

Réunion du comité opérationnel de l'Atelier de la donnée Centre-Val de Loire

20/06/2025

Ordre du jour

- 01 Informations générales concernant l'ADCVL**
- 02 Les groupes de travail de l'ADCVL**
- 03 Restitution de l'enquête sur les codes et logiciels**
- 04 L'écosystème Recherche Data Gouv**
- 05 Questions diverses**

Les informations générales de l'ADCVL

01

Demandes



4 demandes d'accompagnement à la rédaction de plans de gestion de données depuis mars 2025 :

- Elisabeth Herniou (IRBI), projet FLYPATH (biologie des insectes)
- Fabriche Foucher (PRISME), chaire industrielle DELHYCE (ingénierie des produits et des procédés)
- Julie Loiseau (VALLOREM), projet ValPostConso (sciences sociales)
- Caroline Giraudeau (PAVEA), projet ACAPULCO (psychologie)

Indicateurs de l'atelier



Quels indicateurs mettre en place pour l'ADCVL ?

Quels indicateurs pour nos besoins ?

Quels indicateurs pour nos tutelles ?

Comment caractériser l'apport de l'ADCVL ?

- Nombre de sollicitations sur le guichet unique

- Demandes directes

- Nombre de formations / sensibilisations dispensées

- Nombre de personnes formées

- Nombre de présentations de l'ADCVL

- ...

Les ressources humaines ?

- Nombre de personnes impliquées dans l'ADCVL

- Nombre d'ETP?

Comment mesurer la satisfaction des utilisateurs?

Les indicateurs – côté Recherche Data Gouv

Réflexion en cours au sein de Recherche Data Gouv :

- Pour la demande d'inscription sur la feuille de route des infrastructures de recherche 2026-2030
 - ⇒ Utilisation des dossiers de demande de labellisation
- Pour présenter l'activité de Recherche Data Gouv, tous membres de l'écosystème mélangés
 - ⇒ Réflexion sur les services avec un travail d'harmonisation sur le type de service et sur le type d'accompagnement
 - ⇒ Besoin de chiffres globaux et significatifs
 - Le nombre de personnes touchées par les Ateliers (sollicitations + formations + présentations) / nombre de demandes arrivées sur les guichets uniques

Formations



Formations réalisées et à venir

- Formation doctorale « **Comment partager vos données et vos codes et logiciels** » à Tours le 15 mai (3 participants)
 - Annulation de la session d'Orléans car moins de 5 inscrits
- Formation « **GIT de survie** » à Tours le 12 juin et à Orléans le 15 septembre (toutes les deux complètes, intérêt fort)
- Formation « **FAIRisation des données** » en cours de montage, prévue pour l'automne 2025 (2 sessions de 2 heures en visio). Date à déterminer (Evento pour déterminer la date, entre fin octobre et début novembre)
- **Formation pour l'ADCVL : Science ouverte, codes et logiciels** en cours de montage, prévue pour l'automne

Les groupes de travail de l'ADCVL

02

GT FAIRisation des données

Formation en cours de montage :

- Session de 2h sur le F et le A
- Session de 2h sur le I et le R

Pour chaque session :

- définitions des notions connexes aux principes FAIR (données de recherche, plans de gestion de données, vocabulaire contrôlé, données personnelles, entrepôts de données...)
- Bons et mauvais exemples de données FAIR/non FAIR
- Cas pratiques/interactivité avec les participants

GT Plans de gestion de données

Deux supports en cours de préparation :

- Réalisation d'une fiche de sensibilisation dédiée aux PGD par le Service Partenariats Innovations
Valorisation de l'Université de Tours :
 - L'ADCVL a été sollicité pour la relire
 - Renvoi vers l'ADCVL pour les questions sur les PGD
- Fiche pratique sur les délais de transmission des PGD aux financeurs (ANR et Commission Européenne) en cours de rédaction

PLAN DE GESTION DE DONNÉES (PGD/DMP)

Les données c'est comme des pépites que vous trouveriez pendant vos recherches.
Pour savoir quoi en faire il faut les caractériser.
Les pépites d'**or** ça va dans le 🏠
et les pépites de **chocolat** sur les 🍫 que l'on partage avec les autres pour nourrir la communauté scientifique.

Vous l'aurez compris pour savoir comment gérer ses données il vous faut un plan
= le **Plan de Gestion de Données** (ou PGD pour les connaisseurs).

Il vous aide à identifier les données que vous allez **produire**, comment les **organiser**, les **stocker**, les **traiter** et déterminer celles qui pourront être **partagées**

Heureusement, en Région,
**l'Atelier de la donnée
Centre Val de Loire**
est là pour
vous accompagner
vous guider
vous former



le plan de gestion de données s'inscrit dans une démarche de **science ouverte** et est un bon outil pour la **gestion de projet**



Outil permettant de **définir ses données** et d'accéder à l'ensemble des **informations sur la gestion de ses données** (OPiDOR, modèles de PGD, fiches synthétiques..)

- **Rendez-vous initial** pour faire le point sur votre projet et la gestion de vos données
- **Relecture** du plan de gestion des données

CONTACT ✉



Guichet unique de l'ADCVL →

Avec les bons outils, chaque pépite est à sa place : protégée ou partagée
et là, le PGD devient votre allié

Plan de gestion de données : quand et comment le transmettre ?

Cas de l'ANR et de la Commission Européenne

De quoi parle-t-on ?

Le **plan de gestion de données**, appelé PGD ou DMP (Data Management Plan), est un document permettant d'**expliquer comment sont créées, collectées, gérées, stockées et partagées les données dans le cadre d'un projet de recherche** ([En savoir plus](#)). Bien qu'inscrit dans un objectif de science ouverte, l'ensemble des données utilisées dans le cadre du projet de recherche doivent être décrites, **qu'elles soient ouvertes ou non**. Ce document fait partie des livrables demandés par les financeurs publics.

Les exigences de l'ANR

Quand ?

Plusieurs versions du PGD sont à rendre à différentes échéances au cours du projet :

- **Version initiale** dans les 6 premiers mois (à compter de la date de démarrage scientifique du projet)
- **Version intermédiaire** mise à jour lors de l'évaluation à mi-parcours (pour les projets de plus de 30 mois)
- **Version finale** complète à la fin du projet

Les dates de ces échéances sont consultables :

- dans la convention de financement,
- dans l'acte attributif de financement,
- dans le système d'information de l'ANR ([SIM](#)).

Comment ?

Via le système d'information de l'ANR ([SIM](#)).

Quel modèle ?

- **Modèle « Science Europe : modèle structuré »** est celui recommandé par l'ANR

↪ **Outil de rédaction de plan de gestion de données en ligne :** 

Les exigences du programme cadre Horizon Europe

Quand ?

Comment ?

Quel modèle ?

Attention

Si des **accords de consortium** ont été établis dans le cadre du projet, et qu'ils définissent des **conditions particulières en matière de gestion des données**, celles-ci doivent être abordées dans le plan de gestion de données (Ex. propriété des données, règles de confidentialité, propriété intellectuelle...).

GT Plans de gestion de données

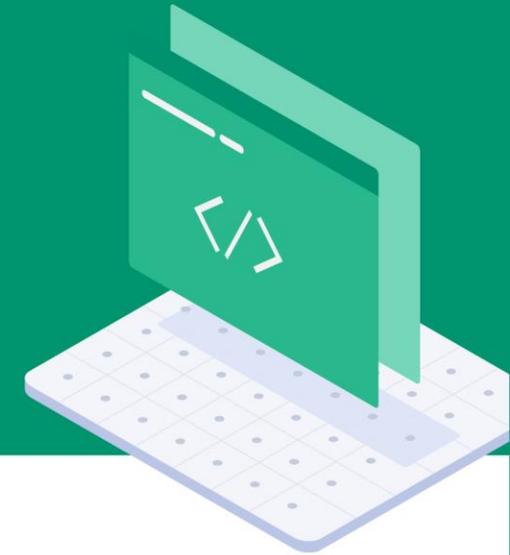
Quelle articulation entre l'ADCVL et les services valorisation dans les institutions ?

- Questions communes pour rédiger un PGD et un accord de consortium
- Cas d'un PGD en contradiction avec l'accord de consortium (Chloé Petit - INRAE)
- Cas des projets en santé avec des données sensibles (articulation avec les services des CHU ?)
- GT PGD dans le cadre du PUI Loire Valley Innov

Restitution de l'enquête sur les codes et logiciels

03

Enquête sur les pratiques et les besoins de la communauté scientifique autour des **codes sources et logiciels**



- Enquête réalisée dans le cadre du **PUI Loire Innov' Valley** s'adressant à **l'ensemble des chercheurs, des doctorants et des personnels d'appui à la recherche** des institutions partenaires : l'Université de Tours, Université d'Orléans, CNRS, CHRU de Tours, CHU d'Orléans, INRAE, BRGM et INSA
- Objectif : mieux connaître les **pratiques** et les **besoins** des acteurs de la recherche autour de la **production**, de la **diffusion** et de la **valorisation des codes source et logiciels, faire remonter les besoins et détecter des foyers de production de code** dans la région Centre-Val de Loire
- Réalisée en ligne sur le site l'ADCVL entre le 1^{er} mars et 18 avril
- Enquête a bénéficié d'un **bon relais dans les institutions** (Newsletter, liste de diffusion, affiches papiers, ...)

Taux de réponses

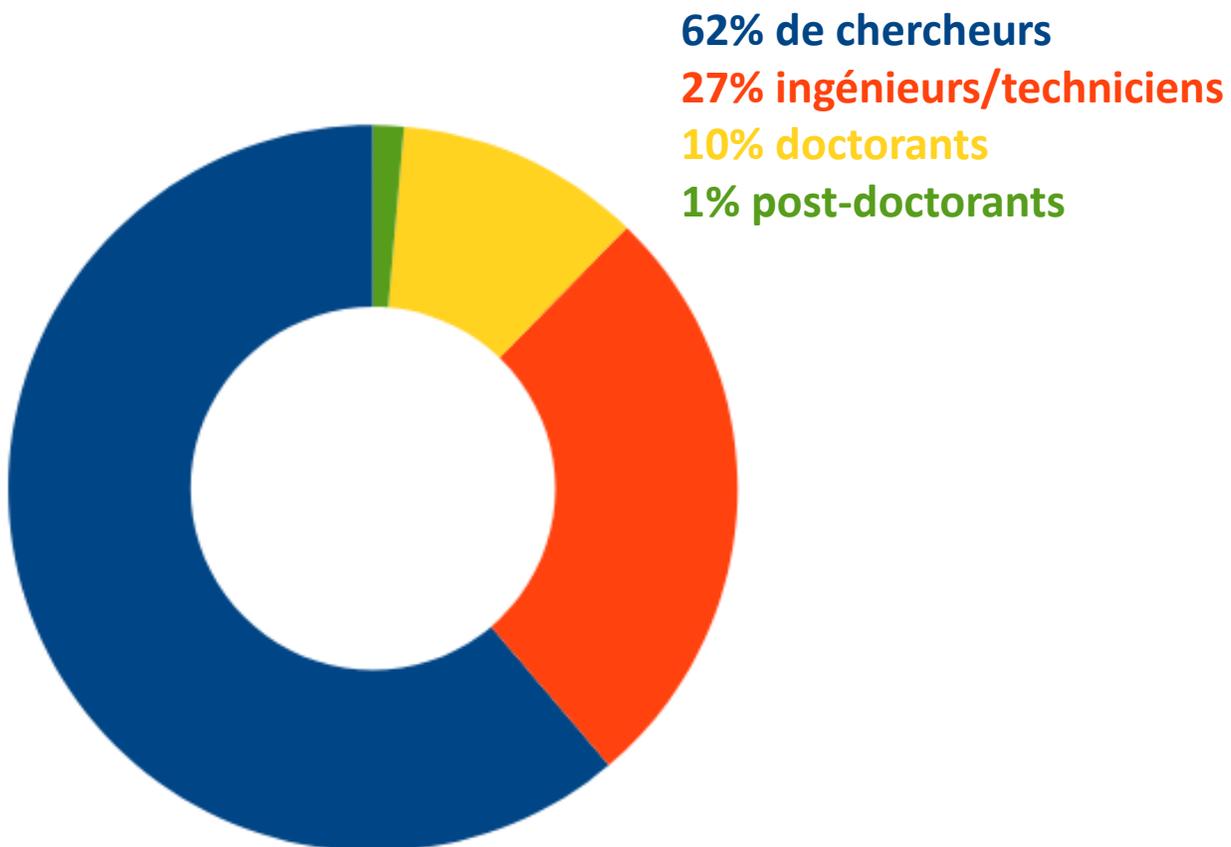
- **147 réponses** en provenance de l'ensemble des institutions partenaires du PUI
 - Nombre pouvant paraître faible face aux 3500 chercheurs identifiés de la région mais sujet plus « niche » que les données
 - Difficulté à faire prendre conscience aux répondants qu'ils sont des « **producteurs de codes** »

-  la mieux représentée avec 55% des répondants l'ayant pour tutelle

- Réponses en provenance de **34 laboratoires différents**, les mieux représentés sont :

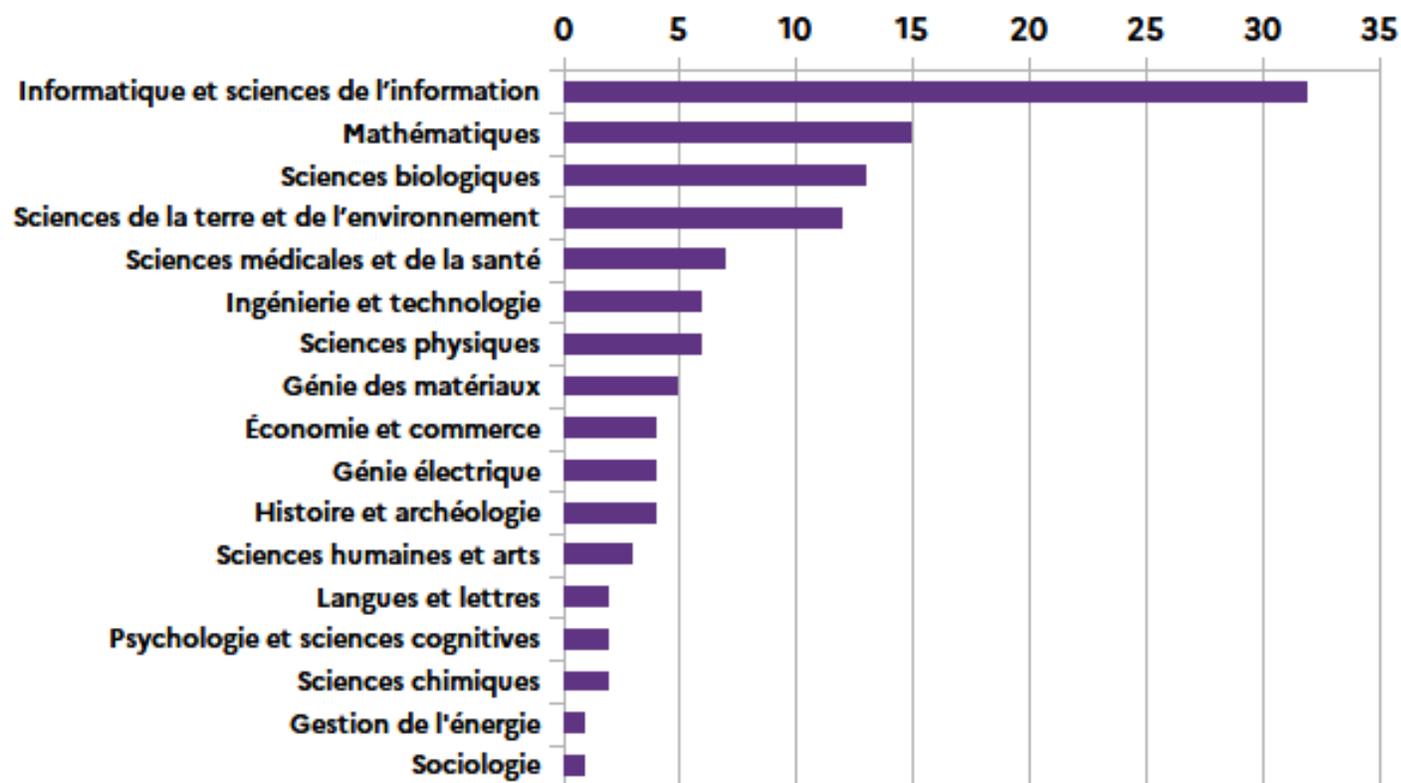


Profil des répondant(e)s : statut



- Majorité de chercheurs dans les répondants
- Le code apparaît bien comme étant un « produit de recherche » au sens le plus strict
- Pas seulement une production de « développeurs » professionnels

Profil des répondant(e)s : domaine scientifique



- Même constat que l'enquête nationale de 2023 sur la valorisation des logiciels de recherche : le code source est **produit dans tous les domaines scientifiques ...**
- ... mais en **proportion différente en fonction des domaines**
- « L'informatique et les sciences de l'information » sont **très bien représentées mais très diverses** :
 - Biostatistiques
 - Humanités numériques
 - Modélisation
 - *etc.*

Profil des répondant(e)s : fréquence de production



35% Tous les jours

33% Au moins une fois par semaine

18% Au moins une fois par mois

9% Au moins une fois par an

7% Jamais

91% des répondants se définissent comme « **producteurs de codes** » et 9% seulement comme des « utilisateurs de codes »

- Profils de **producteurs de codes actifs majoritaires**
- Production de code comme « activité secondaire » plutôt à la marge
- **Objectif de l'enquête sur le profil des répondants atteint**

Bonnes pratiques de la science ouverte

- Un des objectifs de l'enquête : **connaître les habitudes de travail des producteurs de codes** et les **outils qu'ils utilisent**
- Questions assez larges sur les **usages**, pour avoir un panorama objectif, et qui incluaient systématiquement la possibilité de **nommer les outils** :
 - *Avez-vous l'habitude d'utiliser un **outil de travail collaboratif pour la production de code** ?*
 - Une forge ou un gestionnaire de version est-il utilisé ?
 - *Avez-vous des habitudes de travail pour la **conservation du code source au sein de votre équipe de recherche** ?*
 - Le code est-il conservé seulement en local ? Sur un serveur dédié ? Sur une forge locale / institutionnelle / communautaire ?
 - *Avez-vous des habitudes pour la **conservation pérenne du code produit** ?*
 - Software Heritage et HAL sont-ils connus et utilisés ? La forge logicielle est-elle considérée comme une solution de conservation pérenne ?

L'utilisation des forges

- **58 %** des producteurs de code **utilisent une forge logicielle**
- **20%** d'entre eux utilisent **exclusivement une forge privée**, principalement Github
- **17% mentionnent l'utilisation d'une forge institutionnelle**, la forge Mia (INRAE) et la forge de l'UT notamment
- Le **format hybride** avec l'utilisation d'une forge privée et d'autres solutions est majoritaire

- **Forges historiquement utilisées** pour d'autres raisons qu'une volonté de Science ouverte
- Apparition récente des forges institutionnelles et communautaires et pas systématique (UO)
- **L'utilisation d'une forge privée n'empêche pas de mettre en place des miroirs sur une forge institutionnelle**
- Nécessité de **faire connaître les forges institutionnelles et leur intérêt**

La conservation « pérenne » du code

- **24%** des répondants déclarent avoir des habitudes de travail pour conserver de manière pérenne leur code mais seulement **cinq d'entre eux nomment Software Heritage et HAL**
- La forge logicielle est nommée dans la plupart des autres cas ...



HAL
science ouverte



Software Heritage

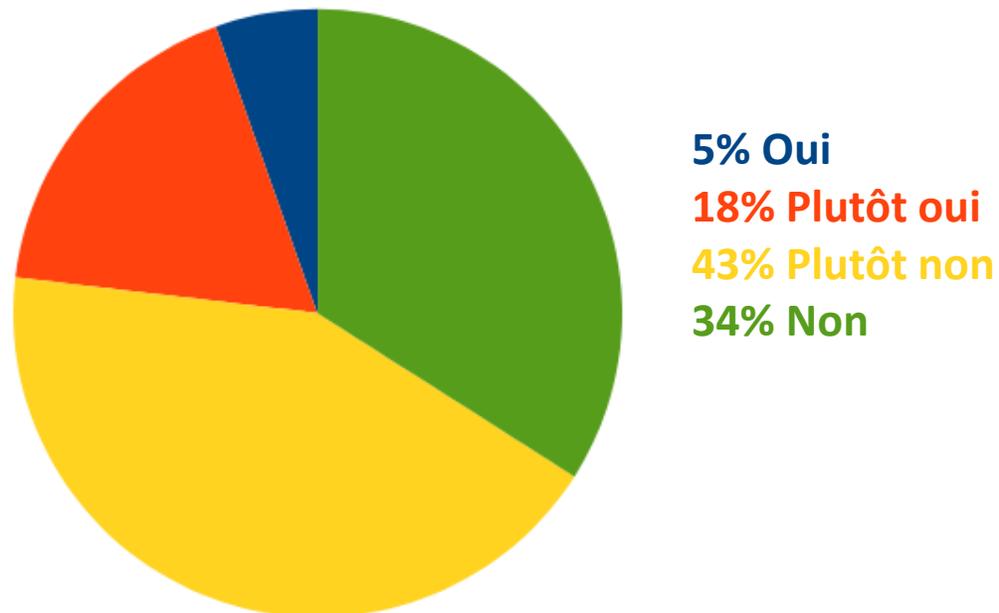
- **Software Heritage archive tous les projets publics** donc beaucoup de producteurs bénéficient de cette solution de conservation pérenne du code **sans le savoir**
- Nécessité de **sensibiliser à l'existence de cet outil** et son utilité pour citer les productions logicielles

Les aspects juridiques liés au code

- Volonté d'évaluer le **niveau de connaissance des répondants sur les problématiques en lien avec la propriété intellectuelle des logiciels** ainsi que les habitudes de travail autour de cette question

14. Avez-vous connaissance du droit et de la politique de votre établissement concernant la production logicielle ?

Oui Plutôt oui Plutôt non Non



Les droits liés aux productions logicielles utilisées et produits

- **Enjeu important sur les droits liés aux logiciels / briques logicielles utilisés** lors du développement d'un nouveau logiciel car cela peut **empêcher une ouverture des codes** ou une **valorisation économique** de celui-ci

15. Lors de la création ou de la réutilisation de codes sources et logiciels, avez-vous pour habitude de vérifier les droits associés à ces productions ?

Toujours Souvent Parfois Jamais

28% Toujours

22% Souvent

23% Parfois

27 % Jamais

- **La Licence** est la manière la plus préconisée de fixer les droits associés à une production logicielle dans le domaine de la Science Ouverte. Il était donc pertinent d'évaluer son utilisation par les répondants.

13. Avez-vous pour habitude que soient attachées des licences à vos productions de codes sources et logiciels ?

Toujours Souvent Parfois Jamais

16% Toujours

18% Souvent

18% Parfois

48% Jamais

Les aspects juridiques liés au code

- **Les aspects juridiques en lien avec le code ont l'air mal connu des répondants**
- **Manque de ressources humaines** spécialisées sur le sujet au niveau local
- **Manque de cadre dans un certain nombre d'établissements** : par exemple au sein de l'Université d'Orléans ou de Tours, il n'y a pas de licence logicielle préconisée par l'institution
- **Nécessité de proposer un accompagnement** sur cet aspect pour permettre une valorisation des codes et ne pas se retrouver avec des codes « inutilisables » en fin de projet

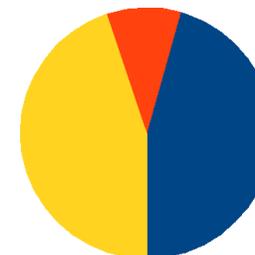
La valorisation de la production logicielle

- Volonté d'évaluer les **habitudes de valorisation des codes produits** par les répondants, trois « types » de valorisations ont été évalués :
 - **La valorisation « scientifique »** : le code est-il publié en lien avec une publication scientifique ?
 - **La valorisation « économique »** : ce type de valorisation est-il envisagé / exploré par les répondants ?
 - **La valorisation « science ouverte »** : ce type de valorisation est-il envisagé / exploré par les répondants ?

Des valorisations pratiquées et envisagées en proportions diverses

16. Avez-vous pour habitude de publier vos codes sources et logiciels et de les associer à des publications scientifiques et/ou des données de recherche ?

Publication et association Publication de codes seuls Pas de publication



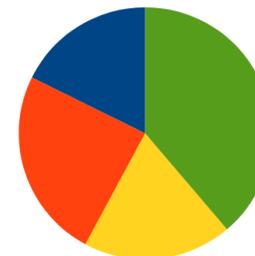
45% publication et association

10% Code publié seul

45% Pas de publication

17. Avez-vous pour habitude de valoriser vos codes sources et logiciels en les ouvrant en accord avec les principes de la Science Ouverte?

Systématiquement Régulièrement Ponctuellement Jamais



18% Systématiquement

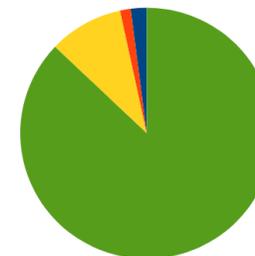
24% Régulièrement

19% Ponctuellement

39% Jamais

18. Avez-vous pour habitude de rechercher une valorisation économique de vos productions logicielles ?

Systématiquement Régulièrement Ponctuellement Jamais



2% Systématiquement

1% Régulièrement

10% Ponctuellement

87% Jamais

Potentiel de valorisation Science Ouverte et économique de la Région

21. Parmi vos productions de codes sources et logiciels au sein de votre équipe de recherche, en identifiez-vous certaines qui pourraient faire l'objet d'une valorisation économique et/ou d'une ouverture via la Science Ouverte ?

- Valorisation économique Valorisation type « Science Ouverte »
 Les deux Non

49% Valorisation « Science Ouverte »

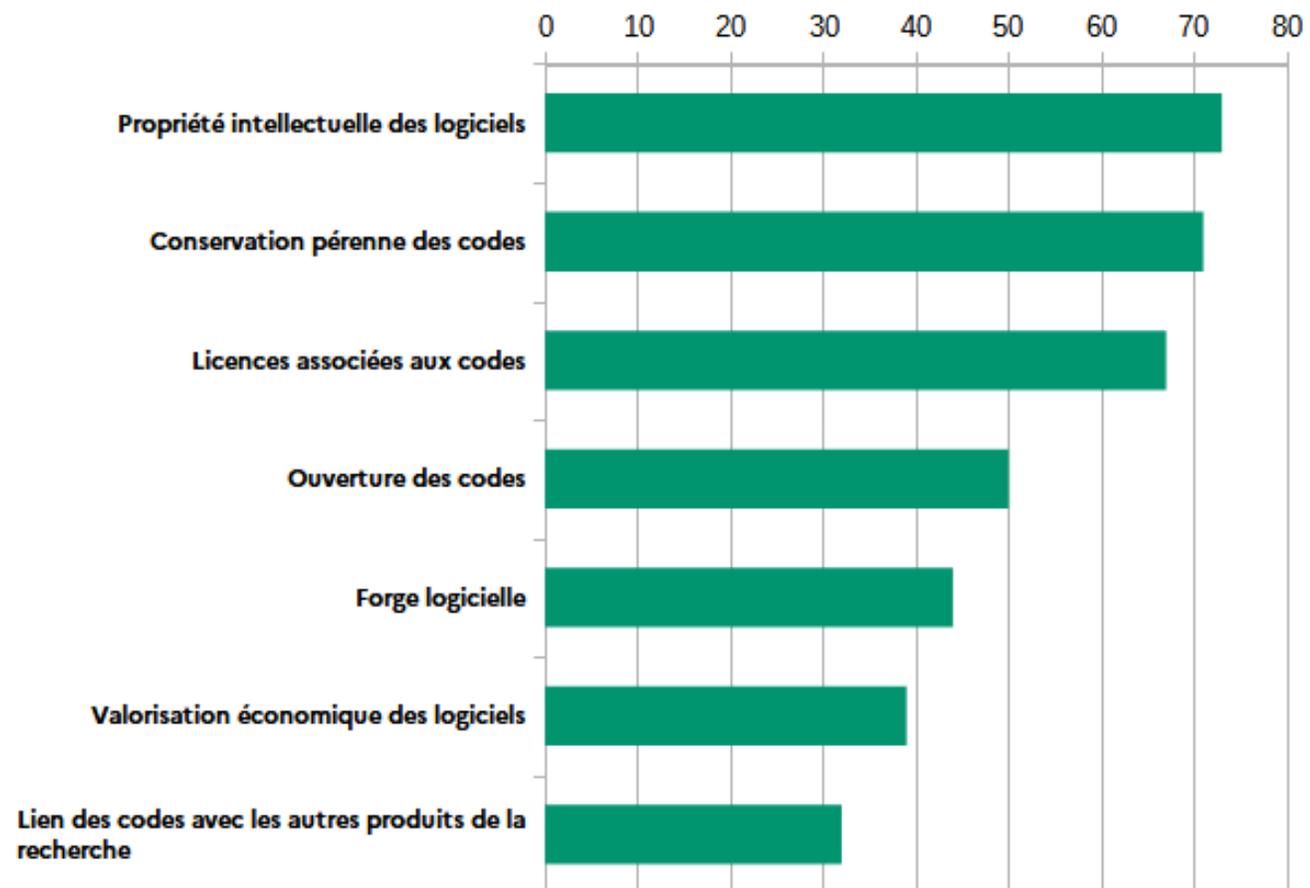
5% Valorisation économique

14% Les deux

33% Non

- **L'habitude de valoriser les codes est encore peu présente**, la valorisation scientifique semble la plus répandue devant la valorisation « Science Ouverte » et la valorisation « économique »
- Nécessité de **mieux accompagner les projets logiciels futurs pour permettre une valorisation**, de faire prendre l'habitude aux chercheurs de penser cet aspect et **d'investiguer sur le patrimoine logiciel déjà créé**
- Ces **différentes valorisations ne sont pas forcément contradictoires**

Besoins en formation et sollicitations



- **25 sollicitations de répondant(e)s** pour être recontactés concernant leurs besoins

- **Besoin en formation très important** qui concerne la plupart des domaines mais particulièrement les questions juridiques
- **Besoin d'accompagnement personnalisé notable** qui demanderait un service dédié

Conclusion

- Enquête qui malgré un nombre de réponses modeste, a su **remplir ses objectifs** :
 - Avoir une **meilleure vision du profil des producteurs de codes**
 - **Connaître les habitudes** de travail de la communauté (outils, pratiques, stratégies de valorisation...)
 - **Découvrir des foyers** de productions de codes
 - Faire remonter les **besoins des producteurs de codes** et **ouvrir un dialogue** avec eux

Perspectives :

- **Rendre public les données** (anonymisées) et terminer de produire un **compte-rendu écrit sur les résultats de l'enquête**
- **Répondre aux sollicitations des chercheurs**
- Réfléchir au niveau global sur la **manière de répondre aux besoins de formations** des producteurs de codes

L'écosystème Recherche Data Gouv

04

Écosystème Recherche Data Gouv

Journées Recherche Data Gouv – 16 au 18 juin 2025 à Lille

- Possibilité de participer en visio pour les plénières
En avez-vous profité ?

Perspectives

Inscription dans la stratégie nationale des infrastructures

Candidature feuille de route nationale des infrastructures

Développement technique

Portail avancé, outils de curation, intégration IA

Coopération renforcée

Centres de référence thématiques, infrastructures spécifiques

Impact attendu

Visibilité accrue, optimisation des ressources, science ouverte à tous



Représentation
des besoins ;
coordination
des partenaires

Validation
feuille de
route et
grandes
orientations

Qualité et
pertinence
des
orientations
scientifiques

Conseil des membres

Comité de pilotage

Conseil scientifique
international

Stratégie
globale et
pilotage des
actions

Direction Recherche Data Gov

Isabelle Blanc – Directrice

Véronique Stoll
adpte stratégie & partenariat

Gilles Mathieu
adjt opérations & coordination

Richard Moreno
Responsable
plateforme

Collège de direction

Nadège Joly
Responsable
communication

Club
utilisateurs

Plateforme
des données

Dispositifs d'accompagnement en
proximité géographique et thématique

Un écosystème pérennisé

Organisation formalisée

Un document cadre formalisant l'organisation.

Gouvernance stabilisée

Une gouvernance stabilisée pour l'écosystème.

Conventions pour les partenaires

Conventions pour les partenaires impliqués dans l'écosystème.

Modèle économique

Un modèle économique à l'étude pour déployer de nouveaux services.

Lettres de mission

Lettres de mission spécifiques pour valoriser l'engagement des membres.



Informations science ouverte au-delà de l'écosystème Recherche Data Gouv

Stratégie nationale de science ouverte

- Prendra la suite des Plans nationaux
- En cours de finalisation
- Permettra de montrer les bénéfices du partage des données, codes et logiciels : rayonnement, reproductibilité, souveraineté scientifique, innovation, ouverture au-delà du monde académique

Open Source Programme Office OSPO

- Accompagnement sur les codes et logiciels
- Ne relèvera pas des Ateliers de la donnée mais une labellisation par le ministère de la Recherche
 - 1^{er} à Grenoble : Inauguration de l'Open Source Programme Office de l'UGA les 23 et 24 septembre 2025

Avez-vous des **questions ?**

Merci

 <https://adcvl.pu-cvl.fr/>

 @Atelier de la donnée Centre-Val de Loire