

Enoch Peserico
Université de Padoue, Italie

EN INFORMATIQUE, IMPOSSIBLE D'IMAGINER UNE SCIENCE NON OUVERTE

*propos recueillis et traduits de l'italien
par Francesca Musiani (Mines-ParisTech)*

Quelles sont les principales caractéristiques de la notion de « science ouverte » en informatique ?

Enoch Peserico : Pour un chercheur en informatique, la notion de « science ouverte » renvoie à plusieurs aspects qui en première approximation peuvent se distinguer ainsi : les collections de données ; la littérature ; les implémentations (surtout de logiciels, mais aussi de matériel). Les collections de données sont l'instrument qui permet la réalisation de tests reproductibles mesurant l'efficacité d'un nouvel algorithme, d'un programme ou d'un système. Il s'agit par exemple de *snapshot* de réseaux (la photographie, à un moment précis, de l'état de leurs connexions), du journal de fonctionnement d'un système, de traces d'exécution des programmes. Construire une collection de données est généralement compliqué et/ou coûteux, et il convient d'investir une seule fois dans l'effort de créer une base commune. Si la collection est ouverte, elle permet ensuite un accès plus facile au domaine de recherche concerné

– et cela entraîne, pour les créateurs de la collection, un nombre plus important de citations !

La littérature est peut-être l'aspect où les pratiques de science ouverte sont les plus présentes en informatique. Il est maintenant de coutume d'avoir une page Web personnelle avec ses publications : le coût de maintien de ces portails est très faible, et la pression des utilisateurs a joué un rôle essentiel pour convaincre même les plus récalcitrants des éditeurs d'autoriser cette forme d'autopublication des textes. Il y a encore relativement peu de revues en accès totalement libre, mais on observe une tendance croissante à utiliser des archives *pre-print* comme ArXiv¹, car publier un article au moyen de ces outils n'empêche pas une publication ultérieure dans des actes de conférence ou dans une revue ; qui plus est, cette publication ultérieure n'entraînera pas la suppression de l'archive ouverte. Pour l'instant, il est plutôt rare en informatique de publier une contribution uniquement dans une archive ouverte. Mais pour que

la contribution soit acceptée et publiée dans un livre ou une revue, les délais sont beaucoup plus longs.

L'avantage de l'archive ouverte est donc de faire rapidement connaître les sujets sur lesquels on est en train de travailler, ainsi que les résultats obtenus au fil de la recherche?

Oui, cela permet de rendre disponible de façon instantanée un résultat pendant la longue procédure de révision par les pairs: pas seulement afin d'établir une antériorité, mais aussi pour rendre la plus efficace possible la diffusion des résultats – ce qui est, après tout, la première motivation pour publier! Les archives ouvertes fournissent ainsi une possibilité d'accès immédiat: elles suppriment certaines contraintes (par exemple concernant le nombre de mots ou de pages) et éliminent la nécessité de recherches multiples sur un même sujet dans les différentes bases de données payantes. En somme, les principaux avantages de l'ouverture sont justement le confort et la rapidité. Cela me fait penser à l'Europe du traité de Schengen par rapport à l'Europe médiévale des taxes, douanes et droits d'entrée.

Les implémentations, c'est-à-dire la réalisation de logiciels et parfois de matériel, sont souvent, pour les chercheurs en informatique, la preuve concrète du fonctionnement d'un concept imaginé. Il y en a donc beaucoup dans les milieux universitaires, et elles sont généralement ouvertes. Dans beaucoup de cas non seulement elles fonctionnent bien, mais elles rivalisent avec les meilleurs produits du marché; elles concernent des projets de dimension très variables: des plus vastes à d'autres réduits et spécifiques, destinés à être incorporés dans d'autres projets. Il faut aussi noter que la définition d'ouverture est différente dans ce cas par rapport aux précédents: un article est ouvert si on peut le lire gratuitement; un logiciel est ouvert si on peut le lire, le démonter et le reconstruire.

Peut-on désormais envisager une science non ouverte?

À titre personnel, ma vie de chercheur n'a pas été beaucoup changée par les pratiques de science ouverte car, en tant que scientifique, je suis né et j'ai grandi avec. Au cours des dix années de mon «histoire scientifique», les concepts et pratiques de la science ouverte ont sûrement évolué, mais en même temps que moi. N'ayant jamais vécu sans science (au moins partiellement) ouverte, je peux seulement imaginer comme les choses se passeraient en son absence.

J'ai collaboré avec des secteurs où la disponibilité des données sous forme ouverte est presque nulle. Il y a différentes formules, soit la disponibilité exclusivement payante, soit un échantillon disponible gratuitement avec l'accès à une version plus large de la collection possible en échange d'une contribution, en termes monétaires ou en termes de collaboration au travail. Dans ces cas-là, j'ai constaté qu'il faut beaucoup plus de temps à la fois pour acheter des données, mais aussi pour décider si on les acquiert ou pas. Ce délai n'est pas toujours excessif car la récolte de nombreuses données nécessite de toute façon un effort *ad hoc*, à partir de questions spécifiques. Mais cela entrave néanmoins le travail; les phases préliminaires à la recherche – comme prendre l'initiative de «donner un coup d'œil» dans un champ qu'on ne connaît encore pas bien – sont pour le moins inhibées, quand elles ne sont pas carrément rendues impossibles.

S'il n'y avait pas de science ouverte en matière de développement des implémentations, cela aurait surtout, me semble-t-il, deux effets. Premièrement, si j'avais besoin d'un logiciel comme outil de recherche, je devrais avoir la disponibilité financière pour l'acheter. Ce qui implique non seulement de trouver les moyens nécessaires (c'est un investissement de taille quand il s'agit de logiciels qui servent d'instruments scientifiques), mais aussi d'avoir moins de liberté de choix: si le logiciel est ouvert, je peux en essayer dix

types différents et choisir celui qui s'adapte le mieux à mes besoins.

Deuxièmement, si la recherche consiste dans le développement de logiciels, ne pas en avoir d'autres auxquels on se confronte et qui vous aident à apprendre est un désastre. Si le logiciel est ouvert, chacun procède à partir du travail des autres et cela permet l'existence de grands systèmes qui n'existeraient pas autrement. Il est donc fondamental que les bases de cette construction soient librement accessibles.

Pour ce qui est des logiciels, l'ouverture paraît donc presque congénitale au développement même de la science. Et pour ce qui est de la littérature?

Sans archives ouvertes, il faudrait avoir accès par le biais de son institution à toutes les principales archives payantes (Springer, Elsevier, ACM, IEEE...). Si l'accès n'était pas fourni ou ne l'était que partiellement, cela paralyserait le travail jusqu'à ce que l'on décide de payer les abonnements – généralement chers – décidés par les éditeurs, avec les gênes additionnelles que cela implique: procédures de paiement compliquées, recherches séparées archive après archive, absence des mises à jour les plus récentes... Comme pour les collections de données, la phase de recherche qui serait la plus compliquée serait celle, par ailleurs cruciale, de la récolte d'informations préliminaires sur un sujet que l'on envisage de prospecter. Pour l'informatique, ceci est encore plus vrai aujourd'hui qu'il y a dix ans: avec plus de chercheurs qui publient et qui s'occupent d'une plus grande variété de questions, il est plus difficile de se tenir constamment informé des nouveautés.

Science ouverte, un avantage compétitif?

On ne peut donc pas se permettre de ne pas avoir de science ouverte. Une économie sans science ouverte serait destinée à bouger au ralenti: ne pas pouvoir faire usage, d'une façon ouverte, des résultats qui nous ont précédés entraînerait des coûts qui dépassent la simple cotisation monétaire que représente l'accès. Ainsi, même en supposant que cette cotisation soit destinée dans son entièreté à financer la recherche (ce qui n'est souvent pas le cas), ce coût différentiel se traduirait en une perte évidente pour la communauté scientifique.

En outre, l'ouverture a démontré être un modèle durable. C'est surtout vrai pour le logiciel, mais aussi pour les pratiques de publication: l'assertion selon laquelle «si quelque chose est gratuit maintenant, on ne pourra pas le financer dans le futur» s'est révélée une erreur.

La science ouverte est un avantage compétitif. Faire de la science aujourd'hui est déjà assez complexe sans que l'on doive faire face à des corvées évitables, telles celles qu'entraînerait le manque d'ouverture. Éviter ces obstacles – qui, même sans être incontournables, causent un important gaspillage d'efforts – est fondamental afin de rester compétitif dans un marché comme celui de la science, toujours plus dynamique, toujours plus actif et avec des barrières basses à l'entrée. Une science permettant l'ouverture est une science qui accélère et qui permet de se concentrer sur la création d'un savoir nouveau – plutôt que sur les façons de recueillir un savoir déjà élaboré par quelqu'un d'autre.

NOTE

1. Plus de 500000 articles en *Open Access* dans les domaines des mathématiques, physique, informatique (*computer science*),

biologie quantitative, finance quantitative et statistiques, hébergés par la bibliothèque de l'Université Cornell (État de New York).