

**Danièle Bourcier**

*Cersa, Université Paris II*

## **SCIENCE COMMONS : NOUVELLES RÈGLES, NOUVELLES PRATIQUES**

Nous sommes tous conscients qu'Internet va changer les façons de *faire* de la science. Les nouveaux modes d'accès et de circulation de l'information scientifique sur le Web ne sont qu'une des facettes de ce changement. La seule diffusion de la communication ne conduira pas à des sauts qualitatifs décisifs. Pour comprendre le véritable changement, il est nécessaire d'avoir une vue de la façon dont travaillent les professionnels, quels qu'ils soient, et surtout de repérer et de dépasser les blocages, tous les types de blocage, qui empêchent cette communication d'être réellement efficace.

Dans un article sur les rapports entre droit et technologies de l'information, j'avais fait une distinction entre les interactions fortes et faibles qui existent entre deux domaines de savoirs (Bourcier, 2006). L'informatique a en effet influencé toutes les pratiques dans les disciplines et les métiers: c'est ainsi que de nouvelles disciplines comme l'informatique juridique ou l'informatique médicale ont vu le jour. Aujourd'hui,

la physique ne se conçoit plus que liée à l'informatique et aux réseaux. Mais l'informatique, si elle a changé des façons de travailler, n'a pas toujours affecté comme elle le fait actuellement le cœur du métier lui-même. Quand on parle des transformations du travail scientifique à travers les techniques du Web et les politiques qui les soutiennent (accès ouvert, par exemple), on transforme les *conditions institutionnelles de la production* au-delà des modes de communication scientifique. Et on crée les conditions pour en renouveler les résultats.

### **Les conditions d'une transformation de la communication scientifique**

Les sciences dépendent de l'accès et de l'utilisation de données, qu'elles soient textuelles ou factuelles.

Améliorée par des développements dans le stockage électronique et la capacité de calcul, la recherche scientifique gère de plus en plus de données dans toutes les disciplines. Les avancées rapides dans les technologies numériques et les réseaux au cours des deux dernières décades ont modifié significativement la façon dont les données et les informations peuvent être produites, diffusées, gérées et utilisées dans les sciences et les autres domaines de l'activité humaine. Des opportunités nouvelles ont été créées, accélérant les progrès de la science et de l'innovation. Ces développements viennent principalement du mouvement qui a émergé de la *peer production* (co-production entre pairs) formelle et informelle, c'est-à-dire de la «co-production» à travers des plates-formes adaptées, et de la diffusion globale de l'information en mobilisant la coopération – dans un réseau distribué et ouvert – de communautés de connaissances. La science dans l'environnement numérique est le résultat d'un mélange complexe de partages et d'échanges sociaux entre pairs.

### ***Faire un Web scientifique***

Dans le travail scientifique, ce n'est plus seulement l'accès aux données qui importe, mais les nouvelles formes de partages de données. La «science ouverte» vise à aller au-delà de l'accès ouvert aux résultats de la recherche, en facilitant un partage non seulement des données et de l'information mais des données brutes et des outils logiciels. L'examen systématique des preuves disponibles, nécessaire aux expériences en cours, ainsi que la mise à disposition de méthodes adaptées au contexte de ces mêmes expériences, visent à l'interprétation immédiate des résultats, y compris inattendus.

### ***Dépasser les obstacles juridiques***

Dans beaucoup de disciplines, les processus de recherche dépendent de décisions politiques, de lois et de pratiques, qui sont pour la plupart opaques aux scientifiques eux-mêmes. Au cours des dernières décades, les lois sur la propriété intellectuelle ont évolué dans une direction de fermeture et non d'ouverture des droits. Par exemple, auparavant, la protection ne portait pas sur les données brutes; on brevetait le piège mais non les données sur le comportement de la souris; un article était protégé mais pas les données sur lesquelles il s'appuyait. Dans certains pays comme les États-Unis, pays de *Common Law* où les droits moraux ne sont reconnus qu'exceptionnellement, les travaux scientifiques subventionnés par le gouvernement fédéral tombaient immédiatement dans le domaine public, sans droits d'auteur.

Mais la tendance a changé radicalement. Dans le domaine de la génétique, le droit des brevets a créé un droit sur les faits bruts, comme les séquences de gènes. Dans d'autres domaines, des contrats complexes ont modifié la propriété intellectuelle sur les bases de données, avec de nombreuses limitations d'usages et de réutilisation. L'auto-archivage par les chercheurs est devenu difficile. Ces tendances sont parfaitement opposées aux fondements de la propriété intellectuelle, qui prévoyait qu'on ne pouvait s'approprier des faits ou des idées mais seulement des inventions et des créations originales et formalisées. Le paradoxe est que, à une époque où les technologies permettent un accès mondialisé et un traitement distribué des données, des restrictions juridiques empêchent de lier ces mêmes données entre les banques de données.

## **Qu'est-ce que Science Commons ?**

Science Commons est un projet lancé en 2005 par des scientifiques, des juristes et des entrepreneurs du Web qui, après avoir repéré précisément les blocages de la communication dans le monde scientifique, ont appliqué une philosophie issue des Creative Commons et ont expérimenté à travers le Web, de façon pragmatique, les outils et méthodes qui pouvaient réduire ces obstacles. En France, l'Institut Pasteur a suivi le mouvement et met désormais à la disposition des industriels et des organismes de recherche extérieurs un vaste portefeuille de matériels biologiques issus des travaux de recherche menés par divers laboratoires pasteuriens ; ces matériels sont disponibles sur simple requête (Institut Pasteur, 2010).

### ***De Creative Commons à Science Commons***

En réaction au mouvement de fermeture des droits de la propriété intellectuelle s'est ainsi créé le mouvement Science Commons (SC), inspiré du mouvement Creative Commons (CC). Deux facteurs les séparent : tout d'abord, les licences CC concernent les créateurs individuels et leurs droits d'auteur, alors que le projet SC a une portée plus large, incluant les institutions dont dépendent les scientifiques. De plus, les revues scientifiques demandent de plus en plus que leurs auteurs renoncent à leur *copyright* et ils le font pour être cités dans des publications à *indice d'impact* élevé. Il y a donc un problème d'action collective : personne n'a d'incitations directes à changer le système (Wilbanks, 2005).

Science Commons, aidé par un groupe d'experts issus des sciences et du droit, a retenu la philosophie du partage dans le respect du droit des auteurs qui

avait fait le succès de CC dans le domaine des arts et de la culture. Les deux projets ont de nombreux points communs, à commencer par les acteurs. On retrouve ainsi dans le projet des membres du conseil exécutif de Creative Commons : Hal Abelson, chercheur au MIT ; James Boyle, Michael Carroll, Lawrence Lessig, experts en propriété intellectuelle ; John Wilbanks, entrepreneur en bio-informatique.

### ***Pourquoi Science Commons ?***

Science Commons a pour objectif de créer des stratégies et des outils pour rendre le Web scientifique plus rapide. Trois principes ont été fixés. Le premier est de *rendre la recherche réutilisable* grâce à des politiques et des outils qui aident les individus et les organisations à rendre leur production accessible, et à indiquer, sur leurs résultats et leurs données, cette possibilité de réutilisation ; un Protocole *Open Data* permet à la communauté scientifique mondiale de mettre en commun leurs ressources, quel que soit leur statut juridique. Le deuxième est de *permettre l'accès immédiat à des outils* grâce à des contrats standard qui apportent des économies d'échelle aux chercheurs de sorte que ces mêmes chercheurs puissent dupliquer, vérifier et étendre leur recherche. Le troisième, enfin, est d'*intégrer des sources fragmentées* d'information en fournissant aux chercheurs les moyens de trouver, d'analyser et d'utiliser des données de sources disparates en marquant et intégrant l'information à travers un langage commun et traduisible dans la machine.

Le programme consiste ainsi à offrir concrètement un spectre d'outils et de ressources susceptibles d'être traités en accès ouvert : le transfert de matériaux biologiques (ADN, lignées de cellules, anticorps, plasmides...) ; une plate-forme de gestion de connaissances en accès ouvert pour la recherche en neurosciences ; un

*Common* sanitaire, coalition d'acteurs intéressés à considérer la santé comme bien commun.

Fer de lance du projet, le site Web de Science Commons propose divers services (Science Commons, 2010). On peut y marquer les données pour les réutiliser de façon légale (*Scholar's Copyright Project*). On y apprend à faciliter le transfert des matériaux physiques (*Biological Materials Transfer Project*), à intégrer les résultats, les données, les matériaux et les services émanant de sources disparates (*NeuroCommons* et *Health Commons*), etc. Des standards ouverts sont développés pour faciliter l'émergence d'une infrastructure collaborative qui accélère la traduction de données en découvertes.

### ***Les instruments du partage de données***

Quand recherches et données sont rendues publiques, des régimes propriétaires ou des contrats empêchent souvent de changer les formats ou les langages, et d'intégrer les données, l'enrichissement sémantique, ou la fouille de données. Ces restrictions limitent l'impact des recherches publiées et empêchent d'exploiter les potentialités du Web pour accélérer les découvertes scientifiques.

En réaction, le *Scholar's Copyright Project* propose des outils et des ressources pour étendre et améliorer l'accès ouvert. La plupart des systèmes et formats existant actuellement sont issus de la civilisation pré-numérique. L'accès ouvert est un *pré-requis* pour trouver de nouvelles façons d'exploiter la valeur de vastes gisements de recherches. Le *Science Commons Protocol for Implementing Open Access Data* est une méthode qui garantit que la base de données utilisée peut être légalement intégrée à une autre, eu égard au pays d'origine. Le protocole n'est pas une licence ou un outil juridique, mais une méthodologie et un document de

«bonnes pratiques» pour créer de tels outils juridiques et marquer les données comme appartenant au domaine public. Deux outils juridiques utilisent ce même protocole: l'*Open Data Commons Public Domain Dedication and License* et l'option *Creative Commons CC0*.

De plus en plus souvent, les fondations et agences qui alimentent la recherche demandent d'adopter l'accès ouvert pour les résultats scientifiques. À la fin de 2008, plus de 1 000 journaux scientifiques *peer-reviewed* mettaient en œuvre une philosophie proche de Creative Commons, ce qu'illustre la *Public Library of Science* (plos.org). Tous utilisent la licence Creative Commons «Attribution» (BY).

Les éditeurs de journaux juridiques sont intéressés par le Programme d'accès ouvert (Open Access Law Program – cf. Science Commons, 2010). 35 journaux s'y sont engagés depuis son lancement. Ce programme fournit un ensemble de ressources proposant l'accès ouvert en science. Auteurs et éditeurs s'engagent sur ces principes et un modèle gratuit de contrat est à la disposition des auteurs et des journaux.

Baucoup de scientifiques font de l'auto-archivage après avoir publié dans des revues. Cependant les rapports avec les éditeurs sur cette restriction sont parfois confus. Science Commons a créé un outil d'aide au copyright *Scholar's Copyright Addendum Engine* (Science Commons, 2010) qui aide les auteurs à négocier les droits qu'ils veulent. Cet outil, appelé *Addendum Engine*, met à disposition une interface pour générer une annexe au contrat (1 page) prête à signer. Ce document garantit que vous pouvez mettre votre recherche librement sur le Web, pour la publication, la pré-publication, ou après une période de temps déterminée. Les institutions peuvent intégrer cet outil sur leur site.

Ainsi, par la licence «Accès-réutilisation», «vous conservez les droits nécessaires pour accorder au public une licence Creative Commons Attribution - non commercial (BY-NC) ou une licence similaire qui

permet au public la réutilisation ou la rediffusion de votre article, tant que vous êtes crédité comme l'auteur et aussi longtemps que l'usage est non-commercial».

Vous pouvez également choisir d'autres options, tels « Accès immédiat » ou « Accès retardé » :

– « *Accès immédiat*. Vous conservez les droits suffisants pour afficher une copie de la version publiée de votre article (généralement sous forme de PDF) en ligne immédiatement sur un site qui ne fait pas payer pour l'accès à l'article. »

– « *Accès retardé*. Vous avez également le droit d'offrir votre article à la consultation, immédiatement après la version finale de votre texte, et après examen par les pairs, sur un site qui ne fait pas payer l'accès à l'article, mais vous ne devez mettre la version publiée à la disposition du public que six mois après la date de publication. »

### ***Le projet Neurocommons***

Le projet Neurocommons a créé une plate-forme de gestion de connaissances *Open Source* pour la recherche biologique, et particulièrement les neurosciences. La première phase a été testée en 2007. La deuxième phase met en place un système informatique d'analyse de données. Ces deux éléments représentent une plate-forme standard avec des contenus ouverts sur le Web.

On cherche à provoquer des liens entre les agences à travers le *Biological Materials Transfer Project*. Neurocommons combine édition et annotation fondées sur les *commons* et création d'une communauté de logiciel libre autour d'une plate-forme d'analyse (pour les aspects techniques, se reporter à Science Commons, 2010).

Neurocommons a émergé de l'histoire de Creative Commons et du Web sémantique. John Wilbanks avait créé le *Semantic Web for Life Sciences* au *World Wide*

*Web Consortium* (w3c.org) tout en dirigeant une *start-up* en bio-informatique. En effet, la portée de la connaissance issue du secteur public et du secteur commercial avait conduit certains experts dans le champ de la gestion de la connaissance pharmaceutique à construire des méthodes de production fondées sur des *commons* et des partages de connaissances avant leur exploitation industrielle. Aucune société pharmaceutique ne peut en effet capturer, représenter et utiliser toute la connaissance disponible sur le Web. La recherche publique, avec l'aide de cette même industrie s'est investie dans un grand projet de collecte ontologique des relations et des termes communs à partir des textes et des bases de données non structurés pour fournir une "carte" de la sémantique implicite, partagée entre les secteurs scientifiques, et à l'intérieur des domaines. Les sociétés Pfizer et Bionogen ont apporté des éléments dès les premières discussions et le système a été modélisé et porté sur un service (hélas propriétaire) qui appartient à Novartis.

Actuellement l'équipe Neurocommons vise à rendre publique et à améliorer la base de données ouverte d'annotations en les étendant aux résumés d'articles biomédicaux à travers une base de codes ouverts pour la biologie informatique. Les plus grands fonds en neurosciences sont intégrés dans le graphe d'annotation. Avec ce système, les scientifiques peuvent charger, dans des listes de gènes, des données venant des robots des laboratoires et revenir à ces listes avec une information pertinente fondée sur une connaissance partagée. Ils peuvent ainsi trouver les articles et fragments de données d'où vient cette information, bien plus rapidement qu'avec Google ou tout autre outil de recherche documentaire, à cause du système qui permet de créer des liens avec la structure sous-jacente. Enfin, tout le monde est incité à mettre ses publications dans le même système et à le rendre plus performant.

L'échelle de l'effort demandé pour cartographier la connaissance est immense, d'autant plus que ce travail

requiert l'intervention humaine à un certain moment du processus (la fouille de textes est nécessaire mais non suffisante). Le schéma général est fondé sur le modèle Wikipedia, incrémental et décentralisé. La deuxième phase est le logiciel d'analyse de données fourni par Science Commons sous la licence BSD *Open Source*. Sans ce logiciel, le Web n'aurait ni moteur ni fonction de recherche.

## Le futur des commons scientifiques

Le recours aux *commons* scientifiques<sup>1</sup> s'inscrit dans une évolution qui dépasse le monde scientifique lui-même. Le prix Nobel d'économie a été attribué en 2009 à une ardente avocate des *commons*, Elinor Ostrom : elle s'intéressait depuis plusieurs décades aux nouveaux modes de gouvernance des ressources naturelles (eau, forêt), caractérisées par définition par leur rareté et un besoin de préservation (Ostrom, 1990). Les *commons* de la connaissance et de la création, aidés par les technologies du numérique, suivent une tendance résolument parallèle : dans un univers scientifique caractérisé par des obstacles juridiques et techniques mais aussi par des exigences d'efficacité sinon de rareté (tragédies dites *anti-commons*), il faut inventer des institutions et des normes fondées sur la non-exclusivité des ressources. Les *commons* impliquent donc, à côté de ces ressources partagées, une gestion commune, ni publique ni privée, pour une meilleure utilisation d'un patrimoine qui intéresse toute l'humanité. Le projet SC transpose la philosophie et la méthodologie du projet CC dans le monde de la recherche scientifique. L'objectif de SC est de dégager la voie juridique et technique pour accélérer les découvertes scientifiques à travers le monde. Le processus

de production scientifique a toujours été soutenu par une série de politiques, de lois et de pratiques qui ont été en grande partie invisibles, même pour ceux qui travaillaient dans les sciences elles-mêmes.

Tentons un résumé. Premièrement, les données brutes doivent rester dans le domaine des *commons* scientifiques. Deuxièmement, les résultats issus de subventions du gouvernement et des collectivités, ou de ressources publiques, ont vocation à être en accès ouvert au public. Troisièmement, dans les sciences elles-mêmes, et notamment dans les universités, une tradition sociologique forte – parfois appelée la tradition *mertonienne* de la science ouverte – devrait décourager l'exploitation exclusive de données. Ces trois principes fondamentaux existaient même avant la révolution industrielle. De même, la publication scientifique a une tradition de longue date. Les technologies modernes, notamment l'utilisation du *World Wide Web*, ont changé les mécanismes de livraison et de duplication des documents. Dans de nombreux domaines, les résultats sont publiés presque aussi vite qu'ils sont découverts. Mais le droit d'auteur a évolué à un rythme différent. Les progrès dans les technologies modernes, combinés à un système juridique qui a été conçu pour l'ère analogique, vont provoquer des obstacles artificiels et freiner l'innovation.

Pour faire contrepoids, une communauté mondiale importante et dynamique a mis sur pied, pour appuyer le libre accès à la littérature scientifique, des plates-formes rendant les données « numériques, en ligne, gratuites et libres de la plupart des droits d'auteur et des restrictions de licence ». En Amérique du Nord, des fondations de recherche importantes, comme le *Howard Hughes Medical Institute*, ont adopté des politiques novatrices pour exiger l'accès libre aux résultats de recherche, les *Institutes of Health* exigent maintenant un accès ouvert à la recherche financée, des établissements d'enseignement (Harvard, MIT) recommandent que les travaux

soient publiés «ouvertement», la plupart des grands journaux ont accordé aux auteurs le droit à l'autopublication des versions successives de leurs articles.

Pourtant, dans de nombreux cas, les questions juridiques restent en suspens. Comment faciliter la réutilisation de la recherche tout en assurant que les auteurs et les éditeurs puissent en conserver l'attribution ? Quel est le meilleur moyen de permettre l'intégration des données recueillies dans différents systèmes juridiques ? De quels types d'infrastructures juridiques et politiques avons-nous besoin pour faciliter le transfert du matériel nécessaire, vérifier les résultats et enrichir la recherche ?

Nous avons essayé à partir d'un projet de Web scientifique ouvert de montrer comment de nouvelles interactions fortes entre disciplines (l'informatique et la science par exemple) pouvaient changer les façons de communiquer, à condition de se situer dans un monde ouvert. En France, les derniers débats sur Google et les œuvres du patrimoine, ainsi que sur la loi dite Hadopi, ont montré l'intérêt immense de l'opinion publique et des scientifiques sur ces sujets et les oppositions voire les

incompréhensions qu'ils mettaient en évidence. Ce n'est plus seulement l'accès aux données qui importe mais l'évolution des formes de production et de communication scientifique, les politiques de partages et d'échanges d'informations scientifiques, l'*Open Access* et l'innovation. Enfin, le droit doit s'adapter aux nouvelles pratiques du monde scientifique pour éviter des effets contre-productifs pour les avancées scientifiques.

L'éthique ne peut être seule à intervenir, en créant une *soft law*, dans des domaines si compétitifs. Depuis quelques années, les débats éthiques se sont déplacés sur le terrain de la communication, alors que c'est le droit lui-même qui est interpellé comme instrument pour des politiques scientifiques clairement débattues. L'exemple des cellules souches montre que le défi concerne moins le statut de l'embryon que les règles du jeu à imposer à la communauté scientifique et à l'industrie «bio-tech». Ces règles portent désormais sur la gratuité, sur l'accès aux connaissances, sur la brevetabilité, sur la protection des données et des personnes, règles qui peuvent traditionnellement être traitées par le droit.

## NOTE

1. Nous avons gardé résolument le mot anglais *commons* car la notion de biens communs recouvre une acception moins «juridique» (Bourcier, 2009).

## RÉFÉRENCES BIBLIOGRAPHIQUES

BOURCIER, D., «Droit, administration et technologies de l'information et de la communication», in GANASCIA, J.-G. (dir.), *Communication et connaissance, supports et médiations à l'âge de l'information*, Paris, CNRS Éditions, 2006.

BOURCIER, D., «Digital Commons Works: Thinking governance», in BOURCIER, D., CASANOVAS, P., DULONG DE ROSNAY, M., MARACKE, C. (dir.), *Intelligent Multimedia. Sharing Creative Works in a Digital World*, Florence, European Press Academic Publishing, 2009.

INSTITUT PASTEUR, *Technology Transfer. Matériel biologique*, 2010. En ligne sur <<http://www.pasteur.fr/ip/easysite/go/03b-000012-005/produits-et-services/materiel-biologique>>.

OSTROM, E., *Governing the Commons. The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.

SCIENCE COMMONS, *Biological Materials Transfer Project. Building the Clearinghouse for Research Tools*. 2010. En ligne sur <<http://sciencecommons.org/projects/licensing>>.

SCIENCE COMMONS, *Health Commons. "It's time to bring the same efficiencies to human health that the network brought to commerce and culture. And to do that, it takes a Commons"*, 2010. En ligne sur <<http://sciencecommons.org/projects/healthcommons>> ou <<http://www.healthcommons.net>>.

SCIENCE COMMONS, *Neurocommons Technical Overview*, 2010. En ligne sur <[http://sciencecommons.org/projects/data/nc\\_technical\\_overview](http://sciencecommons.org/projects/data/nc_technical_overview)>.

SCIENCE COMMONS, *NeuroCommons*, 2010. En ligne sur <<http://sciencecommons.org/projects/data>> ou <<http://neurocommons.org>>.

SCIENCE COMMONS, *Open Access Law Program*, 2010. En ligne sur <<http://www.sciencecommons.org/projects/publishing/oalaw>>.

SCIENCE COMMONS, *Scholar's Copyright Addendum Engine*, 2010. En ligne sur <<http://scholars.sciencecommons.org/#form>>.

SCIENCE COMMONS, *Scholar's Copyright Project. Making Research and Data "Re-Useful"*, 2010. En ligne sur <<http://sciencecommons.org/projects/publishing>>.

WILBANKS, J., «What is Science Commons?», 2005. En ligne sur <<http://creativecommons.org/weblog/entry/5695>>.