



HAL
open science

Étude relative à l'ouverture des codes sources au sein de l'Enseignement Supérieur et de la Recherche (ESR) : Considérations en termes d'usage et de valeur.

Célyla Gruson-Daniel, Benjamin Jean

► To cite this version:

Célyla Gruson-Daniel, Benjamin Jean. Étude relative à l'ouverture des codes sources au sein de l'Enseignement Supérieur et de la Recherche (ESR) : Considérations en termes d'usage et de valeur.. [Rapport de recherche] INNO3; Etalab; Comité pour la Science Ouverte. 2021. hal-03125456

HAL Id: hal-03125456

<https://hal.science/hal-03125456v1>

Submitted on 29 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etude relative à l'ouverture des codes sources au sein de l'Enseignement Supérieur et de la Recherche (ESR)

Considérations en termes d'usage et de valeur

etalab gouv.fr

CUBE
inno

Sommaire

A Introduction et méthodologie.....	4
A.1 CONTEXTE ET OBJECTIFS.....	4
A.2 MÉTHODOLOGIE.....	4
B La publication de codes sources : une pratique partagée entre les différents établissements de l'ESR.....	6
B.1 UNE PRISE EN COMPTE TRANSVERSE AU SEIN DE L'ESR.....	6
B.2 LA DISTINCTION ENTRE LES ACTIVITÉS DE RECHERCHE ET DE DÉVELOPPEMENT : FONCTION, DOMAINE ET RESPONSABILITÉ.....	7
B.3 MODE DE PRODUCTION SCIENTIFIQUE ET OPEN SOURCE : DES MOTIVATIONS COMMUNES ET DES CULTURES INTIMEMENT LIÉES.....	9
B.4 UNE MULTITUDE DE PROJETS DE L'ESR EN OPEN SOURCE ET UN RECENSEMENT COMPLEXE À METTRE EN ŒUVRE.....	11
C L'ouverture des codes sources : pour un enseignement supérieur et une recherche publique fiable.....	14
C.1 LE CADRE LÉGAL APPLICABLE AUX ÉTABLISSEMENTS DE L'ESR.....	14
C.2 SCIENCE OUVERTE, REPRODUCTIBILITÉ ET INTÉGRITÉ SCIENTIFIQUE.....	15
C.3 OPEN SOURCE ET ENSEIGNEMENT SUPÉRIEUR.....	18
C.4 LES INFRASTRUCTURES SUPPORT DE L'ESR.....	20
D Valorisation et pratiques open source : de l'injonction contradictoire aux articulations fécondes.....	22
D.1 UN ÉCART MANIFESTE ENTRE PRATIQUES OPEN SOURCE ET MODALITÉS DE VALORISATION.....	22
D.1.1 APPRÉHENSIONS ET CRITIQUES ENVERS LES SERVICES DE VALORISATION ET MATURATION	22
D.1.2 DES POLITIQUES INSTITUTIONNELLES OPEN SOURCE ENCORE BIEN DISCRÈTES.....	25
D.2 CONJUGUER OPEN SOURCE ET VALORISATION : ÉLÉMENTS COMMUNS ET SPÉCIFIQUES AU SEIN DE L'ESR.....	27
D.2.1 PENSER LA VALORISATION DANS UNE DÉFINITION PLUS LARGE ET PLURIELLE.....	27
D.2.2 DES ÉLÉMENTS DE VALEURS COMMUNS AUX ACTEURS DE L'ESR ADAPTÉS À DES MODÈLES OUVERTS.....	28
D.2.3 LES ÉLÉMENTS DE VALORISATION SPÉCIFIQUES AUX DIFFÉRENTS ACTEURS DE L'ESR ET DE LEURS MISSIONS.....	30
E Stratégie nationale : soutien à la publication et au développement de projets logiciels au sein de l'ESR.....	32
E.1 INSCRIRE LE DÉVELOPPEMENT ET LE MAINTIEN DE PROJETS LOGICIELS ET OPEN SOURCE AU SEIN DES POLITIQUES INSTITUTIONNELLES ET DE L'ÉVALUATION DES CHERCHEURS.....	32
E.1.1 LEVER LES FREINS À LA PUBLICATION DES CODES SOURCES : QUALITÉ DU CODE, COMPÉTITIVITÉ, MANQUE DE VALORISATION.....	32
E.1.2 UN RENFORCEMENT ET UNE COORDINATION DES INITIATIVES DE POLITIQUES PUBLIQUES DÉJÀ EN COURS.....	33
E.1.3 UN RÉFÉRENCIEMENT DES INITIATIVES EXISTANTES ET LA CONSTRUCTION RAISONNÉE D'INDICATEURS.....	33
E.2 LE BESOIN D'UNE D'UNE DYNAMIQUE DE MUTUALISATION ET DE CONTRIBUTION SOUTENUE AU SEIN DE L'ESR.....	34
E.3 SERVICES EXTERNES : ÉCONOMIE D'ÉCHELLE ET RÉDUCTION DE COÛT.....	35
E.4 BESOIN HUMAIN : UNE RÉÉVALUATION DES PLANS DE CARRIÈRE ET UNE PÉRENNISATION DES POSTES AU SEIN DE L'ESR DÉDIÉS AU DÉVELOPPEMENT ET À LA MAINTENANCE.....	36
E.5 BESOIN D'INFRASTRUCTURES ET DE FORMATION POUR DE BONNES PRATIQUES D'OUVERTURE DU CODE SOURCE.....	39
E.5.1 INFRASTRUCTURE NÉCESSAIRE À LA PUBLICATION DE CODE DE L'ESR.....	39

E.5.2 FACILITER LE DÉVELOPPEMENT DE LOGICIELS AU CODE SOURCE PÉRENNE ET DE QUALITÉ : DOCUMENTATION, ARCHIVAGE ET RÉFÉRENCIEMENT.....	42
F Pour aller plus loin.....	45
G Synthèse des recommandations.....	46
H Annexes.....	49
H.1 PRÉSENTATION DU QUESTIONNAIRE ET ENTRETIENS RÉALISÉS.....	49
H.2 DÉMARCHE DE SCIENCE OUVERTE ET LICENCES.....	49
H.3 LISTES DE LOGICIELS OPEN SOURCE ET ESR.....	49
H.3.1 SÉLECTION DE PROJETS RECENSÉS (SECTION O) ET NON-RÉPERTORIÉS SUR CODE.ETALAB.GOUV.FR.....	49
H.3.2 SÉLECTION D'AUTRES PROJETS MENTIONNÉS (SECTION B) ET NON-RÉPERTORIÉS SUR CODE.ETALAB.GOUV.FR.....	52
H.3.3 LOGICIELS OPEN SOURCE MAJEURS.....	53
H.4 COLLECTE D'INFORMATION.....	55
H.5 BIBLIOGRAPHIE.....	55
H.6 SUIVI DE VERSION.....	57

A | Introduction et méthodologie

A.1 Contexte et objectifs

Commandée par Etalab, l'étude sur l'« ouverture des codes sources au sein de l'Enseignement supérieur et la Recherche » (ESR) vise à :

1. dresser un premier panorama des pratiques de publication de codes sources¹ au sein de l'ESR et des enjeux, freins et besoins associés ;
2. recenser des projets diffusés en open source² ou qui auraient vocation à l'être et de les caractériser ;
3. décrire les positionnements et politiques des établissements en matière d'open source en articulation avec les stratégies de valorisation et d'autres perspectives (souveraineté, open data, open access, etc.).

Elle permettra à court terme d'alimenter la contribution d'Etalab auprès de la mission Bothorel sur « la politique de la donnée et des codes sources » et d'envisager quelques actions susceptibles d'être structurantes pour les années à venir.

A.2 Méthodologie

Réalisée courant septembre 2020, cette étude s'appuie sur un questionnaire³ intitulé « pratiques de publication de codes sources au sein de l'Enseignement supérieur et de la Recherche (ESR) » (panorama des pratiques), complété par une vingtaine d'entretiens d'acteurs de la recherche et de l'enseignement supérieur et plusieurs autres études menées récemment sur des sujets connexes⁴. Elle s'appuie et intègre également les recommandations déjà existantes issues notamment du Collège « Logiciels libres et open source » du CoSO (Comité pour la Science Ouverte) et du projet PLUME⁵.

1 Nous employons dans ce rapport le terme de code source de manière générique. L'étude s'intéresse aux pratiques de développement de logiciels tout autant qu'aux autres briques essentielles au déploiement d'infrastructures et de services associées à cette production (documentation, etc.).

2 Nous employons le terme open source pour désigner à la fois les logiciels libres et open source regroupés communément sous le vocable de FOSS (*Free Open Source Software*).

3 Accessible sur : <https://questionnaire.inno3.cricket/index.php?r=survey/index&sid=592911>

4 Voir Annexe [H.4 Documents de référence](#)

5 Le projet PLUME ([Promouvoir les Logiciels Utiles Maîtrisés et Économiques dans l'Enseignement Supérieur et la Recherche](#)) a été mené de 2006 à 2013. Plusieurs des travaux de ce projet sont mentionnés en note de bas de page et en fin de document.

Points clefs questionnaire :

- Le questionnaire a été largement communiqué et a fait l'objet d'une diffusion ciblée auprès de différents établissements de recherche et de l'enseignement supérieur.
- Il a permis de recueillir **223 réponses** en une vingtaine de jours.
- Une partie optionnelle en fin de questionnaire a aidé à recenser **24 projets** dont les codes sources sont publiés (ou qui auraient vocation à l'être) et de les caractériser.
- En parallèle de sa diffusion, une vingtaine d'entretiens a été réalisée auprès de personnes contribuant à la production de codes sources pour la recherche et l'enseignement supérieur afin de mieux comprendre les pratiques de développement de codes sources, les types de code source/logiciels produits, leurs finalités et modalités de valorisation.
- D'autres interlocuteurs privilégiés ont été les responsables de la valorisation, des juristes au sein d'instituts de recherche.

Enfin, l'étude s'est aussi appuyée sur la plateforme <http://code.etalab.gouv.fr> développée au sein du département Etalab de la direction interministérielle du numérique (DINUM) pour recenser les codes sources issus de l'administration publique, y compris un certain nombre de projets issus de la recherche publique. À ce jour, **58 organismes de l'ESR sont répertoriés pour un total de 1500 dépôts** – à noter que trois des dix organismes comptant le plus de dépôts sur code.etalab.gouv.fr proviennent de l'ESR⁶.

L'ensemble des informations recueillies dans le cadre de cette étude offre un premier panorama des pratiques de diffusion⁷ publication des codes sources, permet de noter la richesse des initiatives à l'œuvre au sein de l'ESR et de souligner les éléments d'articulation et de synergies à développer pour intégrer l'open source comme levier de valorisation et d'impact pour l'ESR à l'échelle nationale, européenne et internationale. Elle permet de relever certains éléments utiles pour alimenter immédiatement les réflexions en matière de valorisation et identifie un certain nombre de « pistes à explorer » ultérieurement afin d'approfondir le sujet.

6 Les trois organismes ayant le plus grand nombre de dépôts sont : le médialab (laboratoire de recherche interdisciplinaire de Sciences Po), Esup-Portail (le consortium porteur du projet national « Espace Numérique de Travail ») et l'Inist-CNRS (Institut de l'Information Scientifique et Technique du CNRS).

7 Le terme « diffusion » est employé dans le texte ainsi que celui de publication. Dans les milieux scientifiques, il est important de noter que la publication prend une signification particulière. Elle s'accompagne d'un processus d'évaluation par les pairs et d'une mise à disposition dans un cadre déterminé (publication dans un journal avec une date de soumission et de publication notamment). Pour plus d'informations sur les procédures de publication des logiciels de recherche, consultez Gomez-Diaz, Teresa, et Tomas Recio. 2019. « On the Evaluation of Research Software : The CDUR Procedure ». *F1000Research* 8 (novembre) : 1353. <https://doi.org/10.12688/f1000research.19994.2>.

B | La publication de codes sources : une pratique partagée entre les différents établissements de l'ESR

B.1 Une prise en compte transverse au sein de l'ESR

Les établissements de rattachement des personnes ayant répondu au questionnaire sont représentatifs de la diversité des types d'organismes regroupés sous le terme d'ESR⁸ et répartis géographiquement sur toute la France. On y trouve ainsi à la fois :

- des **établissements de recherche** (CNRS essentiellement, Institut Pasteur, INRIA, INRAE, CEA, IGN, Institut Curie, CEREMA⁹) ;
- 27 **Universités** et 7 **Grandes Écoles** (Sorbonne Université, Université de Paris, Paris Dauphine, UPEM, universités de Bordeaux, Strasbourg, Poitiers, Franche-Comté, Côte d'Azur, Versailles, Lille, Toulouse, Caen, La Réunion, Rennes, Grenoble, École polytechnique, Centrale, CNAM, INSA Rennes, ENI Tarbes) ;
- d'autres **services et groupements d'intérêt public (GIP)** et organisations de soutien à la recherche (ABES, INIST-CNRS, GIP FUN, GIP RECIA, GIP RENATER) ;
- ainsi que la participation de ministères (MESRI, ministère de la Santé et des Solidarités) et d'une direction rattachée à l'éducation nationale (DSDEN) ;
- quelques acteurs privés (entreprises ou indépendants et prestataires externes) ont également répondu au questionnaire tout autant que des acteurs ayant travaillé préalablement au sein de l'ESR.

Points clefs questionnaire :

- Les réponses proviennent de plus d'une centaine d'institutions différentes. Le CNRS est l'institution la plus représentée (avec 41 personnes affiliées au CNRS contre 18 pour l'INRIA).
- Les personnes ayant majoritairement répondu sont essentiellement des postes d'ingénieurs (de recherche ou d'étude).

8 Structure des organismes de recherche du [ministère](#), y compris des acteurs numériques spécifiques de l'ESR tels ceux identifiés par l'[AMUE](#).

9 Le CEREMA a répondu à un autre questionnaire dans le cadre de la mission politique publique de la donnée confiée à Mr. Bothorel et coordonnée par Mr. Vedel et Mme. Combes et plusieurs corps d'inspection. Un entretien a été réalisé par inno³ avec Pascal Berteaud, directeur général du CEREMA.

B.2 La distinction entre les activités de recherche et de développement : fonction, domaine et responsabilité

Parmi les personnes ayant répondu, les fonctions occupées sont les suivantes :

- ingénieurs de recherche et d'ingénieurs d'études (plus de 50 % – dont 27,6 % IR et 25,8 % IE) ;
- maîtres de conférences (12 %) ;
- doctorants (8 %) ;
- chargés de recherche (6,5 %) ;
- d'autres profils ont été indiqués notamment des post-doctorants, des bibliothécaires, des vacataires, des ingénieurs sur contrat et des développeurs/formateurs indépendants.

En revanche, les responsables des Systèmes d'information (DSI) n'ont été que peu nombreux à répondre au questionnaire (3 %) tout comme les chercheurs dans le secteur privé (2,3 %) et les responsables de la valorisation et de l'innovation (1,8 %).

Les rôles majoritaires tenus dans les projets et de façon non exclusive concernent tout d'abord le développement (88,6 %) avec des fonctions d'architecture et de conception (77,8 %), mais aussi de documentation (72,4 %) et de maintenance (70,6 %). Fait intéressant, peu de personnes ont joué un rôle dans la valorisation du projet auquel elles ont participé (34,8 %). Quelques rôles supplémentaires ont été repérés dans le questionnaire en commentaires notamment celui de test, de coordination et d'encadrement du développement ainsi que de formation et de transmission des connaissances, ce qui permet d'enrichir la palette des activités nécessaires à la vie d'un projet « logiciel » ou de développement de codes sources.

Points clefs questionnaire :

- La majorité des personnes ayant répondu au questionnaire l'ont fait en leur nom propre.
- Seulement ~9 % de réponses ont été données au nom d'une équipe et ~ 2 % pour un établissement.

	Nb réponses	Pourcentage
Vous-même	195	87.1%
Votre équipe	21	9.4%
Votre établissement	5	2.2%
Autre	3	1.3%

Figure 1: À quel titre répondez vous au questionnaire ?

— La plupart des initiatives développées sont majoritairement menées dans le cadre de missions professionnelles (93 %).

— Les répondants sont majoritairement des hommes¹⁰.- La majorité des réponses proviennent de personnes dont les fonctions sont associées à des projets de développement. Environ la moitié participent également à titre personnel à de tels projets (voir ci-dessous)

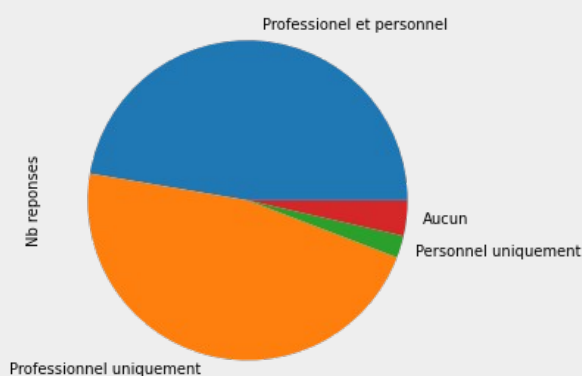


Figure 2: Dans quel cadre développez-vous des codes sources ?

Peu de personnes ont joué un rôle dans la valorisation du projet auquel elles ont participé.

— Une grande majorité des personnes affirment avoir partagé le code source publiquement.

Les personnes ayant répondu ont majoritairement une pratique de production de code source et une familiarité avec les principes open source de par des valeurs et modes d'organisation qui sont connexes. En effet, l'open source rejoint des pratiques et des modalités de fonctionnement ancrées au sein de certaines cultures disciplinaires de l'ESR (partage, relecture par les pairs, mutualisation, organisation en réseau).

10 L'échantillon ayant répondu au questionnaire majoritairement masculin semble être représentatif d'une population de programmeurs principalement masculine.

B.3 Mode de production scientifique et open source : des motivations communes et des cultures intimement liées

Il est important de rappeler que les principes du logiciel libre et de l'open source sont nés au cœur des universités au début des années 1970. Face à un mouvement d'appropriation des logiciels – pleinement matérialisé par la consécration d'un droit d'auteur spécifique dans les années 80 – les universités ont ainsi joué un rôle crucial – en s'appuyant sur leur force économique et intellectuelle – pour s'assurer de l'existence de logiciels libres et open source nécessaire à leur autonomie. Au-delà de leur contribution technique, les universités et centres de recherche ont aussi fortement contribué dans les années qui suivirent à renforcer les outils juridiques associés aux pratiques de l'open source¹¹.

Ainsi, au classement des motivations à publier les projets en open source dans le questionnaire, la première motivation est celle de « favoriser la mutualisation du travail d'équipe » (classement 1 : 41,2 %). Ensuite, il s'agit « d'encourager la réutilisation du code source » (classement 2 : 30,3 %). La « pérennisation du code source » et « la valorisation logicielle » ne se positionnent qu'en troisième position (20,4 % et 16,6 % respectivement). Enfin « favoriser la citation et la reconnaissance des contributions » ne vient qu'en dernier lieu (classement 4 : 18,1 %)¹².

Points clefs questionnaire :

— La mutualisation et l'encouragement à la réutilisation des codes sont les premières motivations pour publier des codes sources, la pérennisation, la reconnaissance et la valorisation venant en dernier lieu.

Un chercheur¹³ note ainsi que l'open source s'est démocratisé, devenant un usage commun là où il était l'exception. Cela renvoie à une dimension générationnelle forte confirmée par ailleurs, les nouveaux chercheurs et ingénieurs de recherche étant beaucoup plus enclins à partager leur production et familiers avec les outils et processus de l'open source.

Aujourd'hui c'est devenu la normalité pour les plus jeunes. Au début il fallait se battre et maintenant [c'est] un sujet connu notamment des ingénieurs de recherche/ "

11 Patrice Flichy, « Internet ou la communauté scientifique idéale », *Réseaux* 17, n° 97 (1999) : 77-120, <https://doi.org/10.3406/reso.1999.2168> ; Célya Gruson-Daniel, « Chapitre 1 – Open : Les Différentes Facettes Du 'Numérique' », in *Numérique et Régime Français Des Savoirs En-action : L'open En Sciences. Le Cas de La Consultation République Numérique (2015)*, 2018, <https://phd-cgd.pubpub.org/pub/facettes-numerique-fr> ; Sébastien Broca, *Utopie du logiciel libre* (Neuvy-en-Champagne : Le Passager Clandestin, 2013).

12 Une enquête menée par le projet PLUME visait à recenser les motivations pour diffuser un logiciel dans l'ESR et apportent quelques précisions et ou motivations supplémentaires à celles proposées dans le questionnaire. Voir : Baudin Véronique, « Pourquoi diffuser un logiciel développé dans un laboratoire ou une université avec une licence libre ? | Ressource PLUME », text (PLUME, 140909), <https://projet-plume.org/ressource/pourquoi-diffuser-en-libre#1>.

13 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

Au départ, la personne note que l'open source n'était accepté que s'il n'y avait pas de valorisation¹⁴.

DevLog : Un exemple de réseau d'acteurs du Développement LOGiciel au sein de l'Enseignement Supérieur et de la Recherche

Sa mission essentielle est de favoriser les échanges, le soutien aux réseaux régionaux dans leurs actions, des offres complémentaires d'actions et de formations. Le réseau vise à faciliter le lien entre la communauté et les tutelles afin de remonter les réalités du terrain.

Voir <http://devlog.cnrs.fr/>

Le partage du code source pour favoriser les mutualisations, le travail collaboratif, mais aussi la vérification par d'autres personnes contribuant au projet fait écho à l'idéal de la démarche scientifique. Néanmoins, les pratiques open source diffèrent en fonction des communautés disciplinaires qui n'ont pas les mêmes cultures de recherche (communautés épistémiques)¹⁵. Néanmoins, l'adoption de pratiques d'ouverture au sein de l'ESR¹⁶ évolue ces dernières années en faveur d'un rôle de plus en plus important reconnu aux données, leur gestion, leur analyse et leur mise à disposition (science ouverte, science des données)¹⁷ au sein d'un champ disciplinaire élargi (notamment en SHS¹⁸). Cette même idée se retrouve dans le questionnaire comme le souligne ce commentaire : « *Les pratiques évoluent très vite ces dernières années, avec un usage de plus en plus tôt de forges et de code ouvert* ». De fait, les pratiques sont encore hétérogènes de par la nature « hybride » des logiciels¹⁹ et naviguent entre :

- un code source diffusé « de fait » dans des disciplines telles que les mathématiques ou l'informatique, car la production de logiciels est considérée comme un résultat même de recherche ou bien un objet d'étude (comme artefact)²⁰ ;
- un code source graduellement partagé pour des logiciels considérés comme moteurs et outils de traitement d'analyse au sein de travaux de recherche, faisant également de l'open source une pratique de plus en plus fréquente au sein de nouvelles disciplines scientifiques touchées par l'essor du numérique. En effet, ces dernières années, les utilisations de langages de

14 Ce même chercheur faisait ainsi remarquer qu'« au niveau des IR le message est bien passé, tout le monde connaît les bonnes techniques et solutions depuis 5 ans c'est bon, il y a un grand progrès au niveau des professionnels les plus techniques. On peut s'adresser à eux ».

15 Morgan Meyer et Susan Molyneux-Hodgson, « « Communautés épistémiques » : une notion utile pour théoriser les collectifs en sciences ? », *Terrains travaux* n° 18, n° 1 (18 août 2011) : 141-54.

16 En France, en 2018, le CoSO (comité pour la science ouverte) a été fondé pour assurer la mise en œuvre d'une politique de soutien à l'ouverture des publications et des données de la recherche. Les groupes de travail Collège « logiciels libres et open source » et « Collège Données de la recherche » ont été notamment constitués.

17 Gruson-Daniel Célya, « Numérique et régime français des savoirs en~action : l'open en sciences. Le cas de la consultation République numérique (2015) » (Université Paris Descartes, 2018), <https://doi.org/10.5281/zenodo.1491292>.

18 Des distinctions sont néanmoins présentes en fonction des disciplines. Une personne interrogée notait par exemple que l'économie ne participait que peu à des projets opens source.

19 Comme le rappelle la note « [valorisation des logiciels issus de la recherche](#) » du groupe de travail du CoSO sur les logiciels libres, le code source peut être à la fois un résultat (comme preuve d'existence d'une solution), un objet d'études (comme artefact). Il peut également être un moteur, c'est-à-dire un outil permettant la collecte, le traitement de données.

20 Collège « logiciels libres et open source » (CoSO), « Note d'opportunité sur la valorisation des logiciels issus de la recherche », Ouvrir la Science (CoSO), 2019, <https://www.ouvriirlascience.fr/note-dopportunite-sur-la-valorisation-des-logiciels-issus-de-la-recherche>.

programmation open source – dont les deux plus répandus sont Python et R – ont été de plus en plus fréquentes avec une offre de formation grandissante au sein des instituts, écoles doctorales, etc. Les sciences humaines et sociales développent également des outils open source : dans le cadre de l'analyse de réseau, par exemple, ou encore du traitement naturel du langage (*Natural Language Processing*).

Le développement de bibliothèques autour des langages de programmation Python et R en fonction des projets et disciplines permet des usages spécifiques et adaptés à certaines disciplines ou bien, tout au contraire, des développements transverses. On peut citer par exemple le projet phare [SciKitlearn](#) (bibliothèque Python de *machine learning*) avec des usages qui dépassent largement la communauté académique (utilisé massivement dans le milieu de la data science). Pour le traitement statistique, une personne interrogée²¹ notait le remplacement du logiciel STATA ou encore de Matlab de statistique propriétaire par l'usage de R dans les enseignements : « ce qui joue beaucoup c'est la montée conjointe des langages R et Python dans les domaines où Stata et Matlab étaient habituellement utilisés ». ²²

Recommandation 1 : Reconnaître le rôle et la place de l'open source dans l'enseignement supérieur et la recherche française et situer la France dans une perspective internationale.

B.4 Une multitude de projets de l'ESR en open source et un recensement complexe à mettre en œuvre

Une multitude de projets ²³ ont été recensés lors de cette étude, qui pour certains jouent rôle majeur à l'échelle des communautés de recherche (développement d'outil phare dans des disciplines), mais aussi plus globalement par un tissu d'acteurs socio-économiques et administratifs, car le code source a été ouvert.

Dresser un panorama exhaustif de ces projets se révèle complexe compte tenu de l'éclatement des initiatives²⁴ et en l'absence d'inventaire à l'échelle des instituts de l'ESR. Faisant écho à une culture de la recherche qui s'appuie sur des communautés de chercheurs distribuées et mondiales, les projets open source et leurs écosystèmes sont souvent difficiles à circonscrire et ne rentrent pas dans les processus généralement utilisés par les instituts de l'ESR pour dénombrer et valoriser les

21 Entretien avec Guillaume Plique, ingénieur de recherche au Medialab (SciencePo Paris).

22 Dans la même optique [Scilab](#) est un logiciel qui a été cofinancé par le CNES et l'INRIA pour remplacer MatLab et qui lui même aujourd'hui est concurrencé par R & Python.

23 Cf Annexe H.2 [Listes de logiciels open Source et ESR](#)

24 Les initiatives sont souvent individuelles à l'échelle de laboratoire ou d'une équipe ou bien même d'un doctorant développant son propre logiciel pour ses recherches.

logiciels produits. Ainsi, seuls les projets logiciels les plus importants sont visibles, comme le mentionne un commentaire du questionnaire :

Du fait de cette décentralisation et non-planification, les universités, labos et EPST sont souvent incapables de lister les logiciels qui y sont produits, exception faite des plus visibles.

Comme le souligne une personne interrogée²⁵, certains projets²⁶ n'auraient pas connu un succès international s'ils n'avaient pas été en open source. D'autres logiciels spécifiques sont aussi développés pour les besoins de communautés données. On trouve également des infrastructures numériques clefs pour le soutien à l'enseignement supérieur et à la recherche²⁷.

Des projets de recensement logiciels sont néanmoins à mentionner à l'échelle de laboratoires ou d'institut. Le laboratoire LGIM par exemple possède une politique de patrimoine logiciel afin d'étudier la mise en place de services (suivi de versions, publications) et de favoriser leur visibilité.²⁸ L'INRIA a par exemple mis en place une base de données de ses logiciels (BIL)²⁹, mais comme le note une personne en commentaire du questionnaire : « *elle est probablement incomplète et datée, car mise à jour simplement en fonction de besoins ponctuels (création de startups, dépôt APP, demande de moyens)* ». De même, le projet PLUME³⁰ initié au CNRS – dont le but était de promouvoir les logiciels libres, de mutualiser et de valoriser les compétences sur les logiciels – s'est arrêté en 2013 après 7 ans d'existence et un travail conséquent de référencement logiciels avec la production de 406 fiches descriptives de logiciels validés dont 95 développés dans la communauté de l'ESR³¹.

Plus encore, les entretiens ont révélé que la nature même de l'open source complique l'identification des logiciels qui « passent au travers des mailles » des processus de recensement et de valorisation. Point important, les processus envisagés pour identifier de tels projets sont susceptibles de générer un formalisme ou une charge complémentaire qui est considérée comme inutile par les projets ne souhaitant pas se tourner vers une valorisation économique. Par ailleurs,

25 Entretien avec Roberto Di Cosmo, professeur d'informatique, détaché chez l'INRIA, membre du laboratoire PPS (IRIF/Université de Paris) et directeur de l'initiative Software Heritage.

26 On peut citer par exemple SciKitLearn ou d'autres logiciels phares dans certaines disciplines telles Gephi en analyse de réseaux ou encore SageMath soutenu par l'initiative européenne OpenDreamKit <https://opendreamkit.org/project/software/> qui a aidé également le déploiement de Jupyter Notebook.

27 Une liste de logiciels Open Source produits ou utilisés par l'ESR figure en annexe « Listes de logiciels open Source et ESR ». Les projets recensés par le questionnaire sont souvent développés dans le cadre de recherches en STM (astronomie, chimie, hydrologie, agronomie, mathématiques, ingénierie informatique, neurosciences). Le médialab et le CNRS (avec Mobiliscope) font figure de proue en ce qui concerne le développement d'outils propres aux sciences sociales.

28 Consultez le poster « Patrimoine logiciel LIGM) http://igm.univ-mlv.fr/~teresa/2013octPostersAERES/PatrimoineLogicielLIGM/LogicielsLIGMPlume2013_FR.pdf et le guide laboratoire pour recenser ses développements logiciels du projet PLUME : <https://www.projet-plume.org/ressource/guide-laboratoire-recensement-developpements-logiciels>

29 Voir plus loin E.1.3 Un référencement des initiatives existantes et la construction raisonnée d'indicateurs

30 cf. <https://hal.archives-ouvertes.fr/hal-02069359/document>

31 Les fiches descriptives de logiciels validés correspondent des logiciels utilisés sur au moins 3 sites dont 1 site correspond à un laboratoire de recherche ou une université. Le projet PLUME a référencé également des fiches descriptives des développements au sein de l'ESR (358 en français et 116 en anglais), qui décrivent succinctement des logiciels développés ou en cours de développement. Elles ont été conçues dans le cadre du projet RELIER cf. https://projet-plume.org/types-de-fiches#logiciel_valide

toute démarche administrative supplémentaire ajoutée au travail des chercheurs est perçue négativement et risque d'être contre-productive comme le souligne une des personnes interrogées³².

Il faut trouver un processus qui combine un gain de temps et des bénéfices immédiats pour les chercheurs - de sorte que l'évaluation de l'open source ne soit pas perçue comme générant une étape bureaucratique supplémentaire.

Recommandation 2 : Améliorer le recensement actuellement existant au sein de l'ESR des projets open source afin de construire une politique publique représentative de ces usages.

Recommandation 3 : Corréler davantage les processus que mettent en place les établissements de l'ESR pour piloter leur activité avec les bénéfices et avantages que retirent les projets.

32 Entretien avec Roberto Di Cosmo, professeur d'informatique, détaché chez l'INRIA, membre du laboratoire PPS (IRIF/Université de Paris) et directeur de l'initiative Software Heritage.

C | L'ouverture des codes sources : pour un enseignement supérieur et une recherche publique fiable

Une question du questionnaire visait à savoir si les personnes avaient connaissance de codes sources qui devraient être publiés, mais qui ne le sont pas encore. Même si peu de personnes donnent des exemples de logiciels non publiés (62,4 %), les personnes ayant répondu sont favorables à une ouverture du code source. Plusieurs raisons sont citées qui rejoignent des caractéristiques fortes constitutives de l'enseignement supérieur et de la recherche en tant que service public, mais également comme institution garante de productions de savoirs de qualité.

C.1 Le cadre légal applicable aux établissements de l'ESR

Les instituts de l'ESR sont soumis à un cadre légal favorable à l'open source :

- **en tant qu'administration** : tout logiciel, en tant que document produit ou reçu par un acteur de l'ESR dans le cadre de ses missions de service public de recherche, est soumis par principe au régime de l'open data. En effet, depuis la loi « Valter » de 2015, les centres de recherche ne peuvent plus appliquer leurs propres conditions de réutilisation des documents administratifs. Ils sont ainsi pleinement soumis au régime général de l'open data revu par la Loi pour une République numérique (LPRN) de 2016 qui a étendu le champ d'application de l'open data aux codes sources. Dès lors, la diffusion des codes sources devra se faire sous une licence open source et dans un standard ouvert, sans que des restrictions techniques ou juridiques puissent entraver son accès, sa modification, son analyse et sa redistribution. Au-delà de l'application du régime standard de l'« open data à la demande », les logiciels entrent dans la catégorie dite de l'« open data par principe » (ou par défaut) en tant que données « dont la publication présente un intérêt économique, social, sanitaire ou environnemental³³ » – ce qui impose un comportement proactif d'ouverture et de diffusion ;
- **en tant qu'acteur de l'ESR** : la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche dispose par ailleurs que les « logiciels libres sont utilisés en priorité » (codifié en l'article L. 123-4-1 du code de l'éducation) par le service public de l'enseignement supérieur dans la mise à disposition de « services et des ressources pédagogiques numériques ». Ces dispositions ont été prévues afin de favoriser une mutualisation dans le financement, l'achat et l'utilisation de logiciels au bénéfice des acteurs de l'enseignement supérieur et de la recherche³⁴.

33 Article L 312-1-1 du CRPA.

34 Un répondant au questionnaire a ainsi rédigé un commentaire en ce sens : « *ParcoursSup, tout l'algorithme d'affectation des élèves aux établissements, commentés et expliqués. Je pense qu'à partir du moment qu'un établissement public passe une commande de logiciels ceux-ci devraient être systématiquement mis sous licence Open Source (à part bien sûr cas vraiment spécifiques). Cela devrait être la condition primordiale.* »

Néanmoins, il semblerait que l'absence de sanction et de recours administratifs entraîne une faible application du cadre prévu par le législateur³⁵.

À noter qu'au niveau juridique, les licences open source CeCILL (CeCILL, CeCILL-B et CeCILL-C) ont été rédigées par le CNRS, le CEA et l'INRIA pour disposer de licences compatibles avec le droit français (face à un paysage open source majoritairement nord-américain) et, ainsi, sécuriser et accompagner les pratiques de l'open source au sein des centres de recherche. Ces licences ne sont aujourd'hui plus portées politiquement et humainement, rendant les chercheurs ou chargés de valorisation démunis devant ce choix, préjudicant aux objectifs initiaux. Un des chercheurs interrogés³⁶ notait qu'il avait utilisé les licences CeCILL mais avait arrêté, car « *cela posait problème avec l'international* ». D'autres pays, telle l'Allemagne, ont aussi historiquement produit leur propre licence, aujourd'hui abandonnée au profit des licences open source qui sont devenues des standards de fait. On peut noter aussi l'existence de la licence EUPL³⁷, qui ne figure pas néanmoins dans les licences autorisées au sein de l'administration française³⁸.

Recommandation 4 : Réaliser une étude pour évaluer la pertinence de l'emploi des licences françaises CeCILL et de la licence EUPL dans l'Enseignement Supérieur et de la Recherche et formuler des règles explicites quant à leur usage.

Outre ces enjeux de transparence, de « redevabilité » et de mutualisation, plusieurs commentaires soulignent également que l'ouverture des codes sources contribue à la qualité de la recherche produite, rejoignant les enjeux actuels portés sur la reproductibilité dans une démarche de science ouverte³⁹.

C.2 Science ouverte, reproductibilité et intégrité scientifique

L'ouverture des codes sources rejoint une problématique majeure au sein de la recherche qui concerne la qualité des savoirs produits et des modalités d'évaluation de la recherche tout autant que de son personnel (enseignant.e.s-chercheur.e.s, etc.). À ce propos, ont été cités également plusieurs fois en commentaires des entretiens et du questionnaire, l'enjeu de la reproductibilité en sciences nécessitant l'ouverture du « code scientifique » tout autant que des formations un système de reconnaissance et d'évaluation prenant en considération ces nouvelles pratiques.

35 Si ce n'est parfois quelques critères parfois favorisant les projets open source dans le cadre de certains investissements du MESRI (notamment dans la gestion de fonds du Programme d'Investissement d'Avenir PIA).

36 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

37 cf. [https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/European%20IPR%20Helpdesk%20Bulletin%20Issue%20\(26\).pdf#page=4](https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/European%20IPR%20Helpdesk%20Bulletin%20Issue%20(26).pdf#page=4)

38 Décret n° 2017-638 du 27 avril 2017 relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation <https://www.legifrance.gouv.fr/codes/id/LEGIARTI000034504991/2017-04-29/>

39 Jeffrey M. Perkel, « Challenge to Scientists : Does Your Ten-Year-Old Code Still Run ? », *Nature* 584, n° 7822 (24 août 2020) : 656-58, <https://doi.org/10.1038/d41586-020-02462-7>.

Points clefs questionnaire :

— Une idée récurrente dans les commentaires les codes sources des logiciels cités dans des articles devraient être publiés. L'enjeu semble être celui de la reproductibilité.

Ce point rejoint de façon plus globale les initiatives actuelles à l'échelle internationale, européenne et nationale de science ouverte (open access, plan de gestion des données, principes FAIR) soulignée notamment par la note d'opportunité du CoSo sur la valorisation des logiciels issus de la recherche. La publication du code source représente en ce sens une des briques du travail scientifique à partager afin de pouvoir tracer l'ensemble de la démarche scientifique :

- ces codes sources ici ne sont pas produits dans une optique de production logicielle ou de valorisation, c'est-à-dire dans le but de développer un outil utile à un grand nombre et qui a vocation d'être maintenu ;
- ces codes sources « *bouts de code* », « *scripts* », « *petits projets* », « *pipeline de données* », comme le rappellent certains commentaires et remarques de personnes interrogées, sont néanmoins la traduction en langage de programmation d'algorithmes nécessaires au traitement, à l'analyse et à la visualisation de données pour la production des résultats et des figures associées aux publications scientifiques⁴⁰.

Bien qu'associée à la notion juridique de logiciels⁴¹, cette distinction est importante à souligner car elle permet de mettre l'accent sur différents publics concernés par la production, diffusion et utilisation de ces éléments. En effet, ces « *scripts* » et « *codes scientifiques* » sont des noms usuels employés par les personnes les produisant (chercheurs, doctorants, etc.) dont les connaissances de développement « *logiciels* » sont différents d'ingénieurs dont le développement est le cœur de métier.

Des journaux scientifiques demandent désormais la publication du codesource (et également des jeux de données) qui ont servi à la production de résultats de recherche⁴². C'est le cas par exemple du journal IPOL⁴³ qui se présente comme un « *Open Science and Reproducible Research journal* ». L'objectif est de souligner l'importance des mathématiques comme source pour le design algorithmique et la reproductibilité en recherche. La revue ReScience⁴⁴, quant à elle encourage la répliation de travaux déjà publiés et promeut de « *nouvelles implémentations en open source* afin

40 Plusieurs citations peuvent être tirées du questionnaire : « Il y a des projets open source à long terme dans le but d'utiliser un projet utile. Et de l'autre côté il y a des bouts de codes pour un projet spécifique et ensuite on les oublie. On le rend public, mais il y a pas d'idées de maintenance et de services » et « Ce petit programme très vite obsolète et très conjoncturel, il a pas vocation d'avoir une longue vie. »

41 Pour un référencement des différentes définitions données aux logiciels de recherche (*research software*) comprenant ces « *scripts* »/codes scientifiques, voir : Teresa Gomez-Diaz et Tomas Recio, « On the Evaluation of Research Software : The CDUR Procedure », *F1000Research* 8 (26 novembre 2019) : 1353, <https://doi.org/10.12688/f1000research.19994.2>.

42 Extrait du questionnaire : « Diffuser le code source, c'est s'assurer que le travail peut être vérifié facilement. C'est une question de transparence, et il me semble que c'est même obligatoire dans certains domaines (au moins de fournir le code aux reviewers) ».

43 <https://www.ipol.im/>

44 <http://rescience.org/x>

de s'assurer que les travaux originaux soient reproductibles ». ⁴⁵ En effet un de ses fondateurs en entretien ⁴⁶ note que les enjeux de reproductibilité nécessitent d'une part d'ouvrir les codes sources, mais également de s'assurer que l'environnement de travail disponible à ce moment puisse être retrouvé, et cela même 10 ans après. Cela implique ainsi des questionnements transverses sur le référencement et l'archivage et la constitution d'un patrimoine logiciel dont Software Heritage ⁴⁷ est une des initiatives clefs.

Recommandation 5 : Adopter une approche différenciée des bénéfices de l'open source en fonction des finalités de production des logiciels (codes scientifiques isolés ou production logicielle).

Recommandation 6 : Sensibiliser au partage des codes sources (et de jeux de données) étant associés aux résultats de recherche comme critères d'évaluation pour la publication dans une revue scientifique.

Pour l'ensemble des codes sources, des problématiques transversales se retrouvent telles : l'archivage, la maintenance, l'évaluation, mais aussi la valorisation possible des codes sources ⁴⁸.

	Production logicielle (outils et support)	Codes scientifiques/scripts
Métiers concernés	Ingénierie logicielle, ingénierie de recherche et d'étude	Métiers associés à la production de résultats scientifiques et évalués sur ces éléments (MCU, PU, DR, Post-doctorat et doctorat)
Finalité du code	Mutualisation et réutilisation du code, contribution et développement de nouvelles fonctionnalités, utilisation massive	Reproductibilité, preuve, traçabilité de la recherche, science ouverte
Enjeux	Valorisation, maintenance et soutenabilité, dynamique d'écosystème, visibilité	Acculturation aux pratiques open source documentation, citation, formation
Enjeux transversaux	Pérennisation, archivage, citation, référencement	
Ressources associées	Infrastructures de publication facilement interfaçables, référencement sur des plateformes, mises en place d'indicateurs, modèle de valorisation basée sur l'open (licences, contrat),	Personnes-ressources dans les laboratoires et en bibliothèque, infrastructures de publication interfacée à des archives, référencement avec identifiant pérenne,

Tableau visant à distinguer différentes finalités de l'ouverture du code source et des personnes concernées au sein de l'ESR

Cette distinction « manichéenne » entre production logicielle à maintenir et code source en soutien d'une démarche de science ouverte nécessite néanmoins d'être nuancée. D'une part, les codes sources servant à la production de résultats peuvent donner naissance lorsqu'ils sont stabilisés ou bien enrichis au développement de logiciel ou bien à des libraires supplémentaires pour des langages de programmation. D'autre part, ces deux éléments sont intimement liés pour la reproductibilité (le code source ne pouvant souvent qu'être reproduit si l'on retrouve le même environnement de travail (version du logiciel utilisé, dépendances, etc.).

45 *ReScience/ten-years* (2019 ; repr., ReScience organization, 2020), <https://github.com/ReScience/ten-years> ; Konrad Hinsén et Nicolas P. Rougier, « ReScience », in *Open science, transparence et évaluation. Perspectives et enjeux pour les chercheurs* (Bordeaux, France : URFIST Bordeaux, 2017), <https://hal.archives-ouvertes.fr/hal-01573262>.

46 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

47 <https://www.softwareheritage.org/?lang=fr>

48 Pour plus d'informations, voir également la note d'opportunité du CoSO sur la valorisation des logiciels issus de la recherche décrit en détail les différents enjeux associés à la contribution, à la citation (métadonnées) à l'archivage.

Plusieurs commentaires du questionnaire et des entretiens mentionnent en ce sens que les enjeux de reproductibilité ou encore d'archivage sont des sujets et des pratiques nouvelles nécessitant un accompagnement. Un manque de méthodologie est souligné avec la nécessité de former les chercheurs à ces pratiques et aux enjeux de reproductibilité (environnement de travail, documentation du code, etc.). L'ouverture du « code scientifique » dans une démarche de science ouverte vise l'ensemble des communautés de chercheurs. Si les métiers d'ingénierie de recherche et d'étude et de développement logiciel sont de plus en plus familiarisés à l'open source et aux enjeux juridiques associés au sein de l'ESR, ce n'est pas le cas souvent des chercheur.e.s et étudiant.e.s (MCU, DR, PU, post-doctorat, doctorant) qui manquent de connaissances à ce sujet et de clefs de compréhension⁴⁹. À ce sujet, la procédure CDUR⁵⁰ a été proposée pour évaluer les logiciels de la recherche en quatre étapes : Citation, Dissémination (Dissemination), Utilisation (Use), Recherche (Research), pour analyser les conditions d'identification et de citation, les bonnes pratiques de diffusion et de (ré-)utilisation et l'évaluation de la recherche associée à ce type de logiciel.

Un accompagnement à ces pratiques (connaissances organisationnelles, juridiques, mais aussi génie logiciel (maintenance, suivi de version, test, etc.) est ainsi recherché. Le projet PLUME a ainsi proposé des formations et des ateliers pour sensibiliser à ces enjeux⁵¹. Un MOOC recherche reproductible a été également développé sur la plateforme FUN et soutenue par l'INRIA⁵². Un de ses créateurs⁵³ souligne néanmoins que les écoles doctorales ne sont souvent pas réceptives à ce nouveau genre de formation. Or l'open source joue également un rôle majeur dans la formation des nouvelles générations.

Recommandation 7 : Mettre à disposition des personnels de la recherche des formations, bonnes pratiques et outils en matière de publication de code source et les accompagner dans cette appropriation.

C.3 Open source et enseignement supérieur

De nombreux projets de recherche visent aussi à fournir les outils aux chercheur.e.s et ingénieur.e.s de demain. Outre l'utilisation de plus en plus fréquente de R et de Python dans les enseignements⁵⁴ d'informatique, de traitement et d'analyse des données (data science) (cf. B.3), plusieurs projets issus de travaux de recherche sont particulièrement connus⁵⁵ =et utilisés dans le cadre de formation, que le projet provienne :

49 Une personne responsable de l'initiative Reproducible Research Oxford contactée par mail soulignait ainsi « *From what I have seen, it seemed to assume respondents would know that pieces of software they share on e.g. GitHub need a licence to allow reuse. From what we've seen (anecdotally), this may not be the case : people tend to think that pieces of software shared publicly are automatically Free and Open Source Software (FOSS) ».*

50 Gomez-Diaz et Recio, « On the Evaluation of Research Software ».

51 Les formations sont référencées sur le site du projet PLUME <https://www.projet-plume.org/patrimoine-logiciel-laboratoire>

52 Voir <https://www.fun-mooc.fr/courses/course-v1:inria+41016+self-paced/about>

53 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

54 Des formations sont offertes au sein des universités et des Écoles mais aussi sous forme de MOOC accessibles en ligne.

55 Un certain nombre de logiciels employés au sein de l'ESR sont mentionnés sur le projet PLUME : `fiches_logiciels_dev_internes`

- des écoles (que ce soit à l’initiative des élèves tel VidéoLan, initialement développé par le bureau des élèves de l’École Centrale Paris, ou des instituts tels CHOCO, GPAC ou Open AIR développés par l’IMT) ;
- de centres de recherche, notamment du fait de l’activité d’enseignement réalisée par bon nombre de chercheurs (tels OrfeoToolBox⁵⁶ dans le domaine du géospatial, Gephi pour l’analyse de réseau, etc.).

L’implication dans la conception, le développement et l’usage de logiciels sur la base desquels sont formées les nouvelles générations (recherche, ingénierie) est déterminante pour participer au rayonnement de la recherche française. La diffusion en open source dans le cadre d’une gouvernance ouverte est une garantie facilitant l’implication et contribution de chercheurs étrangers. Ainsi la dimension communautaire internationale du logiciel open source GAMA Platform a-été décisive dans le choix de se former à cet outil par plusieurs équipes de recherche japonaises (Kyoto University et Ritsumeikan University).

On peut noter également dans le champ de l’enseignement à la programmation des projets déployés de façon indépendante en open source par des enseignants et chercheurs. C’est le cas de la plateforme hackinscience⁵⁷, un espace d’apprentissage de python avec la proposition de correction automatique d’exercices. Un des fondateurs⁵⁸ observe notamment son utilisation par des enseignants-chercheurs au sein d’Université aujourd’hui. Le projet OpenDreamKit⁵⁹ (projet européen H2020) gagne à être aussi souligné car il illustre l’importance de financement pour la création et le maintien d’environnement de recherche virtuel et collaboratif. Un des projets phares soutenus est Jupyter Notebook⁶⁰, un environnement de recherche jouant le rôle de « cahier de laboratoire », d’espace de documentation et de publication interactive de code pour des projets computationnels et de traitement de données. Ces environnements de travail apportent des briques opérationnelles afin de faciliter la reproductibilité de ce type de projet scientifique.

Plus globalement dans le champ de l’enseignement en ligne, le déploiement des MOOC a été majeur ces dernières années. En France, la plateforme FUN (France Université Numérique) développée à partir de 2013 repose sur la solution open EdX⁶¹ fondé par Harvard et le MIT en 2012. Celle-ci proposant selon un ancien développeur chez FUN⁶² « des avantages techniques et d’utilisation que Moodle⁶³ [projet open source d’apprentissage en ligne] négligeait un peu comme la problématique des interfaces utilisateurs qui sont souvent mal pensées dans le champ de l’open source ». Un travail conséquent a été nécessaire au début pour FUN afin d’adapter le projet pour « s’affranchir de la partie du code qui utilisait Amazon et YouTube. »⁶⁴. Ce même développeur notait des collaborations limitées avec la communauté EdX qui dénote d’un projet initial où le code est diffusé en open source mais sans être associé à une dynamique de contribution et communautaire comme cela peut-être le cas au sein de la communauté Moodle. À ce sujet, la

56 <https://www.orfeo-toolbox.org/>

57 <https://www.hackinscience.org/>

58 Entretien avec Julien Palard, co-fondateur de Hackinscience, développeur et formateur indépendant Python.

59 <https://opendreamkit.org/about/about/>

60 <https://jupyter.org/>

61 <https://open.edx.org/>

62 Entretien avec Laurent David, ancien développeur au sein du GIP FUN (2014-2017) et actuellement président de CALL Learning.

63 <https://moodle.org/?lang=fr>

64 Le code source de FUN est disponible sur Github : <https://github.com/openfun>

problématique du développement et du maintien de projets logiciels comme support de l'ESR a été adressée et implique de considérer l'open source comme élément clef des infrastructures de l'enseignement supérieur et de la recherche.

C.4 Les infrastructures support de l'ESR

Plusieurs commentaires insistent sur le fait que la production logicielle est issue de l'investissement d'argent public et devrait ainsi être ouverte. Cela a été mentionné pour le cas de la plateforme d'archivage HAL en commentaire du questionnaire⁶⁵. Un autre exemple cité était l'ouverture des algorithmes entiers de ParcoursSup en précisant l'importance d'une documentation associée pour expliciter ces algorithmes⁶⁶. Plusieurs réponses soulignent également la nécessité de logiciels libres pour les infrastructures de recherche et d'enseignement supérieur. Cela concerne aussi bien des logiciels employés au sein des laboratoires ou de l'enseignement, qu'ils soient généralistes (à la place des suites Microsoft) ou bien plus spécifiques.

Le logiciel, ainsi plus qu'un objet de recherche, un artefact ou un outil, est aussi à la base même des infrastructures de soutien au fonctionnement de l'ESR, qui représenterait une quatrième fonction, ou catégorie, majeure à considérer. En anglais, le terme d'« *Open Scholarly Infrastructures* » ou d'« *Open Source Publishing Tools and Platforms* » est employé pour mentionner les infrastructures soutien à la publication, à l'archivage des produits de la recherche (données, publications, etc.)^{67 68}. Cette prise en considération est devenue majeure depuis quelques années avec la thématique plus générale des « *Digital Infrastructures* » et de leur durabilité (*sustainability*) qui bénéficie de financement conséquent notamment aux États-Unis⁶⁹. Outre le GIP (Groupement d'intérêt public) FUN en open source, le GIP AMUE (Agence de Mutualisation des Universités et Établissements) dédié à la mutualisation des outils support des domaines d'activités de l'ESR se positionne d'ores et déjà dans une approche de co-construction et de mutualisation avec les universités pour développer des projets open source⁷⁰. L'open source étant un levier pour favoriser une mutualisation des

65 Le CCSD a récemment publié le code de HAL, désormais disponible sur [github](#). De nouveaux usages au sein de l'administration sont en cours d'exploration. La gouvernance de HAL est actuellement en cours de reconfiguration pour développer une gouvernance partagée et un modèle de financement pérenne : <https://www.enseignementsup-recherche.gouv.fr/cid154220/hal-une-gouvernance-partagee-et-un-modele-de-financement-perenne.html>

66 Avant ParcoursSup, la publication du code de APB (Admission Post-Bac) avait elle-même fait très largement réagir, la transparence offerte par l'open source rendant en effet visible des prises de décisions automatisées par l'outil (aujourd'hui remplacées par des interventions humaines). La documentation et le code sont ouverts sur framagit : <https://framagit.org/parcoursup/algorithmes-de-parcoursup>

67 John W Maxwell et al., *Mind the Gap : A Landscape Analysis of Open Source Publishing Tools and Platforms*, 1^{re} éd. (PubPub, 2019), <https://doi.org/10.21428/6bc8b38c.2e2f6c3f>.

68 Neylon Cameron, « Principles for Open Scholarly Infrastructures », 2015, <https://cameronneylon.net/blog/principles-for-open-scholarly-infrastructures/>.

69 La Ford Foundation par exemple propose des financements de projets portant que les infrastructures numériques. Le livre de Nadia Eghbal, *Roads and Bridges : The Unseen Labor Behind Our Digital Infrastructure* (Ford Foundation, 2016).

70 Un groupe de travail de veille et de prospective a ainsi été constitué pour réfléchir sur les choix de technologies à mettre en œuvre compatible open source tout en questionnant le modèle économique associé. Ce point nécessite une transition économique pour ce Groupement d'Intérêt Public afin de s'adapter, mais également repenser le rapport avec les établissements adhérents et autres acteurs numériques de l'ESR selon David Rongeat (responsable numérique AMUE) et Bertrand Mocquet (expert numérique AMUE). Voir la [fiche « co-construction et mutualisation »](#) de l'AMUE.

initiatives et proposer une gouvernance numérique de l'ESR basée sur ces principes. Plus encore, l'ensemble des logiciels administratifs (gestion, comptabilité, édition) et leur « composante métier » pourraient bénéficier d'une telle mutualisation.

Recommandation 8 : Mutualiser les moyens logiciels mis en œuvre dans le développement et la maintenance des infrastructures nécessaires à la publication des codes sources de l'ESR.

D | Valorisation et pratiques open source : de l'injonction contradictoire aux articulations fécondes

D.1 Un écart manifeste entre pratiques open source et modalités de valorisation

À la lecture du questionnaire, rares sont les cas où les services de valorisation ont accompagné la publication d'un code source : ainsi, 72 % des personnes ayant répondu n'ont pas été accompagnés et seulement 13 % témoignent d'un accompagnement.

D.1.1 Appréhensions et critiques envers les services de valorisation et maturation

Les commentaires ainsi que certains des entretiens montrent un décalage entre les pratiques d'ouverture du code source (avec l'apposition d'une licence) et l'accompagnement de la part des services de valorisation et des SATT⁷¹. Faire appel à ces services est vu comme un **frein à l'avancement du projet** et parfois même **synonyme de son arrêt**. Les processus sont en effet considérés comme particulièrement longs et non adaptés aux enjeux de collaboration qui caractérisent la recherche. Ainsi, le dépôt auprès de l'Agence pour la protection des programmes (dits « dépôts APP »)⁷² fait partie intégrante des « Déclarations Logicielles » et sera à la base de la conception d'un dossier qui devra ensuite être soutenu auprès des SATT (devant une commission dédiée) ou des services de valorisation. Plusieurs commentaires du questionnaire et retours des interviews soulignent que les délais de réponse et traitement sont souvent lents, voire inexistant, amenant les producteurs du code source à faire sans, de peur de se voir retarder.

Par ailleurs, l'open source est souvent décidé et géré seul par le projet – les questions spécifiques autour du choix des licences restent également sans préconisations précises. Ainsi, comme le souligne un chercheur interrogé le choix des licences se fait souvent par les équipes des projets par mimétisme et pour des raisons pragmatiques. Du point de vue de ce chercheur⁷³ contribuant par exemple au développement du langage de programmation Python, le choix des licences se fait par rapport aux usages déjà existants pour faciliter la collaboration et la création d'une dynamique de contribution et de maintien du logiciel utilisé :

L a philosophie est de rester compatible dans votre écosystème. Le contenu de la licence ne compte peu. Le plus important c'est de dégager les chercheurs de l'obligation d'y penser. Tout ce qui est en python ça part en BSD.

71 Des injonctions contradictoires sont relatées comme le souligne ce commentaire en réponse au questionnaire : « Entre les SATT qui ne savent à peine ce qu'est le logiciel libre (et qui ne voient aucun intérêt à accompagner une diffusion sous forme de logiciels libres, car elles ne connaissent pas les modes de valorisation économiques des LL) et les financeurs qui demandent une diffusion des codes sources, on fait le grand-écart. »

72 Une fiche sur l'Agence de protection des programmes et le registre inter dépôt digital number a été produite <https://www.projet-plume.org/ressource/app-et-iddn>

73 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

Réalisé ainsi sans bénéficier d'une implication du service de valorisation ou juridique de l'organisme de rattachement, ce choix méconnaît très souvent les considérations juridiques sous-jacentes et leurs articulations avec des éléments économiques et de valorisation. Dans les rares situations où sont prévus et proposés des processus d'accompagnement à l'open source (incluant le choix de licence), ceux-ci sont considérés comme trop longs et trop peu agiles pour répondre aux besoins des chercheurs – voire trop chers lorsque le projet doit procéder à ses frais à un audit de codes open source. De ce fait, les considérations de qualité juridique et de valorisation économique qui accompagnent généralement ces démarches sont souvent absentes ou partiellement traitées, générant un risque pour le projet et pour l'organisme sous la responsabilité duquel le projet est diffusé.

Points clefs questionnaire :

- Le questionnaire met en évidence que les choix juridiques et économiques sont actuellement réalisés sans impliquer les cellules et acteurs de la valorisation. Par exemple, la majorité des projets ont été publiés avec l'apposition d'une licence (82 %).
- La majorité des personnes ayant répondu au questionnaire ont su indiquer le type de licence, ce qui est représentatif du public du questionnaire ayant une culture de l'open source forte⁷⁴.
- En termes de choix de licence, l'on retrouve à la fois des licences dites « avec obligation de réciprocité⁷⁵ » et des licences « permissives ».
- Elles ne figurent pas toutes dans la [liste](#), fixée par décret, de licences pouvant être utilisées par les administrations pour la réutilisation à titre gratuit de leurs informations publiques⁷⁶.

De la même façon que les démarches open source s'opposent souvent aux logiques de valorisation et sécurisation des organismes, les logiques de financement internes et externes semblent aussi mal s'articuler avec les logiques de diffusion open source des projets de la recherche. Ainsi, le choix de l'open source peut constituer un véritable frein lorsque le projet souhaite solliciter un accompagnement par l'institut ou une SATT en termes de valorisation ou maturation du projet. Plus encore, en cas de refus de maturation, le logiciel est parfois publié en open source « par défaut », sans réelle stratégie ou accompagnement. Ainsi, le bénéfice d'un accompagnement induit automatiquement le retour de la recherche d'une voie de « retour sur investissements » par le biais de transferts de propriété intellectuelle, ce qui s'oppose à une logique d'open source. Ainsi, les réponses apportées par les services de valorisation internes ou les SATT sont jugées non adaptées aux spécificités de l'open source. Les chercheurs sont soumis à des injonctions contradictoires entre valorisation financière et diffusion open source. Comme le souligne une personne interrogée⁷⁷ :

L e circuit proposé aujourd'hui pour les logiciels par les services de valorisation sont adaptés à des logiciels à vocation commerciale et non pas à des logiciels open source, ce qui dissuade n'importe qui de le faire.

74 Ce qui contraste avec les projets référencés sur code.etalab.gouv, majoritairement sans licence.

75 Aussi dite « copyleft », ces licences imposent que toute modification apportée au logiciel initial soit reversée selon les mêmes termes, créant de ce fait une sorte de « commun » partagé entre les différents utilisateurs-contributeurs du projet.

76 Article D. 323-2-1 du code des relations entre le public et l'administration (CRPA).

77 Entretien avec Antoine Blanchard, responsable du bureau impulsion, service Innovation. Université de Bordeaux.

Pour la plupart des personnes ayant répondu au questionnaire ou entendues lors d'entretiens, les services de valorisation et de l'innovation freinent ainsi l'open source pour des **raisons commerciales au détriment des bénéfices que pourrait en tirer le projet** :

[Il y a] beaucoup d'injonctions contradictoires entre la direction scientifique, qui favorise la diffusion libre, et la direction de l'innovation, qui pense que si le source est propriétaire ça se vendra mieux.

Le manque d'articulation entre l'ouverture des codes sources et des modalités de valorisation adaptée à l'open source est un élément majeur souligné dans le questionnaire, mais également par un ensemble d'entretiens menés avec des juristes et responsables de service de valorisation. Cette absence d'articulation résulte d'un côté de la grande liberté dont disposent les chercheurs quant à la diffusion de leur projet en open source, et d'un autre côté par l'absence de ligne directrice claire à ce sujet en provenance des services de valorisation et des SATT. Le schéma économique de ces services étant complexifié dans le cas d'un recours à l'open source, qui peut amener pour ces structures à des impasses de rentabilité. Par ailleurs, peu de politiques institutionnelles sont mises en œuvre aujourd'hui pour soutenir l'open source.

Recommandation 9 : Identifier les procédures actuellement imposées pour la valorisation en termes de dépôt et publication et les articuler avec les bénéfices apportés par une démarche de publication par défaut de code source au sein de l'ESR.

Recommandation 10 : S'appuyer sur les services de valorisation des établissements de l'ESR pour contribuer aux objectifs de non-multiplication des licences applicables aux logiciels.

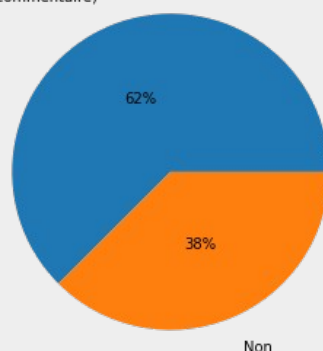
Recommandation 11 : Réaliser une étude complémentaire venant préciser les processus de diffusion open source dans l'ESR et mettant en perspective l'intervention des services de valorisation et organismes de maturation.

D.1.2 Des politiques institutionnelles open source encore bien discrètes

La majorité des répondants (61 %) n'a pas connaissance de politiques de publication des codes sources au sein de leur établissement. Concernant les établissements, on note la plupart du temps des positionnements informels ancrés dans une culture d'un laboratoire (principes éthiques et philosophiques) sans pour autant de politique institutionnelle validée.

Points clefs questionnaire :

Oui (précisez en commentaire)



Non

Figure 3: Avez-vous connaissance d'une politique de publication de codes sources pour votre établissement ?

La majorité des répondants n'ont pas connaissance d'une politique de publication des codes sources au sein de leur institution, mais les personnes qui en ont semblent indiquer deux tendances institutionnelles : l'une ouverte aux logiciels libres, l'autre poussant à l'utilisation de licences permissives souvent jugées problématiques dans les commentaires.

Recommandation 12 : Formaliser des politiques de valorisation au sein des établissements de l'ESR en s'appuyant sur la politique de contribution aux logiciels libres de l'État et en étendant le plan national pour la science ouverte aux enjeux de l'open source.

Recommandation 13 : Informer les personnes travaillant ou étudiant au sein de l'ESR des politiques de diffusion de codes sources au sein de leur établissement ainsi que des recommandations nationales.

Quelques instituts néanmoins, ont mis en œuvre une politique de développement informatique et un positionnement concernant l'open source. Sans vocation à être exhaustif, voici quelques politiques

remontées par le biais du questionnaire et des entretiens en lien avec différents établissements français⁷⁸.

	Existence d'une politique institutionnelle	Pratiques constatées au sein d'un institut, d'un laboratoire, d'une équipe
CEA	Non connue	Il s'agirait d'un dépôt à l'APP et une autorisation de publication de code.
CNES (Centre National d'Etudes Spatiales)	Non connue	Gouvernance s'appuyant sur la CBL (Commission Brevets & Logiciels) pour définir et valider les choix et stratégies de licences open source
CNRS	Feuille de route du CNRS pour la science ouverte du 18 novembre 2019	Codes sources et logiciels sont mentionnés dans la section 2/ données de la recherche ⁷⁹ .
INRIA	Non connue	Plusieurs outils et initiatives. Voir notamment https://www.allistene.fr/cvstene/ et https://www.inria.fr/fr/inriasoft-pour-la-diffusion-des-logiciels-open-source
INRAE	Note depuis 2017 concernant notamment des conseils pour le choix des licences](https://www6.inrae.fr/datapartage/content/download/3544/36521/version/1/file/Note_choix_licence_logiciel_decret2017_VF.pdf)	Cette note s'appuie sur la loi République Numérique et l'inclusion des codes sources comme document administratif elle oblige le dépôt des codes sources (sur SourceSup) si elle ne rentre pas dans le cas d'exceptions. Le choix est fait de respecter la compatibilité avec des composants tiers et si le développement est fait de zéro de privilégier une licence permissive.
Telecom Paris, Laboratoire COSI	Non connu	Création en 2018 du Centre pour l'innovation en matière de logiciels libres au sein du Laboratoire Traitement et Communication de l'Information (LTCI)..
BRGM	Non connu	De multiples projets structurants (not. De plateformes) en open source. Voir par exemple https://www.brgm.fr/projet/collaviz-plate-forme-open-source-traitement-collaboratif-multi-domaines-distance
ABES	L'Association des Bibliothèques d'Enseignement supérieur propose une politique de développement informatique complète et détaillée dont une partie concerne la gestion du code source .	La mention de l'open source est clairement définie avec un développement par défaut en open source et une publication sur GitHub dont le but est de faciliter la contribution d'autres acteurs.
Medialab	Non connu	À propos du site : En accord avec notre volonté d'ouverture et de transparence, les productions du laboratoire sont par ailleurs dans leur grande majorité diffusées librement, c'est-à-dire sous conditions d'Open Access pour les publications scientifiques (voir le médialab sur HAL et

78 Cette étude se limite aux politiques institutionnelles françaises. À l'international, d'autres politiques sont développées et nécessiteraient d'être détaillées et comparées (cf. F Pour aller plus loin). Pour donner un exemple, l'Université d'Oxford s'est doté d'une initiative « Reproducible Research Oxford » traitant notamment la question des usages des logiciels libres et open source. Une politique concernant les licences a ainsi été mise en œuvre <https://researchsupport.admin.ox.ac.uk/>.

79 « Extrait de la feuille de route du CNRS pour la science ouverte (novembre 2018) « Les données (données brutes, textes et documents, codes sources et logiciels) produites par les chercheurs et les chercheuses CNRS ou avec des moyens mis en œuvre par le CNRS doivent être, dans la mesure du possible, rendues accessibles et ré-utilisables selon les principes FAIR pour une consolidation des connaissances essentielle au développement d'une science plus efficiente. « Les données doivent être aussi ouvertes que possibles, et fermées autant que nécessaire ». https://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaqueette_ScienceOuverte.pdf

	sur SPIRE) et en logiciels libres open source pour les développements informatiques (voir le compte GitHub du laboratoire). Dans cette dynamique, le médialab soutient également le développement de logiciels créés par la communauté open source.
--	---

Mais de telles politiques nécessitent d'être construites et coordonnées avec l'ensemble des acteurs politiques impliqués au sein de l'ESR, en concertation avec les autres administrations. Cela implique et permet également de penser la valorisation de manière plus large : à l'échelle de la société, de l'administration, de l'ESR ou encore de certaines recherches disciplines scientifiques. Une telle vision élargie permet d'intégrer les dynamiques propres de l'ESR en tant que service public, de mesurer son impact pour la société et l'environnement, d'identifier les spécificités et la diversité d'instituts et de pratiques qui composent la recherche.

D.2 Conjuguer open source et valorisation : éléments communs et spécifiques au sein de l'ESR

D.2.1 Penser la valorisation dans une définition plus large et plurielle

Aujourd'hui, l'évaluation de l'ESR et sa « valorisation » (entendu par valorisation financière se basant sur un ensemble d'indicateurs majoritairement liés à la propriété intellectuelle et industrielle (brevet, etc.) et au « transfert » possible avec les milieux industriels). Outre un modèle de valorisation orienté essentiellement sur les Sciences Techniques et Médicales (STM), sans prendre en considération les spécificités des Sciences Humaines et Sociales (SHS), plusieurs personnes interrogées soulignent que ces indicateurs n'apporteraient pas une véritable plus-value financière (redevance). Les brevets ou licences serviraient plutôt des critères d'évaluation d'une recherche performante.

Ainsi, avant même de réfléchir à la valorisation possible par l'open source, l'ensemble des réponses montrent la nécessité de repenser la notion même de valorisation. Une des personnes interrogées chargées de l'innovation au sein d'une université notait ainsi⁸⁰ :

Pour les instituts, plus que de parler de valorisation la notion d'impact semble plus adaptée sachant que la valeur créée ne se résume pas seulement à une valeur financière, mais également d'action sur des enjeux telle que la transition écologique, la durabilité, le dynamisme des universités, etc.

Au cours des entretiens et par l'apport de missions précédemment menées auprès d'instituts de recherche et université, plusieurs éléments de valorisation clefs ressortent afin de pouvoir conjuguer la dynamique des communautés de l'ESR produisant et ouvrant leur code source et une valorisation

80 Entretien avec Antoine Blanchard, responsable du bureau impulsion, service Innovation. Université de Bordeaux.

respectueuse de ses principes. Pour cela, il s'agit de bien comprendre les éléments de valeurs possibles et leurs synergies.

D.2.2 Des éléments de valeurs communs aux acteurs de l'ESR adaptés à des modèles ouverts

Plusieurs éléments de valorisation ressortent des échanges menés avec les instituts de recherche⁸¹ :

- **l'usage** : il est généralement considéré par les instituts de l'ESR que l'usage fait d'un logiciel a plus d'importance que les redevances qui en sont tirées. Un logiciel est valorisé dès lors qu'il est utilisé, quel que soit le domaine ou le type de personne (startup, acteurs locaux ou nationaux). Le nombre de personnes et le type d'utilisation sont autant d'indices à prendre en compte. La réutilisation des recherches par un autre laboratoire est une valorisation évidente) ;
- **la mutualisation et les économies d'échelle** : la mutualisation avec d'autres acteurs peut-être génératrice d'économies parfois plus intéressantes que de potentielles licences ;
- **l'indépendance technologique** : le développement de technologies open source par les centres de recherche contribue aussi à une volonté commune au monde de la recherche d'éviter toute appropriation par un seul acteur dans les domaines clefs de la société (retours perçus dans des domaines comme l'optique, le traitement d'image, la mécanique spatiale, la télécommunication, etc.) ;
- **le rayonnement** : valorisation, reconnaissance à l'extérieur du savoir-faire, bénéfique en termes d'image. Plusieurs projets bénéficient ainsi d'une retombée médiatique (notamment en lorsqu'ils participent à l'information des citoyens) par les analyses et visualisations qu'ils produisent sur des sujets d'actualités et de société.⁸²

Au-delà des bénéfices directs parfois inexistant pour l'institut de recherche ou projet diffuseur, la publication et l'ouverture d'un projet peuvent reposer sur le bénéfice que cette ouverture est susceptible de générer pour d'autres acteurs (motivation extrinsèque). En plus des valeurs habituellement évaluées (diffusion et usage, développement de l'activité économique, soutien à la recherche et rayonnement scientifique, mutualisation externe, facilitation de partenariats et collaborations, services associés, rayonnement et marque), plusieurs valeurs non évaluables sont à considérer (innovation et agilité, transparence, qualité et mutualisation interne) tout autant que de des problématiques prenant un poids majeur aujourd'hui associées aux objectifs de développement durable, de constitution de territoire résilient⁸³.

81 Voir notamment les éléments développés dans le rapport pour le CNES par le cabinet inno3.

82 L'outil open source GarganText a ainsi été employé dans divers projets de recherche en lien avec l'actualité (élection présidentielle 2017 avec le projet politoscope <https://politoscope.org/> ou encore cartographie des recherches en lien avec la covid-19 <https://iscpif.fr/chavalarias/?p=1712>). L'IRD a développé également le projet COMOKIT sur la base de l'outil open source GAMA.

83 À ce propos, l'initiative ACT Campus (PIA3 2020) au sein des campus de l'Université de Bordeaux vise « à développer, tester, valider et diffuser de nouvelles façons d'aborder les grands problèmes environnementaux, sociaux et de transition économique ». Voir : <https://www.u-bordeaux.fr/Universite/Strategie/Initiatives-transverses/ACT-Campus-experimental>

Tous ces critères sont complexes à mesurer pour l'open source. Ils permettent *in fine* d'identifier les situations où les ordres de grandeur de revenus de redevances seraient moindres que les externalités impliquées par une diffusion sous licence libre. Plus encore, ils amènent à prolonger la politique publique au-delà de la seule recherche, de sorte à s'assurer que le rôle économique du secteur public tient compte aussi de cette valorisation par l'open source.

Recommandation 14 : Évaluer les bénéfices pour l'ESR d'une politique d'acquisition ou de subvention publiques favorisant l'open source, à l'échelle de l'ESR et au-delà.

D.2.3 Les éléments de valorisation spécifiques aux différents acteurs de l'ESR et de leurs missions

L'ESR est composé d'une multitude d'acteurs ayant des missions et finalités variées. Les statuts des différents établissements amènent également de prendre en considération les spécificités de chacun dans des logiques de valorisation possible.

Concernant les instituts de recherche par exemple tous n'ont pas nécessairement les mêmes missions et n'impliquant la production de savoirs du même ordre. Le « transfert de connaissances » en fonction des domaines diffère largement du tissu d'acteurs socio-économique propre. Ceci explique par exemple les différentes politiques de valorisation de la recherche existantes lorsque l'on s'approche des logiques d'industrialisation des travaux de la recherche. Cet écart entre les différentes « fonctions » occupées par les centres de recherche n'est pas toujours parfaitement considéré par les ministères de tutelle qui tendent à n'appliquer que des cadres communs.

Ainsi et sans surprise, la recherche fondamentale est beaucoup plus partagée que la recherche qui produit des résultats avec un haut niveau de maturité technologique (aussi dit TRL pour *technology readiness level*). Cependant, certains centres de recherche associés à des établissements publics (tels que le CNRM, Unité Mixte de Recherche sous la tutelle de Météo-France et du CNRS), opèrent cette fonction d'ouverture vers l'extérieur et de relation avec tous les grands organismes de recherche et laboratoires au niveau national et international.

En cela, la recherche opère un rôle de facilitateur complémentaire à aux logiques de concurrence susceptible d'exister par ailleurs. Inversement, le rapport de chaque institut avec l'industrie est ainsi particulièrement structurant sur les choix de valorisation :

- **en termes de licences associées au projet.** L'usage de licence permissive, facilitant la reprise du logiciel au sein de produits commerciaux « fermés », est généralement l'une des incitations fortes en prévenance des industriels souhaitant réutiliser un projet open source de l'ESR⁸⁴ ;
- **en termes de gouvernance.** À noter que le contrôle de l'évolution du projet par les équipes de recherche repose généralement de la dynamique du projet : tant que les équipes restent impliquées dans l'évolution et la maintenance du code source, elles restent généralement décideuses de cette évolution. À l'inverse, l'implication d'autres acteurs, notamment acteurs de la recherche, doit être accompagnée et anticipée afin qu'elles ne viennent pas déséquilibrer et briser ce lien initial entre l'équipe et son projet.

Les enjeux de « soutien à la filière » convergent naturellement vers l'open source : ce dernier posant un cadre commun à tous les acteurs économiques, favorisant ainsi la libre concurrence entre eux⁸⁵.

84 Le passage à une licence plus permissive (Cecill vers Apache-2.0) a ainsi été demandé par les industriels dans le cadre d'OrféoToolbox.

85 Ainsi, en proposant une alternative de qualité plus efficace économiquement (OTB, CelestLab), le CNES remplit sa mission de soutien aux utilisateurs d'observation de la terre.

La situation est d'autant plus complexe lorsque les établissements cumulent les rôles. Par exemple, le CEREMA est particulièrement enclin à partager ses projets de recherche aboutissant à du logiciel (ce qui représente 10 % de son activité), mais la question est plus complexe concernant les autres productions issues du CEREMA, ne relevant pas de la recherche, mais dont l'essentiel est disponible en *open data*⁸⁶.

L'ensemble des réponses recueillies montre ainsi la nécessité de repenser la notion même de valorisation pour prendre en considération les spécificités de l'open source et plus largement de modèles basés sur l'ouverture et la circulation de ressources (open access, open data, etc.). Cela implique

- une définition de la valeur élargie, ne prenant pas seulement en considération la valeur financière, mais d'autres types de valeur (impact social, rayonnement et visibilité, soutien organisationnel, etc.). Ces éléments sont plus difficilement quantifiables et des indicateurs éventuels doivent veiller à ne pas s'imposer au détriment des logiques d'écosystème associées ;
- cela s'appuie sur la mise en œuvre de modèles économiques qui prennent en considération les spécificités des modèles ouverts et des dynamiques vertueuses des écosystèmes qui peuvent être distinctes au vu de la diversité d'acteurs et de missions présentes au sein de l'ESR ;
- par ailleurs cela se construit par un appui mutuel de la science ouverte et des initiatives à l'œuvre concernant les infrastructures numériques.

Pour y parvenir, plusieurs éléments sont envisageables mêlant à la fois un soutien politique et national cohérent pour référencer et valoriser les initiatives existantes et ainsi favoriser des mutualisations possibles (échelles macro, méso) inter-établissements tout autant que de renforcer les dynamiques communautaires des projets open source en les accompagnant et en les soutenant (personnes-ressources, infrastructures dédiées à la publication de code source, etc.).

Recommandation 15 : Définir une stratégie open source différenciée pour la recherche en fonction du niveau de maturité technologique des projets.

86 Entretien avec Pascal Berteaud, directeur général du CEREMA.

E | Stratégie nationale : soutien à la publication et au développement de projets logiciels au sein de l'ESR

E.1 Inscrire le développement et le maintien de projets logiciels et open source au sein des politiques institutionnelles et de l'évaluation des chercheurs

E.1.1 Lever les freins à la publication des codes sources : qualité du code, compétitivité, manque de valorisation

La question des freins à la publication de code source a été abordée au sein du questionnaire. Le manque de temps (30 %) est une raison majeure mentionnée. Néanmoins, d'autres raisons ont été notées se rattachant d'une part des modalités d'évaluation actuelle des chercheurs et des projets de recherche (compétitivité, etc.) et, d'autre part, du manque de soutien de la part des institutions vis-à-vis des initiatives d'ouverture de code source.

Points clefs questionnaire :

Dans les raisons pour la non-publication, la plus plébiscitée est celle du manque de temps. Dans les commentaires, les répondants évoquent beaucoup de sentiments improductifs, voire négatifs en lien à la culture de recherche actuelle. Les idées les plus récurrentes concernent la peur de l'évaluation des pairs en rapport avec la qualité/propreté du code d'une part et la rivalité entre équipes de recherche. Le manque d'incitation politique institutionnelle et de valorisation pour les chercheurs de l'ouverture des codes sources est évoqué.

Plusieurs remarques en lien avec la non-publication du code source rejoignent les réticences mentionnées à d'autres initiatives associées à la science ouverte (mise à disposition des données, publication en open access). Il s'agit à la fois de la peur de la critique et du jugement sur la qualité du code jugée trop faible, la reconnaissance par les pairs étant clef dans l'évaluation des chercheurs. Plusieurs personnes en commentaire du questionnaire notent qu'elles ne souhaitent pas exposer du code qui n'est pas « propre » ou avaient peur des répercussions négatives associées. La peur de publier les codes est aussi associée à des enjeux de compétitivité et une rivalité entre équipes et recherche comme le résumant ces commentaires.

« Volonté de nettoyer le code avant. » « Peur d'être jugé sur la beauté du code. » « L'égo et l'évaluation des pairs en somme. » « Peur de la visibilité et des retours négatifs, impact sur l'image de l'organisme/service (j'ai connu ça il y a quelques années). »

Recommandation 16 : Identifier les freins personnels rencontrés face à la publication de codes, spécifiques ou non à l'ESR.

E.1.2 Un renforcement et une coordination des initiatives de politiques publiques déjà en cours

Le manque d'incitation de stratégie de la part des institutions a été aussi relevé avec des leviers de valorisation absents pour les personnes produisant ces codes sources au regard du risque encouru pour leur carrière ou éventuelle promotion. En résumé, alors même qu'une motivation intrinsèque est forte au sein des personnes participant à l'ouverture des codes sources, le système institutionnel en place ne crée pas les incitations nécessaires permettant de potentialiser ces initiatives.

Plusieurs besoins ont ainsi été remontés et dessinent le rôle clef de politique institutionnelle en soutien aux initiatives open source existant et aux communautés qui y participent. Ces actions peuvent s'appuyer sur les politiques déjà à l'œuvre au sein du Ministère de L'Enseignement supérieur de la Recherche et de l'Innovation (MESRI). Le Comité pour la science ouverte et un nouveau comité pour les infrastructures numériques (en cours de constitution) sont des instances clefs (pilotage, groupe de travail) pour la mise en place de politiques cohérentes au sein des différences institutions de l'ESR. Plus que des enjeux de science ouverte, il s'agit avant tout d'aider à la coordination d'acteurs de l'ESR en transition avec de nouveaux enjeux numériques (modèles économiques, infrastructures à déployer, adoption de nouvelles pratiques, etc.).

Recommandation 17 : Confier au CoSO, via son Collège logiciels libres, une action de coordination des politiques de publication des codes sources au sein de l'ESR.

E.1.3 Un référencement des initiatives existantes et la construction raisonnée d'indicateurs

L'un des problèmes identifiés est celui de la bonne identification des logiciels produits. Au-delà des difficultés que cela présente en termes d'accompagnement et de valorisation, l'absence de ces projets est préjudiciable dans l'évaluation et le pilotage de la recherche.

S'appuyant sur l'initiative de l'INRIA⁸⁷, une BIL généralisée simplifierait grandement l'action des ONR permettrait une plus grande mutualisation. Néanmoins la construction des indicateurs doit se faire avec prudence car plus que la diffusion d'un code source ce qui fait la réussite d'un projet sont les dynamiques de communautés sous-jacentes qui ne se mesurent que difficilement. Comme l'a montré le domaine de la scientométrie et des indicateurs des publications scientifiques, le champ des « métriques » peut amener à des mésusages et des résultats contre-productifs par une

87 Pour l'identification des logiciels et la valorisation des contributions des chercheurs, Inria s'est doté d'un inventaire nommé « BIL » : base d'information sur les logiciels : <https://bil.inria.fr/>.

quantification excessive. Il est donc important dans le référencement de ces projets de réfléchir à une évaluation qui laisse aussi la part au qualitatif⁸⁸.

Recommandation 18 : Réaliser une étude spécifique sur l'évaluation de la valorisation des logiciels dans l'ESR proposant des critères adaptés aux pratiques de l'open source.

Recommandation 19 : Établir un baromètre de l'open source dans l'ESR qui s'articule ou complète le baromètre de la science ouverte.

E.2 Le besoin d'une dynamique de mutualisation et de contribution soutenue au sein de l'ESR

Le référencement des projets peut avoir pour effet une meilleure visibilité et ainsi faciliter une mutualisation des initiatives en cours. En effet, ce point est essentiel, car l'open source afin de bénéficier de ces dynamiques ne se limite pas à la publication du code source, mais à la mise en œuvre de logique de contribution entre plusieurs acteurs. Une des personnes interrogées et ayant participé au développement d'une plateforme open source de l'enseignement supérieur notait par exemple l'importance de réussir à créer une dynamique non plus seulement de service, mais de contribution.

Les bénéfices attendus à l'échelle d'un projet ou d'une organisation ne sont qu'une partie d'un bénéfice beaucoup plus important susceptible d'être réalisé à une échelle globale. Une des personnes interrogées⁸⁹ notait en ce sens le besoin d'intégrer également comme coût d'amortissement la maintenance afin de pouvoir continuer à développer les logiciels open source. Face au numérique, bien souvent une structure seule ne peut pas couvrir une somme de coûts qui augmenterait continuellement. Les logiciels open source, dès lors qu'ils sont suffisamment génériques pour être transposables d'un usage à un autre, d'une administration à une autre, peuvent permettre une réelle mutualisation et répartition des coûts afin de les faire tomber vers zéro.

Ainsi, différents types de mutualisation peuvent être envisagés :

- une **mutualisation entre les centres de recherche**, selon des modalités à définir (au travers de consortiums, de structures de mutualisation ou de gouvernance complètement ouverte). Des projets tels que Scilab, OrféoToolbox ou MMG3D sont ainsi des exemples de mutualisation importants (avec le CEREMA, la Maison de la télédétection ou encore l'IRSTEA). De même le projet Regards du CNES a été diffusé en open source compte tenu

88 À ce propos : Moulin, Camille, et Célya Gruson-Daniel. « Metrics and valuation of open source projects : A crossover between open source and open science ». Présenté à CHAOSScon2020, Bruxelles, 31 janvier 2020 ainsi que l'article <https://inno3.fr/actualite/conjuguer-open-source-et-science-ouverte-opportunités-et-leviers-d'action>

89 Entretien avec Laurent David, ancien développeur au sein du GIP FUN (2014-2017) et actuellement président de CALL Learning.

des économies d'échelle qu'une diffusion de ce type pouvait permettre à l'échelle de nombreux laboratoires⁹⁰. Au-delà, l'open source peut favoriser une mutualisation dans le cadre d'une recherche internationale, comme ce fut le cas pour le projet SWOT (*Surface Water and Ocean Topography*, réalisé en collaboration étroite par la France et les USA) ou le projet GAMA (entre la France, l'Asie et les USA) ;

- une **mutualisation avec les autres acteurs publics**. C'est notamment le cas d'EDF, qui reste très proche de la recherche universitaire et participe à l'industrialisation de certains projets. Plusieurs agences s'appuient également sur les résultats de la recherche pour mener à bien leurs propres missions⁹¹ ;
- une **mutualisation globale**, sous réserve d'une vraie pratique d'industrialisation pour intéresser les acteurs industriels et réduire l'incertitude relative à la pérennité du projet. L'ouverture du code favorise aussi des collaborations, notamment publiques-privées, sans cela. Le constat est que les échanges sont plus subtils et complexes avec des opérateurs privés que pour le monde académique au sein duquel les recherches sont généralement ouvertes et partagées. Les industriels imposent en effet une confidentialité dans les échanges, ainsi que la prise en considération d'intérêts sociaux économiques.

Une telle vision politique et démarche de mutualisation est ainsi évoquée comme alternative viable face à la domination des »big tech « (GAFAM ou BATX).

Recommandation 20 : Ajouter la réutilisation et la contribution à des logiciels provenant d'autres établissements de l'ESR comme critère d'évaluation vis-à-vis de leurs tutelles afin de favoriser les mutualisations.

Recommandation 21 : Définir une stratégie open source de l'ESR français bénéficiant d'un cadre d'évaluation et de pilotage national.

E.3 Services externes : économie d'échelle et réduction de coût

Deux sources de services externes ont été mentionnées dans le cadre de la valorisation de projets logiciels comme étant possiblement partiellement remplaçables dans le cadre d'une démarche de publication de codes sources :

90 En ce sens, le CNES forme régulièrement à l'usage de CelestLab – par exemple au profit de laboratoires ou universités qui développent des nano-sat.

91 Voir par exemple l'usage réalisé par l'ONF du logiciel open source OTB : « Suivi de l'occupation des sols et de la déforestation : nouvelle publication de notre équipe télédétection », <https://www.onfinternational.org/suivi-de-l'occupation-des-sols-et-de-la-deforestation-nouvelle-publication-de-notre-equipe-teledetection/>

- l'**audit de code** nécessaire à la mise en conformité d'un logiciel de recherche à l'égard des licences des composants tiers utilisés (d'autant plus nécessaire que le projet est exploité industriellement). Les licences des logiciels d'audit sont particulièrement élevées (d'une dizaine à une vingtaine de milliers d'euros par projet) et représentent un coût conséquent à l'échelle nationale. L'accompagnement des projets de recherche dans le cadre d'une publication de codes sources permettrait de distiller un certain nombre de bonnes pratiques notamment d'apprentissage de développement de projets collaboratifs (production de code de qualité, standardisation, test) tout autant que de renforcer la qualité juridique des projets. Au-delà, une mutualisation pourrait aussi être envisagée en s'appuyant sur une solution d'audit de code open source mutualisée entre les établissements⁹² ;
- les « **dépôts logiciels** » réalisés auprès de l'Agence pour la Protection des Programmes (APP) à de multiples reprises lors de la conception, du transfert, de la valorisation et de la maturation d'un projet logiciel. Cet enregistrement permet d'identifier précisément l'objet de la maturation et d'obtenir un numéro unique d'enregistrement auquel renverra le contrat avec le partenaire (une SATT en cas de maturation, tout autre partenaire industriel en cas de valorisation directement réalisée par l'institut de recherche). Chaque dépôt étant payant, il serait opportun d'évaluer le coût complet de ce service tiers, pour dans un second temps évaluer les démarches alternatives. Dans ce contexte, la publication généralisée des codes sources pourrait générer une économie importante pour les établissements tout en répondant aux enjeux de preuve d'antériorité et de traçabilité.

Recommandation 22 : Identifier les économies susceptibles d'être générées par une démarche accompagnée de publication par principe des codes sources.

E.4 Besoin humain : une réévaluation des plans de carrière et une pérennisation des postes au sein de l'ESR dédiés au développement et à la maintenance

Le questionnaire et les entretiens menés avec des chercheurs, ingénieurs de recherche et responsables de productions logiciels dans des GIP ont permis de faire remonter différents besoins pour favoriser la publication et l'ouverture des codes sources. La problématique des ressources humaines allouées au développement de projet logiciel ou bien de leur maintenance a été abordée à plusieurs reprises. Cela concerne aussi bien des laboratoires de recherche qui subissent le manque de poste fixe pour pouvoir assurer le maintien et le suivi des projets tout autant que le recrutement de développeurs dans des structures telles que les GIP. Une des personnes interrogées⁹³ mentionnait que, dans le champ de l'enseignement supérieur, le nombre d'étudiants croît en même temps que

92 À noter à ce sujet l'échec du projet Antelink, financé par des fonds européens et de l'INRIA, dans une logique de concurrencer les éditeurs américains sur un modèle similaire de licence commerciale. En l'absence de logique de mutualisation, la solution n'a pas obtenu l'adoption et le succès commercial souhaité. Voir également le projet QualiPSo et la proposition d'un cadre de traçabilité des droits de propriétés intellectuelles (Intellectual Property Rights Tracking IPRT) https://fossbazaar.org/system/files/methodologie_IPR_A1.D2.1.4-1.pdf

93 Entretien avec David Rongeat, responsable numérique AMUE et Bertrand Mocquet, expert numérique AMUE.

celui des demandes de nouvelles fonctionnalités numériques – mais les ressources n’augmentent pas. Les réponses relatent ainsi des moyens insuffisants pour prendre soin du développement, mais aussi de la maintenance des infrastructures, des codes sources.

Les instituts de recherche devraient promouvoir activement la reproductibilité des recherches effectuées par leurs agents. Pour cela, les agents faisant l’effort de le faire pourraient être plus soutenus que les autres, par exemple en financement récurrent. De plus, les directions des instituts pourraient avoir une véritable stratégie de développement, en favorisant le partage de code à partir des initiatives individuelles des agents.

Points clefs questionnaire :

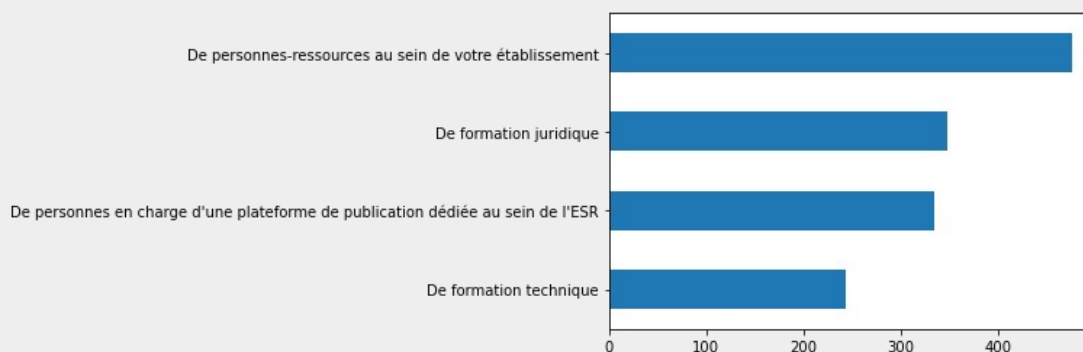


Figure 4: Pour publier des codes sources, de quoi auriez-vous besoin ?

Les répondants auraient besoin de plus d’accompagnement au sein de leur institution, et moins comme un service public généralisé à travers l’ESR.

Au-delà des postes d’ingénierie de recherche, d’étude ou d’enseignement-recherche, il est important de noter que les entretiens et les commentaires au questionnaire mettent en avant un ensemble d’autres personnes souvent contractuelles de l’ESR (post-doctorat, vacataire)⁹⁴ ainsi que des personnes indépendantes travaillant avec l’ESR (formation, développement en statut d’indépendant) contribuant au développement de projets open source. Plusieurs entretiens ont par exemple souligné la précarité des contrats (CDD) amenant au départ de personnes contribuant à des projets de développement logiciel. La publication du code source dans certains cas a été le seul moyen de pouvoir continuer des projets de développement en cours qui s’arrêterait à la fin du contrat comme le note la personne interrogée⁹⁵ :

La question, critique, était celle du devenir du projet une fois l’ingénieur en charge parti. À la fois parce qu’il voulait continuer à contribuer au

94 Une liste des personnes pouvant contribuer au développement logiciel est précisé dans la fiche PLUME <https://projet-plume.org/fr/ressource/diffuser-logiciel-recomm-juridiques-admin> Elle distingue les personnes en régime salarié soit en poste permanent tel que les enseignants-chercheurs, les ingénieurs soit avec un statut contractuel tels que les post-doctorants, les ATER, etc. et les personnes n’ayant pas la qualité d’employé (professeur émérite, retraité, stagiaire, boursier, etc.)

95 Entretien avec Antoine Blanchard, responsable du bureau impulsion, service Innovation. Université de Bordeaux.

développement, et parce que ses compétences allaient manquer à l'université pour envisager toute exploitation. La solution était donc une licence libre..

Néanmoins, l'open source ne permet pas d'asseoir ces projets de manière durable, car il s'appuie sur un flou entre contribution personnelle et professionnelle. En effet, dans le cadre des développements open source, les personnes contribuant à ces projets le font souvent de façon bénévole et mêlent à la fois le développement à titre de mission professionnelle, mais aussi à titre de loisir comme le souligne un des ingénieurs de recherche interrogé⁹⁶..

***L**a frontière entre la contribution et les missions personnelles est floue. Il y a une très forte publication pour des projets institutionnels qui sont personnels.*

Valoriser la production open source et les fonctions de maintenance des productions logicielles dans les fiches de poste et l'évolution de carrière est une manière de pérenniser ces projets. Un des commentaires du questionnaire mentionne ainsi :

J'aimerais que la carrière de développeur soit mieux valorisée au sein du CNRS et pas bloquée par l'obligation de devenir chef de projet ou de service pour pouvoir monter en grade. À cause de cette politique, il y a peu de développeurs expérimentés et donc la qualité du code n'est pas bonne et du coup les gens ne publient pas leur code par peur des critiques ou par peur de leaker des informations confidentielles (password, adresse email, adresse ip).

L'open source peut devenir également un élément d'attractivité pour des développeurs face à une concurrence forte du privé (meilleur salaire, condition de travail) : comme le souligne une des personnes interrogées « avec l'open source, les gens vont retrouver pourquoi ils sont dans l'ESR. »

Recommandation 23 : Recruter et pérenniser les postes de fonctions support et personnes-ressources mobilisables sur le développement et la maintenance de projets open source.

Recommandation 24 : Sensibiliser les communautés de recherche sur les compétences nécessaires à mobiliser pour pérenniser un projet open source.

Recommandation 25 : Apporter un cadre aux contributions apportées au sein des projets logiciels qui dépassent les missions de recherche et d'enseignement (notamment contributions personnelles réalisées en nom propre).

96 Entretien avec Guillaume Plique, ingénieur de recherche au Medialab (SciencePo Paris).

Recommandation 26 : Faire valoir le développement de code source dans les critères de recrutement et d'évaluation des professionnels de la recherche, et développer des mesures incitatives pour faciliter le développement de code source de qualité.

E.5 Besoin d'infrastructures et de formation pour de bonnes pratiques d'ouverture du code source

E.5.1 Infrastructure nécessaire à la publication de code de l'ESR

Une des questions du questionnaire portait sur le lieu de publication des codes sources et les modalités de leur partage (public, privé, compte d'un projet ou personnel). La plupart des projets proposés ont été publiés publiquement sur un compte d'organisation dédié au projet (71 %) ou sur un compte personnel (43 %)⁹⁷. La plateforme GitHub est essentiellement utilisée pour le partage des codes sources (68,8 %).

Les réponses mettent également en évidence l'existence de nombreuses forges basées sur GitLab⁹⁸.

— L'utilisation de forge basée sur GitLab illustre en soi les logiques d'open source soit de faciliter la réutilisation d'un outil et de mutualiser le développement de nouvelles fonctionnalités sans avoir à développer de zéro un nouvel outil.

— En ce sens, il est important de rappeler que l'open source ne se limite pas à la publication de code source. Le succès du développement logiciel open source ne se mesure pas seulement à la capacité à produire un logiciel, mais surtout sur sa capacité à créer une communauté permettant de maintenir le projet, et de l'adapter à de nouveaux usages apparaissant au quotidien.

D'autres forges sont aussi employées (25 %), notamment la forge de Renater [SourceSup](#). Cette plateforme web de gestion de projets à destination de l'Enseignement supérieur et des laboratoires de recherche française développé par Renater, a été citée à plusieurs reprises dans les commentaires du questionnaire et mentionnée au cours d'entretiens. Plusieurs critiques lui sont adressées comme un usage difficile pour un public de « non développeur », là où GitHub permet de partager aisément du code source sans connaissance poussée en informatique. Une autre raison est le « cloisonnement » qu'imposent ces plateformes. Si ces forges ont l'avantage de garantir des critères de sécurité attendus au sein de l'ESR, elles rendent difficiles les logiques de contribution souvent internationale en recherche. Une personne⁹⁹ mentionnait ainsi :

97 25 % seulement des projets ont été partagés en privé sur un compte dédié au projet et 11,3 % sur un compte personnel.

98 Les forges basées sur Gitlab mentionnées sont celles de l'INRIA, de l'IN2P3, d'HumaNum, de l'IRSTEA. la forge proposée par l'association framasoftware, framagit recueille également de nombreux projets. Des projets sont développés également directement sur Gitlab (20,3 %)

99 Entretien avec Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans).

*A*vant j'étais sur SourceSup qui était pas mal au départ. J'ai commencé les projets là-dessus, mais cela a été problématique pour donner accès aux collaborateurs non français. Ensuite c'était bureaucratique à l'extrême.

GitHub reste ainsi la plateforme avec une utilisation la plus massive avec plusieurs raisons citées : « visibilité, praticité et intégration avec d'autres outils (npm registry, docker hub, travis...) ». L'utilisation de cette plateforme suscite néanmoins des inquiétudes (rachat par Microsoft). Comme le note un commentaire :

*A*u final on a le choix entre utiliser des plates-formes commerciales fiables et visibles, ou se battre contre beaucoup de lourdeurs administratives pour des outils bricolés avec les moyens du bord.

La visibilité et l'effet de réseau qu'offre cette solution centralisée et commerciale rendant peut-être à terme captif ses utilisateurs.

Si des personnes plébiscitent l'existence d'une « forge nationale de qualité basée sur Gitlab », plusieurs commentaires nuancent l'existence d'une solution centralisée nationale. Ces personnes notent que la centralisation irait à l'encontre d'une culture en recherche largement décentralisée à l'image des critiques faites à des plateformes telles que HAL (atteinte à l'indépendance des chercheurs, renforcement des mesures d'évaluation, etc.). Par ailleurs, la proposition d'une plateforme centralisée (une alternative nationale à un GitHub) risquerait de ne pas être adoptée pour des raisons pratiques comme le souligne ce commentaire :

*J*e ne sais pas s'il faudrait un GitHub français souverain pour des raisons pratiques on veut pas se retrouver dans une solution de devoir pusher sur différentes forges remote. Il faut une seule source de vérité pour le code...

À défaut, l'accès aux fonctionnalités d'une plateforme moderne, ou une modernisation de SourceSup conformément à ces standards, répondrait aux besoins des projets open source non ou insuffisamment outillés.

Une solution proposée qui rejoint code.etalab.gouv est le développement d'une vitrine de visibilité des codes produits moissonnés automatiquement avec la publication du code source sur une liste de forges recommandées. Cela permettrait d'intégrer les productions logicielles dans les éléments d'impact des universités et instituts.¹⁰⁰

Une plateforme de partage de code source doit ainsi à la fois disposer de fonctionnalités adaptées aux chercheurs et à des modes de travail décentralisé tout en assurant la visibilité des projets. Outre le partage du code source, ces plateformes doivent pouvoir s'intégrer (interopérabilité) avec un ensemble d'autres outils et d'initiatives à l'œuvre sur le référencement, l'indexation, l'attachement à des identifiants pérennes, le référencement dans des plateformes de type HAL¹⁰¹, l'archivage du code source nécessitant également un accompagnement à ces principes pour garantir la

100 Le CRI, Centre de Recherche Interdisciplinaires développe actuellement la plateforme *Projects* pour permettre de recenser l'ensemble des projets réalisés au CRI <https://projects.cri-paris.org/repo/5cf4c2b4bec2168a6b809557>. Le code source devrait être mis en ligne prochainement.

101 Roberto Di Cosmo, « Archiving and Referencing Source Code with Software Heritage », in *Mathematical Software – ICMS 2020*, éd. par Anna Maria Bigatti et al., Lecture Notes in Computer Science (Cham : Springer International Publishing, 2020), 362-73, https://doi.org/10.1007/978-3-030-52200-1_36.

pérennisation de projets¹⁰² ¹⁰³. Le travail sur la citation et le référencement est en cours, avec des solutions d'interfaçage qu'il s'agit de continuer.

Recommandation 27 : Moderniser les services offerts aux projets de l'ESR pour le développement et la publication de code source.

Recommandation 28 : Favoriser l'interopérabilité entre les différentes plateformes de développement et partage de code source à l'échelle de l'ESR et à l'échelle nationale.

Recommandation 29 : Évaluer les besoins de plateformes d'hébergement ou de mise en visibilité des codes sources de l'ESR et penser les conditions de leurs interfaçages avec les autres plateformes pré – existantes (Software Heritage, HAL, etc.).

102 La différence entre sauvegarde et archivage peu être tenue et mal distingué amenant à penser que la publication de code source sur une plateforme telle que GitHub est suffisante pour la pérennisation.

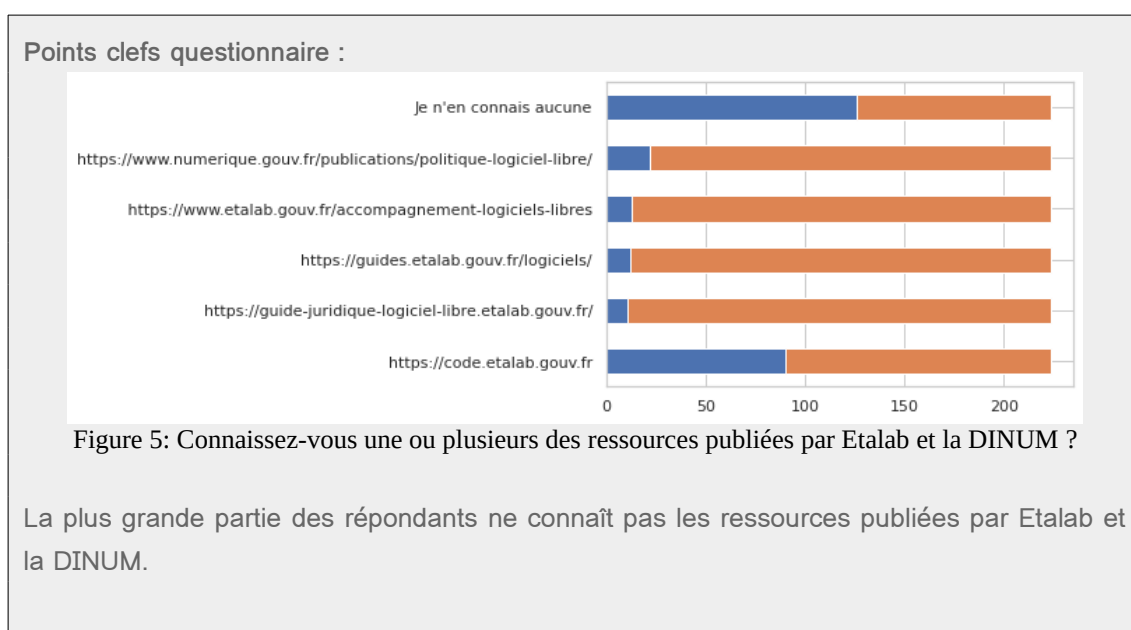
103 Un partenariat entre la DINUM et Software Heritage a été mis en place pour assurer l'archivage de tous les codes sources référencés sur code.etalab.gouv.fr.

E.5.2 Faciliter le développement de logiciels au code source pérenne et de qualité : documentation, archivage et référencement

Les besoins de formation ont été remontés par les personnes ayant répondu au questionnaire, à l'image de ce commentaire :

Il serait intéressant de disposer d'un accompagnement méthodologique et parfois technique pour aider à la diffusion des codes sources notamment dans les petits établissements comme le mien qui ne peuvent pas élaborer un tel accompagnement par manque de compétences et de moyens.

Autre point : la majorité des personnes ayant répondu au questionnaire ne connaissent pas les ressources mises à disposition par Etalab au sujet de la publication de code source.



Recommandation 30 : Faire travailler le CoSO et Etalab de façon plus étroite pour la valorisation de ce contenu et l'évaluation régulière de leur pertinence.

La documentation du code, l'apprentissage au développement collaboratif et à l'ingénierie logicielle (« au-delà du code prototype, plein de bugs, sans tests ni doc. ») ont été mentionnés. Des initiatives étrangères telles que le « Software Sustainability Institute », le « Reproducible Research Oxford » ou encore l'initiative « Essential Open Source for Science » de la fondation Chan Zuckerberg apparaissent comme des exemples à suivre¹⁰⁴. Le développement d'open badges¹⁰⁵ pour faciliter des dynamiques communautaires et la production de code de qualité est un point soulevé pour favoriser

104 Pour des exemples supplémentaires, consultez notamment Gomez-Diaz T and Recio T. On the evaluation of research software : the CDUR procedure, F1000Research 2019, 8:1353 (26 novembre 2019) [version 2 ; peer review : 2 approved], <https://doi.org/10.12688/f1000research.19994.2>

105 Voir notamment <https://openbadges.info/>

« une meilleure valorisation et ouvrant la voie à la création d'un écosystème dynamique. ». Certains commentaires proposent d'aller plus loin avec la proposition de labels de qualité. À l'image des plans de gestion de données (DMP) nécessaires désormais pour le financement de projets de recherche (ERC, ANR), un plan de gestion logiciel tel que le modèle PRESOFT (research software management plan¹⁰⁶¹⁰⁷) pourrait être également un outil employé dès la conception d'un projet logiciel.

Des différentes réponses au questionnaire, des entretiens réalisés tout autant que du recensement de projets open source, il en ressort une dynamique communautaire forte au sein des communautés de recherche de part des pratiques open source proches des valeurs soutenues au sein de l'Enseignement Supérieur et de la Recherche. Ces initiatives souvent développées par les communautés de recherche elles-mêmes gagneraient à être mises en visibilité pour favoriser des dynamiques de mutualisation et de méthodes ouvertes collaboratives. Ces dernières peuvent se construire et se renforcer non pas en opposition aux processus de « valorisation » mises en œuvre au sein de l'ESR mais en se combinant. Pour cela, une définition élargie de « la valeur » et une réflexion sur les indicateurs associés est une piste féconde pour à la fois souligner l'impact de tels projets (financiers, sociaux, environnementaux, etc.) tout en se basant sur le développement d'une évaluation centrée sur la vitalité de tels écosystèmes.

Messages clefs de l'article Conjuguer open source et science ouverte opportunités et leviers d'action¹⁰⁸

1. Valoriser par l'*open source* nécessite de prendre de la hauteur, d'opérer un changement de vision et de logique par rapport au mode de valorisation habituel. Évaluer le plein potentiel de ces projets implique de prendre en considération des dynamiques multi-échelles (vision macro dépassant le périmètre usuel de l'institut de recherche). L'*open source* tire sa force de dynamiques de contribution et mutualisation qu'il s'agit de faciliter et maintenir tout en trouvant des mesures de valorisation adaptées aux logiques écosystémiques.
2. L'*open source* répond au besoin de repenser les relations privé/public en les insérant dans un réseau d'acteurs aux priorités différentes (recherche, économique, etc.) coopérant pour la réussite du projet. Le succès escompté implique de mettre en place des règles de gouvernance soutenues par un cadre contractuel discuté et établi en amont du projet puis durant ses étapes de développement, ce qu'offrent pleinement les outils contractuels de l'*open source*. Mettre en place un tel cadre et partager ces « bonnes pratiques » apparaissent comme des catalyseurs afin de mieux articuler le rôle des parties prenantes dans différentes phases et potentialiser les

106 Teresa Gomez-Diaz et Genevieve Romier, « Research Software Management Plan template, V3.2 » (avril 2018), <https://hal.archives-ouvertes.fr/hal-01802565>.

107 Le modèle PRESOFT pour réaliser des plans de gestion des logiciels de la recherche est aussi disponible sur le service DMP OPIDoR de l'INIST du CNRS, qui propose également la réalisation des plans de gestion des données suivant différents modèles. <https://dmp.opidor.fr/>

108 Gruson-Daniel Célya et Benjamin Jean, « Conjuguer open source et science ouverte : opportunités et leviers d'action », Inno³, 2019, <https://inno3.fr/actualite/conjuguer-open-source-et-science-ouverte-opportunités-et-leviers-daction>.

bénéfices propres et communs à chaque secteur (instituts de recherche, entreprises partenaires, etc.)

3. **La valorisation par l'*open source* se base sur le développement d'une évaluation centrée sur la vitalité d'un écosystème.** Aux mesures usuelles économiques et technologiques, s'ajoutent des éléments de valeur ancrés dans des dynamiques contributives saines. Tenir compte de valeurs spécifiques aux établissements de recherche et aux communautés amène à évaluer des usages, complexes par nature et différemment quantifiables. Valoriser par l'*open source* nécessite une attention toute particulière au développement d'indicateurs et de métriques se construisant avec les communautés et acteurs concernés. Il s'agit en premier lieu de comprendre les raisons des usages existants et de les faire évoluer, si besoin est, vers d'autres pratiques plus adaptées et pertinentes pour la valorisation tout autant que pour la communauté en elle-même qui gagne en réflexivité sur ces propres usages.
4. **Les politiques publiques pour une science ouverte constituent un tremplin pour l'adoption de l'*open source* au sein des instituts de recherche.** Il semble néanmoins que les opportunités croisées entre ces deux domaines ne soient pas encore exploitées à leur juste mesure. Les axes principaux du plan national de la science ouverte mettent en priorité les questions de l'accès aux publications et aux données. Si les logiciels libres et *open source* font partie d'une thématique d'un groupe de travail du CoSO en qualité d'« outil », « objet de recherche » ou « résultat au service de la recherche » (qualité, reproductibilité, avancée scientifique), il semble aussi que la généralisation de l'*open source* participe à l'émergence nécessaire des infrastructures (ouvertes) de la science ouverte (édition, partage des données). Penser la science ouverte à l'aune des dynamiques open source en s'appuyant sur le renversement de logique et de culture que les logiciels libres et *open source* matérialisent est aussi un moyen de questionner les modalités économiques et de gouvernance de la production des savoirs de la recherche publique. Toutes et tous, acteurs de la recherche ou industriels, acteurs privés ou publics, organisations et individus, gagneront à l'émergence et à la gouvernance partagée d'une telle infrastructure ouverte produite par et pour la recherche. Un chantier encore à développer...

F | Pour aller plus loin

Cette étude réalisée sur un temps court sur la base des réponses à un questionnaire, d'entretiens et de ressources bibliographiques est un premier jalon qui nécessiterait d'être approfondi sur une période plus longue. Plusieurs projets pourraient être menés dans le champ des STS (*Science & Technologies Studies*), des *Software Studies* en favorisant des groupes de travail interdisciplinaires et regroupant différentes professions de l'ESR (recherche et enseignement, information et documentation, soutien aux infrastructures, ingénierie de recherche, etc.).

Plus particulièrement, plusieurs thématiques clefs pourraient faire l'objet de travaux dédiés :

- étude sur l'évolution des pratiques open source au sein de l'ESR (niveau de familiarité, évolution des outils employés, connaissances juridiques) et leviers d'appropriation (formation proposée, accompagnement, etc.) en prenant en considération les différentes communautés épistémiques et de pratiques ;
- étude sur les démarches de recensement des projets et des contributions open source en articulation avec des modalités d'évaluation « vertueuse » pour les communautés les développant ;
- étude sur le développement et la maintenance d'infrastructures numériques de la recherche mutualisées et des dynamiques associées (modèle économique, gouvernance) en s'appuyant sur les travaux en cours sur les communs au sein de l'ESR (édition numérique) ou en dehors de ce secteur ;
- étude sur les modalités de valorisation par l'open source et les évolutions des modèles économiques des structures de valorisation de l'ESR ;
- étude sur l'évaluation de la valorisation par l'open source à une échelle nationale ou européenne ;
- analyse comparative de pratiques et politiques de soutien aux logiciels libres et à l'open source dans différents pays (Europe, UK, États-Unis, Canada, Asie, etc.) en prenant en considération les spécificités législatives, administratives et économiques de chacun.

G | Synthèse des recommandations

Recommandation 1 : Reconnaître le rôle et la place de l'open source dans l'enseignement supérieur et la recherche française et situer la France dans une perspective internationale.

Recommandation 2 : Améliorer le recensement actuellement existant au sein de l'ESR des projets open source afin de construire une politique publique représentative de ces usages.

Recommandation 3 : Corréler davantage les processus que mettent en place les établissements de l'ESR pour piloter leur activité avec les bénéfices et avantages que retirent les projets.

Recommandation 4 : Réaliser une étude pour évaluer la pertinence de l'emploi des licences françaises CeCILL et de la licence EUPL dans l'Enseignement Supérieur et de la Recherche et formuler des règles explicites quant à leur usage.

Recommandation 5 : Adopter une approche différenciée des bénéfices de l'open source en fonction des finalités de production des logiciels (codes scientifiques isolés ou production logicielle).

Recommandation 6 : Sensibiliser au partage des codes sources (et de jeux de données) étant associés aux résultats de recherche comme critères d'évaluation pour la publication dans une revue scientifique.

Recommandation 7 : Mettre à disposition des personnels de la recherche des formations, bonnes pratiques et outils en matière de publication de code source et les accompagner dans cette appropriation.

Recommandation 8 : Mutualiser les moyens logiciels mis en œuvre dans le développement et la maintenance des infrastructures nécessaires à la publication des codes sources de l'ESR.

Recommandation 9 : Identifier les procédures actuellement imposées pour la valorisation en termes de dépôt et publication et les articuler avec les bénéfices apportés par une démarche de publication par défaut de code source au sein de l'ESR.

Recommandation 10 : S'appuyer sur les services de valorisation des établissements de l'ESR pour contribuer aux objectifs de non-multiplication des licences applicables aux logiciels.

Recommandation 11 : Réaliser une étude complémentaire venant préciser les processus de diffusion open source dans l'ESR et mettant en perspective l'intervention des services de valorisation et organismes de maturation.

Recommandation 12 : Formaliser des politiques de valorisation au sein des établissements de l'ESR en s'appuyant sur la politique de contribution aux logiciels libres de l'État et en étendant le plan national pour la science ouverte aux enjeux de l'open source.

Recommandation 13 : Informer les personnes travaillant ou étudiant au sein de l'ESR des politiques de diffusion de codes sources au sein de leur établissement ainsi que des recommandations nationales.

Recommandation 14 : Évaluer les bénéfices pour l'ESR d'une politique d'acquisition ou de subvention publiques favorisant l'open source, à l'échelle de l'ESR et au-delà.

Recommandation 15 : Définir une stratégie open source différenciée pour la recherche en fonction du niveau de maturité technologique des projets.

Recommandation 16 : Identifier les freins personnels rencontrés face à la publication de codes, spécifiques ou non à l'ESR.

Recommandation 17 : Confier au CoSO, via son Collège logiciels libres, une action de coordination des politiques de publication des codes sources au sein de l'ESR.

Recommandation 18 : Réaliser une étude spécifique sur l'évaluation de la valorisation des logiciels dans l'ESR proposant des critères adaptés aux pratiques de l'open source.

Recommandation 19 : Établir un baromètre de l'open source dans l'ESR qui s'articule ou complète le baromètre de la science ouverte.

Recommandation 20 : Ajouter la réutilisation et la contribution à des logiciels provenant d'autres établissements de l'ESR comme critère d'évaluation vis-à-vis de leurs tutelles afin de favoriser les mutualisations.

Recommandation 21 : Définir une stratégie open source de l'ESR français bénéficiant d'un cadre d'évaluation et de pilotage national.

Recommandation 22 : Identifier les économies susceptibles d'être générées par une démarche accompagnée de publication par principe des codes sources.

Recommandation 23 : Recruter et pérenniser les postes de fonctions support et personnes-ressources mobilisables sur le développement et la maintenance de projets open source.

Recommandation 24 : Sensibiliser les communautés de recherche sur les compétences nécessaires à mobiliser pour pérenniser un projet open source.

Recommandation 25 : Apporter un cadre aux contributions apportées au sein des projets logiciels qui dépassent les missions de recherche et d'enseignement (notamment contributions personnelles réalisées en nom propre).

Recommandation 26 : Faire valoir le développement de code source dans les critères de recrutement et d'évaluation des professionnels de la recherche, et développer des mesures incitatives pour faciliter le développement de code source de qualité.

Recommandation 27 : Moderniser les services offerts aux projets de l'ESR pour le développement et la publication de code source.

Recommandation 28 : Favoriser l'interopérabilité entre les différentes plateformes de développement et partage de code source à l'échelle de l'ESR et à l'échelle nationale.

Recommandation 29 : Évaluer les besoins de plateformes d'hébergement ou de mise en visibilité des codes sources de l'ESR et penser les conditions de leurs interfaçages avec les autres plateformes pré – existantes (Software Heritage, HAL, etc.).

Recommandation 30 : Faire travailler le CoSO et Etalab de façon plus étroite pour la valorisation de ce contenu et l'évaluation régulière de leur pertinence.

H | Annexes

H.1 Présentation du questionnaire et entretiens réalisés

Le questionnaire vise à mieux comprendre les pratiques de publication des codes sources au sein de l'ESR et à définir les freins et les besoins associés à leur partage. Il était composé de 33 questions, dont 15 optionnelles (recensement d'un projet).

Le questionnaire a été diffusé sur une période très courte (du 3 au 20 septembre) et a pu recueillir dans ce laps de temps plus de **223 réponses complètes et 457 réponses incomplètes** (c'est à dire des réponses au questionnaire interrompues avant soumission finale). L'analyse se porte sur les réponses complètes, c'est-à-dire soumises, mais en intégrant néanmoins des informations supplémentaires issues des réponses dites incomplètes.

L'analyse du questionnaire a été complétée par la tenue d'entretiens avec différentes personnes impliquées dans la publication de codes sources et de profils variés (voir en annexe la liste des personnes interviewées).

La partie optionnelle de recensement de projets a été complétée par 24 personnes et pourra alimenter dans un second temps l'initiative <http://code.etalab.gouv.fr>.

H.2 Démarche de science ouverte et licences

Le questionnaire a été réalisé via limesurvey. Les données complètes ont été exportées en csv et anonymisées (suppression des colonnes dont les informations permettent de retrouver les personnes ayant répondu au questionnaire). L'analyse des données et l'obtention des graphiques ont été réalisés avec le langage de programmation Python. Un Jupyter Notebook est mis à disposition sur framagit <https://framagit.org/inno3/open-source-esr>.

Le rapport est accessible en ligne sur le site <https://opensource-esr.pubpub.org/>

Le rapport est publié en [licence CC-BY](#) et [licence Ouverte 2.0](#), les données et les scripts sont publiés en licence Ouverte 2.0.

H.3 Listes de logiciels open source et ESR

H.3.1 Sélection de projets recensés (Section O) et non-répertoriés sur code.etalab.gouv.fr

Le questionnaire a permis de recenser 24 projets open source publiés ou qui auraient vocation à l'être.

Parmi ces projets, on compte 9 organismes de l'ESR n'ayant pas encore ouvert de compte sur code.etalab.gouv.fr pour y déposer les codes sources des logiciels libres qu'ils produisent.

Nom du groupe	Plateforme	Dépôts url	Nombre de dépôts	Info url	Description
GNU Data Language / Observatoire de Paris et CNRS	github	https://github.com/gnudatalanguage	4	-	GDL is a domain-specific programming language and a data analysis environment.
Group of neuroanatomy	github	https://github.com/neuroanatomy	30	http://neuroanatomy.github.io	Neuroanatomy applied and theoretical ProWhy est un logiciel libre et gratuit, support des processus de résolution de problèmes en entreprise.
École Nationale d'Ingénieurs de Tarbes (ENIT)	Prowhy.org	http://www.prowhy.org/redmine/projects/prowhy/repository	1	http://www.prowhy.org/redmine/projects/prowhy	A multilingual Natural Language Processing (NLP) suite
LIMA / CEA LIST	github	https://github.com/aymara	23	https://aymara.github.io/lima/	CNVmap (logiciel libre sous forme d'un paquet R disponible à tous) cartographier des variants structuraux à partir de données servant à la construction de cartes génétiques
CNVMap / INRAE	sourcesup	https://sourcesup.renater.fr/projects/cnvmap	?	http://moulon.inrae.fr/news/2020/01/cnvmap-ou-le-recyclage-de-donn%C3%A9es-pour-localiser-les-duplications-dans-les-g%C3%A9nomes/	hackinscience.org source code repository
hackinscience	framagit	https://framagit.org/hackinscience	3	https://www.hackinscience.org/about/	
Brian INSERM, Sorbonne, Imperial College London	github	https://github.com/brian-team	23	https://briansimulator.org/	Brian is a free, open source simulator for spiking neural networks.
LMGC90 CNRS- Université Montpellier	gitlab	https://git-xen.lmgc.univ-montp2.fr/lmgc90/lmgc90_user	1	https://hal.archives-ouvertes.fr/hal-01899254/document	Open platform dedicated to the modeling of large collections of interacting objects (2D/3D).
Hacketafac	sourcesup	https://	1	http://hacketafac.u-	Thème WordPress

Nom du groupe	Plateforme	Dépôts url	Nombre de dépôts	Info url	Description
Université de Bordeaux		sourcesup.renater.fr/projects/hacketafac		bordeaux.fr/	développé pour l'organisation du concours d'innovation étudiant de l'université de Bordeaux « Hacketafac - Transforme ton campus », sur la base du thème Hackathon conçu par Phuse

Bien que le code du Scikit-learn-MOOC soit déjà répertorié sur code.etalab.gouv par l'INRIA, le code source de Scikit-learn ne s'y trouve pas encore :

Scikit-learn / INRIA	github	https://github.com/scikit-learn	9	https://scikit-learn.org/stable/	machine learning in Python
----------------------	--------	---------------------------------	---	----------------------------------	----------------------------

Plusieurs dépôts de code recensés par le questionnaire sont hébergés sur des comptes personnels.

code.etalab.gouv.fr étant destiné au référencement des comptes institutionnels, ces projets ne peuvent être répertoriés malgré leur rattachement à l'ESR. Cela montre une des difficultés pour le suivi des productions logicielles de l'ESR.

Ci-dessous le tableau de 4 projets concernés :

Nom du projet/orga	Plateforme	Dépôts url	Nombre de dépôts	Info url	description
FedilabTube / Fedilab	Framagit	https://framagit.org/tom79/fedilab-tube	2	-	This project groups two different apps. TubeAcad, a Peertube Android app for French academic authorities. All is in French. Its use is limited to some instances. The other app is TubeLab a Peertube Android app working for all instances.
PANDORAE / médialab Sciences Po	github	https://github.com/Guillaume-Levrier/PANDORAE	1	https://guillaume-levrier.github.io/PANDORAE/	A data retrieval & exploration protocol designed to investigate science and policy processes
pecebl / CNRS	github	https://github.com/	1	-	eBeam Lithography simulation and

Nom du projet/orga	Plateforme	Dépôts url	Nombre de dépôts	Info url	description
		looninho/ pecebl			Proximity Effect Correction
OnLine Analysis Form (OLAF) / Institut Charles Gerhardt Montpellier (ICGM) et Université de Montpellier	github	https:// gitlab.com/ fboyrie/olaf	-	https:// Imp.edu.umontpellie r.fr/2020/09/02/ demandes- danalyses/	Logiciel de traitement de demandes du laboratoire de mesures physiques de l'université de Montpellier

H.3.2 Sélection d'autres projets mentionnés (Section B) et non-répertoriés sur code.etalab.gouv.fr

Le questionnaire a permis de repérer d'autres projets open source développés au sein de l'ESR déjà publiés mais non-répertoriés sur code.etalab.gouv.fr même si les organismes de rattachements y sont déjà présents :

Nom du projet/orga	Plateforme	Dépôts url	Nombre de dépôts	Info url	description
pm2 / INRIA	INRIA forge	https:// gforge.inria.fr/ projects/pm2/	5	http:// pm2.gforge.inria.fr/	PM2 is a low level generic runtime system which integrates multithreading management (Marcel) and a high performance multi-cluster communication library (Madeleine).
MAS4Data / SMAC (Univ. Lille et INRIA)	github	https:// github.com/ cristal-smac/ mas4data	1	https://github.com/ cristal-smac	Multiagent systems for processing very large datasets
ScaIA / SMAC (Univ. Lille et INRIA)	github	https:// github.com/ cristal-smac/ ScaIA	1	https://github.com/ cristal-smac	library of algorithms which aim at forming coalitions of individuals around some activities
SageMath / UPEM	sagemath	https:// www.sagemat h.org/ download- source.html	1	https:// www.sagemath.org/ index.html	logiciel libre de mathématiques sous licence GPL. Il combine la puissance de nombreux programmes libres dans une interface commune basée sur le langage de programmation Python.
PLaTon / UPEM	github	https:// github.com/ PremierLanga ge	22	https:// premierlangage.gith ub.io/PLaTon-web/	PLaTon est une plateforme moderne d'exercices multi-matières avec auto-correction

Nom du projet/orga	Plateforme	Dépôts url	Nombre de dépôts	Info url	description
Gaspard2 / Univ. Lille et INRIA	INRIA forge	https://gforge.inria.fr/projects/gaspard2	-	-	Integrated Development Environment (IDE) for SoC visual co-modeling. It allows modeling, simulation and code generation of SoC applications and hardware architectures.
N2S3 / Univ. Lille	Sourcesup	https://sourcesup.renater.fr/projects/n2s3/	5	https://sourcesup.renater.fr/wiki/n2s3/start	Neural Network Scalable Spiking Simulator

Projets mentionnés en Section B développés au sein de l'ESR mais hébergés sur des comptes personnels :

Nom du projet/orga	Plateforme	Dépôts url	Nombre de dépôts	Info url	description
Ponio / Univ. Rennes 1	github	https://github.com/kivvix/miMas	1	https://pypi.org/project/ponio/	Python library for study of numerical integrators for solve linear transport equation
MetaShARK / MNHN	github	https://github.com/earnaud/MetaShARK-v2/	1	https://metashark.pndb.fr/	MetaShARK (Metadata Shiny Automated Resource & Knowledge) is a R shiny app allowing the user to get information about Ecological Metadata Language and to fill in metadata for datasets according to this standard.

Par ailleurs, le projet PLUME (20062013) fourni 406 fiches descriptives et validées de logiciels dont 95 logiciels issus de l'Enseignement Supérieur et de la Recherche.¹⁰⁹

H.3.3 Logiciels open source majeurs

Outils	Description
	SciKitLearn : Il s'agit d'une bibliothèque Python spécialisée dans le cas du machine learning. Python langage de programmation open source est employé largement dans le milieu académique et plus largement des data science, c'est-à-dire tout le site entrepreneurial de la donnée et de sa gestion. SciKitLearn est donc une librairie clef dans le développement des travaux d'IA.
	GarganText : Logiciel libre pour l'exploration interactive de grands corpus numériques. Il est développé à l'ISC-PIF par l'équipe "Digital Humanities" en partenariat avec plusieurs institutions et projets.
	Gephi : Logiciel pour visualiser, analyser et explorer en temps réel les graphes (aussi appelés réseaux ou données relationnelles) de tout type. Sorte de Photoshop pour les réseaux, l'utilisateur interagit avec la représentation graphique, manipule les structures, formes et couleurs pour en révéler les propriétés cachées via des saillances visuelles.
	OpenFLUID (Software environment for modelling Fluxes In Landscapes) : plate-forme

109 cf. https://projet-plume.org/logiciels_valides et https://projet-plume.org/fiches_logiciels_dev_internes

	<p>logicielle de simulation de flux dans le paysage. Elle permet la construction de modèles et l'exécution de simulations. Dédiée à la modélisation des flux dans les paysages complexes, OpenFLUID est basé sur une représentation topologique de l'espace sous forme d'unités spatiales connectées et une modélisation des processus basée sur un couplage de simulateurs.</p> <p>ORFEOToolBox : logiciel de traitement de données de télédétection impulsé par le CNES en 2002. Le projet a évolué avec un tournant important en 2006 concernant les modalités de gouvernance. Pour faire vivre le projet et faciliter l'enrichissement par des contributions externes du logiciel, les membres du projet ont noté l'intérêt de détailler les processus de prise de décision et de tenir au courant des choix émis dans un souci de transparence et d'équité entre l'ensemble des contributeurs. C'est ainsi qu'une gouvernance ouverte s'est développée avec la constitution d'un comité directeur au sein du projet, mais aussi le rattachement à la fondation OSGeo. Un tel rattachement s'est traduit par un mentoring, bénéfique à la professionnalisation du projet, et à sa diffusion accrue¹¹⁰.</p>
Infrastructures spécifiques à l'ESR	<p>OpendreamKit : Infrastructure de recherche qui vise à créer et renforcer des environnements de recherche virtuels. Développée dans un premier temps pour soutenir la recherche en mathématique, cette e-infrastructure peut s'adapter à de nombreux autres domaines de recherche. Ce projet coordonné par des scientifiques du Laboratoire de Recherche en Informatique - LRI (UPSud/CNRS) a été financé par l'Europe dans le cadre d'un programme H2020</p>
	<p>FUN-MOOC : Plateforme qui offre plus de 540 cours conçus par des professeurs d'universités et d'écoles françaises et par leurs partenaires académiques internationaux. Cette plateforme a été lancée par le Ministère de l'Enseignement supérieur et de la Recherche en octobre 2013, cette initiative vise à accompagner le développement des formations tirant pleinement profit du levier numérique et accessibles au plus grand nombre. Elle se base sur le code source de la plateforme nord-américaine EdX. FUN est aujourd'hui un GIP (groupement d'intérêt public). [à ajouter partie analyse de FUN les manquements, etc. entretien Laurent David]</p>
	<p>HumaNum : Très grande infrastructure de recherche (TGIR) visant à faciliter le tournant numérique de la recherche en sciences humaines et sociales. Elle met à disposition un ensemble de services pour le stockage, le traitement, l'exposition, le signalement, la diffusion et la conservation sur le long terme des données numériques de la recherche. HumaNum propose par exemple plusieurs services basés sur des instances open source (Gitlab, QGIS)</p>
	<p>ISIA (Information System for Infrastructure Administration of AnaEE- France) : dans le cadre d'AnaEE France, le CNRS a développé une plate-forme ISIA (Information System for Infrastructure Administration of AnaEE- France) diffusée publiquement sous une licence Libre et open source.</p>
Infrastructure générique	<p>Les compilateurs (not. gcc) & Git sont des composants essentiels de la recherche aujourd'hui. Disponibles pour tout le monde, ces outils sont un véritable levier de productivité. « Il est nécessaire de financer les projets d'infrastructure ». Si on veut baser une stratégie souveraine et pérenne autour des LL, il faut être capable d'identifier les projets qui ont un impact substantiel sur l'activité de recherche. (François Pellegrini)</p>
Autres	<p>Scenari Plateforme : Scenari est un environnement de conception open source de solutions éditoriales et documentaires métiers pour créer et structurer des contenus, les publier sur différents canaux et les exploiter. Il a été conçu au sein de l'UTC et est aujourd'hui maintenu au sein d'une structure associative soutenue par une SpinOff de l'UTC (Kelis).</p>
	<p>VidéoLAN (VLC) : développé initialement au sein du bureau des étudiants de l'École Centrale Paris</p>
	<p>CHOCO - Solveur de programmation par contraintes Choco (bibliothèque open source, librairie Java)</p>
	<p>GPAC - Plateforme open source de création, distribution et lecture de contenus multimédias interactifs <https://www.imt.fr/Item-portfolio/gpac-plateforme-open-source-de-creation-distribution-et-lecture-de-contenus-multimedias-interactifs/></p>
	<p>OpenAir - Plateforme de développement open source hardware/software sur les</p>

110 cf. [rapport et article Inno](#)³

H.4 Collecte d'information

L'étude s'appuie par ailleurs sur des informations collectées auprès de personnes provenant de différentes organisations attachées à l'enseignement supérieur et à la recherche : MESRI, CEREMA, INSHS, CNRS, INRAE, AMUE, Université de Bordeaux, Software Heritage,, Médialab (SciencesPo Paris), CNRM, CEA, URFIST, HackinScience, CNES, une SATT. s.

Listes des personnes interrogées ou ayant contribué à la réalisation du rapport

Voici la liste des personnes interrogées dont les propos ont été cités et ayant validé la mention de leur nom dans le rapport.

— Pascal Berteaud, directeur général du CEREMA.

— Antoine Blanchard, responsable du bureau impulsion, service Innovation. Université de Bordeaux.

— Laurent David, ancien développeur au sein du GIP FUN (France Université Numérique) de 2014-2017 et actuellement président de CALL Learning.

— Roberto Di Cosmo, professeur d'informatique, détaché chez l'INRIA, membre du laboratoire PPS (IRIF/Université de Paris) et directeur de l'initiative Software Heritage.

- Konrad Hinsén, directeur de recherche au CNRS (Centre de Biophysique Moléculaire, Orléans.

— Julien Palard, co-fondateur de Hackinscience, développeur et formateur indépendant Python.

— Guillaume Plique, ingénieur de recherche au Medialab (SciencePo Paris).

— David Rongeat, responsable numérique AMUE (Agence de Mutualisation des Universités et Établissements) et Bertarnd Mocquet, expert numérique AMUE.

Ce rapport a été relu et annoté par Bastien Guerry (Etalab/DINUM) en premier lieu. Par la suite, il a fait également l'objet d'une relecture attentive de Lionel Maurel, Roberto Di Cosmo, Teresa Gomez-Diaz et François Pellegrini que nous remercions pour leurs précieux commentaires.

H.5 Bibliographie

Baudin Véronique. « Pourquoi diffuser un logiciel développé dans un laboratoire ou une université avec une licence libre ? | Ressource PLUME ». Text. PLUME, 140909.

<https://projet-plume.org/ressource/pourquoi-diffuser-en-libre#1>.

Broca, Sébastien. *Utopie du logiciel libre*. Neuvy-en-Champagne : Le Passager clandestin, 2013.

Collège « logiciels libres et open source » (CoSO). « Note d'opportunité sur la valorisation des logiciels issus de la recherche ». Ouvrir la Science (CoSO), 2019.

<https://www.ouvrirlascience.fr/note-dopportunit e-sur-la-valorisation-des-logiciels-issus-de-la-recherche>.

Di Cosmo, Roberto. « Archiving and Referencing Source Code with Software Heritage ». In *Mathematical Software – ICMS 2020*,  dit e par Anna Maria Bigatti, Jacques Carette, James H. Davenport, Michael Joswig, et Timo de Wolff, 362-73. Lecture Notes in Computer Science. Cham : Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-52200-1_36.

Eghbal, Nadia. *Roads and Bridges : The Unseen Labor Behind Our Digital Infrastructure*. Ford Foundation, 2016.

Flichy, Patrice. « Internet ou la communaut e scientifique id eale ». *R eseaux* 17, n o 97 (1999) : 77-120. <https://doi.org/10.3406/reso.1999.2168>.

Gomez-Diaz, Teresa, et Tomas Recio. « On the Evaluation of Research Software : The CDUR Procedure ». *F1000Research* 8 (26 novembre 2019) : 1353. <https://doi.org/10.12688/f1000research.19994.2>.

Gomez-Diaz, Teresa, et Genevieve Romier. « Research Software Management Plan template, V3.2 », avril 2018. <https://hal.archives-ouvertes.fr/hal-01802565>.

Gruson-Daniel, C elya. « Chapitre 1 – Open : Les Diff erentes Facettes Du ‘Num erique’ ». In *Num erique et R egime Fran ais Des Savoirs En~action : L’open En Sciences. Le Cas de La Consultation R epublique Num erique (2015)*, 2018. <https://phd-cgd.pubpub.org/pub/facettes-numerique-fr>.

Gruson-Daniel C elya. « Num erique et r egime fran ais des savoirs en~action : l’open en sciences. Le cas de la consultation R epublique num erique (2015) ». Universit e Paris Descartes, 2018. <https://doi.org/10.5281/zenodo.1491292>.

Gruson-Daniel C elya et Benjamin Jean. « Conjuguer open source et science ouverte : opportunit es et leviers d’action ». Inno³, 2019. <https://inno3.fr/actualite/conjuguer-open-source-et-science-ouverte-opportunit es-et-leviers-daction>.

Hinsen, Konrad, et Nicolas P. Rougier. « ReScience ». In *Open science, transparence et  valuation. Perspectives et enjeux pour les chercheurs*. Bordeaux, France : URFIST Bordeaux, 2017. <https://hal.archives-ouvertes.fr/hal-01573262>.

Meyer, Morgan, et Susan Molyneux-Hodgson. « « Communaut es  pist miques » : une notion utile pour th oriser les collectifs en sciences ? » *Terrains travaux* n o 18, n o 1 (18 ao t 2011) : 141-54.

Neylon Cameron. « Principles for Open Scholarly Infrastructures », 2015. <https://cameronneylon.net/blog/principles-for-open-scholarly-infrastructures/>.

Perkel, Jeffrey M. « Challenge to Scientists : Does Your Ten-Year-Old Code Still Run ? » *Nature* 584, n o 7822 (24 ao t 2020) : 656-58. <https://doi.org/10.1038/d41586-020-02462-7>.

ReScience/ten-years. 2019. Reprint, ReScience organization, 2020. <https://github.com/ReScience/ten-years>.

W. Maxwell, John, Erik Hanson, Leena Desai, Carmen Tiampo, Kim O'Donnell, Avvai Ketheeswaran, Melody Sun, Emma Walter, et Ellen Michelle. *Mind the Gap : A Landscape Analysis of Open Source Publishing Tools and Platforms*. 1^{re} éd. PubPub, 2019. <https://doi.org/10.21428/6bc8b38c.2e2f6c3f>.

H.6 Suivi de version

Version	Date	Auteur	Commentaires
V1.1	28/09/2020	Gruson-Daniel Célya, Benjamin Jean , Camille Moulin, Maya Anderson-Conzales	Version envoyée à Etalab
V1.2	19/10/2020	Célya Gruson-Daniel	Rajout d'informations par Célya Gruson-Daniel
V1.3	02/11/2020	Benjamin Jean	Validation par Benjamin Jean et envoi aux interviewés
V1.4	01/12/2020	Célya Gruson-Daniel	Validation et reformulation si besoin des citations des personnes interrogées
V1.5	03/12/2020	Célya Gruson-Daniel	Intégration commentaires de Teresa Gomez-Diaz
V1.6	09/12/2020	Célya Gruson-Daniel	Intégration commentaires SATT ajout recommandation
V1.7	14/12/2020	Célya Gruson-Daniel	Intégration graphique
V 1.8	23/12/2020	Célya Gruson-Daniel et Benjamin Jean	Retravail recommandations pour intégration remarques
V1.9	23/12/2020	Célya Gruson-Daniel	Relecture finale avant publication
V2.0	23/12/2020	Célya Gruson-Daniel	Préparation publi markdown
V2.1	24/12/2020	Célya Gruson-Daniel et Benjamin Jean	Relecture finale
V2.2	26/01/2021	Célya Gruson-Daniel	Insertion commentaires et correction François Pellegrini. Ajout licences.