

ÊTRE CHERCHEUR AU 21^{ÈME} SIÈCLE DANS UN ENVIRONNEMENT NUMÉRIQUE

TO BE A 21TH CENTURY RESEARCHER IN A DIGITAL ENVIRONMENT

Par Serge-Georges ROSOLEN^{1,2}, Jean-Pierre HUGOT³, Frédéric VEILLEROT⁴ & Amandine PLUCHET⁵
(Communication présentée le 21 Juin 2018,
Manuscrit accepté le 2 Août 2018)

AVERTISSEMENT AUX LECTEURS

Les URL cités dans ce texte rédigé en juin 2018 doivent être vérifiés régulièrement afin de s'assurer de leur validité. De plus, les auteurs précisent que les sources citées le sont à titre d'exemples sans prétention à l'exhaustivité et en toute neutralité.

RÉSUMÉ

La révolution numérique a considérablement modifié l'environnement du chercheur. Avec le développement du libre accès et des archives ouvertes, la surabondance d'informations, la diversité des supports et les mutations concomitantes des modèles juridiques et économiques appliqués à la publication scientifique, le chercheur est confronté à un environnement foisonnant en perpétuelle évolution. Il doit adopter une démarche stratégique pour effectuer ses recherches documentaires afin de dresser un inventaire de l'état de l'art le plus exhaustif possible au sein de cette « infobésité » croissante. Il doit également construire son identité numérique pour valoriser ses travaux et protéger ses droits. Enfin il doit référencer et diffuser largement ses données *via* des réseaux professionnels et institutionnels. Afin de pouvoir gérer cette complexité, le chercheur du 21^{ème} siècle a besoin de travailler en collaboration avec des experts de l'information scientifique et technique. Il est en effet nécessaire de savoir prendre en charge l'indexation et la gestion des métadonnées, leur référencement dans des moteurs de recherche ainsi que leur diffusion grâce à un archivage électronique pérenne. Ceci tout en tenant compte du droit de la propriété intellectuelle et des relations contractuelles avec les éditeurs puisque ces experts assurent également une veille sur le contexte juridique, les directives européennes et l'adaptation aux systèmes juridiques étrangers. Qu'il soit doctorant ou confirmé, le chercheur du 21^{ème} siècle ne peut ignorer cet environnement, notamment les outils et la démarche méthodologique qui permettent de s'y adapter. Cet article, qui contient un glossaire et une liste des sites internet d'intérêt, s'emploie à les résumer.

Mots-clés : doctorant, chercheur, environnement numérique, veille documentaire, données de la recherche, plan de gestion de données, libre accès, archives ouvertes, identité numérique.

(1) Centre de Recherche Institut de la Vision UMR-S968 Inserm/UPMC/CHNO des XV-XX 75012 Paris
Courriel : serge.rosolen@inserm.fr

(2) Clinique Vétérinaire Voltaire, 92600, Asnières

(3) Muséum National d'Histoire Naturelle UMR 7205 CNRS MNHN UPMC EPHE, 75231 Paris cedex 05

(4) Université Paris-Sud, Service Commun de la Documentation, Bibliothèque des sciences / Cellule formation des usagers, 91405 Orsay

(5) Université Paris-Sud, Service Commun de la Documentation, responsable des services à la recherche, 91405 Orsay Cedex

ABSTRACT

With the development of open access and open archives, and the resulting glut of information, the digital revolution significantly changed the environment of researchers. In front of the diversity of media and the concomitant changes in legal and economic models applied to scientific publication, the researcher faced an exuberant environment in perpetual evolution. In order to draw up a comprehensive inventory of the state of the art, he must adopt a strategic approach to carry out his documentary research. He also must build his digital identity to actualize his work and protect his own rights. Finally he must reference and widely distribute his work via professional and institutional networks. In order to manage this complexity, the 21st century researcher needs to work in collaboration with experts in scientific and technical information. It is indeed necessary to know how to manage the indexing and the handling of metadata, their referencing in search engines and their diffusion thanks to long-lasting electronic archiving. In addition, he has to take into account the legal context, including the intellectual property laws, the contractual relations with the publishers, the European directives and the adaptation to foreign legal systems. Whether being a doctoral student or a senior scientist, he cannot ignore this environment, including the related tools and methodological approaches. The present article, which contains both a glossary and a list of web sites of interest, strives to summarize these.

Key words: *PhD student, researcher, digital environment, documentary intelligence, research data, data management plan, open access, open archives, digital identity.*

INTRODUCTION

L'essor du numérique à partir des années 1990 a considérablement modifié l'environnement du chercheur. Dans un contexte de surabondance d'informations (« infobésité ») et de diversité des supports (revues, web, réseaux sociaux, etc), le chercheur doit savoir repérer efficacement les publications en lien avec son sujet, enregistrer et organiser ses références bibliographiques, connaître les droits d'auteur qui s'appliquent aux publications scientifiques. En outre, il doit savoir déposer un document dans une archive ouverte et gérer ses données de recherche et son identité numérique. Pour exploiter et diffuser ses données, il doit donc élaborer une véritable stratégie. Dans cet environnement complexe et en perpétuelle évolution, le chercheur doit aussi être un veilleur. La collaboration avec les experts de l'information scientifique et technique, la connaissance de leurs outils et de leur démarche méthodologique s'avèrent par conséquent essentielles.

L'objet de cet article est de proposer à un jeune chercheur une méthodologie applicable à ses premiers travaux et en particulier à sa thèse. On listera d'abord les étapes permettant de chercher, d'identifier et de collecter des documents relatifs à un sujet de recherche. La deuxième partie sera consacrée à la valorisation et à la diffusion des travaux (publications & données de la recherche). Enfin, la troisième partie sera consacrée à l'identité numérique du chercheur.

RECHERCHE ET VEILLE DOCUMENTAIRE**Première étape : délimiter le contexte de la recherche et identifier les mots-clés significatifs**

Une fois le sujet bien délimité (en répondant à des questions du type : « Qui ? Quoi ? Quand ? Où ? Comment ? Pour qui ? Pourquoi ? », le chercheur doit traduire les concepts en mots-clés, et chercher l'origine et la définition des termes, dans des dictionnaires encyclopédiques généraux (<http://www.larousse.fr/dictionnaires>) et des articles de synthèse, etc. Il doit aussi s'intéresser à la variation du sens des mots, relever les synonymes, rechercher les termes équivalents dans des langues étrangères (<http://www.gdt.oqlf.gouv.qc.ca/> ; <http://iate.europa.eu/SearchByQueryLoad.do?method=load>) et noter les mots-clés mentionnés dans les documents au fur et à mesure de leur consultation. L'utilisation du thésaurus⁶ MeSH (Medical Subjects Headings) de la National Library of Medicine (<http://www.ncbi.nlm.nih.gov/mesh>) est le moyen le plus efficace pour construire un référentiel de mots-clés. Le site de l'INSERM assure la traduction en français du MeSH (<http://mesh.inserm.fr/mesh/>). Le portail TERMSCIENCES, développé par l'Institut National de l'Information Scientifique et Technique (INIST) a pour but de valoriser et de mutualiser les ressources terminologiques des organismes publics de recherche et d'enseignement supérieur pour aboutir à la constitution d'un référentiel commun (<http://www.term-sciences.fr/>). Le chercheur consultera et utilisera aussi avec profit les mots-clés RAMEAU (Répertoire d'Autorité-Matière

(6) Répertoire structuré de termes (mots-clés) pour l'analyse de contenu et le classement de documents servant de référence internationale dans le domaine des sciences biomédicales.

Encyclopédique et Alphabétique Unifié : <http://catalogue.bnf.fr/recherche-sujets.do?pageRech=rsu>) thésaurus géré par la Bibliothèque Nationale de France (BNF) ; ce sont les mots-clés issus de RAMEAU qui seront utilisés par les bibliothécaires pour décrire la thèse du doctorant. RAMEAU présente également l'avantage de renvoyer aux équivalents MeSH du terme ainsi qu'au LCSH (Library of Congress Subject Headings : <http://id.loc.gov/authorities/subjects.html>), le thésaurus de la Bibliothèque du Congrès à Washington. Une fois le mot-clé validé et replacé dans son environnement structuré, ce mot devient un « mot-sujet descripteur » permettant d'optimiser les recherches dans les catalogues de bibliothèques ou dans les moteurs de recherche des grandes bases de données qui utilisent un langage contrôlé comme *Medline*. Ces mots-clés correctement référencés (mots-sujet descripteur) lui permettront de faire des requêtes pertinentes dans les moteurs de recherche et les bases de données scientifiques. Plusieurs possibilités s'offrent alors ; utiliser ces mots clés en recherche simple ou bien en recherche avancée avec l'utilisation d'opérateurs booléens (AND, OR, NOT, etc...) et de troncatrices. Le chercheur doit aussi savoir adapter les termes de sa recherche à la langue de la base de données qu'il consulte en particulier : une bonne maîtrise de l'anglais devient nécessaire.

Deuxième étape : connaître les outils pour rechercher l'information

a) Rechercher des pages web dans des moteurs de recherche (en libre accès) : le web visible?

Les principaux moteurs de recherche généralistes sont Google, Bing, Yahoo!, Qwant et Mozbot. Il existe aussi des métamoteurs⁸ comme Dogpile, Yippy, Ixquick ou Seek. L'indexation est faite automatiquement par des logiciels robots utilisant des algorithmes, par exemple, *PageRank* pour Google (<http://www.pagerank.fr/>), fondé entre autres sur le nombre de liens entre pages et d'occurrences des termes dans les pages. Ces moteurs privilégient toutefois les pages commerciales et ne signalent pas (ou très peu) des données se trouvant dans le web invisible : par exemple les catalogues de bibliothèques ou les bases de données consultables sur abonnement.

Il existe des moteurs de recherche spécialisés gratuits pour le public universitaire. Citons *Google Scholar* (<http://scholar.google.fr/schhp?hl=fr>) qui indexe le contenu de périodiques et d'ouvrages en ligne, académiques et institutionnels. Cependant, il faut noter que *Google Scholar* donne accès à des travaux à divers états d'avancement qui ne sont parfois pas encore validés par les pairs. Attention, *Google Scholar* n'est pas une garantie d'obtenir les nombreux documents contenus dans les bases de données spécialisées payantes dont l'accès est réservé aux seuls abonnés. *Semantic Scholar* (<https://www.semanticscholar.org/>) présente l'intérêt d'être fondé sur une intelligence sémantique artificielle. *Worldwide science* (<https://worldwidescience.org>), *Science Open* (<https://scienceopen.com/search>), *Microsoft Academic* (<https://academic.microsoft.com>) sont des sites également utilisés.

Isidore (<https://www.rechercheisidore.fr/>) est à la fois un moteur de recherche et une plateforme de recherche dans le domaine des sciences humaines qui s'adresse en particulier aux enseignants, chercheurs, doctorants et étudiants. Il s'appuie sur les principes du web de données et permet de consulter des données en accès libre (« open access ») : bibliothèques numériques (Gallica de la BnF, Bibliothèque Inter-Universitaire de Médecine), archives ouvertes (HAL-SHS ; theses.fr ; etc.), plateformes (Cairn.info ; Persée ; Revues.org ; etc.). Citons également, pour les neurosciences, le moteur de recherche du Centre Thématique de Recherche en Neurosciences de l'Université de Laval (<https://neuro.ulaval.ca>).

Isidore (<https://www.rechercheisidore.fr/>) est à la fois un moteur de recherche et une plateforme de recherche dans le domaine des sciences humaines qui s'adresse en particulier aux enseignants, chercheurs, doctorants et étudiants. Il s'appuie sur les principes du web de données et permet de consulter des données en accès libre (« open access ») : bibliothèques numériques (Gallica de la BnF, Bibliothèque Inter-Universitaire de Médecine), archives ouvertes (HAL-SHS ; theses.fr ; etc.), plateformes (Cairn.info ; Persée ; Revues.org ; etc.). Citons également, pour les neurosciences, le moteur de recherche du Centre Thématique de Recherche en Neurosciences de l'Université de Laval (<https://neuro.ulaval.ca>).

b) Rechercher des références d'ouvrages dans des catalogues de bibliothèques

Le *Système Universitaire de Documentation* (SUDOC) est un catalogue collectif, alimenté par l'ensemble des bibliothèques universitaires françaises et autres établissements documentaires utiles pour l'enseignement supérieur et la recherche (<http://www.sudoc.abes.fr/>). Il recense 12 millions de notices de tous types de documents (revues, livres, e-books ainsi que toutes les thèses françaises imprimées ou en version électronique).

WorldCat est la base de données bibliographiques en ligne de OCLC (Online Computer Library Center) (<http://0-www.worldcat.org.novacat.nova.edu/?lang=fr>) qui contient des données relatives à plus de 72 000 bibliothèques publiques et privées dans le monde. C'est la plus importante base bibliographique mondiale avec 271 millions d'enregistrements pointant sur plus de 2 milliards de références.

c) Rechercher des thèses, des rapports, de la littérature grise

Plusieurs sites spécialisés permettent de rechercher des thèses :

- *Thèses.fr* (<http://www.theses.fr/>), moteur de recherche pour les thèses de doctorat françaises, fournit un point d'entrée à toutes les thèses de doctorat en cours de préparation ou soutenues depuis 1985 en France, quel que soit leur support, ainsi qu'aux personnes et organismes en lien avec ces thèses. On y trouve des thèses non diffusées sur internet, des thèses sous embargo ou en accès libre, en fonction de l'option choisie par le doctorant.
- Le serveur TEL (thèses-en-ligne : <http://tel.archives-ouvertes.fr/>) a pour objectif de promouvoir l'auto-archivage en ligne des thèses de doctorat et d'Habilitation à Diriger des Recherches (HDR). TEL est un environnement particulier de l'archive ouverte HAL (*Hyper Articles en Ligne*) (<http://hal.archives-ouvertes.fr/>).

(7) Catalogues de bases de données consultables en libre accès.

(8) Outils soumettant une même requête à plusieurs moteurs de recherche.

- DART-Europe ETheses Portal (<http://www.dart-europe.eu/basic-search.php>) est un moteur de recherche qui propose en libre accès (open access) des thèses provenant de plusieurs centaines d'universités européennes.
- Quelques sites sont spécialisés pour trouver des rapports : la Bibliothèque des rapports publics (<http://www.ladocumentationfrancaise.fr/rapports-publics>), la base LARA éditée par l'INIST (<http://lara.inist.fr/>) et le site Open Grey (<http://www.opengrey.eu/>) qui recense la littérature grise (rapports techniques, actes de colloques, thèses,...) en Europe.

d) Rechercher des articles en texte intégral ou des références d'articles dans des bases de données en libre accès (voir définition plus bas), à comité éditorial.

- Le Directory of Open Access Journals (DOAJ) (<https://doaj.org>) est le principal portail pour les revues en libre accès (voie dorée, voir ci-dessous).
- Le Directory of Open Access Repositories (OpenDOAR) (<http://www.opendoar.org>) donne accès à des dépôts d'archives ouvertes institutionnelles de niveau académique.
- Le Directory of Open Access Books (DOAB : <https://www.doabooks.org>) signale les e-books en libre accès publiés par des éditeurs académiques à comité de lecture.
- Pour les articles, le Bielefeld Academic Search Engine (BASE : <https://www.base-search.net/?l=fr>) donne accès à des ressources académiques sur le web (100 millions de documents dans 5 000 sources) avec le texte intégral dans 60% des cas.
- L'Open Access Infrastructure for Research in Europe (<https://www.openaire.eu/search>) donne accès à des ressources académiques européennes.
- Dans le domaine scientifique et médical, Medline via l'interface PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) contient 26 millions de références. PubMed Central est un service d'archives ouvertes proposé par la National Library of Medicine (<http://www.ncbi.nlm.nih.gov/pmc/>).
- La Public Library of Science (PloS) édite la revue PloSOne (<http://www.plosone.org/>) en libre accès sous licence Creative Commons, qui est devenue le premier périodique scientifique au monde en terme de volume de publication. PloS édite également un blog (<http://blogs.plos.org/everyone/>) et d'autres revues : PLOS Biology, PLOS Computational Biology, PLOS Genetics, PLOS Medicine, PLOS Neglected Tropical Diseases, PLOS Pathogens.

e) Rechercher des articles dans des bouquets de revues proposées par des éditeurs

- Cambridge University Press (<https://www.cambridge.org>) correspond à un bouquet de revues traitant de toutes les disciplines scientifiques. La couverture temporelle de la collection est large : les parutions les plus anciennes datent de 1770.

- Science Direct (<http://www.sciencedirect.com>) est un bouquet de l'éditeur Elsevier. La base comprend plus de 13 millions d'articles mais la recherche ne repose pas sur des mots-clés contrôlés.
- Springer (<http://link.springer.com>) est un bouquet de revues (2 100 titres) et livres en ligne (46 000) dans toutes les disciplines académiques : sciences, médecine, pharmacie, sciences humaines, psychologie, économie et gestion, droit.
- Web of Science fédère plusieurs bases de données couvrant 8700 publications scientifiques de référence. L'accès se fait via ISI Web of Knowledge (<https://login.webofknowledge.com>), service d'information universitaire en ligne produit par la société ISI Institut for Scientific Information de Clarivate Analytics. L'intérêt de WoS est qu'il est aussi un outil bibliométrique ; par le biais de l'option "Times cited" (= nombre de fois cité), il est possible de connaître le nombre de citation d'un article, mais aussi de classer et de sélectionner les résultats de sa recherche documentaire en fonction de ce paramètre. Particulièrement intéressante, l'option "Analyze results" (= analyser les résultats) permet notamment de connaître les thématiques (Web of Science Categories) majoritairement représentées dans les résultats et surtout les auteurs les plus productifs dans les domaines recherchés par l'utilisateur, notamment sous la forme d'un graphique très clair.

Il existe deux moyens d'obtenir le texte en ligne d'un article figurant dans une base de données commerciale. Si l'université, la grande école ou l'institut de recherche dont dépend le chercheur a souscrit un abonnement à la base de données, l'accès au texte intégral sera gratuit à condition que l'utilisateur s'identifie sur le portail de l'institution. Sinon, l'accès est payant, à partir du site de l'éditeur. Pour autant, l'accès à une base de données d'éditeur par son institution ne signifie pas l'accès à toutes les revues qui y figurent : les abonnements électroniques fonctionnent par sélections de titres qu'on appelle « bouquets ». Enfin, Le chercheur peut également installer sur son navigateur l'add-on ISTEEX (<https://addons.istex.fr>) qui lui permettra d'accéder à des ressources payées par le Ministère de l'Enseignement Supérieur et de la Recherche pour l'ensemble des institutions académiques françaises, par le biais de son institution.

Troisième étape : évaluer les résultats de la recherche faite sur internet – mise en place d'une veille informationnelle

Lorsqu'un site internet n'est pas d'origine institutionnelle, il faut prendre en compte les critères suivants : sa fiabilité, sa pertinence, la traçabilité des informations (l'auteur cite-t-il des documents source ?) et leur fraîcheur (les informations sont-elles encore d'actualité ?). La bibliométrie permet l'évaluation de la production scientifique d'un chercheur à partir du nombre de ses publications, du prestige des revues dans lesquelles elles ont été publiées et des citations auxquelles elles ont donné lieu. Elle repose sur le calcul de divers indicateurs à partir des bases de données bibliographiques. Le facteur d'impact (Impact factor) est un critère d'évaluation d'une revue et reflète donc son importance au sein de la communauté

des chercheurs et de la recherche. Le facteur d'impact est le calcul du nombre moyen de citations de chaque article de la revue. Plus le facteur d'impact est élevé, plus la revue est considérée comme importante dans son domaine. L'Eigenfactor (EF) est un indicateur de notoriété des revues indexées dans la base de données *Web of Science* (WoS) du producteur américain Clarivate Analytics, qui indexe plus de 12 000 revues toutes disciplines confondues incluant les sciences humaines et sociales. Les revues indexées dans le WoS et dotées d'un facteur d'impact sont enregistrées dans le *Journal Citation Reports* (JCR) de Thomson Reuters. Le H-Index est un indicateur qui reflète à la fois le nombre de publications d'un auteur et le nombre de citations par publication. Il permet d'aller plus loin que la simple mesure du nombre d'articles ou de citations et d'évaluer la production scientifique d'un auteur. Le H-index est notamment utilisé par Clarivate Analytics ; les onglets "InCites" et "Journal Citation Reports" de WoS permettent notamment de déterminer quelles revues et quels auteurs sont le plus cités. L'indicateur H-index est également utilisé par Google Metrics (<https://scholar.google.com/intl/en/scholar/metrics.html>) ; il est possible d'inclure un calculateur de H-index directement dans sa fenêtre de recherche *Google Scholar* en installant l'add-on "Scholar H-Index Calculator for Google Chrome™" (<https://chrome.google.com/webstore/detail/scholar-h-index-calculator/cdpobfbhbdplpblocckjgkjkjmifbng?hl=fr>) dans le navigateur Google-Chrome. Enfin, Scopus (<https://www.scopus.com/home.uri>), l'outil bibliométrique de l'éditeur Elsevier, utilise aussi le H-index.

Le chercheur peut enregistrer et conserver les résultats de sa recherche. En outre, il peut déclencher dans la base de données une veille sur sa requête : à chaque fois qu'un nouvel article correspondant paraîtra il en sera informé.

Deux solutions s'offrent à lui :

- recevoir des alertes-mail
- déclencher une veille par flux RSS (option qui n'est malheureusement pas présente sur toutes les bases de données) ; ce format d'information peut être notamment réceptionné et lu par des logiciels de messagerie tels que : Microsoft-Outlook, Thunderbird ou Zimbra ; le logiciel Zotero, que nous allons voir ci-dessous, permet également d'agréger les flux RSS.

Quatrième étape : constituer et gérer sa propre bibliographie

Les données fiables et complètes recueillies selon les recommandations décrites ci-dessus permettent au chercheur/doctorant de constituer sa propre base de données bibliographiques qu'il doit gérer grâce à un logiciel de gestion bibliographique. Un logiciel bibliographique lui permet de :

- 1) récupérer, importer, mémoriser des documents lus publiés ou en sa possession ;
- 2) gérer, modifier, retrouver et partager ces références et ces documents ;
- 3) éditer, mettre en forme, rédiger, réutiliser des données existantes et insérer de nouvelles données.

Le chercheur/doctorant doit apporter un soin particulier à ce travail qui lui servira tout au long de sa vie professionnelle. Pour cela il existe plusieurs types de logiciels, dont trois particulièrement connus :

- *EndNote* est facile à utiliser directement lors de la rédaction d'un texte et facilite l'importation de références provenant de catalogues de bibliothèques mais il est payant (<https://endnote.com>).
- *Zotero* est un logiciel libre, gratuit, très utilisé dans les universités et bien adapté pour faire de la veille (<https://www.zotero.org>),
- *Mendeley* est également un logiciel gratuit du milieu académique. Son site est très collaboratif avec plus de 175 millions d'articles déjà référencés, 2 millions d'inscrits (<https://www.mendeley.com>).

VALORISATION ET DIFFUSION DES DONNÉES DE LA RECHERCHE

Les données de recherche : définition, enjeux et craintes

Quelques définitions

La définition la plus usuelle des données de la recherche (« research data ») provient de l'OCDE : (<http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=159&Lang=fr&Book=False>) : « 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 valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche ». On peut classer les types de données en quatre catégories : les données d'observation (capturées en temps réel, généralement uniques et impossibles à reproduire), les données expérimentales (obtenues à partir d'équipements de laboratoires, souvent reproductibles mais coûteuses), les données de simulations numériques (générées par des modèles informatiques, souvent reproductibles si le modèle est correctement documenté) et les données dérivées ou compilées (issues du traitement ou de la combinaison de données brutes, souvent reproductibles mais coûteuses). Le développement de nouvelles technologies produit de grandes quantités de données numériques qu'il faut gérer, stocker, partager et mettre à disposition de la communauté scientifique. Il s'agit d'une autre façon de faire de la science qu'on désigne par Science Ouverte [Livre blanc - une science ouverte dans une république numérique: (<http://www.cnrs.fr/dist/z-outils/documents/livre-blanc-loi-numerique.pdf>)]. Le législateur a intégré cette nouvelle approche dans l'article 17 art.L.533-4 de la loi pour une République Numérique en rendant obligatoire la libre diffusion des données issues d'une activité de recherche financée par des dotations de l'État.

Pourquoi donner accès à ces données ? La recherche moderne se construit sur des échanges approfondis entre scientifiques et progresse en s'appuyant sur des travaux antérieurs. Par conséquent, connaître les résultats (positifs ou négatifs) de recherches

antérieures est essentiel. Avec l'essor de l'informatique et des nouvelles technologies liées au numérique, le stockage, l'exposition et la diffusion des *données de la recherche* n'ont plus de limites. La révolution numérique permet ainsi de concevoir les travaux de recherche comme des élaborations et des traitements successifs de données, sur un continuum qui va des données brutes produites par l'instrument de recherche aux résultats publiés dans des articles, des livres ou de la littérature grise. La détermination de certaines données comme *nécessaires pour valider les résultats de la recherche* est elle-même un enjeu scientifique. Cependant cette obligation de mise à disposition du public d'une telle quantité d'information pose un certain nombre de problèmes. D'un point de vue *juridique et éthique*, le droit de propriété intellectuelle et de la protection des données personnelles doit être garanti. Les *infrastructures matérielles et les outils existants* sont encore insuffisants et onéreux aussi bien en termes de ressources humaines que d'infrastructures.

b) Gérer ses données

Les données scientifiques ont un cycle de vie qui comprend les étapes de gestion, de conservation, de diffusion et de réutilisation (**figure 1**). Il est donc nécessaire d'établir un *plan de gestion de données* (DMP = « Data Management Plan ») qui s'inscrit dans une stratégie de leur mise à disposition et de leur diffusion (**figure 2**). Ce DMP doit être élaboré dès le début du projet de recherche afin que le chercheur définisse la destinée de ses données pendant et après ses travaux. Il existe des guides qui permettent de réaliser un DMP, en français (<https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>) ou, en anglais (<https://library.sydney.edu.au/research/data-management/index.html>). Citons également OPIDOR, un outil en ligne (<https://dmp.opidor.fr>). Ce DMP est un document évolutif qui doit être continuellement mis à jour tout au long

du projet. Certaines agences de financement et organismes de recherche l'exigent (par exemple les projets européens financés dans le cadre du programme Horizon 2020) (<http://www.horizon2020.gouv.fr>). C'est un outil de travail qui permet de gagner du temps, d'éviter la duplication des efforts et la perte de données. Il permet de faciliter la conservation des données, mais également d'assurer l'intégrité et la reproductibilité de la recherche. De plus, il favorise la réutilisation et l'innovation en augmentant la visibilité de la recherche. Enfin, il facilite la rédaction d'un *data paper* et son dépôt dans un entrepôt.

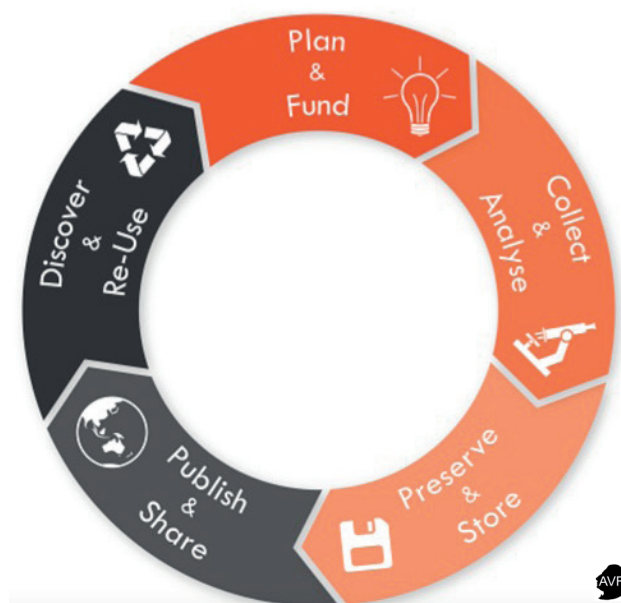


Figure 1 : cycle de vie des données scientifiques d'après : <https://library.sydney.edu.au/research/data-management/research-data-management.html>. C'est l'ensemble des étapes de gestion, de conservation, de diffusion et de réutilisation des données liées aux activités de recherche.

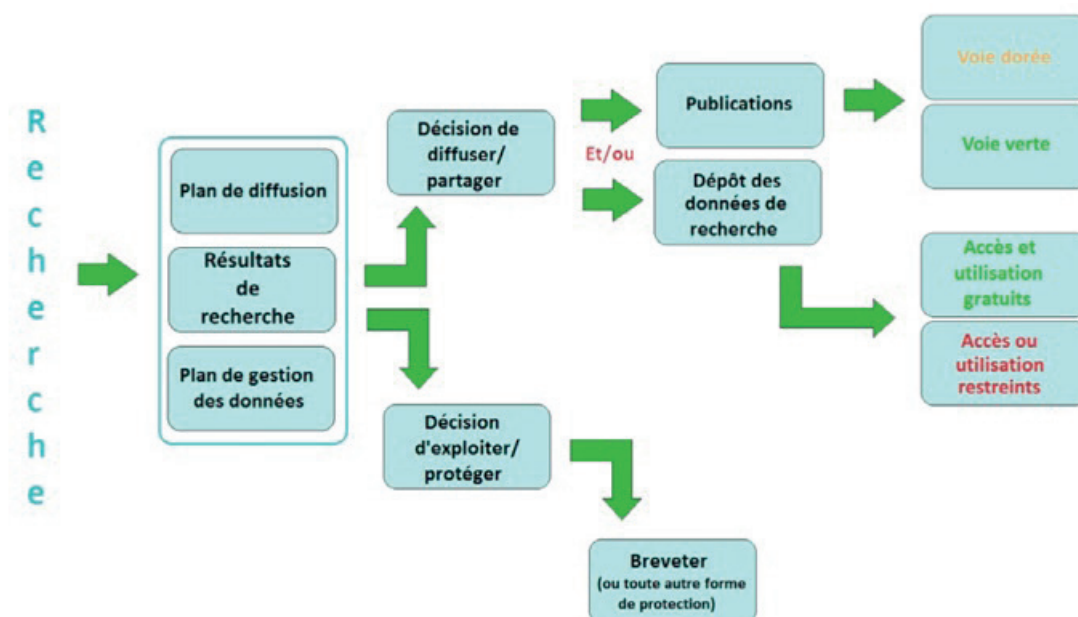


Figure 2 : libre accès aux publications scientifiques et aux données de la recherche d'après les Lignes directrices sur les règles pour le libre accès aux publications scientifiques et pour le libre accès aux données de la recherche dans Horizon 2020 http://openaccess.inist.fr/IMG/pdf/lignes-directrices_libre-acces_horizon2020_version3.1_tr-fr.pdf

Afin que les données de la recherche soient réutilisables, le contexte de leur production doit être documenté de manière précise et intelligible, en particulier à l'aide de métadonnées. Les métadonnées sont interprétables par des machines et doivent permettre l'interopérabilité entre machines, le repérage et le « moissonnage » des données par le biais des moteurs de recherche. Elles doivent être complétées tout au long du cycle de vie de la donnée et permettre de répondre aux questions suivantes : Qui l'a créée ? Que contient elle ? Où est elle ? Quand a-t-elle été créée ? Pourquoi a-t-elle été créée ?

c) Stocker & conserver ses données

Stocker et conserver ses données sur une longue durée a un coût qu'il faut prendre en considération. Plusieurs critères doivent être envisagés : peuvent-elles être réutilisées, doivent-elles être conservées pour des raisons juridiques et/ou leur valeur potentielle ? On doit prévoir une conservation de longue durée pour les données uniques et difficiles à reproduire. Les formats ouverts, non-proprétaires et d'un usage répandu au sein d'une communauté de recherche seront privilégiés. Des exemples sont disponibles sur le site HAL (Hyper Article en Ligne : <https://hal.archives-ouvertes.fr>) créé par le Centre pour la Communication Scientifique et Directe (CCSD). Ces données sont stockées dans des entrepôts (« repositories ») pour une recherche ou une réutilisation. Un identifiant pérenne ou un numéro d'accès est attribué à chaque jeu de données. Citons SEANOE qui est un entrepôt disciplinaire de l'Ifremer (<https://data.ifremer.fr>), quelques entrepôts multidisciplinaires comme ZENODO (<https://zenodo.org>) ou institutionnels comme Edinburg DataShare (<https://datashare.is.ed.ac.uk>) et 4TU Datacentrum (<http://researchdata.4tu.nl/en/home/>). Il existe aussi des annuaires d'entrepôts multidisciplinaires, comme RE3data-Registry-of-data-repositories (<https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/re3data>) ou spécialisés en science de la vie et biomédicales comme FairSharing (<https://fairsharing.org>).

d) Partager et diffuser ses données

Ces données peuvent être publiées dans des « data papers » (publications dont le but est de décrire un jeu de données ou un ensemble de jeux de données brutes avec des métadonnées, mais sans interprétation des résultats) et dans les « fichiers supplémentaires » (*supplementary data*) associés à un article. L'accès en ligne aux données décrites dans le data paper se fait par lien pérenne [Digital Object Identifier (DOI) ; Uniform Resource Locator (URL)] pointant vers l'entrepôt de données où elles ont été stockées, ou sous forme de fichiers en annexe de l'article (exemple : des vidéos).

Cependant, avant de diffuser ses données, le chercheur doit prendre quelques précautions : s'assurer qu'il en est bien le seul auteur et qu'il détient l'autorisation du fournisseur (vérifier la licence d'utilisation). Il doit également anonymiser les données personnelles et identifier celles qu'il ne doit pas diffuser (confidentialité des informations commerciales, secret-défense natio-

nale, sécurité publique, etc.). Il devra placer ses données sous licence Creative Commons et obtenir des identifiants pérennes comme le DOI, le PURL (Persistent Uniform Resource Locator), l'Handle System (HANDLE) ou l'ePIC (European Persistent Identifier Consortium). Ces identifiants permettront de lier les publications aux jeux de données.

e) Diffuser ses données en libre accès (Open Access)

La Déclaration de Berlin sur le libre accès à la connaissance (2003) (<http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>) stipule que « les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de données brutes et de métadonnées, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédia ». Cette déclaration complète l'initiative de Budapest de 2002 (<http://www.budapestopenaccessinitiative.org/read>). On entend par libre accès à cette littérature et à ces données leur mise à disposition gratuite sur l'Internet public, permettant à chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers l'intégralité des textes et données de la recherche, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et de l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine, est de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités (INIST : <http://openaccess.inist.fr/?Libre-Acces-des-outils-pour-vous>). Enfin, en 2017, un collectif français représentatif des chercheurs et des professionnels de l'édition scientifique lance l'appel de Jussieu (<http://openaccess.inist.fr/?L-Appel-de-Jussieu>) pour la Science ouverte et la bibliodiversité dont le but est d'élaborer et de mettre en œuvre des modèles alternatifs adaptés aux objectifs de la science ouverte en affirmant la nécessité : de soutenir l'innovation pour une rénovation profonde des fonctions éditoriales : réformer les systèmes d'évaluation de la recherche; sécuriser et stabiliser le cadre juridique dans les différents pays. Le collectif appelle à la constitution d'un consortium international d'acteurs pour fédérer les initiatives.

Les deux voies pour le libre accès.

L'auto-archivage (ou « voie verte ») permet à l'auteur (ou l'un de ses représentants) d'archiver (ou déposer) l'article publié, ou le manuscrit final validé par ses pairs, dans un réservoir en ligne avant, pendant ou après sa publication. L'auto-archivage s'effectue dans des archives institutionnelles, nationales ou disciplinaires. Si le contenu n'a jamais été publié dans une revue, il peut être déposé immédiatement dans l'archive ouverte. Si le contenu a déjà été publié par un éditeur, celui-ci peut le libérer pour le rendre accessible gratuitement. Toutefois, afin que les utilisateurs continuent de souscrire aux abonnements courants, les éditeurs appliquent sur les contenus une période d'embargo, qui ne peut excéder 2 ans.

La publication en libre accès (ou « voie dorée ») permet la mise en libre accès d'un article dès sa publication. Le paiement des frais de publication (APC = Article Processing Charges) n'est pas à la charge du lecteur mais de l'auteur. Les APC sont payés par l'université, par l'Institut de recherche auquel l'auteur de l'article est rattaché ou par l'organisme de financement soutenant les recherches. La politique des éditeurs en matière de libre accès est disponible dans SHERPA/RoMEO (<http://www.sherpa.ac.uk/romeo/>). Il est important de préciser que le libre accès n'empêche pas d'exploiter les résultats de la recherche de façon commerciale, via un brevet, par exemple.

Les archives ouvertes : trouver des documents en libre accès.

Les chercheurs déposent leurs articles en texte intégral dans des réservoirs ou entrepôts de documents (Repository) après avoir pris connaissance de leur droit d'auteur. Les archives ouvertes suivent le protocole d'échange des données OALPMH (*Open Archives Initiative Protocol for Metadata Harvesting*) qui permet, pour chaque requête, une interrogation simultanée de tous les entrepôts par les moteurs de recherche spécialisés dans le moissonnage des serveurs d'archives ouvertes (« Harvesters »).

Outre les données décrites ci-dessus, on y trouve également des actes de conférences, des rapports, des thèses, des cours, des communications affichées (posters) et des articles à différents stades de publication (*preprints*, *postprints*, articles en version éditeur). La diffusion dans une archive ouverte permet ainsi au chercheur d'organiser des conversations scientifiques (circulation des *preprints*), d'établir publiquement une antériorité et accroître sa visibilité. La diffusion des découvertes scientifiques constitue une démarche citoyenne en mettant à disposition d'une large communauté les résultats de recherches financées par des fonds publics.

Les archives ouvertes sont organisées en collections, suivant le plan de classement d'une institution ou par thématiques. La page de consultation d'une collection peut être personnalisée tant sur le graphisme que sur le contenu. Les archives ouvertes peuvent également être accessibles via des portails ; ce sont des collections avec des fonctions supplémentaires comme l'utilisation d'un nom de domaine (URL) ou l'intégration d'un thesaurus de mots-clés. Il est également possible qu'une collection agrège sur le critère de l'affiliation tous les dépôts faits dans un portail, pour une meilleure visibilité.

Le site DOAJ (voir ci-dessous : <https://doaj.org>) répertorie les revues en *voie dorée*. La liste des archives ouvertes pour la voie verte (auto-archivage) se trouve sur le site OpenDOAR (voir ci-dessous : <http://www.opendoar.org>) et le *Registry of Open Access Repositories* (ROAR : <https://roar-map.eprints.org>). Il existe une infrastructure de libre accès pour la recherche en Europe (*Open Access Infrastructure for*

Research in Europe = OpenAIRE : <https://www.openaire.eu/search>). On trouve différents types d'archives ouvertes : des archives institutionnelles (Institut National en Recherche en Informatique et en Automatique, Sciences Po), des archives disciplinaires (*ArXiv* : <https://arxiv.org>) pour les mathématiques et la physique, *PubMedCentral* (<https://www.ncbi.nlm.nih.gov/pmc/>) pour le biomédical et les sciences de la vie) et des archives nationales multidisciplinaires comme HAL (voir ci-dessous). On trouve sur le site HAL (<https://hal.archives-ouvertes.fr>) un guide du dépôt et du bon usage.

Quel est l'intérêt du dépôt dans des archives ouvertes ?

Le dépôt en archives ouvertes augmente la *visibilité du chercheur*, de son laboratoire et de son établissement. Cette *visibilité* des équipes favorise l'attractivité vis-à-vis des financeurs. Notamment pour les jeunes chercheurs, déposer leur thèse dans des archives ouvertes *via* le serveur TEL (voir ci-dessous : <https://tel.archives-ouvertes.fr>) présente de multiples avantages : archivage quotidien et pérenne, avec des URL stables et un format de fichier évolutif dans le temps beaucoup plus sécurisé que dans le cas d'un document en format pdf déposé sur un réseau social et qui peut disparaître lors de remaniements du site ou lorsque les logiciels ne permettent plus de lire cette version de fichier pdf.

Bien qu'initialement non conçues dans ce dessein, les archives ouvertes sont de plus en plus utilisées à des fins bibliométriques car elles sont moissonnées par des moteurs de recherche scientifique comme *Google Scholar* ou *Worldcat*.

Qui valide l'entrée du document dans une archive ouverte ?

La plupart des plateformes sont des hébergeurs. L'auteur est responsable du contenu de son dépôt. Il doit donc s'assurer de détenir toutes les autorisations concernant notamment les contrats de recherche, les co-auteurs, les droits des éditeurs [(vérifier la politique des principaux éditeurs sur le site SHERPA/RoMEO (voir ci-dessous : <http://www.sherpa.ac.uk/romeo/>)], la confidentialité. Il est conseillé aux auteurs d'associer à leurs dépôts une licence Creative Commons (<https://creativecommons.org> ; <http://creativecommons.fr>). Il s'agit d'un ensemble de licences gratuites dans lesquelles les auteurs déterminent les droits attachés à leur travaux en combinant quatre éléments : la paternité, la modification, l'utilisation commerciale et la réutilisation dans les mêmes conditions que celles de la licence initiale. L'utilisation de ces licences permet notamment aux auteurs de présenter leurs travaux préliminaires, quel que soit le support (présentation orale, communication affichée, etc.) dans des congrès, colloques, etc., d'assurer leur diffusion tout en préservant leur paternité.

Le **tableau 1** présente les différents sites accessibles en libre accès.

Tableau 1 : principaux sites utiles en libre accès

- pour trouver et utiliser des mots-clés du MeSH (Medical Subject Headings) via le site de l'INSERM : <http://mesh.inserm.fr/mesh/>
- pour trouver des références dans des catalogues de bibliothèques :
SUDOC : <http://www.sudoc.abes.fr/>
WORLDCAT : <http://0-www.worldcat.org.novacat.nova.edu/?lang=fr>
- pour trouver des thèses, des rapports, de la littérature grise :
HAL : <http://hal.archives-ouvertes.fr/>
Pour les thèses : TEL : <http://tel.archives-ouvertes.fr/>
Pour les thèses européennes : DART-Europe : <http://www.dart-europe.eu/basic-search.php>
Pour les rapports : LARA : <http://lara.inist.fr/>
Pour la littérature grise : Open Grey : <http://www.opengrey.eu/>
- pour retrouver des articles en texte intégral dans des bases de données en libre accès :
DOAJ : <https://doaj.org>
OpenDOAR : <http://www.opendoar.org>
OPENAIRE : <https://www.openaire.eu/search>
Medline via l'interface PubMed : <http://www.ncbi.nlm.nih.gov/pubmed/>
PlosOne : <http://www.plosone.org/>
- pour optimiser l'utilisation du Web Of Science :
http://australe.upmc.fr/access/content/group/BUPMC_04/midi_biblio_05_WoS.pdf
- pour réaliser un plan de gestion de données (Data Management Plan) :
en français : <https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>
en anglais : <https://library.sydney.edu.au/research/data-management/index.html>
- pour stocker ses données dans un entrepôt :
Entrepôt multidisciplinaire ZENODO : <https://zenodo.org>
- pour connaître la politique des éditeurs en matière d'open access :
SHERPA/RoMEO : <http://www.sherpa.ac.uk/romeo/>
- pour tout savoir sur les licences Creative Commons :
en anglais : <http://creativecommons.org>
en français : <http://creativecommons.fr>
- Pour préparer, diffuser & stocker sa thèse :
Mode d'emploi pour les doctorants : <http://www.abes.fr/Theses/Pour-les-doctorants>
Vérifier qu'un projet de thèse sur le même sujet n'est pas en cours : <http://www.theses.fr>
Vérifier l'archivabilité par le Centre Informatique National de l'Enseignement Supérieur (CINES) : <https://facile.cines.fr>
- Logiciels de gestion de données :
Télécharger Zotero : <https://www.zotero.org>
Télécharger Mendeley : <https://www.mendeley.com>

L'IDENTITÉ NUMÉRIQUE DU CHERCHEUR ET LA GESTION DE SON e-IMAGE

Le chercheur doit construire son identité numérique

L'identité numérique est la représentation virtuelle et technologique d'une entité réelle (personne, structure, société). Elle peut se définir comme l'ensemble des traces qu'on laisse en ligne, volontairement ou non. On peut différencier l'identité

déclarative, l'identité active et l'identité passive. Le chercheur gère sa présence sur le web en choisissant d'être visible ou le plus anonyme possible. L'e-réputation est la conséquence de ce que l'on met en ligne : c'est la création d'une image de cette identité numérique. La visibilité sur le web permet au chercheur de promouvoir son travail de recherche et d'être bien identifié comme spécialiste dans un domaine. Elle permet d'assurer la promotion de ses articles en augmentant leur visibilité, d'échanger avec ses pairs, de trouver de nouveaux partenaires et de chercher un emploi en valorisant son CV. Gérer son image sur le web c'est choisir de rester le plus anonyme possible, ou au contraire être visible, voire le plus visible possible ou bien encore visible pour seulement une partie de ses activités. Les chercheurs soucieux de limiter leur visibilité, utilisent des pseudos, cloisonnent leurs identités en utilisant plusieurs comptes et remplissent les formulaires à minima. Ceux qui souhaitent être le plus visible participent à des blogs et à des réseaux sociaux.

a) Les identifiants chercheurs : ORCID et RESEARCHER ID

Un identifiant chercheur est un code numérique permanent et unique attribué à un chercheur. Il permet donc d'identifier le chercheur dans un environnement numérique, des bases de données, des archives ouvertes et de retrouver tous les documents dont il est l'auteur : articles mais aussi brevets, etc. Grâce à l'identifiant le chercheur évite les homonymies. D'autre part les différentes identités sous lesquelles il apparaît (nom de jeune fille, nom d'usage, prénom ou simple initiale) et sous lesquelles il signe (ou a signé) ses publications se trouvent liées. Cela lui évite aussi de ressaisir les informations le concernant lors d'exportations de références d'une base à l'autre.

b) Les blogs

Libres de forme et sans limites de caractères, ils sont accessibles à tous et permettent de se faire connaître hors de sa communauté. Les échanges se font par des billets. On distinguera des outils individuels sur lesquels le chercheur présente ses publications et des outils collectifs pour faire de la veille ou communiquer avec des collègues. Certains blogs peuvent se voir attribuer un *International Standard Serial Number* (ISSN). Très souples d'utilisation, les blogs sont des outils d'échange et de visibilité qui permettent aussi de s'entraîner à l'argumentation et à l'écriture de façon régulière. Tenir seul un blog est difficile, c'est pourquoi on a pu observer la création de nombreux portails de blogs institutionnels. Citons, par exemple la plateforme *Hypothèses* (<https://cleo.openedition.org/openedition/hypotheses>) du portail *OpenEdition* (<https://cleo.openedition.org>) proposé par le Centre pour L'Édition Électronique Ouverte (CLEO : <https://cleo.openedition.org/presentation/dates>). Le site indépendant *The Conversation* (<https://theconversation.com/fr>) permet à des chercheurs de publier des articles dans leur domaine d'expertise et ils peuvent ainsi se faire connaître et valoriser leurs travaux.

c) Les réseaux sociaux

Le microblogging avec Twitter

Twitter est un réseau social de microblogging fondé sur le principe d'un échange de messages courts (limités à 280 caractères). Certains chercheurs se sont emparés de cet outil très populaire afin d'échanger avec leurs pairs, mais aussi de faire connaître leurs travaux et leur environnement de travail au-delà de leur communauté. L'aspect interactif de Twitter peut générer des échanges nombreux et rapides, mais il faut bien définir l'usage que l'on souhaite en faire (communication scientifique ? échange avec d'autres chercheurs ? communication personnelle ?) et consacrer du temps à ses propres publications ainsi qu'au suivi d'autres usagers.

Twitter dispose d'un vocabulaire spécifique : un tweet (« gazouillis ») est un message posté sur Twitter, un retweet consiste à rediffuser le message d'un autre abonné et un hashtag (symbole # immédiatement suivi d'un mot-clé) est un tag ou mot-clé permettant de catégoriser un tweet et donc de faciliter la recherche.

Les réseaux sociaux professionnels

Ce sont des outils bien référencés sur *Google*. Bien que peu calibrés pour la recherche, ils permettent de s'ouvrir sur le monde industriel et commercial, de garder des contacts et de se créer des réseaux grâce à des groupes thématiques, de faire de la veille et de participer à des forums de discussion. Les plus connus sont *LinkedIn* et *Viadeo*.

Les réseaux sociaux académiques

Les deux principaux réseaux sont *Academia* avec plus de 32 millions de comptes et *ResearchGate* avec plus de 8 millions de comptes. Le premier est plus ciblé « sciences humaines » et le second, « sciences dures ». Le premier n'a pas de thesaurus alors que le second en possède un (liste fermée) ce qui limite l'éparpillement.

Ces réseaux sociaux, qu'ils soient professionnels ou académiques présentent des aspects positifs en termes de visibilité, de mise à jour des CV, de veille, d'échanges d'articles. Certains permettent de communiquer sur des publications de recherche et sur des expériences négatives ou infructueuses comme *FIGSHARE* (<https://figshare.com>). Cependant il ne faut pas oublier que ce sont des réseaux commerciaux dont la pérennité n'est pas garantie et qui sont axés uniquement sur les activités de publications. Le chercheur doit veiller au respect du droit d'auteur et des termes des contrats éditoriaux qu'il a signés.

d) Jeune chercheur doctorant : un parcours fléché et codifié depuis 2016

Depuis le 1^{er} septembre 2016, tout établissement habilité à délivrer le doctorat doit passer au dépôt électronique des thèses. Dans la plupart des cas, la thèse n'existe plus que sous sa forme électronique. Signaler une thèse soutenue est une obligation légale (arrêté du 25 Mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme nation-

nal de doctorat) (<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000032587086>).

La première démarche du doctorant consiste à prendre contact avec le coordinateur de thèses de l'Établissement d'Enseignement Supérieur (EES) dont il dépend.

Dans un deuxième temps, le doctorant doit vérifier l'originalité du sujet de sa thèse en utilisant le moteur de recherche des thèses françaises (<http://www.theses.fr>). Ce moteur de recherche, librement interrogeable sur le web, recense les thèses en préparation et soutenues. Le site de l'Agence Bibliographique de l'Enseignement Supérieur (ABES : <http://www.abes.fr>) renseigne sur les formalités à accomplir pour le dépôt de la thèse (<http://www.abes.fr/Theses/Pour-les-doctorants>). Le signalement du projet de recherche permet au doctorant d'obtenir une page web avec une URL valable pour la durée de sa thèse. Le moteur de recherche *Theses.fr* est aussi une interface qui est efficacement indexée par les moteurs de recherche du web. Cet outil numérique évolutif lui permet de faire figurer l'avancement de ses travaux et d'être éventuellement sollicité pour participer à des colloques en lien avec son sujet.

Dans un travail académique, le chercheur doit citer ses sources. En effet, si une information est accessible librement sur le web (extrait de texte, données chiffrées, schéma, images...) le chercheur doit mentionner explicitement sa source. L'omission de la source, volontaire ou non, constitue un acte de plagiat. Les travaux universitaires entachés de plagiat peuvent être sanctionnés et invalidés. Selon l'article L112-1 du Code de la propriété intellectuelle (CPI), toute représentation ou reproduction intégrale ou partielle effectuée sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque (Article Art. L 122-4 du CPI). Les universités se dotent de logiciels d'aide à la détection du plagiat comme *Compilatio* (<https://www.compilatio.net>) ou *Scribbr* (<https://www.scribbr.fr/logiciel-anti-plagiat/>), par exemple.

La troisième étape consiste à suivre les dispositions propres au dépôt sur support électronique : la feuille de style et la bibliographie. Il est important de fournir des formats archivables c'est à dire normalisés pour une publication large. Il est possible de tester la validité des fichiers en les soumettant au test FACILE mis en place par le Centre Informatique National de l'Enseignement Supérieur (CINES : <https://facile.cines.fr>). Ces dispositions sont propres à chaque établissement et doivent être remplies avec précision pour bien valoriser la thèse. En effet, plus un texte est référencé et visible sur internet, plus il est difficile de le plagier.

CONCLUSION

La révolution numérique ne se limite pas uniquement à un passage du papier à un support informatique. La quantité de données s'est considérablement accrue. Les modèles juridiques et économiques appliqués à la publication scientifique sont en mutation, notamment avec le développement de l'Open Access

et des Archives ouvertes. Le chercheur doit s'inscrire dans un environnement numérique foisonnant et en perpétuelle évolution, trouver des repères, des critères d'évaluation, adopter une démarche stratégique concernant aussi bien ses recherches documentaires que la construction de son identité numérique, la valorisation de ses travaux ou la protection au long terme des données de sa recherche.

Mais il n'est pas seul face à ces défis. Les organismes de recherche, comme le CNRS et son unité de services (l'INIST), l'Agence bibliographique de l'enseignement supérieur (ABES), les services à la recherche des universités et grandes écoles s'emploient à faciliter et accompagner le travail des chercheurs et acteurs de la recherche. Ils assurent une veille sur le contexte juridique, l'application des directives européennes, l'adaptation aux systèmes juridiques étrangers. Ils entretiennent des outils collaboratifs et créent des plateformes d'archivage. Sous la pression sociétale qui demande de plus en plus de transparence, les pouvoirs publics ont pris des mesures nationales pour réglementer la mise à disposition et la diffusion des données de la recherche financée par des fonds publics et à l'échelon européen. Des collectifs regroupent chercheurs et professionnels de l'édition pour la défense d'une science ouverte et de la « bibliodiversité ». Des professionnels de la documentation prennent en charge l'indexation, la gestion des

métadonnées et le référencement dans les moteurs de recherche, la diffusion, l'archivage électronique pérenne et la mise à disposition de cette masse de données, en tenant compte du droit de la propriété intellectuelle et des relations contractuelles avec les éditeurs. Tous ces partenaires mesurent l'importance de la formation des doctorants et chercheurs à l'Information scientifique et technique (IST). Ils multiplient les initiatives en ce sens (webtutoriels, séminaires, ateliers dans le cadre des écoles doctorales...). Le chercheur bénéficie aussi d'un environnement documentaire en ligne d'une richesse exceptionnelle. Les moteurs de recherche, catalogues, plateformes, sites web institutionnels ou commerciaux, qu'ils soient dans le web visible ou dans le web invisible, constituent pour le chercheur un atout considérable lui permettant de dresser un « état de l'art » le plus exhaustif possible : documenter et contextualiser sa recherche, varier les points de vue, identifier d'autres chercheurs liés à son domaine, effectuer une veille, évaluer des travaux ou des revues.

Le chercheur s'inscrit désormais au cœur de vastes réseaux professionnels et institutionnels dont les enjeux tournent autour de la diffusion, de la valorisation, de l'archivage, à l'échelon national et international. Son indetification numérique en tant que chercheur et le référencement de ses travaux constituent donc les portes d'entrée indispensables dans cet univers hyperconnecté.

REMERCIEMENTS

Les auteurs remercient Agnès Rosolen (Service Commun à la Documentation, Université Paris-Sud) pour son aide à la rédaction du manuscrit et sa relecture attentive.

BIBLIOGRAPHIE ET SITOOTHÈQUE PAR ORDRE D'APPARITION DANS LE TEXTE

Partie : Recherche et veille documentaire

- Dictionnaires encyclopédiques généraux en ligne : <http://www.larousse.fr/dictionnaires>
- Grand dictionnaire terminologique : <http://www.gdt.oqlf.gouv.qc.ca/>
- Terminologie interactive pour l'Europe : <http://iate.europa.eu/SearchByQueryLoad.do?method=load>
- Thesaurus de la National Library of Medicine : <http://www.ncbi.nlm.nih.gov/mesh>
- MeSH (traduction, site INSERM) : <http://mesh.inserm.fr/mesh/>
- TermSciences : référentiel commun à l'INIST, PASCAL, FRANCIS, MeSH, INRA, BDSP, etc... : <http://www.termosciences.fr/>
- Catalogue de la BnF : <http://catalogue.bnf.fr/recherche-sujets.do?pageRech=rsu>
- Library of Congress Subject headings : <http://id.loc.gov/authorities/subjects.html>
- PageRank, algorithme de Google : <http://www.pagerank.fr/>
- Google Scholar : <http://scholar.google.fr/schhp?hl=fr>
- Semantic Scholar : <https://www.semanticscholar.org/>
- Worldwide science : <https://worldwidescience.org>
- Science Open : <https://scienceopen.com/search>
- Microsoft Academic : <https://academic.microsoft.com>
- ISIDORE (plateforme) : <https://www.rechercheisidore.fr/>
- Centre Thématique de Recherche en Neurosciences de l'Université de Laval (<https://neuro.ulaval.ca>)
- SUDOC : <http://www.sudoc.abes.fr/>
- WorldCat : <http://0-www.worldcat.org.nova-cat.nova.edu/?lang=fr>
- Thèses : <http://www.theses.fr/>
- TEL : <http://tel.archives-ouvertes.fr/>
- HAL : <http://hal.archives-ouvertes.fr/>
- DART-Europe E-Theses Portal : <http://www.dart-europe.eu/basic-search.php>
- Bibliothèque des rapports publics : <http://www.ladocumentationfrancaise.fr/rapports-publics>
- LARA : <http://lara.inist.fr/>
- Open Grey : <http://www.opengrey.eu/>
- DOAJ (Directory of Open Access Journals) : <https://doaj.org>
- OpenDOAR (Directory of Open Access Repositories) : <http://www.opendoar.org>
- DOAB (Directory of Open Access Books) : <https://www.doabooks.org>
- Site du Bielefeld Academic Search Engine (BASE) : <https://www.base-search.net/?l=fr>
- Open Access Infrastructure for Research in Europe : <https://www.openaire.eu/search>
- Medline via l'interface PubMed : <http://www.ncbi.nlm.nih.gov/pubmed/>
- PubMed Central : <http://www.ncbi.nlm.nih.gov/pmc/>
- Public Library of Science (PloS) : <http://www.plosone.org/>
- Public Library of Science (blog) : <http://blogs.plos.org/evereveryone/>
- Cambridge University Press : <http://www.cambridge.org>
- Springer : <http://link.springer.com>

- Web of Science : <https://login.webofknowledge.com>
 - ISTEEX (add-on à installer sur son navigateur) : <https://addons.istex.fr>
 - Google Metrics : <https://scholar.google.com/intl/en/scholar/metrics.html>
 - Scholar H-Index Calculator for Google Chrome™ (add-on à installer) : <https://chrome.google.com/webstore/detail/scholar-h-index-calculator/cdpobfbhbdlpbllocjokjgkjmifbng?hl=fr>
 - Scopus : <https://www.scopus.com/home.uri>
 - Endnote : <https://www.endnote.com>
 - Zotero : <https://www.zotero.org>
 - Mendeley : <https://www.mendeley.com>
- Partie : Valorisation et diffusion des données de la recherche**
- Définition des données de la recherche (research data) de l'OCDE : <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=159&Lang=fr&Book=False>
 - Texte du livre blanc - une science ouverte dans une république numérique: <http://www.cnrs.fr/dist/z-outils/documents/livre-blanc-loi-numerique.pdf>
 - Site de l'université Paris-Descartes permettant de réaliser un DMP en français : <https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>
 - Site de l'Université de Sydney permettant de réaliser un DMP en anglais <https://library.sydney.edu.au/research/data-management/index.html>
 - Opidor (pour réaliser un DMP en ligne) : <https://dmp.opidor.fr>
 - Horizon 2020 : <http://www.horizon2020.gouv.fr>
 - SEANOE (Ifremer) : <https://data.ifremer.fr>
 - ZENODO : <https://zenodo.org>
 - Edinburg DataShare : <https://datashare.is.ed.ac.uk>
 - 4TU Datacentrum : <http://researchdata.4tu.nl/en/home/>
 - RE3 data Registry of data repositories : <https://www.eui.eu/Research/Library/ResearchGuides/Economics/Statistics/DataPortal/re3data>
 - FairSharing : <https://fairsharing.org>
 - Texte de la déclaration de Berlin (2003) sur le libre accès : <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>
 - Texte de l'initiative de Budapest (2002) sur le libre accès : <http://www.budapestopenaccessinitiative.org/read>
 - Site de l'INIST pour tout savoir sur l'open access : <http://openaccess.inist.fr/?Libre-Acces-des-outils-pour-vous>
 - Texte de l'appel de Jussieu (2017) sur le libre accès : <http://openaccess.inist.fr/?L-Appel-de-Jussieu>
 - Connaître les politiques des éditeurs sur l'open access – site sherpa/Romeo : <http://www.sherpa.ac.uk/romeo/>
 - ROAR (Registry of Open Access Repositories) : <https://roarmap.eprints.org>
 - OpenAIRE (Open Access Infrastructure for Research in Europe) : <https://www.openaire.eu/search>
 - ArXiv : <https://arxiv.org>
 - PubMedCentral : <https://www.ncbi.nlm.nih.gov/pmc/>
 - Creative Commons (en anglais) : <https://creativecommons.org>
 - Creative Commons (en français) : <http://creativecommons.fr>
- Partie : L'identité numérique du chercheur et la gestion de son e-image**
- Hypothèses : <https://cleo.openedition.org/openedition/hypotheses>
 - OpenEdition : <https://cleo.openedition.org>
 - CLEO (Centre pour L'édition Electronique Ouverte) : <https://cleo.openedition.org/presentation/dates>
 - The Conversation : <https://theconversation.com/fr>
 - FIGSHARE (expériences négatives) : <https://figshare.com>
 - Texte de l'arrêté du 25 Mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national de doctorat : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORF-TEXT000032587086>
 - Thèses de doctorat françaises : <http://www.theses.fr>
 - ABES (Agence Bibliographique de l'Enseignement Supérieur) : <http://www.abes.fr>
 - ABES ; aide aux Doctorants : <http://www.abes.fr/Theses/Pour-les-doctorants>
 - Compilatio (site antiplagiat) : <https://www.compilatio.net>
 - Scribbr (site antiplagiat) : <https://www.scribbr.fr/logiciel-anti-plagiat/>
 - CINES (Centre Informatique National de l'Enseignement Supérieur) pour effectuer le test FACILE : <https://facile.cines.fr>

GLOSSAIRE

- DOI : Digital Object Identifier – identifiant permettant l'identification d'une ressource numérique.
- Flux RSS (Rich Site Summary » ou « Really Simple Syndication) : famille de formats de données utilisés pour la syndication de contenu Web. Les Flux RSS sont des fichiers XML.
- ISBN : International Standard Book Number – numéro d'identification d'un livre.
- ISSN : International Standard Serial Number – numéro d'identification d'une publication en série (revue).
- ISTEEX : Initiative d'excellence en information Scientifique et Technique. Le projet ISTEEX a pour principal objectif d'offrir, à l'ensemble de la communauté de l'enseignement supérieur et de la recherche, un accès en ligne aux collections rétrospectives de la littérature. L'URL de téléchargement pour les navigateurs Mozilla Firefox ou Chrome est : <https://addons.istex.fr/>.
- Littérature grise : tout document dactylographié ou imprimé, souvent à caractère provisoire, reproduit et diffusé à un nombre d'exemplaires inférieur au millier, en dehors des circuits commerciaux de l'édition et de la diffusion.
- Métadonnées : Caractéristique formelle normalisée et structurée utilisée pour la description et le traitement des contenus des ressources numériques. Les métadonnées sont des données qui en décrivent d'autres. On distinguera des métadonnées de propriété, des métadonnées descriptives et des métadonnées d'enrichissement.
- Moteurs de recherche : outils soumettant une même requête à plusieurs moteurs de recherche.
- Moteur de recherche : application informatique permettant de rechercher une ressource.
- Notice bibliographique : fiche qui décrit un objet (document) manipulé dans une bibliothèque, traditionnelle ou numérique : livre, fascicule, article, etc. Elle comporte un ensemble d'indications permettant de l'identifier, de le localiser ou de le décrire sommairement.
- Opérateurs booléens : les opérateurs booléens (AND, OR, NOT / ET, OU, SAUF) précèdent un mot ou une expression pour limiter ou élargir une recherche. Par exemple : pomme ET poire pour limiter une recherche – pomme OU poire pour élargir une recherche – fruit SAUF poire pour une requête contenant le terme fruit mais pas le terme poire.
- ORCID (Open Researcher and Contributor ID) : identifiant ouvert pour chercheur et contributeur. C'est un code alpha-numérique propriétaire, qui permet d'identifier de manière unique les chercheurs et auteurs de contributions académiques et scientifiques.

- Plateforme numérique : tout lieu d'échanges ou se rencontrent l'offre et la demande de façon dématérialisée.
- PMID : Pubmed Identification – numéro de notice Pubmed.
- Thésaurus MeSH (Medical Subject Headings) : répertoire structuré de termes (mots-clés) pour l'analyse de contenu et le classement de documents servant de référence internationale dans le domaine des sciences biomédicales.
- Thésaurus RAMEAU (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié) : thésaurus géré par la Bibliothèque Nationale de France (BNF) renvoie aux équivalents MeSH du terme ainsi qu'au LCSH.
- Thésaurus LCSH (Library of Congress Subject Headings) : thésaurus de la Bibliothèque du Congrès à Washington.
- Troncature : une troncature est utilisée pour remplacer des lettres manquantes. Elle est généralement représentée par un astérisque. Par exemple : handicap* pour récupérer les documents contenant handicapS, handicapÉ, handicapÉS, handicapÉES, handicapANT, etc.
- Web visible : catalogues de bases de données consultables en libre accès.