

# Comité opérationnel de l'Atelier de la donnée Centre-Val de Loire du 20 juin 2025

---

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

**Horaires** : 14h00 – 16h00

**Participants** : Anne AZANZA (Direction des bibliothèques & des archives, Université de Tours), 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), Florian HIVERT (MSH Val de Loire), Christine LE BAS (Correspondant données, INRAE), Clara MONMARCHÉ (ADCVL), Nathalie MOULARD (OSUC), 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), Alain RAVAZ (MSH Val de Loire)

**Ordre du jour** : 1. Informations générales concernant l'ADCVL – 2. Les groupes de travail de l'ADCVL - 3. Restitution de l'enquête sur les 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 02/09/2025]*

## Compte-rendu

---

Laurence Rageot introduit la séance en invitant Nathalie Moulard, nouvelle membre du comité opérationnel de l'ADCVL, à se présenter. Elle travaille avec Nathalie Pothier à l'OSUC, à Orléans, au CNRS. Elle est en charge des systèmes d'information et de la valorisation des données. S'en suit un tour d'écran et de table afin que chaque membre du comité opérationnel se présente.

### I. Informations générales concernant l'ADCVL

#### a. Site et guichet unique – demandes

Clara Monmarché indique que, depuis mars, l'atelier a surtout reçu des demandes d'accompagnement à la rédaction de PGD :

- Elisabeth Herniou (IRBI), projet ANR FLYPATH (biologie des insectes), rendez-vous initial
- Fabrice Foucher (PRISME), chaire industrielle DELHYCE (ingénierie des produits et des procédés), rendez-vous initial fait par Clara Monmarché et Pascale Solon
- Julie Loiseau (VALLOREM), projet ValPostConso (sciences sociales), rendez-vous initial, définition des produits de recherche et relecture de la première version du PGD
- Caroline Giraudeau (PAVEA), projet ANR SAPS ACAPULCO (psychologie), rendez-vous initial fait par Clara Monmarché et Mélanie Fauconnier. Cette chercheuse travaillait également sur la rédaction des accords de consortium, cela a donc été l'occasion de tester l'accompagnement mutualisé sur ces deux sujets pour lesquels il y a des questions communes
- Thaïs Nunez et une doctorante du LEO sur un projet européen

Clara Monmarché précise que la majorité des personnes ayant contacté l'Atelier n'avaient pas participé à la formation PGD proposée en début d'année.

## b. Indicateurs de l'Atelier

Laurence Rageot évoque les réflexions sur les indicateurs à mettre en place pour témoigner de l'activité de l'Atelier, à la fois pour les besoins de l'atelier en lui-même, mais aussi pour ses tutelles, le prochain comité de pilotage étant en juillet. Laurence Rageot, Rémi Ossant et Clara Monmarché ont commencé à y réfléchir.

Certains indicateurs sont importants et relativement aisés à mettre en place :

- le nombre de sollicitations sur le guichet unique,
- les demandes directes qui arrivent à Clara Monmarché et Laurence Rageot, et peut être également aux autres membres du comité opérationnel. Par rapport au plan de gestion de données qui sont en cours d'accompagnement, il est intéressant de savoir si les sollicitations font suite à la formation
- le nombre de formations
- le nombre de personnes formées
- le nombre de présentations de l'Atelier qui peuvent se faire dans différents cadres. Laurence Rageot indique qu'une présentation est prévue en juillet devant la commission recherche santé de l'université de Tours.

La question des ressources humaines se pose également :

- Le nombre de personnes impliquées dans l'atelier : nous pouvons nous baser sur le nombre de membres du comité opérationnel, sachant qu'il y a des personnes qui s'impliquent plus ou moins
- Le nombre d'ETP est plus compliqué à définir, à part pour Rémi Ossant et Clara Monmarché, qui sont respectivement à 50% et 100% pour l'Atelier. Laurence Rageot indique qu'il lui est difficile de définir le temps qu'à titre personnel elle consacre à l'Atelier et elle ne sait comment comptabiliser la participation des autres membres du comité opérationnel. Lors de la constitution du dossier de labellisation, une liste des personnels impliqués et de leur quotité était demandée. Les noms avaient été fournis par chaque partenaire et elle avait attribué par défaut 0,1% d'ETP à chacun. Cette répartition était d'autant plus artificielle que l'Atelier n'existait pas. Pour Recherche Data Gouv, le nombre d'ETP de l'ADCVL correspond à cette estimation.

Se pose aussi la question de comment mesurer la satisfaction des utilisateurs.

Laurence Rageot invite les membres à donner leur avis sur les indicateurs évoqués, notamment sur la façon de comptabiliser les ETP.

Mélanie Fauconnier indique que pour C-VaLo, auquel elle contribue, il lui est demandé de remplir un Excel avec une granularité à la demi-journée, ce qui peut être assez lourd. Elle propose que la participation aux réunions du comité opérationnel de l'ADCVL soit comptabilisée pour mesurer l'implication des personnes, de tracer les autres réunions du périmètre de l'ADCVL et d'en faire une remontée à Laurence Rageot (une fois tous les 3 ou 6 mois).

Anne Azanza évoque le suivi de l'accompagnement des usagers pour le service Micado : le temps passé est noté ainsi que la personne et la date. A la fin de l'année, il est possible d'avoir une quantification. Stéphanie Puel confirme que c'est peut-être la meilleure façon de procéder.

Mélanie Fauconnier demande comment sont fait les bilans et quels en sont les enjeux. Y- avait-il des objectifs au moment du dépôt du dossier? Par exemple, le nombre d'étudiants sensibilisés à la science ouverte, est-ce qu'on parlait d'un indicateur à T0 et il fallait démontrer une progression ? Ensuite quelles sont les personnes qui lisent ces indicateurs et quels sont les enjeux ? Est-ce que c'est par rapport à des tranches de financement sur l'ADCVL ? Quel est le « risque » et donc quel est le niveau d'exigence qu'on se fixe dans la très forte précision des indicateurs qu'on remplit ? On peut, par exemple, tracer, dans la plateforme, quel service de l'ADCVL a été sollicité, et puis affecter un pourcentage à tous les membres de ce service, un quart de jour à chaque fois qu'il y a une demande, par exemple. On peut avoir quelque chose qui se fait un peu automatiquement.

Laurence Rageot répond que ces indicateurs n'ont pas de rapport avec des tranches de financement de l'ADCVL. Elle précise qu'un tableau reprenant toutes les demandes arrivées par le guichet unique

et celles qui sont arrivées directement par mail (voir présentation) est déjà complété. Elle détaille les informations recensées ainsi que pour le tableau sur les formations.

Sur les enjeux, elle ajoute que le premier fichier de suivi des demandes a été initié pour un usage interne de suivi des dossiers. Ensuite, elle a reçu une demande de Recherche Data Gouv, qui travaille à son inscription sur la feuille de route des infrastructures de recherche, et doit fournir des indicateurs sur tous les membres de l'écosystème. Après avoir sollicité le bureau des ateliers, les responsables de Recherche Data Gouv ont décidé d'utiliser les éléments fournis dans les dossiers de demande de labellisation, bien que certains dossiers datent de 2021, celui de l'ADCVL est de 2022. Ils ont pris cette décision car ils ont été confrontés à différents problèmes. En premier lieu, le besoin d'indicateurs pour tous les membres de l'écosystème et pas uniquement les ateliers de la donnée. D'un autre côté, au niveau des ateliers, il est difficile de savoir quels sont les services offerts. En effet, chaque atelier a des services un peu différents, ou tout du moins avec des noms différents. De plus, le périmètre des services ne sont pas identiques. Elle prend pour exemple le service plan de gestion de données, que recouvre-t-il ? Une sensibilisation pour renseigner sur le contenu du PGD ? Une relecture du PGD ? Une participation à l'écriture du PGD, certains ateliers proposant d'écrire avec le chercheur le plan de gestion de données ? Ces trois services sont très différents. Cette réflexion sur les indicateurs rejoint un travail en cours sur l'harmonisation des services et du niveau d'accompagnement pour chacun réalisé par les responsables d'atelier. Le sujet a été discuté au cours des trois derniers séminaires Recherche Data Gouv, sans vraiment réussir à se mettre d'accord, notamment sur le type d'accompagnement. Entre sensibilisation et formation et entre accompagnement et mise en œuvre les définitions ne sont pas identiques selon les ateliers. Enfin, Recherche Data Gouv cherche des chiffres d'activité extrêmement globaux. Ils ne sont pas intéressés par le chiffre de sollicitations reçues via le guichet unique mais par le nombre de personnes touchées par les ateliers. Cela revient à regrouper les sollicitations, les formations, les présentations. Ils souhaitent des chiffres suffisamment significatifs pour justifier leur labellisation en tant qu'infrastructure.

En revanche, Laurence Rageot pense qu'il faut aller plus dans le détail au niveau local.

Mélanie Fauconnier constate que le besoin de Recherche Data Gouv est très macro et que ce sont les ateliers qui les aident dans leur propre démarche de reconnaissance. A l'inverse, le besoin de l'atelier est plutôt micro. Elle suggère donc de partir du besoin de l'atelier car il sera plus facile d'agréger des données fines en grosses masses. Il faut donc réfléchir aux besoins, pour nos établissements ou pour nous-mêmes, du type d'indicateurs pour suivre l'activité sans pour autant que ce soit trop chronophage.

Laurence Rageot répond qu'elle ne souhaite pas aller vers un niveau de détail trop important mais elle suppose que les partenaires de l'Atelier, universités, CNRS, INRAE... auront besoin de quantifier le temps passé par leurs agents. Elle propose donc de compter les temps de réunion du comité opérationnel et d'y ajouter les réunions spécifiques, comme celles évoquées par Clara Monmarché sur l'accompagnement PGD. Pour les autres activités, elle propose d'envoyer un message une fois par trimestre pour que chacun indique s'il a passé du temps pour l'atelier, sans descendre à une précision d'une demi-heure. Ces informations seront consignées dans un tableau à destination de l'Atelier uniquement mais qui permettra d'avoir des chiffres plus fins quand les tutelles se manifesteront.

Christine Le Bas pose la question des activités à associer à l'ADCVL. Etant aussi RDO (Research Data Officer) au niveau d'INRAE, doit-elle comptabiliser les activités menées à ce titre aussi comme des activités ADCVL ou ne comptabiliser que les activités spécifiques ADCVL comme les réunions ou la participation à des formations ? Elle ne connaît pas l'activité des autres agents RDO de INRAE. Laurence Rageot évoque Édouard Guiton qui a participé à certaines réunions. Elle propose de remonter les deux activités en spécifiant si c'est au titre de RDO ou de l'ADCVL. Christine Le Bas accepte.

### c. Formations

- **Formation doctorale « Comment partager vos données et vos codes et logiciels »**

Clara Monmarché indique que le 15 mai a eu lieu la formation doctorale « comment partager vos données, codes et logiciels ». Il n'y avait que trois participants mais la formation a été maintenue afin

de tester le format avec une partie théorique et une partie pratique où les participants étaient invités à vérifier des données et du code et à les partager. Elle se concluait par la démonstration du lien entre les données, le code et la publication au travers des différents outils. Une session était aussi proposée à Orléans, mais elle a été annulée parce qu'il y avait moins de 5 participants (4 inscrits).

Laurence Rageot indique que la communication avec les écoles doctorales n'a pas été simple.

- **Formation « GIT de survie »**

Le 12 juin, a eu lieu la formation « GIT de survie » à Tours. Une deuxième session est proposée le 15 septembre à Orléans et a été complète très rapidement.

Clément Plancq indique que la formation s'est bien passée. Il l'a animé avec Barthélemy Serres (MSH CETU ILIAD3). Deux personnes ne sont pas venues, ce qui est dommage car ils avaient dû refuser des inscriptions. Les participants ont été satisfaits et ont exprimé le souhait d'une suite. Une formation GitLab a déjà été faite avec Barthélemy Serres et Guillaume Blin (DSI de l'université de Tours). Ils essayeront d'en programmer une nouvelle session.

Il ajoute que, pour la formation du 15 septembre, Rémi Ossant les accompagnera en tant que formateur afin d'avoir plus de flexibilité sur l'organisation.

Laurence Rageot précise que pour la session d'Orléans une dizaine de personnes sont sur liste d'attente. Lors d'une réunion sur la science ouverte à l'université d'Orléans avec la nouvelle VP Recherche, le VP numérique, la chargée de mission sciences ouvertes et le SCD (Pascale Solon et Philippe Serisier), Carine Lucas avait indiqué un besoin sur GIT pour les gens du mésocentre CaSciModOT. L'information leur a donc été transmise et ils se sont inscrits dans la journée. Aussi quand la communication a été faite sur les autres canaux, elle était déjà quasiment complète. Il faudra donc prévoir de proposer une nouvelle date à Orléans.

Clément Plancq ajoute que cette formation a été l'occasion de sortir de leur public habituel des SHS, car parmi les participants toutes les disciplines étaient représentées, avec des pratiques et des questions différentes. De plus, ils ont eu des questions sur les licences, l'ouverture des codes, les forges logicielles. Ils ont essayé de répondre mais n'ont pas forcément répondu à tout et cela dépend aussi des plateformes technologiques qui sont proposées par les institutions dont dépendent les participants. Il s'interroge sur la possibilité de faire remonter des questions aux gestionnaires de la plateforme GitLab de l'Université de Tours.

Laurence Rageot rappelle qu'il n'y a pas de plateforme à Orléans. Maryse Castan ajoute que le CNRS propose une plateforme Git vers laquelle elle oriente les chercheurs car le RSSI est très exigeant sur la façon dont les codes sont produits et diffusés. La DSI de la délégation a diffusé l'annonce de la formation aux ASR (Administrateurs Systèmes et Réseaux) du campus et de l'Université d'Orléans car elle ne propose pas ce genre de formation. Elle signale qu'elle participe aux réunions de l'ADCVL pour avoir de l'information sur ce qui est fait. De plus, elle peut orienter les gens, lors de ses contacts un peu élargis dans les labos, notamment Tours, Orléans, jusque Poitiers et Limoges. Cependant, elle ne sait pas toujours comment se positionner.

Nathalie Moulard indique qu'un GitLab est ouvert aux unités qui font partie de l'OSUC, dont le LPC2E, l'ISTO et l'Observatoire de Nançay. Il est réservé aux projets de ces unités. Cependant, il est possible, au travers de projets, d'accéder à des GitLab, d'avoir accès à des outils de versionnement d'autres institutions, par exemple, quand on travaille sur un projet CNES, on va avoir accès au Git du CNES. Ainsi, si les formations Git sont suffisamment généralistes, elles donnent une couche d'informations qui permettent de rentrer dans les GitLab spécifiques avec des pratiques et des règles spécifiques. Le GitLab de l'OSUC n'a pas les mêmes règles de dépôt que sur celui du CNRS car il y a certains types de données qui ne sont pas acceptées car trop volumineuses. Par exemple, celui du CNRS va des notebooks au contraire de celui de l'OSUC car ils peuvent comprendre des images et sont trop volumineux à versionner. Cela dépend de la politique de chaque GitLab et de chaque établissement.

Clément Plancq ajoute que la formation « Git de survie » aborde surtout la première étape, c'est-à-dire l'outil de versions Git. Ils abordent un peu les forges et prennent l'exemple du GitLab de l'Université de Tours qui est ouvert à toutes les personnes qui ont un compte sur la fédération d'entités Renater. C'est un GitLab assez ouvert, qui est utilisé à titre d'exemple pour que tout le monde travaille sur les mêmes fichiers. Mais après il peut y avoir des variantes selon les forges. Dans le monde académique et universitaire, c'est surtout GitLab qui est utilisé, et effectivement, il y a des règles d'utilisation en fonction des établissements.

Maryse Castan répond que c'est pour cela que la formation a intéressé beaucoup de personnes, car ça permet d'avoir une première couche et après de s'adapter au Git qu'on a « sous la main ».

Laurence Rageot ajoute que pour la communication sur cette formation, sur conseil des membres du comité opérationnel, elle avait contacté les services formation permanente des institutions. Le service de l'université de Tours a accepté de la mettre au catalogue des formations. La session du 12 juin a donc été gérée par eux avec une inscription sur le site de l'ADCVL et via Geforp. Ces deux canaux d'inscription n'ont pas été compliqués à gérer, il a cependant fallu se synchroniser avec la collègue pour ne pas avoir avec trop de monde. Elle n'a pas eu de réponse du service formation permanente de l'université d'Orléans. Côté CNRS, l'annonce de chaque session a été envoyée aux correspondants formation dans les laboratoires, charge à lui de communiquer ou pas à ses collègues. Cependant, la session de septembre était déjà complète quand l'information a été envoyée. De plus, elle a été diffusée sur un périmètre plus large que celui de l'ADCVL et un chercheur de Limoges a contacté l'Atelier pour savoir s'il pouvait suivre la formation. Pour INRAE, Christine le Bas avait donné le contact d'un collègue qui fait une lettre hebdomadaire sur les formations. L'information a ainsi été transmise et une personne d'INRAE a participé à la session de Tours et plusieurs sont inscrites pour la session à Orléans. Christine le Bas précise que pour la session d'Orléans, la communication via CaSciModOT ayant devancé celle d'INRAE, elle a rediffusé elle-même l'information auprès des agents INRAE d'Orléans. Laurence Rageot ajoute qu'il y a des personnes du BRGM qui sont inscrites sur la session de septembre, et qu'ils ont probablement eu l'information via CaSciModOT.

- **Formation « FAIRisation des données »**

Clara Monmarché évoque la formation « FAIRisation des données » en cours de préparation par le GT FAIR. Elle est prévue à l'automne 2025, la date restant à déterminer. Rémi Ossant a envoyé un événement afin de définir une date entre fin octobre et début novembre. Cette formation sera composée de deux sessions de deux heures en visio. Ce sujet sera abordé plus en détail dans le point dédié aux groupes de travail.

- **Formation pour les membres de l'ADCVL « Science ouverte, codes et logiciels »**

Rémi Ossant indique que, suite à la discussion de la dernière réunion du comité opérationnel, il a travaillé sur le montage d'une formation pour les membres de l'ADCVL sur les codes et logiciels. Il assurera une partie de la présentation, sur les aspects génériques autour des codes et logiciels, de la science ouverte, etc. En revanche, un intervenant extérieur parlera des aspects de propriété intellectuelle autour du logiciel, car c'est un sujet très spécifique, pour lequel il est préférable de faire appel à un spécialiste. Il est en discussion avec plusieurs personnes, notamment du CNRS.

Cette formation serait prévue soit sur une journée, soit sur une demi-journée, avec :

- une première partie sur tous les aspects logiciels classiques assurée par Rémi Ossant : qu'est-ce qu'un logiciel de la recherche ? Qu'est-ce qu'on entend par les bonnes pratiques de la science ouverte pour le développement logiciel ? Comment le valorise-t-on selon les principes de la science ouverte ? Quels entrepôts ? Comment dépose-t-on ? L'idée est de partir des questions qui se posent pour les données de la recherche et de se demander si elles existent aussi pour les codes et logiciels, quelles solutions existent, et lesquelles sont privilégiées par le Ministère ou la communauté scientifique.
- Une deuxième partie axée propriété intellectuelle avec notamment les licences, pour lesquelles il y a beaucoup de demandes. L'idée n'étant pas de faire de nous des spécialistes du

domaine mais d'avoir une base de connaissances communes et d'être capable de répondre à des questions de base, de savoir si une question est hors de nos compétences et qu'il faut s'adresser un juriste ou un spécialiste du logiciel.

Des rendez-vous sont prévus la semaine suivante avec des spécialistes des questions juridiques et il espère avoir la capacité de proposer des dates avant les congés d'été .

## **II. Les groupes de travail de l'ADCVL**

### **a. GT FAIRisation des données**

Clara Monmarché revient sur la formation FAIRisation des données, qui sera proposée sur 2 sessions de 2h, la première sur le F et le A, la deuxième sur le I et le R. Pour chaque session, il y aura des définitions des notions connexes aux principes FAIR et des exemples de bonnes et de mauvaises pratiques. Ils vont essayer d'intégrer des cas pratiques pour générer de l'interactivité avec les participants, sachant que ces formations se feront en visio. Le plan de la formation est prêt et le prochain rendez-vous de préparation est prévu le 4 juillet.

Nathalie Pothier demande si pour les cas pratiques ou les bons et les mauvais exemples plusieurs disciplines seront représentées et suggère de faire appel aux membres du comité opérationnel pour en trouver.

### **b. GT Plans de gestion de données**

Clara Monmarché indique que deux supports sont en cours de préparation :

- **Fiche produite par le SPIV de l'université de Tours**

L'ADCVL a été sollicité par le service partenariat innovation valorisation de l'université de Tours (SPIV) qui souhaitait faire une fiche d'information à destination des chercheurs sur les PGD en renvoyant vers l'ADCVL pour l'accompagnement. Ils souhaitaient l'avis de l'Atelier sur le contenu de cette fiche. Ils ont fait le choix de la tourner d'une façon ludique, en comparant les données à des pépites, les pépites d'or étant celles qu'on ne peut pas partager, et les pépites de chocolat, celles qu'on partage. Elle indique qu'après discussion avec Rémi Ossant et Laurence Rageot, elle a émis quelques réserves sur le fond de cette fiche, et elle invite les membres du comité opérationnel à donner leur avis.

Mélanie Fauconnier souhaite préciser le contexte de cette fiche et indique qu'elle est encore largement modifiable. Le service SPIV accueille une stagiaire (Riyane Gharbi) qui va partir dans peu de temps, mais qui avait comme mission d'aider à la communication sur tous les aspects partenariat, valorisation, propriété intellectuelle à destination des chercheurs et éventuellement des doctorants. Elle a commencé par faire une enquête auprès des directeurs d'unités et d'un certain nombre de chercheurs dont les chargés d'affaires lui ont fourni la liste, pour qu'elle repère les sujets sur lesquels ils s'interrogeaient régulièrement. Certains sujets identifiés concernaient l'expertise du SPIV et d'autres non, comme la science ouverte ou les PGD. Il a ainsi été décidé de faire une fiche que les chargés d'affaires puissent distribuer lors de leurs échanges lors des montages de projets, fiche qui renvoie sur l'ADCVL présenté comme ressource sur le sujet des PGD. En revanche, le fait d'avoir un ton humoristique peut être mis en question. Le service communication a trouvé que c'était trop ludique, que le style n'était pas adapté à une fiche pratique. Ensuite, on peut avoir une idée sur le parti pris par rapport à « or on ferme et chocolat, on partage dans les cookies ».

Anne Azanza indique que le ton humoristique ne la dérange pas, en revanche, elle trouve que la métaphore or/chocolat n'est pas complètement adaptée. Si le document est fait pour renvoyer vers l'ADCVL, vers des gens qui vont entrer plus dans le détail et dans le côté sérieux des choses, c'est très bien. Le but étant d'attirer l'attention, la fiche remplit ici bien son rôle tout en la faisant tiquer sur la distinction or/chocolat. Stéphanie Puel rejoint Anne Azanza sur le bénéfice du côté ludique et sur la métaphore des pépites, car les pépites d'or se partagent aussi selon elle. Elle ajoute qu'il faudrait avoir un retour utilisateur, pour voir comment un chercheur a reçu la fiche et comment il l'a prise.

Mélanie Fauconnier indique que Riyane Gharbi l'a soumise à un panel de chercheurs mais elle ne connaît pas le résultat. Sur le fond elle a une interrogation sur la nécessité de renvoyer sur Doranum

dans la mesure où le site de l'ADCVL a une partie ressources. Elle préférerait que le message soit directement dirigé vers l'ADCVL. Nathalie Pothier répond qu'il n'est pas inutile de faire une piqûre de rappel sur Doranum. Clara Monmarché indique que la version présentée ici est la deuxième version de la fiche, qu'une première relecture avait déjà été faite et qu'elle avait suggéré de faire plus de liens vers Doranum ou DMP OPIDoR. Elle indique qu'elle n'a pas commenté le ton ou le niveau d'information de la fiche car elle considérait que cela relevait du choix éditorial. Par contre, elle avait déjà discuté du choix de la notion de pépite d'or / pépite de chocolat lors de la première relecture. La réponse du SPIV a été que la notion pépite d'or renvoie à des données qui peuvent être valorisées économiquement, donc qui ne peuvent être partagées. Les pépites de chocolat correspondent à l'aspect nourricier. Anne Azanza ajoute que le PGD est utile aussi aux chercheurs, le chocolat nourrit mais rend aussi intelligent et donne du magnésium. Ainsi, si les données sont mieux organisées, le chercheur va gagner du temps dans sa recherche et va être en mesure de mieux utiliser et valoriser les données avant même de les partager. Elle trouve que c'est aussi un argument qui pourrait figurer sur la fiche car le PGD n'est pas seulement utile pour dire ce qu'on partage ou pas.

- **Fiche pratique de l'ADCVL sur les PGD**

Clara Monmarché indique qu'elle a travaillé sur une fiche pratique sur les plans de gestion de données car, lors de ses rendez-vous avec des chercheurs elle a eu des questions sur les délais et moyens de transmission des PGD, et notamment de la date à partir de laquelle courent ces délais. Elle a donc décidé de faire une fiche qui résume quand et comment doit être transmis le plan de gestion de données en fonction des exigences de l'ANR et du programme cadre Horizon Europe. Elle a sollicité les services de valorisation. Mélanie Fauconnier lui a fait un retour pour les exigences de l'ANR Franziska Metzinger et Maëlie Olouman de la Cellule Mutualisée Europe Recherche (CMER) doivent lui faire un retour pour Horizon Europe.

Une piqûre de rappel est faite sur les accords de consortium, par exemple, s'il y a des conditions juridiques particulières à respecter qui doivent aussi être mentionnées dans le plan de gestion de données.

- **Quelle articulation entre l'ADCVL et les services valorisation dans les institutions ?**

Clara Monmarché indique que des questions communes se posent quand on rédige un plan de gestion de données et des accords de consortium.

La question a été discutée lors d'une réunion du PUI Loire-Vallée Innov' et Chloé Petit du service valorisation d'INRAE a indiqué avoir eu un cas dans lequel un PGD était en contradiction avec ce qui avait été défini dans les accords de consortium, le PGD a dû être modifié et retransmis à l'ANR. Elle a ainsi demandé à Laurence Rageot qui participait à la réunion, que l'ADCVL lui signale les sollicitations de collègues d'INRAE pour un accompagnement sur un plan de gestion de données de façon à voir si un accord de consortium a été conclu. La demande a également été faite par le CHU de Tours pour la question des données sensibles mais le CHU n'étant pas membre de l'ADCVL, aucune demande de chercheurs du CHU n'est traitée.

Rémi Ossant et l'un de ses collègues de C-VaLo ont cherché à savoir dans quelle mesure il était possible de détecter les innovations au travers des plans de gestion de données car l'un des objectifs du PUI est de détecter l'innovation. Ils ont conclu qu'il est possible de détecter un potentiel d'innovation et de transfert vers la société civile d'un projet de recherche via le PGD mais, par contre, ils se posent la question du moyen d'accès aux PGD. Ils pensent que ce sont les personnels d'accompagnement qui doivent être sensibilisés. Aussi, pendant la journée PUI du 16 juin, a été émise l'idée de monter un groupe de travail PGD au sein du PUI dont l'objectif serait de développer des outils, des grilles d'analyse pour être capable de détecter ces potentiels d'innovation.

Christine le Bas indique être surprise par la demande de Chloé Petit car tous les PGD INRAE ne vont pas faire l'objet d'une demande à l'ADCVL ; si elle veut avoir accès aux PGD des projets faits par des agents INRAE, il faut qu'elle communique avec eux. Pour sa part, elle indique n'avoir jamais reçu aucune information de la part du service partenariat d'INRAE.

Laurence Rageot revient sur l'échange avec Chloé Petit et indique qu'elle a accepté qu'en cas de demande d'accompagnement sur un PGD par un collègue INRAE, la question de l'existence d'un accord de consortium soit posée au service valorisation car cela permet aussi de répondre à certains champs du PGD. Mélanie Fauconnier ajoute qu'il n'est pas question que les services partenariats et les services « ouverture des données » soient en opposition, mais qu'il est important de se donner les points de raccordement nécessaires pour éviter des incohérences. L'idée est de voir comment on peut travailler ensemble. C'est ce qu'elle expérimente avec Clara Monmarché sur un des PGD. Il y a des rubriques du PGD sur lesquelles elle est incompétente mais pour lesquelles soit l'ADCVL, soit d'autres services, comme le service informatique, peuvent répondre. Sur la rubrique questions éthiques et juridiques, il faut que Cloé Freulon (DPO de l'université de Tours) d'un côté et le SPIV de l'autre soient contactés pour vérifier que si un accord de consortium est en cours de négociation en parallèle et que les réponses soient en cohérence.

Laurence Rageot revient sur la demande du collègue du CHU et indique qu'il y a une réflexion sur une possible demande de participation à l'ADCVL. La question a été abordée par Hélène Blasco, ancienne VP recherche du CHU, qui s'est renseignée sur l'atelier. Mélanie Fauconnier ajoute qu'un certain nombre d'enseignants-chercheurs sont aussi praticiens au CHU et qu'il peut y avoir confusion dans l'accompagnement des projets de recherche. Les projets de recherche clinique peuvent ressembler à des projets portés par une unité de recherche mais dans un service hospitalier. La question des données sensibles, des données de santé, et la façon dont le CHU doit être impliqué se pose.

Laurence Rageot conclut en insistant sur les liens à créer avec les collègues des services valorisation et les DPO pour que la réflexion ne soit pas faite deux fois, ou qu'on se retrouve en désaccord avec le contenu des accords de consortium.

### **III. Restitution de l'enquête sur les codes et logiciels**

Rémi Ossant rappelle le contexte et les objectifs de l'enquête qui concernait tous les membres de l'ADCVL et du PUI. Les objectifs de l'enquête étaient de

- mieux connaître les pratiques autour de la production, la diffusion, la valorisation des codes sources et logiciels,
- de faire remonter les besoins en formation, en accompagnement
- détecter des foyers de production de codes, ou des actifs potentiels qui seraient déjà préexistants au sein de la région.

L'enquête a été ouverte entre le 1er mars et le 18 avril. Elle a bénéficié d'un bon relai auprès des institutions par différents canaux : newsletters, listes de diffusion, affiches papier. Il indique s'être appuyé sur des personnes ressources comme le réseau de développeurs RDEV, ou des petites communautés axées sur différentes thématiques autour du logiciel.

L'enquête a reçu 147 réponses de l'ensemble des institutions partenaires du PUI. Le chiffre peut paraître un peu faible comparé aux 3500 chercheurs de la Région, plus les ingénieurs et les doctorants, mais c'est un sujet de niche comparé aux données. De plus, il est s'est avéré difficile de faire prendre conscience aux personnes qu'elles sont des producteurs de codes. Comme l'a évoqué plusieurs fois Laurent Beunèche personne ne considère qu'il est développeur, tout le monde pense faire un peu de code sans être développeur. Peu de personnes se sentent légitimes en tant que producteur de code et il a fallu insister sur le fait que, dès qu'on produit un tout peu de code, même très ponctuellement, on est pertinent pour répondre à l'enquête.

L'institution la plus représentée est l'université de Tours avec 55% des répondants qui l'ont comme tutelle de leur unité de recherche. Les réponses proviennent de 34 laboratoires. Les laboratoires les plus représentés sont le LIFAT, l'Institut Denis Poisson, le GREMAN, le LIFO et PRISM, des laboratoires d'informatique ou de mathématiques appliquées. Ensuite on trouve CITERES (SHS) et N2Cox (médecine) qui sortent du champ de l'informatique et de l'ingénierie.

62% des répondants sont des chercheurs, ce qui démontre que le code n'est pas un produit de développeurs ou d'ingénieurs au service des chercheurs, qu'il existe une communauté de chercheurs

qui produisent du code au sein de la Région. Le code est un vrai produit de recherche au sens strict, à côté des données et des publications. 10% des répondants sont des doctorants et 30% des ingénieurs. La question sur le domaine scientifique étant une question ouverte, Rémi Ossant a dû retravailler les données, il s'est appuyé sur les domaines scientifiques de l'OCDE. Le domaine Informatique et sciences de l'Information est majoritaire mais il regroupe, en plus de l'informatique, des champs très divers comme les humanités numériques, la biostatistique et tous les domaines à la frontière entre l'informatique et un autre scientifique. Une quinzaine de personnes ont répondu travailler dans l'informatique.

Néanmoins, il apparaît que le logiciel est produit dans tous les domaines scientifiques car sont représentés les domaines de la santé, les sciences humaines et sociales, les sciences physiques, l'ingénierie, l'économie, les langues, la psychologie, la chimie, etc.

Cette diversité avait déjà été mise en avant dans une enquête nationale faite par le ministère de la recherche sur la valorisation des logiciels de recherche. Les codes et logiciels sont un produit de recherche de tous les domaines scientifiques, bien que dans des proportions différentes.

Concernant la fréquence de production, nous avons beaucoup insisté pour que les producteurs occasionnels répondent, il était donc important de savoir si on avait réussi. Il apparaît que la majorité des répondants sont des producteurs de code assez actifs :

- 35% déclarent produire du code tous les jours,
- 33% au moins une fois par semaine.

Dans une autre question, il était demandé s'ils se considéraient comme des producteurs ou des utilisateurs de codes. 91% se définissent plus comme des producteurs. Cela permet de s'assurer que la population qui a répondu est pertinente pour cette enquête.

Un des enjeux de l'enquête était de connaître les habitudes de travail des producteurs de codes, les outils utilisés. Le choix avait été fait de questions assez larges pour ne pas trop orienter, les forges, les outils de versioning ou les entrepôts comme Software Heritage ou HAL n'étaient pas cités.

Concernant les outils collaboratifs pour la production, quasiment 60% des producteurs déclarent utiliser une forge logicielle. 20% d'entre eux utilisent une forge privée, majoritairement GitHub. 17% mentionnent l'utilisation d'une forge institutionnelle, principalement la forge MIA de l'INRAE ou la forge institutionnelle de l'université de Tours. La majorité utilisent plusieurs forges : GitHub et un GitLab, mais sans préciser lequel.

A noter que les forges sont utilisées depuis longtemps et ne sont pas propres à la science ouverte. Elles sont très utilisées pour du travail collaboratif, pour s'inscrire dans une démarche de qualité logicielle. Les forges institutionnelles sont assez récentes et pas systématiques. De plus, lors de l'utilisation d'une forge privée un miroir peut être dans une forge institutionnelle et inversement. Une communication sur les forges institutionnelles peut être utile.

Sur la conservation pérenne du code qui diffère de l'archivage au sens des données, 25 % des répondants ont déclaré avoir des habitudes de travail pour la conservation pérenne de leurs codes. Concernant les moyens d'y parvenir, seulement 5 ont nommé Software Heritage et HAL, les autres évoquant les forges logicielles alors qu'elles ne sont pas considérées comme un moyen de conservation pérenne. Il apparaît clairement qu'un travail de sensibilisation sur ces questions est à faire d'autant que Software Heritage archive tous les projets publics qui sont sur une forge.

Venaient ensuite des questions sur les aspects juridiques car ils influent sur les possibilités de valorisation qui peuvent être mis en œuvre que ce soit en science ouverte ou avec une valorisation économique.

A la question de leur connaissance du droit et de la politique de l'établissement autour de la production logicielle, seuls 5% répondent « oui » et 18% « plutôt oui ».

Sur la réutilisation des codes sources et logiciels et sur le fait de vérifier les droits qui y sont associés, les réponses sont plus équilibrées : quasiment 30% vérifient toujours, 22% souvent, 23% parfois et 27% jamais.

Concernant les licences, quasiment 50% disent ne jamais mettre de licence sur leur logiciel. Sans licence sur un logiciel, on considère qu'il est complètement fermé et que personne n'a le droit d'utiliser. Les aspects juridiques semblent assez mal connus des répondants mais il faut noter qu'il y a un manque de ressources humaines spécialisées sur le sujet au niveau local et que certains établissements n'ont pas de cadre bien défini. Par exemple les universités d'Orléans et de Tours n'ont pas de politique préconisant une licence logicielle pour les logiciels ouverts. Cela démontre la nécessité de proposer un service d'accompagnement sur ces questions car c'est un gros enjeu sur le devenir des codes produits mais cela nécessite des ressources.

La partie suivante concernait la valorisation des productions logicielles. trois « types » avaient été définis :

- la valorisation scientifique (codes publiés en lien avec une publication scientifique)
- la valorisation économique
- La valorisation science ouverte.

Il apparaît que les codes sont régulièrement associés aux publications : 45% des répondants déclarent publier leurs codes et les associer à des publications scientifiques. Cependant, 45% déclarent qu'ils ne publient jamais leurs codes.

Près de 20% valorisent systématiquement leur code source et logiciel selon les principes de la science ouverte et 25% régulièrement.

Seulement 2% recherchent systématiquement une valorisation économique et 1% de façon régulière. Sur le potentiel de valorisation des codes sources et logiciels produits au sein de leur équipe de recherche, 50% identifient une valorisation en science ouverte possible, 5% une valorisation économique et 14% les deux. A noter que 33% n'identifient aucun potentiel de valorisation.

Il apparaît que l'habitude de valoriser les codes est assez peu présente, même si la valorisation scientifique semble être la plus répandue. L'intérêt de proposer une sensibilisation et un accompagnement sur les aspects de valorisation semble clair.

La dernière partie de l'enquête concernait les besoins en formation et en accompagnement. Les thématiques de formation qui ont été plébiscitées sont la propriété intellectuelle, la conservation pérenne et les licences associées aux codes.

25 personnes ont laissé leurs coordonnées pour être recontactées pour un accompagnement personnalisé.

Rémi Ossant va désormais rendre publiques les données, en les anonymisant et produire une synthèse. Il faudra ensuite répondre aux sollicitations des chercheurs qui en ont exprimé le souhait et réfléchir à la façon de répondre aux besoins de formation.

Alain Ravaz note qu'environ un tiers des répondants souhaite être recontacté et demande si on a une idée de la répartition afin de cibler les endroits où faire de la sensibilisation. Rémi Ossant répond qu'il est possible de croiser l'information avec les laboratoires ou équipes d'appui dont ils sont issus. Pour ceux-là il serait pertinent d'interpréter lesquels semblent demandeurs de formations. Les données collectées permettent cette analyse.

Laurence Rageot ajoute que le besoin en formation est très important et qu'on ne peut se limiter à de la sensibilisation ciblée auprès de ceux qui ont demandé un accompagnement. Alain Ravaz répond qu'il pensait que les DSI pouvaient être intéressées par un moyen d'identifier la population de développeurs.

Mélanie Fauconnier revient sur l'identification des ressources pour faire de la formation à la propriété intellectuelle sur les logiciels. Elle indique que l'INPI (Institut national de la propriété intellectuelle) peut être sollicité. Le SPIV avait fait une demande de formation, cela avait été très court mais il est possible de travailler avec eux sur le format.

Laurence Rageot répond que la formation en cours de montage pour l'ADCVL n'est pas à destination des chercheurs mais des membres de l'atelier et qu'il faudra ensuite réfléchir à la formation à monter

pour les chercheurs, ou les producteurs du code. Des contacts ont été noués avec CNRS Valorisation et Lionel Morel, ancien directeur scientifique adjoint de CNRS-SHS sur la science ouverte et maintenant à la DDOR (Direction des Données Ouvertes de la Recherche). Nathalie Pothier ajoute qu'il a été recruté pour travailler sur les questions juridiques et sur les données.

#### IV. L'écosystème Recherche Data Gouv

Laurence Rageot revient sur les deux jours de séminaire Recherche Data Gouv des 16 au 18 juin à Lille, auquel elle a assisté avec Rémi Ossant et Clara Monmarché.

Les perspectives qui ont été présentées sont :

- l'inscription de l'écosystème Recherche Data Gouv dans la stratégie nationale des infrastructures de recherche
- le développement technique du portail comprenant l'entrepôt et le catalogue qui est désormais au stade de prototype et dans lequel est intégré beaucoup d'IA qui nécessite de très nombreux tests.
- des outils de curation des données pour que ce soit plus fluide et que ça aille plus vite
- un renforcement des coopérations notamment avec les centres de référence thématiques
- Un renforcement de la visibilité, l'optimisation des ressources et la réalisation de la science ouverte à tous.

L'organigramme stabilisé de Recherche Data Gouv a également été présenté. A noter qu'un conseil des membres est en cours de montage. Il regroupera les institutions porteuses de toutes les membres de l'écosystème. C'est l'université de Tours qui siègera au conseil des membres pour l'ADCVL.

Un consortium pour tous les membres avait été prévu, mais cela aurait impliqué qu'à chaque départ ou arrivée d'un membre l'accord doive être revu et signé par tous les membres. Avec 70 membres actuellement, cela n'aurait pas été gérable. Une convention cadre a donc été définie et un contrat sera proposé à chaque membre.

L'autre nouveauté sur l'organigramme est la création d'un club des utilisateurs pour la plateforme des données. Il se réunira une fois par an en AG pour présentation des évolutions de la plateforme et retour des utilisateurs.

Pour la première fois, la question du modèle économique a été très succinctement abordée. Il a été précisé que le modèle économique ne concernait ni le catalogue, ni tout ce qui était en place aujourd'hui. Seuls des nouveaux services personnalisés seraient concernés, question à suivre.

Enfin des lettres de mission spécifiques pour valoriser l'engagement des membres sont prévues mais aucun détail n'a été fourni. Concerneront-elles tous les membres des ateliers ?

Concernant les informations sur la science ouverte dans son ensemble, Isabelle Blanc a évoqué une stratégie nationale, qui serait en cours de finalisation. Elle a indiqué que l'instabilité ministérielle a ralenti le travail. Cette stratégie prendra la suite des plans nationaux pour la science ouverte et devrait permettre de montrer les bénéfices du partage des données, des codes et logiciels, notamment en termes de rayonnement, de reproductibilité de la recherche, de souveraineté scientifique, d'innovation et d'ouverture au-delà du monde académique.

Enfin, une demi-journée a été consacrée à l'accompagnement sur les codes et logiciels. Une présentation de l'Open Source Program Office (OSPO) en cours de mise en place à l'université de Grenoble a été faite. Ce modèle va être préconisé par le ministère de la recherche et une labellisation spécifique pour les OSPO est prévue. Isabelle Blanc a précisé que l'accompagnement sur les codes et logiciels ne dépendait pas des ateliers de la donnée mais qu'il serait confié aux OSPO.

Laurence Rageot indique qu'elle s'interroge sur ce que recouvrent les OSPO. Est-ce une personne ou un service ? Faudra-t-il un OSPO par institution ou pour une Région ? Rémi Ossant ira à l'inauguration de l'OSPO de Grenoble et pour tenter d'avoir plus de précisions. Il indique que c'est la volonté du ministère, mais que cela ne lui semble pas encore bien défini au niveau opérationnel.

Anne Azanza demande si le sujet des ateliers de la donnée et leur financement a été abordé car on arrive bientôt au terme des financements de la création de l'ADCVL. Laurence Rageot répond que, dès la labellisation, il était très clair que le ministère ne financerait pas les ateliers. Le FNSO finance la mise en œuvre des ateliers, mais ensuite c'est aux institutions membres de financer. C'est la raison pour laquelle, le financement des postes de Clara Monmarché et Rémi Ossant a été envisagé sur le PUI. Elle a abordé la question avec Daniel Alquier et Patrick Vourc'h (VP Recherche et chargé de mission science ouverte de l'université de Tours) quand elle leur a présenté l'ADCVL. Daniel Alquier lui a répondu que le ministère favorisait les PUI et qu'il pensait que c'était la piste de financement pour l'atelier. Elle ajoute que la question sera abordée lors du prochain comité de pilotage de l'ADCVL.

Mélanie Fauconnier rejoint l'inquiétude concernant le financement et le positionnement sur les PUI. Il y a des réflexions pour que le PUI soit le modèle qui perdure et que les sociétés d'accélération du transfert de technologies n'aient pas la pérennité prévue il y a 10 ans à leur création. Elle s'interroge sur les OSPO, ne vont-ils pas être rattachés aux filiales ? L'UGA est une des universités les plus dynamiques et pionnière sur les questions de valorisation socio-économique. Une filiale a été créée. Ils ont non seulement développé des choses, mais structurellement, ils se sont organisés avec des entités juridiques propres pour gérer les activités industrielles et commerciales. Elle a l'impression qu'il y a une forme de glissement. Cela rejoint les discussions des réunions précédentes sur la possibilité d'associer la science ouverte et la valorisation socio-économique.

Laurence Rageot précise que l'OSPO de Grenoble est positionné au méso-centre car la cellule data et l'atelier de la donnée sont des émanations du méso-centre, sans forcément une orientation socio-économique. Mélanie Fauconnier répond que c'est rassurant dans le sens où stratégiquement, il garde une forme d'autonomie sur la manière de développer les services aux chercheurs, etc. Laurence Rageot ajoute que la présentation faite par Violaine Louvet ne lui a pas laissé penser que c'était lié à l'aspect valorisation. Elle l'a présenté comme une partie de la cellule data, élément en contradiction avec le discours d'Isabelle Blanc qui sort l'accompagnement sur les codes et logiciels du périmètre des ateliers de la donnée. Il lui semble important que Rémi Ossant assiste à l'inauguration pour discuter et tenter de mieux comprendre le positionnement de l'OSPO.

Stéphanie Puel demande s'il y a des informations ou un positionnement du côté du collège codes sources logiciels du COSO. Rémi Ossant répond qu'il y avait une réunion du COSO le 24 juin à laquelle il n'a pas pu participer mais qu'il vérifiera dans le compte-rendu. Cela n'a jamais été évoqué dans les précédentes réunions.

Maryse Castan ajoute en remarque générale que cela reflète la volonté du CNRS de promouvoir l'utilisation de logiciels libres.

Laurence Rageot ajoute que les OSPO sont très présents dans les entreprises privées, et peu développés côté recherche académique. Rémi Ossant ajoute que dans ces entreprises, l'intérêt est de pouvoir mettre en avant la politique concernant le logiciel libre et de se coordonner au niveau international. A noter que les GAFA ont leur OSPO : Amazon, Intel...

Stéphanie Puel indique que cela rejoint en partie des éléments qui ont été abordés lors du GTSO (Groupe de travail science ouverte) Couperin. Deux réunions plénières sont organisées par an et un sous-groupe, piloté par Laëticia Bracco de l'université de Lorraine, travaille sur les données. Elle a évoqué le fait d'être à la croisée des chemins, de ne pas trop savoir où on va en politique sur la question des codes sources et logiciels, notamment sur le lien avec les ateliers de la donnée. Le GTSO de Couperin se posait la question de savoir s'il fallait créer un sous-groupe de travail sur les logiciels.

Laurence Rageot indique qu'Isabelle Blanc ayant la double casquette avec sa position d'Administratrice ministérielle des données, des algorithmes et des codes sources, elle doit savoir vers où elle veut aller, mais elle ne communique pas clairement pour le moment. Par contre, il était très clair dans son discours que les codes sources et logiciels n'entraient pas dans le champ d'action des ateliers de la donnée. Beaucoup d'ateliers avaient la crainte de devoir prendre en charge l'accompagnement sur les codes sources et logiciels en plus des données. Elle a probablement voulu rassurer ces collègues.

Georges Fins indique que le risque est que le modèle économique positionne le développement du logiciel comme base et que les données deviennent les éléments sur lesquels est fait le retour sur

investissement. Les entreprises privées sont sorties du modèle distinguant codes et données car cela modifie la qualité du code même en termes d'architecture.