

Comité opérationnel de l'Atelier de la donnée Centre-Val de Loire du 10 décembre 2024

Lieu : MSH Val de Loire, UFR Droit Économie Gestion Université d'Orléans et visioconférence

Horaires : 15h00 – 17h00

Participants : Anne AZANZA (Direction des bibliothèques & des archives - Université de Tours), Thierry BISSI-YANDIA (OSUC), Mélanie FAUCONNIER (Service Partenariat Innovation Valorisation - Université de Tours), Jorge FINS (MSH Val de Loire), Céline GIRAULT (Archiviste / Direction Pilotage Prospective Amélioration continue, Université d'Orléans), Edouard GUITTON (Correspondant données, INRAE), Florian HIVERT (MSH Val de Loire), Clara MONMARCHÉ (ADCVL), Julienne NEUHAUS (Déléguée à la Protection des Données - Université d'Orléans), Rémi OSSANT (ADCVL), Clément PLANQ (MSH Val de Loire), Nathalie POTHIER (OSUC), Stéphanie PUEL (Direction des bibliothèques & des archives, Université de Tours), Laurence RAGEOT (ADCVL), Pascale SOLON (SCD, Université d'Orléans)

Ordre du jour : 1. Informations générales concernant l'ADCVL – 2. Les groupes de travail de l'ADCVL - 3. L'enquête « Codes et logiciels » – 4. L'écosystème Recherche Data Gouv – 5. Questions diverses

[Compte-rendu rédigé par Rémi Ossant et Clara Monmarché, relu par Laurence Rageot et envoyé aux membres du comité opérationnel le 7 janvier 2025]

Compte-rendu

En introduction Laurence Rageot présente Julienne Neuhaus, nouvelle DPO et chargée d'affaires juridiques au sein du service des affaires juridiques de l'Université d'Orléans, et invite les membres présents à se présenter.

I. Informations générales concernant l'ADCVL

a. Communication

À la suite du départ de Lisa Bertrand qui avait la charge de la communication de l'ADCVL, Laurence Rageot, présente la nouvelle organisation et la répartition des actions de communication entre Rémi Ossant (LinkedIn), Clara Monmarché (Newsletter) et toute l'équipe pour les Actualités et agenda du site. Elle précise qu'il n'y aura pas de newsletter en décembre.

De nouveaux destinataires à la Newsletter ont été ajoutés en novembre :

- Les membres du comité de pilotage de l'ADCVL à la suite de la réunion du 1^{er} octobre 2024.
- La bibliothécaire du Centre d'Études Supérieures de la Renaissance de Tours) qui fait et envoie aux membres de son unité, une fois par mois, une lettre science ouverte centrée sur les publications.
- La cheffe de projet PUI, projet sur lequel sont financés les contrats de Clara Monmarché et de Rémi Ossant.

Laurence Rageot invite les membres du comité à faire part de leurs idées de contenus pour alimenter les canaux de l'ADCVL. Nathalie Pothier mentionne la fiche pédagogique sur EasyData et Rémi Ossant confirme qu'elle a été mentionnée dans les actualités du site.

b. Site et guichet unique : utilisation et évolutions

• Refonte de la page d'accueil

Rémi Ossant évoque la refonte de la page d'accueil du site de l'ADCVL, proposée par Lisa Bertrand, et mise en place depuis le dernier comité opérationnel. Une page Actualités et agenda a été ajoutée pour diffuser, en parallèle des publications sur LinkedIn :

- Les informations en rapport avec la science ouverte,
- Les actualités des différentes institutions partenaires de l'Atelier,
- Les actions de l'Atelier

Il invite les membres du comité à consulter cette page et à faire part de leurs suggestions. Il évoque aussi la possibilité de les former pour publier eux-mêmes des actualités sur cette page. Nathalie Pothier alerte sur d'éventuels doublons et demande s'il y a un modérateur. Rémi Ossant précise que si les membres veulent publier des contenus, il devra leur octroyer des droits supplémentaires. Ce point sera à discuter au cas par cas avant d'envisager un éventuel système de modération.

• Statistiques du guichet unique

Rémi Ossant présente les statistiques du guichet unique. Depuis le comité opérationnel du 26 septembre, deux demandes officielles ont été enregistrées, concernant l'accompagnement au dépôt de jeux de données dans un entrepôt SHS (l'une sur Progedo et l'autre sur Nakala) et une demande non-officielle sur la reproductibilité.

Thierry Bissi-Yandia interroge sur la notion de demande non-officielle. Laurence Rageot répond qu'elle a été formulée par mail à l'adresse de l'ADCVL mais que, pour garder trace de la réponse, elle a créé un ticket sur le guichet. La demande émanait d'un chercheur de l'INSERM qui souhaitait conduire un projet évaluant la reproductibilité dans son domaine de recherche et recherchait des pistes de contacts d'instituts qui pourraient être intéressés par la thématique. Elle ajoute qu'elle a décliné cette demande, la reproductibilité ne relevant pas du périmètre d'expertise premier de l'ADCVL et ne connaissant pas d'unités qui pourraient correspondre.

c. Plans de gestion de données

Clara Monmarché indique qu'un accompagnement à la rédaction de PGD est en cours (projet ERC PRIMA – SHS).

Une sensibilisation aux bonnes pratiques de gestion des données de la recherche a eu lieu en octobre pour des doctorants du Laboratoire Ligérien Linguistique à Orléans.

Une formation/sensibilisation à la rédaction d'un plan de gestion de données est en cours de préparation. Cette formation est à destination des lauréats ANR. Nathalie Pothier demande s'il est prévu de se limiter aux projets ANR ou s'il est possible de cibler également les projets européens car les conditions sont les mêmes en ce qui concerne les PGD. Thierry Bissi-Yandia demande quel est le périmètre de la formation. Clara Monmarché précise que, sur la forme, il s'agira d'une sensibilisation d'une durée d'environ 1h30 et que, sur le fond, elle ne sera pas tournée spécifiquement vers une discipline, du fait notamment que les lauréats ne sont pas encore connus dans leur totalité.

Thierry Bissi-Yandia pense que les lauréats de l'OSUC sont déjà connus. Nathalie Pothier ajoute que les projets financés par l'ERC ont été publiés. Elle pense qu'il faudrait proposer cette formation aux lauréats des projets ANR et européens. Laurence Rageot répond que l'on peut l'ouvrir aux deux mais que la question qui se pose actuellement concerne la façon dont on récupère les informations sur les projets financés. Pour les lauréats ANR de l'université de Tours, il a été discuté avec Mélanie Fauconnier du meilleur moment pour obtenir leurs coordonnées pour pouvoir les contacter. Si les calendriers correspondent avec les projets européens et que les informations sont disponibles, la formation peut leur être proposée. Nathalie Pothier ajoute que, s'il s'agit d'une formation-sensibilisation multidisciplinaire, elle sera adaptée aux projets européens. Laurence Rageot confirme et ajoute que l'objectif de la sensibilisation, est d'expliquer ce qu'est qu'un plan de gestion de données et d'inciter les chercheurs à contacter l'ADCVL pour un accompagnement plus spécifique. Elle se conçoit donc aussi bien pour les projets ANR que pour les projets ERC. S'il a été décidé de cibler, dans un premier temps les projets ANR, c'était dû aux moyens de récupérer la liste des projets financés et au calendrier. L'ADCVL va recontacter, dès le mois de janvier, les services de valorisation et partenariats pour essayer d'avoir la liste la plus complète possible des lauréats.

Nathalie Pothier demande si, aussi bien à Orléans qu'à Tours, le nombre de projets ANR et européens est connu. Mélanie Fauconnier répond que, pour ce qui est de l'appel à projet générique de l'ANR, le nom des lauréats est connu mais les phases de signature des actes attributifs sont encore en cours. Il avait donc été convenu de faire des extractions des bases d'actes attributifs ANR une seule fois en janvier, plutôt que tous les mois au fil de l'eau, pour être sûr d'avoir l'intégralité. Pour ressortir les autres appels à projets ANR il faut le faire au fil de l'eau car les calendriers ne sont pas nécessairement identiques. Il est aussi possible de requêter pour obtenir les lauréats des projets européens, voire tous les projets dont les financeurs demandent un PGD comme l'ANSES. Nathalie Pothier demande s'il est possible d'extraire aussi les PEPR. Mélanie Fauconnier précise que, quand elle évoque les autres projets, elle ne pense pas uniquement à la programmation ANR. Cependant, cela dépend de la manière dont sont structurées les bases de données et sont gérés les contrats dans les différentes institutions membres de l'ADCVL. Pour l'Université de Tours, il est aussi possible de faire un requêtage sur les PEPR.

Pour les PEPR, Nathalie Pothier évoque les projets OneWater et FairCarboN, pour lesquels il serait intéressant de voir s'ils ressortent dans les requêtes.

Mélanie Fauconnier précise que la base de l'Université de Tours n'a pas de champ spécifique pour indiquer si un PGD est obligatoire. Aussi, le choix de se concentrer sur les projets génériques ANR avait été fait car le calendrier est maîtrisé : en janvier de chaque année, on peut avoir la vision exhaustive de tous les projets financés l'année N-1. Elle suggère à Laurence Rageot de se réunir en janvier 2025 afin de regarder spécifiquement, au vu de la base de données de contrats, comment faire ressortir de manière plus large, indépendamment de l'ANR, les ERC ou d'autres types de financements européens.

Laurence Rageot ajoute que l'Atelier prendra également contact, en début d'année 2025, avec les services valorisation et partenariats de l'Université d'Orléans et du CNRS qui sont mutualisés pour essayer d'avoir les mêmes informations. Elle demande à Édouard Guittou si une politique de sensibilisation sur les PGD est faite au niveau national à l'INRAE. Il répond qu'il n'y a pas de politique générale à l'INRAE. Cela se fait au cas par cas, par centre. Il y a de la sensibilisation, mais il n'y a pas de recueil systématique des projets.

Laurence Rageot répond que l'ADCVL pourra faire passer l'information à INRAE Centre-Val de Loire pour que les personnes intéressées puissent participer aux sessions de sensibilisation.

d. Comité de pilotage du 1^{er} octobre 2024

Laurence Rageot fait un retour sur le comité de pilotage de l'ADCVL qui s'est déroulé le 1^{er} octobre 2024. Elle commence par présenter la liste des personnes présentes. Elle précise que personne n'était présent pour le BRGM, INRAE et INSA. A la demande de Edouard Guitton, elle précise que c'est le président de INRAE Centre-Val de Loire, Marc Guérin, qui est invité.

- **Les actions de l'ADCLV**

Dans un premier temps Laurence Rageot a fait un panorama des actions (*cf. support de présentation joint à ce compte-rendu*) qui ont eu lieu depuis la précédente réunion du comité de pilotage en septembre 2023. Les discussions ont principalement porté sur les aspects communication et sur l'enquête sur les codes et logiciels.

En effet, il a été demandé aux membres du comité de pilotage des voies de relais de la communication dans leurs institutions. Cela a donné lieu à une discussion et un intérêt marqué pour la newsletter, notamment au niveau de l'INSERM. Après la réunion, tous les membres du comité de pilotage ont été inscrits pour recevoir la newsletter et celle de novembre a été transmise à tous les laboratoires INSERM de la Région par Stéphane Cesbron.

Il y a ensuite eu une discussion sur les codes et logiciels et notamment sur le lien entre codes et logiciels et Ateliers de la donnée. Est-ce que les Ateliers de la donnée doivent se concentrer sur les données ? Ou doivent-ils aussi avoir une action sur les codes et logiciels ? Est-ce qu'un chercheur viendra chercher des informations sur les codes et logiciels dans un Atelier de la donnée ? Mme Lucas, chargée de mission science ouverte de l'Université d'Orléans, a indiqué, qu'en tant que chercheur, elle n'irait pas vers un Atelier de la donnée pour chercher des informations sur les codes et logiciels. Cela ne lui semble pas être le bon interlocuteur. D'autres membres du comité de pilotage ont répondu que les Ateliers de la donnée s'appellent « données », mais qu'ils traitent de la science ouverte avec des actions plus génériques. A ce titre, les codes et logiciels entrent dans le champ des Ateliers, ainsi plusieurs Ateliers de la donnée s'ouvrent aux codes et logiciels.

- **La mise en place de référents données pour l'ADCVL**

A l'ordre du jour de ce comité de pilotage, était inscrite la mise en place d'un réseau de référents données. Laurence Rageot a commencé par présenter la situation, institution par institution, comme cela avait été discuté lors de la réunion du comité opérationnel du 26 septembre. Carine Lucas a indiqué qu'elle était également en train d'établir une liste de correspondants science ouverte, et non uniquement données, dans les laboratoires avec des personnes volontaires. Elle a contacté les directeurs de laboratoire pour obtenir un correspondant dans chaque unité.

Nathalie Pothier indique avoir discuté avec son directeur, Christophe Tournassat, des correspondants science ouverte de l'OSUC qui fédère plusieurs unités. Il considère que c'est à lui de nommer ou pas un ou des correspondants. Concernant les autres laboratoires du campus, elle ne sait pas s'ils sont prêts ou pas à nommer un correspondant. Cependant, si Carine Lucas établit une liste de correspondants science ouverte cela pourrait faire doublon pour les UMR. Elle suggère de dresser une liste commune Université d'Orléans - CNRS. Pascale Solon est d'accord avec cette proposition car les UMR ne vont pas désigner deux personnes. Nathalie Pothier ajoute que pour l'OSUC il n'y aurait qu'une seule personne et non trois : UAR, LPC2E et ISTO. Ce sera un gros travail mais il ne reposera que sur une seule personne.

Laurence Rageot précise que l'objectif des correspondants est de transmettre de l'information dans les deux sens. Il faudra leur donner connaissance de ce que fait l'ADCVL pour qu'ils puissent orienter les chercheurs de leur unité et transmettre les propositions de formations ou actions proposées par l'ADCVL. De plus, leur rôle sera également de faire remonter à l'ADCVL les besoins de leur unité sur les différents aspects de la science ouverte. Il faudra donc voir, une fois que Carine Lucas aura donné la liste des correspondants, quelles sont les unités propres du CNRS dont il faudra solliciter les directeurs et leur expliquer la démarche. Nathalie Pothier ajoute que la liste de Carine Lucas pourrait être discutée en conseil des directeurs d'unité. Elle suggère aussi qu'une présentation de l'ADCVL soit faite en conseil des directeurs. Laurence Rageot indique que Emmanuelle Huver, vice-présidente recherche de l'Université de Tours, l'a proposé pour le prochain conseil des directeurs d'unités qui doit avoir lieu au printemps. Mais depuis cette réunion, le vice-président recherche a changé, il faudra donc en rediscuter. Par ailleurs, Emmanuelle Huver a proposé que, pour l'Université de Tours, ce soient les correspondants HAL qui voient leurs missions s'élargir aux données pour devenir correspondants Science Ouverte. Laurence Rageot prendra contact avec Anne Azanza et Stéphanie Puel pour en discuter et voir comment mettre cela en œuvre.

Enfin, à la demande de Nathalie Pothier, Laurence Rageot confirme que le BRGM était présent à l'inauguration de l'ADCVL : Christelle Loiselet (membre du comité opérationnel) et Michaël Chelle représentant DataTerra. De plus, elle indique avoir participé à une réunion au BRGM sur la mise en place du Datacentre régional et lors de cette réunion, le DSI du BRGM l'a informée qu'il allait, à partir de fin de l'année 2024, intégrer la Science Ouverte à sa direction. Laurence Rageot reprendra donc contact avec lui en début d'année prochaine. Elle espère pouvoir constituer une première liste de correspondants dans les premiers mois de 2025.

II. Groupes de travail

Laurence Rageot précise qu'il n'y a pas eu beaucoup d'évolution concernant les groupes de travail exceptés pour deux d'entre eux.

a. GT Plan de gestion de données

Clara Monmarché indique que le groupe s'est réuni le 12 novembre pour discuter de l'organisation des réponses aux sollicitations, étant donné qu'elle travaille à temps plein sur les plans de gestion de données. Il a été décidé de travailler à deux sur les demandes : Clara Monmarché et une autre personne du groupe.

A la demande de Nathalie Pothier, Clara Monmarché indique qu'elle n'a pas rencontré de difficulté particulière pour le moment. Elle a traité les demandes en binôme avec Laurence Rageot et avec Pascale Solon pour l'une d'entre elles.

b. GT conseils éthiques et juridiques

Laurence Rageot rappelle que l'Université d'Orléans a désormais une déléguée à la protection des données et que l'ADCVL serait très heureux de l'accueillir dans le groupe si elle le souhaite. En revanche, Morgane Guibert quittera l'Université de Tours et donc le groupe fin décembre. Elle est spécialiste de la propriété intellectuelle et donc des licences dans le service valorisation et partenariat. Elle devrait être remplacée mais nous ne savons si la personne prendra part au groupe de travail. Mélanie Fauconnier indique qu'il y a un recrutement en cours, mais qu'il n'y a pas de visibilité, ni sur le nom de la personne, ni sur sa date d'arrivée.

Elle ne sait pas non plus s'il a été prévu qu'elle reprenne la participation à ADCVL. Laurence Rageot précise avoir rencontré, avec Rémi Ossant, Éric Bricout, responsable du service, qui avait indiqué que la personne y participerait peut-être mais qu'il fallait d'abord qu'elle prenne en charge tous les dossiers du quotidien que Morgane Guibert traite.

Mélanie Fauconnier ajoute que Cloé Freulon, DPO de l'Université de Tours, participe à ce groupe de travail.

Laurence Rageot explique le rôle et la problématique du groupe de travail pour Julienne Neuhaus : impossibilité de permettre les sollicitations via le guichet unique par manque de représentants des différentes institutions. Quelques questions ont cependant été soulevées notamment dans le cadre des plans de gestion de données. Un premier niveau de réponse a été apporté avec l'indication du service à contacter au sein de l'institution du demandeur.

Nathalie Pothier ajoute qu'il s'agit d'un service très important. Dans le cadre de ses fonctions nationales, elle a reçu des demandes et constate se sentir rapidement démunie. Laurence Rageot ajoute que Recherche Data Gouv annonce un site portail sur les aspects juridiques qui devrait sortir début 2025. Cependant, Nathalie Pothier remarque que le chercheur pose toujours des questions très pointues à laquelle on n'est pas forcément en mesure de répondre. Par exemple, elle a été obligée de solliciter Lionel Maurel, Directeur adjoint scientifique de l'Institut national des Sciences Humaines et Sociales du CNRS, sur une question juridique bien particulière.

Mélanie Fauconnier ajoute que ce sont les services partenariat valorisation qui ont les compétences pour les conseils éthiques et juridiques, et spécifiquement sur les questions de propriété intellectuelle et de licences. Les compétences existent dans chacun des établissements. Par ailleurs, le conseil dépend aussi des politiques d'établissement. Aussi, même s'il existe de nombreuses sources d'informations ou de documentation au niveau national, l'application concrète sur les questions juridiques et de propriété intellectuelle, relèvent aussi parfois de décisions des conseils d'administration des universités.

Laurence Rageot confirme que pour l'instant, l'ADCVL ne peut que renvoyer vers le service partenariat et valorisation de l'institution du chercheur.

III. L'enquête Codes sources et logiciels

Rémi Ossant présente le projet d'enquête sur les pratiques et les besoins de la communauté scientifique autour des codes sur ces logiciels (*cf. enquête envoyée au préalable aux membres du comité opérationnel et jointe à ce compte-rendu*). Elle est réalisée dans le cadre du Pôle Universitaire d'Innovation Loire Valley Innov', projet qui vise à accélérer l'innovation sur le territoire et le transfert de technologies de la recherche publique vers le monde social et économique. Elle servira à l'ADCVL mais aussi C-VaLo, structure qui fait de l'investissement pour du transfert des technologies.

Le public visé par l'enquête correspond à l'ensemble des chercheurs, des doctorants et des personnels d'appui à la recherche des institutions partenaires du PUI (Université de Tours, Université d'Orléans, CNRS, INRAE, BRGM, INSA, CHRU de Tours et CHU d'Orléans). Elle veut toucher plus spécifiquement les producteurs de codes et, dans une moindre mesure, les utilisateurs de codes pour cerner le mieux possible les pratiques et les besoins des acteurs de la recherche autour de la production, de la diffusion et de la valorisation des codes sources et logiciels. L'enquête sera sous forme numérique ce qui permettra de la diffuser sur le maximum de canaux de communication (mails, LinkedIn, listes de diffusion, etc.).

Rémi Ossant rappelle le contexte général de production des codes sources et logiciels au sein du milieu de la recherche et les problèmes qui sont liés à leur production, leur conservation, leur maintien et leur diffusion. Une meilleure valorisation des productions logicielles peut permettre de palier à ces problèmes, de leur donner plus de visibilité et d'augmenter leur usage par la communauté. Cette valorisation implique des aspects techniques (diffusion, sauvegarde, archivage, maintenabilité et documentation) et des aspects juridiques (propriété intellectuelle et licences attachées).

Il existe deux voies de valorisation qui impliquent des questionnements et des pratiques communes :

- La voie Science Ouverte, portée par l'ADCVL, qui vise à ouvrir les codes et logiciels au maximum
- La voie de la valorisation économique, qui intéresse C-VaLo, qui vise à transférer une création, une production scientifique vers le monde économique global.

L'enquête comprend une vingtaine de question organisées autour de sept grandes thématiques :

- Les informations générales sur le répondant
- Le statut du répondant (producteur et/ou utilisateur de codes et logiciels)
- Les motivations à produire des codes et logiciels de la part du répondant
- Le devenir des codes et logiciels
- Les aspects juridiques
- Les voies de valorisation
- Les besoins des répondants

Elle n'est pas très longue pour ne pas décourager les répondants et se veut la plus inclusive et transversale possible. Elle donne la possibilité de faire remonter des besoins et d'être recontacté. L'un des enjeux du questionnaire est aussi de balayer les éléments importants sur la production et la valorisation des codes sources et logiciels pour avoir un panorama des pratiques. Elle devrait ainsi permettre d'obtenir des profils types de producteurs et d'utilisateurs, de détecter des foyers de production, de lister les besoins des répondants et de réfléchir à la manière dont l'ADCVL pourrait y répondre et, enfin, d'ouvrir un dialogue avec les répondants qui souhaiteraient un accompagnement.

Nathalie Pothier indique que certains termes techniques ou outils mentionnés dans le questionnaire risquent de ne pas parler aux chercheurs. Rémi Ossant répond que des définitions et des renvois seront intégrés au questionnaire pour clarifier au maximum ces points.

Nathalie Pothier et Jorge Fins s'interrogent sur la meilleure manière d'interroger les répondants sur leurs pratiques de conservation des codes sources, les solutions mises en avant comme *Software Heritage* risquant de ne pas être connues. Jorge Fins propose de laisser la question sur la manière dont les chercheurs ouvrent leurs codes la plus ouverte possible pour avoir la vision la plus réaliste des pratiques. Il propose également d'ouvrir la question sur l'origine des codes utilisés par les répondants en ajoutant un champ « Autre » pour laisser le plus de liberté possible. Même remarque concernant la question sur les besoins en accompagnement et sensibilisation des répondants. Sur cette même question, Thierry Bissi-Yandia, Nathalie Pothier et Céline Girault proposent de revoir la formulation et les réponses proposées.

Rémi Ossant termine en encourageant les membres de l'ADCVL à continuer à lui faire des retours sur l'enquête.

IV. L'écosystème Recherche Data Gouv

Laurence Rageot évoque quelques informations obtenues lors des journées Recherche Data Gouv des 20 au 22 novembre 2024 à Nancy.

Des chiffres clés sur le dispositif ont été présentés : 36 dispositifs d'accompagnement et 370 personnels dans les Ateliers de la donnée et les centres de référence thématique ; 1,5 million de téléchargements sur l'entrepôt Recherche Data Gouv ; 3000 jeux de données disponibles. La structuration de Recherche Data Gouv a également été évoquée, l'objectif étant de postuler pour devenir infrastructure de recherche. Une organisation opérationnelle a donc été mise en place avec un organigramme : pilotage avec Isabelle Blanc et deux adjoints (Gilles Mathieu – adjoint opérations et coordination - et Véronique Stoll - adjointe stratégie et partenariat) ; coordination avec le responsable de l'UMS INRAE gérant l'entrepôt Recherche Data Gouv et le futur catalogue, le bureau des Ateliers, Gilles Mathieu, Véronique Stoll et la chargée de communication.

Un consortium Recherche Data Gouv va être créé pour formaliser le lien et la fédération des 76 institutions qui contribuent à Recherche Data Gouv. L'accord de consortium sera probablement envoyé fin 2024 ou début 2025. Pour l'ADCVL c'est l'Université de Tours qui recevra cet accord.

Il est prévu la création d'un comité de pilotage Recherche Data Gouv auquel participera le conseil des membres du consortium.

Nathalie Pothier indique que DataTerra étant infrastructure de recherche, il lui semble normal que Recherche Data Gouv le soit également. Elle ajoute qu'être infrastructure de recherche est une vraie labellisation qui permet, par exemple, de mettre le code ministériel d'infrastructure de recherche dans HAL. Laurence Rageot évoque aussi le fait que cela permettra à Recherche Data Gouv d'être partenaire de projets notamment européens ce qui est compliqué actuellement car ce n'est pas une institution. Il y a des vrais enjeux à ce cette évolution.

Laurence Rageot évoque ensuite une intervention d'Isabelle Blanc sur les principes FAIR pendant laquelle elle est revenue sur ce que prévoit la loi et ce que l'on peut conseiller en termes d'ouverture des données. Elle a apporté des précisions sur le « aussi ouvert que possible, aussi fermé que nécessaire ».

Nathalie Pothier indique qu'elle explique aux chercheurs que ce n'est pas parce que les données sont accessibles qu'elles sont réutilisables. Elle pense que c'est une nuance importante et un message à faire passer. Mélanie Fauconnier demande quel était le fond du message d'Isabelle Blanc dans cette présentation.

Au regard du discours d'Isabelle Blanc, Laurence Rageot indique avoir eu l'impression de faire des raccourcis en disant par exemple que certaines choses étaient obligatoires alors qu'elles ne le sont pas, notamment sur le fait que la loi n'oblige pas à partager les données. Clara Monmarché ajoute qu'elle a compris le propos d'Isabelle Blanc dans le sens où la loi autorise l'ouverture des données mais qu'elle n'oblige pas à ouvrir, car il y a de nombreux cas dans lesquels ce n'est pas possible.

Nathalie Pothier ajoute qu'il ne faut pas oublier que le maître d'œuvre, c'est le chercheur, que c'est lui qui définit l'ouverture ou pas des données. Mélanie Fauconnier ajoute qu'il y a également des questions de propriété intellectuelle. Le sujet est toujours de définir ce qui

prévaut. Est-ce que c'est « on ouvre, sauf si » ou bien faire attention à ce qui doit être fermé? Les chercheurs font effectivement les choix de diffusion mais il y a aussi des contextes dans lesquels ils n'ont pas le choix, par exemple, dans un contrat de collaboration de recherche avec un industriel. Elle ajoute que la question n'est pas toujours simple, car souvent, ils sollicitent l'industriel parce qu'ils ont besoin d'argent ou d'un matériel, etc. Mais, d'expérience, une fois les contrats signés, le chercheur n'a pas toujours en tête ce qu'il a le droit de faire ou pas. Elle se demande si ces précisions apportées par Isabelle Blanc ne montrent pas une forme d'affaiblissement de la politique de Science Ouverte à la suite d'un constat que des données ont été trop ouvertes et ont nui à des possibilités d'innovation, de valorisation.

Nathalie Pothier ajoute qu'il y a une politique IA qui moissonne une partie des données sans demander obligatoirement des autorisations. Un cas vient d'être signalé sur des auteurs d'un ouvrage. Céline Girault rejoint Nathalie Pothier et ajoute que le croisement entre les effets de l'IA, les obligations contractuelles et d'autres environnements réglementaires qui peuvent s'appliquer, fait que finalement, certaines politiques peuvent se réajuster, comme avec la Science Ouverte, qui nécessite de remettre quelques barrières. Elle prend l'exemple des entrepôts de données qui dans un premier temps incitaient à tout déposer et désormais précisent qu'ils ne sont que des diffuseurs qui ne conservent les données que pendant cinq ans. Nathalie Pothier ajoute que le problème des entrepôts de données est le recrutement et le non-renouvellement des CDD. Mélanie Fauconnier ajoute qu'il faut des moyens pour pérenniser et maintenir les dispositifs.

Nathalie Pothier indique que, dans un premier temps, l'objectif était d'habituer le chercheur, de lui faire acquérir un vocabulaire. Désormais, il faut être plutôt prudent et ne pas tout déposer, il faut habituer le chercheur à sélectionner les données. La philosophie de Recherche Data Gouv et des entrepôts de données, est de lier les données aux publications et non pas de tout déposer, il faut donc peut-être faire attention à recentrer nos discours. Georges Fins demande si l'un des intérêts des Ateliers de la donnée pouvait être de repréciser ce qu'implique un financement ou une contractualisation.

Stéphanie Puel demande si des informations concernant un 3^{ème} Plan National pour la Science Ouverte avaient été données pendant les journées Recherche Data Gouv car, lors de l'Assemblée des partenaires HAL à laquelle elle a participé au mois de novembre, la question a été posée plusieurs fois. Laurence Rageot répond que la question n'a pas été posée.

Elle ajoute que pendant les journées Recherche Data Gouv de nombreuses discussions entre Ateliers de la donnée ont porté sur la façon convaincre les chercheurs de partager leurs données. Dans certains cas, on peut être amené à aller plus loin que les obligations légales. Rémi Ossant ajoute qu'il pense qu'Isabelle Blanc voulait alerter sur des amalgames entre ce qui est encouragé par la politique Science Ouverte du ministère et ce qui est légalement obligatoire ou un devoir pour le chercheur. Il y a des amalgames entre ce qui est fortement encouragé et ce qui est fondamentalement obligatoire.

Clément Plancq demande si l'ouverture des publications liées à la recherche est obligatoire. Nathalie Pothier répond que le dépôt des publications est obligatoire mais que pour les données, c'est plus subtil. Il n'y a pas d'obligation dans la loi, c'est fortement conseillé. Elle constate que, quand elle accompagne Thierry Bissi-Yandia, le binôme fonctionne bien car elle peut repérer des données brutes liées à un article. Ils sollicitent le chercheur et travaillent avec lui pour sélectionner les données qu'ils déposent soit sur Recherche Data Gouv, soit sur un entrepôt de données thématique type DataTerra. Thierry Bissi-Yandia ajoute qu'il y a un travail de communication avec le chercheur. Il explique qu'il y a peu de données déposées car il y a

beaucoup de contraintes au niveau de la modération si les données brutes ne sont pas associées à un article. Nathalie Pothier ajoute qu'une difficulté et une perte de temps pour le chercheur est de retrouver les données brutes. En revanche, l'un des chercheurs accompagnés était satisfait d'avoir un DOI pour ses données. Thierry Bissi-Yandia ajoute que les indicateurs de consultation sur Recherche Data Gouv sont aussi des éléments encourageants pour le chercheur.

Mélanie Fauconnier indique qu'elle trouve intéressant le positionnement de Rémi Ossant qui est à la fois dans l'ADCVL et dans un contexte de pôle universitaire d'innovation. Elle serait intéressée par un retour d'expérience sur la frontière entre la politique d'innovation dans laquelle on cherche à fermer des choses, la logique d'investissement et d'intervention de C-VALO, et la politique publique d'ouverture des données. Rémi Ossant répond qu'il lit beaucoup de littérature sur le sujet de l'innovation et de la Science Ouverte car il a découvert le monde de la valorisation en arrivant à C-VALO. Il propose aux membres du comité un retour d'expérience sur les points de friction entre Science Ouverte et valorisation, mais aussi sur les choses qui peuvent fonctionner de pair, même si les finalités sont différentes. Nathalie Pothier suggère qu'il rédige un article sur le sujet. Il est convenu d'évoquer le sujet lors de la prochaine réunion du comité opérationnel.

V. Questions diverses

Céline Girault souhaite évoquer un post LinkedIn de l'ADCVL sur la distinction stockage/archivage. Certaines phrases l'ont interpellée, notamment celle concernant le CINES. Elle ne comprend pas pourquoi le CINES est présenté comme étant mandaté pour stocker les archives de l'enseignement supérieur et de la recherche, alors qu'il s'agit d'un « tiers archiveur », à part sur la partie HAL. Cela occulte complètement la sphère du contrôle scientifique et technique des archives et du Service Interministériel des Archives de France (SIAF). Elle souhaite savoir qu'elle était l'initiative derrière ce post et le public visé car elle n'est pas certaine d'adhérer à la totalité du message.

Laurence Rageot répond que l'objectif était d'expliquer la différence entre stockage, partage et archivage qui est souvent très floue pour les chercheurs, de définir ces trois notions. Elle reconnaît qu'un raccourci a peut-être été fait sur le rôle du CINES, mais elle s'interroge sur ce qu'il faut conseiller à un chercheur qui voudrait archiver ses données. Faut-il l'adresser au SIAF? Céline Girault indique qu'il y a une question de financement avec le CINES, l'établissement du chercheur doit contractualiser avec le CINES pour qu'il puisse déposer des archives car le dépôt est payant. Alors qu'avec les archives nationales et les archives départementales, le dépôt est gratuit car c'est une obligation légale.

Laurence Rageot demande si les archives nationales et les archives départementales prennent les données numériques. Céline Girault répond que l'identification des interlocuteurs est un sujet primordial du groupe inter-ateliers sur l'archivage. La loi indique que les interlocuteurs sont les archives nationales et les archives départementales mais le maillage territorial fait que les archives départementales ont des financements plus ou moins équitables suivant les départements. De ce fait les départements sont plus ou moins bien pourvus en systèmes d'archivage électronique, dans certains départements il n'y en a aucun. C'est tout l'enjeu porté par les archivistes des universités et des organismes de recherche et dans les Ateliers de données par ce groupe de travail. Il faut porter cette réalité du territoire au niveau national pour que des solutions soient apportées.

Elle ajoute que le groupe de travail rencontre un problème avec Isabelle Blanc qui est très ouverte sur les livrables des groupes mais qui refuse de solliciter le SIAF pour avoir des

conseils. Pour Isabelle Blanc, c'est le COSO qui pilote la stratégie, la politique nationale de la Science Ouverte et donc il n'y a pas d'interaction entre les Ateliers de la donnée et les institutions.

Laurence Rageot demande que doit faire un chercheur de l'Université d'Orléans qui voudrait archiver ses données numériques. Céline Girault répond qu'elle n'a pas de solution, car la logique voudrait que l'interlocuteur soit le service d'archivage électronique des archives départementales mais, à l'heure actuelle, ce n'est techniquement pas réalisable. De plus, on ne peut décorréliser les données de la recherche des archives de la recherche, des publications et des laboratoires dans lesquels elles ont été produites. Elle espère que le livrable produit par le groupe de travail recevra un accueil positif auprès du COSO, car il y a des enjeux. Elle explique la chaîne de validation du livrable et s'interroge sur les liens entre Isabelle Blanc, Recherche Data Gouv et le COSO. Rémi Ossant précise que ce sont des institutions compartimentées qui n'ont officiellement pas de lien administratif mais des personnes sont membres des deux. Une des missions du COSO est de faire des propositions directement au Ministère de la Recherche sur tous les aspects autour de la Science Ouverte, c'est l'interlocuteur du ministère. Céline Girault ajoute que Véronique Stoll et Isabelle Blanc seraient membres du COSO en plus de piloter Recherche Data Gouv.

Laurence Rageot précise que, jusqu'ici, les livrables des groupes de travail ont été mis sur le site de recherche Data Gouv et qu'elle n'a pas entendu parler de censure d'un livrable. Elle pense que le livrable sera publié mais elle ne sait si le COSO s'en emparera. Rémi Ossant précise que tous les livrables des groupes de travail n'ont pas la même portée. Dans le groupe de travail sur les outils communs, les livrables restent à la disposition des Ateliers.

Pour terminer sur le sujet du post LinkedIn, Rémi Ossant propose à Céline Girault de discuter du fond pour corriger les éléments qui ne lui sembleraient pas correctes et reconnaît qu'ils auraient dû le faire relire. Céline Girault répond qu'elle ne veut pas avoir de visa à donner sur les posts, mais que si l'Atelier le souhaite, elle accepte d'être consultée.