

DIALOGU'IST#10 De la politique à l'opérationnel, 12 février 2021

Foire aux questions

Vous trouverez ci-après des éléments de réponses aux questions posées aux intervenants.

Interventions de Odile Contat et de Claire Leymonerie (MESRI)

Une question portant sur la signification du GIS - Groupement d'Intérêt Scientifique - nous amène à vous indiquer l'article « Le CNRS simplifie ses structures de recherche » que vous pouvez trouver ici :

<https://www.cnrs.fr/fr/cnrsinfo/le-cnrs-simplifie-ses-structures-de-recherche>

Qu'appellez-vous structure éditoriale ?

Les structures éditoriales sont les maisons d'édition privées ou publiques, établissement ou universités ou associations éditeur.

Service générique d'accueil et de diffusion de données simples

« Le groupe « [Service générique d'accueil et de diffusion de données simples](#) » du Comité pour la science ouverte a pour objectif d'évaluer la faisabilité d'un entrepôt générique national. »

Ce service sera dédié aux données pour lesquelles aucun des entrepôts disciplinaires existants ou à venir ne constituerait une solution adaptée. Ce service n'a pas de nom particulier pour le moment.

Plus d'informations sur les entrepôts des données :

<https://www.ouvrirlascience.fr/etude-comparative-des-services-nationaux-de-donnees-de-recherche-facteurs-de-reussite/>

Réseau d'administrateurs de Données et Codes

Comment se déclinera le Réseau d'administrateurs de Données et Codes au CNRS? Par instituts ? Par région ? Ou transversalement ?

Isabelle Blanc a été nommée Chief Data Officer (administratrice ministérielle de données de la recherche) au Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Chaque établissement est amené à nommer un administrateur de données et des codes sources. Sylvie Rousset est l'administratrice de données qui représentera le CNRS au sein du réseau d'administrateurs de données de la recherche en cours de constitution. L'objectif est de mettre en place la gouvernance et la valorisation de des données de la recherche.

Un baromètre de la science ouverte existe-il au niveau des établissements ?

Le code-source du Baromètre national étant en libre accès, certains établissements s'en sont emparés pour construire leur propre baromètre institutionnel. C'est le cas par exemple de l'Université de Lorraine, de l'Université de Strasbourg ou de l'Université Grenoble Alpes.

Le code est accessible sur gitlab :

https://gitlab.com/Cthulhus_Queen/barometre_scienceouverte_universitedelorraine/-/blob/master/nettoyage_donnees.ipynb

En Lorraine, à Clermont, Brest et Strasbourg on travaille à la mise en place d'indicateurs sur les APC à inclure dans les baromètres.

Pour plus d'infos voir ici : <https://openapc.net/>

Et ici : <https://www.couperin.org/site-content/261-a-la-une/1341-de-nouvelles-donnees-apc-pour-la-france-fournies-par-couperin-org>

Voir aussi le Guide pratique pour le suivi des dépenses de frais de publication de Couperin (<https://www.couperin.org/images/stories/AO/2020-01-22-COUPERIN-Guide-gerer-APC.pdf>)

Question au sujet des modèles économiques liés aux acquisitions de ressources. Entre les abonnements, les APC et les modèles dits transformant, quel est la position du MESRI sur le débat qui agite actuellement la communauté ?

L'ouvrage de Nicolas Chevassus-au-Louis, "Malscience. De la fraude dans les labos" :

<https://www.seuil.com/ouvrage/malscience-nicolas-chevassus-au-louis/9782021175950>

HAL

Pourquoi dupliquer les dépôts (les ressaisir dans HAL) lorsqu'ils sont déjà sur une autre plateforme ouverte ? (Augmentation des coûts, impact environnemental)

La motivation première c'est l'archivage pérenne. Il y a un risque de disparition de certaines archives). HAL garantit l'archivage pérenne. Le CNRS travaille sur l'amélioration de l'interopérabilité entre les systèmes pour ne pas obliger les chercheurs à dupliquer leurs dépôts.

Question à propos de la définition des données de la recherche : pour les mathématiques, que peut-on considérer comme données de la recherche ?

Tout élément qui permet de faire un raisonnement scientifique : l'algorithme, la méthode etc.

Concernant les données dans des disciplines comme les maths, il y avait eu une discussion intéressante sur la liste de diffusion gestion des données du JISC. Les archives sont publiques (objet de la discussion "I have no data").

Quelle politique pour l'affectation des RH associés à ces nouveaux métiers pour l'accompagnement sur le terrain : au sein des laboratoires/UMR ? ou au sein des UMS ?

Tout reste à faire. Le Copil de la DDOR discutera de ces questions. Une enquête menée par la DDOR auprès des directeurs d'unité a démontré que plus de 60% d'entre eux ne disposent pas de personnels dédiés. Il sera important de s'entendre au niveau national avec d'autres établissements pour mettre en commun les ressources locales.

Peut-on en savoir plus sur le profil "data steward" ?

Un « data steward » est un « coordonnateur des données ». Sa mission est la gestion des données dans son ensemble – il/elle veille à la qualité des données, s'occupe de leur stockage et de leur accessibilité.

Pour plus d'infos voir : les démarches initiales au Royaume Uni (université de Cambridge) et aussi aux Pays Bas TU Delft, ou encore Ecole CODATA.

Les Data steward sont les personnes ressources de différents profils (génériques ou spécialisés), pour accompagner les scientifiques en proximité. Ces profils existent déjà à l'INRAe (aussi Data Champions).

<https://openworking.wordpress.com/data-stewardship/>

Qu'est-ce qu'un data paper ?

Un article qui rassemble des jeux des données mis à disposition à la communauté.

Ci-dessous quelques ressources :

- Un récent webinar tenu par le GTinter réseaux Atelier Données a été consacré à la question des data papers.

<https://mi-gt-donnees.pages.math.unistra.fr/site/index.html>

- Et aussi le sujet de data papers et abordé dans le Guide des bonnes pratiques sur la gestion des données de la recherche :

<https://mi-gt-donnees.pages.math.unistra.fr/guide/07-publier.html#publier-un-datapaper-pour-valoriser-et-expliciter-les-donnees>

<https://ist.inrae.fr/produit/data-papers-formation/>

Comment la DDOR et/ou le MESRI voient les choses pour ne plus prendre en compte les indicateurs de revues (impact factor) dans les évaluations des chercheurs ?

En pratique, les chercheurs, notamment les doctorants, sont "contraints" de choisir des revues avec fort impact pour leur évaluation, leur carrière et les recrutements.

Les scientifiques expérimentés doivent donner l'exemple pour amorcer un cercle vertueux.

Intervention de Fabien Borget, enseignant-chercheur, chargé de mission « Science ouverte » à Aix-Marseille Université (AMU)

Combien d'actions liées à la science ouverte avez-vous dans votre plan d'actions HRS4R ?

Plus de 100 actions ont été proposées par les 5 Groupes de Travail.

En savoir plus : <https://www.univ-amu.fr/fr/public/plan-dactions-hrs4r-et-mise-en-oeuvre>

Qu'est-ce qui a incité la gouvernance d'Aix-Marseille Université à mettre en valeur la science ouverte ?

Le tournant vers les publications ouvertes a été pris dès 2014 avec une politique incitative auprès des collègues pour déposer leurs publications dans HAL-AMU. La science ouverte s'impose également aujourd'hui comme le garant d'une science qui s'adresse au citoyen et qui rétablit la relation de confiance.

De plus, le contexte national et international, notamment les injonctions de la communauté européenne et de l'ANR pour la mise en œuvre de plans de gestion de données obligatoires permet d'accélérer la transition débutée. Aujourd'hui, nous sommes de plus impliqués dans CIVIS, une université civique européenne, <https://www.univ-amu.fr/fr/public/civis-une-universite-civique-europeenne>, et les valeurs de la science ouverte sont inscrites dans la dynamique européenne. Notons également la volonté de l'établissement d'obtenir le label européen HRS4R qui possède un volet important dans le cadre de la science ouverte.

Quelles sont les "mesures incitatives" pour déposer dans HAL ?

La mesure est le Bonus Qualité Recherche attribué aux laboratoires en fonction de leur production déposée dans HAL AMU (bonne affiliation et texte intégral). Le Bonus peut représenter jusqu'à 10% du montant du budget alloué au laboratoire par l'Université

En savoir plus : <https://bu.univ-amu.libguides.com/openaccess/bonus>

Comment s'articule l'interaction AMU - UMRs/laboratoires ? Services communs ouverts et gratuits (stockage, calcul) ? Conventions ? Projets communs ? Data center certifié ?

Pour le moment le data center ne se veut pas comme un entrepôt institutionnel mais comme la possibilité d'amener une offre de service. Ce centre, à vocation régionale impliquant de nombreux partenaires dont l'Université de Nice, l'Inserm... est en cours de déploiement et va permettre la mise en place de solution de stockage et de calcul. Cela permettra le développement de nouveaux services en relation avec la science ouverte.

Dans la stratégie développée pour la relation avec les laboratoires, nous comptons nous appuyer sur des projets PIA 3 structurants obtenus par l'Université ainsi que sur un réseau de collègues sensibilisés au sein des laboratoires permettant d'identifier un interlocuteur particulier capable d'aider au montage d'un projet incluant le cadre des actions science ouverte.

Avez-vous un entrepôt institutionnel AMU ?

<https://hal-amu.archives-ouvertes.fr/>

Au sujet de l'offre de formation, quel service s'en charge au niveau de l'AMU ?

Le Service Commun à la Documentation effectue des interventions ponctuelles à la demande dans les laboratoires et régulières au niveau du collège doctoral.

En savoir plus : <https://oaamu.hypotheses.org/category/formation>

De plus nous allons développer une offre de formation Science Ouverte sur la base de ce qui est déjà proposé afin de toucher l'ensemble des publics concernés dès le master. Nous pourrions alors affiner la détection et le déploiement de nouvelles compétences et ainsi anticiper l'apparition de certains nouveaux métiers en relation directe avec la science ouverte.

Interventions de Laetitia Bracco et de Thomas Jouneau (Université de Lorraine)

Quelles métadonnées pour l'entrepôt des données à l'université de Lorraine (pluridisciplinarité) ?

Dataverse est basé sur les métadonnées Datacite. Des schémas complémentaires (DDI pour les sciences sociales par ex.) sont également "livrées" avec le logiciel. Nous réfléchissons à la prise en compte de métadonnées complémentaires mais pensons que la réflexion doit être collective.

Quelles sont les contraintes en taille des jeux des données dans l'entrepôt ?

L'entrepôt est adossé à un stockage de type S3 qui est le stockage préexistant de l'infrastructure UL. Il n'y a pas de limite "technique" de volume mais nous voulons le tempérer par deux mécanismes : déclenchement d'un accompagnement pour les jeux > 1Go (taille indicative) et la sensibilisation à la curation dans le cadre de l'accompagnement (laquelle aboutit souvent à une baisse en taille significative).

Quelles sont les missions/activités envisagées pour les ambassadeurs de données ?

L'idée est qu'ils soient des relais de communication du [CODR](#) (Comité Opérationnel pour les Données de la recherche) auprès de leurs collègues, qu'ils soient formés de manière à pouvoir répondre aux questions de 1er niveau sur les données. Dans l'idéal, qu'ils parlent plan de gestion de données à la machine à café ;-)

Est-ce que vous contrôlez les éventuels cas de dépôts multiples pour le même jeu de données (plusieurs DOIs pour un même jeu des données) ?

Nous cherchons à fournir un service si et quand c'est nécessaire. La pratique encouragée sera donc la représentation des jeux de données déposés ailleurs, plutôt qu'un nouveau dépôt.

Comment articulez-vous entrepôt local et entrepôts thématiques ? Ou comment ne pas dupliquer le travail, se retrouver avec beaucoup de data centres et se retrouver devant une offre multiple et complexe à terme et pas forcément très efficace ?

C'est une question importante qui appelle un complément à la réponse ci-dessus sur la duplication. Le fait est que les entrepôts thématiques et généralistes ne couvrent pas, loin de

là, l'ensemble des besoins disciplinaires de notre établissement, et qu'un espace institutionnel nous semble pouvoir apporter une plus-value pour la plupart des publications de données de l'UL. Il est possible que la couverture et le périmètre de l'entrepôt varient au fil du temps en fonction des nouveaux outils qui émergent (comme l'entrepôt national, par exemple). En l'absence de visibilité sur ce point, nous avançons sur un outil en misant sur son interopérabilité avec l'environnement actuel et futur des entrepôts de données.

Comment s'y retrouver parmi tous ces entrepôts pour conseiller au mieux ses chercheurs ?

Garder une harmonisation pour l'ensemble du laboratoire en choisissant un seul entrepôt pour tous et/ou ne pas hésiter à laisser le choix d'utiliser différents dépôts suivant les thématiques et les besoins des chercheurs.

Quelle articulation prévue entre HAL-UL et le Datverse UL ?

Il faut penser au lien réciproque dans les métadonnées dans un premier temps. Une veille sur les dépôts d'un côté comme de l'autre sera active en ce sens.

Success stories n°3 : comment vous êtes-vous formés pour les questions de droits sur les données ?

Une collègue de la MSH Lorraine, Florence Bouchet Moneret, est spécialiste des ces questions. Pour ma part c'est de l'auto-formation.

Combien de personnels (bibliothèque, direction du Numérique, etc.) ont été mobilisés pour la mise en œuvre de cet entrepôt de données ?

Actuellement 2 techniciens informatiques, et les personnes du CODR avec des quotités de travail qui varient selon les membres.

Intervention de Marie-Claude Quidoz (Centre d'Ecologie Fonctionnelle et Evolutive – CEFÉ)

Comment a été initié l'axe transverse au CEFÉ <https://www.cefe.cnrs.fr/> ?

C'est la plateforme SIE, dont je suis la responsable technique, qui a fait cette proposition lors de la réflexion de l'unité pour le projet HCERES 2021-2025.

Lors des premières réunions, la direction a été un peu sceptique, non pas sur l'intérêt de ce thème mais sur la forme car les axes transversaux étaient pensés pour compléter les thèmes des départements. Mais nous sommes arrivés à la convaincre de son intérêt et l'axe est présenté maintenant comme venant s'articuler avec chacun des quatre autres axes (<https://www.cefe.cnrs.fr/fr/recherche/axes-transversaux>)

Était-ce une demande de la direction ou une proposition des personnels ?

C'est une proposition qui a été portée par le personnel technique des plateformes. Il est d'ailleurs piloté par trois personnes techniques (une biologiste moléculaire, une documentaliste et une gestionnaire de données)

Quel temps cela vous prend-il par rapport à vos autres missions (avant création de l'axe) ?

Dans cette phase de lancement, le temps consacré à cet axe est variable selon les connaissances déjà acquises et le métier exercé des animatrices. Une fois la phase d'acquisition de connaissances faites, ça devrait représenter 10% par semaine.

Intervention de M. Burnel et de Mme Batifol et M. Johany

L'affiche « Cycle de vie des données » montrée dans la présentation est-elle en licence CC ?

L'affiche est en cours de finalisation et sera prochainement partagée !

L'action Qualinous « Retour d'expérience sur la mise en place d'une animation ciblée sur la qualité des données de la recherche au sein du Département "Sciences pour l'Action et le Développement (SAD)" de l'INRAe : <https://hal.inrae.fr/hal-02059782>

Intervention de Pierre-Yves Lacour

Référence de Pierre-Yves Lacour :

<https://commondatafr.wordpress.com/2018/05/30/pierre-yves-lacour-lexistence-precaire-des-donnees-scientifiques-pour-lhistorien/>

Dans quelle mesure le droit lié aux données diffère du droit d'auteur (propriété, diffusion, moral etc.) ou au contraire s'en rapproche ?

Le législateur n'a pas encore légiféré sur ces questions-là.

Je publie quels sont mes droits ?, <https://www.ouvrirlascience.fr/je-publie-quels-sont-mes-droits/>

Guide pour la recherche RGPD, février 2021,

https://inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf

Le guide "Ouverture des données de recherche. Guide d'analyse du cadre juridique en France" a été mis à jour (V2)., <https://datapartage.inrae.fr/Actualites/Guide-juridique-nouvelle-version>

<https://hal.inrae.fr/hal-02791224/document>

Des pistes de réponses : <https://mate-shs.cnrs.fr/actions/tutomate/tuto25-propriete-donnees-lionel-maurel/>