

Research software : qui es-tu ?

Une conversation scientifique et juridique
sur les logiciels de la recherche

Teresa Gomez-Diaz

Université Gustave Eiffel-LIGM-CNRS

Tomas Recio

Université Antonio de Nebrija, Madrid, Espagne

Les logiciels de la recherche sont désormais reconnus comme l'un des piliers de la science ouverte, intégrés, par exemple, dans le *Deuxième plan national pour la science ouverte* en France et dans la *Recommandation de l'UNESCO sur une science ouverte*. Plusieurs équipes ont proposé des définitions de *logiciel de la recherche*, néanmoins ce concept a encore besoin d'être analysé comme le recommande un rapport récent de Science Europe¹.

La méthodologie. Ce travail sélectionne trois des formulations proposées récemment dans trois contextes différents et établit un cadre de comparaison qui aidera à clarifier ce concept et à consolider une formulation pour sa définition. Ces trois formulations incluent la définition proposée par les auteurs de ce texte, définition développée dans le cadre de la construction des protocoles CDUR (citation, dissemination, use, research) d'évaluation des logiciels de la recherche.

Les résultats. Il est vérifié que le cadre comparatif peut être correctement appliqué à ces trois définitions et que les critères utilisés dans ledit cadre permettent d'analyser avec précision les trois formulations sélectionnées.

Cette étude confirme la définition que nous avons proposée et consolide notre compréhension, acquise dans des travaux précédemment publiés, de

1. <https://scienceeurope.org>.

cet objet complexe, le *logiciel de la recherche*, en vue des analyses effectuées ici. Ceci contribue à mieux répondre aux défis qu'il soulève dans le contexte de la science ouverte et à faciliter la conception, la mise en place et l'adoption des politiques scientifiques dont il relève.

Introduction

La recherche a pour objet de produire et de diffuser des connaissances nouvelles dans tous les domaines dans lesquels la méthode scientifique est applicable. Pour produire ces nouvelles connaissances, il est souvent nécessaire de collecter, manipuler et analyser des ensembles de données, qui seront ensuite étudiés et manipulés avec des logiciels. Les articles, les présentations dans des conférences, et les publications en général sont les moyens le plus couramment utilisés pour diffuser ces connaissances nouvelles. Dans la plupart de nos laboratoires, trois objets sont les plus importants dans la production et la diffusion des nouvelles connaissances, et cela dans presque toutes les matières scientifiques : les articles (ou publications en général), les données et les logiciels.

Ces logiciels sont souvent produits par des équipes de recherche dans un contexte précis d'analyse d'un ensemble de données, ou bien de façon tout à fait indépendante, hors de ce contexte d'étude. La diffusion de ces données et de ces logiciels facilite la transmission de ces nouvelles connaissances, la transparence dans les processus de la recherche, la reproductibilité et la validation des nouveaux résultats scientifiques obtenus avec ces objets. Selon l'encyclopédie en ligne Wikipédia² :

un logiciel est un ensemble de séquences d'instructions interprétables par une machine et d'un jeu de données nécessaires à ces opérations. Le logiciel détermine donc les tâches qui peuvent être effectuées par la machine, ordonne son fonctionnement et lui procure ainsi son utilité fonctionnelle. Les séquences d'instructions, appelées programmes, ainsi que les données du logiciel sont ordinairement structurées en fichiers.

«Logiciel» est la traduction de *software*, ce qui a été officialisé par un arrêté en 1974 et confirmé en 1981 [68] (voir la section 3).

Le manuscrit présenté ici continue les études existantes sur la définition de *logiciel de la recherche*, en sélectionnant et analysant trois des définitions récemment proposées dans des contextes différents [39], [42], [51], [66], [71]. Parmi ces trois formulations nous avons inclus celle que nous avons développée dans le cadre de la construction des protocoles CDUR d'évaluation des logiciels de la recherche [39], [42]. Il est important de bien comprendre cet objet puisque les logiciels de la recherche sont désormais reconnus comme

2. <https://fr.wikipedia.org/wiki/Logiciel>.

l'un des piliers de la science ouverte [12], [40], [65], [78], et les universités et autres organismes de recherche ont besoin de préciser ce qui relève du champ d'application de leurs politiques scientifiques dans ce contexte [6], [40], [47], [76].

Les sections suivantes seront consacrées à l'examen de ces trois définitions en développant les avantages et les inconvénients de chacune. Pour cela nous avons construit un cadre de comparaison qui fournira la méthode d'analyse utilisée pour étudier le contexte d'élaboration de la définition et la qualité de la formulation proposée. Il sera également analysé si la définition reflète une compréhension correcte des logiciels de la recherche en tant que production scientifique, numérique ou en tant qu'objet juridique, ce qui sera complété avec deux points supplémentaires relatifs aux *conundrum challenges*³ [9], [46] et la comparaison avec une autre production scientifique importante, les données de la recherche [41], [42].

Voici les trois définitions qui seront analysées en suivant les critères proposés :

D1-COSO

Définition proposée par le collège Codes sources et logiciels du Comité pour la science ouverte (COSO) en 2022 [66], [71].

D2-RDA

Définition proposée par le groupe de réflexion FAIR for Research Software (FAIR4RS) WG de la Research Data Alliance (RDA) en 2021 [51].

D3-CDUR

Définition que nous avons proposée en 2019 dans le cadre de la construction des protocoles CDUR d'évaluation des logiciels de la recherche [39], [42].

Comme indiqué dans [11] en relation avec les données de la recherche :

*A clear conceptualization is essential for proper data management, dissemination, and reuse. Yet, ambiguity surrounding the term continues to generate inconsistencies in how research results are reported and how data policies are implemented*⁴.

3. Quel est le logiciel à partager? par qui? avec qui? dans quelles conditions? pourquoi et avec quels effets?

4. Dans toute la suite, nous associons systématiquement une traduction en français à chaque citation donnée dans son anglais d'origine. Nous ne le signalons explicitement que pour cette première citation en anglais, toutes les autres seront traitées de la même façon. « Une conceptualisation claire est essentielle pour une gestion, une diffusion et une réutilisation adéquates des données. Pourtant, l'ambiguïté qui entoure ce terme continue de générer des incohérences dans la manière dont les résultats de la recherche sont présentés et dont les politiques relatives aux données sont mises en œuvre ».

De manière analogue, la gestion des logiciels de la recherche doit se faire correctement, et cela en considérant les logiciels de plusieurs points de vue : en tant que productions scientifiques, en tant qu'objets juridiques, et sans négliger les aspects techniques qui sont importants pour définir les infrastructures et les services qui sont ou qui seront utilisés pour leur diffusion et leur réutilisation, ce qui montre le besoin de l'analyse approfondie présentée ici, et motive les critères sélectionnés pour réaliser les comparaisons.

Ainsi, les analyses effectuées seront sans doute utiles afin de détecter les défis qui se présentent pour fixer une définition et contribuer à mieux comprendre cette production scientifique dans le cadre de la science ouverte. Ils faciliteront, entre autres, la conception, la mise en place et l'adoption des politiques scientifiques qui sont pertinentes pour les logiciels de la recherche.

Le plan de ce texte est le suivant. Nous expliquerons en premier lieu le cadre de comparaison et les critères qui le composent. Ensuite, pour élaborer nos analyses de ces critères appliqués aux trois définitions sélectionnées, nous avons inclus deux sections dédiées au regroupement des informations sur les logiciels en tant que productions scientifiques et en tant qu'objets juridiques, suivies d'une section dédiée à la question de l'usage comme critère à retenir pour la formulation d'une définition. Ces sections sont suivies de l'analyse des trois définitions mentionnées ci-dessus et d'une discussion finale.

1. Le cadre de comparaison

Cette section est dédiée à expliquer le cadre de comparaison que nous allons utiliser pour analyser ces trois définitions de logiciel de la recherche, qui ont été proposées dans des contextes différents : D1-COSO [66], [71], D2-RDA [51], D3-CDUR [39], [42]. Il doit donc inclure, entre autres points, l'étude des contextes dans lesquels elles sont proposées, et doit montrer que cet objet complexe est bien compris dans les diverses dimensions qui le composent. Après avoir énuméré les points de ce cadre nous les détaillons et argumentons.

Voici donc les critères qui composent le cadre de comparaison que nous allons utiliser :

- cadre d'élaboration de la formulation, bibliographie consultée ;
- qualité de la formulation : s'agit-il d'une définition précise ?
- compréhension de l'objet en tant que production numérique ;
- compréhension de l'objet en tant que production scientifique ;
- positionnement de l'objet dans un contexte de droits d'auteur et de propriété intellectuelle ;
- capacité de répondre aux questions du *conundrum challenges* ;
- cadre de comparaison avec le concept de données de la recherche.

Les deux premiers points sont relatifs à la formulation de la définition et à la qualité de cette formulation : le cadre dans lequel elle est proposée, au sein de quelle équipe, et quels sont les travaux consultés.

En relation avec la qualité de la formulation d'une définition nous avons consulté [18], dont le chapitre 6 est dédié aux définitions et explique les caractéristiques d'une *bonne définition*. L'utilité d'une définition est de préciser ce qui est inclus ou non dans la définition d'un terme, c'est-à-dire de délimiter correctement une partie d'un ensemble : une bonne définition du terme X doit éviter toute imprécision et toute obscurité. L'objectif d'une définition est la clarté ; il ne devrait pas être plus difficile pour le public cible de comprendre la définition de X que de comprendre le mot X lui-même. Une bonne définition doit comporter plusieurs caractéristiques importantes. Par ailleurs, comme argumenté dans [18], il existe un lien étroit entre définition et classification, mais la classification ne fait pas partie de la formulation de la définition⁵.

Pour déterminer les caractéristiques les plus importantes, la formulation d'une définition doit aussi montrer que l'objet défini est bien compris dans les multiples dimensions qui le constituent. Dans notre cas de logiciel de la recherche, la première de ces dimensions correspond, évidemment, à l'objet numérique *logiciel*, c'est-à-dire, *comme l'ensemble de séquences d'instructions interprétables par une machine qui sont incluses dans un ensemble de fichiers*.

Du fait de l'importance de bien identifier ces fichiers, nous souhaitons pointer ici les règles de nommage de fichiers⁶, mais nous ne traiterons pas certaines questions relatives aux aspects de cette dimension numérique [1] comme le cycle de vie, l'organisation de l'équipe de développement, la gestion des différentes versions, les bonnes pratiques de test, de procédures d'installation, de diffusion, de sauvegarde et d'archivage, et beaucoup d'autres sujets qui échappent à l'objectif principal de ce texte, l'étude de trois définitions et la méthodologie construite pour réaliser les comparaisons entre ces définitions. Ces différents points relèvent de la prise de décision au sein de l'équipe de développement, ce qui peut être facilité, par exemple, avec les plans de gestion des logiciels [48], [50], [63].

L'utilisation de ces aides à la planification de la vie d'un logiciel montre combien il est important de bien identifier les logiciels avec des métadonnées

5. Par exemple, une définition de l'objet *chaise* n'inclut pas le fait qu'il y a des chaises vertes et de chaises roses...

6. Voir, par exemple, la ressource *Stockage et archivage : Comment bien nommer ses fichiers ?* de DoRANum (https://doranum.fr/stockage-archivage/comment-nommer-fichiers_10_13143_wgqw-aa59/) ou la ressource *File Naming Conventions* de Harvard (<https://datamanagement.hms.harvard.edu/plan-design/file-naming-conventions>).

au moment de leur diffusion, de favoriser des pratiques de citation⁷ [3], [35], [39]. Elle montre également que l'utilisation d'une forge peut faciliter la gestion des versions, le développement collaboratif ou la diffusion d'un logiciel. À noter que la mise en œuvre de ces différentes bonnes pratiques peut être évaluée avec les étapes (C) citation, (D) dissémination et (U) use du protocole CDUR pour les logiciels de la recherche [39].

D'autres dimensions importantes des logiciels de la recherche comprennent une vision en tant que production scientifique et aussi en tant qu'objet juridique. Nous analysons dans les deux sections suivantes (section 2 et section 3) les questions relatives à ces deux dimensions, avec des descriptions détaillées des caractéristiques de ces types d'objets. La question de l'usage comme critère à retenir dans les caractéristiques d'une bonne définition est soulevée régulièrement, et la section 4 est dédiée à l'étude de ce point en détail. Même s'il s'agit d'un point qui doit être compris correctement, il n'est pas inclus dans notre cadre de comparaison, puisque les trois formulations sélectionnées le mentionnent de façon assez similaire.

Pour continuer avec la description du cadre de comparaison, la compréhension correcte de ces différentes dimensions peut être confirmée avec la facilité et la précision avec lesquelles ces trois formulations vont fournir des réponses aux questions du *conundrum challenges* [9], [46], ou par la clarté qu'elles apportent dans un contexte de comparaison entre logiciels et données de la recherche [42].

En relation avec les questions du *conundrum challenges*, rappelons ici qu'elles ont été initialement soulevées par C. Borgman dans le cadre des données de la recherche [9] :

*Data sharing is thus a conundrum. [...] The challenges are to understand which data might be shared, by whom, with whom, under what conditions, why, and to what effects. Answers will inform data policy and practice*⁸.

Les publications de C. Borgman sont bien connues et très consultées dans le domaine des données de la recherche, dans plusieurs domaines scientifiques et en relation avec la science ouverte⁹. Nous étudions dans les sections qui suivent si les définitions analysées apportent des réponses à ces questions.

7. En utilisant, par exemple, les CITATION.cff files proposés dans <https://citation-file-format.github.io/> ou avec des dépôts dans HAL (<https://doc.hal.science/deposer/deposer-le-code-source/>), dans Software Heritage (<https://www.softwareheritage.org/>), ou dans Zenodo (<https://zenodo.org/>).

8. *Le partage des données est une énigme. [...] Les défis consistent à comprendre quelles données peuvent être partagées, par qui, avec qui, dans quelles conditions, pourquoi et à quels effets. Les réponses éclaireront la politique et la pratique des données.*

9. Voir, par exemple, la liste des résultats de la recherche de "Christine L. Borgman" avec le rôle d'*author* dans la plateforme OpenAlex (<https://openalex.org/>) : <https://openalex.org/works?page=1&filter=authorships.author.id:a5015900970>.

Finalement, depuis le début de nos travaux sur les logiciels de la recherche [35], [39], nous avons trouvé nécessaire et utile de réaliser des comparaisons entre ces différents objets : articles, données, logiciels, ce qui nous a permis de mieux comprendre les questions que ces productions soulèvent et d'être en mesure d'apporter certaines réponses. En particulier, nous souhaitons compléter le cadre de comparaison proposé ici avec l'étude de l'adaptation des définitions proposées aux données de la recherche, ce qui permet d'exhiber, selon nous, des critères sur la qualité des trois formulations.

Tel est le cadre complet de comparaison qui fournit la méthodologie que nous allons utiliser pour analyser les trois définitions de logiciels de la recherche, ce qui sera fait après les trois sections suivantes dédiées aux questions relatives aux productions scientifiques, au contexte juridique des logiciels de la recherche, et à la question du critère de l'usage.

2. Logiciel de la recherche, une production scientifique

Logiciel de recherche? Code logiciel de la recherche? Logiciel dans la recherche? Code scientifique? Logiciel d'un laboratoire de recherche? Logiciels issus de la recherche? Codes sources produits par la recherche? Logiciels académiques? La terminologie utilisée en anglais, *research software*, semble désormais bien fixée dans la littérature [3], [5], [6], [7], [21], [34], [39], [44], [46], [47], [51], [52], [53], [54], [56], [62], [63], [76] (et les références qui s'y trouvent) et c'est ce terme que nous adoptons ici, traduit en français comme «logiciel de la recherche¹⁰».

Comme nous l'avons indiqué dans l'introduction, un logiciel est un ensemble d'instructions utilisé sur une machine pour gérer un ensemble de données. Ainsi, il peut réaliser un ensemble d'opérations, de calculs ou de simulations. Désormais, cet ensemble d'instructions et de codes sources peut être écrit par des outils d'intelligence artificielle [72], [74]. Cela soulève des interrogations sur l'utilisation de ces outils, sur l'accès exclusif de certains acteurs aux codes sources des logiciels qui sont, à leur tour, source des outils de l'intelligence artificielle, et sur la possibilité d'exploitation des résultats obtenus avec ces outils, sans connaître qui a produit ces nouveaux codes, dans quel contexte, avec quel objectif, avec quels outils, avec quelles sources... Les questions juridiques, éthiques et de natures diverses qui se posent dans ce cadre sont en cours d'étude (voir, par exemple, [15], [17], [49], [55], [58], [64]), en conséquence nous ne prenons pas en compte ce cas dans ce texte. Les travaux rapportés ici sont donc dédiés aux logiciels produits par

10. Nous observons ici la différence entre les termes *search software*, traduit comme logiciel de recherche, et *research software*, traduit comme logiciel de la recherche. Cette question se pose également pour les données : *search data* ou données de recherche n'a pas la même signification que *research data* ou données de la recherche. À noter que cette problématique ne semble pas se poser pour les termes laboratoire de recherche ou article ou publication de recherche.

des personnes, membres d'une équipe de recherche. La production, notamment la diffusion de cet objet, un logiciel de la recherche, soulève des questions juridiques et de politique scientifique [35], questions que nous avons étudiées dans les années 2006–2013, au sein du laboratoire d'informatique Gaspard-Monge¹¹, au sein du projet national CNRS PLUME (2007–2013) [31], [36], [37], [38], et, depuis 2018, dans plusieurs publications [39], [40], [41], [42], [43], [44], [45], [46], [47].

Le premier point traité dans [35], et l'un des plus importants, est celui de la définition du point de vue juridique et de celui de la politique scientifique en tant que production d'un laboratoire, que nous traiterons dans les sections suivantes, et qui est le sujet d'étude principal de ce document. D'autre part, la première partie de [35] étudie les questions juridiques, qui font ici l'objet de la section 3. La deuxième partie est dédiée à l'étude des décisions qui interviennent dans la diffusion de ces logiciels. Elles relèvent de niveaux décisionnaires différents, qui vont de la personne qui participe à un projet logiciel (chercheurs et chercheuses, ingénieurs et ingénieures, membres d'un laboratoire, responsables, ou bien, simplement membres d'une équipe avec un projet logiciel) aux responsables des tutelles et du financement des logiciels, en passant par la direction du laboratoire.

Dans [35], [39] et [40], nous avons tenu compte également d'un niveau de décision qui dépasse les laboratoires et les institutions, celui de *la communauté scientifique internationale*, au sens large, ce qui comprend, entre autres, la politique scientifique européenne ou les politiques des revues scientifiques. Il inclut également les communautés scientifiques associées à un domaine scientifique spécifique, puisqu'il peut y avoir des pratiques propres à chaque communauté, comme, par exemple, le traitement de la signature des articles [13], [73].

Ce niveau peut interférer, parfois de façon surprenante, sur les pratiques de publication et de diffusion des nouvelles connaissances, comme la renommée d'une revue où l'on souhaite faire publier ses articles et les contrats qui sont signés avec les éditeurs (*publishers*) responsables de ces revues [25] :

... contracts concluded by publishers based on such rights are not disclosed, since they are subject to the exemption provided by the European directive 93/37/CEE.^{12 13}

11. <https://ligm.univ-eiffel.fr/>.

12. ... les contrats conclus par les éditeurs sur la base de ces droits ne sont pas divulgués, car ils sont soumis à l'exemption prévue par la directive européenne 93/37/CEE.

13. Cette directive a évolué depuis 1993. La Directive 93/37/CEE du Conseil du 14 juin 1993 portant coordination des procédures de passation des marchés publics de travaux est abrogée par la *Directive 2004/18/CE du Parlement européen et du Conseil relative à la coordination des procédures de passation des marchés publics de travaux, de fournitures et de services*, remplacée également par la *directive 2014/24/UE du Parlement européen et du Conseil sur la passation des marchés publics*, et qui a une dernière version datée 01/01/2026.

À la différence des questions juridiques de base que nous avons traitées dans [35], et qui sont liées principalement au droit d’auteur et à l’utilisation des licences des logiciels libres ou *open source* (FLOSS : *free/libre/open source software*), les points de réflexion soulevés dans la deuxième partie de [35] trouvent, depuis 2015, des avancées considérables, tant en France qu’ailleurs. Ces points de réflexion sont les suivants : définition, signature, références, liste des œuvres d’un laboratoire, libre accès, validation, qualité et évaluation, motivation, objet.

Nous souhaitons mentionner ici le deuxième plan national pour la science ouverte [65] publié en juillet 2021, dont le quatrième axe inclut l’ouverture et la promotion des codes sources produits par la recherche ainsi que la création du Collège codes sources et logiciels¹⁴ au sein du Comité pour la science ouverte (COSO)¹⁵. Les travaux en cours au COSO¹⁶ (et ailleurs, de façon plus générale) placent sur la table de discussion tous ces points de réflexion et d’autres comme la proposition des procédures de diffusion [33], [34], [56], [77] et d’évaluation [39] de ces productions scientifiques, et la question des plans de gestion de logiciels [48], [50], [63], les *conundrum challenges* [9], [46], les *FAIR principles* [5], [42], [43], [79].

Ces dernières références nous montrent le long chemin parcouru par les logiciels de la recherche, entre le fait d’être intégrés dans un plan national pour la science ouverte du ministère de la Recherche [65] en 2021 et une vision de *savoir tacite* en 1994 [69, p. 508] qui pouvait peut-être générer des retombées positives :

*What is alluded to here is that there may be important positive spillovers across projects in the form of ‘learning effects’. [...] which often remain in the region of tacit knowledge [...] including the development of generic computer software for performing data processing, storage, retrieval and network transmission*¹⁷.

L’analyse d’un point important dans toute production scientifique correspond à celui de l’équipe de production. En effet, il est très courant que les logiciels de ce type soient écrits, documentés et diffusés par une unique personne, mais l’équipe de production peut s’avérer complexe, et inclure des membres avec des statuts divers, qui peuvent contribuer à un logiciel avec des rôles très différents [39], [45]. Du point de vue du droit d’auteur, il est important de distinguer les auteurs et autrices du code, qui génèrent des

14. <https://www.ouvrirlascience.fr/college-codes-sources-et-logiciels/>.

15. <https://www.ouvrirlascience.fr/accueil/>.

16. Voir les productions du collège Codes sources et logiciels sur <https://www.ouvrirlascience.fr/college-codes-sources-et-logiciels/?menu=3>.

17. Il est ici sous-entendu qu’il peut y avoir d’importantes retombées positives entre les projets, sous forme d’effets d’apprentissage’. [...] qui restent souvent dans le domaine du savoir tacite [...] y compris le développement de logiciels génériques pour le traitement, le stockage, la récupération et la transmission de données en réseau.

droits d’auteur, des contributions mineures. Mais dans le contexte académique il arrive très régulièrement que les responsables du projet ne participent pas de façon importante à l’écriture du code. Par contre, ce rôle de responsable est à souligner en tant qu’activité, et souvent responsabilité académique. C’est pour cela que dans [39] nous avons sélectionné les trois rôles suivants pour présenter les trois types de contributions les plus importantes à un logiciel de la recherche, tout en sachant qu’une seule personne peut cumuler ces trois rôles, ou bien que les limites peuvent rester très floues :

- responsabilité de l’équipe ou du projet «logiciel de la recherche» ;
- contributions importantes (liées à l’écriture du code) ;
- contributions mineures (liées à des petites contributions à l’écriture du code ou à d’autres participations, comme la conception, la documentation, les tests...).

Comme expliqué dans [45], ces trois rôles peuvent être également mis en lumière pour la production des publications ou des données de la recherche, ce qui peut participer à faciliter la gestion des informations de *authorship* et *contributorship* du système CRediT, adopté par beaucoup de publications scientifiques [61].

Un autre point qui peut s’avérer important dans la compréhension de ces productions scientifiques correspond à la réalisation des catalogues de logiciels d’un laboratoire, comme ceux élaborés entre 2008 et 2018 au laboratoire d’informatique Gaspard-Monge (LIGM) par la première autrice de ce texte, dans le cadre des évaluations HCERES en 2013 et en 2019, [37], [38] (où nous avons utilisé les études [31], [33], [35]). Lorsque les documents du laboratoire sont préparés pour ces évaluations, il est relativement facile de lister les membres du laboratoire pour la période évaluée, en éliminant les personnels qui ne font plus partie du laboratoire, et de répertorier les publications de ces personnels associées aux dates de la période évaluée. Mais comment faire pour réaliser le catalogue des logiciels? Nous rappelons ici que les logiciels ne disposent pas d’une procédure de publication comme celle des articles scientifiques, qui associe, de façon précise, une référence, avec un titre, des auteurs et autrices, des dates, un lieu de publication, etc. à ces articles.

Pour illustrer ces difficultés, nous reprenons ici l’exemple du catalogue de logiciels LIGM, dont la première version (2008–2013) a été réalisée en utilisant les fiches RELIER [36] sur le serveur du projet CNRS PLUME¹⁸. La question se pose alors de déterminer les logiciels de ce premier catalogue à inclure dans le catalogue suivant pour la période 2013–2018. Pour résoudre ce problème, nous avons utilisé le concept de *logiciel actif* dans les textes de préparation de l’HCERES 2019, défini comme suit.

18. Le serveur du projet PLUME a été arrêté en 2022. Une trace du catalogue de logiciels LIGM avant 2013 est archivée sur Zenodo : <https://doi.org/10.5281/zenodo.7063170>.

Depuis 2013, un logiciel est considéré comme «*actif*» s'il vérifie au moins une des conditions suivantes :

- une partie du développement est réalisée depuis 2013 par un membre du laboratoire ;
- il existe des publications associées au logiciel, qui présentent des résultats obtenus avec ce logiciel, datées depuis janvier 2013, ou bien qui sont en préparation, ce qui veut dire que le logiciel est utilisé, revu, corrigé... et il est possible d'y associer des exemples d'utilisation ;
- il existe des activités de maintenance : corrections de bugs, échanges avec les utilisateurs...

Ceci reflète la durée de vie du logiciel, dont la fin arrive lorsque le développement et la maintenance du logiciel s'arrêtent, sans savoir, souvent, si ces activités vont reprendre ou pas dans un futur proche ou distant. Elle est différente de celle des articles et autres publications : en général on arrête de s'occuper d'un article lorsqu'il est publié, mais quand arrête-t-on de s'occuper d'un logiciel ?

Finalement, nous souhaitons mentionner les documents suivants [3], [5], [6], [7], [8], [16], [21], [40], [47], [52], [57], [65], [66], [67], [76], [77] pour d'autres questions de politique scientifique qui sont liées aux logiciels, ce qui complète cette section.

3. Logiciel, objet juridique

Cette section est dédiée aux bases juridiques de l'objet logiciel, liées principalement au droit d'auteur et à l'utilisation des licences des logiciels FLOSS [2], [4], [14], [30], [34], [35]. Le Code de la propriété intellectuelle [30] indique que le logiciel est une œuvre de l'esprit protégée par le droit d'auteur, et l'arrêté du ministère de l'Industrie du 22 décembre 1981 relatif à l'enrichissement du vocabulaire de l'informatique [68] donne la définition suivante.

Logiciel : ensemble des programmes, procédés et règles, et éventuellement de la documentation, relatifs au fonctionnement d'un ensemble de traitement de données (en anglais : software).

En tant qu'œuvre de l'esprit, il s'agit d'une œuvre écrite, et ce qui est écrit sont les programmes, ou le *code source* du logiciel. Nous parlerons ici de cet objet logiciel d'une manière globale, comme un ensemble (de fichiers) qui contient son code source et qui peut contenir d'autres objets comme le code compilé (s'il y en a), une documentation (ou plusieurs : documentation pour les utilisateurs, pour les possibles contributeurs, des exemples d'utilisation...), et peut-être aussi d'autres objets comme des tests, des procédures d'installation, des données... C'est en général un objet qui évolue dans le

temps et qui peut avoir plusieurs versions selon les nouvelles informations, fonctionnalités ou composantes qui y sont ajoutées, il peut y avoir une maintenance faite par l'équipe (ou l'entreprise¹⁹) responsable du logiciel qui se charge de la correction de bugs, de la maintenance et des nouveaux développements.

Comme indiqué dans la section précédente, un logiciel a une certaine durée de vie, dont la fin arrive lorsque le développement et la maintenance du logiciel s'arrêtent. Il peut être repris par la même équipe des années plus tard, ou bien par des nouvelles équipes, ce qui est favorisé par la diffusion de ces logiciels avec des licences de logiciels FLOSS.

Comme pour toute œuvre écrite, il est difficile d'éviter l'expression de la personnalité de ses auteurs ou autrices. Mais la question de l'originalité de cette œuvre a été révisée (réduite) par l'arrêt Pachot [10], du 7 mars 1986, selon le terme de «apport intellectuel» pour faire preuve :

[d'un] effort personnalisé allant au-delà de la simple mise en œuvre d'une logique automatique et contraignante.

Il y a eu plusieurs débats récents sur cette question de l'originalité d'un logiciel, notamment ceux organisés par la Compagnie nationale des experts de justice en informatique et techniques associées (CNEJITA) [23], [24], mais les travaux de P. Depaz [19], [20], [60] et de F. Pellegrini [70] montrent que l'originalité d'un logiciel n'est plus l'objet de l'arrêt Pachot.

Comme nous l'avons mentionné dans la section 2, le contexte juridique de base étudié dans [35] a peu évolué depuis 2015, mais nous souhaitons mentionner ici les points suivants qui vont compléter cette étude juridique, dont le premier fait référence aux droits patrimoniaux dans le cas des stages dans les établissements de recherche qui participent à des projets logiciels.

L'article 2 de l'ordonnance n° 2021-1658 du 15 décembre 2021 relative à la dévolution des droits de propriété intellectuelle sur les actifs obtenus par des auteurs de logiciels ou inventeurs non salariés ni agents publics accueillis par une personne morale réalisant de la recherche [29], [75] indique :

Art. L. 113-9-1. Sauf stipulations contraires, lorsque des personnes qui ne relèvent pas de l'article L. 113-9 et qui sont accueillies dans le cadre d'une convention par une personne morale de droit privé ou de droit public réalisant de la recherche créent des logiciels dans l'exercice de leurs missions ou d'après les instructions de la structure d'accueil, leurs droits patrimoniaux sur ces logiciels et leur documentation sont dévolus à cette structure d'accueil, seule habilitée à les exercer, si elles se trouvent à l'égard de cette structure dans une situation où elles perçoivent une

19. Ce document ne prend pas en compte le cas des entreprises ou des start-ups, y compris si elles ont une origine académique.

contrepartie et où elles sont placées sous l'autorité d'un responsable de ladite structure.

Par ailleurs, les logiciels issus de la recherche menée au sein des établissements publics dont les statuts prévoient une mission de recherche font l'objet de la section II de l'article 163 de la loi n° 2022-217 du 21 février 2022 relative à la différenciation, la décentralisation, la déconcentration, et portant diverses mesures de simplification de l'action publique locale [27] :

Le Gouvernement remet au Parlement, dans un délai de dix-huit mois à compter de la promulgation de la présente loi, un rapport sur la production et la valorisation des logiciels libres et propriétaires issus de la recherche menée au sein des établissements publics d'enseignement supérieur, des établissements publics à caractère scientifique et technologique, des établissements publics à caractère industriel et commercial et des autres établissements publics à caractère administratif dont les statuts prévoient une mission de recherche.

Ils font également l'objet de l'article 2 du décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche [28] :

Les établissements publics et fondations reconnues d'utilité publique mentionnés au troisième alinéa de l'article L. 211-2 du code de la recherche : [...] 3° Promeuvent la diffusion des publications en accès ouvert et la mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche afin d'en garantir la traçabilité et la reproductibilité.

À noter que le seul accès aux codes sources sans avoir accès aux autres composantes du logiciel, comme par exemple la documentation, peut empêcher la reproductibilité qui est au centre des intérêts de cet article 2.

En ce qui concerne les licences de réutilisation, la loi de la République numérique [26] a prévu la création d'une liste de licences qui peuvent être utilisées par les administrations pour la réutilisation à titre gratuit de leurs informations publiques, liste qui contient des licences de logiciels FLOSS, et qui a été fixée par deux décrets : le décret n° 2017-638 du 27 avril 2017 et le décret n° 2021-1559 du 1^{er} décembre 2021 (voir la plateforme des données publiques françaises qui publie cette liste²⁰).

20. <https://www.data.gouv.fr/pages/legal/licences/>.

4. La question de l'usage ou « usage *vs* production »

Comme il est mentionné dans l'introduction, la question de l'*usage* se pose régulièrement au milieu des réflexions et des discussions qui ont conduit aux formulations des différentes définitions, et cette section est consacrée à analyser plus en détail ce sujet.

Par exemple, dans [51], les discussions intègrent un spectre qui va d'une définition inclusive, où l'usage de tout type de logiciel dans une activité de recherche lui donne une caractéristique de logiciel de la recherche, à une définition exclusive, qui fera partie de la formulation finale retenue (voir la section 6).

Ceci est également pris en considération dans [71], qui mentionne la caractéristique *production de la recherche* comme un premier critère pour soutenir la définition qui y est proposée, mais qui est déclaré comme non pertinent puisque de nombreux logiciels couramment utilisés, s'ils sont bien issus de travaux de recherche ne sont pas (ou plus) considérés comme des logiciels de la recherche. Le deuxième critère pour soutenir ladite définition pourrait être l'utilisation du logiciel au sein d'un environnement de recherche, critère qui soulève également des questions.

Nous souhaitons donc, ici, bien faire la distinction entre la production et l'usage, et les actions qui peuvent en dériver, puisque faire cette distinction nous semble essentiel pour mieux comprendre notre objet d'étude.

Comme il a été vu dans la section 3, la production d'un logiciel en milieu de recherche génère des droits d'auteur [30], [35] et donc un patrimoine immatériel auquel il est possible d'associer des actions de valorisation et de protection [8], [14], [57], [77]. Il nous semble plutôt difficile, voire impossible, d'associer des actions de valorisation au fait d'utiliser un logiciel, dans le milieu scientifique ou autre.

Par ailleurs, et comme argumenté dans [71], la question de l'usage peut soulever des questions. Mais, à notre avis, le fait d'utiliser un logiciel n'a pas d'incidence sur sa qualité en tant que logiciel de la recherche.

De nombreux logiciels sont utilisés, de façon consciente ou pas, à partir du moment où un chercheur ou une chercheuse se sert d'un ordinateur ou d'un téléphone portable pour ses activités quotidiennes de recherche ; ces logiciels font partie des niveaux mentionnés dans [53].

Quelle est la différence entre utiliser un logiciel comme Tex²¹ pour l'édition d'un texte (scientifique), dont la provenance du milieu académique ne pose pas de question mais dont les multiples successeurs et produits dérivés sont difficilement classés comme logiciels de la recherche, avec l'utilisation des logiciels de calcul scientifique comme Maple²², qui, de façon similaire, a une

21. <https://fr.wikipedia.org/wiki/Tex>.

22. <https://fr.wikipedia.org/wiki/Maple>.

origine académique dans l'université de Waterloo, mais qui est devenu un produit propriétaire, dont la production et la maintenance dépassent largement le milieu scientifique initial ? À notre avis, l'usage de ces deux logiciels, l'un pour l'écriture d'un texte scientifique, l'autre pour faire des calculs scientifiques, ne permet pas de distinguer la nature scientifique de leur jeunesse — que nous pouvons qualifier en tant que logiciel de la recherche — ni de distinguer les arguments pour ou contre la qualification de logiciel de la recherche de la vie actuelle des multiples produits dérivés. Un autre exemple évoqué en [71] correspond aux logiciels de cryptographie garantissant la sécurité des échanges numériques, avec également une provenance du milieu scientifique et dont l'usage dépasse largement les milieux académiques.

D'un autre côté, il serait difficile de demander de rendre au Parlement des rapports sur l'usage des logiciels (de la recherche ou autres) afin d'éclairer le législateur sur les pratiques relatives au logiciel en vue de valoriser un certain patrimoine. Il pourrait être également compliqué de demander aux établissements publics de prendre des mesures pour la diffusion en accès ouvert des logiciels utilisés dans les activités de recherche et qui ne sont pas produits par les équipes hébergées dans ces établissements.

Pour compléter cette vision sur l'usage en milieu de recherche, nous reprenons ici la méthodologie de comparaison exploitée dans [35] : pour mieux comprendre les logiciels produits dans les laboratoires de recherche, faisons la comparaison entre la production des articles et des logiciels.

Il nous semble raisonnable d'affirmer que, en général, dans les milieux scientifiques, le concept de publication scientifique, la nature d'un article de recherche, ne soulève pas de question.

La science se construit en se tenant debout sur les épaules de géants²³... Que veut dire utiliser des articles pour produire des nouvelles connaissances présentées dans un nouvel article ? En général, les scientifiques étudient une bibliographie pour traiter un sujet d'étude, et ces travaux sont mentionnés dans le nouveau texte en forme de références et de citations. Mais les scientifiques ne réduisent pas les matériels à étudier, et à utiliser, aux seuls textes scientifiques. Prenons l'exemple des romans de Victor Hugo, qui ont été et sont toujours analysés dans des domaines scientifiques divers (sociologie, histoire, linguistique, littérature...), mais leur étude et leur utilisation dans ces contextes scientifiques ne leur font pas perdre leur nature initiale de roman, et ne peuvent les doter d'une nature de texte scientifique.

Nous pouvons également affirmer que les publications réalisées par des journalistes dans les journaux quotidiens ou hebdomadaires ne sont pas facilement classifiées comme des publications scientifiques, mais, néanmoins,

23. Isaac Newton, dans sa lettre à Robert Hooke le 5 février 1675, note cet aphorisme devenu célèbre : *Si j'ai vu plus loin, c'est en me tenant debout sur les épaules de géants...*

ceci n'est jamais un obstacle pour leur utilisation par des scientifiques de différents domaines comme source d'information.

Pour continuer avec notre méthodologie de comparaison, cette fois-ci entre logiciels et données de la recherche, nous souhaitons mentionner la définition de données de la recherche proposée dans [41], où la question du contexte de production scientifique rigoureux a été également soulevée. Comme pour les articles et les logiciels, ce contexte n'exclut en aucun cas l'utilisation des données d'origines diverses dans un contexte scientifique. Nous soulignons simplement l'importance de bien montrer l'origine des divers ensembles de données qui ont été utilisés dans un projet ou dans une étude, et d'expliquer, par exemple, si ces ensembles ont été modifiés pour la réalisation des travaux de recherche ; ceci afin de promouvoir les bonnes pratiques de citation, de transparence et d'intégrité qui facilitent la reproductibilité et la validation des résultats scientifiques obtenus avec ces données.

Par ailleurs, pour compléter et renforcer cette vision, nous faisons référence aux travaux juridiques de C. Domange [22], où il est affirmé :

Les informations publiques sont considérées comme des informations culturelles dès lors qu'elles sont détenues ou élaborées par une entité dont l'activité est directement culturelle. Le seul exercice d'une activité dans le secteur culturel ne suffit pas.

Ceci nous montre l'importance de l'origine des données pour pouvoir les qualifier de *données culturelles*, et nous pouvons donc faire une affirmation similaire pour les données et pour les logiciels de la recherche.

Nous pouvons ainsi écarter le seul critère de l'utilisation en tant que critère important pour définir un logiciel comme logiciel de la recherche : il doit être aussi élaboré dans un contexte de recherche. Il est bien sûr possible d'externaliser des parties du développement, ou bien de recruter des personnels en CDD pour faire des développements, mais cette activité restera sous le contrôle des scientifiques qui vont utiliser le logiciel, et nous pouvons, dans ce cas, interpréter les équipes de recherche d'une façon plus large (et sûrement temporaire) pour inclure ces activités. Voir [45] pour plus de réflexion sur les équipes de recherche.

Les sections 2, 3, 4 qui précèdent complètent le cadre de comparaison de la section 1. Les trois sections qui suivent sont dédiées à l'analyse des trois définitions sélectionnées comme nous l'avons annoncé dans l'introduction.

5. D1-COSO. Définition 1 : la définition COSO

Cette section est dédiée à l'étude des travaux présentés dans [71], faisant référence à la définition proposée par le collège Codes sources et logiciels (blog du Comité pour la science ouverte) en 2022 [66] :

Les logiciels de recherche sont développés pour répondre à des besoins spécifiques de la science. Ils sont conçus, maintenus, et utilisés par des scientifiques (chercheurs et ingénieurs) et institutions de recherche, éventuellement dans une dimension internationale. Ils peuvent découler de travaux de recherche comme ils peuvent les favoriser, notamment par des publications avant / sur / autour / avec le logiciel. Ceux-ci peuvent se formaliser de différentes façons (une plateforme, un intergiciel, un workflow ou une bibliothèque, module ou greffon d'un autre logiciel) et être ainsi en interaction dans un écosystème ou au contraire plus autonomes.

et les critères qui y sont analysés pour soutenir la proposition de cette formulation.

L'équipe de ce collège est composée, à la date d'écriture de ce texte, d'une quarantaine de personnes²⁴ et est organisée en 5 groupes de travail, mais nous n'avons pas trouvé d'information sur les personnes qui ont rédigé cette définition en 2022. Ce qui est mentionné dans [66] s'appuie sur un historique existant :

S'appuyant sur un historique de partage ancien et fortement lié aux pratiques communautaires développées dans le contexte des logiciels libres, la démarche de reconnaissance du logiciel en tant qu'élément incontournable de la production scientifique est bien entamée.

mais il ne semble pas y avoir d'autre information sur l'historique mentionné ou sur les travaux consultés. L'article [71], publié dans la revue juridique Lamy *Droit de l'immatériel*, mentionne deux références : [16], [66].

Cette définition place l'activité scientifique au centre de sa formulation et mentionne les activités associées (logiciels conçus, maintenus, et utilisés) par des scientifiques et des institutions de recherche. À noter que c'est la mise en forme, c'est-à-dire l'écriture des programmes (ie. du code source), qui génère des droits, et que l'étape de conception ne génère pas de droits sans cette mise en forme [30], [35]. Par ailleurs, cette définition n'indique pas si ces différentes activités sont réalisées par une même équipe, au sein d'une institution, ou bien s'il y a différentes équipes, liées à des institutions différentes. La possible dimension internationale est mentionnée, sans clarifier si elle s'applique à l'équipe qui développe et maintient le logiciel, ou bien au contexte d'utilisation. Il semble également un peu étrange d'associer des activités de conception, de maintenance ou d'utilisation à une institution.

Par ailleurs, dans [71], on trouve l'argument :

Il découle de cette analyse que des logiciels, tant issus de la recherche proprement dite que servant de façon spécifique à une activité de

24. <https://www.ouvrirlascience.fr/college-codes-sources-et-logiciels/?menu=4>.

recherche, ne peuvent être qualifiés de “logiciels de recherche” que dans la mesure où ils peuvent être liés à une activité mesurable (création, maintenance, modification à fin d’usage spécifique) et récente prenant place dans le monde de la recherche.

L’analyse de la section 4 nous a permis d’écarter les «logiciels servant de façon spécifique à une activité de recherche qui ne sont pas issus de la recherche proprement dite». Par ailleurs, une équipe qui produit un logiciel de la recherche le fait en relation avec une activité et un objectif de recherche, pour réaliser une étude, pour obtenir ou vérifier un résultat, pour faciliter la compréhension d’un objet. Mais cet objectif, qui s’inscrit généralement dans les activités quotidiennes de la recherche, peut évoluer dans le temps, l’équipe peut se doter de nouveaux objectifs et faire évoluer le logiciel dans un certain contexte, ou bien elle peut mettre ce logiciel de côté pour privilégier d’autres activités de recherche. Le logiciel peut être maintenu ou pas, rester *actif* ou pas. Il peut être réactivé des années plus tard par l’équipe initiale de développement ou bien par une nouvelle équipe.

Donc, l’activité de recherche et l’usage prévu du logiciel sont, à notre avis, en général, spécifiques au moment de la création, et peuvent se poursuivre ou pas, par des périodes alternées, suivant les évolutions de l’équipe et de ses activités de recherche, souvent en fonction des financements obtenus. Mais, comme toute activité de recherche, cette spécificité est évolutive, et il peut y avoir une grande variabilité. Dans [71] il est indiqué que :

... dans la mesure où ils peuvent être liés à une activité mesurable (création, maintenance, modification à fin d’usage spécifique) et récente prenant place dans le monde de la recherche.

Mais ceci nous semble soulever également des questions. Une bonne partie des logiciels répertoriés au LIGM entre 2008 et 2018 [37], [38] ne peut plus être associée à des activités mesurables et récentes de la recherche, et ceci ne peut pas avoir de conséquence sur leur qualité de logiciel de la recherche. Par ailleurs, les activités de la recherche sont parfois difficilement mesurables, les priorités peuvent varier énormément dans une année, peuvent dépendre des financements obtenus et de bien d’autres paramètres. Le concept de logiciel *actif* utilisé pour réaliser le catalogue des logiciels LIGM nous semble mieux décrire tous ces aléas.

Cette définition de logiciel de la recherche indique la formalisation des logiciels en tant que plateforme, intergiciel, *workflow*... ce qui reflète partiellement la dimension numérique de l’objet défini, mais qui implique aussi une certaine classification qui semble superflue.

Pour conclure cette section, il nous semble que plus de précision sur l’objet en soi, en tant qu’objet numérique et en tant que production scientifique, pourrait améliorer la clarté de la définition, ce qui pourrait, en conséquence,

apporter des réponses aux questions du type : quel est l'objet partagé? par qui? avec qui?... De plus, cette formulation de la définition D1-COSO semble également trop complexe pour pouvoir réaliser le parallèle avec la définition des données de la recherche.

6. D2-RDA. Définition 2 : la définition RDA

La définition que nous appelons ici D2-RDA a été proposée en 2021 au sein du groupe de travail FAIR4RS²⁵ de la Research Data Alliance (RDA)²⁶. Elle est publiée dans un document preprint sur Zenodo [51] :

Research Software includes source code files, algorithms, scripts, computational workflows and executables that were created during the research process or for a research purpose. Software components (e.g., operating systems, libraries, dependencies, packages, scripts, etc.) that are used for research but were not created during or with a clear research intent should be considered software in research and not Research Software. This differentiation may vary between disciplines. The minimal requirement for achieving computational reproducibility is that all the computational components (Research Software, software used in research, documentation and hardware) used during the research are identified, described, and made accessible to the extent that is possible²⁷.

Le groupe de travail associé à [51] est composé d'une trentaine de personnes avec des rôles différents : auteurs (*signataires*) de [51], ou bien membres du groupe FAIR4RS, ou bien relecteurs du document, et cette définition est le résultat d'amples discussions au sein de ce groupe. Le document fait référence à une vingtaine de publications (y compris [3], [39], [53], [59]) et à un workshop en 2021.

Cette définition place au centre la dimension numérique de l'objet, en listant les fichiers source du logiciel, les exécutables, les flux de calcul (*computational workflows*), les scripts et les algorithmes qui ont été créés tout le long d'un travail de recherche, avec un objectif de contribuer à cette recherche.

25. <https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg>.

26. <https://www.rd-alliance.org/>.

27. *Les logiciels de la recherche comprennent les fichiers de code source, les algorithmes, les scripts, les flux de calcul et les exécutables créés au cours du processus de recherche ou à des fins de recherche. Les composants logiciels (systèmes d'exploitation, bibliothèques, dépendances, packages, scripts, etc.) utilisés pour la recherche, mais n'ayant pas été créés pendant ou avec une intention de recherche claire, doivent être considérés comme des logiciels dans la recherche et non comme des logiciels de la recherche. Cette distinction peut varier selon les disciplines. L'exigence minimale pour atteindre la reproductibilité informatique est que tous les composants informatiques (logiciels de la recherche, logiciels utilisés en recherche, documentation et matériel) utilisés pendant la recherche soient identifiés, décrits et rendus accessibles dans la mesure du possible.*

Il nous semble que la terminologie de *computational workflow* a besoin de plus de clarté, nous faisons ici référence à la page Wikipédia *Scientific workflow system*²⁸, qui indique, dans sa section *Scientific workflows* :

*The simplest computerized scientific workflows are scripts that call in data, programs, and other inputs and produce outputs that might include visualizations and analytical results*²⁹.

Quelle est donc la différence entre logiciel, *computational workflow* et script ? Si nous comprenons correctement le contexte juridique, il n'a pas de distinction entre ces trois objets, qui sont protégés de la même façon en tant qu'œuvres de l'esprit.

Cette définition donne également un rôle central à l'activité de recherche, mais elle ne mentionne pas qui fait cette recherche, et donc elle manque la liaison avec l'objet juridique et les droits qui y sont générés, dont certains sont associés à ses auteurs et autrices et aux institutions qui sont leurs employeurs [14], [30], [35].

Par ailleurs, les algorithmes programmés dans un logiciel peuvent être créés au cours d'une activité de recherche spécifique, mais ils peuvent aussi être réalisés par d'autres équipes, tout à fait indépendantes de l'équipe associée au développement du logiciel. Également, l'objectif d'un logiciel peut être la comparaison de plusieurs algorithmes développés par plusieurs équipes, et donc ces algorithmes ne seront pas créés en même temps que le logiciel au sein d'une activité de recherche spécifique.

Il nous semble que cette définition manque de clarté dans la liste des éléments d'un logiciel, puisqu'elle ne propose pas une vision précise des objets logiciel, script ou *computational workflow*. Alors que la liaison avec l'activité de recherche est clairement établie, cette formulation ne mentionne pas l'équipe qui produit le logiciel ou la recherche, et donc ne facilite pas la réponse à la question « *qui partagera le logiciel ?* » Il nous semble donc que les dimensions de l'objet juridique et de l'objet numérique pourraient nécessiter de plus de précision.

De plus, une bonne formulation d'une définition doit éviter d'inclure la distinction entre les logiciels dans la recherche (ce qui fait référence à l'utilisation, comme analysé dans la section 4) et les logiciels de la recherche. Elle ne devrait pas inclure, non plus, l'objectif de la reproductibilité de la recherche, qui, bien qu'étant un objectif important dans la question « *pourquoi et à quels effets le logiciel est partagé ?* », n'est pas le seul objectif important [9], [41], [46].

Par ailleurs, nous ne voyons pas de façon simple comment faire le parallèle avec une possible définition des données de la recherche en suivant cette formulation.

28. https://en.wikipedia.org/wiki/Scientific_workflow_system.

29. *Les flux de travail scientifiques informatisés les plus simples sont des scripts qui font appel à des données, des programmes et d'autres entrées et produisent des sorties pouvant inclure des visualisations et des résultats analytiques.*

7. D3-CDUR. Définition 3 : la définition CDUR

Pour terminer, cette section est dédiée à la définition suivante de logiciel de la recherche que nous avons proposée en 2019 [39] :

*Research software is a well identified set of code that has been written by a (again, well identified) research team. It is software that has been built and used to produce a result published or disseminated in some article or scientific contribution. Each research software encloses a set (of files) that contains the source code and the compiled code. It can also include other elements as the documentation, specifications, use cases, a test suite, examples of input data and corresponding output data, and even preparatory material*³⁰.

Cette définition trouve son origine dans l'expérience acquise tout le long des carrières scientifiques des deux auteurs en tant que développeurs de logiciels ou participants dans des projets et activités très diverses en relation avec des projets logiciels³¹. Cette expérience s'est particulièrement enrichie avec la *Mission logiciels LIGM* [32], [37], [38] et la participation au projet national CNRS PLUME [4], [31], [33], [35], [36], études et travaux qui se poursuivent, depuis 2018, dans plusieurs publications [39], [41], [42], [45], [46], [47].

Toute cette expérience accumulée nous a permis d'observer, de façon directe, les problématiques que les logiciels soulèvent dans nos laboratoires et nos projets de recherche, et cela au niveau local et national, international ou au sein d'une équipe, dans des communautés scientifiques diverses.

Cette définition a fait l'objet d'une publication [39] dans le journal *F1000-Research*³² et a bénéficié d'une procédure d'*open peer review* où les rapports des *referees* sont publiés avec l'article. Cette publication a deux versions, la deuxième prenant en compte les commentaires et les propositions d'amélioration des rapporteurs; elle inclut un total de 80 références, dont 22 sont mentionnées dans sa section 2.1, *Research software definition*.

Cette définition place l'équipe de développement au centre de la formulation, faisant ainsi la liaison avec la définition juridique de logiciel en tant qu'œuvre de l'esprit qui a des auteurs ou autrices et des droits associés. Mais elle souligne aussi les aspects relatifs à une production scientifique, faite par une équipe, et indique clairement l'objectif d'une activité de recherche, ce qui

30. Un logiciel de la recherche est un code bien identifié, écrit par une équipe de recherche (là encore bien identifiée). Il s'agit d'un logiciel développé et utilisé pour produire un résultat publié ou diffusé dans un article ou une contribution scientifique. Chaque logiciel de recherche inclut un ensemble (de fichiers) contenant le code source et le code compilé. Il peut également inclure d'autres éléments tels que la documentation, les spécifications, les cas d'utilisation, une suite de tests, des exemples de données d'entrée et de données de sortie correspondantes, et même du matériel préparatoire.

31. Voir <https://openalex.org/works?page=1&filter=authorships.author.id:a5029421675>, par exemple.

32. <https://f1000research.com/>.

est aussi souligné dans [59] (voir aussi [53]), où D. Kelly propose la définition suivante :

*Scientific software is defined by three characteristics: (1) it is developed to answer a scientific question; (2) it relies on the close involvement of an expert in its scientific domain; and (3) it provides data to be examined by the person who will answer that question [...]*³³

Cette formulation de D. Kelly, publiée en 2011, montre déjà les différents rôles qui peuvent intervenir dans un développement logiciel, où les personnes qui développent le code peuvent être différentes de l'expert qui traitera la question scientifique; des réflexions qui ont été également étudiées dans [33] et [35] (voir la section 2).

Finalement, cette définition D3-CDUR explique le logiciel de la recherche comme un sous-ensemble de l'ensemble *logiciel*, et traite ainsi le contexte numérique lorsqu'il y est mentionné l'ensemble de fichiers qui peut contenir des objets de natures différentes : code source, code compilé, documentation et peut-être d'autres éléments.

Basés sur cette vision, nous avons proposé des procédures de diffusion [33], [34], les protocoles CDUR d'évaluation [39], [42], des réponses aux *conundrum challenges* (quels logiciels sont destinés à être partagés, par qui, avec qui, comment, où, dans quelles conditions, pourquoi et à quels effets...?) [46] et des études sur l'adoption des politiques de *free/open source software* dans des institutions scientifiques [40], [47].

Comme déjà mentionné, depuis nos premiers travaux sur les logiciels [35], [39] nous utilisons une méthodologie de comparaison entre des productions diverses, ce qui nous a aidé à mieux comprendre ces objets. En particulier, cette méthodologie nous a permis de proposer une définition de données de la recherche (*research data*) qui trouve une formulation similaire à celle des logiciels de la recherche, travaux qui sont analysés en profondeur dans [41], [42].

Research data is a well identified set of data that has been produced (collected, processed, analyzed, shared & disseminated) by a (again, well identified) research team. The data has been collected, processed and analyzed to produce a result published or disseminated in some article or scientific contribution. Each research data encloses a set (of files) that contains the dataset maybe organized as a database, and it can also include other elements as the documentation, specifications, use cases, and any other useful material as provenance information, instrument

33. Les logiciels scientifiques se définissent par trois caractéristiques : (1) ils sont développés pour répondre à une question scientifique; (2) ils reposent sur l'implication étroite d'un expert dans leur domaine scientifique; et (3) ils fournissent des données à examiner par la personne qui répondra à cette question [...]

*information, etc. It can include the research software that has been developed to manipulate the dataset (from short scripts to research software of larger size) or give the references to the software that is necessary to manipulate the data (developed or not in an academic context)*³⁴.

Cette définition de données de la recherche a été estimée par R. Cerda-Cosme et E. Mendez l'une des formulations les plus complètes [11].

En résumé, il nous semble que cette formulation de définition D3-CDUR correspond aux critères d'une bonne définition [18] : elle spécifie ce qui est ou pas inclus dans l'objet à définir sans imprécision ni obscurité. Elle montre également les caractéristiques importantes relatives aux logiciels de la recherche en tant que objets numériques, scientifiques et juridiques, et n'inclut pas de classification. Par ailleurs, elle fournit une base solide pour les réponses que nous avons proposées aux *conundrum challenges* [46] et facilite une formulation similaire pour définir les données de la recherche [41], [42].

8. Discussion

Cette section est dédiée à la comparaison entre les trois définitions étudiées dans les sections 5, 6 et 7 en suivant la méthode proposée dans la section 1. Nous faisons remarquer que cette discussion correspond à notre perception du sujet des logiciels de la recherche, qui est supportée par notre expertise diverse et nos travaux précédemment publiés. Cette discussion est résumée dans la table sur la figure 1.

34. *Les données de la recherche constituent un ensemble de données bien identifié, qui a été produit (données collectées, traitées, analysées, partagées et diffusées) par une équipe de recherche (également bien identifiée). Ces données ont été collectées, traitées et analysées afin de produire un résultat publié ou diffusé dans un article ou une contribution scientifique. Chaque ensemble de données de la recherche comprend un ensemble de fichiers contenant les données organisées, éventuellement, sous forme de base de données. Il peut également inclure d'autres éléments tels que la documentation, les spécifications, les cas d'utilisation et tout autre élément utile, comme des informations sur la provenance, les instruments, etc. Il peut inclure le logiciel de la recherche développé pour manipuler les données (de courts scripts à des logiciels de la recherche plus volumineux) ou fournir les références aux logiciels nécessaires à la manipulation des données (développés ou non dans un contexte académique).*

	D1-COSO	D2-RDA	D3-CDUR
Cadre d'élaboration de la formulation :			
- équipe	nationale	internationale	internationale
- nb. de personnes	?	21	2
- type de publication	article	preprint	article, avec <i>open peer review</i>
- nb. de références	2	18	22
Qualité de la formulation	faible	faible	forte
Compréhension de l'objet numérique	faible	forte	forte
Compréhension de la production scientifique	forte	forte	forte
Compréhension de l'objet juridique	faible	faible	forte
Réponses aux <i>conundrum challenges</i>	partielles	partielles	complètes
Compréhension de données de la recherche	faible	faible	forte

Fig. 1. Les dix points du cadre de comparaison des trois définitions.

Les trois définitions étudiées (D1-COSO, D2-RDA, D3-CDUR) ont été proposées dans trois contextes différents : le premier correspond à une équipe nationale (en France), mais sans spécifier les auteurs et autrices du texte ; le deuxième correspond à une équipe internationale et il inclut les discussions d'une trentaine d'experts d'horizons divers ; le troisième correspond à une équipe de deux experts avec des expériences complémentaires.

Les équipes du COSO et de RDA sont composées par un grand nombre de membres, qui ont étudié une large bibliographie, mais cette bibliographie n'est pas précisée dans le cas du COSO, qui fait référence à un historique de partage ancien. Deux personnes ont formé l'équipe CDUR, mais l'origine de ce travail se place dans plusieurs contextes de collaboration, et comprennent, entre autres, le laboratoire d'informatique LIGM [37], [38] et le projet national CNRS PLUME [31], [35], [36]. Cette équipe a consulté également une large bibliographie. La définition D3-CDUR semble être la seule qui a bénéficié d'un cadre de comparaison entre deux types de productions scientifiques, les articles et les logiciels.

Ces trois définitions sont diffusées dans des documents librement accessibles. La définition D2-RDA est diffusée dans un preprint disponible sur Zenodo [51], les définitions D1-COSO et D3-CDUR ont fait l'objet de publications scientifiques avec une procédure de *referee* [39], [42], [71]. Les publications de l'équipe CDUR ont fait l'objet d'un *open peer review*. Il nous semble que la qualité de la formulation, la clarté et la précision peuvent être améliorées dans le cas des définitions D1-COSO et D2-RDA.



Comme il a été argumenté dans la section 1, une *bonne définition* de l'objet logiciel de la recherche doit appréhender de façon claire trois aspects importants : la production scientifique et numérique, mais sans oublier le contexte juridique qui apparaît naturellement dans la création des œuvres de l'esprit. Les programmes (ie. le code source) de ces logiciels sont écrits par des personnes, ce qui va générer des droits mais, également, doter cet objet de sa nature de production scientifique. Ce point est correctement pris en considération par les définitions D1-COSO et D3-CDUR, mais peut être amélioré dans D2-RDA. La mention des éléments qui composent le logiciel en tant qu'objet numérique peut être améliorée dans la définition D1-COSO.

La capacité de répondre aux questions du *conundrum challenges* (quel est le logiciel à partager, par qui, avec qui, dans quelles conditions, pourquoi et à quels effets) qui est suscitée par les définitions étudiées nous semble un complément important qui montre la justesse de chaque définition, ainsi que le fondement scientifique solide du travail effectué pour comprendre cet objet et pour éviter les ambiguïtés qui peuvent encore l'entourer. De ce point de vue, il nous semble que les travaux associés à la définition D3-CDUR sont ceux qui sont les plus complets, puisqu'ils bénéficient de ce cadre de comparaison avec la production d'articles dans leur phase initiale [35], [39] mais qui ont été, aussi, poursuivis dans le cadre de comparaison avec les données

de la recherche [41], [42], ce qui a entraîné une étude complète des *conundrum challenges* dans [46].

En résumant, cette étude fournit un cadre de comparaison et une méthodologie qui nous a permis d'analyser en détail le concept de logiciel de la recherche et de comparer trois formulations récemment proposées. Il confirme et consolide notre compréhension de cet objet complexe, ce qui nous permet de réaffirmer la définition D3-CDUR proposée dans [39].

Nous espérons ainsi pouvoir contribuer à la conception, la mise en œuvre et à l'adoption des politiques scientifiques qui sont pertinentes pour cet objet. Ces politiques portent sur la gestion d'un patrimoine scientifique et sur les méthodes de diffusion et d'évaluation qui seront choisies dans le cadre de la science ouverte. Elles portent également sur la conception, la construction et la mise en place des services et des infrastructures qui vont participer à rendre cette production plus visible, accessible et réutilisable [40].

Remerciements.

Nous remercions les équipes du collège Codes sources et logiciels et du groupe de travail RDA FAIR4RS pour avoir inspiré les réflexions présentées ici. La version finale de ce texte a bénéficié des commentaires et corrections de plusieurs personnes que nous remercions chaleureusement.

Références

- [1] K. S. Ackroyd, S. H. Kinder, G. R. Mant et al. 2008. Scientific Software Development at a Research Facility. *IEEE Software* 25, 4: 44-51. <https://doi.org/10.1109/MS.2008.93>.
- [2] T. Aimé. 2010. Guide pratique d'usage des logiciels libres dans les administrations. *DGFiP - Ministère du Budget, des Comptes Publics et de la Reforme de l'État*. <https://doi.org/10.5281/zenodo.7191385>.
- [3] P. Alliez, R. Di Cosmo, B. Guedj et al. 2020. Attributing and Referencing (Research) Software: Best Practices and Outlook From Inria. *Computing in Science & Engineering* 22, 1: 39-52. <https://doi.org/10.1109/MCSE.2019.2949413>.
- [4] J.-L. Archimbaud & T. Gomez-Diaz. 2009. FAQ : licence & copyright pour les développements de logiciels libres de laboratoires de recherche. *Projet PLUME (2007–2013)*. <https://doi.org/10.5281/zenodo.7063146>.
- [5] M. Barker, N. P. Chue Hong, D. S. Katz et al. 2022. Introducing the FAIR Principles for research software. *Scientific Data* 9, 622. <https://doi.org/10.1038/s41597-022-01710-x>.
- [6] M. Barker, J. Cohen, P. Hernández Serrano et al. 2025. Institutional Policy Pathways for Supporting Research Software: Global Trends and Local Practices. *Journal of Open Research Software* 13, 39. <https://doi.org/10.5334/jors.625>.
- [7] I. Blanc Catala, R. Di Cosmo, M. Giraud et al. 2023. Establishing a national research software award [version 1; peer review: 2 approved]. *Open Research Europe* 3, 185. <https://doi.org/10.12688/openreseurope.16069.1>.

- [8] I. Blanc, P. Boulet, G. Courbebaisse et al. 2024. *Production et valorisation des logiciels issus de la recherche publique française*. Comité pour la science ouverte. <https://doi.org/10.52949/47>.
- [9] C. L. Borgman. 2012. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63, 6: 1059-1078. <https://doi.org/10.1002/asi.22634>.
- [10] Cour de cassation. 1986. Assemblée Plénière 83-10.477 du 7 mars 1986. Publié au bulletin. Consulté à l'adresse <https://www.legifrance.gouv.fr/juri/id/JURITEXT000007016934>.
- [11] R. Cerda-Cosme & E. Mendez. 2025. Scoping review on research data in Spain: Academic output and developing trends. *The Journal of Academic Librarianship* 51, 6. <https://doi.org/10.1016/j.acalib.2025.103151>.
- [12] Y. Chen, X. Lu & Y.-F. Chen. 2014. Close connections between open science and open-source software. *PeerJ PrePrints* 2, e235v1. <https://doi.org/10.7287/peerj.preprints.235v1>.
- [13] Z. Chinchilla-Rodríguez, R. Costas, V. Larivière et al. 2022. The relationship between corresponding authorship and author position. <https://doi.org/10.5281/zenodo.6957638>.
- [14] M. Clément-Fontaine. 2024. Le cadre juridique des codes sources et des logiciels en Science ouverte. (Dossier spécial : La science ouverte (partie 1) : Les logiciels et codes sources de recherche). *Revue Lamy Droit de l'immatériel* 210. Consulté à l'adresse <https://hal.science/hal-05009877>.
- [15] M. Clément-Fontaine. 2024. L'accès aux connaissances logiciels : enjeux et perspectives pour l'intelligence artificielle et la Science ouverte. *Propriété industrielle* 4, 4. Consulté à l'adresse <https://hal.science/hal-04884965>.
- [16] M. Clément-Fontaine, R. Di Cosmo, B. Guerry et al. 2019. *Note d'opportunité sur la valorisation des logiciels issus de la recherche*. Comité pour la science ouverte. <https://doi.org/10.52949/17>.
- [17] European Commission. 2024. Living guidelines on the responsible use of generative AI in research. Consulté à l'adresse https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf.
- [18] E. Dayton & K. Rodier. 2024. *Critical Thinking, Logic, and Argument: An Introduction*. (French Google translation available at: https://read-aupress-ca.translate.google.com/read/critical-thinking-logic-and-argument/section/6d3e9906-c0a5-42c4-b841-f69295b75560?_x_tr_sl=en&_x_tr_tl=fr&_x_tr_hl=fr&_x_tr_pto=rq). Athabasca University Press, Athabasca, AB Canada.
- [19] P. Depaz. 2021. The Craft of Code. Consulté à l'adresse <https://hal.science/hal-04365925>.
- [20] P. Depaz. 2023. The role of aesthetics in understanding source code. Université de la Sorbonne nouvelle - Paris III. Consulté à l'adresse <https://theses.hal.science/tel-04588711>.
- [21] R. Di Cosmo, S. Granger, K. Hinsin et al. 2025. Stop treating code like an afterthought: record, share and value it. *Nature* 646. <https://doi.org/10.1038/d41586-025-03196-0>.
- [22] C. Dommange. 2013. *Rapporteur. Guide Data Culture. Pour une stratégie numérique de diffusion et de réutilisation des données publiques numériques du secteur culture*. Ministère de la Culture et de la Communication. Secrétariat Général. Consulté à l'adresse <https://www.enssib.fr/bibliotheque-numerique/documents/60533-guide-data-culture.pdf>.

- [23] Compagnie nationale des experts de justice en informatique et techniques associées (CNEJITA). 2021. L'originalité du logiciel en question, JFC du 13 avril 2021. Consulté à l'adresse <https://www.cnejita.org/jfc-du-13-04-2021-14h-17h-sur-zoom-loriginalite-du-logiciel-en-question>.
- [24] Compagnie nationale des experts de justice en informatique et techniques associées (CNEJITA). 2023. Quelle place reste-t-il pour l'originalité du logiciel? Colloque du 3 avril 2023. Consulté à l'adresse <https://www.cnejita.org/product/colloque-3-avril-2023>.
- [25] M. Farge. 2017. Scholarly publishing and peer-reviewing in open access. In *Europe's Future: Open Innovation, Open Science, Open to the World. Reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group*. European Commission: Directorate-General for Research; Innovation, 73-81. <https://doi.org/10.2777/348700>.
- [26] République française. 2016. Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. Consulté à l'adresse <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>.
- [27] République française. 2022. Loi n° 2022-217 du 21 février 2022. Article 163. Loi relative à la différenciation, la décentralisation, la déconcentration et portant diverses mesures de simplification de l'action publique locale. Consulté à l'adresse http://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000045197616.
- [28] République française. 2021. Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique. Consulté à l'adresse <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>.
- [29] République française. 2021. Ordonnance n° 2021-1658 du 15 décembre 2021 relative à la dévolution des droits de propriété intellectuelle sur les actifs obtenus par des auteurs de logiciels ou inventeurs non-salariés ni agents publics accueillis par une personne morale réalisant de la recherche. Consulté à l'adresse <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000044501327>.
- [30] République française. 2026. Code de la Propriété intellectuelle. (Version en vigueur au 26 mars 2026). Consulté à l'adresse https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEX000006069414.
- [31] T. Gomez-Diaz. 2007–2013. Le thème PLUME « Patrimoine logiciel d'un laboratoire ». (Zenodo Community). *Projet PLUME (2007–2013)*. Consulté à l'adresse <https://zenodo.org/communities/plume-patrimoine-logiciel-laboratoire>.
- [32] T. Gomez-Diaz. 2007–2026. Page web with presentations, articles, posters and other productions since 2007. (Laboratoire d'informatique Gaspard-Monge (LIGM)). Consulté à l'adresse <https://ligm.univ-eiffel.fr/~teresa/logicielsLIGM>.
- [33] T. Gomez-Diaz. 2010. Diffuser un logiciel de laboratoire : recommandations juridiques et administratives. *Projet PLUME (2007–2013)*. <https://doi.org/10.5281/zenodo.7096216>.
- [34] T. Gomez-Diaz. 2014. Free software, Open source software, licenses. A short presentation including a procedure for research software and data dissemination. (Preprint). Presented at the Workshop on open licenses: Data licencing and policies, EGI Conference 2015, Lisbon, May 2015, <https://indico.egi.eu/event/2452/sessions/1522/>. <https://doi.org/10.5281/zenodo.7096216>.
- [35] T. Gomez-Diaz. 2015. Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels. 1024 : *Bulletin de la Société Informatique de France* 5: 119-140. <https://doi.org/10.48556/SIF.1024.5.119>.

- [36] T. Gomez-Diaz. 2019. Le Projet PLUME et le paysage actuel des logiciels de la recherche dans la science ouverte. (Preprint). <https://doi.org/10.5281/zenodo.2591474>.
- [37] T. Gomez-Diaz. 2021. Free/Open source Research Software production at the Gaspard-Monge Computer Science laboratory (LIGM). Lessons learnt. In *FOSDEM'21*. Consulté à l'adresse https://fosdem.org/2021/schedule/event/open_research_gaspard_monge.
- [38] T. Gomez-Diaz. 2021. Sur la production de logiciels libres au laboratoire d'informa-tique Gaspard-Monge (LIGM) : ce que nous avons appris. In *Atelier BlueHats*. <https://doi.org/10.5281/zenodo.10970140>.
- [39] T. Gomez-Diaz & T. Recio. 2019. On the evaluation of research software: the CDUR procedure [version 2; peer review: 2 approved]. *F1000Research* 8, 1353. <https://doi.org/10.12688/f1000research.19994.2>.
- [40] T. Gomez-Diaz & T. Recio. 2020–2021. Towards an Open Science definition as a poli-tical and legal framework: on the sharing and dissemination of research outputs. (Version 3 dated 28 February 2021 available at <https://doi.org/10.5281/zenodo.4577065>). *POLIS* 19. <https://doi.org/10.58944/yuro5734>.
- [41] T. Gomez-Diaz & T. Recio. 2022. Research Software vs. Research Data I: Towards a Research Data definition in the Open Science context [version 2; peer review: 3 appro-ved]. *F1000Research* 11, 118. <https://doi.org/10.12688/f1000research.78195.2>.
- [42] T. Gomez-Diaz & T. Recio. 2022. Research Software vs. Research Data II: Protocols for Research Data dissemination and evaluation in the Open Science context [version 2; peer review: 2 approved]. *F1000Research* 11, 117. <https://doi.org/10.12688/f1000research.78459.2>.
- [43] T. Gomez-Diaz & T. Recio. 2023. How to achieve FAIRER research data by studying evaluation & assessment protocols. (Poster). Open Science FAIR 2023. <https://doi.org/10.5281/zenodo.8398431>.
- [44] T. Gomez-Diaz & T. Recio. 2024. Open comments on the « Open Letter: Establishing a national research software award ». *Open Research Europe*. <https://doi.org/10.12688/openreseurope.16069.1>.
- [45] T. Gomez-Diaz & T. Recio. 2024. Articles, software, data: An Open Science etholo-gical study. *Maple Transactions* 3, 4. <https://doi.org/10.5206/mt.v3i4.17132>.
- [46] T. Gomez-Diaz & T. Recio. 2024. The Conundrum Challenges for Research Software in Open Science. *Computers* 13, 11. <https://doi.org/10.3390/computers13110302>.
- [47] T. Gomez-Diaz & T. Recio. 2025. Perceptions on the adoption of Free/Open Source Software policies by a Scientific Institution. The case study of the NIH. *Maple Tran-sactions* 5, 3. <https://doi.org/10.5206/mt.v5i3.22904>.
- [48] T. Gomez-Diaz & G. Romier. 2018. Research Software Management Plan template, V3.2. PRESOFT project: Preservation for REsearch SOFTware. (Document bilingue (FR/EN)). <https://doi.org/10.5281/zenodo.1405614>.
- [49] R. Gröpler, S. Klepke, J. Johns et al. 2025. The Future of Generative AI in Software Engineering: A Vision from Industry and Academia in the European GENIUS Project. (Preprint). Consulté à l'adresse <https://arxiv.org/abs/2511.01348>.
- [50] Y. Vincent Grossmann, G. Lanza, K. Biernacka et al. 2024. Software Management Plans – Current Concepts, Tools, and Application. *The CODATA Data Science Journal* 23, 43. <https://doi.org/10.5334/dsj-2024-043>.
- [51] M. Gruenpeter, D. S. Katz, A.-L. Lamprecht et al. 2021. Defining Research Software: a controversial discussion. (Preprint). <https://doi.org/10.5281/zenodo.5504016>.

- [52] S. Hettrick. 2016. *Research software sustainability: report on a knowledge exchange workshop*. Knowledge Exchange. Consulté à l'adresse https://repository.jisc.ac.uk/6332/1/Research_Software_Sustainability_Report_on_KE_Workshop_Feb_2016_FINAL.pdf.
- [53] K. Hinsén. 2019. Dealing With Software Collapse. *Computing in Science and Engineering* 21, 3: 104-108. <https://doi.org/10.1109/MCSE.2019.2900945>.
- [54] K. Hinsén. 2025. Reviewing Research Software. *Computing in Science and Engineering* 27, 3: 64-66. <https://doi.org/10.1109/MCSE.2025.3601909>.
- [55] C. Hudelot. 2025. Vers des LLMs ouverts et transparents pour l'Europe : conception, adaptation et explicabilité. In *Assises Nationales des Données de la Recherche (ANDOR)*. Consulté à l'adresse https://andor2025.sciencesconf.org/data/pages/Keynote_LLM_Celine_Hudelot_PPT_Andor.pdf.
- [56] R. C. Jiménez, M. Kuzak, M. Alhamdoosh et al. 2017. Four simple recommendations to encourage best practices in research software [version 1; peer review: 3 approved]. *F1000Research* 6, 876. <https://doi.org/10.12688/f1000research.11407.1>.
- [57] N. Jullien. 2024. La valorisation économique des logiciels scientifiques. (Dossier spécial : La science ouverte (partie 1) : Les logiciels et codes sources de recherche). *Revue Lamy Droit de l'immatériel* 210. Consulté à l'adresse <https://hal.science/hal-04473408>.
- [58] P. Kamocki, T. Bond, K. Lindén et al. 2024. Mind the Ownership Gap? Copyright in AI-generated Language Data. In *CLARIN Annual Conference*. Consulté à l'adresse <https://www.ecp.ep.liu.se/index.php/clarin/article/download/1024/931/1053>.
- [59] D. Kelly. 2024. An Analysis of Process Characteristics for Developing Scientific Software. *Journal of Organizational and End User Computing (JOEUC)* 23, 4. <https://doi.org/10.4018/joeuc.2011100105>.
- [60] X. de La Porte. 2024. France Inter podcast du 7 juin 2024. Le code a changé. « Esthétique du code ». Consulté à l'adresse <https://www.radiofrance.fr/franceinter/podcasts/le-code-a-change/le-code-a-change-11-9775071>.
- [61] V. Larivière, D. Pontille & C. R. Sugimoto. 2021. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies* 2, 1: 111-128. https://doi.org/10.1162/qss_a_00097.
- [62] Md. Ariful Islam Malik, J. C. Carver, & N. U. Eisty. 2025. Peer code review in research software development: The research software engineer perspective. *Empirical Software Engineering* 31, 2. <https://doi.org/10.1007/s10664-025-10785-0>.
- [63] M. G. Santangelo, V. Louvet, S. Chachay et al. 2025. Recommandations Recherche Data Gouv – Collège Codes sources et logiciels pour le questionnaire dédié aux codes sources et logiciels de recherche. Consulté à l'adresse https://dmp.opidor.fr/guidance_group_export/1160.pdf.
- [64] B. Mermet. 2025. IA et développement logiciel : où va-t-on ? 1024 : *Bulletin de la Société Informatique de France* 25: 189-199. <https://doi.org/10.48556/SIF.1024.25.189>.
- [65] Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. 2021. Deuxième Plan national pour la science ouverte. Généraliser la science ouverte en France 2021–2024. Consulté à l'adresse <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525>.
- [66] Ministère de l'Enseignement supérieur et de la Recherche. Collège Codes sources et logiciels. 2022. Le logiciel de recherche, un pilier de la recherche scientifique ouverte. Blog du Comité pour la science ouverte. Consulté à l'adresse <https://www.ouvrirlascience.fr/le-logiciel-de-recherche-un-pilier-de-la-recherche-scientifique-ouverte>.

- [67] Ministère de l'Enseignement supérieur et de la Recherche. Comité pour la science ouverte. 2022. Science ouverte – Codes et logiciels. Consulté à l'adresse <https://www.ouvrirlascience.fr/science-ouverte-codes-et-logiciels>.
- [68] Ministère de l'Industrie. 1981. Arrêté du 22 décembre 1981 du ministère de l'Industrie relatif à l'enrichissement du vocabulaire de l'informatique. Consulté à l'adresse <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000842020>.
- [69] D. Partha & P. A. David. 1994. Toward a new economics of science. (Special Issue in Honor of Nathan Rosenberg). *Research Policy* 23, 5: 487-521. [https://doi.org/10.1016/0048-7333\(94\)01002-1](https://doi.org/10.1016/0048-7333(94)01002-1).
- [70] F. Pellegrini. 2016. L'originalité des œuvres logicielles. *Rapport de recherche RR-8945*. Inria Bordeaux Sud-Ouest, université de Bordeaux. 2016, pp. 13. Consulté à l'adresse <https://inria.hal.science/hal-01352700>.
- [71] F. Pellegrini. 2024. Qu'est-ce qu'un logiciel de recherche? (Dossier spécial : La science ouverte (partie 1) : Les logiciels et codes sources de recherche). *Revue Lamy Droit de l'immatériel* 210. Consulté à l'adresse <https://inria.hal.science/hal-04476872>.
- [72] R. Arikrishna Perumal. 2025. The role of AI tools like ChatGPT and Copilot in revolutionizing software development and user experiences. *International Journal of Advance Research, Ideas and Innovations in Technology* 11, 1: 125-131.
- [73] D. Pontille. 2002. La signature scientifique. Authentification et valeur marchande. *Actes de la Recherche en Sciences Sociales* 2, 141: 72-78. <https://doi.org/10.3917/arss.141.0072>.
- [74] A. S. Pothukuchi, L. V. Kota, & V. Mallikarjunaradhya. 2023. Impact of Generative AI on the Software Development Lifecycle (SDLC). *International Journal of Creative Research Thoughts* 1, 8. Consulté à l'adresse <https://ssrn.com/abstract=4536700>.
- [75] A. Robin. 2024. Exploitation de l'innovation. *Cahiers Droit, Sciences & Technologies* 18. <https://doi.org/10.4000/1271j>.
- [76] B. Saenen. 2025. *Developing and Aligning Policies on Research Software: Recommendations for Research Funding and Research Performing Organisations*. Science Europe. <https://doi.org/10.5281/zenodo.14622355>.
- [77] S. Tonda-Goldstein, M. Clément-Fontaine, N. Jullien et al. 2025. *Quelles stratégies de diffusion et pérennité pour les logiciels de recherche dans les établissements publics français?* Comité pour la science ouverte. <https://doi.org/10.52949/82>.
- [78] UNESCO. 2021. Recommandation de l'UNESCO sur une science ouverte. Consulté à l'adresse https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre.
- [79] M. D. Wilkinson, M. Dumontier, IJ. J. Aalbersberg et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1. <https://doi.org/10.1038/sdata.2016.18>.

