

Comité opérationnel de l'Atelier de la donnée Centre-Val de Loire du 18 mars 2025

Lieu : MSH Val de Loire à Tours, OSUC à Orléans et visioconférence

Horaires : 14h00 – 16h00

Participants : Laurent BEUNECHE (RSSI - Université de Tours), Thierry BISSI-YANDIA (OSUC), Maryse CASTAN (Correspondante archive DR08 / Service Informatique DR08 CNRS), Mélanie FAUCONNIER (Service Partenariat Innovation Valorisation - Université de Tours), Jorge FINS (MSH Val de Loire), Cloé FREULON (Déléguée à la Protection des Données - Université de Tours), Céline GIRAULT (Archiviste / Direction Pilotage Prospective Amélioration continue, Université d'Orléans), Florian HIVERT (MSH Val de Loire), Christine LE BAS (Correspondant données, INRAE), 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. Innovation versus/et ouverture des données– 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 15 avril 2025]

Compte-rendu

En introduction, Laurence Rageot informe les membres du comité qu'elle a rencontré deux vice-présidents de l'Université de Tours le matin même : le Vice-président Recherche Daniel Alquier et le vice-président Écoles doctorales/SAPS, Patrick Vourc'h, qui est également référent Science Ouverte pour l'université de Tours. L'objectif du rendez-vous était de leur présenter l'Atelier de la donnée et de répondre à leurs questions sur son fonctionnement. Elle indique également avoir évoqué avec Camélia Turcu, nouvelle vice-présidente Recherche de l'Université d'Orléans, la possibilité d'organiser une réunion avec le Vice-président Numérique Mathieu Exbrayat et la chargée de mission Science Ouverte de l'Université d'Orléans, Carine Lucas, afin de présenter également l'Atelier de la donnée.

I. Informations générales concernant l'ADCVL

a. Site et guichet unique – demandes

L'Atelier a reçu 5 demandes depuis le 14 décembre (date du dernier comité opérationnel).

Deux concernent l'accompagnement au dépôt de jeux de données dans un entrepôt :

- Rémi Ossant évoque la demande de Hovig Ter Minassian, chercheur au laboratoire CITERES qui a sollicité l'Atelier car il voulait un accompagnement pour déposer ses données dans l'entrepôt de données Progedo. Le jeu de données est issu d'une

enquête en sociologie de 2014. Rémi Ossant précise que la démarche est intéressante car le chercheur souhaitait aussi anticiper le dépôt d'un nouveau jeu de données similaire pour mise à disposition à des étudiants. Le chercheur a indiqué que son jeu de données ne contenait pas de données personnelles, mais après vérification auprès de Cloé Freulon, il s'est avéré que certaines données étaient identifiantes. Les données ne pourront donc pas être ouvertes. La demande suit son cours, car le chercheur doit aussi vérifier certains aspects contractuels, l'enquête ayant été faite par une société privée, il faut s'assurer que le contrat autorise la diffusion des données. La question a été posée au Service Partenariat Innovation Valorisation.

- Georges Fins évoque la demande de Dominique Andrieu, cartographe à la MSH Val de Loire, pour une aide au dépôt dans Nakala de données d'un atlas cartographique qu'il a créé dans le cadre d'un projet en géographie (LAPTER – Labels Patrimoniaux et Touristiques en Région Centre Val de Loire : une ressource TERRitoriale ?). Dominique Andrieu a suivi une formation faite par Huma-Num mais avait besoin d'accompagnement pour gérer les spécificités de ses données. Georges Fins l'a conseillé et orienté pour se poser les bonnes questions notamment sur les catégories existantes.

Clara Monmarché développe les demandes traitées concernant les plans de gestion de données :

- Demande de Dominique Certon (GREMAN) qui a obtenu un projet ANR et a sollicité l'Atelier avant de débiter la rédaction du PGD. Elle lui a expliqué les attendus du document et comme le projet est en partenariat avec des entreprises privées, elle lui conseillé de se baser sur les accords de consortium pour certaines questions.
- Demande de relecture du PGD du projet ANR PSIND (musicologie) par Curtis Dumont du CESR (Tours).
- Sollicitation de Curtis Dumont pour le PGD du PEPR ICCARE (musicologie) mais il s'est avéré que le CESR ne se chargerait pas de sa rédaction. Cependant, Clara Monmarché se demande s'il faut rédiger un PGD pour l'ensemble du PEPR ou uniquement pour les sous-projets. Nathalie Pothier répond qu'un PGD doit être rédigé par le coordinateur du PEPR. Enfin, Clara Monmarché indique que, pour l'un des sous-projets du PEPR ICCARE, les porteurs ont été rappelés à l'ordre par l'ANR pour cause de retard dans le dépôt de la version initiale du PGD, ce qui montre que l'agence effectue un suivi actif de ces livrables.
- Relecture de la version initiale du PGD du projet ERC Prima (littérature) pour donner suite à la demande de Elena Pierazzo (CESR) faite à Laurence Rageot.
- Relecture de la version intermédiaire du PGD du projet ANR CelticBronzeCoins (histoire) pour donner suite à la demande de Sylvia Nietto (IRAMAT) faite à Laurence Rageot.
- Rendez-vous avec Nicolas Labroche (LIFAT) pour le projet ANR AIDA dont le BRGM est partenaire. Le projet a démarré en janvier 2025 et ils ont fait le point sur les attendus du PGD.

Les autres demandes traitées :

- Une demande de stage à laquelle l'Atelier n'a pas répondu favorablement.
- Une demande transmise à l'Atelier par Julienne Nehauss d'un doctorant du LLL (linguistique) qui se posait des questions sur la réutilisation de données issues de

Cocoon. Laurence Rageot précise que Cocoon est un entrepôt d'Huma-Num pour les données orales.

b. Codes sources et logiciels - Enquête sur les pratiques autour des codes sources et logiciels : bilan à mi-parcours

Rémi Ossant présente l'état d'avancement de l'enquête sur les pratiques autour des codes sources et logiciels qui est ouverte du 3 au 31 mars. Il remercie l'ensemble des membres de l'Atelier pour avoir relayé l'information dans leurs réseaux professionnels. Il indique que les institutions membres du PUI ont également bien relayé l'information. Il a aussi mobilisé des personnes ressources et des réseaux de professionnels au sein d'unités de recherche, via des structures comme la MSH Val de Loire ou des groupements de professionnels comme le réseau R'Dev (réseau des développeurs des universités de la Région).

L'enquête a reçu, à la date du comité, une centaine de réponses en provenance d'une quarantaine de laboratoires différents ce qui indique qu'il y a eu une bonne diffusion de l'information mais qu'au sein de chaque laboratoire, peu de personnes se sont senties concernées ou légitimes de répondre. Rémi Ossant indique qu'il a été contacté par des personnes n'ayant pas répondu à l'enquête car elles pensaient ne pas produire « assez de code » pour que leur réponse soit pertinente. Il indique donc travailler à une nouvelle communication pour les inciter à répondre.

Il présente ensuite quelques tendances : les réponses présentent une grande variété d'origine, les chercheurs sont très bien représentés (60% des répondants) et 20% des répondants ont laissé leurs coordonnées pour être recontactés et accompagnés sur ce sujet.

Nathalie Pothier indique qu'il aurait été pertinent de diffuser une version anglaise de l'enquête car de nombreux chercheurs, post-doctorants et doctorants des unités de recherche sont anglophones. Rémi Ossant répond qu'il n'est malheureusement plus possible de modifier l'enquête en cours mais qu'une version bilingue sera envisagée pour les prochaines.

Alain Ravaz demande si les personnes ayant répondu appartiennent au réseau R'Dev. Rémi Ossant répond qu'il n'a pas analysé suffisamment les réponses pour le savoir.

Thierry Bissi-Yandia demande si la variété d'origine des répondants concerne les équipes de recherche ou les domaines scientifiques. Rémi Ossant répond que les réponses présentent une bonne représentation des équipes de recherche (environ 40 représentées sur 70 dans la région) ainsi que des domaines scientifiques avec des informaticiens, des chercheurs en SHS, des biologistes, des écologues, des juristes, ...

Laurent Beunèche insiste sur l'importance du réseau R'Dev qui permet de regrouper des personnes produisant des codes et logiciels et d'échanger sur les pratiques. Il demande quel sera le devenir de cette enquête une fois terminée. Rémi Ossant indique qu'un bilan permettra faire ressortir les tendances dans les pratiques et les besoins afin de définir une offre d'accompagnement adaptée tant par C-VaLo que par l'ADCVL. Le bilan devrait aussi permettre d'identifier les foyers de production de codes. Enfin, les répondants qui ont exprimé le besoin d'être accompagnés seront recontactés.

c. Formations

Formations réalisées

Laurence Rageot présente les formations que l'ADCVL a animées ou auxquelles il a participé depuis le dernier comité opérationnel.

En janvier, Laurence Rageot et Rémi Ossant sont intervenus dans deux conférences de sensibilisation à la science ouverte pour les écoles doctorales, sur sollicitation de Stéphanie Puel. Elles portaient sur les publications, les données de recherche ainsi que les codes sources et logiciels. Ces conférences sont obligatoires pour les doctorants de Tours et ont rassemblé de 50 et 80 personnes. Le bilan est positif, les doctorants ont participé activement.

Clara Monmarché présente le cycle de huit sessions de webinaires sur les plans de gestion de données qu'elle a animé entre le 25 février et le 6 mars. Cette formation était proposée aux lauréats ANR et ERC financés en 2024 et a réuni 48 personnes. Elle a reçu des retours très positifs. La formation abordait les questions de définition du PGD, de son utilité ainsi que les bonnes pratiques de gestion des données.

Formations à venir

Laurence Rageot indique qu'elle travaille, avec Clara Monmarché et Rémi Ossant, à une formation destinée aux écoles doctorales et portant sur le partage des données et des codes et logiciels. En plus de parties théoriques sur la bonne gestion des données, des codes et logiciels, elle proposera des ateliers pratiques de dépôt de données dans l'entrepôt Recherche Data Gouv et de dépôt de code dans la forge institutionnelle de l'université de Tours. L'objectif de cette formation est de montrer aux doctorants comment lier leurs publications, leurs données et leurs codes sources.

Clément Plancq présente la formation sur l'outil de gestion de versions git prévue le 12 juin et qu'il a conçu avec Barthélemy Serres (Responsable CETU ILIAD3 et chercheur associé au LIFAT). C'est une formation pratique, sur une journée, qui permet de prendre en main l'outil très utilisé dans le monde académique.

Florian Hivert présente le cycle de deux formations en cours de montage sur la FAIRisation des données, prévu pour la rentrée universitaire 2025-2026. La première session abordera les aspects « Facile à trouver » et « Accessible » et la seconde portera sur les aspects « Interopérable » et « Réutilisable ». Ces formations devraient durer deux heures, en distanciel, et se veulent transdisciplinaires. Jorge Fins ajoute que le groupe de travail « FAIRisation des données » travaille actuellement sur le contenu notamment sur des exercices, des quizz, etc. et se prépare à un filage.

Rémi Ossant présente la dernière formation en cours de production sur les codes sources et logiciels comme produits de recherche. Elle aura pour but d'informer sur les enjeux et la manière de les valoriser dans le cadre de la recherche. Cette formation est prévue sur deux heures en distanciel.

Besoins de formation des membres de l'ADCVL

Laurence Rageot interroge les membres du comité sur leurs besoins en formation dans le cadre de la montée en compétences commune. Elle rappelle que l'Atelier a organisé en 2024 deux formations en s'appuyant sur les services de l'URFIST Paris : une première assez généraliste sur la science ouverte et une seconde sur les plans de gestion de données.

Nathalie Pothier indique qu'elle trouverait pertinent de prévoir une formation sur les codes sources et logiciels dans la science ouverte afin d'être capable de répondre aux sollicitations des chercheurs. Elle évoque les questions des licences, de Software Heritage, de HAL, et des *software papers*.

Thierry Bissy-Yandia indique que Patrick Moreau, CNRS Physique, est très compétent sur la valorisation des codes sources et des logiciels et qu'il assure des formations sur le sujet.

Jorge Fins mentionne le sujet des API de l'entrepôt Recherche Data Gouv. Laurence Rageot et Rémi Ossant lui répondent que Recherche Data Gouv anime régulièrement des formations très complètes pour la prise en main de l'outil.

Céline Girault évoque la problématique des métiers autour de la donnée (data analyst, data librarian, etc.) qui ont émergé ces dernières années et parmi lesquels il est parfois difficile de se retrouver. Laurence Rageot répond qu'une étude sur le sujet réalisée par le Ministère de la recherche sera présentée dans les prochaines semaines : « les métiers de la data ». Elle précise qu'elle fournira le lien de ce webinaire aux membres de l'Atelier.

Les réseaux européens, les outils et les structures internationales autour de la science ouverte comme EOSC sont également évoqués par Nathalie Pothier et Christine Le Bas.

Laurence Rageot indique qu'elle pensait proposer au comité de suivre la formation autour des codes sources et logiciels préparée par Rémi Ossant mais que les besoins semblent aller au-delà et qu'elle va réfléchir à un format plus complet sur le sujet. Elle note aussi l'intérêt pour les aspects européens et les métiers de la donnée.

II. Les groupes de travail de l'ADCVL

Accompagnement au dépôt des jeux de données dans un entrepôt

Dans le cadre de la MSH, Laurence Rageot a été sollicitée par Huma-Num pour créer un réseau de modérateurs pour Nakala. Clara Monmarché, Rémi Ossant et Valérie Prévitali (MSH Val de Loire) se sont portés volontaires.

Clara Monmarché précise que la modération porte sur la qualité documentaire du dépôt, les métadonnées. L'objectif à terme est que les données modérées soient moissonnées par le catalogue Recherche Data Gouv. Tous les modérateurs sont membres d'une liste de diffusion. Elle décrit le circuit de modération : lors d'un dépôt sur NAKALA, le déposant peut en demander la modération en sélectionnant le modérateur qu'il souhaite solliciter suivant un critère de périmètre géographique. Une fois la modération assignée à un modérateur, les échanges sont libres et en dehors de la plateforme. Le modérateur évalue le dépôt au moyen d'une grille de modération établie par Huma-Num. Il revient ensuite au déposant d'effectuer les éventuelles modifications. Une fois les données modérées, un label « Données modérées » est affiché sur le dépôt.

Aucune sollicitation n'est encore parvenue à l'équipe.

De plus, lors de toute nouvelle demande d'accès à Nakala, Huma-Num quels sont les modérateurs qui peuvent être sollicités.

En réponse à une remarque de Christine Le Bas, Clara Monmarché confirme qu'il s'agit d'une forme de curation des données. Thierry Bissi-Yandia demande si la modération se traduit par la publication du jeu de données. Rémi Ossant répond que les demandes de modération peuvent aussi se faire sur des données déjà publiées.

Laurence Rageot précise que Nakala a d'abord été un espace de stockage puis est devenu un entrepôt, mais sur lequel il n'y avait aucune modération. Aujourd'hui, Nakala est reconnu entrepôt de confiance aussi Huma-Num met-il en place un circuit de modération. La volonté

affichée est de faire monter la qualité des données mais sans rendre la modération obligatoire. Afin d'inciter les chercheurs à une plus grande qualité documentaire, la modération est désormais proposée et, à terme, seules les données modérées devraient être présentes dans le catalogue Recherche Data Gouv.

Christine Le Bas précise qu'il y a aussi un historique dans l'entrepôt Recherche Data Gouv car la curation n'était pas mise en place dès l'ouverture. Les premières formations à la curation n'ont eu lieu qu'en 2024. Aussi, une reprise de tous les jeux de données non modérés est en cours. Laurence Rageot répond que dans Nakala, il y a plus de dix ans d'antériorité et, qu'à son avis, il ne serait pas réaliste de tout reprendre.

Rémi Ossant précise que ce projet de modération est évoqué par Huma-Num depuis quelques mois mais n'a été mise en œuvre que depuis fin janvier.

III. Innovation versus/et ouverture des données

Rémi Ossant précise que, contrairement à l'intitulé dans l'ordre du jour, il ne va pas parler d'innovation au sens strict mais plutôt de valorisation.

Il rappelle que le sujet avait été évoqué lors de la dernière réunion du comité opérationnel et que des interrogations sur le concept de valorisation et son lien avec la science ouverte avaient été soulevées. Étant donné son positionnement Atelier de la donnée et C-VaLo, il a acquis des compétences sur ces deux aspects.

Il commence par rappeler la définition de la science ouverte : ouverture des procédés (souvent par le biais des publications), des données et des outils de la recherche (dont les logiciels). Cette ouverture des produits de recherche se fait par l'application d'une licence qui définit la réutilisation, le partage, les possibilités de modification, etc.

Il définit ensuite la valorisation comme le fait de donner le plus de visibilité et d'usages possibles à un objet. Dans cette acceptation, cela équivaut au concept de science ouverte parce qu'en ouvrant au maximum les produits de recherche ils gagnent en visibilité et usages. Néanmoins, le concept de valorisation dans le domaine académique comprend d'autres aspects. Ainsi, l'HCERES définit la valorisation dans le domaine académique comme « rendre utilisables ou commercialisables les connaissances et les compétences de la recherche ». Un enjeu financier entre donc en ligne de compte et arrive en opposition avec le concept de science ouverte.

La valorisation économique dans le milieu de la recherche académique consiste à trouver des débouchés économiques dans la société civile pour des innovations de recherche, donc à transposer du monde académique vers la société civile des procédés de recherche en les transformant en actifs économiques. Ainsi peut-on penser que ce type de valorisation est incompatible avec la science ouverte car il repose sur le droit d'interdire à d'autres acteurs d'utiliser cette technologie, ce savoir, cet outil, en vendant un droit de propriété exclusif et secret à un acteur. Cependant, au sein des institutions de l'enseignement supérieur, les deux politiques cohabitent.

Il précise que des droits spécifiques s'appliquent aux produits de recherche au sein des institutions. Ainsi, si un chercheur met au point un procédé scientifique et décide de le décrire dans une publication, l'institution n'a pas forcément le pouvoir de l'en empêcher alors qu'il va potentiellement divulguer un procédé qui aurait peut-être pu être transféré et donner lieu à un dépôt de brevet. En revanche, les institutions ont un droit de regard sur certaines procédures de recherche dont les logiciels.

Le logiciel, comme produit de recherche, est valorisable de deux manières :

- Via une valorisation économique qui peut se faire de différentes façons : par exemple dépôt du code du logiciel auprès de l'Agence pour la Protection des Programmes (APP). Cela permet d'horodater et de certifier le code, d'en prouver la paternité, l'enjeu étant de protéger l'aspect innovant, original du logiciel. D'autres actions comme le dépôt de brevet peuvent ensuite être engagées.
- Via la science ouverte, cela implique de faire en sorte que les droits associés au logiciel permettent de l'utiliser, de le modifier et de le redistribuer de manière libre. Pour cela il faut obtenir l'accord de l'institution. Le chercheur n'a pas le droit d'ouvrir un logiciel sans l'accord de l'institution car un logiciel est une œuvre de l'esprit protégée par le droit d'auteur. Dans le cadre de la recherche publique, les droits patrimoniaux sur les œuvres de l'esprit sont transférés à l'institution employant le chercheur qui conserve les droits moraux. Ainsi, seule l'institution a le pouvoir de décider si ce logiciel peut être ouvert ou non. Dans la réalité, les chercheurs ou les ingénieurs ayant la maîtrise technique du logiciel, décident souvent d'ouvrir les codes et logiciels ou de s'orienter vers une stratégie de valorisation économique.

Les institutions de l'enseignement supérieur et de la recherche ont des politiques de science ouverte et de valorisation économique et peuvent décider, en fonction des possibilités et des cas, laquelle mettre en œuvre suivant le caractère innovant et la possibilité de valorisation économique du produit de recherche.

De plus, les deux politiques peuvent coexister sur un même produit de recherche. Ainsi, Rémi Ossant indique avoir assisté à une présentation d'un chercheur du Commissariat à l'énergie atomique et aux énergies alternatives, Christophe Calvin, sur la politique institutionnelle sur les produits de recherche au sein du CEA. Le CEA a une tradition de valorisation économique mais réfléchit à la coupler à une politique de science ouverte alors que le domaine atomique a un fort enjeu de secret des données. Il a insisté sur l'importance de la volonté politique de l'institution car les politiques de valorisation économique et de science ouverte sont extrêmement cloisonnées.

Rémi Ossant ajoute qu'il a remarqué que ses collègues de C-VaLo considèrent systématiquement que si un projet innovant a une dimension science ouverte, la technologie est divulguée et que le projet ne mérite plus d'investissement.

Certains modèles économiques allient la valorisation économique et science ouverte. Un logiciel se valorise à la hauteur du service qu'il rend, l'enjeu est de savoir créer une structure autour du service rendu, même s'il est open source, pour réussir à créer de la valeur. Comme pour toute valorisation économique, il faut effectuer une analyse de marché, une analyse de valeur, etc.

Il n'y a pas de modèle homogène pour une valorisation économique d'un outil de recherche en open source, mais plusieurs exemples types peuvent être distingués :

- Création d'offre d'accompagnement, de formation ou de déploiement de l'outil open source. Autour d'un logiciel open source accessible et documenté, peut être créée une offre d'accompagnement et de déploiement payante. Cette formule permet à certains d'utiliser le logiciel en l'état, voire de le modifier, et à ceux qui n'ont pas les moyens ou les compétences techniques, d'avoir un accompagnement personnalisé. Cela peut aussi permettre d'adapter une solution open source à un domaine scientifique ou économique différent.
- Création d'un consortium dont l'entrée est payante et qui va permettre d'accéder à une communauté d'utilisateurs qui maîtrisent l'outil et l'utilisent. L'accès à ce « club »

permet d'échanger avec la communauté et d'avoir une maîtrise plus avancée de l'outil ou d'accéder à une assistance.

- Mise en place de projets dérivés adaptés à un domaine issu du monde socio-économique particulier. Cela va permettre de proposer des briques logicielles à implémenter sur un logiciel open source préexistant pour qu'il intègre de nouvelles fonctions spécifiques à une demande particulière. Cette brique logicielle peut être payante.
- « Dual licensing » qui consiste à mettre deux licences sur un seul outil. C'est donc un logiciel qui va avoir à la fois une licence open source et à la fois une licence propriétaire.

Enfin, Rémi Ossant évoque quelques éléments sur la science ouverte et la valorisation économique en général :

- La science ouverte possède une importante force d'attractivité en donnant à voir tout le procédé menant à un outil, un procédé de recherche, etc. Cela permet de détailler la création, les méthodes de travail et donne l'opportunité de créer des nouveaux projets dérivés mais aussi une souplesse pour le modifier, l'adapter en fonction des besoins. La science ouverte est donc créatrice de collaborations de recherche. À partir d'un projet qui n'était pas économiquement valorisé, on peut en créer de nouveaux qui peuvent avoir une vocation économique.
- Une production open source apporte une sécurité. Par exemple quand Airbus commande des productions logicielles à des prestataires, il est systématiquement demandé dans le cahier des charges que les logiciels soient livrés en open source afin que le logiciel puisse être transféré à une autre entreprise si l'entreprise qui l'a produit disparaît, modifie le produit ou perd les compétences. Du fait de l'open source, toute la documentation et tout le code sont accessibles et peuvent être modifiés. C'est une sécurité pour Airbus et cela permet de lutter contre l'effet « boîte noire logicielle ».
- La science ouverte permet la souveraineté. Ouvrir les produits de recherche selon les principes de la science ouverte, est la meilleure façon pour un pays d'être souverain, en revendiquant la paternité sur ses productions de recherche. De plus, contrairement au dépôt de titre classique qui peut avoir de très longs délais (plusieurs années parfois), l'ouverture d'un code, d'un logiciel, d'une base de données, des procédés de recherche peut être très rapide et donc assurer la paternité sur les productions de recherche.

Laurent Beunèche fait remarquer qu'il est intéressant de comparer la science ouverte et le code ouvert. Par exemple, il note une différence entre demander à un prestataire le code source du logiciel et le conserver au sein de l'entreprise et faire du code ouvert. La question est « à qui on ouvre ? ». Cela se contractualise et, dans ce cas, quel est le rapport avec la science ouverte ? Utiliser et ouvrir du code source ne fonctionne que s'il y a une communauté qui se crée et qu'elle a le temps d'étudier le code, d'assurer sa maintenance et de le stocker au bon endroit. Enfin, il soulève la question de la souveraineté car si on ouvre le code et qu'on le stocke sur GitHub, cela revient à le partager avec tout le monde. La question de l'ouverture nécessite un débat. À qui on ouvre ? Doit-on le rendre public ? De qui souhaite-t-on se protéger ? Et de qui on veut protéger son code ou ses données ? Actuellement, il constate qu'il y a trois ennemis : les états qui veulent faire de l'ingérence et de la déstabilisation dont la Russie et les États-Unis ; les groupes d'attaquants qui sont très organisés et très riches ; enfin les big tech. Les questions de sécurité et de souveraineté sont désormais très liées,

beaucoup plus qu'auparavant. Nathalie Pothier indique être d'accord avec Laurent Beunèche sur l'évolution du contexte politique international et la prudence qu'il doit amener.

Mélanie Fauconnier donne un lien vers une formation sur la valorisation de la recherche publique, au sens des services Valorisation : <https://mooc-innover-recherche-publique.fr/> Nathalie Pothier rappelle une intervention d'Isabelle Blanc dans laquelle elle disait « Ouvrir les données, oui mais... ». Laurence Rageot répond que le discours d'Isabelle Blanc était sur le fait que l'on ne pouvait dire qu'il faut ouvrir toutes les données. Il n'y a pas d'obligation d'ouverture des données aussi faut-il se poser la question des données que l'on ouvre et notamment de celles qui pourraient être brevetées ou qui répondent au RGPD. Il y a différents types de données que l'on ne peut pas ouvrir dont celles sur lesquelles il peut y avoir une valorisation économique.

Mélanie Fauconnier ajoute que ce sujet pourrait faire l'objet d'une présentation par un juriste spécialisé en propriété intellectuelle. La question de la valorisation vers le monde socio-économique dépend d'un certain nombre de réglementations dont le droit de la propriété intellectuelle et le droit de la concurrence. De plus, en tant qu'agent public, ce qu'un chercheur est autorisé à faire en termes de gestion de sa propriété intellectuelle est encadré. Donc ce n'est pas qu'une question de politique, ni d'établissement, ni de contexte sociopolitique, mais cela dépend d'un cadrage fort et très ancien. Elle ajoute que la notion de valorisation socio-économique n'est pas qu'une question de valorisation avec retour financier, c'est plus généralement établir un lien avec des tiers non académiques.

Pour résumer, la politique de science ouverte vise à ce que les données et la connaissance, soient utilisables par le plus grand nombre et la valorisation socio-économique encadre les modalités d'utilisation. Tout le cheminement est encadré, du résultat produit dans un laboratoire académique, à son utilisation concrète par un acteur, une entreprise, une collectivité territoriale (...). La propriété intellectuelle est l'outil qui permet de protéger ou de fermer en partie et de licencier ensuite des droits à des tiers. La finalité de cette licence peut être de rapporter des financements au laboratoire, mais pas toujours. Par exemple, en psychologie du travail, l'intérêt est d'obtenir de nouvelles données issues des terrains industriels.

Elle conclut que l'enjeu est dès que la connaissance ou la donnée est produite, les questions de l'ouverture de ces données soient posées. A-t-on intérêt à les ouvrir ? Pour quelles raisons ? À qui ? Comment ? A-t-on intérêt à les fermer ? A-t-on l'obligation de les fermer ? Par exemple, dans le cas d'un projet ANR avec un industriel, il y a signature d'un accord de consortium avec une clause de propriété intellectuelle qui indique si les résultats sont en copropriété, fermés ou pas. La situation lui semble plus complexe que ce qui a été présenté.

Christine Le Bas indique que les équipes INRAE sont coachés par un juriste qui a participé à l'écriture de la loi pour une république numérique. Il les incite à ouvrir les données et n'évoque pas la valorisation comme un moyen de les fermer. Mélanie Fauconnier répond que ce n'est pas un moyen de fermer les données mais une obligation. Elle reprend l'exemple de la collaboration avec un industriel avec qui le chercheur s'est contractuellement engagé à fermer les données, la question ne se pose donc plus.

Christine Le Bas ajoute qu'il leur a dit que le droit de propriété intellectuelle des organismes publics n'est pas un moyen de fermeture des données. Mélanie Fauconnier indique que, dans le cas d'une invention, l'inventeur public est dans l'obligation de prévenir son employeur. C'est indiqué dans le code de la recherche et dans le droit de la propriété intellectuelle. Le fait de

publier un certain nombre d'éléments peut empêcher, dans certains cas, de déposer des brevets. Par exemple, si on a inventé un procédé, y compris d'intérêt général comme pour soigner une maladie, on ne pourra jamais le transformer en médicament parce qu'aucun industriel ne se lancera dans son développement.

Pour Christine Le Bas, la difficulté est d'arriver à savoir quand cet aspect valorisation intervient. Mélanie Fauconnier indique que le plan de gestion de données devrait permettre de s'interroger sur ce qu'on a le droit de faire ou pas avec les données d'un projet. Elle regrette que le groupe sur les questions juridiques n'ait pas pu se mettre en œuvre sur les questions de propriété intellectuelle, de droit de la concurrence et de droit des contrats, qui sont très spécifiques et différentes du droit des données.

Christine Le Bas ajoute que le coach leur a indiqué que le droit de propriété intellectuelle de tiers ne signifiait pas impossibilité d'ouvrir. Il fait une différence entre ouverture des données et réutilisation. Des données peuvent être ouvertes en indiquant au réutilisateur qu'elles répondent à des droits de propriété intellectuelle et qu'il ne pourra pas les réutiliser librement. Elle pensait qu'à partir du moment où la réutilisation était limitée, cela empêchait la diffusion mais d'après lui ce n'est pas le cas. Concernant la valorisation des logiciels, elle n'arrive pas à savoir quel texte de loi fait référence, lors des discussions avec des juristes, cette question n'est pas abordée. Elle s'interroge sur la justification juridique pour fermer ou ouvrir un code. Elle souhaiterait être aidée sur ces questions-là, soit par un membre de l'atelier compétent dans ce domaine, soit par un extérieur. Nathalie Pothier relie ce sujet à la question des formations évoquée précédemment. Cela nécessiterait l'intervention d'un intervenant de l'URFIST spécialisé en droit par exemple.

Christine Le bas précise que dans certains domaines on ferme d'abord et on ouvre « si » comme dans le domaine de la santé. Alors que dans de nombreux autres domaines, on est plutôt dans une obligation d'ouverture. Mais elle revient sur l'argument juridique qui permet de dire qu'un logiciel ne doit pas être ouvert car il peut être valorisé.

Laurent Beunèche ajoute que pendant le processus, la donnée passe par des états différents avec des contraintes différentes. Il s'interroge, par exemple, sur les répercussions, pendant le dépôt de brevet, du vol ou de la divulgation d'une donnée ? De même, suivant l'état les données elles-mêmes ne sont pas forcément les mêmes. On peut, par exemple, enlever des données sensibles à un certain moment du processus. Il pense qu'il serait préférable de voir les choses en termes de processus. On peut être amené à fermer au début puis à ouvrir ensuite. Tout ouvrir du début à la fin du processus de recherche, lui paraît être une démarche étrange.

Georges Fins pense qu'il faut entrer dans la catégorisation et dans l'exemple concret pour se rendre compte de la réalité économique, de sa relation avec la science et en particulier avec les logiciels. Il lui semble difficile de se mettre d'accord quand on a des réalités différentes. L'innovation est l'objectif de tout le monde mais c'est peut-être ce qui est le plus rare. Il prend pour exemple une application pour recenser les espèces de la faune et de la flore dans les forêts, la relation entre logiciel, recherche, études et évaluation est visible. Il faut définir un périmètre viable et qui doit le rester viable. La question de la valorisation économique est évidente. De plus, il y a des aspects de « méthode » pour aller vers une baisse des coûts et des gains en précision. Il faut de la recherche et développement pour y arriver. Ensuite, il ne faut pas oublier, la question du coût de maintenance pour définir le coût de revient, même si ce n'est pas lié à la science ouverte. A chaque question, des catégories différentes entrent en jeu

et permettent de se rendre compte qu'il y a des relations entre science de façon générale, données en particulier et réalité économique, ou gestion d'un capital ou d'une ressource pour qu'elle puisse, au minimum, ne pas disparaître. Il ajoute que la science ouverte est toujours présente car tout a vocation à tomber dans le domaine public mais il faut prendre en compte la réalité de maintenance qui est une réalité économique. Par ailleurs, il s'interroge sur les catégories particulières de consortiums que sont les associations et les sociétés coopératives. Il prend l'exemple d'un logiciel produit dans une logique publique-privée par une université française. Il fonctionne avec une société éditrice et une association mais aussi par l'organisation de formations, un master professionnel... Il lui semblerait intéressant d'étudier des exemples de ce type pour voir ce qui a fonctionné ou pas. Il conclut sur le fait qu'il n'y a pas que l'État d'un côté et les entreprises classiques de l'autre.

Rémi Ossant ajoute qu'il ne faut pas rester dans une logique de chapelles et aller discuter avec les services de valorisation, pour éclaircir ces questions : « Peut-on faire de la science ouverte ? Est-ce qu'on peut faire de la valorisation économique ? Quel degré de souplesse peut-on avoir ? »

IV. L'écosystème Recherche Data Gouv

Laurence Rageot évoque les prochaines journées Recherche Data Gouv de juin à Lille. Elles seront sur le même format que les précédentes et seront portées par les centres de référence thématiques. Elles porteront sur les « dimensions disciplinaires et thématiques des données de recherche ». Elles ont pour ambition de discuter de la diversité des données et voir, suivant la discipline ou la thématique, quelles sont les ressources qui peuvent être sollicitées, quels centres de référence sont experts. Enfin, elles permettront de discuter des points de vigilance à prendre en compte quand on accompagne des chercheurs qui produisent des données comme les données personnelles, les données sensibles, les données 3D... Le programme n'est pas encore connu.

Par ailleurs, Laurence Rageot rappelle que Recherche Data Gouv demande sa labellisation en tant qu'infrastructure de recherche et souhaite être inscrite sur la feuille de route nationale des infrastructures de recherche en cours de réécriture. Les responsables d'Ateliers de la donnée ont été informés que des indicateurs allaient leur être demandés pour alimenter ce dossier. Les indicateurs envisagés devraient porter sur les ressources, les ETP, les personnes et les services impliqués ; les services proposés, leur nombre, l'organisation, les périmètres ; les usages, le nombre de personnes accompagnées, le nombre de personnes qui ont suivi des formations, le niveau de satisfaction. Ce dernier point semble beaucoup plus difficile à quantifier, aussi les responsables d'Atelier ont décidé de proposer des indicateurs à Recherche Data Gouv qui soient représentatifs des activités et aisément calculables par chacun. Pour cela, ils finalisent un travail initié depuis près d'un an pour obtenir une typologie de services qui soit commune quel que soit l'intitulé donné par chacun.