

Recherche Data Gouv : « une vue d'ensemble du patrimoine de données ouvertes partagées » (S. Retailleau)

News Tank Éducation & Recherche -
Paris - Actualité n°258054 - Publié le 08/07/2022 à 17:56

Imprimé par Xavier Teissedre - abonné #13929 - le 13/07/2022 à 10:20



Sylvie Retailleau à l'Université Paris Cité, le 08/07/2022 - © D.R.

« J'ai l'honneur aujourd'hui d'inaugurer le lancement de Recherche Data Gouv pensé pour soutenir ces équipes de recherche qui en ont tant besoin dans leur travail de structuration des données pour les rendre faciles à trouver, accessibles, interopérables, réutilisables, c'est-à-dire conformes au principe dit Fair (Facile à trouver, accessible, interopérable, réutilisable) », déclare [Sylvie Retailleau](#), ministre de l'ESR (Enseignement supérieur et recherche), lors de l'inauguration de Recherche Data Gouv à Université Paris Cité, le 08/07/2022.

Recherche Data Gouv :

- « généralise ce qui se faisait déjà dans certaines disciplines, en s'appuyant sur des communautés d'établissements déjà organisées et en intégrant ceux qui veulent se lancer dans le développement de dispositifs d'accompagnement » ;
- « fédère, amplifie et structure de manière complémentaire ces dispositifs : c'est le point d'accès vers tout l'écosystème français de service et d'accompagnement pour ces systèmes de partage et d'ouverture des données », précise-t-elle.

Au-delà de l'enjeu de préparation des données, la ministre évoque aussi le « besoin de mutualiser certains coûts : Recherche Data Gouv offre un entrepôt souverain de données pluridisciplinaires à toutes nos communautés de recherche qui sont pour l'instant dépourvues de solutions pour partager, pour justement ouvrir aussi leurs données ».

Et Sylvie Retailleau de conclure : « Avec Recherche Data Gouv, la recherche française se dote d'une vue d'ensemble de son patrimoine de données ouvertes et partagées. Et je voudrais en particulier remercier l'[Inrae \(Institut national de recherche pour l'agriculture, l'alimentation et l'environnement\)](#) d'avoir accepté de faire bénéficier à tous de son expérience en la matière, en transformant sa plateforme en une solution nationale capable de fédérer nos

données ouvertes de recherche ».

Le portail [recherche.data.gouv.fr](#)

Inscrit dans le deuxième PNSO (Plan national science ouverte), ce projet collectif réunissant organismes de recherche et universités est, pour les trois premières années, piloté par le MESR et financé par le FNSO (Fonds national pour la science ouverte) à hauteur de 7 M€ ré-partis entre

- le développement de l'entrepôt et du catalogue de données
- et la mise en place d'ateliers de la donnée « au plus près des chercheurs, sur tout le territoire », indique le MESR.

Il met en œuvre un portail web - [recherche.data.gouv.fr](#) - révélé le 08/07 qui regroupe des informations sur les premiers services disponibles, sur les membres de l'écosystème, sur l'entrepôt pluridisciplinaire et sur le catalogue pluridisciplinaire des données de recherche rassemblant déjà plus de 21 000 fichiers de données partagés, selon le ministère.

Ce portail constitue une branche sectorielle de la plateforme ouverte des données publiques française [data.gouv.fr](#).

Pilotage par Inrae

Cet entrepôt pluridisciplinaire dont le développement a été confié à Inrae associé à sept autres établissements - Université de Grenoble Alpes, Université de Lille, Université de Lorraine, Université Paris Cité, Université Paris Nanterre, Université de Strasbourg et le CNRS -, « est une solution souveraine pour le partage et l'ouverture des données de la recherche ».

« Cette offre est destinée aux communautés qui ne disposeraient pas encore d'un entrepôt thématique reconnu. Il sera complété en 2023 par un catalogue de données permettant de repérer et signaler les jeux de données d'entrepôts externes nationaux ou internationaux. »

Trois dispositifs pour favoriser l'appropriation et l'usage du portail

« Ce portail en place, il s'agit d'en faciliter l'appropriation et d'en assurer l'usage, en s'appuyant sur trois dispositifs :

- les ateliers de la donnée ;
- les centres de référence thématiques ;
- et les centres de ressources », précise Sylvie Retailleau.

Ateliers de la donnée

Les ateliers de la donnée sont « des structures de proximité dans les territoires qui soutiennent les actions de laboratoires(...). Cette action de proximité est indispensable, illustrée aussi par tout ce qui se fait au niveau des établissements et des services.

Six premiers ateliers sont aujourd'hui labellisés(...) : ils seront rapidement rejoints dès cet automne par de nouveaux ateliers ».

Le MESR précise que huit autres ateliers de la donnée sont « sur la trajectoire de labellisation, répartis au sein des établissements d'enseignement supérieur sur le territoire, sélectionnés en juin 2022 par le Comité de pilotage de la science ouverte - pour apporter aux équipes de recherche une première expertise de proximité dans la gestion raisonnée de leurs données de recherche ».

Centres de référence thématiques

« Ces ateliers de la donnée s'appuient sur l'expertise des domaines scientifiques pionniers dans cette ouverture, dans ce partage des données [avec] six centres de référence thématique : Humanum, Progedo, l'institut français de bioinformatique, Data Terra, le Pôle national de biodiversité et le Centre de données astronomiques de Strasbourg », indique la ministre.

Ces six centres de référence thématiques « ont été identifiés parmi les infrastructures de recherche existantes pour diffuser les bonnes pratiques et les standards internationaux de gestion, traitement et diffusion des données dans leurs domaines scientifiques respectifs », précise le MESR.

Quatre centres de ressources complètent l'accompagnement local par des services en ligne et des formations.

Prochaines étapes et enjeux

Les prochaines étapes mèneront à des évolutions fonctionnelles de l'entrepôt de données, à l'ouverture ou au partage progressif de nouveaux jeux de données, et à la création, en 2023, du catalogue de données repérant et signalant les jeux de données disponibles dans des entrepôts thématiques nationaux et internationaux.

Le volet accompagnement des équipes de recherche se dotera quant à lui de nouveaux ateliers de la donnée au fil des appels à manifestation d'intérêt et des labellisations. Enfin, s'il est d'abord un écosystème qui structure à l'échelle nationale les services autour des données, l'ambition de Recherche Data Gouv est aussi de devenir un service de l'Eosc (European Open Science Cloud).

« Recherche Data Gouv constitue donc une des infrastructures de recherche de notre prochaine stratégie nationale et nous nous emploierons à ce qu'elle devienne un des services aussi de l'European open science cloud », confirme la ministre.

L'enjeu de la formation

Sylvie Retailleau précise que le GIS (Groupement d'intérêt scientifique) Urfist « engagé sur les enjeux de formation depuis plusieurs années proposera aussi des parcours de montée en compétence ». Elle souligne aussi « l'ambition et l'effort de structuration de ces compétences entre établissements auxquelles plus de 85 établissements ont répondu en un temps très court ».

La crainte d'être pillé

« J'entends encore parmi certaines communautés scientifiques la crainte d'être pillées en ouvrant ces données : il faut l'entendre et un article qui est enrichi par les données voit son taux de citation qui peut être augmenté en moyenne de 25 % », ajoute la ministre.

Elle précise aussi que « la science ouverte ne doit pas se limiter aux succès. Les échecs sont tout aussi importants puisque, renseignés, ils évitent à nos collègues finalement de s'y égarer ou d'y retourner à nouveau ; c'est aussi quelque chose qu'on pourrait apprendre à savoir faire (Comité Service et Infrastructure Numérique) dans l'évaluation et la valorisation. »

« Joindre le Health data hub avec le Recherche.data.gouv »

Christine Clerici, présidente de l'Université Paris Cité

« Il était important de disposer d'une plateforme nationale unique et simple d'utilisation qui comprend à la fois un entrepôt de données et un catalogue de données publiques. Ce type de plateformes sera sans nul doute une grande aide pour nos chercheurs dans l'organisation et l'optimisation de leurs recherches. »

« D'autres enjeux sont à venir : celui de la mutualisation des infrastructures, mais également la nécessité de complémentarité entre les infrastructures ; et peut-être mon côté médecin me fait dire qu'il faudrait à un certain moment arriver à joindre le Health data hub avec recherche.data.gouv : c'est peut-être complètement absurde mais j'assume complètement ».

« En tant que présidente d'université, je me dois de dire que de nombreuses questions subsistent au sein de nos établissements, notamment celles des investissements nécessaires à réaliser pour que ce tournant numérique soit pris et bien pris, tant en infrastructures qu'en maintenance et gestion ».

Plusieurs opportunités ouvertes par Recherche Data Gouv

Carole Caranta, DGD (Directeur(trice) général(e) délégué(e)) science et innovation d'Inrae

« Contribuer à Recherche Data Gouv est une des dix actions phares de notre plan d'action science ouverte. Nous avons pu relever ce défi de rendre accessible en moins d'un an Recherche Data Gouv pour le volet catalogue/entre�ot parce que le projet a clairement bénéficié de l'expertise de nos équipes. »

Selon elle, en mettant son expérience au service de ce projet national, Inrae y a aussi vu plusieurs opportunités :

- « fédérer les expériences et les compétences avec la possibilité de développer des services plus riches, évolutifs et plus robustes : il a d'ores et déjà été possible de mettre en place une sauvegarde déportée dans les infrastructures de l'Unistra (Université de Strasbourg) afin de renforcer la sécurisation des données » ;
- « l'adoption de pratiques communes sur le dépôt et la publication des données avec le développement de chartes, de guides et de formations adaptées grâce à une diversité de compétences et de points de vue » ;
- « répondre aussi à des besoins de pérennité de ce type d'infrastructure assurée par plusieurs partenaires » ;
- « découvrir les données des autres et favoriser le partage et la réutilisation de ces données avec un large panel d'acteurs » ;
- « répondre aux enjeux de visibilité des données produites par la recherche française notamment au niveau européen en se connectant dans l'Eosc ».

« Ce projet va encore se poursuivre pendant deux ans et nous aurons évidemment besoin de tous les partenaires au-delà de la phase projet : nous ne sommes finalement qu'au début de cette formidable aventure ! ».

« Réussir cela en moins d'un an c'était un challenge ! »

Antoine Petit, P-DG du CNRS (Centre national de la recherche scientifique)

« Avant tout, je voudrais qu'on se félicite d'être allés aussi vite : on a été super bons puisqu'on a quand même décidé cela en juillet 2021 et moins d'un an après on est là tous réunis. »

« On a positionné cinq centres thématiques, cinq grandes infrastructures dont le CNRS est tutelle ou cotutelle, qui figurent sur la feuille de route des infrastructures de recherche publiée par le MESR, dans l'écosystème de la plateforme Recherche Data Gouv :

- Le Centre des données astronomiques de Strasbourg ;
- Data Terra pour les sciences de la Terre ;
- l'IFB (Institut français de bioinformatique) pour la biologie ;

- *Humanum et Progedo dans le domaine des SHS (Sciences humaines et sociales)* »

« Nous travaillons également avec l'IN2P3 (l'Institut national de physique nucléaire et de physique des particules) qui est un très gros producteur de données ».

« Des moyens humains ont été engagés directement dans la réalisation de la plateforme et nous avons proposé le pilotage ou copilotage de deux centres de ressources :

- le premier s'appellera Opidor (Optimisation du partage et de l'interopérabilité des données de la recherche) et s'appuiera sur l'expérience de l'équipe éponyme de l'Inist (Institut de l'Information Scientifique et Technique du CNRS) reconnu depuis longtemps pour son expertise et son aide aux chercheurs dans la réalisation des DMP (Data management plan) et l'attribution d'identifiants pérennes pour les données (...);
- le 2^e est le Centre de ressources Doranum (Données de la recherche : apprentissage numérique à la gestion et au partage) dont l'équipe de l'Inist est reconnue pour avoir développé des modules de formation disponibles sur internet autour d'un outil appelé lui aussi Doranum, qui sera largement utilisé dans le cadre des centres de ressources et proposera des supports de formation copilotés par liste et les Urfist (Unité Régionale de Formation à l'Information Scientifique et Technique). »

« Le choix de ne pas créer notre propre atelier de la donnée »

« Enfin, pour les ateliers de la donnée, on a fait le choix de ne pas créer notre propre atelier, mais de participer en tant que partenaire aux diverses initiatives qu'elles soient locales ou régionales. C'est le cas par exemple avec l'Université de Grenoble Alpes, l'Université de Lorraine et l'Université de Strasbourg pour les ateliers déjà labellisés. Mais nous sommes également partie prenante dans les initiatives retenues comme étant en voie de labellisation à Lille, Toulouse, Montpellier et dans la Région Centre Val de Loire », ajoute Antoine Petit.

Il rappelle enfin que le CNRS a été « un peu précurseur » en matière de gouvernance pour la science ouverte en créant la Ddor, la direction des données ouvertes de la recherche qui réunit les problématiques de sciences ouvertes mais aussi de calculs intensifs et des données de la recherche. « Si je peux me permettre, ce serait presque un peu comme si on réunissait le Coso (Comité pour la science ouverte) et le Cosin... je dis ça, je ne dis rien... ».

« Dorel était quand même un petit peu trop avancé pour qu'on puisse l'arrêter »

Hélène Boulanger, présidente de l'Université de Lorraine

« La dynamique des sciences ouvertes est en train d'emporter une mutation de nos activités fondamentales dans les universités et les organismes nationaux de recherche comme on en a rarement vu ces derniers temps. »

« Une université toute seule ne peut pas porter une transformation aussi importante à son échelle : l'impératif de mutualisation s'impose aux établissements, aux universités sur le sujet de la science ouverte encore plus que sur les autres. »

« Mais au moment où on a aperçu l'émergence de Recherche Data Gouv, l'entrepôt lorrain de données Dorel était quand même un petit peu trop avancé pour qu'on puisse l'arrêter et basculer sur le projet national. Toutefois, le travail en commun a permis d'assurer une forme de compatibilité des dispositifs : aujourd'hui Dorel est inclus dans Recherche Data Gouv et nous visons en 2024 son intégration complète. »

UGA : « le choix d'arrêter le déploiement de son entrepôt local »

Isabelle Olivier, VP (Vice-président(e)) numérique d'UGA (Université Grenoble Alpes)

« Au printemps 2021 a été lancé le projet Recherche Data Gouv avec une plateforme fédérée des données de la recherche, et en mai 2021 l'Université Grenoble Alpes a fait le choix d'arrêter le déploiement de son entrepôt local pour rejoindre ce projet national. »

« Cette décision a été prise assez facilement et rapidement, car notre université a immédiatement perçu un intérêt à pouvoir bénéficier d'une collection spécifique UGA au sein d'un entrepôt de données au niveau national. Nous investissons d'ailleurs activement des forces dans ce projet national puisque quatre personnes de l'UGA y participent. »

« Ce projet national apparaissait plus raisonnable en termes de délai, de temps de travail nécessaire, mais aussi d'efficacité. ».

« La cohérence de la cellule locale Grenoble Alpes(...) nous a permis d'avoir une reconnaissance via la labellisation atelier de la donnée ».

« Le point important au niveau du site, c'est surtout l'aspect formation »

Rémi Barillon, VP Recherche, formation doctorale et sciences ouvertes à l'Unistra

« Notre atelier des données a été labellisé et donc soutenu par le ministère c'est quelque chose d'extrêmement important. Les premières réunions ont eu lieu en juin dernier : on a lancé cet atelier qui est maintenant opérationnel. »

« Le point important au niveau du site, c'est surtout l'aspect formation, car la science des données ce n'est pas juste un changement de pratiques, mais c'est une vraie révolution. Cela va nous permettre d'aller beaucoup dans notre propre recherche disciplinaire, mais

également dans de nouveaux champs dans le domaine interdisciplinaire en partageant les données. »

L'enjeu est selon lui « *d'acculturer tous les étudiants, et pas seulement les doctorants, dès la première année à partager leurs données* ».



Ministère de l'enseignement supérieur et de la recherche (MESR)

Catégorie : Etat

Adresse du siège

21, rue Descartes
75005 Paris France

Général

| | |
|---------------------------|---|
| Date de création | 2022 |
| Statut | Ministère |
| Implantations, dont siège | Paris (siège) |
| Missions et objectifs | Prépare et met en œuvre la politique du Gouvernement en matière d'enseignement supérieur et de recherche. |
| Composantes | <ul style="list-style-type: none">• Secrétariat général ;• Dgesip (direction générale de l'enseignement supérieur et de l'insertion professionnelle) ;• DGRI (direction générale de la recherche et de l'innovation) ;• Igésr (Inspection générale de l'éducation, du sport et de la recherche). |
| Direction | Directrices générales : <ul style="list-style-type: none">• recherche et innovation : Claire Giry (depuis le 26/05/2021).• enseignement supérieur et insertion professionnelle : Anne-Sophie Barthez (depuis le 17/07/2019). |
| Secrétaire général(e) | Marie-Anne Lévêque |
| Ministre | Sylvie Retailleau (depuis le 20/05/2022) |

Fiche n° 13526, créée le 23/05/2022 à 14:42 - Màj le 02/06/2022 à 09:35

© News Tank Éducation & Recherche - 2022 - **Code de la propriété intellectuelle** : « La contrefaçon (...) est punie de trois ans d'emprisonnement et de 300 000 euros d'amende. Est (...) un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une oeuvre de l'esprit en violation des droits de l'auteur. »