassemble par exemple différentes personnes d'horizons divers ce qui a pour avantage :

- La création d'une complémentarité entre les activités métiers
- Une meilleure connaissance des services et donc une montée en compétences

- Une bonne entente et un esprit collégial du réseau
- Un pilotage du réseau induit par des objectifs partagés.

L'exemple d'EOSC à l'échelle européenne pose la question de l'appréhension des données en mettant en évidence le **besoin d'harmonisation des formations** pour s'emparer de la question des données afin de les transposer dans d'autres contextes et de mieux les appréhender. Il faut proposer des plans de formation basés sur des profils métiers et donc une **offre de formation uniformisée** (exemple des Compétence centers vers un écosystème de formation pour la science ouverte). Les processus doivent aussi être expliqués pour améliorer le dialogue **science et société**.

Enfin, il est important de rappeler que toutes les compétences sont utiles pour se saisir de cette question, cela permet de ne pas être uniquement des usagers de la donnée mais aussi de maîtriser toute la chaîne de production jusqu'au réemploi et l'archivage pérenne (la donnée est considérée comme souveraine quand elle est partagée/acceptée entre les acteurs). Il est nécessaire d'aller au-delà et d'avoir un regard de recherche pour travailler sur un scénario adapté.

Conclusion de la journée

La conclusion d'un tel atelier n'est pas une tâche simple, en effet, on s'est à nouveau rendu-compte que le contexte de la science ouverte a permis une évolution importante des pratiques, des compétences, des métiers, des approches face à la donnée de recherche, face à la publication, aux types de publications et de productions, aux collaborations... Notre atelier a, depuis son origine, accompagné ces évolutions en permettant des retours d'expériences à l'état de l'art.

Cet atelier, en particulier, nous a permis d'illustrer à nouveau que les structurations nationales, avec les ateliers de la donnée et leur déclinaison locale qui sont entrepris sur différents sites permettent de faire collaborer et de mettre en synergie l'ensemble des métiers permettant à la donnée de recherche de se FAIRiser. En effet, ces ateliers sont réellement les endroits où les spécialistes de l'IST, les informaticiens, les référents juridiques et le chercheur peuvent se rencontrer. Cette innovation n'est-elle pas la garantie de l'intégrité scientifique ?

Mais ces mises en réseau que permettent ces ateliers et ces structures n'ont d'intérêt que par les femmes et les hommes qui les font fonctionner, les animent et permettent finalement cette circulation de compétences pour tendre vers une science plus ouverte et transparente.

Toutes ces structures qui mettent en connexion toutes les compétences propres au cycle de vie de la donnée n'assureront leur succès que si elles bâtissent les conditions du dialogue constructif entre tous les intervenants en plaçant l'humain au cœur du processus : cela passe nécessairement par des sensibilisations, des formations, des échanges. Et ceci ne pourra être garanti qu'à travers une approche interdisciplinaire où le partage des compétences est la base de la réussite.

Ce retour sur 10 ans d'échanges nous montre à nouveau que nos métiers évoluent très vite et que ces transformations sont conçues pour que la science et la recherche retrouvent leurs fondements de base et permettent de réunir à nouveau les conditions d'une confiance retrouvée de la part de la société.

Ce retour sur les 10 ans de Dialogu'IST appelle à poursuivre les constructions en cours et à en développer de nouvelles. En effet, à travers cette décennie de souvenirs, cet atelier nous ouvre des perspectives pour mieux préparer l'avenir.