Logiciels libres : à la recherche du bien commun

Edlira Nano, Olivier Langella, Filippo Rusconi

▶ **To cite this version:**

Edlira Nano, Olivier Langella, Filippo Rusconi. Logiciels libres : à la recherche du bien commun. Journées Réseaux de l'Enseignement et de la Recherche JRES 2022, May 2022, Marseille, France. hal-03663215

HAL Id: hal-03663215
https://hal.science/hal-03663215

Submitted on 18 May 2022

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

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

Logiciels libres: à la recherche du bien commun

Edlira Nano
Informaticienne indépendante – Associations April, La Quadrature Du Net
eda@mutu.net / eda@laquadrature.net

Olivier Langella
Ingénieur au CNRS – Plateforme PAPPSO - Laboratoire GQE-Le Moulon
Ferme du Moulon 91190 Gif-sur-Yvette France
olivier.langella@universite-paris-saclay.fr

Filippo Rusconi
Chercheur au CNRS – Plateforme PAPPSO - Laboratoire GQE-Le Moulon
Ferme du Moulon 91190 Gif-sur-Yvette France
filippo.rusconi@universite-paris-saclay.fr

Résumé

Le logiciel libre et la recherche publique partagent un objectif : le bien commun, au service de tous. Cette présentation revient sur ce que sont le bien commun et la science ouverte pour essayer de les analyser à l'aide de l'exemple de la Plateforme scientifique Analyses Protéomiques de Paris Sud-Ouest (PAPPSO).

PAPPSO s'est dotée d'une infrastructure informatique complète basée exclusivement sur du logiciel libre : réseau, serveurs, stockage, calcul et postes personnels. Elle développe plusieurs logiciels scientifiques sous licence libre, dont ceux qui forment la chaîne de traitement des données de spectrométrie de masse. Ce choix naturel facilite la reproductibilité des traitements, apporte la maîtrise des logiciels et permet l'intégration de code source tiers eux-mêmes publiés sous licence libre.

Historiquement, l'apport majeur du logiciel libre à la recherche publique a d'abord résidé dans la mise en place de systèmes informatiques complexes tirant au mieux parti des réseaux. Progressivement, le logiciel libre s'est implanté avec force dans les laboratoires. En effet, d'un point de vue scientifique, l'ouverture du code source et les licences libres garantissent une plus grande réutilisation du code source et sa vérification par les pairs; la science gagnait ainsi en fiabilité. Cet élan a débordé le cadre strict du code source des logiciels pour irriguer le secteur stratégique des formats de fichier. C'est bien le monde du logiciel libre qui a contraint les éditeurs de logiciels à ouvrir les formats de leurs fichiers, jusqu'alors conservés secrets comme « arme » commerciale redoutable. De manière intéressante, l'impact a été encore plus profond, avec la prise de conscience que les données aussi devaient être ouvertes. Code source libre, formats ouverts et données publiques forment un bien commun qui est protégé par des licences. De nombreux laboratoires y contribuent en utilisant ou en produisant des logiciels libres, ou encore en publiant les données en accès libre, comme en témoignent la forge du code source du secteur public (https://code.gouv.fr) et les actions ministérielles spécifiques (https://www.ouvrirlascience.fr).

Comment contribuer efficacement ? Quelles sont les recommandations et obligations pour les établissements publics ? Quelle licence choisir ? Comment une licence copyleft *peut-elle aussi séduire les partenaires privés ? Nous apporterons des réponses et des éléments de réflexion pour corriger quelques fausses croyances et promouvoir la construction collective d'une culture libre, au service du bien commun.*

Mots-clefs

Logiciel libre, logiciel scientifique, biens communs, science ouverte, open source, copyleft, *protéomique, infrastructure*

Une plateforme pour la recherche sous logiciel libre

Contexte scientifique

La protéomique est l'étude de l'ensemble des protéines produites par une cellule, un organe ou un organisme. Elle permet d'analyser sans *a priori* les changements qualitatifs ou quantitatifs dans la composition en protéines en fonction de traitements, du développement ou de variations génétiques, et de comprendre par exemple comment un organisme répond à un stress ou à une maladie.

L'apparition du terme « protéomique » en 1995 coïncide avec le début de l'utilisation de la spectrométrie de masse pour l'identification des protéines puis pour leur quantification. Aujourd'hui la méthode utilisée quasi exclusivement est l'analyse de peptides issus de la digestion triptyque des protéines, par un spectromètre de masse couplé à une chaîne de chromatographie liquide. Les appareils n'ont pas cessé d'évoluer mais au final il s'agit toujours de mesurer la masse moléculaire des peptides ayant pénétré dans le spectromètre et/ou de leurs fragments et de quantifier leur intensité.

Les analyses par spectrométrie de masse produisent de grandes quantités de données, directement en sortie de l'instrument. Chaque fabricant a son propre format de données, qui est le plus souvent propriétaire. Ordinairement, les données aux formats propriétaires sont traitées par les logiciels fournis avec l'instrument, eux aussi propriétaires. Comme souvent dans le domaine des formats de fichiers, ces formats ne sont pas pérennes et les licences d'utilisation des logiciels propriétaires sont très coûteuses. Il y a donc ici un problème majeur d'interopérabilité.

Depuis 2005, la plateforme scientifique « Analyses protéomiques de Paris Sud-Ouest » (PAPPSO) a fait le choix du logiciel libre pour garantir la pérennité de ses chaînes de traitement, la reproductibilité des expériences et la capitalisation de son savoir-faire.

Transition des logiciels propriétaires au logiciel libre

Le passage progressif au logiciel libre a permis une rationalisation de l'utilisation des ressources informatiques. Nous sommes passés de postes dédiés à licence unique pour usage unique à une infrastructure collective combinant stockage, calcul et enchaînement des traitements depuis n'importe quel poste de travail.

Durant la période 2005–2015, notre travail a été facilité par la définition de formats standards en protéomique, l'émergence de nombreux logiciels libres dans le domaine [1] et le développement de nouvelles solutions logicielles sous licence libre (PROTICdb [2], mineXpert2 [3], X!TandemPipeline [4], MassChroQ [5]).

Choix du système d'exploitation

Le choix de PAPPSO[1] s'est porté d'abord sur la distribution GNU/Linux Ubuntu, puis sur Debian. Un groupe de développeurs officiels Debian, dont un auteur de ce rapport, s'attache à fournir dans la distribution de nombreux logiciels pour la chimie, et en particulier pour la spectrométrie de masse (*team debichem*). L'intérêt principal de la distribution Debian est la richesse de son offre logicielle qui permet de disposer d'un socle de fonctionnalités robuste couvrant les exigences « serveur » et les impératifs « bureautique ». Tous les logiciels développés par PAPPSO sont disponibles sous

1 http://pappso.inrae.fr/

forme de paquets Debian dans des dépôts publics. Le déploiement des logiciels, dépendances comprises, sur les serveurs et les postes de travail de l'équipe est ainsi simplifié et totalement automatisé.

Stockage des données

Les besoins en stockage de la plateforme évoluent constamment en fonction des progrès techniques des spectromètres de masse. Chaque nouvelle génération d'instruments apporte des améliorations, en particulier sur la précision de mesure de masse, qui provoquent une augmentation significative du volume des données générées. Cette progression des besoins en volume de stockage sur les quinze dernières années est illustrée dans la figure ci-dessous.

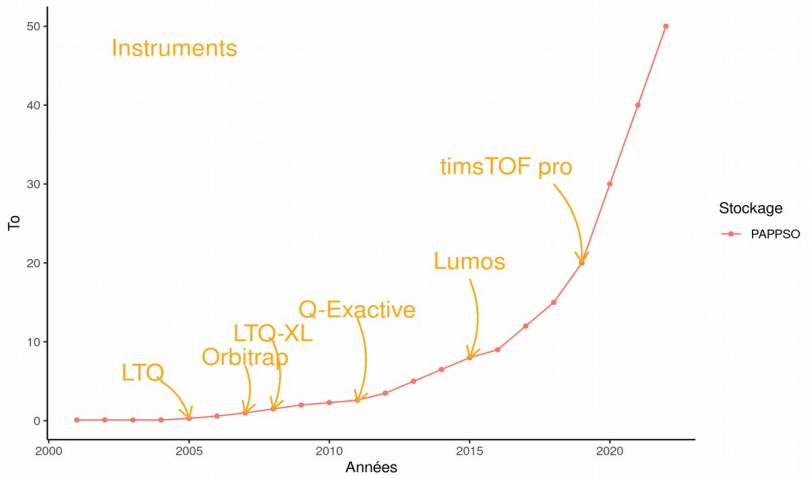

Figure 1 – *Évolution des besoins de stockage de la plateforme en To/an, en relation avec la production des données par les générations d'instruments*

Le système de stockage des données de spectrométrie de masse doit permettre une adaptation en continu de la volumétrie d'espace disque disponible ainsi que les meilleures performances en lecture et écriture. Les solutions classiques de type NAS ont été écartées pour éviter la dépendance matérielle et les problèmes liés au renouvellement des équipements.

Dès 2011, nous avons été parmi les premiers à faire confiance à une solution nouvelle de stockage distribué : Ceph[2]. La principale caractéristique de ce système de stockage est de ne requérir que des serveurs standard. La flexibilité et l'adaptabilité de ce système à des besoins perpétuellement en évolution en ont fait la solution la plus robuste que nous connaissions.

Calcul scientifique

La plateforme PAPPSO est spécialisée dans les traitements en protéomique haut débit (nombreux échantillons à traiter dans les plus brefs délais). Pour assurer la disponibilité de nos moyens de calcul à l'ensemble des utilisateurs, nous utilisons le gestionnaire de processus HTCondor[3].

Les besoins en calcul évoluent eux aussi en fonction des instruments utilisés. Avec l'évolution des techniques, de nouvelles possibilités sont apparues dans le traitement de données en protéomique, exigeant elles aussi des capacités de calcul supplémentaires. De la même manière que pour les capa-

2 https://ceph.io/

3 https://htcondor.org/

cités de stockage, les machines dédiées au calcul doivent être ainsi renouvelées régulièrement et intégrées au fur et à mesure.

Retour d'expérience sur 10 ans

Matériel

L'intégration de nouvelles machines de calcul ou de stockage s'est faite de manière transparente. Nous sommes passés d'une capacité de stockage initiale de 18 To (3 serveurs R515, disques de 3 To) en 2011 à une capacité de 917 To (8 serveurs hétérogènes). Le réseau est passé du 1 Gb cuivre au 10 Gb SFP+. Il n'y a pas eu de transfert de données/migration, pas de modification de l'architecture logique pour les utilisateurs. Le système de fichiers *cephfs* permet un accès direct aux données depuis chaque nœud de calcul. Globalement, les performances ont suivi les évolutions matérielles (augmentation du débit, augmentation des capacités de calcul). La résistance aux pannes a été mise à rude épreuve (panne électrique, disques ou erreurs humaines) et nous n'avons jamais eu de perte de données.

Logiciel

Les systèmes pour les serveurs et pour les postes utilisateurs ont été migrés en 2013 de Ubuntu GNU/Linux vers Debian. Nous y avons gagné en stabilité et en simplicité lors des mises à jour de version. La stratégie consiste à maintenir le parc informatique sous Debian « stable » et effectuer le passage à la version successive dans les mois qui suivent sa publication officielle. Le stockage centralisé est disponible pour tous les postes dans une arborescence commune, via un montage automatique sur les nœuds de calculs (*cephfs* via *systemd* sur les serveurs, *sshfs* sur les postes clients). Les logiciels sont les mêmes sur les serveurs et les postes utilisateurs. L'accès distant au cluster de calcul se fait avec *x2go*, via une clé publique SSH.

Scientifique

Les analyses de la plateforme ont évolué pour passer de la technique chronophage des gels d'électrophorèse 2D[4] vers des analyses uniquement basées sur la spectrométrie de masse. Le traitement des images des gels 2D était majoritairement effectué avec des logiciels propriétaires sous Windows, sur des postes dédiés, ce qui limitait les capacités de traitement. Le passage progressif à des processus analytiques qui faisaient l'économie de l'étape d'électrophorèse a coïncidé avec les apports massifs du logiciel libre dans le domaine scientifique, au milieu des années 2000. Nous avons alors pu commencer la transition vers certains logiciels libres qui émergeaient à cette époque. Cependant, ces logiciels étaient principalement des bibliothèques partagées encore imparfaitement dotées des fonctionnalités requises dans notre domaine. Nous avons alors entrepris le développement de nos logiciels sur la base des besoins scientifiques particuliers à notre plateforme. Le logiciel MassChroQ est né ainsi, de nos besoins en protéomique quantitative. Notre indépendance vis-à-vis des formats de données propriétaires des fabricants nous a permis de produire un logiciel évolutif et pérenne dès le départ, évitant l'effet « boîte noire ». Ainsi, notre offre logicielle a pu être adaptée au fur et à mesure aux nouvelles techniques, à des instruments significativement différents de génération en génération, absorbant ainsi les « chocs » technologiques : doublement des fréquences d'acquisition à chaque génération (3 ans), doublement du pouvoir résolutif (précision des mesures des masses).

La dernière « rupture technologique » a été l'apparition du *timsTOF Pro* du fabricant Bruker. Cet appareil dispose d'une fréquence d'acquisition 10 fois supérieure à celle de la génération précédente

4 Technique visant à séparer les nombreuses protéines d'un échantillon complexe en fonction de leur poids moléculaire, dans l'épaisseur d'un feuillet qui agit comme un tamis moléculaire.

ainsi que d'une nouvelle dimension de séparation des peptides (mobilité ionique). Fait unique depuis les débuts de la protéomique, Bruker nous a confié les spécifications techniques de son format de fichier de données. Cela nous a permis d'adapter MassChroQ pour pouvoir utiliser de manière native les données obtenues sur cet instrument. Notre savoir-faire en développement C++ nous a ainsi permis d'obtenir des gains de performances remarquables par rapport aux logiciels commerciaux, ainsi que de meilleurs résultats scientifiques.

Importance de la culture libre dans la recherche scientifique

Culture libre, communs et recherche publique

La recherche publique vise à produire et à développer des connaissances, à les améliorer sans cesse, puis à les rendre disponibles à la société, au cours des générations, constituant ainsi un corpus de savoirs qui caractérisent notre civilisation. La recherche scientifique est par essence incrémentale, elle se nourrit du savoir établi, pour venir à son tour nourrir cet ensemble commun de savoirs. Pour ce faire, elle a besoin d'accéder aux connaissances établies sans entrave, puis de pouvoir à son tour faire l'objet d'une diffusion libre.

Nous avons là les caractéristiques d'un logiciel libre ou, plus largement, d'un bien libre : quelque chose qu'on peut utiliser librement, distribuer librement autour de nous, mais aussi en modifier le contenu, pour l'adapter ou l'améliorer, pour ensuite le redistribuer tout aussi librement. La culture du libre n'est pas unique aux logiciels. Elle peut concerner du code source, un algorithme, un service informatique en ligne, un document (article scientifique, livre, poème), un produit artistique (œuvre numérique, reproduction numérique d'une œuvre physique d'un musée, photo numérique d'un monument historique), une semence agricole (les graines libres de droit), un médicament ou un vaccin que l'on peut fabriquer librement (libre de brevet).

Cette culture du libre est inhérente à la recherche publique également, et plus largement à ce qu'on appelle les communs. Les communs sont des ressources, naturelles ou culturelles, partagées et gérées collectivement, accessibles et disponibles pour tous, et qui n'appartiennent pas à qui que ce soit au sens de la propriété, qu'elle soit publique ou privée (Elinor Ostrom, prix Nobel d'économie pour ses travaux sur les communs, parle par exemple de « ressources de propriété commune » [6]).

La recherche publique, en ce qu'elle vise à améliorer, à rendre accessibles et disponibles les savoirs communs, peut ainsi être vue comme un processus cumulatif et collectif dont les résultats constituent un « bien commun » [7] .

Vers une science ouverte

La recherche publique en France (voir Figure ci-dessous, sources 2018) est financée à 75 % par des fonds publics français (MIRES, hors MIRES, administrations), et parmi les 25 % restants on ne retrouve que 5 % de ressources contractuelles provenant d'entreprises, les autres 20 % sont un mélange de fonds publics étrangers, notamment européens, ressources propres tels que les prestations de services des structures de recherche elles-mêmes, et enfin des ressources privées provenant d'entreprises étrangères) [8].

En dépit de la très faible proportion de financements privés, la recherche publique fait face à la mise en place progressive de mécanismes de privatisation et de verrouillage de ses productions par des acteurs privés de plus en plus envahissants, malheureusement soutenus par des politiques publiques de plus en plus tournées vers les transferts de technologies à tout prix et la valorisation, de type brevet entre autres. Une autre difficulté est liée au financement sur projet de la recherche publique, à un niveau insuffisant du financement public face à l'augmentation du nombre d'étudiants, au recours massif à des contrats temporaires, qui perturbent grandement son fonctionnement (voir le rapport

2019 du Comité national du CNRS [9], qui souhaite « un soutien public mû par la volonté de faire progresser les connaissances ».

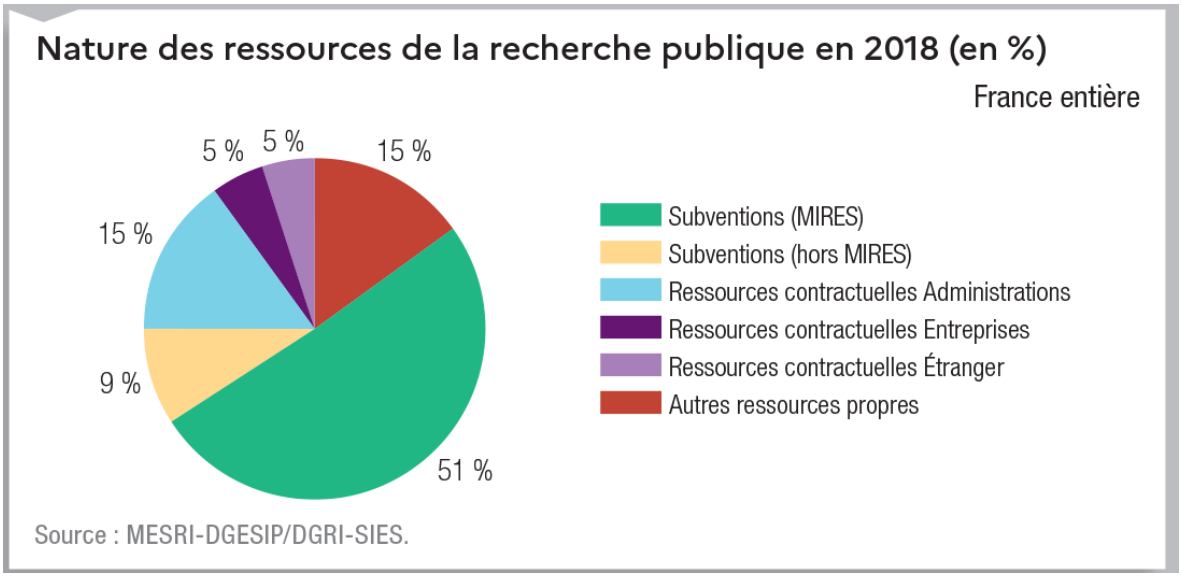

Figure 2 – Nature des ressources de la recherche publique en 2018 (en %)

La culture libre, au service du savoir commun, que la recherche publique incarne et qui la structure, est illustrée par les nombreux mouvements pour une science ouverte, mouvements qui ont souvent vu le jour au sein des organismes de recherche. Ce sont des chercheurs, des étudiants ou des bibliothécaires, confrontés sans doute à cette nécessité d'ouverture, qui œuvrent, parfois leur vie durant, pour faire avancer l'ouverture des productions scientifiques. Ce n'est qu'après ces premières impulsions venant du terrain, qu'on voit s'opérer petit à petit une institutionnalisation progressive de cette ouverture.

Ainsi le mouvement de l'« *open access* » en est un bel exemple. ArXiv[5] a été l'une des premières archives ouvertes de publications scientifiques, créée dès 1991 par le physicien Paul Ginsparg et maintenue par des laboratoires de recherche. L'équivalent français de ArXiv, HAL[6], a été créé en 2001 par des chercheurs et s'est institutionnalisé au cours des années suivantes. Avant HAL, les travaux précurseurs de la bibliothécaire de l'INRAE Hélène Bosc sont également remarquables [10]. Témoignant de l'institutionnalisation progressive de la démarche, la plateforme de publications ouvertes orientée sciences humaines et sociales Open Edition[7] a vu, elle, le jour en 2011 en France.

Du côté des hacktivistes, en cette même année 2011, naissait le projet de captation et de diffusion d'articles scientifiques Sci-Hub[8]. Initié par Alexandra Elbakyan, une étudiante chercheuse, ce projet consiste en un logiciel de captation d'articles scientifiques (y compris ceux fermés dans les plateformes de publications propriétaires), qui sont ensuite diffusés sur des portails web décentralisés et autonomes. Sci-Hub témoigne de cette culture de hacktivisme, dont les projets sont conçus de manière à être résilients aux pannes, aux fermetures et interdictions administratives, de par leur caractère décentralisé : chaque portail d'accès est géré de façon indépendante, mettant en commun les données en mode miroir ; si un portail ferme, d'autres miroirs sont toujours là pour assurer l'accès aux données. Quelques années auparavant, Aaron Swartz, un hacktiviste important de la culture

5 https://fr.wikipedia.org/wiki/ArXiv

6 https://fr.wikipedia.org/wiki/HAL_(archive_ouverte)

7 https://www.openedition.org/

8 https://fr.wikipedia.org/wiki/Sci-Hub

libre, jouait aussi un rôle déterminant dès 2008 avec ses multiples actions de mise en place d'un portail d'accès aux documents publics administratifs d'abord, puis d'accès aux publications scientifiques. Swartz est également à l'origine de la création des licences libres *Creative Commons*, du projet *Open Library*, une des nombreuses fonctionnalités de l'*Internet Archive*, ce multi-projet d'archivage du web. Tout comme Alexandra Elbakyan, il a fait l'objet de poursuites pénales et de pressions administratives conséquentes. Tous deux ont pourtant joué un rôle déterminant de lanceurs d'alerte dans le domaine de l'ouverture des données publiques et de la science ouverte, termes ces derniers qui seront, quelques années après leur condamnation, tant repris dans les discours et les orientations affichées des politiques publiques.

Vers des codes sources sous licences libres

En matière de code source, le combat s'est également déroulé en grande partie dans le milieu de la recherche et par des hacktivistes d'un internet et d'une informatique libres. La *Free Software Foundation* a été créée en 1985 sous l'impulsion du chercheur Richard Stallman, pour tenter de donner un cadre légal de protection et de promotion des logiciels libres. C'est ainsi que la plus emblématique des licences libres dites *copyleft* a vu le jour : la GNU *General Public Licence* (GPL). Cette licence va non seulement introduire les quatre libertés fondamentales qui caractérisent le logiciel libre[9], mais elle va également s'assurer de la pérennité de cette liberté à travers la notion de *copyleft*, également appelée *share-alik*e (le « SA » des licences *Creative Commons*). Le *copyleft* va permettre à un code libre de ne pas se retrouver enfermé dans un code propriétaire non libre au cours de son existence et va ainsi assurer la persistance du caractère libre du code à travers les modifications et les transformations de celui-ci. C'est là un principe protecteur très fort, qui revendique la centralité de la notion éthique de liberté dans l'idéologie et la culture du Libre. Ainsi, le principe de *copyleft* est-il fondamental pour éviter que du code ouvert ne soit enfermé dans du code source propriétaire, comme il advient de plus en plus dans les grandes sociétés privées monopolistiques et privatrices telles que les GAFAM (Google, Amazon, Facebook, Apple et Microsoft).

En 1995, l'*Open Source Initiative* présente la définition de logiciel *Open Source*. Cette définition peu connue en détails, et employée à tort pour désigner tout code ouvert dans l'acception de « accessible », renferme en réalité dix conditions[10] qui assurent entre autres les quatre libertés fondamentales des logiciels libres. Les deux types de licences, libres et open source vont coïncider la plupart du temps, mais le terme de logiciel libre est préféré par ceux qui souhaitent mettre l'accent sur la notion de liberté ainsi que sur le caractère idéologique et politique qu'elle revêt.

En France, l'April est la plus grande association de défense et de promotion les logiciels libres, qui était nommée ainsi au départ en tant qu'acronyme d'Association Pour la Recherche en Informatique Libre. Notons également deux autres grandes associations dans le domaine : Framasoft d'une part, qui met l'accent sur l'éducation populaire via la culture libre au sens général et d'autre part, La Quadrature du Net, qui focalise ses actions sur les libertés numériques. Ces associations jouent un rôle important dans le plaidoyer, l'activisme, et la défense de la culture libre.

Notons également une ressource européenne importante que l'on doit à la campagne « *Public Money, Public Code* » portée par la *Free Software Foundation* Europe, qui résume dans sa brochure les bénéfices et apports multiples de l'ouverture du code source produit et utilisé dans le cadre de la recherche publique [12].

9 https://www.gnu.org/philosophy/

10 https://opensource.org

Vers des obligations légales, mais des pratiques en demi-teinte

En termes de politiques publiques, les expressions science ouverte et données ouvertes sont progressivement entrées dans le vocabulaire et dans les plans des gouvernements et tutelles de la recherche. En France, l'obligation légale s'instaure enfin en 2016 à l'occasion du Plan pour une République numérique[11]. Depuis cette date, les données de la recherche, y compris le code source des logiciels produits, sont considérées comme des données administratives publiques et sont soumises à ce titre à la même obligation de publication que ces dernières, au titre du Code des Relations entre le Public et l'Administration. Ce plan sera renommé en 2018 Plan national pour la science ouverte et porte désormais comme sous-inscription : « les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement ».

Les logiciels libres et *open source* se retrouvent aujourd'hui partout autour de nous, et cela bien souvent de manière non visible : les serveurs, les sites web, les objets intelligents, les services dans le *cloud*, reposent en grande majorité sur des logiciels et des systèmes d'exploitation libres. Ce n'est pas le cas des environnements personnels de travail, ni des services utilisés au quotidien sur ces environnements, dominés et monopolisés par les GAFAM qui proposent des outils la plupart du temps non libres et soumis à des restrictions d'usage diverses et variées (coût élevé, code source fermé, usage unique, DRM[12] intégrés, captation de données personnelles en échange de service, obsolescence, etc.). Ce n'est pas non plus le cas des environnements de travail de l'administration publique, ni de celui des millions d'élèves et d'étudiants qui sont confrontés de plus en plus tôt à un environnement et des services informatiques dominés par les contrats publics passés avec les GAFAM et autres sociétés aux pratiques privatrices et liberticides (citons par exemple le partenariat en 2015 entre Microsoft et l'Éducation Nationale[13], ou encore les tablettes numériques pour tous au collège sous système Windows de Microsoft).

Apports du libre

L'exemple de la plateforme PAPPSO ci-dessus nous démontre l'intérêt scientifique à libérer ses codes sources et son infrastructure de travail. Le logiciel libre et les infrastructures libres en milieu de recherche permettent et facilitent :

- un niveau de technicité et d'autonomie incomparable avec celui des logiciels propriétaires : le paramétrage fin des logiciels libres, leur adaptabilité et l'autonomie dans la réalisation de ces adaptations, l'émulation positive entre développeurs ainsi que l'auto-apprentissage (le code source ouvert et libre est un outil didactique en soi) ;
- une meilleure productivité / efficacité : le mode de développement et de production scientifique libre est agile par construction : les besoins de la recherche peuvent être rapidement assouvis par le paramétrage et l'adaptation du code source, ou par la réutilisation de code source libre tiers. Ce code source bénéficie à son tour de la recherche scientifique pour s'enrichir de nouvelles fonctionnalités lors de cycles développement, test et mise en production courts et nombreux ;
- une meilleure collaboration grâce au caractère ouvert du code source et aux libertés conférées par les licences d'utilisation ;un premier pas indispensable vers la sécurité logicielle : l'accès

11 Loi pour une République numérique.

12 « Digital Rights Management » (gestion [restrictive] des droits numériques, principalement des droits d'auteur, contre laquelle lutte la campagne « Defective by design »).

13 https://www.april.org/microsoft-educ-nat-partenariat-indigne/

au code source permet un audit de sécurité par la communauté et ne nécessite pas de croire sur parole des fabricants aux conflits d'intérêts évidents ;une meilleure reproductibilité des résultats scientifiques : nul besoin de devoir acquérir les droits d'utilisation pour pouvoir reproduire des résultats obtenus avec des logiciels libres ;

- une meilleure interopérabilité : les formats de données des logiciels libres sont ouverts et souvent standardisés. Les logiciels propriétaires reposent souvent sur des formats de données fermés et propres à chaque éditeur, rendant l'interopérabilité impossible ou difficile. Cela provoque des achats forcés et une obsolescence matérielle et logicielle qui a un coût économique et écologique non négligeable.

Le libre en pratique

Etalab est le département de la direction interministérielle du numérique (DINUM) qui coordonne la politique d'ouverture et de partage des données publiques des administrations de l'État. Ce département apporte son appui aux administrations pour leur permettre de répondre aux obligations légales d'ouverture. Il propose et maintient notamment :

- la plateforme https://code.gouv.fr/ qui recense aujourd'hui les dépôts et le code source des logiciels du secteur public et qui invite ceux qui n'y sont pas répertoriés à s'y ajouter ;
- la plateforme https://data.gouv.fr de diffusion de données publiques ;
- le SILL : Socle Interministériel de Logiciels Libres (https://sill.etalab.gouv.fr), un catalogue qui référence les logiciels libres recommandés pour les administrations ;
- un guide légal des obligations des publications administratives : https://guides.etalab.gouv.fr/algorithmes/guide/;
- un accompagnement sur mesure pour assurer les obligations d'ouverture des données ;
- un groupe de travail logiciels libres sur la plateforme https://communs.numerique.gouv.fr/ du Plan d'action logiciels libres et communs numériques. Notons en particulier la gazette et les ateliers vidéo BlueHats[14].

Le portail https://www.ouvrirlascience.fr/ mis en place par le Comité pour la science ouverte du MESRI, propose lui aussi différents guides et recommandations.

Le portail du CNRS pour la science ouverte (https://www.science-ouverte.cnrs.fr/) référence également des dépôts publics de code source, de publications scientifiques et de données de la recherche.

Notons également l'initiative publique *Software Heritage*[15], qui vise à créer une archive universelle des logiciels à laquelle vous pouvez vous-même contribuer en y inscrivant votre code source.

Enfin, l'initiative REUSE[16] de la *Free Software Foundation* Europe vise à faciliter, voire à automatiser la mise sous licence libre de logiciels qui comportent différentes briques de code source, chacune portant potentiellement une licence libre différente. REUSE calcule l'intercompatibilité de ces licences et permet ainsi de les combiner pour une mise sous licence du logiciel complet.

14 https://communs.numerique.gouv.fr/bluehats/

15 https://www.softwareheritage.org

16 https://reuse.software/

Bilan

Le passage au logiciel libre pour tous les besoins informatiques de la plateforme PAPPSO a permis une maîtrise totale de ses outils, depuis la production des données brutes jusqu'à l'interprétation biologique. Les sommes importantes économisées en licences de logiciels propriétaires (20 k€ par an) ont été investies dans la maintenance des ressources de calcul et de stockage. Toutes les analyses sont complètement vérifiables et reproductibles, les logiciels étant tous librement téléchargeables, sous licence GPLv3+, sans demande préalable. Le savoir-faire développé par PAPPSO dans l'analyse protéomique à haut débit est reconnu au niveau international (129 articles citant MassChroQ depuis 2011, publication d'un article de référence en métaprotéomique [11]).

L'exemple de la plateforme PAPPSO n'est pas unique, de multiples autres initiatives locales similaires ont lieu constamment. L'initiative *apps.education.fr*[17], cette plateforme de services libres et éthiques pour travailler à distance née à l'occasion de la crise pandémique que l'Education Nationale est en train de pérenniser et d'étendre, en est un superbe exemple. Le monde de la recherche a toujours joué un rôle de précurseur, se positionnant comme un acteur primordial dans le long processus d'ouverture de la science, des formats de données, des données publiques et de la préservation des savoirs communs. Aidé par les mouvements militants et activistes, le chemin parcouru est considérable. Mais les politiques publiques de financement et d'orientation de la recherche semblent contradictoires et insuffisantes. Les récentes avancées comme la remise « prix science ouverte des données de recherche et du logiciel libre de la recherche » seront peut-être l'occasion de prendre conscience de la richesse de nos productions et de favoriser la cohérence entre les discours et les actes.

Bibliographie

[1] Rusconi F. Free Open Source Software for Protein and Peptide Mass Spectrometry- based Science. Curr Protein Pept Sci, 2 (22) 134-147, 2021 ; https://doi.org/10.2174/1389203722666210118160946

[2] Langella O. , Valot B., Jacob D., Balliau T., Flores R., Hoogland C., Joets J., Zivy M.. (2013) Management and dissemination of MS proteomic data with PROTICdb: Example of a quantitative comparison between methods of protein extraction. Proteomics, 9 (13) 1457-66

[3] Langella O, Rusconi F. mineXpert2: Full-Depth Visualization and Exploration of MSn Mass Spectrometry Data. J. Am. Soc. Mass Spectrom., 4 (32) 1138-114, mars 2021 ; https://doi.org/10.1021/jasms.0c00402

[4] Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, Zivy M. X!TandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. J. Proteome Res., 2 (16) 494-503, décembre 2016 ; https://doi.org/10.1021/acs.jproteome.6b00632

[5] Valot B, Langella O, Nano E, Zivy M. MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics, 17 (11) 3572-3577, juin 2011 ; https://doi.org/10.1002/pmic.201100120

[6] Elinor Ostirom. Governing the Commons : the evolution of institutions for collective actions, 1990, Cambridge University Press ; https://archive.org/details/ElinorOstromGoverningTheCommons

[7] Maurel Lionel. L'ouverture des données de recherche : un retour aux sources de l'Ethos de la Science ? Colloque Intégrité Scientifique et Science Ouverte, avril 2019 ; https://scinfolex.com/2019/06/05/louverture-des-donnees-de-recherche-un-retour-aux-sources-de-lethos-de-la-science

17 https://apps.education.fr/

[8] MESRI, L’état de l'Enseignement supérieur, de la Recherche et de l'Innovation en France n°14, 2021 ; https://publication.enseignementsup-recherche.gouv.fr/eesr

[9] CNRS. La recherche publique en France en 2019 : diagnostic et propositions du Comité national, 2019; https://www.cnrs.fr/comitenational/Actualites/Propositions_Comite-national_Juillet-2019.pdf

[10] Bosc H, Poynder R. The Open Access Interviews, 2009 ; https://www.richardpoynder.co.uk/Helene_Bosc_Interview.pdf

[11] Van Den Bossche T, Kunath BJ, Schallert K, Schäpe SS, Abraham PE, Armengaud J, Arntzen M, Bassignani A, Benndorf D, Fuchs S, Giannone RJ, Griffin TJ, Hagen LH, Halder R, Henry C, Hettich RL, Heyer R, Jagtap P, Jehmlich N, Jensen M, Juste C, Kleiner M, Langella O, Lehmann T, Leith E, May P, Mesuere B, Miotello G, Peters SL, Pible O, Queiros PT, Reichl U, Renard BY, Schiebenhoefer H, Sczyrba A, Tanca A, Trappe K, Trezzi JP, Uzzau S, Verschaffelt P, von Bergen M, Wilmes P, Wolf M, Martens L, Muth T. Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. Nature Communiations, 1 (12) 7305, décembre 2021 ; https://doi.org/10.1038/s41467-021-27542-8

[12] FSFE. Public Money Public Code : Modernising Public Infrastructure with Free Software, 2019; https://download.fsfe.org/campaigns/pmpc/PMPC-Modernising-with-Free-Software.pdf