

Compte-rendu du Comité Opérationnel de l'Atelier de la Donnée Centre Val de Loire du 3 octobre 2023

Lieu : MSH Val de Loire et visioconférence

Horaires : 10h – 12h30

Présents : Azanza Anne (Directrice du SCD, Université de Tours), Thierry Bissi-Yandia (OSUC), Yohann Brossard (OSUC), Gérard Bruère (SCD, Université de Tours), Maryse Castan (Correspondante archive DR08 / Service Informatique DR08 CNRS), Mélanie Fauconnier (Service Partenariats, Innovations et Valorisation, Université de Tours), Jorge Fins (MSH Val de Loire), Cloé Freulon (Chargée des affaires juridiques / Déléguée à la protection des données, Université de Tours), Céline Girault (Archiviste / Direction Pilotage Prospective Amélioration continue, Université d'Orléans), Edouard Guitton (Correspondant données, INRAE), Florian Hivert (MSH Val de Loire), Maéva Icentilăi (ADCVL, MSH Val de Loire), Rémi Ossant (ADCVL, MSH Val de Loire), Clément Plancq (MSH Val de Loire), Nathalie Pothier (OSUC), Stéphanie Puel (SCD, Université de Tours), Laurence Rageot (MSH Val de Loire), Alain Ravaz (MSH Val de Loire), Xavier Rodier (MSH Val de Loire), Philippe Serisier (Directeur SCD, Université d'Orléans), Pascale Solon (SCD, Université d'Orléans)

Ordre du jour : 1. Le Comité Opérationnel et la gouvernance – 2. Le guichet unique – 3. L'actualité des groupes de travail – 4. PUI Loire Valley Innov' – 5. Inauguration de l'ADCVL – 6. Informations - Recherche Data Gouv

[Compte-rendu rédigé par Rémi Ossant et Maéva Icentilăi, relu et amendé par Laurence Rageot et envoyé aux membres du comité opérationnel le 25 octobre 2023]

1. Le comité Opérationnel et la gouvernance

Laurence Rageot rappelle que le comité opérationnel regroupe tous les membres des différents groupes de travail de l'ADCVL. Elle présente le schéma de l'organisation d'ADCVL reprenant la gouvernance (comité de pilotage, comité opérationnel, groupes de travail, bureau exécutif et réseau de correspondants science ouverte), les partenaires avec les services impliqués.

La mise en place d'un réseau de correspondants science ouverte est prévue pour faire le lien entre l'atelier et les chercheurs, avant sa mise en place il était d'abord nécessaire d'avoir une offre de service. Un bureau exécutif a été envisagé, sous la forme d'un comité opérationnel plus restreint en nombre de personnes, mais il n'a pas été mis en place pour l'instant.

Tous les groupes de travail se sont déjà réunis une première fois entre mars et juin 2023.

2. Le guichet unique

Rémi Ossant présente le site web qu'il a développé. Il répond à trois objectifs :

- Donner accès à l'outil de guichet unique qui permet aux chercheurs de poser des questions
- Permettre aux membres de l'atelier de la donnée, via ce même outil, de répondre aux sollicitations des chercheurs
- Apporter des informations et des ressources sur la science ouverte et les domaines d'actions de l'atelier de la donnée

Une page sur la science ouverte a été mise en place, elle détaille les intérêts et aspects pratiques de la science ouverte dans le domaine de la recherche scientifique. Une autre page “boîte à outil numérique” est également en ligne, elle met à disposition des ressources et des informations sur des sujets en lien avec les données de recherche et la science ouverte. Elle a également vocation à porter des ressources numériques créées par les partenaires de l'ADCVL (ex : SecNum de l'Université de Tours).

Laurence Rageot demande aux membres du comité opérationnel de faire un retour sur le contenu du site.

Rémi Ossant présente ensuite les fonctionnalités du guichet unique. Il se présente, pour les chercheurs, sous la forme d'un formulaire qui permet de poser une question aux membres de l'atelier de la donnée. Une fois cette question posée, le suivi se fait par des mails qui donnent accès à l'outil de guichet unique mais il n'y a pas de création de compte pour le chercheur.

Les membres de l'atelier de la donnée accèdent à l'outil de guichet unique en se connectant à la partie “administrateur du site”. L'outil permet de répondre aux questions des chercheurs, d'échanger entre membres de l'atelier sur une demande, d'attribuer une demande à un membre de l'atelier, etc. Des mails automatiques sont envoyés aux membres de l'atelier lors d'une demande. Ces notifications sont conditionnées au sujet de la demande : si elle s'adresse au groupe de travail auquel l'agent appartient alors il est prévenu sinon ce n'est pas le cas.

Les membres de l'atelier de la donnée ne sont pas enregistrés sur l'outil de guichet unique, ils le seront prochainement avec leur accord. La répartition des demandes se fait au sein des groupes en fonction des compétences de chacun et des établissements de rattachement.

Il y a une inconnue concernant le nombre de sollicitations que va recevoir l'atelier, Laurence Rageot a eu des retours d'autres ateliers disant que ce nombre était plutôt faible.

Anne Azanza souligne qu'un même genre d'outil a été mis en place dans le cadre du SCD de l'université de Tours mais qu'à force les demandeurs ont identifié les personnes pouvant répondre à leur question et s'adressent directement à elles. Dans le cadre d'ADCVL, il a été choisi que les échanges se fassent via le site pour éviter que les discussions ne se déplacent vers les adresses mails personnelles des membres de l'ADCVL.

Mélanie Fauconnier demande s'il serait pertinent de demander aux chercheurs qui sollicitent directement les membres de l'ADCVL pour des conseils, de faire une demande sur le guichet unique pour garder une trace. Il apparaît que l'université d'Orléans, l'INRAE et le CNRS n'ont pas de procédure spécifique pour ce type de demande en interne.

Nathalie Pothier demande si le réseau des correspondants a vocation à être intégré à l'outil de guichet unique, pour Laurence Rageot, ce n'est pas le cas, leur rôle est de sensibiliser au sein des institutions et orienter vers l'atelier.

Alain Ravaz demande si les questions / réponses ont vocation à être conservées pour alimenter une base de connaissances. En l'état les tickets sont conservés dans l'outil de ticketing mais ils n'alimentent pas une base extérieure, ce sujet va être exploré.

Pascale Solon pose la question de l'intérêt de fixer un délai de réponse après une question posée à l'ADCVL. Anne Azanza indique que dans le cas du SCD de l'université de Tours, le délai est de 48h (jour ouvrable). Nathalie Pothier indique que sur les questions juridiques : une réponse peut prendre un certain temps donc une annonce de délai peut être pertinente pour prévenir le chercheur. Mélanie Fauconnier propose d'ajouter un champ optionnel dans le formulaire pour indiquer une deadline ou un jalon temporel factuel lié à une demande. Pascale Solon met en avant la charte utilisateur de l'ENSIB comme exemple de ce qui pourrait être mis en place dans le cadre de l'ADCVL. Elle indique également que le SCD d'Orléans a mis en place une charte

de répondant qui développe les possibilités de l'outil, la question des données personnelles, etc. Ce même type de document va être mis en place pour l'outil de guichet unique de l'ADCVL à la fois pour les membres de l'atelier ainsi que les chercheurs.

Mélanie Fauconnier propose d'associer dans l'interface de l'outil de recherche le nom des membres à leur institution et leur fonction pour que les chercheurs sachent qui leur répond.

Pascale Solon propose de rendre possible pour le chercheur de poser une demande sur l'outil de guichet unique en s'inscrivant dans plusieurs catégories. La demande sera donc transmise aux membres de deux groupes de travail correspondant. Cette option va être ajoutée.

Alain Ravaz propose de mettre plus en avant le bouton pour atteindre le guichet unique. Anne Azanza rejoint cet avis, le guichet unique apparaît comme la plus-value du site et cela doit ressortir.

Le service communication de l'Université de Tours travaille actuellement avec l'ADCVL à la mise en place d'une charte graphique pour le site de l'atelier qui viendra remplacer celle actuellement en place.

Ce même service a déjà réalisé un logo duquel découlera la charte graphique. Ce logo viendra remplacer celui actuellement utilisé qui correspond au logo des ateliers de la donnée labellisés. Xavier Rodier demande que l'aspect "atelier de la donnée labellisé" reste présent sur le site pour insister sur cet aspect.

3. L'actualité des groupes de travail

- Groupe de travail "Accompagnement au dépôt de jeux de données dans l'entrepôt approprié"

Le groupe de travail s'était réuni le 1^{er} juin.

Depuis cette date, Maéva Icentilai a travaillé sur l'établissement d'une liste des entrepôts de données par discipline. Les objectifs de ce travail étaient de fournir des ressources utiles aux chercheurs, notamment un état des lieux des entrepôts de données et des fiches pratiques sur ces entrepôts avec le type de données, le volume ou la licence. Dans un premier temps, elle a commencé un recensement de tous les entrepôts de sciences exactes mais devant la multitude, Laurence Rageot lui a demandé de se concentrer sur les entrepôts des centres de référence thématiques car ils devraient être inclus dans le catalogue en cours de constitution par Recherche Data Gouv. Ce catalogue aura pour but de repérer et signaler les données déjà partagées ou ouvertes grâce à un entrepôt thématique français ou international. En juillet, Isabelle Blanc a annoncé la publication de la liste des entrepôts inclus dans le catalogue pour la rentrée mais elle n'a pour l'instant pas été publiée.

Pascale Solon demande s'il y a un lien avec le répertoire des entrepôts <https://www.re3data.org/> qui comprend déjà 128 sites français. Laurence Rageot répond qu'à ce stade il n'y a aucune information sur ce qui sera inclus dans ce catalogue mais que des tests de moissonnage seraient en cours avec Nakala et Data Terra. Elle rappelle que l'entrepôt Recherche Data Gouv n'a pas vocation à remplacer les entrepôts disciplinaires mais à venir en complément pour les disciplines où il n'en existe pas.

Pascale Solon demande qui fera la mise à jour des fiches présentant les entrepôts.

Xavier Rodier rappelle que la complexité réside dans le grand nombre d'entrepôts, environ 3400 ont été identifiés par Recherche Data Gouv. De plus il faut distinguer ceux qui seront labellisés de ceux qui sont certifiés. Au final, très peu de ces entrepôts seront considérés comme dignes de confiance. L'objectif du catalogue n'est pas de recenser les entrepôts, mais plutôt de pointer les nouvelles données déposées par les chercheurs français dans les entrepôts où qu'ils se trouvent. Les données qui seront dans le catalogue auront plus de poids pour la communauté

française. Ainsi, il lui semble évident que les tests commencent avec Data Terra et les entrepôts français connus, et qu'ensuite cela s'élargisse. Ensuite, au niveau régional de l'ADCVL, il faudra réfléchir à la façon de mettre en place un observatoire de la science ouverte régionale qui, comme cela est fait pour les publications, permette de signaler les données partagées par les chercheurs de la Région.

Laurence Rageot demande à Maéva Icentilăi d'envoyer les fiches qu'elle a faites à tous les membres du groupe de travail.

- Groupe de travail "Plan de gestion de données (PGD)"

Le groupe s'était réuni le 14 mars 2023.

Laurence Rageot indique qu'en plus des demandes qui arriveront via le guichet unique, il est possible de connaître les chercheurs qui ont obligation de produire un plan de gestion de données dans le cas d'un financement ANR ou ERC. Cette année Mélanie Fauconnier donne son contact aux chercheurs en SHS de l'université de Tours qui sont porteurs d'un projet financé dès l'envoi de l'acte attributif de l'ANR. Elle a donc demandé à l'ingénieur de la MSH qui est à Orléans, Kiliann Perche de prendre contact avec le service Valorisation de l'université d'Orléans pour avoir la liste de projets SHS financés.

Par ailleurs, elle indique que le groupe de travail sensibilisation et formation avait évoqué en juin dernier la possibilité de mettre en place une formation sur les PGD mais rien n'a encore été entrepris. Cette formation, à monter en commun, permettrait de présenter la partie générique qui permettrait de gagner du temps lors des rendez-vous individuels avec les équipes de recherche qui demandent un accompagnement et une relecture de leur PGD.

Pascale Solon demande si la formation est pensée pour les chercheurs, Laurence Rageot répond par l'affirmative.

Nathalie Pothier demande si c'est uniquement les projets en SHS que Laurence Rageot recherche. Elle acquiesce et précise qu'elle a fait cette démarche, non pas dans le cadre de l'ADCVL, mais car la MSH propose un accompagnement aux chercheurs en SHS.

Mélanie Fauconnier propose de généraliser cette procédure au sein de l'université de Tours et de demander à ses collègues du service SPV d'indiquer une personne ressource pour le PGD.

- Groupe de travail "FAIRisation des données"

Le groupe s'était réuni le 19 juin 2023

Depuis cette réunion, Maéva Icentilăi a réalisé un recensement des guides de bonnes pratiques sur la FAIRisation des données mais cela s'est avéré très complexe car ce sont souvent des ressources très générales avec peu d'informations spécifiques à la discipline.

Laurence Rageot lui a ensuite demandé de réaliser des fiches pratiques sur les identifiants pérennes. Elle a travaillé sur les DOI, ARK et Handle.

Anne Azanza demande à qui ont été diffusées ces fiches et si elles peuvent être envoyées à tous les membres des groupes de travail. Maéva Icentilăi répond qu'elle les a envoyées aux membres du groupe « FAIRisation des données » mais qu'elle va les envoyer à tous les membres du comité opérationnel.

Pascale Solon demande si elles sont à usage interne ou si elles seront incluses dans la boîte à outils. Laurence Rageot répond que l'objectif est de les diffuser dans la boîte à outils mais que pour l'instant elles sont encore en cours de travail ainsi que le tableau récapitulatif/comparatif. Pascale Solon demande si les établissements ont vocation à devenir attributaires de DOI. Anne Azanza répond que la question se pose aussi à l'université de Tours. Pascale Solon se demande si l'ADCVL peut également être un cadre où partager de l'expérience sur ces sujets.

Laurence Rageot rappelle que les jeux de données déposés dans un entrepôt, comme Nakala ou Recherche Data Gouv, reçoivent un DOI et que c'est un critère important dans le choix d'un entrepôt.

Nathalie Pothier demande auprès de quelle institution il faut demander un identifiant ARK. Il lui est répondu que c'est la BNF qui les donne pour la France. Elle signale également qu'il manque l'identifiant ARKIV.

- Groupe de travail "Conseils éthiques et juridiques"

Le groupe s'était réuni le 30 mai.

Laurence Rageot indique que lors de la réunion du 30 mai, seule l'université de Tours était représentée, elle a donc été consacrée à une réflexion sur la façon d'intégrer les autres partenaires, notamment sur les questions de RGPD. Elle demande aux collègues de l'université d'Orléans si le recrutement du délégué à la protection des données (DPO) a abouti. Céline Girault répond que c'est toujours en cours.

Pour l'INRAE, Edouard Guitton indique qu'il y a un DPO au niveau national mais que les chercheurs adressent leurs questions au réseau des correspondants données qui peuvent les transmettre au DPO s'ils ne peuvent répondre.

Laurence Rageot indique qu'au CNRS, il existe un DPO national et que dans chaque délégation il y a un correspondant. Maryse Castan est la correspondante dans la délégation. Maryse Castan précise que les déclarations auprès de la CNIL se font principalement quand il y a des traitements très sensibles, notamment les traitements médicaux. Un registre de déclarations des données est effectué au niveau national.

Laurence Rageot a interrogé les services du DPO national pour savoir s'ils avaient été sollicités par d'autres ateliers de la donnée et si une organisation avait été mise en œuvre. Il lui a été répondu qu'une réflexion sur le mode de fonctionnement avec les ateliers de la donnée avait eu lieu. La DPO souhaite recevoir directement les demandes qui sont faites via le guichet unique afin de répondre aux chercheurs des laboratoires qui l'ont désignée comme DPO.

Nathalie Pothier demande si c'est le service juridique du CNRS qui traite ces demandes. Maryse Castan répond par la négative, c'est le service de la DPO nationale dirigé par Gaëlle Bujan.

Céline Girault indique qu'à l'université d'Orléans le DPO est en cours de recrutement mais il y a un service juridique qui pourrait éventuellement répondre à certaines questions, par contre elle ne sait si les collègues sont informés de l'atelier de la donnée. Cloé Freulon indique les avoir contactés à l'adresse mail générique pour leur poser la question.

Maryse Castan indique qu'elle a contacté Sébastien Coviaux (chef du service des Affaires juridiques de l'université d'Orléans) car un laboratoire avait un problème de déclaration de données et on lui a conseillé, pour l'instant, de déclarer le CNRS comme DPO. Elle rappelle qu'au CNRS, la consigne est que si le directeur d'unité est CNRS, il doit déclarer le DPO du CNRS. Cependant, dans le cas où le directeur d'unité est lié à un établissement universitaire, il est possible de se tourner vers son université.

Mélanie Fauconnier indique qu'on pourrait envisager d'inscrire dans les accords de site le DPO pour chaque unité pour éviter que ce soit lié à l'employeur du directeur d'unité. Xavier Rodier répond que s'il faut changer à chaque direction d'unité, ce n'est pas gérable. Maryse Castan répond qu'une fois un DPO déclaré, il reste même si le directeur d'unité change.

Mélanie Fauconnier indique accompagner de plus en plus de projets en SHS qui nécessitent une réflexion sur les aspects éthiques et juridiques car il y a d'énormes collectes de données personnelles et, contrairement aux sciences médicales, la question de l'ouverture des données se pose. Elle ajoute qu'il y a également des questions éthiques et juridiques de propriété intellectuelle qui ne sont pas traités dans les mêmes services suivant les établissements.

Pascale Solon demande si les responsables intégrité scientifique (RIS) sont intégrés car ils traitent la partie éthique et ont un rôle de sensibilisation.

Laurence Rageot remercie les présents de lui donner de nouveaux contacts pour ce groupe car le guichet unique ne propose pas pour le moment de solliciter ce groupe car il n'est pas en mesure de répondre aux sollicitations de tous les partenaires.

Pascale Solon enverra le lien vers le Comité Intégrité et Déontologie de la Recherche (CIDR) d'Orléans, instance que les chercheurs peuvent solliciter sur ces questions.

Mélanie Fauconnier indique que si le PGD est fait en amont, il peut être un point de départ pour d'autres contractualisations comme le contrat de collaboration de recherche et permet de gagner du temps car les questions de données personnelles sont déjà traitées. C'est un cercle vertueux qui se met en place.

Stéphanie Puel ajoute qu'à la première lecture du PGD, il y a souvent un flou sur les données qui seront utilisées, sur l'hypothèse de départ et par conséquent sur la nécessité de créer de nouvelles données. En sciences et techniques, avec Lucie Lepage, elles sont souvent confrontées à des chercheurs qui ont du mal à savoir s'ils ont des données personnelles et ce que cela recouvre. Elle s'interroge sur les accords de consortium et la paternité des données. Mélanie Fauconnier répond que c'est souvent à la fin des projets quand les résultats sont produits que les accords de consortium sont formalisés.

- Groupe de travail "Préservation et sécurité des données"

Le groupe s'est réuni le 19 juin 2023

Laurence Rageot indique que depuis la réunion du mois de juin, il y a de nouveaux participants dans ce groupe. Comme pour le DPO, au CNRS, il y a une archiviste nationale et un correspondant dans chaque délégation. Elle a contacté l'archiviste nationale qui n'avait pas encore été sollicitée par un atelier de la donnée et qui souhaitait avoir plus d'informations. En Région, c'est Maryse Castan qui est correspondante et qui, à ce titre, participe au groupe de travail.

En ce qui concerne la sécurité numérique, Alain Ravaz a proposé à Pierre Cotizar (responsable de la Sécurité des Systèmes d'Information – RSSI – du BRGM) et à son apprentie, Carla De Sousa, de rejoindre le groupe de travail.

Maryse Castan indique qu'elle pensait que Laurence Rageot devait prendre contact avec le RSSI régional du CNRS pour lui proposer d'intégrer le groupe. Laurence Rageot répond qu'elle allait le faire.

- Groupe de travail "Sensibilisation et formation"

Le groupe s'est réuni le 20 juin 2023.

Lors de la réunion de juin, il avait été décidé de monter une formation sur les plans de gestion de données mais, comme indiqué précédemment, rien n'a été engagé.

Laurence Rageot indique que, pour favoriser la sensibilisation, il est prévu de solliciter des correspondants dans les unités de recherche. Elle n'a pas entrepris de démarche en ce sens car elle attendait la publication du guichet unique pour avoir des éléments à proposer.

Mélanie Fauconnier indique qu'une démarche est entreprise au niveau de la direction de la recherche de l'université de Tours afin de recenser toutes les formations et actions de sensibilisation. Cela devrait permettre d'identifier tout ce qui est fait pour les chercheurs, de repérer les personnes compétentes, notamment en termes d'ingénierie pédagogique.

Enfin, Laurence Rageot indique qu'un chargé de la communication et de l'organisation des actions de formation et de sensibilisation va être recruté pour l'ADCVL dans les prochaines semaines.

4. PUI Loire Valley Innov'

Laurence Rageot présente le projet PUI Loire Valley Innov qui a pour objectif de faire de l'innovation un facteur d'attractivité du territoire en renforçant les liens entre les acteurs académiques de la Région et les acteurs socio-économiques. Les membres fondateurs sont également les partenaires de l'ADCVL avec en plus le CHU de Tours, le CHR d'Orléans et C-Valo plus une vingtaine de partenaires dont l'INSERM.

Éric Bricou qui a rédigé une des parties de ce projet a fait le lien entre la détection de l'innovation et l'atelier de la donnée notamment sur deux thématiques : les PGD qui décrivent le cycle de vie des données et peuvent permettre d'identifier ce qui peut être valorisé et conduire à des transferts technologiques ; les codes et les logiciels, qui représentent un autre pilier de la science ouverte qui n'ai pas traité dans l'ADCVL.

Sur ce PUI, financé pour ans à partir de 2024, deux postes vont être affectés à l'ADCVL. Le premier sur les PGD est prévu durer tout le temps du projet et débuter dès que la contractualisation avec le Ministère de la Recherche sera effective. La personne sera personnel de la MSH et localisée à Orléans. Le second poste, en ingénierie logicielle, est prévu sur 3 ans et 9 mois et sera hébergé à la MSH à Tours. Tous deux travailleront pour tous les partenaires de la Région.

5. Inauguration de l'ADCVL

Une journée d'inauguration officielle de l'ADCVL sera organisée début 2024. Isabelle Blanc (Administratrice ministérielle des données, des algorithmes et des codes sources, Ministère de l'Enseignement Supérieur et de la Recherche) sera invitée pour cette inauguration. Le programme envisagé devrait permettre une mise en perspective de l'atelier dans le panorama national de l'écosystème Recherche Data Gouv général avec des centres de référence thématiques (Huma-Num et Data Terra) et des centres de ressources (INIST pour Doranum et INRAE pour l'entrepôt et le catalogue) puis d'élargir les discussions avec la RDA (Françoise Genova) et l'EOSC (Suzanne Dumouchel).

Laurence Rageot va contacter Isabelle Blanc pour fixer des créneaux de dates qui seront ensuite proposés aux partenaires de l'ADCVL.

6. Informations Recherche Data Gouv

Laurence Rageot indique qu'un séminaire des ateliers de la donnée aura lieu à Reims les 14 et 15 décembre 2023. Ces séminaires ont pour but d'approfondir les liens entre les ateliers et les autres membres de l'écosystème. Celui qui a eu lieu à Lyon début juin, portait sur les centres de référence thématiques et celui de décembre traitera des liens avec les centres de ressources. Afin de préparer les discussions, des présentations en visioconférence des différents centres de ressources (Entrepôt-catalogue, Doranum, Urfist et Opidor) seront proposées à tous les membres de tous les ateliers.

Il a été annoncé que l'URFIST proposerait une offre de formation spécifique pour les ateliers de la donnée et que l'INIST présenterait la nouvelle version de DMP-Opidor.