

# 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'ESR<sup>8</sup> 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, CEREMA<sup>9</sup>) ;
- 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 inno<sup>3</sup> 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
<b>Vous-même</b>	195	87.1%
<b>Votre équipe</b>	21	9.4%
<b>Votre établissement</b>	5	2.2%
<b>Autre</b>	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 hommes<sup>10</sup>.- 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 source<sup>11</sup>.

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 %)<sup>12</sup>.

### 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 chercheur<sup>13</sup> 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.

**A**ujourd'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 valorisation<sup>14</sup>.

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)<sup>15</sup>. Néanmoins, l'adoption de pratiques d'ouverture au sein de l'ESR<sup>16</sup> 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)<sup>17</sup> au sein d'un champ disciplinaire élargi (notamment en SHS<sup>18</sup>). 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 logiciels<sup>19</sup> 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)<sup>20</sup> ;
- 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-dopportunit%C3%A9-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 interrogée<sup>21</sup> 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 ».<sup>22</sup>

**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<sup>23</sup> 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 initiatives<sup>24</sup> 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 interrogée<sup>25</sup>, certains projets<sup>26</sup> 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 recherche<sup>27</sup>.

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.<sup>28</sup> L'INRIA a par exemple mis en place une base de données de ses logiciels (BIL)<sup>29</sup>, 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 PLUME<sup>30</sup> 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'ESR<sup>31</sup>.

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](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 développées 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](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 interrogées<sup>32</sup>.

*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.