

La diffusion et le partage de l'information scientifique

Marc Couture

Chapitre 7 de l'ouvrage

Propriété intellectuelle et université

Entre la privatisation des idées et la libre circulation des savoirs

publié en 2010 aux Presses de l'Université du Québec

Ce document est disponible dans Archipel (UQAM) :

<http://archipel.uqam.ca/3460>

Informations sur la version papier, dans le site des PUQ :
<http://puq.ca/catalogue/collections/propriete-intellectuelle-universite-3699.html>

29 octobre 2010

LA DIFFUSION ET LE PARTAGE DE L'INFORMATION SCIENTIFIQUE

S'il est une valeur que chérissent les scientifiques, c'est bien celle du partage de l'information au sein de leur communauté, sans égard aux frontières politiques, économiques ou sociales qui, dans d'autres sphères de l'activité humaine, entravent souvent ce partage. Cela n'empêche toutefois pas d'observer, au sein même de l'activité quotidienne de la recherche, la présence d'attitudes, de pratiques et de systèmes qui ont pour conséquence de réduire la diffusion ou l'accès à l'information et, par conséquent, son partage. Et c'est la plupart du temps en invoquant la notion de propriété de cette information que l'on justifie ces situations. Mais tout d'abord, il faut définir en quoi précisément consiste cette « information scientifique¹ » que l'on souhaite – ou souhaiterait – mieux partager.

Il convient de distinguer deux types d'information scientifique : d'une part, les données de recherche ou les informations sur celles-ci; d'autre part, les descriptions des recherches. Pour ces deux types d'information, les enjeux sont similaires. Tout d'abord, à qui appartient l'information? A-t-elle même un propriétaire? Le cas échéant, qu'est-ce que cette propriété permet de faire, ou d'empêcher? Comment faire en sorte que les pratiques et les règles touchant la gestion de cette information en favorisent la diffusion et le partage? Finalement, comment tirer profit des technologies de l'information et des communications, notamment l'Internet, qui offrent la possibilité d'accéder à de formidables quantités d'information d'une manière conviviale, sans contrainte de temps ou d'espace?

Nous traiterons dans un premier temps des *descriptions des recherches*, qui se divisent en deux grandes catégories :

- les descriptions des objets, systèmes ou méthodes créés, construits ou mis au point lors de la recherche, que l'on retrouve dans les thèses et les mémoires, les rapports techniques, les demandes de brevets, etc.;
- les descriptions de l'ensemble d'une recherche, incluant les sources, les hypothèses et questions de recherche, les travaux accomplis ainsi que les résultats² et conclusions obtenus, qui font l'objet des divers types de communication scientifique : communications orales prononcées dans les congrès ou colloques, avec ou sans comptes rendus, articles dans des revues scientifiques,

1. Dans ce texte, le qualificatif « scientifique » désigne tant ce qui relève des sciences de la nature (incluant les mathématiques) que des sciences humaines et sociales ainsi que, dans d'autres domaines comme les arts et la littérature, les activités qui ont adopté un mode de fonctionnement similaire, notamment le recours à l'évaluation par les pairs fondée sur des critères faisant consensus dans la discipline concernée.

2. L'expression « résultats de recherche », que l'on rencontre souvent dans les discussions entourant la diffusion des informations scientifiques, est particulièrement élastique. À un extrême, elle désigne uniquement le résultat final du traitement ou de l'analyse des données, qui fonde la discussion et la conclusion des travaux; à l'autre, elle englobe tous les produits (ou extraits) de la recherche, des données brutes aux publications.

chapitres d'ouvrages collectifs, livres et manuels, communications, articles ou livres de vulgarisation, etc.; ces descriptions portent notamment sur le contexte, les justifications, les analyses, les interprétations et les conclusions de la recherche.

Nous examinerons en premier lieu ce qu'il en est des descriptions des recherches, en nous intéressant tout particulièrement aux publications scientifiques (notamment les articles)³. Nous aborderons ensuite la problématique des données de recherche, où les notions de propriété et de gestion peuvent être rattachées non seulement à des considérations de nature juridique, mais aussi à des préoccupations relevant de l'éthique et de la probité scientifiques. Nous verrons que dans les deux cas, de nouveaux modèles de diffusion et d'accès sont en voie de révolutionner la dynamique des échanges d'information scientifique.

La diffusion et l'accessibilité des publications scientifiques

Les scientifiques et, avant eux, les « philosophes » et « savants » qui se sont voués à l'avancement des connaissances, ont compté de tout temps sur l'écrit tant pour établir un dialogue avec leurs pairs que pour s'assurer que leurs idées étaient diffusées auprès de leurs contemporains et léguées à leurs successeurs. Des manuscrits médiévaux patiemment recopiés et enluminés à la main, on est passé aux ouvrages imprimés : livres d'abord, aux XV^e et XVI^e siècles, mais aussi, à compter du milieu du XVII^e siècle, périodiques rendant compte des séances des sociétés savantes comme les *Philosophical Transactions* de la Royal Society anglaise et les *Mémoires* de l'Académie royale des sciences française (Guédon, 2001). C'étaient là les ancêtres des revues scientifiques qui allaient devenir, à partir du XX^e siècle, le véhicule par excellence de la communication scientifique, en sciences de la nature d'abord mais aussi en sciences humaines et sociales, quoique dans ce domaine les monographies continuent d'y jouer un rôle important, sinon dominant (Larivière, Archambault, Gingras et Vignola-Gagné, 2006).

Or, ces publications obéissent à une dynamique complètement différente de celle des œuvres littéraires (au sens courant du terme) de même type – revues, magazines ou livres – que l'on retrouve sur le marché. En effet, ces dernières sont créées par des auteurs qui visent à en tirer un revenu, parce qu'il s'agit là de leur gagne-pain, ou du moins d'une activité analogue à un emploi secondaire auquel ils consacrent une bonne partie de leur temps libre. De leur côté, les maisons d'édition acceptent de publier, parmi les œuvres qui leur sont soumises, celles qui sont susceptibles de générer des revenus suffisants pour, selon le cas, assurer leur simple survie ou générer les profits attendus des entreprises qui les gèrent. On le voit, il s'agit là d'une transaction essentiellement économique entre l'auteur et la maison d'édition, où le droit d'auteur constitue la monnaie d'échange. Cette transaction fait l'objet de négociations dont l'enjeu est avant tout l'importance des redevances versées à l'auteur en échange de la cession de son droit d'auteur ou de l'octroi de la licence permettant à l'éditeur de faire son travail, c'est-à-dire produire et diffuser l'œuvre.

Pour les œuvres à caractère scientifique ou disciplinaire, à quelques exceptions près (manuels, articles de vulgarisation), la situation est complètement différente. Tout d'abord, les auteurs sont en règle générale des employés déjà rémunérés, le plus souvent à partir d'un financement public, pour un travail qui comprend normalement la création de ce genre d'œuvres. On songe bien sûr aux professeurs et chercheurs universitaires, mais on peut aussi inclure les chercheurs des entreprises et

3. La diffusion des descriptions d'inventions et d'autres innovations techniques (comme les obtentions végétales) constitue un enjeu très différent, qui est traité au chapitre 6.

des centres de recherche, de même que les étudiants qui créent ces œuvres dans le cours de leurs études. Ensuite, ces auteurs n'espèrent pas retirer d'avantage financier direct de la publication de ces œuvres; leur intérêt réside plutôt dans la satisfaction des critères présidant à la progression de leur carrière (obtention d'un diplôme, embauche, renouvellement de contrat, octroi de la permanence, promotion, subventions de recherche) et dans l'atteinte d'une notoriété dans leur domaine, qui se manifeste entre autres par toute une panoplie de prix et distinctions. Cette notoriété est favorisée en partie par le nombre d'œuvres publiées, mais aussi par le nombre de citations que celles-ci reçoivent et la réputation des maisons d'édition ou des revues qui les ont acceptées. Ces dernières, par ailleurs, ne cherchent pas directement à sélectionner des œuvres susceptibles individuellement d'augmenter leurs ventes. Ce qu'elles désirent plutôt, c'est convaincre les acheteurs potentiels, essentiellement les bibliothèques des universités et établissements de recherche, de la pertinence et de la qualité générale des œuvres qu'elles publient. Elles tentent d'obtenir cette reconnaissance par la réputation de leurs équipes éditoriales, formées en majorité de chercheurs qui acceptent d'évaluer bénévolement les manuscrits soumis, par leur facteur d'impact (le nombre moyen de citations par article)⁴ et par le taux de rejet de ces manuscrits, qui peut atteindre plus de 90 % pour les revues les mieux cotées.

Dans ces conditions, on peut penser que le recours au droit d'auteur comme monnaie d'échange constitue une simple formalité : l'auteur qui recherche à la fois la plus grande reconnaissance et la plus large diffusion pour ses travaux ne trouve pas *a priori* exorbitant d'avoir à céder son droit d'auteur sans compensation – condition encore exigée dans la plupart des cas – dans la mesure où cela constitue le meilleur, sinon le seul moyen d'atteindre ces objectifs.

Ainsi, malgré certaines critiques portant non pas sur les enjeux relatifs au droit d'auteur mais plutôt sur le processus d'évaluation par les pairs, que l'on estime responsable des longs délais de parution et dont on met parfois en doute la capacité d'assurer la qualité de l'ensemble des œuvres publiées, le système décrit plus haut a fonctionné sans soulever de contestation majeure dans la communauté des chercheurs jusqu'à la fin du XX^e siècle. Une nouvelle conjoncture, culminant avec ce qu'on a appelé la « crise des périodiques », allait alors apporter de l'eau au moulin des partisans de nouveaux modes de diffusion de ces œuvres.

La « crise des périodiques »

Jusqu'au milieu des années 1980, malgré l'augmentation exponentielle du nombre de revues et d'articles publiés, les bibliothèques universitaires, dans le sillage de la croissance générale du financement de la recherche universitaire (voir le chapitre 6), ont pu continuer à assurer aux chercheurs une disponibilité suffisante des revues et monographies disciplinaires et, partant, la large diffusion recherchée par les auteurs. Mais à partir de cette époque on a assisté à une augmentation soutenue du coût moyen des abonnements trois fois supérieure au taux d'inflation⁵ (Dingley, 2005; Van Orsdel et Born, 2008), augmentation que l'accroissement des budgets des bibliothèques n'est

4. Il s'agit du nombre moyen de citations reçues une année donnée par les articles qui y sont parus au cours des deux années précédentes. La pertinence de ce nombre comme mesure de qualité d'une revue, et encore plus d'un chercheur qui y publie ses articles, est fortement contestée (Archambault et Larivière, 2009). Il n'empêche que ce facteur joue, du moins dans certains domaines, un rôle important dans les décisions des chercheurs, que ce soit pour le choix d'une revue ou dans l'évaluation du dossier de publications d'un collègue.

5. Cette situation touche toutes les disciplines, avec des augmentations annuelles moyennes comprises entre deux et quatre fois le taux d'inflation (Dingley, 2005).

pas arrivé à combler, d'autant plus que le nombre de revues, lui, continuait de croître. Les conséquences de cette spirale n'ont pas tardé à se manifester dans les bibliothèques sous forme d'annulations d'abonnements et de coupures dans les achats de monographies, ces dernières entraînant à leur tour une « crise des monographies », peut-être moins criante mais bien réelle (McPherson, 2003). Cette situation inquiétante touche non seulement les établissements moins bien nantis, mais aussi certaines grandes universités traditionnellement bien pourvues à ce chapitre; elle affecte d'une manière particulièrement cruelle les établissements des pays en développement.

Bien que plusieurs facteurs aient été avancés pour expliquer cette situation, on n'a pas tardé à mettre en cause un phénomène nouveau, observé dans bien d'autres sphères : la privatisation de l'édition scientifique et sa concentration, au gré des achats et fusions, entre les mains de quelques grandes maisons d'édition (McCabe, 2002). Comme le montre le tableau 7.1, cinq maisons d'édition, dont les quatre premières ont toutes fait l'objet de fusions ou d'acquisitions entre 2000 et 2007, publient à elles seules environ 7 000 revues, soit près du tiers de toutes les revues scientifiques, dont le nombre est estimé à 24 000 (Björk, Roosr et Lauri, 2008). Dans les domaines regroupés sous l'appellation STM (sciences de la nature, techniques et sciences médicales), la part du marché de ces cinq géants (sur un total de plus de 2 000 éditeurs) est encore plus grande : on l'estimait à plus de 40 % en 2002.

Tableau 7.1

Nombre de revues publiées par les principales maisons d'édition scientifique commerciales et part du marché estimative dans les domaines scientifique, technique et médical (STM).

Éditeur	Tous domaines ^a		STM ^b	
	<i>n</i>	<i>n</i>	<i>n</i>	% du marché
Elsevier	2 350	1 347		18
Springer	1 700	878		11
Wiley-Blackwell	1 400	620		8
Taylor & Francis	1 000	351		5
Sage	500	123		2
Sous-total	7 000	3 319		43
Autres	~ 17 000	4 391 ^c		57
Total	~ 24 000^d	7 710		100

a. Données recueillies en décembre 2008 sur les sites web des éditeurs.

b. Mise à jour, avec prise en compte des fusions et acquisitions subséquentes survenues depuis, de l'estimation présentée dans OFT (2002), sur la base des revues indexées par l'Institute for Scientific Information (ISI).

c. Publiées par 2 028 éditeurs.

d. Données du répertoire Ulrich pour 2007; voir Björk, Roosr et Lauri (2008).

Mentionnons également que beaucoup de revues sont publiées par des organismes à but non lucratif, associations disciplinaires ou presses universitaires, qui offrent cette activité comme un service parmi d'autres à leurs membres. Cette activité est souvent en même temps une source importante de revenus, servant à financer d'autres opérations moins rentables. Ces organisations qui, à deux exceptions près (les presses des universités d'Oxford et de Cambridge), publient moins d'une

centaine de titres, ont elles aussi imposé au cours de cette période des augmentations de tarifs significatives, quoique plus modérées.

Pour justifier leur existence et, pour les plus importantes, leur prix (ou leurs profits), les maisons d'édition invoquent l'importance des quatre fonctions qu'elles remplissent : diffusion bien sûr, mais aussi authentification, évaluation et conservation (Goodman, 2004). Et, il faut l'admettre, bien que l'évaluation par les pairs soit effectuée de manière complètement bénévole, la gestion de tout le processus allant de la réception d'un manuscrit à sa publication entraîne un coût non négligeable⁶. Cependant, le fait que les prix ont augmenté trois fois plus vite que le coût de la vie paraît difficile à justifier, d'autant plus que cela a pour conséquence directe une diminution de l'accès aux descriptions des recherches, donc des retombées de celles-ci.

Face à cette situation jugée de plus en plus intolérable, une solution fondée sur la généralisation de pratiques déjà en place – mais marginales – a été proposée au cours des années 1990 et promue depuis : la diffusion en accès libre⁷ (voir l'annexe 6). Cette proposition, qui a entraîné au début des années 2000 la naissance d'un véritable mouvement, a rapidement commencé à révolutionner le monde de la publication scientifique. S'appuyant sur l'immense potentiel du web, elle favorise deux moyens complémentaires pour atteindre son objectif de rendre l'entièreté des publications scientifiques accessibles gratuitement :

- le dépôt par les chercheurs eux-mêmes (on parle alors d'autoarchivage), ou en leur nom, d'une copie des documents publiés ou en voie de l'être, dans une archive (ou répertoire) numérique garantissant l'accès gratuit;
- la publication dans des revues en ligne offrant l'accès gratuit aux articles qu'elles publient, que ceux-ci soient ou non également diffusés sur support imprimé.

Dans les deux cas, et contrairement à un préjugé largement répandu, l'évaluation par les pairs n'est aucunement remise en question; mieux encore, elle est parfois enrichie par l'ajout d'une discussion ouverte précédant et (ou) suivant l'acceptation des articles. Examinons donc en quoi consistent ces deux voies privilégiées vers l'accès libre qui, combinées, rendaient accessibles gratuitement près de 20 % des articles scientifiques publiés en 2006, selon l'estimation de Björk *et al.* (2008).

Les archives et répertoires de publications scientifiques : des serveurs de prépublications à l'Open Archives Initiative

L'histoire des archives⁸ de publications scientifiques remonte aux années 1960, alors qu'on assiste, dans le domaine de la physique des hautes énergies, à la mise en place d'un système parallèle de

-
6. Dans une étude bien documentée sur l'édition dans le domaine des sciences, technologies et médecine (STM), le Wellcome Trust (2004) estime les coûts de l'évaluation à 175 \$US par manuscrit et les coûts de publication à 600 \$US par article paru, le coût total par article publié se situant entre 1 000 et 2 000 \$US selon le taux d'acceptation des manuscrits.
 7. Aux fins de ce chapitre, nous adopterons la définition d'accès libre promue par Harnad, et qui est la plus couramment employée, soit l'accès (en ligne) gratuit, immédiat et permanent au texte intégral d'une œuvre, délaissant, sauf mention contraire, la dimension « autorisations consenties aux usagers ».
 8. Aux fins de cet ouvrage (mais cette nomenclature n'est pas universelle), on distingue les *archives*, qui hébergent des documents, et les *répertoires*, qui regroupent uniquement des références, plus ou moins complètes (on parle aussi de métadonnées, quand ces références sont analogues à des informations bibliographiques), à des documents hébergés ailleurs.

distribution de prépublications (*preprints*), c'est-à-dire des manuscrits (en version imprimée, à l'époque) soumis à des revues mais non encore acceptés (Till, 2001). Ce système a été enrichi dans les années 1970 par la création d'une base de données bibliographiques accessible en ligne puis celle, au début des années 1990, d'un site d'archivage des versions numériques des documents (ou *eprints*)⁹ hébergé par le Los Alamos National Laboratory, un des plus importants laboratoires de recherche militaire américains (Luce, 2001). L'archive, appelée arXiv, a été depuis transférée sur un serveur de l'Université Cornell. Précisons que dans arXiv, les auteurs ne font pas que déposer des manuscrits soumis à des revues, mais de plus en plus ajoutent (ou y substituent) la version acceptée et révisée après évaluation, appelée postpublication (*postprint*), en indiquant la référence à la revue ayant publié l'article.

Cette archive, qui a dépassé la barre du demi-million de documents en octobre 2008, est encore majoritairement consacrée à la physique théorique et à l'astrophysique – elle est même devenue le moyen privilégié de communication scientifique au sein de groupes de chercheurs de ces domaines. Elle s'est étendue avec le temps à la plupart des domaines de la physique ainsi qu'aux mathématiques, à l'informatique¹⁰ et à la biologie quantitative, sans toutefois y occuper, et de loin, la même place que dans la communauté qui en est à l'origine¹¹.

Deux autres archives, créées au milieu des années 1990, jouent un rôle similaire en sciences humaines et sociales (les nombres indiqués datent de la fin de 2008) :

- SSRN (Social Science Research Network), une archive centralisée comme arXiv, qui héberge près de 200 000 articles en texte intégral (pré- ou postpublications) relevant d'une quinzaine de disciplines;
- RePEc (Research Papers in Economics), un répertoire qui n'héberge pas de documents mais donne accès, via un réseau d'archives institutionnelles, à près de 400 000 articles et 300 000 prépublications (appelées *working papers*) en économie.

Dans le domaine des sciences de la vie, une expérience semblable à celle qui a donné naissance à arXiv a été menée dans les années 1960, pour être rapidement abandonnée à cause de l'opposition des revues traditionnelles et des réticences de nombreux chercheurs du domaine (Till, 2001). Deux initiatives similaires, la première, en chimie, mathématiques et informatique, Preprint Archive, lancée en 2000 par la maison d'édition Elsevier, la seconde en sciences de la vie, cette fois par le *British Medical Journal*, ont été interrompues après quelques années en raison du nombre insuffisant

9. La terminologie en ce domaine n'est pas encore normalisée. Kling (2004), dans une intéressante discussion sur les différentes catégories de documents correspondant aux étapes du processus menant de la recherche à la publication, propose de réserver le terme « prépublication » aux articles acceptés (donc qui *seront* publiés), et suggère d'appeler *e-scripts* ces « manuscrits » électroniques.

10. Il convient de souligner que les chercheurs en informatique pratiquent depuis les années 1980 une forme d'autoarchivage « anarchique », déposant leurs articles dans des sites web personnels ou institutionnels non structurés. Ces articles peuvent par la suite être repérés par des services spécialisés, comme le répertoire CiteSeer^x (voir plus loin), ou généraux, comme les outils de recherche Google et Google Scholar.

11. En guise d'exemple, parmi les 441 documents déposés le 23 octobre 2008 dans arXiv, on comptait 306 articles en physique (69 %), dont les deux-tiers relevaient de trois domaines (astrophysique, matière condensée et physique des hautes énergies), contre 107 en mathématiques et statistique (24 %) et 28 seulement (6 %) dans les autres disciplines (informatique, biologie quantitative). En guise de comparaison, Till (2001) arrivait, pour 1991-2000, à un pourcentage de 90 % en physique (dont les 4/5 dans les trois mêmes disciplines principales) et 9 % en mathématiques.

de dépôts (moins de 100 en quatre ans pour la seconde archive). Une nouvelle tentative (juin 2007) dans le domaine des sciences de la vie, Nature Precedings, de la prestigieuse revue britannique *Nature*, est à suivre. Toutefois, le taux de dépôt de documents, bien que plus élevé que ce qu'ont connu les deux expériences précédentes, n'est peut-être pas de très bon augure : Nature Precedings a reçu durant ses deux premières années environ 1 000 prépublications, soit à peu près le nombre de dépôts effectués en six jours dans arXiv.

S'il n'existe pas d'archives équivalentes à arXiv et RePEc dans d'autres disciplines, notamment en chimie et en sciences de la vie, on retrouve néanmoins quelques ressources remplissant en partie ce rôle, dont notamment deux répertoires :

- E-prints Networks de l'Office of Scientific Information du US Department of Energy, qui permet de chercher parmi 5 millions de documents dans des milliers de sites, tant d'organisations que de chercheurs individuels, dans les domaines d'intérêt du ministère (physique, chimie, biologie, sciences de la vie, sciences des matériaux, etc.);
- CiteSeer^x, qui joue le même rôle en informatique, en donnant accès à plus d'un million de documents.

À partir des années 2000, une communauté militante a mis de l'avant la création d'archives de portée plus restreinte, facilitée par la disponibilité d'outils informatiques conviviaux et gratuits et la mise un point d'un protocole assurant leur interopérabilité¹². Ces archives reposent généralement sur le principe de l'autoarchivage, c'est-à-dire que les auteurs y déposent eux-mêmes leurs documents. La plupart d'entre elles, dites institutionnelles, sont rattachées à un établissement, généralement une université ou un centre de recherche. Le Canada compte ainsi deux douzaines d'archives liées à des établissements universitaires, dont Archipel, à l'Université du Québec à Montréal, Papyrus, à l'Université de Montréal, et T-Space, à l'Université de Toronto.

D'autres archives sont dites centralisées. Certaines sont nationales, telles HAL, mise en place par le CNRS français (voir Baruch, 2007), et NARCIS aux Pays-Bas. D'autres sont liées à une organisation, comme l'archive PubMed Central, mise sur pied par les National Institutes of Health (NIH) américains, en conformité avec les politiques que cet organisme a adoptées au milieu des années 2000, et qui accepte notamment les postpublications présentant les recherches qu'il a financées. Le Canada compte une poignée d'archives de ce type, dont la plus importante (et de loin) est la Bibliothèque numérique du Centre de recherche pour le développement international (CRDI) qui, à la mi-2009, comptait environ 8 000 documents en texte intégral.

Le fait que des manuscrits peuvent être rendus publics avant d'avoir subi une évaluation par les pairs est régulièrement mentionné dans les arguments des opposants à ce mode de diffusion (McMullan, 2008). De fait, il existe tout au plus des systèmes de validation qui veillent à s'assurer de la pertinence ou du caractère « approprié » des documents déposés; dans arXiv, par exemple, ce système de validation est doublé, depuis 2004, d'un système d'endossement des nouveaux déposants. Les gestionnaires des archives n'effectuent toutefois aucune évaluation détaillée du contenu des documents.

12. La plupart de ces archives se conforment au protocole Open Archives Initiative (OAI), qui permet de traiter l'ensemble des archives, aux fins de repérage de documents ou d'exploration de données (*data mining*) notamment, comme s'il s'agissait d'une seule et même archive centralisée.

Pourtant, rien ne permet de croire que le progrès des connaissances dans les domaines où les prépublications jouent un rôle important, voire de premier plan, souffre du fait que les chercheurs prennent connaissance des travaux de leurs collègues avant que ceux-ci n'aient été évalués. Il faut aussi souligner que la plupart des documents déposés dans ces archives finissent par être publiés (donc évalués), quand ils ne l'ont pas déjà été au moment de leur dépôt. Mais on peut sans doute s'interroger sur la pertinence de ce modèle dans d'autres domaines (Pinfield, 2001), entre autres dans la recherche en santé, où l'on invoque les risques associés à l'application possible de résultats qui se révéleraient erronés.

Cependant, cette nouvelle pratique n'est pas sans mettre en cause la pertinence de l'évaluation par les pairs, ou du moins sa consécration comme seul et unique moyen de contrôle de la qualité, un moyen qui fait par ailleurs l'objet de nombreuses critiques et dont les effets réels demeurent difficiles à cerner en dépit de nombreuses études consacrées au sujet (Jefferson, Alderson, Wager et Davidoff, 2002). En fait, cette possibilité de soumettre un manuscrit aux commentaires d'un large public avant (ou en même temps) qu'à une évaluation « traditionnelle » par un comité de lecture anonyme peut être vue comme une simple extension de la pratique de soumettre privément les manuscrits à des collègues. Des systèmes combinant de manière intéressante ces deux types d'évaluation, par exemple en invitant les évaluateurs à prendre part à la discussion publique sur un manuscrit, ont d'ailleurs été instaurés avec succès par un certain nombre de revues (Harnad, 1998; Poschl, 2004).

Outre les prépublications et postpublications, les archives accueillent aussi d'autres types de documents décrivant des recherches, comme des thèses et mémoires, des rapports et notes de recherche, des *works in progress* (appelés « *working papers* » dans certains domaines), des textes de communications (non parus dans des actes), etc. Bien que ces documents soient parfois qualifiés de « littérature grise », une expression à connotation péjorative, on en a souligné à maintes reprises l'intérêt, ne serait-ce parce qu'on peut y trouver des informations plus détaillées sur les travaux et résultats de recherche que ce qui est décrit, souvent de manière succincte, dans les articles. Certaines archives hébergent également des entités qui se rapprochent davantage des données de recherche, comme des photos, des enregistrements audio ou encore des descriptions de molécules. Ainsi, l'archive institutionnelle de l'Université de Cambridge héberge la collection World Wide Molecular Matrix (WWMM), formée de 175 000 fichiers décrivant des molécules, comptant pour plus de 80 % du contenu de l'archive.

À la fin de 2009, on comptait selon les données du Registry of Open Access Repositories (ROAR) plus d'un millier de ces archives ou répertoires, regroupant près de 20 millions d'enregistrements, dont une partie significative¹³ sont associés à un document en texte intégral décrivant des résultats de recherche.

C'est beaucoup, mais peu en même temps si l'on songe que d'après l'estimation de Björk *et al.* (2008), environ 1,5 million d'articles paraissent à chaque année dans les revues scientifiques avec comité de lecture, et qu'une grande partie des archives – surtout les archives institutionnelles –

13. Le nombre exact de documents présentant des résultats de recherche accessibles en texte intégral est cependant très difficile à évaluer compte tenu de la diversité des entités que contiennent les répertoires et archives et de la redondance entre le contenu des répertoires et celui des archives qui, souvent, les alimentent. Une analyse sommaire du contenu des 20 plus grands répertoires ou archives inscrits au Registry of Open Access Repositories (ROAR), comptant pour environ 50 % des enregistrements, indique qu'environ la moitié d'entre eux correspondent à des documents de ce type.

contiennent très peu de documents. En fait, c'est là le problème principal de cette approche : malgré les avantages indéniables qu'ils peuvent en retirer, dont notamment des taux de citations jusqu'à trois fois supérieurs à ce que procure le mode traditionnel (Harnad et Brody, 2004), et le peu de temps que requiert le dépôt des documents¹⁴, les chercheurs sont peu nombreux à utiliser les archives institutionnelles mises à leur disposition. On estime (Björk *et al.*, 2008; Sale, 2005) qu'au plus 15 % d'entre eux vont autoarchiver leurs documents « spontanément » (euphémisme qui cache en réalité tout un travail de sensibilisation). L'expérience de l'auteur de ce chapitre, qui a participé à la mise en place et à la promotion d'une archive institutionnelle dans une université québécoise, montre que même après quelques années d'opération, ce taux modeste peut être encore loin d'être atteint.

Cette situation s'explique en partie par la simple ignorance (Swan et Brown, 2004) ou les préjugés, notamment quant à la difficulté ou la lourdeur appréhendées (mais non fondées) des opérations requises, ou encore (Pinfield, 2001) quant à la légalité de cette opération qui ne peut être effectuée qu'avec l'autorisation du titulaire du droit d'auteur. La crainte que cette diffusion élargie accroisse le risque de voir ses travaux plagés est également invoquée, surtout dans les domaines où, selon l'approche adoptée, l'argumentation peut jouer un rôle aussi important, sinon plus, que les résultats empiriques¹⁵.

Il faut aussi mentionner à cet égard que la politique éditoriale de certaines revues comprend la règle dite d'Ingelfinger¹⁶, selon laquelle un manuscrit ne sera pas considéré si son contenu a déjà fait l'objet d'une publication – ce qui peut inclure la diffusion en ligne au moyen d'un site web personnel ou d'une archive. Bien que ces revues soient globalement peu nombreuses, elles occupent dans certains domaines une place significative. On songe ici notamment à la chimie : les 36 revues de l'American Chemical Society (ACS) appliquent toutes diverses variantes, plus ou moins strictes, de cette règle. Certaines vont jusqu'à préciser qu'une prépublication déposée dans une archive institutionnelle, même « diffusée » sous le mode de l'accès restreint (voir plus loin), constitue une publication. Bien que dans la plupart de ces revues on parle d'une application au cas par cas plutôt que d'une règle automatique, on comprendra facilement la réticence des chimistes à l'égard de l'autoarchivage.

Pour augmenter le taux de dépôt, une sensibilisation des chercheurs à ces enjeux est certes nécessaire, mais en pratique seules deux stratégies se sont révélées efficaces.

La première consiste à fournir une assistance aux chercheurs dans cette tâche, ce qui, à la limite, signifie repérer les articles produits par les chercheurs et les déposer pour eux dans l'archive, avec leur autorisation. C'est la voie qu'a empruntée Archimer, l'archive institutionnelle de l'Institut français de recherche pour l'exploitation de la mer (Ifremer), qui a atteint un taux de dépôt de 80 %, il est vrai au prix d'efforts et de ressources considérables (Merceur, 2008).

La seconde repose sur l'adoption d'une exigence (*mandate*) relative à l'autoarchivage des postpublications¹⁷. Cette exigence peut provenir d'un établissement (centre de recherche ou

14. Selon l'étude de Carr et Harnad (2005), la durée moyenne de l'opération est de 10 minutes par dépôt, excluant le temps pouvant devoir être consacré à la recherche du fichier et à la vérification de son contenu.

15. Ce à quoi on peut rétorquer que s'il est raisonnable de penser que l'accès libre accroît le risque de plagiat, il est tout aussi raisonnable de soutenir qu'il en facilite la détection.

16. Du nom du directeur du *New England Journal of Medicine*, qui l'a promulguée en 1969.

17. Une liste des organismes et des établissements ayant adopté une exigence de dépôt est disponible dans le site Registry of Open Access Repository Material Archiving Policies (ROARMAP).

université) ou d'une composante de celui-ci, comme une faculté ou un département. C'est la voie qu'avaient suivis, à la mi-2009, une cinquantaine d'établissements, européens et américains surtout. Le premier, en 2003, a été la School of Electronics and Computer Science de l'Université de Southampton, au Royaume-Uni. A suivi, en 2004, la première exigence couvrant toute une université, l'Université do Minho, au Portugal. Des universités prestigieuses, Harvard et MIT au premier chef, ont emboîté le pas quelques années plus tard. Selon le cas, l'initiative provient de la direction de l'établissement, comme à l'Université de Liège, dont le recteur Bernard Rentier est un fervent défenseur de l'accès libre, ou du corps professoral lui-même, comme à Harvard et au MIT. L'expérience des premiers établissements ayant adopté cette exigence montre que des taux de 80 % et plus peuvent être atteints, à la condition que la politique rende obligatoire le dépôt, plutôt que de seulement le demander ou l'encourager (Sale, 2005).

L'exigence de dépôt peut aussi provenir des organismes de financement de la recherche; une partie de l'initiative en la matière est venue du côté des sciences biomédicales, sous l'effet de pressions du public (Guédon, 2008). Le premier (en 2005) fut le Wellcome Trust britannique, suivi en 2006-2007 par la majorité des sept conseils subventionnaires du Royaume-Uni. De leur côté, après une période où le dépôt a été simplement recommandé, les NIH américains, le plus grand organisme de financement au monde, et leur pendant canadien, les Instituts de recherche en santé du Canada (IRSC)¹⁸, ont adopté en 2008 des exigences de dépôt dont l'effet s'est rapidement fait sentir (« PubMed Central submissions... », 2008)¹⁹. À la mi-2009, une trentaine d'organismes de financement privés ou publics exigeaient ainsi que les chercheurs déposent dans une archive une copie des publications découlant des travaux qu'ils financent. L'adoption de telles exigences a fait également l'objet d'un projet-pilote de la Commission européenne, mené au sein de son 7^e programme-cadre de recherche (2007-2013), de même que d'une recommandation de l'Association des universités européennes (EUA, 2008).

Il convient de mentionner que les maisons d'édition, du moins aux États-Unis et au Royaume-Uni, ont mené d'importantes opérations de lobbying pour empêcher, sans succès, l'adoption des telles exigences, ou encore en diminuer la portée (Albanese, 2009; Giles, 2005).

Au Canada, l'exemple des organismes de financement du domaine biomédical ne s'est pas étendu aux autres domaines. Le Conseil de recherches en sciences humaines a bien adopté en 2004 une « politique de libre accès », mais celle-ci ne consiste en fait qu'en un énoncé de principe très général dont on ne saisit pas facilement la portée, en l'absence de précision sur ce qu'inclut la notion de « résultats de recherches » (CRSH, 2008) :

La politique de « libre accès » du CRSH vise à accroître l'impact des connaissances issues de la recherche sur la société en rendant l'ensemble des résultats des recherches financées par le gouvernement facilement accessibles au public. Ces résultats sont disponibles gratuitement et en permanence dans Internet.

Soulignons que le CRSH a tout de même adopté en 2007 une mesure concrète en matière d'accès libre, en appuyant la diffusion sous ce mode lors de la révision de son programme d'aide aux revues. Mentionnons également la mise en place en 2008 de l'archive de publications électroniques du Conseil national de recherches du Canada, NPArC (CNRC, n.d.).

18. Suivis peu de temps après par le Fonds de recherche en santé du Québec (FRSQ).

19. Notons que les NIH imposent le dépôt dans l'archive centralisée PubMed Central, alors que les IRSC permettent aux chercheurs de choisir entre cette archive et l'archive institutionnelle de leur établissement.

Bien que plusieurs signes concourent à cette conclusion, il est encore trop tôt pour savoir si cette voie vers l'accès libre, qualifiée de « route verte » par Harnad *et al.* (2004), est celle qui conduira le plus rapidement à l'objectif visé, comme le soutient celui-ci. Car il existe une autre voie, complémentaire : c'est la « route d'or » des revues en accès libre.

L'accès libre et les revues

Avec la création du web au tournant des années 1990 et son explosion à partir du milieu de cette décennie, les perspectives en matière de diffusion des publications de recherche se sont profondément modifiées. D'une part, cette évolution a contribué pour beaucoup à l'extension à d'autres communautés des systèmes de prépublication (et postpublication) en ligne qui, comme on l'a vu plus haut, avaient été jusque-là l'apanage de groupes de chercheurs versés en informatique qui utilisaient déjà massivement les réseaux pour leurs activités de recherche. Ce n'est d'ailleurs pas une coïncidence si l'inventeur du web, Tim Berners-Lee, était chercheur dans l'un des principaux centres de recherche en physique des hautes énergies, le Centre européen de recherche nucléaire (CERN). D'autre part, l'apparition du web a facilité, cette fois dans tous les domaines, la mise sur pied de revues électroniques en accès libre.

Les premières revues en accès libre étaient en bonne partie des expérimentations maintenues à bout de bras par des pionniers, dont certains ont exploré de nouveaux moyens d'enrichir, voire de remplacer l'évaluation par les pairs en mettant à profit l'interactivité du nouveau média (Nadasdy, 1997). Parmi eux, on retrouvait des visionnaires (Harnad, 1991, 1995; Odlyzko, 1995) qui en ont saisi tout le potentiel pour la diffusion des informations scientifiques et, du même souffle, perçu la menace que cela pouvait représenter pour la publication scientifique traditionnelle.

Il faut dire que ce nouveau mode de diffusion heurtait en effet de plein fouet le système traditionnel des revues imprimées. En effet, fidèles à l'esprit communautaire ayant présidé à la création du web, ces revues ne soulevaient pas (du moins explicitement) la question du droit d'auteur. Elles garantissaient ainsi non seulement une très large diffusion des œuvres, tant du fait tant de leur gratuité que de l'accès en ligne, mais aussi la possibilité inégalée de leur réutilisation par leurs auteurs, qui conservaient toute liberté d'action relativement à leur exploitation ultérieure. Cependant, les autorisations relatives à l'utilisation ou la réutilisation des œuvres par le public ne faisaient pas l'objet de préoccupations, explicites du moins.

Malgré l'intérêt qu'il présentait, ce concept a fait son chemin plutôt lentement, en partie par simple inertie, en partie à cause du scepticisme entretenu à l'égard du nouveau média. Quelle était, se demandait-on, la valeur de ces nouvelles revues, dont certaines avaient abandonné ou du moins modifié les procédures d'évaluation des manuscrits? Comment assurer la survie de revues gratuites fonctionnant grâce au bénévolat de leurs artisans et soutenues par un financement incertain? Comment, dans le monde intangible de l'information électronique, assurer l'intégrité et la pérennité de documents qui n'existent qu'en version électronique? Bref, comment assurer le maintien des quatre fonctions exercées par les revues traditionnelles – diffusion, authentification, évaluation et conservation? De fait, plusieurs de ces revues ne résistèrent pas à l'épreuve du temps, et le rythme de création de nouvelles revues se mit à diminuer à la fin des années 1990, au moment où les grandes maisons d'édition commençaient à mettre en ligne – mais pas en accès libre, évidemment – une version électronique de leurs articles (Wells, 1999). Cependant, à la faveur de l'apparition, au tournant des années 2000, d'un véritable mouvement pour l'accès libre, on a assisté à une renaissance de l'intérêt de ces revues, dont le nombre s'est rapidement mis à croître, augmentant à

partir de 2005 de près d'un millier par année. Malgré tout, 15 ans après la généralisation du web et 20 ans après la création des premières revues électroniques, moins de 20 % des revues scientifiques avec évaluation par les pairs, soit 4 200 sur 24 000, étaient diffusées en accès libre²⁰. Et si, dans la plupart des domaines, on retrouve quelques revues dites « prestigieuses » diffusées en accès libre²¹, la grande majorité ont conservé ou adopté le mode d'accès payant.

Aucun modèle économique ne s'est imposé dans le monde des revues en accès libre (Goodman, 2004). À une extrémité, on retrouve le modèle dominant à l'origine, soit le travail bénévole soutenu par de maigres financements ponctuels (par exemple, l'aide des établissements) ou à renouveler régulièrement (subventions à l'édition). Ce modèle demeure largement répandu, du moins pour les « petites revues », qui ne prétendent pas rivaliser avec les géants. À l'autre extrémité, on retrouve des revues prestigieuses, avec des comités scientifiques formés de sommités, qui se sont rapidement hissées aux premiers rangs dans leur domaine (tel que mesuré par le facteur d'impact, par exemple). Ces revues ont en général adopté le modèle auteur-payeur, où l'on demande à l'auteur de défrayer les coûts d'édition.

Mentionnons également l'initiative originale SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics), lancée par le CERN, qui vise à réunir l'ensemble des acteurs du domaine de la physique des hautes énergies (celui-là même qui a été, on se rappellera, à l'origine d'arXiv) pour passer à un mode de diffusion à 100 % en accès gratuit, sans frais pour les auteurs. Difficilement applicable à d'autres domaines, faut-il préciser²², ce mode laisse entrevoir une économie globale pour la communauté concernée. En effet, les revues actuelles n'assumeront plus que la fonction d'évaluation par les pairs, à des coûts qui feront l'objet de négociations; ceux-ci devraient normalement être plus que recouverts par l'économie découlant de l'annulation des abonnements.

De leur côté, les revues traditionnelles n'ont pas tardé à prendre elles aussi le virage du numérique et du web. Depuis le début des années 2000, la grande majorité d'entre elles proposent en ligne une version électronique de leurs articles. Certaines ont épousé, bien que timidement pour la plupart, la cause de l'accès libre, par exemple en rendant disponibles gratuitement leurs articles après une période dite « d'embargo » allant de six mois à deux ans. La plupart ont également autorisé l'autoarchivage²³, que les éditeurs des domaines couverts par les archives et répertoires centralisés toléraient d'ailleurs depuis longtemps. Parallèlement, en réponse aux critiques relatives à la hausse de leurs tarifs et en réaction à la fois à la demande croissante des « consommateurs » et à la

20. Le 20 juin 2009, le site Directory of Open Access Journals (DOAJ) répertoriait 4 227 périodiques en accès libre. Cependant, le pourcentage cité doit être considéré comme un maximum, car divers indices suggèrent que le nombre total de revues scientifiques avec comité de lecture pourrait être largement supérieur au nombre de 24 000 employé pour le calcul.

21. À titre d'exemples, les revues de la Public Library of Science (PLoS), dont notamment *PLoS Biology*.

22. Ce modèle n'est possible qu'en vertu du caractère particulier de ce domaine, où la quasi-totalité des articles sont déjà disponibles dans arXiv et où plus de 85 % des articles se retrouvent dans six revues seulement, publiées par quatre maisons d'édition.

23. À la fin de 2008, on estimait à 95 % la proportion des 10 000 revues scientifiques répertoriées dans le site SHERPA/RoMEO autorisant l'autoarchivage des manuscrits, la plupart du temps la version après évaluation et révision. Ce pourcentage cache cependant de grandes variations selon le domaine et la langue de publication. Il se pourrait notamment que ce nombre soit beaucoup plus bas pour les revues en français, pour lesquelles peu d'informations sont disponibles dans le site.

concurrence grandissante des revues en accès libre, les grandes maisons d'édition commerciales ont proposé de nouvelles modalités :

- l'accès payant à la pièce aux articles (généralement quelques dizaines de dollars, mais parfois avec des mesures de protection qui limitent l'accès à une journée seulement);
- les abonnements en bloc, soit l'accès en ligne à l'ensemble des titres d'une maison d'édition pour un coût inférieur à celui de l'abonnement aux versions imprimées, mais supérieur à ce que paient normalement les bibliothèques, qui ne s'abonnent jamais à l'ensemble des titres offerts par un éditeur;
- l'offre, aux établissements des pays en développement, d'abonnements gratuits ou à coût minime, selon le PIB du pays, à des ensembles de revues dans certains domaines. La plus importante initiative de ce type est HINARI, qui comprend l'accès à plus de 5 000 revues en sciences biomédicales et en sciences sociales²⁴;
- la diffusion immédiate en accès libre des articles dont les auteurs ont déboursé les frais exigés pour cette option (quelques milliers de dollars en général).

Ici aussi, il est trop tôt pour savoir si c'est la « voie d'or » des revues en accès libre qui mènera vers l'accès libre universel. On assiste d'ailleurs à un débat entre les tenants de chacune des stratégies, qui estiment parfois que la promotion de l'autre nuit à la cause, et ceux qui soutiennent que les deux doivent être poursuivies de concert (Pinfield, 2007; Rowlands, Nicholas et Huntington, 2004). Mais quel que soit le rythme auquel se déroulera cette transition, qui paraît maintenant inévitable, le droit d'auteur est au cœur de ce passage d'un mode de diffusion fondé sur une transaction commerciale à celui d'un mode de partage fondé sur le don.

L'accès libre et le droit d'auteur

Comme on l'a vu plus haut, le droit d'auteur sur les œuvres scientifiques – du moins les articles – n'était traditionnellement considéré par les auteurs que comme une simple formalité, ceux-ci signant sans hésiter (ni trop réfléchir) les formulaires par lesquels ils cédaient totalement leur droit d'auteur à leur éditeur.

En fait, ces auteurs semblaient peu conscients, ou du moins peu incommodés de l'étendue des droits qu'ils cédaient ainsi. En effet, l'éditeur, à titre de nouveau titulaire du droit d'auteur, peut dès lors prendre seul toute décision concernant l'exploitation de l'œuvre, par exemple sa publication ultérieure, totale ou partielle, dans une compilation, ou encore pour une adaptation ou une traduction. De son côté, l'auteur doit maintenant obtenir l'autorisation de l'éditeur pour toute utilisation, même personnelle, autre que celles que permettent le contrat d'édition, le cas échéant, et les exceptions prévues dans la législation, par exemple celles du *fair use*, aux États-Unis, ou celles, plus restrictives encore, de l'utilisation équitable (*fair dealing*) au Canada (voir le chapitre 4). Ainsi, au Canada, à moins que le contrat d'édition ne l'autorise explicitement, un professeur-chercheur ne peut

24. Il convient de souligner que l'Inde et la Chine, qui répondent pourtant aux critères fixés pour l'accès gratuit et à coût minime respectivement, en ont été exclues. Cela laisse supposer que le programme ne vise en fait que des pays où le marché pour ces revues est relativement peu important.

vraisemblablement pas reproduire en toute légalité un de ses propres articles pour le distribuer aux étudiants qui assistent à son cours ou à son séminaire²⁵.

Il faut dire qu'avant l'avènement du web, une telle cession n'entraînait guère de conséquences pratiques pour les auteurs : ceux-ci se sentaient tout à fait libres de faire circuler en cercle restreint les projets d'articles ou les manuscrits soumis. Par la suite, lors de la publication de l'article, ils pouvaient obtenir de l'éditeur des tirés à part pour les distribuer aux collègues qui en faisaient la demande, quand ceux-ci voulaient éviter de recourir aux bibliothèques. Le système des demandes de tirés à part constituait en même temps une façon d'établir des contacts à travers le monde avec des collègues s'intéressant aux mêmes sujets. À toutes fins utiles, cette diffusion n'allait pas au-delà de ce que permettent les exceptions de l'utilisation équitable ou du *fair use* et, de toute manière, les maisons d'édition ne trouvaient pas grand intérêt à contrôler ce genre d'utilisation, qui n'affectait pas le potentiel de leurs ventes; les auteurs devaient d'ailleurs généralement payer pour obtenir ces tirés à part.

Mais tout est différent quand les mêmes manuscrits peuvent être facilement rendus accessibles gratuitement à la planète entière dès leur acceptation, voire même, si l'auteur le désire, dès leur soumission à des revues – donc des mois, voire des années avant leur parution. Dans ce contexte, le droit d'auteur n'est plus un objet de négociation individuelle, mais devient un enjeu que se disputent et veulent utiliser à leur avantage les groupes en présence (Miller et Harris, 2004), soit les chercheurs-auteurs, les maisons d'édition et les établissements d'enseignement et de recherche, dont les porte-parole en la matière sont les responsables des bibliothèques.

Du point de vue des chercheurs-auteurs, la situation se présente sous trois perspectives différentes, bien que complémentaires.

Premièrement, comme producteurs de connaissances, ils désirent naturellement que leurs travaux connaissent la plus large diffusion. La problématique des coûts croissants des revues et des annulations, ou du moins de la stagnation des abonnements qui en a résulté, rend l'accès libre de plus en plus intéressant à cet égard. À cet avantage s'ajoutent la diffusion rapide des manuscrits que permet l'archivage des prépublications et la visibilité accrue des articles diffusés en accès libre, dont témoigne l'accroissement du taux de citations mentionné plus haut.

Deuxièmement, comme membres d'une communauté de chercheurs et employés d'un établissement, ils recherchent le prestige et la notoriété que confère la publication d'articles dans des revues de haut niveau, tant à des fins de reconnaissance que pour la progression dans leur carrière. Or, ces revues sont souvent celles qui imposent le plus de restrictions, qui touchent tant l'accès aux articles qu'elles publient (coût élevé des abonnements imprimés ou électroniques, paiement à la pièce pour l'accès en ligne, période d'embargo) que l'utilisation que peut en faire l'auteur lui-même. En effet, malgré l'ouverture manifestée par la majorité de ces revues en ce qui a trait au dépôt des manuscrits dans des archives, celles-ci imposent souvent une période d'embargo, ainsi que des restrictions sévères à toute publication ultérieure, dans un ouvrage collectif ou dans une autre langue, par exemple.

Troisièmement, les chercheurs sont aussi, du moins dans une bonne partie des disciplines scientifiques, les premiers utilisateurs de ces œuvres. Les chercheurs, qui sont très souvent aussi des enseignants, ont donc tout intérêt à pouvoir profiter de la plus grande marge de manœuvre possible

25. Fait cocasse, le même professeur pourrait toutefois en fournir une reproduction manuscrite (au tableau!) ou produire des images des pages de son article et les projeter sur un écran au moyen d'un rétroprojecteur, mais il ne pourrait projeter le contenu d'un site web qui contient l'article (voir l'annexe 1).

dans l'utilisation des œuvres produites par leur communauté scientifique. Cela comprend l'utilisation de leurs propres œuvres qui, lorsqu'ils en ont cédé les droits, requiert l'accord de l'éditeur.

Du point de vue des maisons d'édition, les choses se présentent simplement : elles s'estiment, de manière peu convaincante faut-il préciser²⁶, être les seules à pouvoir bien assurer trois des quatre fonctions qu'elles exercent en matière de publication, soit l'authentification, l'évaluation et la conservation. Elles soutiennent de plus que le contrôle serré de l'exploitation de ces œuvres est une condition nécessaire à leur propre survie.

Finalement, pour les universités, l'enjeu est très simple : offrir à leurs membres le maximum d'informations au moindre coût. Cela peut signifier entre autres refuser de payer le prix fort pour des abonnements à des revues, alors que les universités déboursent consacrent déjà des ressources importantes, notamment les salaires des professeurs-chercheurs, tant pour la réalisation des recherches présentées dans les revues que pour la participation « bénévole » de ces professeurs aux processus d'évaluation et d'édition.

Les deux voies vers l'accès libre permettent aux chercheurs comme aux établissements d'atteindre leurs objectifs; de leur côté, les revues commencent à réaliser qu'elles devront mieux s'adapter à la nouvelle réalité. Ces deux voies présentent cependant des enjeux très différents en matière de droit d'auteur, tant en ce qui touche la diffusion initiale des œuvres que leur exploitation ultérieure.

Diffusion initiale des œuvres

Pour ce qui est du dépôt dans les archives institutionnelles (ou autres), il n'y a aucun enjeu particulier pour les revues qui l'autorisent expressément, ce qui est le cas de la majorité d'entre elles, avec les nuances apportées plus haut. Parmi celles-ci, le tiers n'autorisent que l'archivage de la version soumise (prépublication). Dans ce cas, comme pour la minorité de revues (5 %) qui interdisent tout archivage, la version révisée peut être rendue accessible selon la procédure suivante :

1. Archiver le manuscrit entre le moment où il est soumis pour évaluation et son acceptation. À ce stade, il n'a pas fait l'objet d'une cession de droit d'auteur ou de concession d'une licence.
2. Demander, au moment de remplir le formulaire de cession de droit (ou de concession de licence) proposé par l'éditeur, l'autorisation d'archiver le manuscrit révisé.
3. Si l'éditeur refuse, deux stratégies sont possibles :
 - archiver tout de même le manuscrit révisé en choisissant l'option, dite « accès restreint », qui remplace l'accès direct au document par l'affichage des seules métadonnées (ou données bibliographiques, que le déposant doit fournir quand il dépose son document), accompagnées d'une invitation²⁷ à demander à l'auteur une copie aux fins d'étude privée ou de recherche, qui répond aux critères de l'utilisation équitable;
 - si les modifications apportées après l'évaluation par les pairs sont peu importantes, laisser la prépublication dans l'archive et y ajouter un document auxiliaire, ou *corrigendum*, qui décrit ces modifications.

26. Voir à ce sujet une position typique d'un éditeur (McMullan, 2008) et les arguments d'un des principaux promoteurs de l'accès libre (Suber, 2002, 2008).

27. En pratique, cela se traduit par la présence d'un simple bouton qui automatise l'envoi par courriel d'une demande à l'auteur, qui lui aussi l'autorise (ou la décline) d'un simple clic.

Il est clair qu'à l'étape de la prépublication, l'auteur²⁸ détient tous les droits nécessaires pour diffuser son œuvre comme il l'entend. Cependant, le fait de maintenir en place la prépublication et d'ajouter un *corrigendum* fait intervenir des notions comme l'existence de droits d'auteur distincts sur des œuvres semblables ou le caractère rétroactif d'une violation du droit d'auteur²⁹. Mais ici, comme dans d'autres situations, les démarches (juridiques ou autres) visant à empêcher ou sanctionner de possibles violations du droit d'auteur présentent peu d'intérêt pour les maisons d'édition. En témoignent la tolérance dont ont fait preuve les maisons d'édition des domaines couverts par arXiv et, plus récemment, la généralisation rapide de l'assouplissement des politiques des revues à l'égard de l'autoarchivage.

La question de la responsabilité du gestionnaire de l'archive (généralement l'établissement, par le truchement de sa bibliothèque) se pose également. Le dépôt d'une copie dans l'archive constitue-t-il une publication de l'œuvre, ou encore une communication de celle-ci au public par l'établissement? Ou bien celui-ci, à l'instar des fournisseurs de services Internet, ne joue-t-il qu'un rôle d'intermédiaire? Le droit ne fournit pas encore de réponses claires à ces questions, bien que, comme l'affirme Longdin (2007), les tribunaux aient toujours été très hésitants à tenir les fournisseurs de moyens techniques responsables des usages illégaux, effectués par des tiers, que ces moyens permettent. Mais il n'est pas inutile que les gestionnaires d'archives institutionnelles rappellent aux personnes qui y déposent des documents leurs responsabilités en matière de droit d'auteur. Idéalement, ils devraient obtenir des déposants les déclarations nécessaires pour se dégager de toute responsabilité en cas d'infraction. Ces déclarations pourraient porter notamment sur l'identité des titulaires du droit d'auteur sur l'œuvre et sur la détention des autorisations requises en cas d'inclusion dans celle-ci d'extraits d'œuvres protégées.

Pour ce qui est de la publication dans les revues en accès libre, les choses sont beaucoup plus simples. En effet, comme celles-ci ne retirent aucun revenu de la diffusion des œuvres, elles n'ont pas besoin en principe d'obtenir de l'auteur, par cession ou licence, de droit exclusif sur l'œuvre. Il leur suffit de requérir une simple licence non exclusive, qu'on peut d'ailleurs considérer être accordée implicitement par l'auteur qui soumet un manuscrit pour que la revue le publie. L'auteur conserve alors tous les droits pour l'exploitation ultérieure de son œuvre. Tout comme pour les archives institutionnelles, il est souhaitable que les revues obtiennent de l'auteur le type de déclaration mentionné au paragraphe précédent.

Cependant, comme le montre une analyse d'un échantillon de 300 revues en ligne répertoriées dans le site DOAJ à la fin de 2008 (Couture, 2009), les politiques des revues en accès libre en matière de droit d'auteur sont très variées, allant du maintien de la titularité complète de l'auteur à l'exigence (étonnante pour des revues en accès libre) d'une cession complète du droit à la revue³⁰, sans qu'aucune permission ne lui soit accordée en retour.

-
28. Si, bien sûr, il est titulaire du droit d'auteur, ce qui est généralement le cas pour les professeurs, mais pas nécessairement pour les chercheurs ou étudiants qui ont été embauchés pour la recherche (voir le chapitre 9).
29. Pour une discussion exhaustive (et souvent colorée) sur ce sujet complexe, voir les fils de discussion suivants dans le forum American Scientist Open Access, modéré par nul autre que Stevan Harnad : *Legal ways around copyright for one own's give-away texts* (2000), *Copyright: Form, content, and prepublication incarnations* (2001) et *PostGutenberg copyrights and wrongs for give-away research* (2001 et 2002).
30. L'une de ces revues justifie cette exigence en expliquant qu'elle est... à l'avantage de l'auteur du fait que la revue serait en meilleure position que celui-ci pour réagir à toute infraction en matière de droit d'auteur. Cependant, l'auteur de ces lignes n'a jamais entendu parler d'une revue, même à but commercial, ayant entrepris des démarches à cet égard.

Exploitation ultérieure de l'œuvre

On l'a vu, dans son acception la plus « pure », l'accès libre ne signifie pas seulement l'accès gratuit. Il concerne également les utilisations des œuvres par les usagers, au-delà bien sûr de ce que permettent, d'une part, la licence implicite d'utilisation (dont la portée risque de ne pas aller au-delà du droit de consultation) par un titulaire qui rend ainsi l'œuvre accessible en ligne et, d'autre part, les exceptions à la loi, dont l'utilisation équitale.

L'auteur qui dépose un document dans une archive peut, s'il n'a pas cédé son droit d'auteur ou concédé une licence exclusive à une revue, autoriser tous les usages qu'il juge souhaitables; il doit cependant les mentionner explicitement. Les métadonnées du document comprennent généralement des champs prévus à cette fin, ou encore des champs à usage général où cette information peut être communiquée. Dans le cas contraire, il ne peut autoriser aucun usage supplémentaire à ce que permet l'entente relative au droit d'auteur conclu avec la revue.

Pour sa part, l'auteur qui publie dans une revue en accès libre laisse celle-ci fixer les usages autorisés. L'étude évoquée plus haut a montré qu'en cette matière également, différents modèles cohabitent. Cela va de l'octroi aux usagers d'autorisations supplémentaires plus ou moins généreuses à l'absence de toute autorisation, soit par une déclaration explicite, soit du fait de l'absence de mention relative au droit d'auteur. Fait à signaler, environ un cinquième des revues examinées ont adopté un moyen efficace pour circonscrire et communiquer aux utilisateurs les usages qu'elles autorisent : les licences Creative Commons (voir l'annexe 6).

On le voit, les moyens techniques associés à l'Internet et les pratiques que cette innovation a suscitées ont le potentiel de révolutionner la diffusion des documents décrivant les recherches. Mais, et plus encore pourrait-on dire, elles présentent celui de révolutionner l'accès aux données de recherche elles-mêmes, qui ne bénéficiaient auparavant que de possibilités très limitées de diffusion ou de partage à grande échelle.

Les données de recherche

Les données de recherche peuvent revêtir des formes très diverses (Shores, 2002).

On retrouve tout d'abord les données dites *factuelles* ou *brutes*, recueillies au cours des activités de recherche expérimentale ou empirique, qui peuvent être regroupées en trois catégories :

- les informations quantitatives (nombres) ou qualitatives (caractéristiques diverses, données socio-démographiques, etc.) notées manuellement ou stockées automatiquement, et les informations plus complexes comme des photographies, des enregistrements audio ou vidéo de phénomènes naturels, des schémas (empreintes ou impressions) ou des diagrammes ou graphiques générés « mécaniquement », aujourd'hui de plus en plus souvent numérisées et stockées sur support informatique, sous des formats appropriés (fichiers audio, vidéo, d'image ou fichiers spécifiques à certains domaines ou technologies);
- les observations écrites ou orales, ainsi que les schémas, liés aux manipulations et aux observations, que consignent les chercheurs au cours de l'expérimentation;
- les réponses, explications ou informations, écrites ou orales, que fournissent les sujets participant à une recherche, de même que leur transcription, le cas échéant.

Pour certains auteurs, les données de recherche comprennent deux autres catégories :

- les résultats des traitements plus ou moins complexes effectués sur ces données brutes, une fois l'expérimentation ou la collecte de données terminée, tels les catégorisations, l'affichage à l'aide de tableaux ou de graphiques, les résumés, les codifications, etc.;
- certains types de « ressources de recherche uniques » (« *unique research resources* »; voir Nguyen, 2007, citant un document des NIH américains) servant de base à des expérimentations, comme des produits synthétiques, des organismes, des lignées cellulaires, des virus, des segments d'ADN, des tissus humains, voire même des plantes et des animaux entiers.

Cependant, le type de « données » couvertes par l'avant-dernière catégorie relève plutôt selon nous des résultats de recherche, alors que les entités comprises dans la dernière catégorie, que l'on combine parfois avec les précédentes sous l'appellation « matériel de recherche » (« *research materials* »), ne constituent pas à proprement parler des informations. Nous réserverons donc l'appellation aux « données » factuelles ou brutes.

La « propriété » des données de recherche

Selon la catégorie à laquelle elles appartiennent, les données de recherche peuvent être associées à différentes formes de propriété couvertes par des régimes juridiques distincts.

Les données relevant de la première catégorie (informations quantitatives et qualitatives) n'entrent *a priori* dans aucune des classes d'objets concernées par la propriété intellectuelle. En effet, on peut les considérer simplement comme des faits, qui ne sont pas protégés par le droit d'auteur et ne satisfont pas non plus aux critères de brevetabilité. Cependant, deux situations font en sorte que des droits sur ces données peuvent être reconnus.

- Comme on l'a vu au chapitre 2, si le travail de choix ou d'arrangement des données témoigne de l'expression d'un talent ou d'un jugement suffisant, les documents ou dispositifs (par exemple des bases de données) contenant des ensembles de données pourront être considérés comme des compilations et, du coup, bénéficier de la protection du droit d'auteur. Mais les données elles-mêmes ne sont pas alors protégées, ce qui signifie qu'elles peuvent être copiées et diffusées sans l'autorisation du propriétaire de la compilation. Comme on le mentionnait au même chapitre, la situation est différente dans les pays de l'Union européenne, où le contenu des bases de données peut être protégé, ce qui revient à protéger les données elles-mêmes lorsqu'une base de données constitue le seul moyen d'y accéder.
- Des données recueillies au cours d'une recherche de type contractuel commanditée par une organisation ou une entreprise peuvent être considérées par celle-ci comme des renseignements confidentiels (voir le chapitre 4), dont la protection au Canada relève de la *common law* ou, au Québec, du Code civil. Le contrat contiendra alors des dispositions relatives à l'utilisation des données, dont la diffusion prématurée pourrait éventuellement nuire aux intérêts du bailleur de fonds. Cela ne va pas sans soulever des questions de nature éthique, notamment quand ces données sont recueillies par des étudiants dans le cadre de leurs études (voir le chapitre 9).

De leur côté, les données brutes relevant des deux autres catégories (observations, schémas, réponses des participants) ainsi que les résultats de leur traitement sont susceptibles d'être protégés par le droit d'auteur, dans la mesure où ils sont générés par des humains, témoignent de l'exercice d'un talent et d'un jugement « non négligeable » ou « non banal » (pour reprendre les termes de la Cour

suprême canadienne dans l'affaire *CCH Canadienne Ltée c. Barreau du Haut-Canada*³¹) et n'équivalent pas à un simple enregistrement ou traitement mécanique. Ainsi, tout texte, dessin ou dispositif graphique produit par un chercheur ou par un sujet participant à la recherche est susceptible de constituer une œuvre protégée par le droit d'auteur.

Soulignons en passant le cas des entrevues, fréquent dans plusieurs domaines de recherche. S'il existe dans toutes les juridictions un droit d'auteur sur les enregistrements et les transcriptions, détenu par le producteur de l'enregistrement, il n'existerait pas au Canada de droit d'auteur sur les propos mêmes du locuteur³². Quoi qu'il en soit, les règles d'éthique en recherche en place dans toutes les universités (voir plus loin) obligent les chercheurs à obtenir le consentement des participants au moyen d'une entente signée qui précise notamment les usages auxquels les enregistrements sont destinés.

Si la question de l'existence d'un droit de propriété sur les données présente déjà une certaine complexité, celle de l'attribution de cette propriété revêt une difficulté supplémentaire. En effet, de nombreux acteurs qui interviennent dans une recherche ou sont concernés par elle peuvent être intéressés à jouer un rôle dans la gestion des données qu'elle génère :

- les personnes intervenant directement dans les activités de recherche, qui peuvent détenir divers statuts dans l'établissement ou liens d'emploi avec celui-ci : professeurs ou chercheurs réguliers ou associés, stagiaires postdoctoraux, professionnels de recherche ou techniciens, étudiants des divers cycles, assistants de recherche, etc.;
- l'établissement, qui assume des responsabilités en matière de gestion financière, organisationnelle et éthique de la recherche;
- les entreprises qui commanditent des travaux de recherche et qui, comme on l'a vu, peuvent vouloir protéger des renseignements confidentiels;
- les sujets participant à la recherche, qui fournissent des informations personnelles et des réponses aux questions d'entrevues.³³

Cette difficulté se reflète dans la grande variété des énoncés que l'on retrouve sur le sujet tant dans la littérature spécialisée que dans les politiques des établissements touchant la propriété des données. Ainsi, certains auteurs (voir, par exemple, Fields et Price, 1993; Fishbein, 1991) affirment péremptoirement que sur le plan juridique, les données, quelle qu'en soit semble-t-il la nature, appartiennent aux établissements. Ils invoquent à cet égard soit que les données, parce que produites par des employés, constitueraient une forme de *work for hire*, soit que cette propriété découlerait directement du rôle de récipiendaires des fonds de recherche que jouent les établissements. D'autres (Shores, 2002; Wright, 1998), plus prudents, reconnaissent la complexité de la situation et se contentent de souligner qu'il est dans l'intérêt de tous les acteurs concernés que les établissements se dotent de politiques claires et explicites en la matière, et que tous soient associés aux discussions qui entourent leur élaboration et leur adoption. Cette remarque est certainement pertinente quand on

31. 2004 CSC 13.

32. À la différence des États-Unis et du Royaume-Uni, bien que la jurisprudence ne soit pas complètement cohérente à ce sujet (MacQueen, 2005; Rich, 2004).

33. À ce sujet, mentionnons la suggestion de Parry et Mautner (2004) de s'inspirer des pratiques des chercheurs du domaine de l'histoire orale (Ward, n.d.), qui demandent aux participants une cession de droit ou une licence autorisant les utilisations futures des enregistrements ou comptes rendus de leurs propos.

constate, par exemple, que les politiques des universités canadiennes en matière de données, quand du moins elles existent, sont très variées et souvent loin d'être « claires et explicites » (voir le chapitre 10).

Un des arguments qui justifient cette approche, davantage politique pourrait-on dire, est que la plupart des enjeux entourant l'utilisation locale des données de recherche, incluant celles qui découlent de travaux faisant l'objet de contrats avec des firmes privées, sont de nature éthique. En effet, les organismes publics qui financent la recherche au Canada (IRSC, CRSNG et CRSH, 2005) posent comme critère d'admissibilité que les établissements disposent de politiques et procédures visant à assurer le respect des principes d'éthique et de probité en recherche, à faire enquête sur les allégations d'inconduite scientifique et à résoudre les différends ou conflits qui accompagnent souvent ces situations. Or, l'utilisation des données de recherche est un enjeu dans une bonne partie des situations portées à l'attention des comités d'éthique, qui peuvent relever de trois catégories d'inconduite scientifique : le non-respect de l'équité et des droits individuels, le conflit d'intérêts et la fraude scientifique.

En matière d'équité et de droits individuels, songeons par exemple à la personne, quel que soit son statut, qui publie des résultats fondés sur des données recueillies par d'autres sans les en informer et sans en mentionner l'origine. Songeons encore à celle qui quitte l'établissement en emportant avec elle, parfois sans laisser de copie, des données essentielles au projet auquel elle collaborait. Le non-respect des règles de libre consentement et de respect de la confidentialité dans les recherches avec des sujets humains fait aussi partie de cette catégorie. En matière de conflits d'intérêts, mentionnons le fait que les exigences de confidentialité des bailleurs de fonds privés relativement aux données de recherche pourraient retarder indûment la publication d'articles, voire le dépôt de thèses ou mémoires. Pour ce qui est de la fraude scientifique, on retrouve tous les comportements de fabrication et de falsification de données, rares certes, mais dont les conséquences sont particulièrement néfastes, voire dangereuses pour le public dans certains domaines, comme la recherche biomédicale.

C'est cette responsabilité des établissements, tenus de faire enquête sur les allégations d'inconduite et de mettre en place des mesures visant à prévenir les comportements répréhensibles, qui justifie que leur soit reconnu un rôle de premier plan dans la gestion des données. Ces mesures iront des règles relatives à l'accès et au partage des données entre les membres des équipes de recherche à celles qui touchent leur conservation, le maintien de leur intégrité et leur éventuelle destruction.

Cependant, tout en affirmant ainsi leurs droits, les établissements, par tradition mais aussi par souci d'efficacité, confient aux responsables des projets de recherche l'intendance (*stewardship*) en matière de données, tout comme la responsabilité du respect de l'éthique d'ailleurs. Ce sont eux qui sont appelés à prendre les décisions quotidiennes en ce domaine. C'est d'ailleurs en bonne partie sur la base de cette responsabilité déléguée que les chercheurs se croient souvent, de manière spontanée, « propriétaires » des données et, de ce fait, s'estiment les seuls habilités à en autoriser la diffusion ou le partage.

Le partage et l'accès aux données de recherche

À ces enjeux *locaux* – c'est-à-dire au sein des groupes de recherche et des établissements – touchant l'accès et le partage des données de recherche, correspondent des enjeux globaux, couvrant le pays, voire le monde entier. Depuis le début des années 1980 (Fienberg, Martin et Straf, 1985), et plus particulièrement depuis l'avènement de l'Internet, des organismes et groupes de pression mènent des

réflexions sur la question et promeuvent la diffusion et le partage des données de recherche. En parallèle, un débat à ce sujet avait lieu dans la communauté de recherche en génomique, pour déboucher sur un consensus en faveur du partage des données générées dans les projets dotés d'un financement public (Marshall, 2003).

En 2003, un groupe de travail de l'Organisation de coopération et de développement économiques (OCDE) publiait les résultats d'une vaste étude sur les divers aspects de cette question (Arzberger *et al.*, 2004). Les recommandations du groupe étaient entérinées par les ministres de la science et de la technologie des pays de l'OCDE, dans une déclaration de janvier 2004 « chargeant l'OCDE d'œuvrer en faveur de principes et lignes directrices définis d'un commun accord sur l'accès aux données³⁴ de la recherche financée sur fonds publics » (OCDE, 2004). L'organisation y donna suite trois ans plus tard par la publication d'un document énonçant ces principes et lignes directrices et invitant les organismes subventionnaires publics à les intégrer à leurs politiques (OCDE, 2007; Pilat et Fukasaku, 2007).

Durant la même période, l'Association of Learned and Professional Society Publishers (ALPSP), avec ses 360 membres publiant plus de 10 000 revues, et l'International Association of Scientific, Technical and Medical Publishers, avec une centaine d'éditeurs publiant la majorité des revues dans le domaine, formulaient une déclaration conjointe (ALPSP et STM, 2006) concernant l'accès aux données de recherche :

We believe that, as a general principle, data sets, the raw data outputs of research, and sets or sub-sets of that data which are submitted with a paper to a journal, should wherever possible be made freely accessible to other scholars. We believe that the best practice for scholarly journal publishers is to separate supporting data from the article itself, and not to require any transfer of or ownership in such data or data sets as a condition of publication of the article in question. Further, we believe that when articles are published that have associated data files, it would be highly desirable, whenever feasible, to provide free access to that data, immediately or shortly after publication, whether the data is hosted on the publisher's own site or elsewhere (even when the article itself is published under a business model which does not make it immediately free to all).

Les propositions de l'OCDE ont été suivies par différents organismes nationaux, à des rythmes très divers, faut-il préciser. C'est dans le domaine des sciences biomédicales, où des projets comme le *Human Genome* (NIH, n.d.) ont pavé la voie, que ces recommandations ont été le plus rapidement mises en œuvre. Ainsi les National Institutes of Health exigent dorénavant que toute demande de financement excédant un demi-million de dollars par année comprenne un plan de partage des données de recherche que le projet générera. Une assez grande marge de manœuvre est laissée au chercheur quant au contenu de ce plan; le site des NIH propose différentes consignes et modalités à cet égard (NIH, 2007).

Au Canada, c'est aussi dans ce domaine que des actions ont été menées : la politique d'accès libre adoptée en 2007 par les IRSC (2008), dont les exigences ne sont pas liées à un financement minimal, s'applique aussi aux données de recherche. Du côté des autres grands organismes canadiens de financement de la recherche, les choses vont plus lentement. Le Conseil de recherches en sciences humaines (CRSH), qui avait pourtant adopté dès 1990 une politique sur l'archivage des données de

34. La définition de « données de recherche » retenue par l'OCDE est très semblable à celle qui est présentée au début de ce chapitre.

recherche en vue de leur partage (CRSH, 2007)³⁵, s'est contenté de vagues déclarations de principes. Acteur pourtant important du monde de la recherche, le Conseil de recherches en sciences naturelles et en génie (CRSNG) est tout simplement absent du débat public sur cette question. Une politique générale sur le libre accès, incluant l'accès aux données, annoncée officieusement pour 2009 (Leclerc, 2008) n'était pas encore en vigueur en février 2010; tout au plus le Conseil a-t-il inclus, dans ses règles sur les dépenses admissibles, le coût de la diffusion ou de la publication en accès libre.

Un fort courant se dessine donc en faveur de du partage, à l'échelle planétaire, des données de recherche. Un mouvement parallèle à celui de l'accès libre aux publications scientifiques, prônant le concept de « données libres » (*open data*), a d'ailleurs vu le jour (Hatcher, 2008; Klump *et al.*, 2006; Murray-Rust, 2008). Les propositions relatives au mode de diffusion des données sont également calquées sur celles qui ont été mises de l'avant pour l'accès libre aux publications : archives disciplinaires ou institutionnelles, diffusion en accès libre avec l'article décrivant la recherche, qui doit idéalement être lui aussi en accès libre.

Outre les questions juridiques liées à la protection des bases de données, cette nouvelle réalité pose un certain nombre de défis. Certains sont d'ordre technique et organisationnel, relevant de l'interopérabilité, de la documentation et du contrôle de la qualité des données (Diviacco, 2005; Fitzgerald *et al.*, 2007; Research Information Network, 2008). D'autres, notamment pour les données qualitatives recueillies auprès de sujets humains (Parry et Mautner, 2004), sont d'ordre culturel (Kansa, Schultz et Bissell, 2005) ou éthique (Abowd et Lane, 2004). Les travaux précités laissent toutefois entendre que ces difficultés peuvent être surmontées. Soulignons entre autres l'initiative Open Data Commons, qui propose un système de licences analogues aux licences Creative Commons.

On assiste donc depuis le milieu des années 1990, mais surtout le début des années 2000, dans le domaine de la diffusion des documents décrivant les résultats de recherche – articles scientifiques en tête – et des données de recherche, à l'émergence, voire la consolidation d'une culture du partage, favorisée par le développement des réseaux et des outils permettant de les exploiter. Mais qu'en est-il du côté de l'autre volet de la fonction des professeurs d'université, soit l'enseignement, où l'Internet et les technologies de l'information et des communications ont là aussi causé toute une révolution?

35. Politique qui semble tombée en désuétude, comme semble l'indiquer le fait que l'organisme qui doit y jouer un rôle majeur, l'Association canadienne des utilisateurs de données publiques (ACUDP; cadpu.ca), n'a de toute évidence pas pris le virage du web.