

**2.
Les logiciels
et la démarche
scientifique
ouverte**

L'open science
concerne les processus
de recherche autant que les résultats.
À la jonction des deux,
les logiciels constituent, avec
les publications et les données,
le troisième pilier
de la Science ouverte.

À partir de la diffusion ouverte des résultats de la recherche, une démarche cohérente de Science ouverte consiste à ouvrir toutes les étapes antérieures à l'origine de ces résultats : méthodes, outils, carnets de laboratoire, codes informatiques..., et ce dès l'initiation d'un projet¹

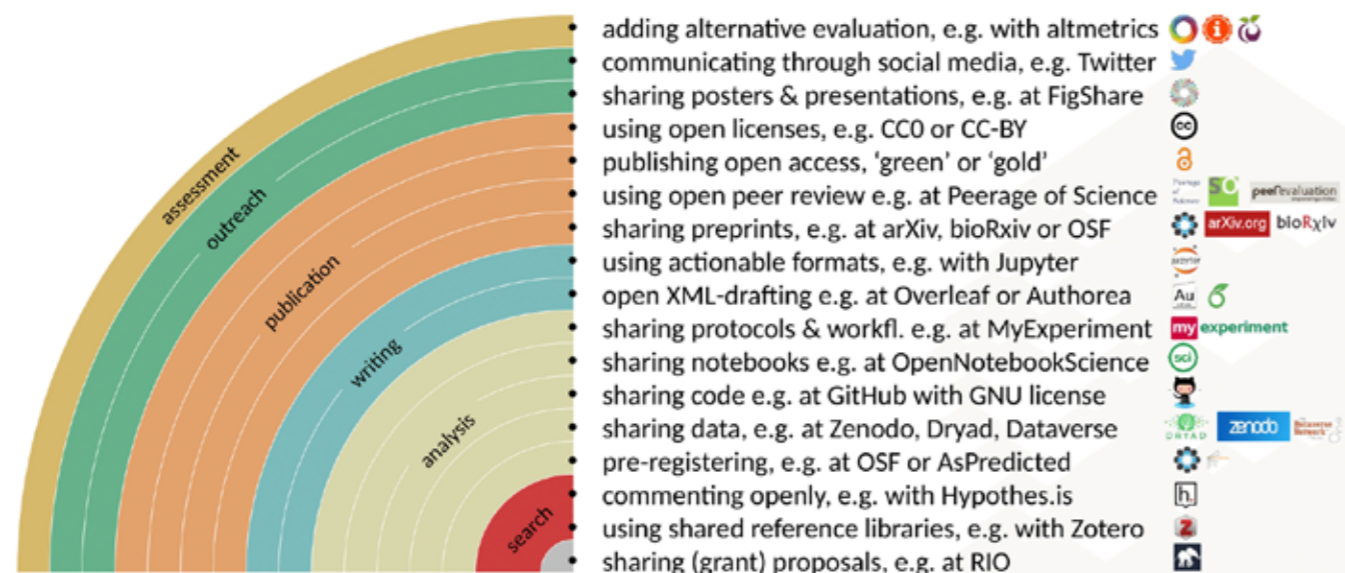
Logiciels et codes sources

Les logiciels sont à la charnière des résultats et des processus. Développés par et pour un projet de recherche, ils font souvent partie des productions et des résultats de la recherche au même titre qu'une publication. Même lorsqu'ils n'ont pas été développés spécifiquement, ils sont à la source de nombreuses données produites ou analysées.

A ce double titre, leur identification et leur archivage sont un enjeu pour la réutilisation d'un jeu de données ou la reproductibilité d'une expérience. Le choix de codes sources ouverts ou de logiciels propriétaires au cours d'un projet doit être évalué sur le long terme et documenté autant que possible.

¹ - "as early as is practical in the discovery process" NIELSEN, Micheal. TED Conference. En ligne. URL : https://www.ted.com/talks/michael_nielsen_open_science_now/transcript?language=fr

Le graphique ci-contre illustre l'enchaînement de logiciels constitutif d'une démarche de recherche, depuis le noyau rouge de conception du projet jusqu'à l'enveloppe externe de son évaluation, ici avec des exemples d'outils gratuits, et libres pour la plupart.



Jeroen Bosman, Bianca Kramer, 2017 | CC BY²

Déposer sur Software Heritage !

*Software Heritage*³ est un entrepôt d'archive logicielle ouvert en septembre 2018 destiné à la collecte et à la préservation du code source des logiciels. Initié en 2014 par des enseignants-chercheurs de l'université Paris Diderot⁴, il est soutenu par l'Inria et l'UNESCO. Près de 4,5 milliards de fichiers sources pour 83 millions de projets y sont déjà archivés, notamment tous ceux disponibles sur la plate-forme de développement *GitHub*⁵. Déposer un logiciel sur HAL, c'est archiver sur *Software Heritage* : un partenariat lie les deux applications !

Questions d'ouverture, aux différentes étapes

Entrepôts et conservation à long terme : comment intégrer cette démarche en amont des projets ? La rédaction du Plan de Gestion de Données (PGD/*Data Management Plan*) formalise une partie de cette réflexion.

Publier, éditer : la soumission, l'édition et la publication des travaux académiques est au cœur du processus de la communication scientifique. Comment contribuer à améliorer ce processus et à le

fonder sur des bonnes pratiques de transparence et d'intégration à un écosystème ouvert ?

La révision par les pairs (*peer reviewing*) est très majoritairement organisée au sein de systèmes propriétaires. Comment ouvrir le processus en même temps que la pratique du *peer reviewing* ? Les réponses et expérimentations varient selon les communautés disciplinaires

La citation : les outils de gestion des données bibliographiques et les pratiques de *l'open citation*⁶ sont indispensables. Comment pleinement intégrer ces outils et ces pratiques au circuit d'écriture et de publication ?

La gestion des données : de la collecte à l'analyse, du stockage à l'archivage. Comment recueillir les besoins des communautés disciplinaires et développer les services adaptés pour y répondre ?

Les pratiques de fouille de textes et de données : pour les lecteurs humains comme pour les machines, les métadonnées — les données caractérisant les données — sont aussi importantes que les données elles-mêmes, indispensable à leur compréhension. Comment les outils ouverts peuvent-ils faciliter et rationaliser leur élaboration ?

Chercher et trouver l'information : *l'open access* contribue à augmenter le volume d'information disponible et le risque d'infobésité. La capacité à apporter la bonne information attendue par les chercheurs est primordiale.

L'identité académique et les identifiants : l'identification non ambiguë des documents et des chercheurs est une problématique centrale dans un système de données liées. Comment créer et utiliser ces identifiants de manière optimale ?

Le travail collaboratif repose parfois sur des outils partagés, du terrain, au laboratoire, à la publication et à la communication : comment peuvent-ils contribuer à faciliter les collaborations ?

Les carnets de laboratoire : l'utilisation de carnets de laboratoire électroniques⁷, leur partage éventuel, leur archivage, constitue une problématique fondamentale, notamment pour la réutilisation de jeux de données ou l'examen de protocoles expérimentaux.

2 - https://ndownloader.figshare.com/files/8572213/preview/8572213/page_8_width_2000.png

3 - <https://www.softwareheritage.org>

4 - Roberto Di Cosmo et Stefani Zacchiroli

5 - <https://github.com/>

6 - <https://i4oc.org/>

7 - <https://openlabnotebooks.org/>