

Bernard Lang

Inria et Aful

DES CORDONNIERS MAL CHAUSSÉS OU LES INFORMATIENS FACE AU LIBRE ACCÈS

entretien avec les coordonnateurs

Quels dispositifs la communauté des chercheurs en informatique a-t-elle mis en place pour rendre ses travaux disponibles et comment fonctionnent-ils ?

Bernard Lang: Les cordonniers sont souvent mal chaussés. La communauté informatique a été plutôt à la traîne. La première grande archive scientifique ouverte, XXX, aujourd'hui appelé ArXiv, fut créée pour les physiciens par Paul Ginsparg à Los Alamos. Il fut rejoint par des mathématiciens, et plus tardivement par des informaticiens du monde de la linguistique informatique. Depuis les années 1980, les informaticiens utilisaient plus facilement le réseau à titre individuel; de ce fait, ils ont sans doute moins senti le besoin d'un système de partage organisé. Par contre, ils ont très tôt introduit le partage ouvert des logiciels. Cela répondait à un besoin d'interaction avec les pairs et de validation par l'usage, mais aussi à l'idée que les logiciels, comme les théories, se peaufinent par des contributions

multiples. Cette façon de faire était même partagée par des acteurs industriels avant 1980.

La mise à disposition des travaux entre scientifiques est une pratique ancienne et fondamentale, mais qui se faisait d'individu à individu, donnant un rôle excessif aux phénomènes de réseaux et aux baronnies installées. Les revues furent une première forme d'ouverture, notamment vers les lecteurs, mais elles restaient sous le contrôle des comités de lecture, dont la lenteur préservait un avantage pour les échanges individuels. La publication numérique directe est une forme de démocratisation qui bénéficie aux auteurs autant qu'aux lecteurs; elle demande des approches nouvelles pour l'évaluation des contributions et leur sélection par le public; ces processus peuvent maintenant être dissociés de celui de publication. Il s'ensuit un très large spectre de modes de diffusion possibles. Cela se décline également pour les logiciels.

Le logiciel est une création particulière qui existe sous deux formes, une forme « source » créée et lisible par l'homme et une forme « objet » (en fait souvent plusieurs) utilisée par les machines. Des programmes spécialisés (les compilateurs) permettent de passer de la forme source à la forme objet, mais l'inverse est une opération artisanale et coûteuse dont les résultats restent difficilement exploitables. Cette dualité fait que la diffusion des logiciels sous forme objet, suffisante pour leur utilisation, ne dissémine en aucune façon les savoirs qui y sont incorporés, et ne laisse pratiquement aucune marge pour l'entretien du logiciel ou son adaptation à des besoins spécifiques. L'accès à la forme source, et le droit de la modifier et de la rediffuser sont donc un aspect essentiel de la science ouverte dans le monde du logiciel. On parle alors de logiciel libre ou *Open Source*. Cette problématique s'étend à d'autres disciplines, quand des logiciels sont partie intégrante de la conduite d'expériences ou de l'analyse et de l'exploitation de leurs résultats.

L'informatique déploie à la fois la recherche la plus ouverte (avec les collections de données ouvertes, les logiciels en « Open Source », l'autopublication fréquente avant la sortie « papier ») et la plus fermée (avec des brevets et des technologies de protection). Les deux logiques s'opposent-elles ou tendent-elles à se combiner plus ou moins harmonieusement ?

Il est difficile de comparer la recherche informatique avec les autres domaines. Le caractère immatériel des créations informatiques fait que la composante production des industries traditionnelles disparaît, et donc que tout développement, amené au stade de prototype utilisable, peut en principe être diffusée comme un produit par simple copie numérique. Il est donc difficile

de comparer avec d'autres secteurs qui, soit ne donnent pas lieu à des réalisations (les mathématiques notamment), soit proposent des réalisations dont la diffusion nécessite une infrastructure industrielle et commerciale, et qui ne relèvent plus vraiment de la recherche.

Ceci dit, il ne me semble pas que les chercheurs en informatique soient plus que d'autres enclins à protéger leurs travaux, ni qu'ils fassent davantage d'efforts pour les rendre publics. Sauf justement en ce qui concerne les logiciels dont la pertinence est souvent établie, ou du moins testée, par l'usage qui en est fait, voire par les extensions qui y sont intégrées par des tiers. Mais la recherche informatique est loin de se cantonner à la réalisation de logiciels. Bien entendu, comme dans tous les domaines, la recherche industrielle tend à être davantage protectionniste, d'autant plus qu'elle est proche de résultats exploitables commercialement. Pour la littérature scientifique, ce protectionnisme accru peut concerner la nature de ce qui est publié, mais non l'accès aux articles.

À ma connaissance, la plupart des chercheurs en informatique – et nombre d'entreprises industrielles – sont hostiles à la brevetabilité du logiciel, dont la légalité en Europe est, au mieux, floue. Le brevet est un mode de régulation économique destiné à favoriser l'innovation. Mais ses effets semblent négatifs dans le cas du logiciel, qui bénéficie d'ailleurs d'une protection par le droit d'auteur, à mon avis bien mieux adaptée¹. L'usage des brevets est sans doute beaucoup plus fréquent dans d'autres secteurs de la recherche. Cela peut avoir une justification économique quand l'exploitation industrielle de certains résultats exige des investissements, parfois lourds et risqués, que l'on peut vouloir protéger, même si la recherche elle-même est initialement publique et donc moins soumise à des contraintes de rentabilité. L'équilibre reste difficile.

S'il est vrai que l'on a de plus en plus de collections de données ouvertes, cela est loin d'avoir toujours

été le cas. La raison en est sans doute que le contrôle de ces collections est un avantage compétitif, pour les chercheurs ou pour les industriels, qui se protègent ainsi de la concurrence. En outre leur constitution est souvent onéreuse, voire quasiment impossible à reproduire pour des raisons techniques ou juridiques; et elle n'est peut-être pas assez considérée comme un produit de la recherche donnant du crédit scientifique. Par exemple la recherche en linguistique informatique a longtemps été lourdement ralentie, handicapée, par une politique restrictive d'accès et d'usage concernant notamment d'indispensables collections de données (corpus, lexiques, grammaires, etc.). Le changement est dû sans doute à une prise de conscience générale de ces problèmes grâce au succès du logiciel libre, et peut-être aussi grâce à la concurrence croissante de collections produites à moindre coût par des méthodes plus modernes, plus mécanisées ou plus coopératives.

De façon générale, du strict point de vue de la recherche, le protectionnisme semble toujours contre-productif, qu'il se fonde sur le droit ou sur la technologie. Cependant l'usage du droit des marques peut favoriser l'ouverture en remplaçant de façon plus souple le droit moral à l'intégrité de l'œuvre qui n'existe fort heureusement pas en matière de logiciel.

Y a-t-il en informatique des problèmes d'allocation des ressources liés à l'ouverture? Quel est le coût à payer de ces pratiques ouvertes?

Il faut savoir de quel point de vue on se place: intérêt général ou particulier. Le protectionnisme avantage souvent des intérêts particuliers au détriment de l'intérêt général. Ceci dit, gérer des logiciels ouverts n'est pas simple dès lors que l'on veut satisfaire une large communauté d'utilisateurs dans la durée. Cela peut représenter un investissement significatif, mais qui reste normalement bien inférieur au coût financier d'une externalisation industrielle, qui s'accompagnerait

d'une appropriation du logiciel introduisant un coût unitaire d'usage et donc limitant sa libre exploitation, que ce soit à des fins de recherche² ou de simple utilisation. Une question est de déterminer quels logiciels méritent un tel effort.

Une bonne solution est de faire émerger une communauté d'utilisateurs qui partagent ce coût. Les industriels pratiquent bien les *pools* de brevets. Mais, une collectivité, nationale ou locale, est elle-même une communauté qui peut déterminer que son intérêt, même simplement financier, est de maîtriser la propriété d'un logiciel pour en garantir l'usage libre et ouvert. Préserver le caractère libre d'un logiciel n'est pas incompatible avec l'externalisation de sa gestion (développement et accès) chez un prestataire privé. Il s'agit de décisions indépendantes.

Le développement et la gestion des logiciels libres est actuellement un thème de recherche en croissance. L'un des concepts clés est la notion de «forge» qui correspond à une infrastructure logicielle gérant la maintenance et l'évolution coopératives des logiciels, ainsi que leur mise à disposition des utilisateurs. Le concept n'est cependant pas limité aux logiciels. Des infrastructures similaires peuvent être utilisées pour la production de documents techniques ou encyclopédiques; Wikipedia en est un exemple.

Toute structure de recherche a pour souci de valoriser ses résultats, financièrement ou autrement. Confrontée aux résultats d'un chercheur, elle utilisera les mécanismes à sa disposition, dont la propriété intellectuelle et l'usage de licences appropriées. Un logiciel intéressant un large public et suffisamment fini pour être directement utilisable pourra être utilement diffusé avec une licence pérennisant son caractère libre (avec une licence dite *copyleft*). À l'inverse, un logiciel n'ayant d'intérêt que par son intégration dans un petit nombre de produits industriels coûteux pourra être plus utilement vendu sous licence propriétaire. Ce qui n'empêche

pas de le laisser en accès libre pour des usages expérimentaux. Tout est affaire d'analyse du contexte, de choix politiques, et d'implémentation juridique et financière de ces choix.

Quelles approches les chercheurs en informatique ont-ils des questions de propriété intellectuelle? Le mouvement d'«Open Science» a-t-il conduit à des problèmes de reconnaissance des droits ou de partage de la propriété?

La recherche est un monde fait de chercheurs, de laboratoires, d'institutions, de disciplines et de pays. Il n'y a aucune raison que les intérêts, les choix, les comportements soient perçus comme les mêmes à chaque niveau, et la carrière d'un chercheur peut être en opposition avec l'intérêt de son laboratoire. Le renforcement de la concurrence (par exemple par des systèmes de primes) peut aussi exacerber ces différences. Ceci dit, le chercheur en informatique est en général assez ignorant des questions de propriété intellectuelle. Il me semble surtout concerné par la protection des idées et particulièrement de leur paternité, mais cela ne relève souvent pas de la propriété intellectuelle. La situation est bien sûr différente en ce qui concerne les laboratoires ou les institutions qui sont amenés à établir des protocoles de coopération avec des industriels ou d'autres entités de recherche, par exemple dans le cadre de finance-

ment de grands projet européens. Pour la coopération Microsoft-Inria, il y avait conflit entre une volonté de laisser ouverts les résultats de la recherche et le désir de privatisation à des fins d'exploitation commerciale. On a trouvé une solution intermédiaire par l'usage d'un type particulier de licence (licence libre, de type BSD, permettant l'appropriation des créations dérivées).

Les structures institutionnelles ont le souci de valoriser la recherche, mais parfois une vue assez naïve (ou idéologique) de ce qu'est la valorisation, trop souvent assimilée aux retours financiers. L'ouverture peut représenter une perte financière, mais un gain en réputation, en attractivité, en influence tant pour les individus que pour les institutions.

Il faut cependant se garder d'une confusion. L'ouverture des résultats de la recherche est le choix d'un mode de travail et de valorisation fait dans l'intérêt du chercheur, des organisations, du public ou de la science elle-même. L'ouverture de l'accès à la littérature scientifique relève de considérations plus bêtement économiques : tout le monde (sauf quelques éditeurs) a avantage à ce que la littérature scientifique soit en libre accès, sous réserve de bien gérer évaluation et sélection. Le contrôle de l'accès aux articles par la propriété intellectuelle permettait paradoxalement de financer leur diffusion. En changeant les structures de coût, la numérisation a rendu ce contrôle inutile, et donc nuisible.

NOTES

1. Bernard LANG, «Brevetabilité du logiciel: le point de vue d'un chercheur en informatique», in Bernard REMICHE (dir.), *Brevet, Innovation, Intérêt général - Le brevet pourquoi et pour faire quoi?*, Actes du Colloque de Louvain-la-Neuve, organisé

par la Chaire Arcelor, Larcier, 2006, p. 385-413. En ligne sur <<http://www.datcha.net/ecrits/liste/arcelor.pdf>>.

2. Il faut garder à l'esprit que la validation d'un travail de recherche en logiciel passe souvent par une mise à la disposition des usagers.