

Préface

Marie Masclet de Barbarin
Maître de conférences HDR à Aix-Marseille Université
Vice-présidente du Conseil d'Administration d'Aix-Marseille Université

Notre monde a changé. C'est un fait incontestable et incontesté. La révolution numérique a opéré un bouleversement profond de nos sociétés du fait du développement de l'informatique et d'Internet qui a peu à peu conduit à abolir les frontières physiques et technologiques et a radicalement transformé notre façon de communiquer, de travailler, de produire et de consommer.

Ce changement s'est effectué sans douleur, s'imposant à chacun d'entre nous avec la discrétion de l'évidence et la séduisante justification de la preuve par l'usage. Il s'est fait de façon progressive et s'est immiscé dans notre quotidien de façon si naturelle qu'il faut à présent s'obliger à regarder en arrière pour prendre l'entière mesure de toute son étendue.

Le monde de la recherche n'a pas échappé à cette mutation, qu'il a d'ailleurs très tôt transformée en opportunité. Opportunité d'échanger, de sauvegarder, de construire les savoirs et de partager des connaissances. Le projet Gutenberg de bibliothèque d'ouvrages numérisés librement accessible initié en 1971 par Michael Hart, bien avant la naissance du World Wide Web, a sans doute constitué une des premières initiatives destinées à faire partager au plus grand nombre l'accès à la connaissance.

L'arrivée d'Internet a par la suite contribué au développement de ce mouvement, avec notamment la création par Stevan Harnad en 1989 de *Psycoloquy*, la première revue intégralement disponible en ligne. La première archive ouverte disciplinaire ArXiv a ensuite vu le jour en 1991, sous l'impulsion de Paul Ginsparg. Alors dénommée the LANL preprint archive, elle n'a fait que développer une pratique qui avait déjà cours chez les physiciens : partager les dernières découvertes, à l'époque par courrier, sans attendre leur publication qui pouvait être soumise à des délais extrêmement longs. S'en sont suivis en 1997 l'archive ouverte Cogprints en psychologie, là encore à l'initiative de Steven Harnad, puis en 2000 PubMed Central dans le domaine des sciences de la vie¹.

L'*open access* était né. Traduit dans un premier temps en français par « accès libre », puis par « accès ouvert » pour tenir compte des restrictions pouvant être

1 Sur la naissance du mouvement de l'*open access*, voir notamment la chronique de Peter Suber, *Timeline of Open Access Movement*, <http://legacy.earlham.edu/~peters/fos/timeline.htm>.

apportées à la circulation de l'article², ce mouvement a été pour la première fois défini dans le cadre de la Déclaration de Budapest, signée le 14 février 2002, dont Peter Suber, ardent défenseur de l'*open access*, a été le principal rédacteur :

[...] par « accès ouvert » à cette littérature, nous entendons sa libre mise à disposition sur l'internet public, permettant à tout usager de lire, télécharger, copier, diffuser et imprimer ces articles, de lancer une recherche dans ces articles, de créer un lien vers leur texte intégral, de les compiler pour les indexer, de les convertir en données pour traitement logiciel, et de les utiliser à toute autre fin légale, sans barrières financières, juridiques ou techniques autres que celles de l'accès à l'internet lui-même. La seule contrainte pour reproduire et diffuser cette littérature et le seul rôle du droit d'auteur dans ce domaine devraient être de donner aux auteurs le moyen de contrôler l'intégrité de leur travail et le droit d'être mentionnés et cités de manière adéquate³.

Les initiatives nées de ce nouveau monde d'opportunités n'ont cependant pas toutes été animées des mêmes bonnes intentions nées de l'amour de la recherche et du partage de la connaissance, et c'est en privilégiant des intérêts privés au détriment de l'intérêt général qu'une partie du monde de l'édition scientifique a profité de ces avancées technologiques pour sortir du modèle alors en crise de l'édition papier et développer une offre de revues numériques de plus en plus riche et oligocentrée, et surtout de plus en plus onéreuse⁴. Cultivant le besoin jusqu'à la dépendance à force d'*impact factor* et autres mesures d'évaluation quantitative de la recherche auto-entretenu par des outils contrôlés par ces mêmes éditeurs, ils ont réussi à créer de nouvelles barrières, financières cette fois, limitant l'accès à la littérature scientifique aux seules institutions en mesure de payer le prix exorbitant et sans cesse en augmentation des abonnements à ces revues numériques. Les premiers témoins de la construction de ce modèle économique d'une redoutable efficacité furent les bibliothécaires, qui ont très tôt alerté la gouvernance des institutions nationales et des universités sur le paradoxe lié au fait que nous dépensions des sommes sans commune mesure avec le service rendu : pouvoir lire les articles rédigés par nos propres chercheurs retraçant les résultats de recherches que nous avons nous-mêmes financés. Ils ont ainsi joué un rôle moteur dans le développement des archives institutionnelles qui ont fait suite à la création des premières archives disciplinaires telles qu'ArXiv ou PubMed⁵.

Dans le même temps, sous l'effet induit de la crise économique, les pouvoirs publics ont pris conscience du rôle majeur que pouvaient jouer les découvertes

2 Voir sur ce point la préface de Marin Dacos à la traduction française de l'ouvrage de Peter Suber, *Qu'est-ce que l'accès ouvert* (2016 : 8 et s.). Voir aussi Mounier et Dacos (2010).

3 <http://openaccess.inist.fr/?Initiative-de-Budapest-pour-l>; voir également <http://www.budapestopenaccessinitiative.org/read> pour l'accès au texte original en anglais. Cette première déclaration a été suivie par la Déclaration de Bethesda sur la publication en accès ouvert, signée le 20 juin 2003 et par la Déclaration de Berlin sur l'accès ouvert aux connaissances dans les sciences et les humanités, signée le 22 octobre 2003.

4 Voir notamment sur ce point Mounier et Dacos (2010).

5 Sur la réflexion mise en œuvre à cet égard au sein d'Aix-Marseille Université, voir Bertin, Dacos, Delhaye, Hug, Masclét de Barbarin et al., (2014), *Vers une archive ouverte pour Aix-Marseille Université. Une démarche en faveur de l'Open Access : conclusions du groupe de travail sur le référencement des articles scientifiques produits par AMU. [Rapport technique]* <https://hal-amu.archives-ouvertes.fr/hal-01226882>.



scientifiques sur la relance de l'économie. Partant du postulat selon lequel « les résultats de la recherche publique doivent irriguer le monde socio-économique et favoriser son développement » (Chartron 2016 : §4), les politiques publiques vont à partir de là converger vers le libre accès aux publications scientifiques issues de projets financés sur fonds publics. Le rapport Finch publié au Royaume-Uni le 27 juin 2012, la recommandation de la Commission européenne du 17 juillet 2012 relative à l'accès aux informations scientifiques et à leur conservation ou encore le mémorandum de l'OST – le bureau de la politique scientifique et technologique de la Maison-Blanche – du 25 mars 2013 visant à améliorer l'accès aux recherches scientifiques recevant des fonds fédéraux, en constituent les principales illustrations.

Cette politique publique a peu à peu porté ses fruits, y compris parmi les chercheurs les plus séduits par l'ergonomie de l'offre éditoriale privée ou auprès des quelques irréductibles réfractaires aux charmes des nouvelles technologies et plus personne aujourd'hui ne songerait à remettre en cause l'intérêt de l'*open access*.

Les dernières évolutions de cette prise de conscience collective s'orientent désormais vers la recherche d'un juste équilibre entre les différentes pratiques éditoriales cohabitant plus ou moins pacifiquement dans le monde de l'édition scientifique. L'appel d'Amsterdam, publié en 2016⁶ ou plus récemment, en 2017, l'appel de Jussieu pour la science ouverte et la bibliodiversité militent ainsi en faveur du soutien à la diversité des acteurs de la publication scientifique pour mettre fin « à la domination par un petit nombre d'entre eux dictant de ce fait leurs conditions aux communautés scientifiques⁷ » : coûts d'abonnements de plus en plus élevés et transferts des charges des abonnements vers les APC (*article processing charges*), ces frais de publication demandés aux chercheurs pour que leurs articles soient disponibles en accès ouverts sur leurs plateformes. Il ne s'agit ni de nier le travail effectué par ces éditeurs privés, ni de leur faire abdiquer leur droit à réaliser un bénéfice sur la diffusion de ces articles, mais simplement de veiller à ce que ce profit soit juste et équitablement réparti et qu'il ne se constitue pas au détriment du financement de la recherche. S'il est difficile de prédire la façon dont va se régler le bras de fer entre les oligopoles de l'édition et les usagers au premier rang desquels figurent les universités, nous pouvons dire sans nous tromper que ce mouvement d'ouverture est en marche et que le débat sur le libre accès aux résultats de la recherche est à ce jour relativement cadré.

Celui qui semble l'être beaucoup moins, alors même qu'il se situe au cœur de ces questionnements, concerne plus directement les données de la recherche. Si l'on se réfère à la définition proposée en 2007 par l'OCDE « les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour

6 <http://openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf>.

7 L'appel de Jussieu a été élaboré par un collectif français représentatif des chercheurs et des professionnels de l'édition scientifique regroupés notamment au sein des segments « *open access* » et « édition scientifique publique » de la BSN (Bibliothèque scientifique numérique). <http://jussieucall.org/index-FR.html>.

valider les résultats de la recherche⁸ ». Or, les règles relatives à la production de ces données, à leur utilisation, à leur transmission et à leur conservation demeurent encore à ce jour en construction, dans un paysage dont nous percevons à peine les contours. Pour preuve les données numériques que nous produisons chaque jour en utilisant Internet et qui pour la plupart d'entre elles sont stockées et réutilisées à des fins commerciales, faisant de chacun d'entre nous les ouvriers bénévoles d'une industrie présente et à venir dont nous ignorons tout ou presque.

Les données de la recherche ne sont pas à cet égard aussi impensées, mais elles n'en constituent pas moins une incroyable richesse susceptible de profiter à d'autres chercheurs et de faire ainsi avancer la science. Parmi celles-ci, les données de la recherche en sciences humaines et sociales occupent une place particulière. La spécificité de leur nature et de leur objet impose en effet un encadrement normatif de nature à les protéger de toute utilisation malveillante ou inopportune susceptible de les détourner de leurs fonctions premières, mais celui-ci ne doit pas non plus avoir pour effet de limiter leur circulation au bénéfice du plus grand nombre. La loi française a commencé à apporter des réponses à ces impératifs d'apparence contradictoires, mais elles demeurent encore à ce jour imparfaites pour embrasser la variété des situations auxquelles se trouvent confrontés de façon quotidienne chercheurs, enseignants, étudiants, spécialistes de l'information scientifique et technique et décideurs académiques.

C'est justement pour répondre à ces questions que l'idée de ce guide de bonnes pratiques éthiques et juridiques pour la diffusion des données en SHS a vu le jour. Il est né de la réflexion d'un groupe de chercheurs, de professionnels de la documentation et de juristes structuré dans le cadre d'un groupe de travail Éthique et Droit, animé depuis sa création en 2011 par Véronique Ginouvès, ingénieure de recherche responsable de la phonothèque de la Maison méditerranéenne des sciences de l'homme (MMSH) d'Aix-en-Provence, avec l'aide depuis 2015 d'Isabelle Gras, conservatrice de bibliothèque au Service commun de la documentation (SCD) d'Aix-Marseille Université. Ce projet, labellisé par la Digital Research Infrastructure for the Arts and Humanities (DARIAH)⁹ en 2015 et financé par le Service commun de documentation (SCD) d'AMU, par la MMSH et par le consortium des ethnologues de la Très Grande Infrastructure de recherche (TGIR) Huma-Num a été supervisé par un comité scientifique qui s'est porté garant de la rigueur scientifique de la démarche¹⁰.

Face au renouvellement des problématiques liées aux droits d'auteurs, à la confidentialité, à la sécurisation et à la validation des données dans un contexte de mutation technologique, le groupe de travail Éthique et Droit a lancé un appel

8 Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics, Paris, 2007, 28 p., <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>.

9 <http://humanum.hypotheses.org/2675>.

10 Le comité scientifique regroupe : Florence Descamps, maître de conférences HDR à l'École pratique des hautes études ; Laurent Dousset, anthropologue, directeur d'étude à l'École des hautes études en sciences sociales ; Mélanie Dulong de Rosnay, chargée de recherche à l'Institut des sciences de la communication du CNRS, Paris Sorbonne, UPMC ; Lionel Maurel, juriste, conservateur des bibliothèques, responsable de l'information scientifique et technique à l'université Paris Lumières ; Philippe Mouron, maître de conférences en droit privé, HDR, à Aix-Marseille Université, spécialiste en droit de la propriété intellectuelle ; Stéphane Pouyllau, ingénieur de recherche au CNRS, directeur adjoint de la TGIR Huma-Num.



à communication ouvert à toutes les personnes s'intéressant aux problématiques éthiques et juridiques en matière de diffusion des données en SHS. Les auteurs ont été invités à proposer soit des retours d'expériences soit des articles portant sur les enjeux éthiques et juridiques en matière de diffusion des données en SHS, en explorant les solutions concrètes envisageables ou envisagées dans le cadre des cinq étapes de diffusion identifiées au préalable : la préparation de la recherche et l'anticipation de l'archivage ; la collecte des données ; le traitement, l'archivage et la description des données ; la diffusion des résultats de la recherche et la réutilisation des données.

Il est important ici de souligner le caractère ambitieux de la démarche, à la fois prospective et réflexive, empirique et comparative, et surtout résolument pragmatique.

La première partie de l'ouvrage, rédigée par des membres du groupe de travail et par des experts reconnus dans leurs domaines, présente ainsi un état des lieux sur les bonnes pratiques éthiques et juridiques en matière de diffusion des données en SHS. Anne-Laure Sternin, qui accompagne les travaux du groupe de travail depuis sa création, résume très bien dans son article introductif « Diffuser des données de la recherche dans le respect du droit et de l'éthique : comment faire lorsqu'on n'est pas juriste ? » la difficulté inhérente à l'encadrement de la diffusion des données en SHS. Les règles de droit susceptibles de s'appliquer sont effectivement extrêmement diverses : droit d'auteur, droit à l'image, droit à la protection des données personnelles, droit au respect de la vie privée, etc. Elles se sont par ailleurs construites pour la plupart sur des objets différents de ceux qui nous occupent aujourd'hui. De l'incubable xylographique aux téraoctets de données en jeu dans les problématiques liées au *big data*, il y a un monde. Et pourtant ce sont avec les mêmes règles de droit que nous devons appréhender les problématiques liées à l'utilisation et à la diffusion de ces objets. Sans compter que ces règles entrent parfois en contradiction l'une avec l'autre. C'est le cas, comme le souligne Émilie Debaets dans son article sur « *Big data* en sciences sociales et protection des données personnelles », lorsque la liberté de la recherche se trouve confrontée à la protection des données personnelles.

Face à ces incohérences et à ce foisonnement de normes juridiques, Anne-Laure Sternin nous donne les clés pour s'y retrouver. Le premier travail auquel devra se livrer le chercheur sera donc d'identifier la règle de droit applicable en fonction du type de matériau utilisé, une étape que connaissent bien les juristes : celle de la qualification juridique. Celle-ci va commencer par l'identification de la donnée en cause, ce qui n'est pas toujours aisé au regard de la diversification typologique de ces objets. Ainsi, en matière d'archives, Jean-François Bert précise dans sa contribution, « Pratiques d'archives, problèmes actuels sur les usages du matériau documentaire », que l'on parle désormais d'archives publiques, privées, intimes, authentiques, apocryphes et que chacun de ces termes fait référence à une législation particulière.

Une fois le type de donnée déterminé, il conviendra d'identifier la règle de droit applicable, du moins lorsqu'elle est suffisamment claire et adaptée à la situation, ce qui n'est pas toujours le cas. Mélanie Dulong de Rosnay, dans son article « Preserving Public Domain Collections: Institutional Policies Best Practices », fournit à cet égard des conseils extrêmement pratiques pour arbitrer les questions liées au droit d'auteur dans la création d'une archive en ligne, en présentant les

différentes licences ouvertes pouvant être utilisées pour que les données soient le plus librement disponibles et réutilisables, au plus proche du domaine public.

Le droit, certes, évolue et comme l'indique Lionel Maurel dans sa contribution sur « La réutilisation des données de la recherche après la loi pour une République numérique » l'adoption en France de la loi Valter, puis de la loi Lemaire ont permis de poser des principes généraux concernant l'utilisation des données produites dans le cadre des activités de recherche des établissements d'enseignement supérieur et de recherche. Mais il existe encore bien des cas où la loi demeure muette ou insuffisamment explicite et dans ce cas seules les règles éthiques et déontologiques peuvent venir au secours du chercheur. Là encore, vous trouverez dans ce guide des définitions précises de ces notions et des recommandations quant au rôle qu'elles peuvent jouer pour encadrer l'activité de recherche. Cette question est particulièrement prégnante s'agissant de la diffusion des matériaux de terrains des ethnologues. Comme vous le constaterez, la contribution de Marie-Dominique Mouton sur « Dématérialisation et valorisation des matériaux de terrain des ethnologues, l'archiviste face aux questions éthiques », fondée à la fois sur les acquis de l'expérience et sur une réflexion menée avec ses pairs, s'avère à cet égard extrêmement éclairante.

La deuxième partie du guide, qui présente des articles sélectionnés dans le cadre de l'appel à contributions, s'articule autour de trois thématiques : les archives orales, les données sensibles de la recherche et l'évolution du droit en matière de numérique. Chacune de ces thématiques est au demeurant introduite par un article d'un des membres du comité scientifique, qui permet à la fois de donner des clés de compréhension et de révéler toute la cohérence de ces différents témoignages.

Ainsi, Florence Descamps explique dans son introduction au chapitre sur les archives orales la place particulière qu'occupent ces ressources documentaires en SHS. Leur émergence progressive, à l'ombre des archives papier, les questionnements qui ont suivi, liés à leur mise en valeur, les enjeux juridiques, éthiques et déontologiques auxquels elles sont désormais confrontées, sont autant de questions mises en lumière par les différentes contributions proposées. Là encore, la démarche est pragmatique. L'article de Raphaëlle Branche, Florence Descamps, Frédéric Saffroy et Maurice Vaïsse a pour objet de présenter les règles juridiques qui s'appliquent à la collecte, au traitement et à l'exploitation des témoignages oraux. Quant aux autres contributions, elles se fondent sur le terreau de l'expérience, seul capable de révéler toutes les subtilités auxquelles le chercheur va se trouver confronté : celle de la mise en ligne d'entrevues d'histoire orale enregistrées dans le cadre d'un projet mené par Leslie Mc Cartney ; celle du consentement éclairé, exposé par Myriam Fellous-Sigrist ; la question des données qualitatives à caractère personnel, évoqué par Marie Huyghe, Laurent Cailly et Nicolas Oppenheim ; ou encore celle relative aux modalités de réutilisation des sources enregistrées exposée par Francesca Biliotti, Silvia Calamai et Véronique Ginouvès à partir de l'exemple comparée de deux phonothèques, celle de l'université de Sienna et celle de la MMSH d'Aix-en-Provence.

Le deuxième chapitre, qui porte sur les données sensibles de la recherche, s'ouvre sur une réflexion introductive qui amène à s'interroger, sous la plume de Laurent Dousset, sur ce qu'est une archive numérique et sur ce qui peut être ou ne pas être archivé, illustré par un cas concret, cruel mais particulièrement éclairant,



permettant de mieux saisir ce que recouvre le terme de donnée sensible. Les quatre contributions qui composent ce chapitre permettent là encore d'éclairer à l'aune de l'expérience les questions les plus délicates posées par la gestion de ces données. Les problèmes d'anonymat et de confidentialité sont ainsi exposés par Selma Bendjaballah, Sarah Cadorel, Émilie Fromont, Guillaume Garcia, Émilie Groshens et Emeline Juillard au travers de leur projet beQuali de constitution d'une banque de données des enquêtes qualitatives qui ambitionne la réutilisation et donc la mise à disposition « neutre » de données « brutes ». Les questions d'éthiques et de déontologie qui doivent présider au traitement documentaire et à la valorisation d'archives orales opérés par l'archiviste-documentaliste sont présentées par Maëlle Meriaux dans l'article « Du remède par les plantes à la sorcellerie ». Sur ces mêmes questions, mais avec une méthodologie différente, Iwona Dudek et Jean-Yves Blaise abordent la problématique de la transmission des protocoles en science patrimoniale au travers de leur projet MEMORIA. Et c'est enfin sur la délicate question du traitement des données d'un défunt dans un contexte de recherche présentée par Jean-Charles Ize que s'achève ce chapitre.

Bien que cette adaptation ne se fasse pas en parfaite instantanéité, le droit, comme nous l'avons évoqué, évolue au gré de l'évolution de la société. C'est donc dans le but de dresser un état des lieux sur l'évolution du droit en matière de numérique que s'est construit ce troisième chapitre, introduit par la contribution de Philippe Mouron, « Droit d'auteur et diffusion numérique des données de la recherche », qui met en exergue le conflit existant entre la liberté de la recherche et le respect du droit d'auteur, tout en exposant les solutions proposées à cet égard par le nouveau cadre juridique posé par la loi pour une République numérique du 7 octobre 2016.

À l'instar des précédents chapitres, nous retrouvons ici le balancement entre le droit et la morale, entre ce qui est permis par la loi et ce que nous dicte notre conscience de chercheur. C'est une des pistes de réflexion que nous livre Isabelle Gras, dans son article sur « Les enjeux éthiques et juridiques du dépôt des travaux scientifiques dans une archive ouverte ». Forte de son expérience au service du développement de l'archive ouverte HAL AMU de notre université, elle nous invite à repenser la question de la privatisation des droits d'auteurs par les éditeurs à l'aune de l'*open access*, sous couvert de l'article 30 de la loi pour une République numérique qui permet à présent au chercheur de mettre gratuitement à disposition par voie numérique ou, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication la version finale de son manuscrit accepté pour publication. Si l'article 38 de cette même loi est moins connu que cet article 30, il n'en a pas moins suscité de nombreux débats, comme nous le rappelle Pierre-Carl Langlais dans sa contribution portant sur « Une exception au "*text & data mining*" : pour quoi faire ? » Il y expose de façon particulièrement claire les enjeux et les principaux obstacles de la pratique de l'extraction de texte et de données, qui n'ont pas pu être totalement résolus par la loi de 2016. Ce chapitre s'achève enfin sur un article de Marie-Luce Demonet relatif à « La confiscation des données issues de l'humanisme numérique : un paradoxe résistant » qui attire notre attention sur les zones d'ombre de ce paysage juridique en construction.

Si vous êtes arrivés au terme de cette préface, je ne devrais plus avoir à vous démontrer l'intérêt de cet ouvrage qui non seulement dresse un état des lieux

extrêmement complet de la législation applicable en matière de diffusion des données en SHS, mais qui va bien au-delà, en exposant les règles éthiques et déontologiques qui doivent s'imposer à chacun d'entre nous dans le silence de la loi.

Ce guide, vous l'avez compris, nous amène à devenir des acteurs responsables et engagés de ce monde de la recherche en pleine mutation. Soyons conscients que chacun de nos actes participe de sa construction, que chaque arbitrage que nous opérerons au profit de l'accès des connaissances au plus grand nombre limitera la place laissée aux marchands de profits et que ce sont nos usages d'aujourd'hui qui fonderont le droit de demain. Cet ouvrage nous y aidera et, qui sait, fera peut-être de vous les rédacteurs de la prochaine édition de demain...