



HAL
open science

20 ans de science ouverte au laboratoire LEGI: Quels succès? Quelles difficultés? Quels enseignements? Un tremplin vers le futur

Cyrille Bonamy, Gabriel Pierre André Moreau, Julien Chauchat, Joël Sommeria

► To cite this version:

Cyrille Bonamy, Gabriel Pierre André Moreau, Julien Chauchat, Joël Sommeria. 20 ans de science ouverte au laboratoire LEGI: Quels succès? Quelles difficultés? Quels enseignements? Un tremplin vers le futur. Congrès JRES: Les Journées Réseaux de l'Enseignement et de la Recherche, RENATER, May 2022, Marseille, France. hal-03608867

HAL Id: hal-03608867

<https://hal.science/hal-03608867v1>

Submitted on 8 Apr 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.

20 ans de science ouverte au laboratoire LEGI : Quels succès ? Quelles difficultés ? Quels enseignements ? Un tremplin vers le futur

Cyrille Bonamy

Gabriel Moreau

Julien Chauchat

Joël Sommeria

Laboratoire LEGI – UMR5519

1209-1211 rue de la piscine

Domaine Universitaire

38 400 Saint-Martin-d'Hères

Résumé

La science ouverte est couramment discutée au sein de nos institutions. Des plans nationaux ont été initiés par le gouvernement et cette problématique soulève énormément de questionnement. La science ouverte implique plusieurs principes : transparence, reproductibilité, efficacité, confiance, sobriété, pérennité. Son application passe ainsi par l'ouverture des données, mais aussi l'ouverture des logiciels et le libre accès aux publications scientifiques. Bien que vécu parfois comme une contrainte, il apparaît que dans la majorité des situations, son application au sein des organismes publics ne peut être que bénéfique, que ce soit pour les initiateurs de la dynamique, les organismes financeurs ou plus généralement nos communautés.

Au laboratoire LEGI, cette stratégie scientifique a été initiée depuis plus de deux décennies. L'objectif initial était de partager les données expérimentales obtenues dans la plateforme CORIOLIS, grand instrument de recherche du CNRS, en respectant les principes FAIR (Findable, Accessible, Interoperable, Reusable). Puis au fil du temps, une partie des experts en calcul scientifique de l'unité ont suivis le mouvement, en partageant des données de modélisation, mais aussi des logiciels. La reproductibilité des activités du laboratoire s'est ainsi améliorée et une méthodologie a été définie. Ces démarches ont été couronnées de succès.

Toutefois, l'ouverture des données, des logiciels et des publications se confronte à des obstacles pratiques, à des problèmes financiers et à certaines réticences. Des chartes incitatives sont sûrement nécessaires afin que les communautés scientifiques prennent plus en considération la science ouverte. Le plan de données ouvertes du CNRS et l'initiative européenne RDA (Research Data Alliance) représentent des avancées intéressantes. Il serait également utile que l'usage et la mise à disposition de logiciels et données ouvertes soient valorisables.

Mots-clefs

Science ouverte, laboratoire, logiciel libre, OpenData, principes FAIR, Open Access, OPeNDAP, Dublin Core

1 Introduction

Le monde de la recherche doit faire face à ses propres maux comme dans toutes les grandes institutions. Par exemple, il connaît régulièrement certaines affaires de tromperie [1][2][3], de plagiat, de texte douteux, *etc.* qui ont rendu les chercheurs, mais aussi le grand public, moins confiants dans la science. Les quantités de données issues de la recherche sont aussi de plus en plus pharaoniques, et issues de processus de traitements complexes. Il est par conséquent délicat d'assurer la reproductibilité et la transparence des avancées scientifiques. Ajoutons à cela le fait que le milieu est très concurrentiel, ce qui incite la communauté à faire du chiffre (toujours plus de publications, plus de contrats) parfois au détriment de la qualité. Enfin, la réduction des moyens financiers, notamment récurrent (et donc humains), impose de « faire plus avec moins ».

Un levier intéressant pour répondre partiellement à ces problématiques est la science ouverte, dont les objectifs sont multiples : transparence, reproductibilité, efficacité, confiance, sobriété, pérennité. Cela passe par l'ouverture des données, mais aussi l'ouverture des logiciels et le libre accès aux publications scientifiques.

Tout d'abord, nous ferons un rappel de la définition de la science ouverte et de ses enjeux. Puis, dans un second temps, nous décrirons les actions menées au LEGI depuis 20 ans pour aller vers plus de science ouverte, à travers le regard conjugué d'un enseignant chercheur, d'un directeur de recherche directeur d'unité, d'un administrateur système et réseaux et d'un expert en calcul scientifique.

Nous exprimerons également les difficultés rencontrées au quotidien, ainsi que celles plus institutionnelles.

Enfin, nous évoquerons les initiatives nous paraissant particulièrement intéressantes ainsi que nos idées d'amélioration.

2 Définition de la science ouverte

La science ouverte (*open science* ou *open research* pour les anglophones) est un principe ancien (XVII^e siècle, création des premières académies : *Royal Society*, 1660 [4], Académie des Sciences, 1666 [5]) ayant pour objectif le partage à tous des données et résultats scientifiques.

Le but premier était d'échanger entre scientifiques. Aujourd'hui, la science ouverte a bien plus d'objectifs : transparence, reproductibilité, efficacité, confiance, sobriété, pérennité. Certains sont assez évidents et font l'unanimité (transparence, reproductibilité, confiance, pérennité) ; d'autres le sont moins et font l'objet de discussion (efficacité, sobriété). Explorons rapidement chacun de ces termes sachant que leur définition exacte fait parfois débat.

– Transparence : dans une étude, rien ne doit être caché, tant ce qui fonctionne que ce qui ne fonctionne pas.

– Reproductibilité : une autre personne dans un autre environnement doit être capable de reproduire (fabriquer à nouveau) les résultats annoncés.

– Efficacité : le partage permet à la recherche en général de ne pas avoir à réinventer la roue et d’être ainsi plus efficace (parvenir au résultat attendu plus rapidement, à un coût humain et en temps moins élevé).

– Confiance : les travaux doivent être revus par les pairs et pouvoir être annotés afin de pouvoir s’appuyer sur ce réseau de confiance pour faire accepter les résultats par le plus grand nombre.

– Sobriété : le partage des données, plutôt que de les reconstruire chacun de son côté, permet de moins consommer et puiser dans les ressources de la planète.

– Pérennité : les résultats doivent être accessibles le lendemain, mais aussi l’année d’après et ainsi de suite. Il doit y avoir une certaine pérennité dans le temps.

Plus concrètement, la science ouverte repose sur l’ouverture à tous de trois « entités » bien différentes : les **données**, les **logiciels** et les **publications scientifiques**.

2.1 Données

Les données constituent la technologie première à laquelle on pense quand on parle de science ouverte. Celles-ci sont la représentation élémentaire d’une information. Les données actuelles sont généralement numériques, et ainsi représentées à l’aide de 0 et de 1. C’est par exemple le cas de la température d’une station météo, enregistrée une fois par heure.

Dans notre milieu de l’enseignement supérieur et de la recherche (ESR) et plus particulièrement au sein du laboratoire LEGI, quand on pense à l’ouverture des données, on ne considère que les données de la recherche. Ces données doivent être mises en forme (métadonnées, considérant notamment les outils d’analyse, utilisation de formats ouverts et efficaces, pré-traitement éventuel...), puis stockées et diffusées. Dit autrement, elles doivent respecter les principes *FAIR* bien connus aujourd’hui (*Findable, Accessible, Interoperable, Reusable*) [6]. Le jeu de mot sur *fair* (équitable, juste) a été choisi pour faire référence au commerce équitable (*fair-trade*) et au fait d’être bon joueur (*fair-play*). Il évoque ainsi un comportement proactif et altruiste afin de rendre les données plus facilement utilisables par tous.

– Faciles à trouver : les données doivent être indexées et référencées.

– Accessibles : les données doivent être en libre accès sur le web et non au fin fond d’un intranet...

– Interopérables : les données doivent être publiées sous un format standard, de préférence ouvert, permettant d’être relues plus tard avec d’autres logiciels.

– Réutilisables : l’accès aux données est défini au travers d’une licence qui précise les conditions d’usage. Dans le cadre de la science ouverte, on considère une licence libre ; les données peuvent être utilisées dans d’autres travaux sans autorisation.

De nombreux guides décrivent plus complètement ces bonnes pratiques et sont à considérer afin de créer, puis d’ouvrir correctement les données, en respectant ces principes [7][8][9].

Il est d’ailleurs important de penser à toutes ces problématiques avant de générer les données. De même, la question de la durée de vie des données doit être traitée dès le début. Cette dernière soulève de nombreuses questions, telles que le coût et l’intérêt de la conservation des données à

long terme. L'expérience au LEGI montre qu'il est très difficile de supprimer une donnée. Elle est pourtant très rarement utilisée après plusieurs années, notamment lorsqu'elle n'est pas clairement structurée et identifiée au travers d'un identifiant unique et pérenne : *doi* (*Digital Object Identifier* [10]).

En agissant ainsi, on peut espérer que les objectifs de la science ouverte (transparence, reproductibilité, efficacité, confiance, sobriété, pérennité) soient en partie atteints. On peut tout de même se demander en quoi des données respectueuses des principes de la science ouverte améliorent l'efficacité et la sobriété. Du point de vue des auteurs, ouvrir les données permet d'éviter que d'autres chercheurs pratiquent la même expérience, et permet à moyen et long terme une meilleure efficacité pour l'exploitation et l'utilisation de ces données, que ce soit par les initiateurs ou les communautés. Ainsi, le coût humain est globalement moindre, et l'efficacité et la sobriété sont bien atteints.

2.2 Logiciels

Le logiciel est la deuxième pierre angulaire de la science ouverte. Le logiciel est un ensemble d'opérations élémentaires permettant de traiter et/ou de générer des données. Il apparaît ainsi logique d'ouvrir les logiciels, en complément des données. Cette démarche permet ainsi d'améliorer la reproductibilité de nos activités de recherche.

Comme pour les données, il convient de considérer des bonnes pratiques [11] et d'initier la démarche au plus tôt dans le développement logiciel (idéalement, avant même l'écriture de celui-ci, et sinon, dès que possible). « Il n'y a pas de petits logiciels ».

Depuis peu, il existe un équivalent aux principes *FAIR* pour les logiciels de recherche : *FAIR4RS* (*FAIR for Research Software*) [12]. Les quatre lettres *FAIR* représentent les mêmes mots et leurs définitions sont juste adaptés au génie logiciel. En résumé (l'article fait 30 pages), F pour avoir un logiciel indexé sur le net avec des versions et des métadonnées identifiées (par exemple les auteurs, la licence), A pour que celui-ci soit facilement téléchargeable au travers d'un protocole ouvert avec une URL unique, I pour un logiciel interopérable donc ayant une API (*Application Programming Interface*, un ensemble de définitions et de protocoles qui facilite la création et l'intégration de logiciels d'applications) claire et documentée et gérant ses entrées sorties selon des protocoles et des formats standards, enfin R pour que le logiciel soit utilisable et réutilisable donc modifiable ce qui entraîne, dans le cadre de la science ouverte, le choix d'une licence libre.

Toutes ces initiatives constituent un cadre clair sur comment partager les logiciels, et sont ainsi une avancée importante pour plus de reproductibilité, principe qui peut être poussé à l'extrême si on considère la reproductibilité computationnelle (reproduire un résultat au bit près) au travers d'outils d'environnements logiciels tels que *nix* ou *guix*. Nous pensons cependant que la reproductibilité computationnelle peut avoir certains travers : reproduire un résultat faux, ne plus bénéficier du débogage que constitue la compatibilité d'un logiciel avec plusieurs compilateurs, etc. Celle-ci est une excellente chose pour la sécurité informatique (projet *Reproducible Builds* dans la distribution Linux Debian [13]), mais elle est souvent questionnable dans le domaine scientifique. En effet, elle peut faire croire qu'il est possible de s'affranchir de l'épineuse question du critère permettant de

dire si oui ou non le résultat est « le même », alors que dans la grande majorité des cas, un résultat sans « tolérance » n'a que peu de sens.

2.3 Publications

L'ouverture à tous des publications est le troisième axe de la science ouverte. Ce principe repose sur le fait de publier dans des journaux en accès libre (*Open Access*) [14], et si possible en considérant un protocole d'évaluation avant publication ouvert et une soumission anonyme. Ce point est cependant délicat dans nos communautés scientifiques, en raison des habitudes historiques, de la main mise sur les publications de rang A des éditeurs imposant des coûts élevés, et d'autre part de certains éditeurs peu scrupuleux de revues secondaires essayant de s'enrichir en profitant de cette dynamique (sacrifiant généralement la qualité des articles publiés).

Il existe cependant des solutions, comme publier librement (sur la plateforme HAL [15] par exemple) et diffuser une version « soumise » (avant évaluation par les pairs) ou encore une version « acceptée » (après évaluation mais non éditée). Enfin, il est dans certains cas possibles de faire de même avec une version « éditée » après une période d'embargo (6 mois pour le domaine des sciences, techniques et médecine [16]). On parle alors de *Green Open Access* [17]. De précieux détails sont fournis dans le guide « *Les preprints : guide pratique* » [18] et sur le site *couperin* de la science ouverte en France [19].

Cette définition de la science ouverte ne correspond qu'à notre interprétation de nombreuses lectures et échanges avec les collègues et représentants de nos tutelles. Nous pouvons notamment mentionner les premier [20] et deuxième [21] plans nationaux pour la science ouverte, le guide pratique à l'usage des doctorants [22] mais aussi le contenu du site *ouvrir la science* [23], et notamment le guide « *je publie ; quels sont mes droits ?* » [24].

3 Science ouverte au LEGI

Depuis 20 ans, le laboratoire LEGI s'est engagé dans la voie de la donnée et de la science ouvertes.

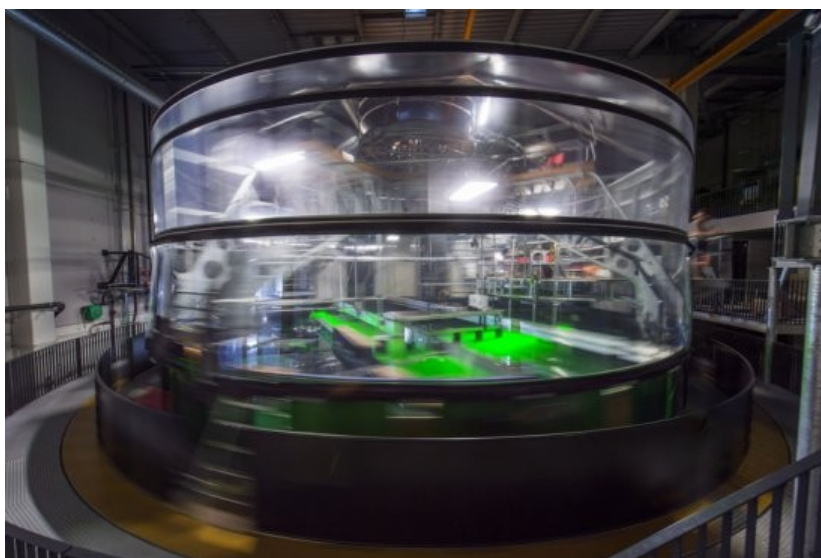


Figure 1 : La plateforme CORIOLIS du LEGI

À l'origine, il s'agissait de partager les résultats d'expérimentations sur la plateforme CORIOLIS entre collaborateurs en respectant les principes *FAIR* (Faciles à trouver, Accessibles, Interopérables, Réutilisables). La plateforme CORIOLIS est une installation unique au monde qui permet la modélisation expérimentale des écoulements géophysiques (figure 1). Des chercheurs de tous les continents viennent y réaliser des expériences et la commission européenne impose la diffusion libre des données acquises dans le cadre de ces programmes après une période d'embargo de deux ans. Ces données sont principalement des images et des champs de vitesse ou autres données organisés en dossiers par numéro d'expérience, enregistrés sous des formats standards, pérennes, multi-plateformes et multi-langages (PNG, NetCDF, XML). Cette organisation permet à une personne extérieure au projet initial de reprendre des données et de les analyser. Un point important pour la compréhension des données est d'associer à chaque projet une page web (wiki) décrivant le cadre des expériences ainsi que le jeu de paramètres de chaque essai. Il est aujourd'hui encore possible de lire et traiter des données datant de l'an 2000 [25]. Cette conservation sur le long terme est utile et nécessaire, car la reproductibilité expérimentale est rarement envisageable du fait de l'unicité de la plateforme et de la lourdeur des protocoles expérimentaux à mettre en place.

La réussite de ce partage de données d'expérimentations nous a amené après plusieurs années à envisager une approche similaire pour le partage de résultats de simulations numériques. Concernant ces données issues de simulations, il n'est pas toujours utile de les conserver sur le très long terme. Le volume de données est souvent important, et dans le cadre de l'utilisation de logiciels libres dont le code historisé est déjà accessible par tous, le partage des données d'entrée des simulations apparaît souvent plus pertinent que le partage de toutes les données issues de la simulation. Cela évite de trop stocker à long terme et suffit à assurer la reproductibilité. Quelques échantillons de données issues de la simulation, bien choisies, et bien documentées (métadonnées) sont cependant conservées et partagées.

Depuis plus de treize ans, les logiciels utilisés au LEGI sont historisés via une forge, initialement déployée en interne au laboratoire (via subversion). Nous utilisons désormais principalement la forge GitLab du site universitaire Grenoblois, ainsi que la forge publique GitHub. Au fil des années, nos techniques de gestion des logiciels ont évolué. Nous y associons une documentation et utilisons l'intégration continue de manière intelligente, afin d'éviter les régressions, mais aussi les surconsommations électriques. Bien sûr, et depuis toujours, nous apposons aux logiciels développés une licence libre (dans le cadre de la science ouverte) ainsi que les fichiers AUTHORS et README.

Plus globalement, nous avons pour objectif d'améliorer la reproductibilité de nos expérimentations (numériques ou non), et ainsi de proposer une méthodologie en termes de gestion globale logiciels-données-publications. Dans le cadre d'une publication sur le développement d'un solveur, sedFoam, basé sur OpenFoam et permettant la modélisation du transport sédimentaire [26], la démarche complète a été réalisée et documentée. Les données ayant permis la réalisation des courbes et des tableaux d'une publication ont été téléversées sur le site Zenodo sous licence libre [27]. Le choix des données pertinentes, leurs mises en format pérenne et l'explication de l'organisation des données ont ainsi permis de développer un certain savoir-faire que nous généralisons aux données issues d'expérimentations.

Les archives Zenodo sont effectivement une solution pour cet archivage des données associées à une publication et permettent à une personne de reprendre les données et de reproduire les courbes. Afin de rendre la construction des archives Zenodo plus robuste et plus facile pour les utilisateurs, nous avons développé un fichier de méta-données, PROJECT-META.yml [28] (exemple de fichier en annexe), de type clef-valeur permettant d'imposer des spécifications, notamment celles du *Dublin Core* [29] et ainsi de générer automatiquement un certain nombre de fichiers (README, AUTHOR, LICENCE...) pour finalement créer les archives Zenodo associées. Le vocabulaire *Dublin Core* est issu d'un consensus international et pluridisciplinaire permettant de décrire avec seulement 15 mots clefs des documents de manière simple et standardisée. En nous basant dessus, nous avons ainsi un socle d'éléments descriptifs structuré nous donnant de bonnes possibilités d'indexation. Ce fichier de métadonnées facilite grandement l'interopérabilité présente ou future.

Les archives Zenodo sont cependant limitées en termes de taille (50 Go) et l'archive complète doit être téléchargée pour avoir accès à l'une de ses données internes. Pour pallier cette problématique, un serveur interne DAP (Data Access Protocol [30]) a été mis en place afin de partager nos données. Notre serveur DAP (en pratique une instance Hyrax OPeNDAP [31]) possède l'avantage de ne pas imposer de quota et de permettre le téléchargement partiel des données ou une lecture directe en ligne. Nous avons ainsi développé un logiciel permettant de visualiser et analyser en local sous Matlab des données directement stockées sur le serveur [32].

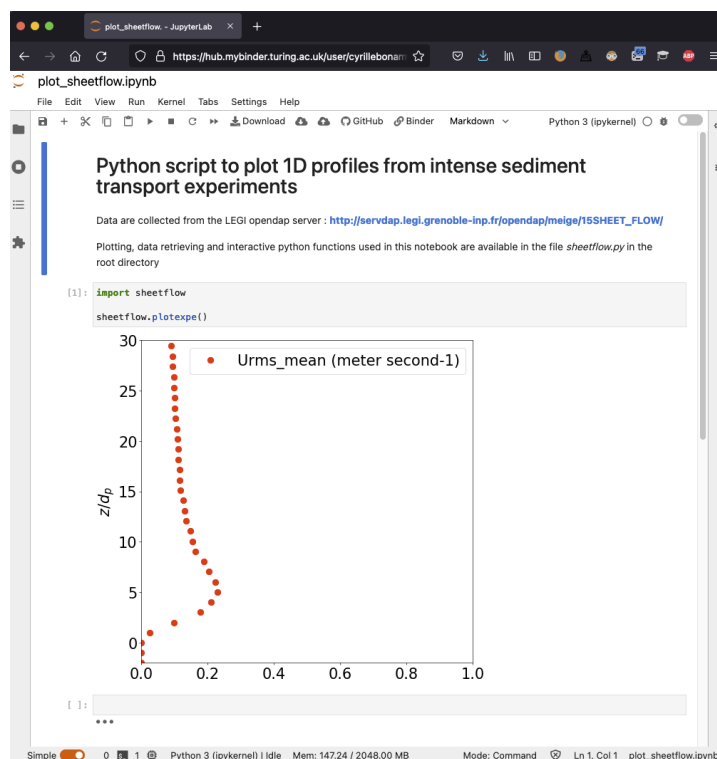


Figure 2 : Exemple d'utilisation de Binder et Notebook

Pour rendre les données diffusées plus compréhensibles et utilisables, nous avons développé des *Notebooks* Python disponibles via GitHub et utilisables facilement avec *Binder* (une console *Jupyter* ouverte à tous). Ainsi, grâce à un outil visuel non hébergé au LEGI (*Jupyter / Binder / Python*), il est possible de visualiser les données libres du LEGI sans aucun téléchargement sur son poste de travail [33] (figure 2).

Enfin, en termes de partage de publications, nous nous attachons à diffuser dès que possible nos articles, principalement en mode *Green Open Access* (§ 2.3), parfois en mode *Gold Open Access* (payant pour publier), quand cela est requis par nos sources de financement (projets européens par exemple, pour lesquels un financement est dédié à ces publications), tout en se protégeant des journaux profiteurs du système (§ 4). Concrètement, les versions soumises de nos productions sont publiées via le site *arxiv* [34], et une fois celles-ci acceptées et la période d’embargo passée, elles sont diffusées via le service HAL.

4 Conclusions, difficultés et pistes pour l’avenir

Toutes ces démarches ont été couronnées de succès. Nous sommes souvent abordés pour parler de notre expérience, et la démarche autour du logiciel *sedFoam* a permis de rendre accessible l’outil et d’attirer bon nombre d’utilisateurs à travers le monde. Il est d’ailleurs prévu une intégration du solveur au sein de la version officielle d’*OpenFOAM.com* (branche principale).

Dans le cadre du projet européen *Hydralab+* (2015-2019) [35], dont *CORIOLIS* était un des grands équipements, un *Joint Research Activity (JRA)* a été dédié aux aspects *OpenData*. Celui-ci a permis d’unifier les pratiques en termes de partage de données expérimentales au sein du consortium, consortium qui réunit 24 partenaires de toute l’Europe travaillant sur la thématique du changement climatique.

Au sein du laboratoire *LEGI*, la démarche est de plus en plus acceptée et considérée ; une grande majorité des membres du laboratoire sont convaincus de son intérêt.

Néanmoins, l’ouverture des données, des logiciels et des publications se confronte à certaines réticences et à des problèmes financiers. Tout d’abord la recherche est un enjeu de puissance et même de survie économique dans un monde concurrentiel. Ainsi les financeurs de la recherche, publics ou privés, attendent un retour sur investissement souvent incompatible avec la diffusion complète des données et logiciels. De larges domaines de la recherche restent cependant propices à la science ouverte. La situation est résumée par la phrase clamée par le CNRS « accessible autant que possible, et fermé autant que nécessaire » [36].

La science ouverte se heurte aussi à des obstacles pratiques même lorsqu’elle est souhaitée. D’une part, certains chercheurs craignent que le fait d’ouvrir leurs logiciels entraîne une mauvaise utilisation (et donc des critiques) ; d’autre part, cela représente un coût important en temps humain pour la documentation et le service aux utilisateurs. Il existe aussi des contraintes vis-à-vis des licences d’origine des codes utilisés. D’anciennes collaborations entreprise-public ont ainsi été contraintes à l’utilisation de licence « fermée », et il est ensuite impossible de faire marche arrière. Enfin, publier en « libre accès » coûte cher (parfois de 2 000 à 3 000 € par publication), et la tentation est donc forte de publier de manière conventionnelle pour des questions de budget, ou bien de se tourner vers des éditeurs « prédateurs » aux pratiques d’expertise douteuses.

Des lois ou chartes incitatives, tel que le récent décret paru au *Journal Officiel* [37], sont sûrement nécessaires afin que les communautés scientifiques prennent en considération la science ouverte. Le plan de données ouvertes du CNRS [38], l’initiative européenne *RDA (Research Data Alliance)* et les dynamiques de « site » d’accompagnement à la science ouverte (exemple à Grenoble : la cellule *data-stewardship* [39]) représentent des avancées intéressantes et nécessaires. Concernant

l'ouverture des publications, on pourrait imaginer des fonds à l'échelle des tutelles (universités, EPST...) afin de permettre aux scientifiques de publier en « accès libre » sans que ce soit imputé à leurs budgets propres (c'est ce qui est fait à l'université de Manchester [40]). Il serait sûrement utile que l'usage et la mise à disposition de logiciels et données ouvertes soient valorisables très concrètement, voire imposés pour certains appels d'offre. Cela reste cependant peu effectif en l'absence d'une pérennisation et documentation de données et logiciels, ce qui nécessite souvent l'appui de personnels dédiés capable d'assurer une continuité de service.

Pour finir, voici une liste non exhaustive d'enseignements, conseils et points saillants issus de notre expérience.

- Le premier bénéficiaire de la « science ouverte » est la personne qui l'initie. En effet, les principes de la science ouverte imposent des principes très utiles à long terme (documentation, format, accessibilité, interopérabilité, visibilité...).
- La démarche doit être initiée dès le montage des projets, comme préconisée par les plans de gestion des données (PGD, ou *DMP – Data Management Plan* en anglais).
- Apposer une durée de vie aux données et ainsi penser à la fin de vie des données est difficilement acceptable par les chercheurs mais apparaît comme indispensable.
- Le choix des licences (que ce soit pour les données ou les logiciels) doit être anticipé dès le début (montage) du projet afin d'éviter tout problème ultérieur.
- Citer des données, que ce soit via un jeu de données scientifiques (*data paper*) ou via un identifiant unique de type *doi* (*Digital Object Identifier [10]*), dans nos publications est indispensable, mais n'est pas quelque chose de naturel pour nos communautés.
- Le coût humain est important à court terme, mais en vaut la peine sur le moyen terme, en réalité souvent dès le stagiaire suivant.
- Science ouverte et éco-responsabilité ne sont pas opposés, bien au contraire.

Licence

Cet article est soumis à la licence libre Art Libre [41]. Avec la Licence Art Libre, l'autorisation est donnée de copier, de diffuser et de transformer librement les œuvres dans le respect des droits de l'auteur. Cette licence ressemble dans l'esprit à la licence GPL (*GNU General Public License* [42]) sur les logiciels, mais elle est plus adaptée aux textes et aux dessins que cette dernière. La GPL est en effet très orientée pour le code source.

Bibliographie

- [1] Dong-Poo Han – <https://nypost.com/2013/12/26/professor-admits-faking-aids-vaccine-to-get-19m-in-grants>
- [2] La controverse Reinhart-Rogoff – https://en.wikipedia.org/wiki/Growth_in_a_Time_of_Debt
- [3] Fraude scientifique, CNRS Le journal, <https://lejournel.cnrs.fr/fraude-scientifique>
- [4] Royal Society – <https://www.gresham.ac.uk/about/history/>

- [5] Académie des Sciences – <https://www.academie-sciences.fr/fr/Documents-a-telecharger/en-savoir-plus-sur-academie.html>
- [6] FAIR (findable, accessible, interoperable, reusable) – https://fr.wikipedia.org/wiki/Fair_data
- [7] *Guide de bonnes pratiques sur la gestion des données de la Recherche*. Christine Hadrossek, Joanna Janik, Maurice Libes, Violaine Louvet, Marie-Claude Quidoz, Alain Rivet, Geneviève Romier, 2021 – <https://hal.archives-ouvertes.fr/hal-03152732/document>
- [8] *Guide de bonnes pratiques Gestion et valorisation des données de la recherche*. Pierre-Yves Arnould et Marie-Christine Jacquemot. [Rapport de recherche] OTELo ; INIST-CNRS. 2016 – <https://hal.archives-ouvertes.fr/hal-01275841>
- [9] *Ouverture des données de recherche, Guide d'analyse du cadre juridique en France*. Nicolas Becard, Céline Castets-Renard, Gauthier, Martin Dantant, Laurence Freyt-Caffin Nathalie Gandon, Caroline Martin, Andrea Martelletti, Alexandra Mendoza-Caminade, Nathalie Morcrette, Claire Neirac. Décembre 2017 – <https://hal.inrae.fr/hal-02791224>
- [10] *Digital object identifier* (DOI, littéralement « identifiant numérique d'objet ») – https://fr.wikipedia.org/wiki/Digital_Object_Identifier
- [11] *Je code : les bonnes pratiques en éco-conception de service numérique à destination des développeurs de logiciels*. Cyrille Bonamy, Cédric Boudinet, Laurent Bourgès, Karin Dassas, Laurent Lefèvre, et al. 2020 – <https://hal.archives-ouvertes.fr/hal-03009741v2>
- [12] *FAIR Principles for Research Software (FAIR4RS Principles)*. Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2021). Research Data Alliance – <https://doi.org/10.15497/RDA00065>
- [13] Projet Reproducible Builds – <https://wiki.debian.org/ReproducibleBuilds/About>
- [14] Libre accès (*Open Access*) – [https://fr.wikipedia.org/wiki/Libre_accès_\(édition_scientifique\)](https://fr.wikipedia.org/wiki/Libre_accès_(édition_scientifique))
- [15] HAL (Hyper Article en Ligne) – <https://hal.archives-ouvertes.fr/>
- [16] Vos dépôts dans HAL : ce qui change avec la loi pour une République Numérique – <https://www.ccsd.cnrs.fr/2016/10/vos-depots-dans-hal-ce-qui-change-avec-la-loi-pour-une-republique-numerique>
- [17] Le libre accès aux publications et aux données de recherche, Horizon 2020 Le portail français du programme européen pour la recherche et l'innovation – <https://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html>
- [18] *Les préprints : guide pratique*. Kristina Hettne, Ron Aardening, Chantal Hukkelhoven, Dirk van Gorp, Nicole Loorbach, et al. 2021, 10.13143/0603-7h42 – <https://halshs.archives-ouvertes.fr/halshs-03470613/>
- [19] Couperin : version et embargos – <https://scienceouverte.couperin.org/versions-et-embargos/>
- [20] Plan national de la science ouverte – https://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf
- [21] Second plan national de la science ouverte – https://www.ouvrirlascience.fr/wp-content/uploads/2021/06/Deuxieme-Plan-National-Science-Ouverte_2021-2024.pdf

- [22] Passeport pour la science ouverte : guide pratique à l'usage des doctorants – <https://www.ouvrirlascience.fr/passeport-pour-la-science-ouverte-guide-pratique-a-lusage-des-doctorants/>
- [23] Ouvrir la science – <https://www.ouvrirlascience.fr/>
- [24] Je publie ? Quels sont mes droits – <https://www.ouvrirlascience.fr/wp-content/uploads/2020/10/JepublieQuelsontmesdroits-Vdef.pdf>
- [25] *Exemple de page wiki d'un projet Coriolis*. J. Sommeria, M.E. Negretti, S. Viboud, T. Valrand – <http://servforge.legi.grenoble-inp.fr/projects/pj-coriolis-17iceshelf/>
- [26] *SedFoam-2.0: a 3-D two-phase flow numerical model for sediment transport*. Chauchat, J., Cheng, Z., Nagel, T., Bonamy, C. & Hsu, T. - J. 2017. *Geoscientific Model Development*, vol. 10(12), p.4367-4392 – <https://www.geosci-model-dev.net/10/4367/2017/>
- [27] *SedFoam/sedfoam: Release 3.2 (3.2)*. Cyrille Bonamy, Julien Chauchat, Tian-Jian Hsu, Eduard Puig-Montella, Remi Chassagne, Antoine Mathieu, Zhen Cheng, Tim Nagel, Ali Salimi-Tarazouj, Benjamin Tsai, & Pablo Higuera. (2021). Zenodo – <https://doi.org/10.5281/zenodo.5095239>
- [28] Logiciel ProjectMeta – <https://gricad-gitlab.univ-grenoble-alpes.fr/legi/soft/trokata/project-meta>
- [29] Spécification Dublin Core – https://fr.wikipedia.org/wiki/Dublin_Core
- [30] DAP / OPeNDAP protocole – <https://en.wikipedia.org/wiki/OPeNDAP>
- [31] Hyrax OPeNDAP serveur – <https://www.opendap.org/software/hyrax-data-server>
- [32] Logiciel UVmat – <http://servforge.legi.grenoble-inp.fr/projects/soft-uvmat>
- [33] Exemple de Notebook Jupyter permettant l'accès à un jeu de données libres – https://mybinder.org/v2/gh/CyrilleBonamy/notebook_opendata/HEAD?labpath=plot_sheetflow.ipynb
- [34] Arxiv – <http://arxiv.org/>
- [35] Projet européen Hydralab – <https://hydralab.eu>
- [36] « accessible autant que possible et fermé autant que nécessaire » – <https://www.science-ouverte.cnrs.fr/>
- [37] Exigences de l'intégrité scientifique – <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>
- [38] Plan données de la recherche du CNRS – https://www.science-ouverte.cnrs.fr/wp-content/uploads/2021/01/Plaquette-Plan-Donnees-Recherche-CNRS_nov2020.pdf
- [39] Cellule *data-stewardship* de GRICAD – <https://gricad.gricad-pages.univ-grenoble-alpes.fr/cellule-data-stewardship/web/>
- [40] Open Access funding – <https://www.library.manchester.ac.uk/using-the-library/staff/research/open-research/access/funding/>
- [41] Licence Art Libre – <https://artlibre.org>
- [42] Licence publique générale GNU, ou *GNU General Public License* – https://fr.wikipedia.org/wiki/Licence_publique_générale_GNU

Annexe : exemple de fichier PROJECT-META.yml

```
---
version: 2
project:
  title: Blockage of saline intrusions in restricted, two-layer exchange flows across a submerged sill obstruction
  identifier:
    acronym: 14CARR
    url-doc: http://servforge.legi.grenoble-inp.fr/projects/pj-coriolis-14carr
  creator:
    - Alan Cuthbertson <a.cuthbertson(A)hw.ac.uk> Heriot Watt University, Edinburgh, UK
    - Janek Laanearu <janek.laanearu(A)ttu.ee> Tallinn University of Technology, Tallinn, Estonia
    - Magda Carr <magda.carr(A)st-and.ac.uk> University of St Andrews, St Andrews, UK
    - Joel Sommeria <Joel.Sommeria(A)univ-grenoble-alpes.fr> LEGI UMR5519, Grenoble, France
    - Samuel Viboud <Samuel.Viboud(A)univ-grenoble-alpes.fr> LEGI UMR5519, Grenoble, France
  subject:
    - Exchange flows
    - Net-barotropic flows
    - Saline intrusion blockage
    - Sill obstructions
    - Stratified flow
    - Mixing and entrainment
    - Internal-flow hydraulic modelling
  contributor:
    - LEGI Coriolis Platform Grenoble
  description: >
    The data in netcdf format (.nc) represent velocity fields in a vertical cut of the gravity current produced on a submerged sill. Two cameras are used, Dalsa1 and Dalsa2. The corresponding velocity field series are stored in the folders named Dalsa1.civ.mproj and Dalsa2.civ(4).mproj respectively. A flag FF=1 indicates non defined data (0 for good data). Folder named Dalsa1.civ.mproj.stat represents the mean velocities over the whole series of Dalsa1.civ.mproj. Vertical salinity profiles are provided in the folder ProbeNetcdf for each data set. See url-doc and publication for details.
  publisher: LEGI UMR 5519 - CNRS, Univ Grenoble Alpes, Grenoble INP
  date: 2018-04-23
  type:
    - dataset
    - image
  format:
    - image/png
    - application/xml
    - application/netcdf
  # if the work is derived from...
  source:
  language: en-US
  relation:
    - doi: 10.1007/s10652-017-9523-2
  coverage: Hydralab, LEGI, Coriolis, Grenoble (2014)
  rights: licence-ouverte-v2.0
```

public-dap:

dap-folder: /fsnet/project/coriolis/app-dap-public

data-set:

- *ESTUARY2_150714/ESTUARY2g/Dalsa1.mproj*
- *ESTUARY2_150714/ESTUARY2g/Dalsa1.civ.mproj*
- *ESTUARY2_150714/ESTUARY2g/Dalsa1.civ.mproj.stat*
- *ESTUARY2_150714/ESTUARY2g/Dalsa2.civ4.mproj*
- *ESTUARY2_150714/ESTUARY2g/Dalsa2.civ4.mproj.stat*
- *ESTUARY2_150714/ProbesNetcdf*
- *ESTUARY3_160714/ProbesNetcdf*