

Au vu des enjeux actuels, les impératifs de science ouverte et de sobriété numérique gagneraient à s'articuler de manière plus étroite. Les professionnels de la documentation peuvent participer à la réflexion sur cet enjeu essentiel et porter des actions concrètes.

La science ouverte à l'épreuve de la sobriété

Opportunité d'ouvrir la science, la mutation numérique constitue aussi un défi pour l'environnement.

Le nouvel impératif de sobriété remet-il en question la promesse d'une « diffusion sans entrave »¹ portée par la science ouverte ? Cette tension sera questionnée dans cet article de réflexion². Le cas des données de la recherche sera en particulier traité, en envisageant diverses pistes d'actions.

UNE CROISSANCE EXPONENTIELLE À PRENDRE EN CONSIDÉRATION

La masse des données générées en ligne croît chaque année de façon exponentielle, avec une estimation à plus de 60 zettaoctets³ (Zo) pour 2020 et des prévisions autour de 600 Zo en 2030⁴. Cette tendance a des conséquences matérielles bien documentées : multiplication des *data centers*,

tailles comparées du fichier d'un article en climatologie et de l'ensemble des simulations informatiques qui l'alimentent. Les données de la recherche s'imposent donc comme un axe spécifique de réflexion. Durant leur cycle de vie, elles vont subir diverses transformations, de la donnée brute aux données traitées et analysées. Comment penser la conservation et le partage de ces différents états, y compris sur le long terme, le tri et l'élimination des versions intermédiaires ? Comment envisager une gestion des données de la recherche alliant durabilité et sobriété ?

AU-DELÀ DE L'OUVERTURE, LA NÉCESSITÉ D'UNE BONNE GESTION DES DONNÉES

Les plans de gestion de données (PGD) existent pour répondre à ces questionnements. En invitant les chercheurs à anticiper les étapes du cycle de vie de leurs données, ils les conduisent à adopter des pratiques rationnelles et possiblement compatibles avec l'exigence de sobriété numérique. Outils de coordination entre les membres d'un projet, ils permettent et organisent la concertation, documentent les redondances des jeux de données, afin de fluidifier les processus et penser la gestion des données sur toute la durée du projet. De la même manière, ils permettent de réfléchir et de justifier l'élimination de jeux de données devenus inutiles, une sélection qui peut être aidée par l'adoption des normes archivistiques⁵. Les PGD facilitent en outre l'adoption des principes FAIR⁶ qui garantissent une structuration optimale des jeux de données et leur pérennité. Sans l'usage de ces standards, les données partagées s'appauvrissent, car elles ne sont pas dotées des métadonnées permettant la compréhension de leur contexte de production, indispensable à leur réutilisation. Les PGD visent enfin à limiter la perte de données, par leur conservation pérenne, afin d'éviter leur recréation, dispendieuse en énergie. En diffusant ces bonnes pratiques, la science ouverte rencontre l'objectif de réduction de l'impact environnemental du numérique. Pour prolonger cette complémentarité, il serait intéressant que chaque discipline et structure de recherche réfléchisse à la définition de critères visant à identifier les données stratégiques de leur

Les données de la recherche, dont le poids dépassent largement celui des publications, s'imposent comme un axe spécifique de réflexion.

trafic accru sur les réseaux informatiques, etc. La recherche scientifique n'échappe pas à ce mouvement, auquel les attentes de la science ouverte peuvent contribuer. En déposant un article sur une archive ouverte, on se trouve en situation de dupliquer des informations déjà accessibles sur le site d'un éditeur. Les données de la recherche peuvent aussi être stockées simultanément dans des environnements différents. Il arrive que cette réplification des jeux de données soit volontaire (sauvegardes) mais elle peut aussi être accidentelle et liée à une complexité dans les processus de traitement. L'ouverture représente alors un risque supplémentaire de redondance.

Dans ces conditions, il est naturel de s'interroger sur le précepte d'ouverture promu par la science ouverte. Il convient d'abord d'introduire la notion d'ordre de grandeur, fondamentale sur les sujets d'impact environnemental. Même sans chiffres précis, on peut convenir que le poids des données de la recherche (et des infrastructures sur lesquelles elles s'appuient) dépasse largement celui des publications. Que l'on songe par exemple aux

[1] www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte

[2] Ce texte reflète la réflexion personnelle de leurs auteurs et n'engage aucunement leurs établissements respectifs.

[3] Un zettaoctet représente 1021 octets, soit un milliard de téraoctets.

[4] Didier Mallarino, Sylvie Le Bras, Cyrille Bonamy. « Les impacts environnementaux et sociétaux des données : un défi pour l'avenir ». Congrès JRES : Les Journées Réseaux de l'Enseignement et de la Recherche, RENATER, mai 2022, Marseille, France. <https://hal.science/hal-03702208v1>, p. 8.



écosystème. En invitant à une sélection exigeante et à une priorisation des données vers lesquelles doit tendre l'effort d'ouverture, ce travail permettrait de concentrer les opérations chronophages d'interopéabilisation et de pérennisation sur les jeux identifiés. Cette démarche pourrait par exemple s'appliquer à l'occasion du départ d'un chercheur : parmi ses données, lesquelles conservent un intérêt stratégique pour sa discipline, l'histoire de sa spécialité ? Étant entendu que « la valeur des données réside dans leur usage »⁷. De façon plus générale, les « PGD structure »⁸ pourraient être employés comme outils de pilotage de la gestion des données d'une unité de recherche. Leur élaboration et leur usage permettraient de mettre en œuvre une approche qualitative du partage et d'inscrire la réflexion sur l'archivage de la recherche dans le temps long de la communication savante.

VERS UN DÉSHERBAGE NUMÉRIQUE ?

Par ailleurs, les données partagées sur les entrepôts doivent faire l'objet d'un réexamen ponctuel : est-il nécessaire de les laisser en accès ouvert ? Doivent-elles faire l'objet d'une autre forme de préservation ? Les entrepôts, notamment aux États-Unis, qui ont une expérience de plus de dix ans dans le partage des données sont aujourd'hui confrontés à ce type de réflexion. Cela invite en outre à réfléchir sur l'opportunité d'un « désherbage » à appliquer aux données de la recherche, même si l'approche paraît contre-intuitive pour des documents numériques et pose des questions nouvelles. Quelles sont les durées de conservation requises ? Qui a la légitimité pour désherber ? Quels sont les critères de désher-

bage à appliquer (données FAIR, importance des publications liées, etc.) ? À défaut de répondre dès à présent à ces questionnements, une planification en deux temps de l'ouverture des données peut être envisagée et mise en œuvre à travers les PGD. Ainsi, les données les plus consultées resteraient accessibles directement dans des entrepôts (données chaudes) alors que les données faisant l'objet d'un usage résiduel ne seraient accessibles qu'à la demande et par conséquent stockées de façon plus économe en énergie (données froides)⁹. Cela renvoie à la répartition, bien connue pour la documentation physique, entre libre accès et magasin.

Il serait en outre intéressant de regarder ce qui est mis en place par d'autres organismes, à l'instar de ce que propose l'Afnor dans son guide d'écoconception des services numériques¹⁰, qui donne des pistes sur la mise en place d'une politique d'archivage, d'expiration et de suppression des données, en s'appuyant notamment sur la rédaction de politiques de gestion du cycle de vie des données et sur la réalisation régulière d'audits des données pour valider leur conservation (des audits qui, dans le cas des données de la recherche, pourraient être considérés au sein des PGD structure).

Au-delà des données elles-mêmes, la réflexion porte aussi sur les infrastructures de la connaissance, essentielles afin de se situer dans une approche globale de sobriété. Le choix de proposer un entrepôt de données national¹¹ se substituant au développement d'entrepôts multiples dans chaque établissement s'inscrit pleinement dans cette logique. Un pilotage

[5] Référentiel de gestion des archives de la recherche : https://doranum.fr/stockage-archivage/referentiel-de-gestion-des-archives-de-la-recherche_10_13143_pcqd-hy47

[6] Les principes FAIR ont été définis dans cet article de 2016 : Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. "The FAIR Guiding Principles for scientific data management and stewardship". *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

[7] BORGMAN, Christine L. 1. Provocations In : *Qu'est-ce que le travail scientifique des données ? Big data, little data, no data* [en ligne]. Marseille : OpenEdition Press, 2020 (généré le 4 février 2023). Disponible sur Internet : <http://books.openedition.org/oep/14727>. DOI : <https://doi.org/10.4000/books.oep.14727>.

[8] Pour une présentation des PGD structure, voir celle de Dominique L'Hostis et Sylvie Cocaud : https://urfirstinfo.hypotheses.org/files/2020/07/PGD-Structure_SC-DLH_GTDMP_25062020.pdf

[9] Cyril Pernet et al., On the long-term archiving of research data, <https://arxiv.org/abs/2301.01189>

[10] www.boutique.afnor.org/fr-fr/norme/afnor-spec-2201/ecoconception-des-services-numeriques/fa203506/323315, cf. Fiche 5.2.4 AFNOR Écoconception des services numériques, p. 24.

[11] <https://recherche.data.gouv.fr>

national des infrastructures permettant l'ouverture et la conservation à long terme des données de la recherche, s'il est bien conçu, les rendra plus solides face aux risques d'obsolescence technologique et de délitement de leur gouvernance.

Nombre d'actions pourraient être portées par les professionnels de la documentation afin de tendre vers l'objectif d'une science ouverte plus sobre. La sensibilisation à cette thématique est un premier levier évident, auprès des équipes de recherche, des directions et de l'ensemble des personnels. Elle peut, par exemple, infuser dans des présentations plus globales sur la science ouverte ou les données de la recherche, en s'inspirant de recommandations déjà émises par d'autres acteurs comme OpenDataFrance¹². Ces actions de sensibilisation pourraient aussi viser des publics étudiants, en s'appuyant sur la nécessité de former aux enjeux de la transition écologique dans l'enseignement supérieur¹³. Ce type d'initiatives serait particulièrement pertinent aux niveaux master et doctorat, auprès de jeunes chercheurs en devenir.

POUR UNE SCIENCE OUVERTE PLUS SOBRE

Le rôle des professionnels de la documentation dans l'accompagnement à la rédaction de PGD s'affirme progressivement. En conseillant les chercheurs, en orientant certaines pratiques de gestion des données, il est possible de favoriser des usages cohérents avec une démarche de sobriété numérique. La rédaction de PGD revêtirait alors un surcroît de sens, en particulier pour des jeunes chercheurs souvent sensibles aux enjeux climatiques. Il conviendra alors d'être en cohérence avec les initiatives prises par le monde académique et par les groupes déjà engagés dans une réflexion sur la réduction de l'empreinte des activités de recherche sur l'environnement, à l'image du collectif Labos1point5¹⁴.

Des supports de communication ou de formation pourraient également compléter les productions existantes afin d'enrichir l'accompagnement des chercheurs sur les volets de sobriété numérique, en insistant sur la priorisation de l'ouverture des données selon leur intérêt. La centralisation de ces actions faciliterait leur reproductibilité et leur adaptation au sein des différents établissements. Enfin, la profession pourrait réfléchir à l'intégration de volets spécifiques à la diffusion des résultats de la recherche dans les plans de sobriété énergétique des établissements ou les bilans carbone des unités de recherche, facilitant ainsi la communication entre services (notamment avec les directions de la recherche ou de l'informatique) et une prise en compte globale des enjeux.

Intégrer les impératifs de sobriété numérique dans la diffusion des productions scientifiques revient *in fine* à développer une approche qualitative, intégrée et raisonnée de leur gestion, des projets de recherche

jusqu'aux infrastructures. Il s'agit de promouvoir des usages plus frugaux, « dimensionnés en fonction du besoin et non des possibilités techniques »¹⁵.

LES MÉTIERS DE LA CONFORMITÉ EN SOUTIEN À LA RECHERCHE

Les professionnels de la documentation ont tout intérêt à s'emparer de ces questionnements, à l'image de ce qui est réalisé par la commission Bibliothèques vertes de l'Association des bibliothécaires de France¹⁶. Pour articuler les impératifs (plus interdépendants qu'opposés) de la science ouverte et de la sobriété numérique, de la protection des données personnelles et des droits d'auteur, ou encore de l'intégrité scientifique, les services de soutien à la recherche, notamment développés en bibliothèque, vont s'inscrire durablement dans les métiers de la conformité, adossés à des référentiels et des règlements dont le croisement nécessitera à la fois un fort développement d'expertises et des actions de sensibilisation et de formation toujours plus systématiques.

ANTHONY MOALIC

Responsable du département Appui à la recherche et Système d'information documentaire, SCD Université Caen Normandie
anthony.moalic@unicaen.fr

ÉLISE LEHOUX

Responsable du service à la recherche et adjointe au département Système d'information et accompagnement à la recherche, SCD Université Paris Nanterre
elehoux@parisnanterre.fr

CHRISTOPHE PION

Directeur de la bibliothèque de l'université Paris 8 Vincennes-Saint-Denis
christophe.pion@univ-paris8.fr

CHRISTOPHE LASNE

Chef de la cellule Infra et Système, Direction des Bibliothèques et Musées de l'université Paris Cité
christophe.lasne@u-paris.fr

[12] <https://opendatafrance.gitbook.io/greendata-pour-un-impact-maitrise-des-donnees/greendata/livrables-greendata>

[13] Voir Jean Jouzel, Luc Abadie, « Sensibiliser et former aux enjeux de la transition écologique dans l'Enseignement supérieur », ministère de l'Enseignement supérieur et de la Recherche, 2022.

www.enseignementsup-recherche.gouv.fr/fr/sensibiliser-et-former-aux-enjeux-de-la-transition-ecologique-dans-l-enseignement-superieur-83888

[14] <https://labos1point5.org>

[15] Fiche concept « Le numérique : concepts et définitions pour un numérique plus écologique », MOOC Impacts environnementaux du numérique (plateforme FUN).

[16] www.abf.asso.fr/4/210/981/ABF/bibliotheques-vertes

