

Le logiciel de recherche, un pilier de la recherche scientifique ouverte

Violaine Louvet

Laboratoire Jean Kuntzmann & collège codes sources et logiciels du COSO

JCAD 2023, 3 octobre 2023

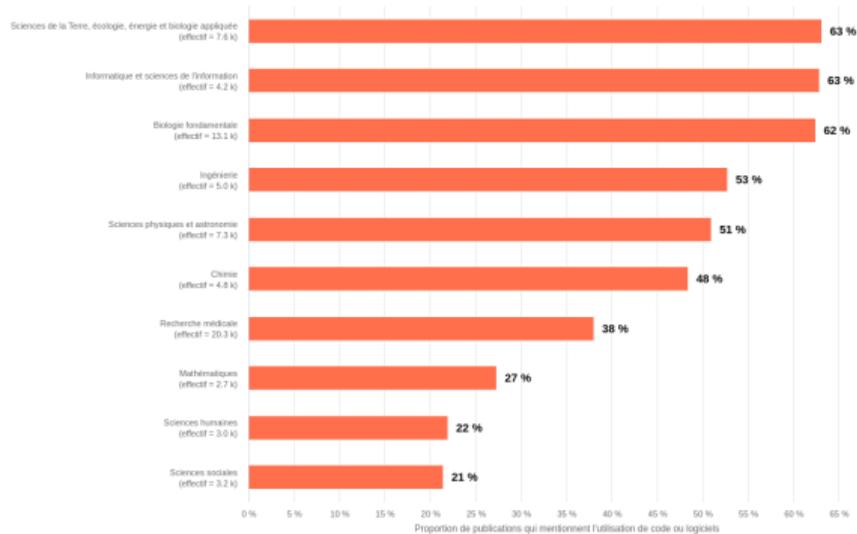
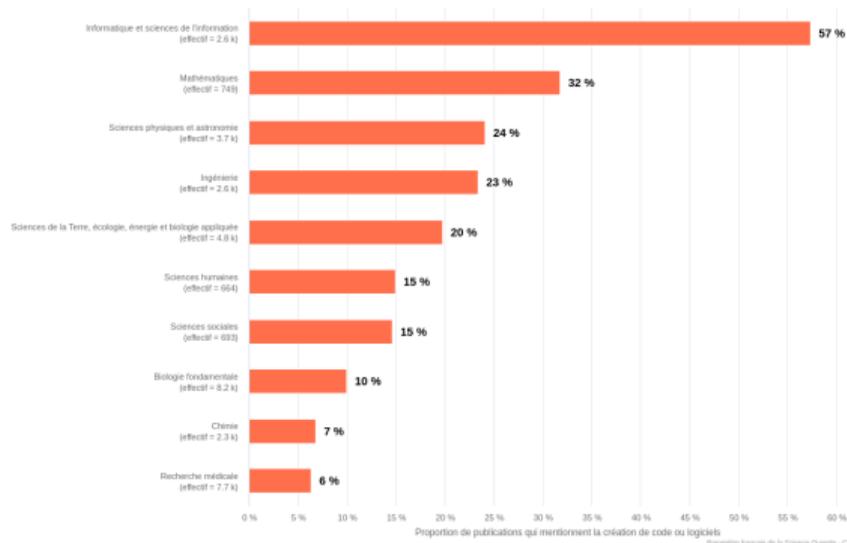


OS Ouvrir
la science !



Toutes les disciplines scientifiques sont concernées

Quelques éléments du baromètre science ouverte 2022



Proportion de publications qui mentionnent la création de codes

Proportion de publications qui mentionnent l'utilisation de codes

Logiciel libre et science ouverte

- Le logiciel libre est né dans la **pratique universitaire**, le partage des connaissances est au coeur de la recherche.
 - En particulier, dans la communauté des **hackers** (« bidouilleurs »), désignant originellement ces jeunes étudiants en informatique du MIT, génies de la programmation, dont Richard Stallman.
 - Il crée la **Free Software Foundation (FSF)** en 1985.
- Une grande majorité des logiciels développés dans la recherche le sont sous **licence libre** (cf rapport à venir MESR)
- Finalement, le logiciel est un **précurseur de la science ouverte** depuis toujours.
- Mais n'entre en jeu que **récemment dans les politiques publiques** ...



Plan National Science Ouverte V2 (1/2)

Intégration du logiciel dans la **politique nationale lors de la publication du PNSO V2** :

- Définir et promouvoir une **politique en matière de logiciels libres**
 - Produire une **charte nationale pour les logiciels libres** issus de l'enseignement supérieur, de la recherche et de l'innovation
 - Développer le **lien entre les données et les logiciels** à travers un réseau d'administrateurs des données et des codes de recherche dans les différentes universités et organismes performants de recherche
 - Développer les **modèles économiques du logiciel libre** et les faire connaître au sein des services de valorisation
 - Soutenir le projet **Software Heritage** et le recommander pour l'archivage et le référencement du code source

Plan National Science Ouverte V2 (2/2)

- Reconnaître le code source comme une **contribution à la recherche**
 - Créer un **prix du logiciel libre de la recherche**
 - Accorder une plus grande **reconnaissance** à la production de logiciels dans la carrière des chercheurs et du personnel de soutien à la recherche
- Créer un **écosystème qui relie le code, les données et les publications**
 - Développer une **coordination** adéquate entre les forges logicielles, les archives de publications ouvertes, les dépôts de données et le secteur de l'édition scientifique



La création du collège codes sources et logiciels libres

Au sein du comité pour la science ouverte (COSO)

Nommé en avril 2022 par le MESR.

Co-pilotes : Roberto Di Cosmo et François Pellegrini

- **Suivre, analyser et produire des avis** sur les sujets d'actualité relatifs aux codes source et aux logiciels dans le cadre de la science ouverte
 - Production de notes, recommandations, documents collaboratifs, etc.
- **Impulser, réaliser ou piloter des projets opérationnels**, qui contribuent aux chantiers identifiés dans l'axe logiciel du PNSO2
- **Se coordonner** avec les autres collèges, la DINUM, et l'écosystème national et international

Thèmes de travail du collège

- Thème 1 : **Identification et mise en avant de la production logicielle** de l'ESR
 - Catalogue des logiciels
- Thème 2 : **Outils et bonnes pratiques** techniques et sociales
 - Forges logicielles
- Thème 3 : **Valorisation et durabilité**
 - Politiques de valorisation
- Thème 4 : Liaison et animation **nationale, Europe, international**
 - Politique logicielle
- Thème 5 : **Reconnaissance et carrières**
 - Prix du logiciel libre de recherche

Focus catalogue logiciels : quelques éléments

- Elaboration d'un état des lieux
 - **Catalogues existants** : contenu, processus de construction (et de mise à jour), pratiques des utilisateurs ...
 - Tryptique **Software Heritage / HAL / CodeMeta**
- Compréhension des **usages** et des problématiques de description (métadonnées)
 - En particulier, comment optimiser pour le chercheur : référencement, mise à jour, ...
 - Faire le lien avec les forges
- Réflexions autour de l'**accompagnement** des communautés scientifiques
- **Enquête du ministère** sur l'étude d'impact de la production logicielle de l'ESR (attendu fin novembre)

→ Livrable prévu fin d'année

Focus sur les travaux autour des forges logicielles

- Rapport public sur un **état des lieux des forges dans l'ESR** (disponible également en anglais) :
 - Une quarantaine de **forges auto-hébergées** identifiées
 - Quasiment toutes sont des instances de **gitlab**
 - **Rapport accessible sur HAL.**
 - Rapport ouvert aux compléments (en particulier forge non encore listée) :
<https://gitlab.in2p3.fr/coso-college-codes-sources-et-logiciels/forges-esr>
- Identification de **problématiques** :
 - Accès le plus souvent lié à l'institution qui héberge
 - Fonctionnalités offertes très variables (CI, pages ...)
 - Maintenance et support variables
 - Utilisation importante des forges commerciales (en particulier github)
- Travail en cours sur des **préconisations** basées sur l'état des lieux.

Prix du logiciel libre de la recherche

- Décerné depuis 2022
- **Récompense des logiciels** qui contribuent à l'avancée de la connaissance scientifique selon plusieurs catégories :
 - la qualité scientifique et technique du logiciel,
 - la construction d'une communauté active de contributeurs et utilisateurs,
 - l'effort indispensable de fournir une documentation qui facilite l'usage et l'appropriation des logiciels,
 - et une catégorie « coup de coeur » du jury.
- Prix **attribués collectivement** aux conceptrices/concepteurs, développeuses/développeurs et chargées/chargés de la maintenance du logiciel et à sa communauté.
- **Lauréat 2022** :
 - Catégorie scientifique et technique : Coq proof assistant (accessit CORIOLIS VLSI CAD TOOLS)
 - Catégorie communauté : Scikit-learn (accessit Vidjil et WebObs)
 - Catégorie documentation : Faust (accessit OpenVibe)
 - Prix du jury : Gammapy (accessit GAMA et SPPAS)
- Remise des **prix 2023** le 29 novembre.

Le code n'est pas une donnée comme les autres

- Les données de recherche sont plutôt passives, les codes sont **intrinsèquement vivants**
 - On ne change en général pas les données, collectées dans un contexte bien défini
 - On change éventuellement la façon dont on les traite et on les analyse (grâce à des codes)
 - Les codes sont associés à une (ou des) **action(s)** : création de connaissances, transformation d'informations, visualisation, ...
 - Le code peut être réutilisé tel que, en reproduisant son environnement et toutes ses dépendances mais on a surtout envie de le **modifier pour l'adapter** à nos besoins propres ou **l'enrichir de nouvelles fonctionnalités**
- Les codes représentent un travail de création, et correspondent à un **cadre juridique différent de celui des données**
- Les codes s'appuient sur des **dépendances et tout un environnement logiciel et matériel** qui évolue sans cesse
 - Cela complexifie les questions de **reproductibilité**

Problématique de reproductibilité

- Au coeur de la **science ouverte**
- Un mouvement qui prend de l'ampleur et qui concerne des sujets à **différents niveaux** :
 - l'**ouverture** des différents objets : données, codes, publications
 - la **description /documentation** claire des objets (dont des métadonnées riches)
 - la maîtrise de l'**environnement logiciel**
 - Environnements reproductible comme guix, nix ...
 - la connaissance du **matériel** sous-jacent
 - Reproductibilité d'un calcul HPC ??
 - ...

Réseau recherche reproductible

- Rassemble les personnes intéressées par l'étude des facteurs qui contribuent à la robustesse de la recherche, la promotion des activités de formation et la diffusion des bonnes pratiques ainsi que des recommandations.
- Toutes les disciplines sont concernées.
- <https://www.recherche-reproductible.fr/>

Accompagnement sur les questions du logiciel

Besoins en formation

- A tous les niveaux : masters, doctorants, ingénieurs et chercheurs
 - Sur des sujets techniques : bonnes pratiques de développement, forges logicielles, packaging, ...
 - Mais aussi sur des sujets liés à la science ouverte : archivage, métadonnées, licences, ...
-
- S'appuyer sur la **dynamique autour des données**
 - **Maillage national** avec les ateliers de la donnée au plus proche des communautés scientifiques (cf la table ronde ce soir !)
 - Nécessité d'intégrer les compétences techniques associées
 - Intégrer la **dimension codes et logiciels** dans cet accompagnement
 - Les **centres de calcul** ont un rôle important à jouer : expertises techniques, proximité avec les laboratoires, formations déjà existantes ...
 - Nécessité de s'adapter à des communautés qui sont loin du HPC

Conclusions

- Le logiciel prend enfin une place reconnue dans les **politiques de science ouverte**, au niveau national mais aussi dans d'autres pays
 - Mais il a un grand temps d'avance !
 - Il faut que les **établissements et les organismes** s'emparent des problématiques associées en particulier :
 - La valorisation de l'activité de développement,
 - Les questions de maintenance, de support, de RH associés,
 - Les outils pour le développement,
 - L'incitation à l'archivage et au référencement ...
- La question de la **formation et de l'accompagnement** est critique
 - Il faut profiter de la dynamique autour des données pour que le logiciel ne soit pas oublié
 - Il est nécessaire que les personnes qui comprennent et maîtrisent l'objet s'investissent
 - Il ne faut pas oublier non plus la formation des services de valorisation (un des groupes du collège travaille sur le sujet)
- **SPOS 2023** : 29 novembre, matinée consacrée à la présentation du rapport logiciel, et à la remise des prix science ouverte 2023 du logiciel libre de recherche et des données de la recherche. Après midi autour du logiciel avec un programme à venir.