
La visibilité des mémoires déposés sur DUMAS : une question de discipline

Matthieu Cisel et Nicolas Laudier



Édition électronique

URL : <https://journals.openedition.org/rfsic/16772>

DOI : 10.4000/13ift

ISSN : 2263-0856

Éditeur

Société Française de Sciences de l'Information et de la Communication

Ce document vous est fourni par Ecole nationale des sciences de l'information et des bibliothèques (ENSSIB)



Référence électronique

Matthieu Cisel et Nicolas Laudier, « La visibilité des mémoires déposés sur DUMAS : une question de discipline », *Revue française des sciences de l'information et de la communication* [En ligne], 29 | 2025, mis en ligne le 02 mai 2025, consulté le 14 mai 2025. URL : <http://journals.openedition.org/rfsic/16772> ; DOI : <https://doi.org/10.4000/13ift>

Ce document a été généré automatiquement le 2 mai 2025.



Le texte seul est utilisable sous licence CC BY-NC-SA 4.0. Les autres éléments (illustrations, fichiers annexes importés) susceptibles d'être soumis à des autorisations d'usage spécifiques.

La visibilité des mémoires déposés sur DUMAS : une question de discipline

Matthieu Cisel et Nicolas Laudier

Introduction

L'archivage en ligne de mémoires d'étudiants

- 1 Chaque année en France, des milliers de mémoires d'étudiants de Master réifient plusieurs mois de réflexions (Chauvin *et al.*, 2010 ; Gendre *et al.*, 2023) dans des textes qui ne contribuent globalement que modérément à l'avancée des connaissances scientifiques, faute d'être publiés dans des revues, ou même d'être recensés dans des archives scientifiques (Schöpfel et Stock, 2009). Or, dans le cas où le travail effectué est de qualité, ces écrits sont susceptibles d'apporter des éclairages intéressants sur leur sujet d'étude (Hawkins *et al.*, 2013), et la mise en ligne de leur travail permet aux étudiant.es de contribuer à leur échelle à un décroisement des travaux de recherche ; pour citer Gendre *et al.* (2023, p. 49) : « Valoriser ces travaux en archives ouvertes, c'est à la fois donner une forme de reconnaissance au travail de recherche effectué pendant au moins un an, et permettre à un large public d'en prendre connaissance, peut-être pour mieux en explorer certaines pistes ou établir des collaborations. C'est aussi et surtout une manière de participer au mouvement national de la science ouverte ».
- 2 Ces mémoires ont en tout cas suscité un intérêt suffisamment important pour que des entreprises anglo-saxonnes comme *Proquest* cherchent à en monétiser l'accès (Clement et Rascoe, 2013), rendant payant le téléchargement d'une grande partie d'entre eux. Dans les pays qui ont opté pour une approche moins mercantile de l'accès à ces documents, l'on trouve des répertoires offrant un accès gratuit et centralisé aux documents électroniques de mémoires d'étudiants et de thèses à différentes échelles,

une dynamique que certains auteurs voudraient obligatoire (Hawkins *et al.*, 2013). Dans certains cas, cet archivage en ligne se fait essentiellement à l'échelle d'une institution ou d'un groupement d'institutions, comme l'étude Ferrera-Fernandes (2013) pour l'Université de Salamanque en Espagne. Néanmoins, d'autres pays font le choix d'encourager la mutualisation des travaux d'étudiants à l'échelle nationale, comme c'est le cas en France avec le site theses.fr pour les thèses de doctorat (Martin, 2015 ; Schöpfel, Zendulkova, et Fatemi, 2014), et DUMAS pour les mémoires de niveau Bac+4 et au-delà (Gendre *et al.*, 2023).

- 3 DUMAS correspond aux initiales de Dépôt Universitaire de Mémoires Après Soutenance¹. Il ne s'agit pas d'une simple solution d'autoarchivage, dans la mesure où la mise en ligne passe par les documentalistes ou bibliothécaires de l'établissement. Par ailleurs, une note minimale est nécessaire, de façon à ce qu'un premier tri sur la qualité des manuscrits archivés soit effectué (Gendre *et al.*, 2023). Plus de 50 000 écrits avaient été déposés en 2023 : thèses d'exercice de médecine, mémoires de master, etc. – à travers une démarche centralisatrice ouverte à l'ensemble des établissements de l'enseignement supérieur français. On retrouve quelques écrits datant de 1960, mais ils relèvent de l'anecdote. L'essor de l'archive date de la fin des années 2000, lorsque le rythme annuel de dépôts passe de quelques centaines à plusieurs milliers de manuscrits. Des sites spécialisés ont par ailleurs été créés à côté de DUMAS pour certaines disciplines ; l'on peut par exemple mentionner la base TRHAA, pour Travaux de Recherche en Histoire de l'Art et en Archéologie², maintenue par l'Institut National de l'Histoire de l'Art. Il s'agit là d'une différence notable avec les thèses de doctorat, qui sont toutes centralisées en France via le site theses.fr (Martin, 2015).
- 4 Parmi les nombreuses questions de recherche que peut soulever l'émergence de telles archives, l'on peut se poser en premier lieu celle des différentes cultures disciplinaires eu égard au partage de telles travaux. Certains travaux ont par exemple documenté l'appétence plus prononcée chez les sciences naturelles que chez les sciences humaines et sociales eu égard à la publication en *open access* (Eger *et al.*, 2015 ; Look et Pinter, 2010). Néanmoins, pour pouvoir aborder cette problématique sur la base d'analyses de répertoires, encore faudrait-il disposer d'une vue d'ensemble de l'exhaustivité des archives utilisées, ce qui, au regard de la multiplicité des sites, constitue un défi de taille.
- 5 Le problème de la réappropriation des mémoires constitue également une question importante, qui fait écho à un ensemble de recherches sur l'impact sociétal de ce que les anglo-saxons nomment l'*open access*. Coexistent plusieurs interprétations concurrentes de la signification de ce syntagme, l'une reposant sur la « notion de réutilisation empruntée à la culture des logiciels libres et à la culture libre, et l'autre plutôt sur la notion d'accessibilité associée à la culture éditoriale et militant pour des répertoires ouverts » (Chartron et Schöpfel, 2017) ; c'est cette dernière acception que nous retenons ici. ElSabry (2017) a fait dans le milieu des années 2010 un point d'étape sur la cinquantaine de publications qui avaient alors été réalisées sur la réappropriation de recherches académiques en *open access* par différents profils, des politiciens aux chercheurs en passant par les étudiants. La thèse d'Albarin (2015) sur les répertoires latino-américains y est citée en exemple, de par son travail quantitatif d'identification du public, fondé sur un questionnaire.
- 6 Une lecture attentive de cette littérature montre que la question de la réappropriation n'est pas déclinée par discipline, alors que cette dimension peut être considérée comme

primordiale pour appréhender les dynamiques à l'œuvre. En effet, l'on peut supposer qu'un travail de master sur un sujet très précis de biologie moléculaire n'intéressera a priori que les spécialistes du domaine, là où un mémoire de médecine pourra éventuellement intéresser aussi bien le corps médical que des patients, ce qui n'est a priori pas sans conséquences sur la visibilité du travail réalisé. A défaut de pouvoir mener comme Albarin (2015) une enquête auprès des visiteurs de DUMAS, nous nous proposons de mobiliser les données de suivi des usages, récoltées en temps réel, pour visualiser les contrastes entre disciplines. Appréhender de telles différences de visibilité potentielle et effective représente selon nous un point essentiel si l'on veut saisir les enjeux associés au maintien et à l'expansion de l'archive. La première question qui structure ce travail est de nature descriptive, et sert d'ancrage empirique à notre propos : Quelles sont les disciplines de DUMAS pour lesquelles le nombre de téléchargements en ligne de manuscrits est le plus élevé ?

- 7 La seule analyse des données de téléchargements, même lorsqu'on les croise avec des métadonnées associées au mémoire, ne donne qu'un aperçu limité des mécanismes susceptibles d'expliquer les différences observées entre disciplines. Nous nous proposons dès lors d'aborder dans un second temps la question des conséquences possibles de la visibilité des mémoires. Elle est de nature plus théorique, au sens où elle s'appuie essentiellement sur une revue de la littérature et sur le cadre conceptuel que représente la théorie de l'activité d'Engeström (1987), plus que sur une collecte de données. Plusieurs auteurs ont travaillé sur le devenir des manuscrits de thèse juste après la soutenance et la perception que les étudiants en avaient (Ferrerias-Fernández *et al.*, 2016 ; Stanton et Liew, 2011). Ils ont montré que le choix de l'*open access* sur un répertoire comme theses.fr augmentait considérablement la visibilité de la recherche, notamment sur le plan bibliométrique, et constituait une incitation importante à partager leur travail.
- 8 Comme ces auteurs, nous partons du principe que la possibilité de voir son travail consulté et repris constitue a priori pour les étudiants de master une raison de mettre en libre accès leur travail sur DUMAS. Notre hypothèse est que l'on peut s'attendre à ce que certains individus sensibles à la question de la diffusion de leur mémoire, à plus forte raison s'ils ne se projettent pas dans une carrière académique, soient pris entre la volonté d'être reconnus au sein du milieu académique d'une part, ce qui signifie se plier aux normes afférentes, et la recherche de visibilité auprès du plus grand nombre d'autre part, au-delà du seul milieu académique. Cette considération nous amène à la question suivante : Quelles contradictions potentielles, au sens de Engeström (1987), peuvent-elles émerger de l'opportunité, pour les étudiants, de voir leur mémoire diffusé aussi bien au sein du milieu académique qu'en dehors de celui-ci ? Le terme *potentielles* utilisé ici vise à préciser le fait que l'on se concentre sur les conséquences possibles de certaines tendances que nous allons mettre au jour empiriquement, conséquences suggérées par la littérature scientifique et qu'il faudra démontrer lors de recherches ultérieures.

Revue de littérature et cadre théorique

Travaux sur les répertoires de mémoires et question des altmetrics

- 9 La littérature francophone comporte des travaux sur les archives ouvertes et leurs usages (Schöpfel et Prost, 2010, Schöpfel *et al.*, 2024). S'agissant de la réappropriation

des mémoires, les données bibliométriques comme le nombre de citations représentent selon toute logique un indicateur peu pertinent. En effet, ces écrits ne sont que rarement cités, que cela soit dans les articles de revues ou dans d'autres formes de productions scientifiques, et l'on ne peut dès lors reprendre directement pour les mémoires de DUMAS les méthodes adoptées dans étudier quantitativement le devenir d'autres types d'écrits. En effet, les manuscrits ne sont pas passés à travers le processus d'évaluation anonymisé qui caractérise revues et conférences scientifiques, de sorte qu'ils correspondent à une forme de littérature grise (Schöpfel, 2012), dès lors moins citée.

- 10 Néanmoins, un niveau faible de citations peut être contrebalancé par un niveau élevé de consultations. Les chiffres parlent d'eux-mêmes : depuis sa création, 10 456 821 visionnages et 64 092 295 téléchargements de mémoires avaient été recensés, au 23 décembre 2023, par les outils de suivi d'usages de DUMAS. Cela revient à plus de mille téléchargements en moyenne par mémoire – certains manuscrits totalisent à eux seuls plusieurs centaines de milliers de téléchargements, un volume considérable qui invite à se pencher sur les métriques alternatives, ou *altmetrics* dans la littérature anglo-saxonne (Haustein, 2016 ; Lin et Fenner, 2013).
- 11 Ces métriques visent à mesurer l'impact d'un travail scientifique au-delà d'indicateurs bibliométriques traditionnels. Shema *et al.* (2014, p.1018) définissent ce terme par rapport à la diffusion de l'article sur Internet : « *Altmetrics* [...] est un terme désignant les mesures basées sur le Web pour l'impact du matériel scientifique, en mettant l'accent sur les médias sociaux comme sources de données » (traduction libre). Bornmann (2014, p. 895), quant à lui, fournit un certain nombre d'indicateurs afférents, sans en désigner un de manière préférentielle, ce qui permet d'appréhender l'hétérogénéité des métriques qu'il est possible de retenir : « Les vues, les téléchargements, les clics, les notes, enregistrements, tweets, partages, likes, recommandations, balises, publications, hyperliens, discussions, signets et les commentaires sont comptés, plutôt que de simples citations d'un article dans une base de données telle que Scopus ».

La théorie de l'activité au service de l'analyse des contradictions

- 12 Nous nous proposons de combiner dans la présente recherche d'une part le cadre conceptuel proposé par Engeström (1987), et d'autre part des résultats issus de travaux s'inscrivant dans la continuité de la théorie de l'échange social (Homans, 1958 ; Wang et Noe, 2010). Du fait de son caractère généraliste, la théorie de l'activité gagne en effet à s'enrichir de notions issues d'autres champs scientifiques. Dans la continuité de Homans (1958), un certain nombre de chercheurs ont visé à identifier les leviers et les freins consubstantiels au partage de ressources, notamment au sein de répertoires électroniques (Bordia *et al.*, 2006).
- 13 Les travaux sur les leviers du partage s'intéressent en particulier à la distinction entre motivations intrinsèques et motivations extrinsèques, les études empiriques consacrées à la question débouchant sur des résultats contrastés (Wang *et al.*, 2005). D'un côté, certaines recherches soulignent l'impact positif des récompenses extrinsèques sur la volonté de partager ses connaissances, tandis que d'autres en pointent les limites (Kwok et Gao, 2005). Des auteurs comme Coleman (1990) et Wellman et Gulia (1999) montrent que ceux qui se considèrent comme des contributeurs au bien commun, sont

animés par une motivation intrinsèque que représente la participation à une dynamique collective. Celle-ci peut se suffire à elle-même, ou vient s'ajouter éventuellement à des facteurs extrinsèques comme les gains réputationnels.

- 14 Cette dichotomie entre motivations extrinsèques et intrinsèques a été également analysée dans le contexte de Wikipédia (Anthony *et al.*, 2005, 2009). Les auteurs montrent que si, au sein de l'encyclopédie, un compte associé à un pseudonyme peut gagner en réputation et en autorité de par son engagement, un certain nombre d'individus décident de participer de manière anonyme. Certains des principaux contributeurs ne sont dès lors identifiables que par leur adresse IP, ce qui souligne pour ces derniers l'importance de la motivation intrinsèque dans la dynamique de partage des connaissances. En substance, nous nous inscrivons dans la lignée de ces auteurs pour suggérer que ces deux types de motivation peuvent coexister dans le contexte de DUMAS, avec, à terme, des conséquences contrastées selon les disciplines en matière d'impact sur le comportement des étudiants, phénomène que nous allons aborder au prisme de la théorie de l'activité d'Engeström (1987).
- 15 Cette théorie représente un cadre interdisciplinaire utilisé dans des contextes contrastés, allant du champ de l'activité professionnelle au sein du milieu hospitalier aux dynamiques de partage de ressources pédagogiques au sein de l'enseignement supérieur (Engeström, 2001). En substance, l'activité est représentée par un modèle constitué de six pôles en interaction, qui constituent alors le système d'activité (SA) : le sujet de l'activité (S), son objet (O), qui englobe les visées de l'activité, les règles qui la régissent (R), la communauté (C), la division du travail (DT), et les instruments (I). Tout travail mobilisant ce cadre nécessite de délimiter le système d'activité d'intérêt sur la base de ces différents pôles, ce que nous ferons dans la section consacrée aux résultats.
- 16 Un système d'activité évolue notamment sous la pression de contradictions, porteuses de transformations. Avoir intégré ce concept de contradiction proposé par Ilyenkov (1974) constitue d'ailleurs l'une des caractéristiques saillantes de la version de la théorie de l'activité d'Engeström (Karanasios, 2018). La recherche de dépassement des tensions associées notamment à ces évolutions constitue l'un des mécanismes à l'origine de l'évolution de l'activité. Engeström (1987) distingue plusieurs niveaux au sein de ces contradictions. Les primaires correspondent aux tensions au sein d'un pôle donné. Par exemple, s'agissant du pôle *Objet*, plusieurs visées que le sujet attribue à l'activité peuvent entrer en tension. Les contradictions secondaires découlent des tensions entre deux pôles, tandis que les tertiaires correspondent aux tensions entre anciennes et nouvelles versions d'un système d'activité. Enfin, les contradictions quaternaires correspondent aux tensions entre systèmes d'activité distincts. Dans le cadre de la présente recherche, nous nous cantonnerons aux contradictions primaires et secondaires.

Méthodologie

Nature des métadonnées disponibles et méthode de collecte

- 17 Pour la grande majorité des manuscrits qu'elle recense, l'archive DUMAS fournit les métadonnées suivantes : auteur(s), titre du mémoire, type de document, date de soutenance ou de validation, type de mémoire (Master 2, de fin d'études médicales, etc.), mention (ex. : métiers de l'enseignement, de l'éducation et de la formation),

spécialité (ex. : premier degré pour les mémoires consacrés au système scolaire, par rapport au second degré), domaine(s), mots-clés, langue du document, licence et droits d'utilisation (par exemple, licence en Creative Commons), identifiant unique, nombre de pages, identité du directeur / de la directrice de mémoire ou du maître de stage, entre autres. Par ailleurs, deux éléments sont fournis de manière dynamique sur la page de présentation d'un manuscrit : le nombre de consultations de la notice, et le nombre de téléchargements du document. Les métadonnées disponibles ont été récoltées le 23 décembre 2023, pour la totalité des 52 787 manuscrits, via un algorithme de type web scraping codé en Python et fondé sur la librairie *Selenium*. En effet, l'outil d'export des données de l'archive, qui permet théoriquement de collecter sous la forme d'une table l'intégralité des données de la base, était dysfonctionnel. Python a été utilisé pour le scraping ; le traitement des données et l'analyse statistique ont été réalisés avec l'environnement logiciel R (version 4.2).

Croissance de l'archive Dumas dans le courant des années 2010 et choix des métriques

- 18 Pour identifier la thématique d'un manuscrit, la personne en charge de poster le document doit choisir parmi une liste de labels préexistants, sachant qu'elle peut en sélectionner plusieurs – notamment dans le cas d'un travail interdisciplinaire – et que l'on peut décider du niveau de précision de la caractérisation via une arborescence. Ainsi, l'on peut rester à la base de l'arborescence, en choisissant sciences humaines et sociales (SHS) – de niveau 0, ou choisir un label plus détaillé, comme SHS/Sciences de l'éducation – de niveau 1. La Table 1 présente l'importance relative des principaux champs de niveau 0, après recodage, et que nous regrouperons par la suite de manière sensiblement différente.

Table 1. Répartition des documents en fonction des principaux champs disciplinaires de DUMAS au 23 décembre 2023. N = 52 783

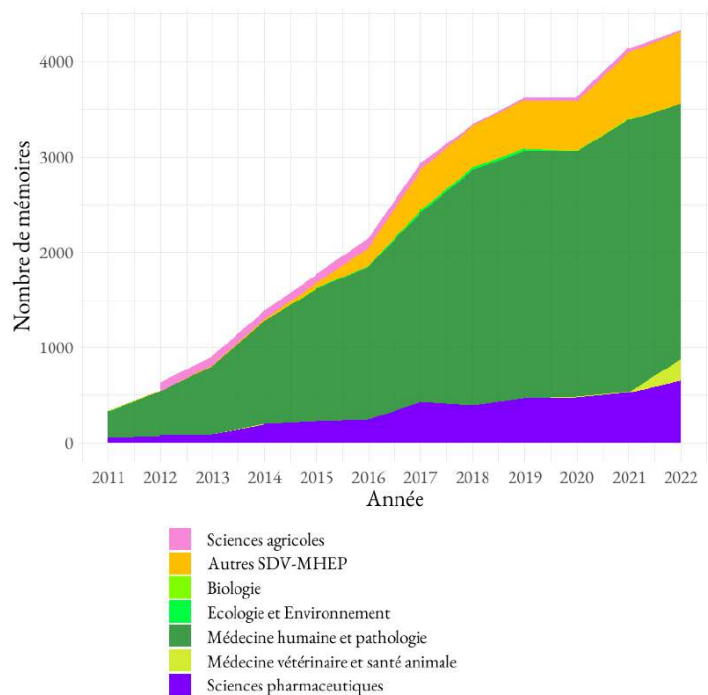
Domaine	Valeur (%)
Économie - Sciences de Gestion	1.7
Autres sciences dures (Biologie, Chimie, Environnement, etc.)	0.5
Physique - Sciences de l'Ingénieur	2.5
Informatique	1.1
Mathématiques Appliquées	0.3
Sciences de la Santé (Médecine humaine, vétérinaire, pharmacie)	54.2
SHS, Droit, Architecture	40.8

- 19 Au regard de la composition de l'archive, il est apparu pertinent de distinguer trois grands domaines et de regrouper les mémoires en conséquence : ceux relevant des sciences de la santé (pharmacie, médecine humaine et pathologie), ceux relevant des sciences dures – au sens de Biglan (1973) (biologie, chimie, informatique, mathématiques appliquées, etc.), et ceux relevant des SHS au sens large. Les disciplines relevant des SHS au sein de DUMAS sont précisées en figure 2.
- 20 L'analyse de l'évolution de la représentation des différentes disciplines au sein de l'archive constitue un préalable à celle des utilisations qui se développent autour des manuscrits. En effet, des métriques comme le nombre de visionnages et de téléchargements correspondent à l'activité cumulée des internautes depuis la mise à

disposition du document. Nous n'avons pas, pour un mémoire donné, d'informations sur l'évolution des visionnages et des téléchargements au fil du temps, mais seulement sur le cumul de ces actions jusqu'au 23 décembre 2023, date de l'extraction des données. Ainsi, toutes choses égales par ailleurs, un manuscrit posté en 2015 cumulera plus de vues ou de téléchargements qu'un manuscrit posté en 2022. Dès lors, si une discipline a fait son apparition sur l'archive peu avant l'extraction, ce qui est le cas par exemple de l'essentiel des travaux de la catégorie « Médecine vétérinaire et de la santé animale » (figure 1), des problèmes d'interprétation apparaissent potentiellement.

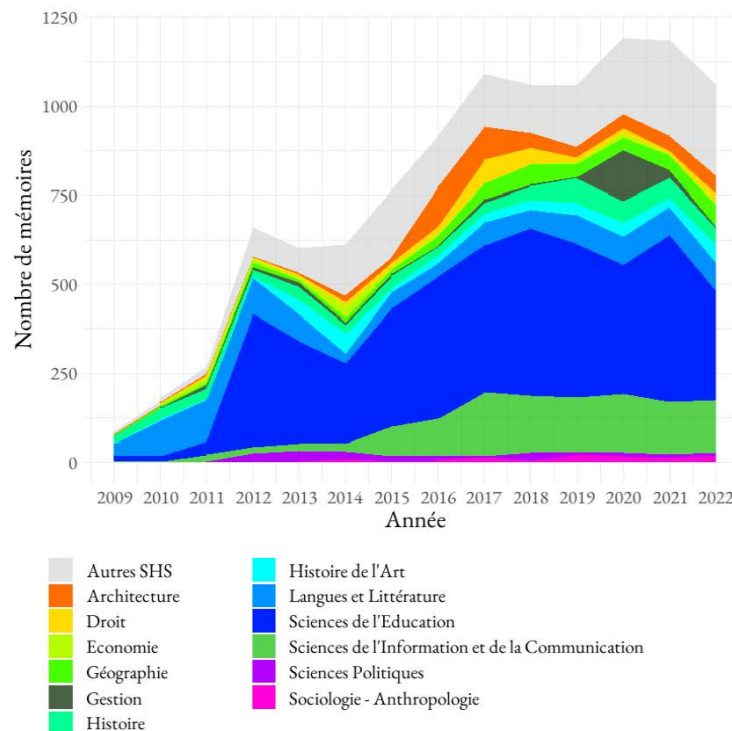
- 21 Un nombre faible d'utilisations pourrait à tort être attribué à un faible intérêt pour la discipline au sein de l'archive, alors que cela peut découler d'une fenêtre de disponibilité trop courte sur le site pour que les manuscrits de cette discipline aient accumulé de nombreux téléchargements. Les données confirment cette tendance. Les documents publiés en 2012 cumulent individuellement un nombre moyen de 746 ± 187 vues, contre 85 ± 140 vues pour ceux publiés en 2022, avec une chute à peu près linéaire du nombre moyen de vues entre ces deux dates. Une tendance similaire est observée pour les téléchargements. De manière générale, la comparaison des métriques d'utilisation entre disciplines doit se faire à durée de disponibilité équivalentes.
- 22 Pour corriger les biais liés au fait que les disciplines ne sont pas apparues de manière simultanée en quantités analogues sur la plateforme, nous avons employé une stratégie qui consiste à comparer le nombre de téléchargements à durée d'exposition équivalente, par exemple, après dix ans sur l'archive. Plus précisément, nous avons procédé à un calcul de rang, année par année, qui permet de déterminer, pour une année de production du mémoire donnée, quelle est la discipline pour laquelle le nombre de téléchargements par mémoire a été le plus important. Il convient de noter que nous avons retenu le nombre de téléchargements, et non le nombre de consultations de la notice depuis le site, ce dernier étant toujours inférieur au premier : le ratio entre les deux étant de 5.6 (± 10.0), en faveur des téléchargements. Ceci tient a priori au fait que les mémoires sont découverts et téléchargés après une recherche sur un moteur comme Google, sans passer par le site DUMAS.

Figure 1. Evolution du nombre de manuscrits disponibles sur Dumas dans les sous-disciplines des champs Sciences du Vivant / Médecine Humaine et Pathologie (SDV-MHEP)



- 23 Dans les deux figures présentées dans cette section, nous avons regroupé d'une part les disciplines relevant des sciences de la vie et de la santé (figure 1) et d'autre part celles relevant des sciences humaines et sociales (figure 2), qui constituent l'essentiel de la base de données. Dans le premier cas, l'on constate une très large prédominance des mémoires relevant des sciences de la santé, avec en premier lieu les thèses professionnelles de médecine humaine et pathologie, et dans une moindre mesure celles des étudiant.es de pharmacie. Les travaux de biologie (notamment de biologie végétale) ou de sciences agricoles existent, mais ils restent dans une large mesure anecdotiques sur le plan quantitatif.

Figure 2. Évolution du nombre de manuscrits disponibles sur Dumas dans les sous-disciplines catégorisées comme relevant, dans la base, des sciences humaines et sociales (SHS)



- 24 Dans le champ des sciences humaines et sociales (figure 2), les sciences de l'éducation dominent largement, tandis que les mémoires en sciences de l'information et de la communication ont pris leur essor à partir de 2014 pour se stabiliser à un peu moins de 200 manuscrits postés annuellement. Les trajectoires ne sont pas systématiquement ascendantes : plusieurs disciplines apparaissent sur l'archive avant de se raréfier par la suite. C'est le cas par exemple des sciences de gestion, qui connaissent un pic en 2020 avant de virtuellement disparaître les années suivantes. On notera qu'il existe des différences significatives entre disciplines en termes de nombre de page du manuscrit ou de la propension à écrire en anglais (figures 5 et 6, Annexes). En sciences de l'univers, le mémoire dépasse en moyenne rarement plus de la cinquantaine de pages, tandis qu'il est d'environ 170 pages pour un écrit en histoire ou histoire de l'art. Avec 56 % de manuscrits en anglais, les travaux en économie sont les seuls à être rédigés majoritairement dans cette langue.

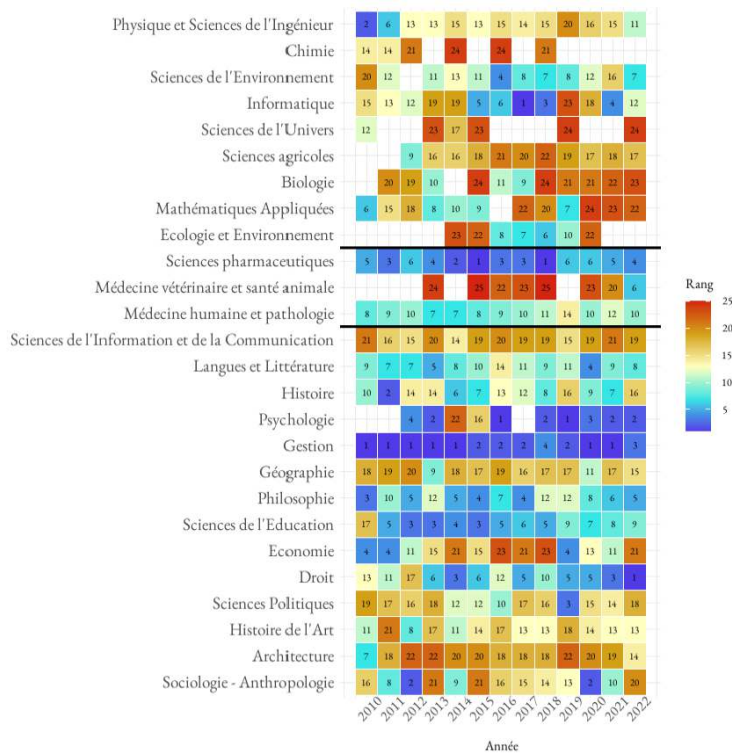
Résultats

Analyse quantitative des téléchargements

- 25 En nous focalisant sur le nombre de téléchargements moyen des mémoires, nous avons constaté que pour certaines disciplines comme les sciences de gestion concentraient, le nombre de téléchargement par mémoire était bien au-dessus de la moyenne des autres disciplines, et ce de manière constante au fil des années (figure 3). Les mémoires de cette thématique, pourtant très peu représentée au sein de la base, était ainsi les plus téléchargés sur huit des treize années rapportées sur la figure, le rang moyen étant de 1,7. Alors qu'ils ne représentent que 1,3 % de l'archive, ils sont responsables de 5,1 %

des téléchargements. Les résultats sont comparables lorsque l'on se base sur le nombre de téléchargement médian, pour éviter les biais liés aux valeurs extrêmes.

Figure 3. Évolution du rang de vingt-cinq disciplines en matière de nombre de téléchargements des mémoires de DUMAS, ce rang étant calculé pour une année de production de mémoire donnée



- 26 Les écrits de la catégorie *Médecine et pathologie humaine* sont certes responsables de 36,1 % des téléchargements à l'échelle de l'archive, mais il en constituent 44,0 % des documents. Il n'est dès lors pas surprenant de constater qu'ils oscillent entre le septième et le quatorzième rang lorsque l'on se focalise sur le nombre moyen de téléchargements par mémoire. Autre cas intéressant, celui de la pharmacie, qui, alors qu'elle ne constitue que 9,5 % des mémoires, correspond à 14,6 % des téléchargements, primauté que l'on peut observer en figure 3. Nous y avons distingué les trois grands domaines que sont les matières correspondant aux sciences dures, aux sciences de la santé (humaine ou animale), et aux sciences humaines et sociales et considérées comme telles dans la taxonomie de DUMAS.
- 27 Un test non-paramétrique de Friedman réalisé sur les rangs présentés dans ce graphe confirme le caractère statistiquement significatif des différences entre disciplines en matière de téléchargements ($X^2 = 186.1$, ddl = 25, p-value < 0.001). Ce test statistique est adapté au caractère répété des mesures. La relative stabilité des rangs, dont notamment la place prééminente des *sciences de gestion*, est moins affirmée lorsque l'on se focalise non plus sur le nombre moyen de téléchargements, mais sur le nombre médian ; le test de Friedman donne néanmoins des résultats relativement similaires ($X^2 = 169.1$, ddl = 25, p-value < 0.001). La différence entre rang moyen et rang médian tient à l'existence de valeurs extrêmes très nombreuses, en particulier pour les mémoires relevant des sciences de gestion. Cette considération nous a amené à rechercher les écrits ayant cumulé au fil des années un nombre important de téléchargements, afin d'approcher la question sous un angle plus qualitatif. En choisissant arbitrairement le

seuil de 100 000, neuf manuscrits issus de disciplines aussi distinctes que les sciences agricoles ou les sciences de gestion répondent aux critères (table 2).

Table 2. Titre et discipline des mémoires ayant cumulé plus de 100 000 téléchargements sur l'archive DUMAS à date du 23 décembre 2023.

Titre	Discipline	Téléchargements
Les troubles urinaires et vésicaux induits par les médicaments	Sciences pharmaceutiques	209953
Mise en place d'un plan de maintenance préventive sur un site de production pharmaceutique	Sciences pharmaceutiques	111607
L'enseignant face aux élèves en difficulté	Sciences de l'Education	111009
Comment optimiser l'organisation et le suivi de chantier dans une entreprise ayant un rayon d'action étendu ?	Sciences agricoles	108186
Les métrorragies du premier trimestre de la grossesse : complications obstétricales et néonatales	Médecine humaine et pathologie	126487
The 1950's and the 1960's and the American Woman : the transition from the "housewife" to the feminist	Histoire	283925
La motivation des salariés et la performance dans les entreprises	Gestion	402890
La fidélisation des salariés : le cas de l'entreprise X (spécialisée dans le bâtiment)	Gestion	126065
Les enjeux stratégiques de l'intégration des nouveaux collaborateurs	Gestion	106685

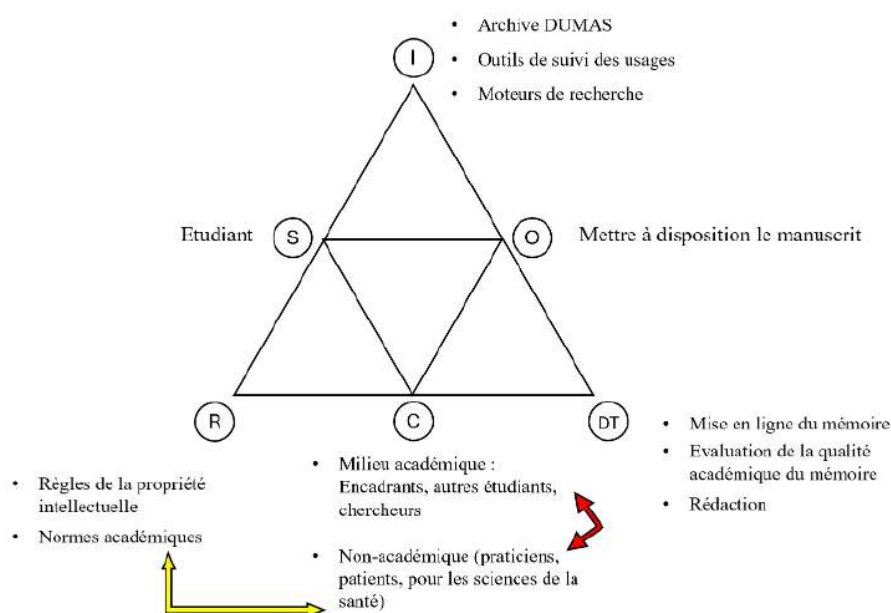
- 28 Sur ces neuf mémoires à succès, trois relèvent des sciences de gestion et portent, si l'on se cantonne au titre, sur des questions plutôt générales susceptibles d'intéresser des individus en charge du bon fonctionnement d'une entreprise : intégration, fidélisation et motivation des salariés ou collaborateurs, selon la terminologie utilisée. De manière intéressante, plusieurs autres mémoires classés dans d'autres disciplines (sciences agricoles ou pharmaceutiques) semblent également relever de questions de gestion, avec par exemple un mémoire de sciences pharmaceutiques centré sur la production de médicaments. Le mémoire d'histoire, focalisé sur les femmes au foyer, fait office d'exception à cet égard. La possibilité grâce au dépôt de mémoires d'atteindre aussi bien un public académique que non-académique est source de tensions potentielles que nous nous proposons de modéliser maintenant au prisme de la théorie d'Engeström (1987), la première étape consistant à délimiter le système d'activité en précisant la nature de ses pôles (figure 4).

Une modélisation des contradictions

- 29 Dans la modélisation que nous proposons, le *Sujet* de l'activité est l'étudiant. Le pôle *Communauté* correspond tant aux enseignants avec qui il est en interaction au sein du milieu académique (encadrants, évaluateurs, etc.), qu'à l'audience de l'archive et des mémoires qu'elle héberge. Le pôle *Instruments* inclut l'ensemble des outils permettant d'assurer la visibilité et l'accès aux mémoires des étudiants, avec d'une part l'archive DUMAS et ses outils de suivi de trafic, et d'autre part les moteurs de recherche comme

Google, qui permettent d'accéder directement au PDF d'un mémoire sans avoir à naviguer via l'archive. Le pôle *Règles* inclut tant les normes qui régissent la rédaction d'un mémoire académique – du choix de la thématique à la formulation du titre, en passant par les considérations stylistiques. S'agissant de la *Division du Travail*, l'étudiant est responsable de la qualité de l'écrit, les documentalistes et bibliothécaires se chargent de la mise en ligne, tandis que les encadrants se chargent du respect des normes de l'écriture académique et de son évaluation. Enfin, le pôle *Objet* correspond aux visées recherchées. Dans la mesure où nous n'avons pas de données empiriques sur les motivations individuelles, nous nous cantonnerons à décrire l'objectif comme étant le fait de « mettre à disposition le manuscrit ».

Figure 4. Modélisation du système d'activité et des contradictions qui le traversent sur la base du modèle d'Engeström (1987)



- 30 S'agissant des tensions qui traversent le système d'activité figure 4), nous proposons de les représenter sous la forme d'une contradiction primaire au sein du pôle *Communauté* (flèche rouge), qui peut évoluer dans certains cas en une contradiction secondaire entre le pôle *Communauté* et le pôle *Règles*. La contradiction primaire découle, dans notre modélisation, de l'existence de deux audiences potentielles : les individus qui évoluent dans le milieu académique (chercheurs, encadrants, étudiants), et ceux qui évoluent en dehors de celui-ci (professionnels exerçant dans des entreprises, médecins, patients, etc.). Nous considérons qu'il y a ici une contradiction, car ces différents profils ne vont pas chercher dans les mémoires, *a priori*, les mêmes types d'information, entre autres.
- 31 Ainsi, par exemple, là où un étudiant ou un chercheur pourra être intéressé par les références bibliographiques ou par le développement d'un cadre théorique, en plus des résultats, un salarié d'une organisation ou un praticien de santé se concentrera sans doute davantage sur des résultats saillants en lien avec une situation concrète rencontrée dans le cadre de son activité professionnelle, ou un patient une information en lien avec sa maladie. Concrètement, cela peut se traduire par le fait que le public venant d'un milieu non académique estime que des parties importantes du mémoire ne sont pas en adéquation avec ses besoins. Cette tension reste une contradiction primaire

tant que l'existence de ce public potentiel n'a pas d'influence sur la réalisation du travail par l'étudiant.

- 32 Nous proposons de modéliser comme contradiction secondaire (flèche jaune) la tension qui apparaît du fait de l'exposition d'un public non-académique à un écrit qui, lui, doit respecter les normes académiques, et de la prise en compte de cette tension par l'étudiant. Le caractère plus ou moins marqué de cette contradiction dépend selon nous de la motivation qui anime ce dernier au moment de la mise en ligne de son travail, et, dans une certaine mesure, de la thématique dans il traite dans son travail. La possibilité d'acquérir de la visibilité à grande échelle, semble-t-il plus marquée pour des disciplines comme les sciences de gestion, pourrait alors devenir pour certains étudiants motivés par des gains réputationnels une incitation à concevoir leur mémoire de sorte à être connu en dehors du seul milieu académique. Se pose alors la question de la tension que fait naître une telle situation : une thématique peut n'intéresser que le seul milieu universitaire. Le nécessaire respect des normes de l'écriture académique, qui permet une reconnaissance au sein du monde universitaire, est susceptible de nuire à la visibilité en dehors de celui-ci, considérations que nous nous proposons d'approfondir en discussion.

Discussion

- 33 L'analyse de la base de données nous a permis de mettre en évidence une grande visibilité relative de certaines disciplines pourtant sous-représentées dans l'archive. Tant les titres des manuscrits les plus plébiscités que le fait que les téléchargements dominant sur les consultations suggèrent qu'une audience non-académique passant par un moteur comme Google est à l'origine du succès des sciences de gestion, par exemple. Ces résultats embryonnaires suggèrent que les travaux les plus plébiscités sont vraisemblablement consultés par de nombreux professionnels s'inscrivant dans une logique d'apprentissage auto-dirigé, selon une logique que des auteurs comme Philippe Carré (2001, p. 22) qualifieraient d'opératoire professionnelle : « il s'agit d'acquérir des compétences (connaissances, habiletés, comportements) perçues comme nécessaires à la réalisation d'activités spécifiques sur le champ du travail, afin d'anticiper ou de s'adapter à des changements techniques, de découvrir ou de perfectionner des pratiques, toujours avec un objectif de performance précis ».
- 34 Une analyse rigoureuse des causes de cette logique opératoire professionnelle imposerait une enquête dédiée. Aussi préférons-nous nous pencher sur les conséquences possibles de ce phénomène. Elles sont multiples, et nous nous proposons d'explorer certaines d'entre elles au prisme de la théorie de l'activité, en distinguant plusieurs configurations correspondant à des motivations distinctes. L'existence de ces divers cas de figure est suggérée par les travaux antérieurs relevant de la théorie de l'échange social (Homans, 1958) ainsi que par ceux réalisés sur la publication en *open access* des thèses de doctorat (Ferrera-Fernandes *et al.*, 2016 ; Stanton et Liew, 2011). C'est dès lors sous l'angle de la théorie que nous abordons nos résultats, en nous intéressant aux aboutissants de la situation, plus qu'à ses tenants.
- 35 La focale principale de cette discussion est la contradiction secondaire que nous avons modélisée entre le pôle *Règles* et le pôle *Communauté*. Premier cas de figure, la possibilité de voir son mémoire diffusé auprès d'une audience plus large que celle du milieu académique n'affecte en aucune manière la conception du manuscrit, que l'on

parle du choix de la thématique, de la formulation du titre, du style adopté, ou de la structuration du contenu. La contradiction entre ces deux pôles du système d'activité est alors peu marquée. L'étudiant qui met son travail en ligne peut être animé par une volonté désintéressée de partage (Coleman, 1990 ; Wellman et Gulia, 1999), analogue à celle identifiée par Anthony *et al.* (2009) chez les contributeurs anonymes de Wikipedia. Sa démarche correspond alors en premier lieu à la participation, à son échelle, au mouvement de science ouverte (Gendre *et al.*, 2023).

- 36 Satisfaire aux exigences du milieu académique constitue ici la principale focale, puisqu'elle conditionne la note obtenue, et les éventuels gains réputationnels recherchés par une diffusion de l'écrit au sein du milieu académique, s'ils sont envisagés, ne constituent qu'un bénéfice secondaire. C'est ainsi que nous pourrions interpréter le cas du mémoire sur le féminisme aux États-Unis dans les années 1950 et 1960. L'étudiant n'a d'ailleurs pas nécessairement conscience de la visibilité potentielle éventuellement acquise par son travail au fil des ans. Dans ce cas de figure, la tension entre les pôles du système d'activité reste limitée, et ce sont avant tout des intérêts personnels, ou des interactions avec les encadrants, des contraintes fixées par ces derniers, qui déterminent le choix du sujet de mémoire et imposent la formulation du titre et le style rédactionnel. Pour un travail en biologie moléculaire par exemple, l'étudiant aura sans doute moins de liberté dans l'orientation qu'il impulse à son travail, car il s'inscrira plus vraisemblablement dans un travail de recherche initié par son laboratoire, à plus forte raison si les activités développées coûteuses.
- 37 Deuxième cas de figure, l'étudiant est sensible au gain réputationnel potentiel dans le milieu professionnel, par exemple parce qu'il n'envisage pas de rester au sein du monde universitaire à l'issue de son master. On s'éloigne alors de la logique de partage désintéressé que Gendre *et al.* (2023) décrit. L'on peut alors distinguer plusieurs situations, qui correspondent alors à une contradiction secondaire plus ou moins marquée entre les deux pôles du système d'activité. Soit l'individu donne malgré tout, la priorité aux contraintes imposées par le milieu académique, le temps du mémoire, et la réception du mémoire hors du milieu universitaire n'influence pas le choix du sujet traité, ou la formulation du titre par exemple. Soit l'étudiant dispose d'une certaine agentivité, et le mémoire peut alors être conçu de sorte à choisir une thématique valorisée en dehors du milieu universitaire, et dont l'approfondissement peut lui conférer un avantage sur le plan professionnel après ses études. Le nombre de téléchargements peut être un bénéfice explicitement recherché au moment de la mise en ligne, au même titre qu'une compétence et une légitimité sur une thématique jugée porteuse par l'étudiant, sans qu'il soit possible à ce stade de distinguer l'effet sur son comportement des deux formes de gains réputationnels potentiels.
- 38 Cette dernière configuration explique aisément le choix qui a été fait, parmi les auteurs des mémoires les plus lus, de s'orienter vers des thématiques généralistes peu originales sur le plan de la recherche. En sciences de gestion, nous avons relevé des titres qui peuvent correspondre à des recherches fréquentes dans un moteur comme Google : la motivation, la performance ou la fidélisation des salariés. Ces thématiques sont susceptibles de concerner tous les secteurs d'activité, et donc de déboucher sur un nombre élevé de téléchargements en dehors du seul milieu académique. L'intensité de la contradiction avec le pôle *Règles* dépend alors des exigences des encadrants et membres du jury de master. La situation est différente si ces derniers attendent un travail original sur le plan de la recherche – ce qui exclut par exemple d'aborder de

manière superficielle un sujet comme les leviers de motivation dans l'activité professionnelle, pour reprendre l'exemple d'un mémoire à succès – et si l'étudiant est en droit de produire un travail peu novateur sur le plan académique.

- 39 Le premier cas, où les contraintes de l'activité de recherche prédominent, serait par exemple celui d'un travail de biologie moléculaire. L'étudiant ne dispose pas, *a priori*, alors de l'agentivité permettant de formuler un titre susceptible d'intéresser en dehors du milieu universitaire. Le second cas pourrait être d'un mémoire professionnel hors master de recherche, où une réflexion praxéologique sur un sujet comme « L'enseignant face aux élèves en difficulté », pour reprendre notre exemple, pourrait être considéré comme bénéfique pour l'étudiant à défaut de contribuer à l'avancée des connaissances scientifiques, tant le sujet a déjà été traité. Libéré de la contrainte que constitue l'originalité du mémoire sur le plan scientifique, l'individu peut alors s'orienter vers le sujet de son choix, et donc sélectionner une thématique d'actualité à même de correspondre à ses intérêts et de lui apporter et de la visibilité par la suite, sur Internet ou pour sa carrière, si du moins le gain réputationnel constitue pour lui un levier de motivation. La contradiction s'accroît en revanche, si l'individu évolue dans un milieu universitaire dont les attentes sont incompatibles avec une telle orientation du mémoire, ce qui le place dans une situation de double contrainte caractéristique des contradictions que décrit Engeström (1987).
- 40 A travers notre argumentaire, l'on voit se dessiner peu à peu le faisceau de contraintes associées aux différents pôles du système d'activité, et susceptibles de peser sur les décisions individuelles, que cela soit la mise en ligne du mémoire, mais aussi le choix de la thématique, ou la formulation du titre, sujets qui mériteraient des recherches complémentaires : exigences des jurys du master, contraintes liées au fonctionnement du laboratoire d'accueil, dans le cas des stages de recherche, stratégies à adopter pour faciliter l'insertion professionnelle à la suite du master, visibilité potentielle du manuscrit. Selon les disciplines et selon la nature des formations suivies, ces différentes contraintes peuvent tantôt fonctionner en synergie, tantôt s'opposer et mettre en porte-à-faux l'étudiant. Celui-ci peut avoir à choisir entre des stratégies qui le valorisent au sein du milieu universitaire, et celles qui le valorisent en dehors de celui-ci : c'est la contradiction que nous avons pointée entre les pôles *Communauté* et *Règles*.
- 41 La manière dont les individus prennent en considération ces différentes contraintes pour arrêter leurs choix, en particulier sur la question de la mise en ligne, est vraisemblablement dictée par la façon de se positionner quant aux pratiques de partage : la recherche de bénéfices individuels, sous la forme notamment de la récompense symbolique que constitue un nombre élevé de téléchargements, source de tensions potentielles, ou la volonté de s'appropriier les valeurs d'un partage désintéressé (Anthony *et al.*, 2009 ; Coleman, 1990 ; Wellman et Gulia, 1999). Il a été démontré que le rapport à la publication d'articles en *open access*, et à la science ouverte en général, différerait d'une discipline à l'autre (Eger *et al.*, 2015 ; Look et Pinter, 2010). Si l'on admet que l'existence de cultures disciplinaires dépasse le seul cas des chercheurs confirmés, il est vraisemblable que de tels contrastes existent alors au niveau d'étudiants de master, jouant sur la nature de la motivation qui peut les pousser à mettre en ligne leur mémoire, et donc sur les contradictions afférentes. Les résultats de Eger *et al.* (2015) ne font néanmoins que pointer l'importance qu'il y a à étudier de telles

différences, mais ne permettent pas de prédire aisément la manière dont elles pourraient se traduire pour des étudiants de master en France.

- 42 Pour reprendre une idée centrale au travail d'Ilyenkov (1974), reprise par Engeström (1987), les contradictions sont porteuses de transformations du système d'activité. Reste à donc à identifier les éventuelles évolutions que pourrait engendrer la contradiction secondaire que nous avons modélisée. Celles que nous envisageons reposent sur la prise de conscience, par les différents acteurs en présence, des possibilités de diffusion considérables apportées par DUMAS, prise de conscience qui n'est pas encore prouvée empiriquement. L'on peut imaginer que les institutions académiques, chiffres à l'appui, mettent en avant l'intérêt que présente pour l'étudiant la mise en ligne de son travail sur l'archive, comme valorisation d'un travail académique, visibilité dont peut même bénéficier indirectement l'établissement où est réalisé le mémoire.
- 43 À l'inverse, des encadrants pourraient redouter que cette opportunité ne pousse des étudiants à sacrifier certaines des normes de l'écriture académique, ou à choisir des thématiques d'actualité mais présentant peu d'intérêt du point de vue de la recherche, pour accroître la portée de leur travail auprès d'un public non universitaire. Une telle crainte pourrait alors déboucher sur l'introduction de nouvelles règles et formes d'interventions réalisées durant la formation, dont la fonction serait de découpler l'exercice d'écriture du manuscrit des éventuelles retombées associées à sa mise en ligne. Un approfondissement de la recherche est nécessaire pour prédire les évolutions que pourrait impulser la visibilité importante de l'archive, notamment sur un moteur de recherche comme Google.

Conclusion

- 44 Ce travail souffre des limites consubstantielles de toute recherche visant à appréhender un phénomène sur la base des seules approches quantitatives. Si la mise au jour de régularités empiriques constitue une étape essentielle pour appréhender les dynamiques de partage de productions scientifiques, il manque à cette enquête un volet qualitatif qui viendrait compléter le travail que nous avons présenté ici. Des entretiens semi-directifs seraient nécessaires pour capturer la perspective des étudiants quant à l'effet incitatif de la visibilité apportée par l'archive sur les pratiques de partage. En effet, nous ignorons dans quelle mesure les individus sont conscients de l'apport potentiel de DUMAS en termes de visibilité, de l'intérêt que les internautes portent aux mémoires de certaines disciplines, et dans quelle mesure une éventuelle connaissance de ces dynamiques est susceptible d'influer leur décision de déposer leur travail au sein de l'archive. En définitive, un certain nombre de décisions reposent sur un bénéfice perçu, or nos données apportent uniquement des informations sur les bénéfices effectifs en termes de visibilité. La prolongation de cette enquête par la mise en place de méthodes mixtes semble s'imposer.
- 45 En guise de perspective, nous proposons de poursuivre la réflexion sur la diffusion des mémoires, en nous centrant cette fois sur l'audience du site. Il pourrait être intéressant de dupliquer la méthodologie mise en place par Alperin (2015) pour collecter des données sur la profession des visiteurs et la fonction de leur visite, mais en s'attachant, comme dans la présente recherche, à contraster la situation selon les disciplines. Une telle démarche permettrait d'apporter des éclairages intéressants pour les étudiants

qui envisagent de mettre à disposition librement le fruit de leur travail, en leur fournissant davantage d'informations quant à la portée de la démarche.

BIBLIOGRAPHIE

- ALPERIN, J.-P. (2015). *The public impact of Latin America's approach to open access*. Stanford University. <https://purl.stanford.edu/jr256tk1194>
- ANTHONY, D., SMITH, S. W., et WILLIAMSON, T. (2005). *Explaining quality in internet collective goods : Zealots and good samaritans in the case of wikipedia*. Dartmouth College.
- ANTHONY, D., SMITH, S. W., et WILLIAMSON, T. (2009). Reputation and reliability in collective goods : The case of the online encyclopedia Wikipedia. *Rationality and society*, 21(3), 283-306. <https://doi.org/10.1177/1043463109336804>
- BIGLAN, A. (1973). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, 57(3), 204-213. <https://doi.org/10.1037/h0034699>
- BORDIA, P., IRMER, B. E., et ABUSAH, D. (2006). Differences in sharing knowledge interpersonally and via databases : The role of evaluation apprehension and perceived benefits. *European journal of work and organizational psychology*, 15(3), 262-280. <https://doi.org/10.1080/13594320500417784>
- BORNMANN, L. (2014). Do altmetrics point to the broader impact of research ? An overview of benefits and disadvantages of altmetrics. *Journal of informetrics*, 8(4), 895-903. <https://doi.org/10.1016/j.joi.2014.09.005>
- CARRE, P. (2002). *De la motivation à la formation*. L'Harmattan
- CHARTRON, G., et SCHÖPFEL, J. (2017). Open access et Open science en débat. *Revue française des sciences de l'information et de la communication*, 11, Article 11. <http://journals.openedition.org/rfsic/3331>
- CHAUVIN, S., GALLEZOT, G., et SCHÖPFEL, J. (2010). *Les mémoires de master dans les archives ouvertes*, Rapport de recherche, Université Lille 3, 49 pages https://hal.science/sic_00497390
Consulté le 26 février 2024
- CLEMENT, G. P., et RASCOE, F. (2013). ETD management et publishing in the ProQuest system and the university repository : A comparative analysis. *Journal of Librarianship and Scholarly Communication*, 1(4). <http://hdl.handle.net/1853/48713>
- DAWSON, D., et LANGRELL, K. (2023). More Obstacles for the Graduate Student Author : Open Access ETDs Trigger Plagiarism Detectors. *Journal of Graduate Librarianship*, 1(1). <https://doi.org/10.59942/2995-9063.1002>
- EGER, T., SCHEUFEN, M., et MEIERRIEKS, D. (2015). The determinants of open access publishing: Survey evidence from Germany. *European journal of law and economics*, 39, 475-503.
- ENGSTRÖM, Y. (1987) *Learning by Expanding. An Activity-Theoretical Approach to Developmental Research*. Orienta-Konsultit.

- ENGESTRÖM, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of education and work*, 14(1), 133-156.
- ELSABRY, E. (2017). Who needs access to research ? Exploring the societal impact of open access. *Revue Française Des Sciences de l'information et de La Communication*, 11, Article 11. <https://doi.org/10.4000/rfsic.3271>
- FERRERAS-FERNANDEZ, T., GARCIA-PENÁLVO, F., MERLO-VEGA, J. A., et MARTIN-RODERO, H. (2016). Providing open access to PhD theses: visibility and citation benefits. *Program*, 50(4), 399-416. <https://doi.org/10.1108/PROG-04-2016-0039>
- GENDRE, F., HUGOT, C., SCHÖPFEL, J., et SOUCHON, A. (2023). La diffusion des travaux d'étudiants : le portail d'archives ouvertes DUMAS. *Culture et Recherche*, 144, 49-51 <https://shorturl.at/RYZ16>
- HAUSTEIN, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, 108, 413-423. <https://doi.org/10.1007/s11192-016-1910-9>
- HAWKINS, A. R., KIMBALL, M. A., et IVES, M. (2013). Mandatory Open Access Publishing for Electronic Theses and Dissertations: Ethics and Enthusiasm. *The Journal of Academic Librarianship*, 39(1), 32-60. <https://doi.org/10.1016/j.acalib.2012.12.003>
- HOMANS, G. C. (1958). Social behavior as exchange. *American Journal of Sociology*, 63(6), 597-606. <https://doi.org/10.1086/222355>
- ILYENKOV, E. (1974), *Dialectical Logic: Essays on its History and Theory*, Progress Publishers.
- KARANASIOS, S. (2018). Toward a unified view of technology and activity : The contribution of activity theory to information systems research. *Information Technology et People*, 31(1), 134-155. <https://doi.org/10.1108/ITP-04-2016-0074>
- KWOK, S. H., et Gao, S. (2005). Attitude towards knowledge sharing behavior. *The Journal of Computer Information Systems*, 46(2), 45-51. <https://doi.org/10.1080/08874417.2006.11645882>
- LIN, J., et FENNER, M. (2013). Altmetrics in evolution : Defining and redefining the ontology of article-level metrics. *Information standards quarterly*, 25(2), 20.
- LOOK, H., et PINTER, F. (2010). Open access and humanities and social science monograph publishing. *New Review of Academic Librarianship*, 16, 90-97. <https://doi.org/10.1080/13614533.2010.512244>
- MAHE, A., et PRIME-CLAVERIE, C. (2017). "Qui dépose quoi sur Hal-SHS ? Pratiques de dépôts en libre accès en sciences humaines et sociales". *Revue française des sciences de l'information et de la communication*, (11). <https://journals.openedition.org/rfsic/3315>
- MARTIN, I. (2015). "Le signalement des thèses de doctorat". *I2D - Information, données documents*, 52(1), 46-47.
- O'KEEFE, J., WILLINSKY, J., et MAGGIO, L. (2011). Public access and use of health research : an exploratory study of the National Institutes of Health (NIH) Public Access Policy using interviews and surveys of health personnel. *Journal of medical Internet research*, 13(4), e1827. <https://doi.org/10.2196/jmir.1827>
- PARK, E. G., et RICHARD, M. (2011). Metadata assessment in e-theses and dissertations of Canadian institutional repositories. *The Electronic Library*, 29(3), 394-407. <https://doi.org/10.1108/02640471111141124>

PERRIN, J. M., WINKLER, H. M., et YANG, L. (2015). Digital Preservation Challenges with an ETD Collection—A Case Study at Texas Tech University. *The Journal of Academic Librarianship*, 41(1), 98–104. <https://doi.org/10.1016/j.acalib.2014.11.002>

RAMIREZ, M L., MCMILLAN, G., DALTON, J. T., HANLON, A., SMITH, H. S., et KERN, C. (2013). "Do open access electronic theses and dissertations diminish publishing opportunities in the sciences ?" *College and Research Libraries*, 74, 368-380. <http://hdl.handle.net/10919/23911>

SCHÖPFEL, J., et PROST, H. (2010). *Les statistiques d'utilisation d'archives ouvertes-État de l'art*. https://archivesic.ccsd.cnrs.fr/file/index/docid/480538/filename/EPEF_Schopfel_Prost_5.0.pdf

SCHÖPFEL, J., PROST, H., JACQUEMIN, B., KERGOSIEN, É., et THIAULT, F. (2024). *La valorisation de HAL. Finalités, outils et process*. <https://hal.science/hal-04450898/>

Schöpfel, J., et Stock, C. (2009). "Les archives ouvertes en France : Un potentiel documentaire pour la formation à distance". *Distances et savoirs*, 7(3), 443-456. <https://www-cairn-info.bibdocs.u-cergy.fr/revue-distances-et-savoirs-2009-3-page-443.htm>

SCHÖPFEL, J., ZENDULKOVA, D., et FATEMI, O. (2014). Electronic Theses and Dissertations in CRIS. *Procedia Computer Science*, 33, 110–117. <https://doi.org/10.1016/j.procs.2014.06.018>

SHEMA, H., BAR-ILAN, J., et THELWALL, M. (2014). "Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics". *Journal of the Association for Information Science and Technology*, 65(5), 1018-1027. <https://doi.org/10.1002/asi.23037>

STANTON, K. V., et LIEW, C. L. (2011). "Open Access Theses in Institutional Repositories: An Exploratory Study of the Perceptions of Doctoral Students". *Information Research : An International Electronic Journal*, 16(4). <https://eric.ed.gov/?id=EJ971947>

WANG, S. et NOE, R. A. (2010). "Knowledge sharing: A review and directions for future research". *Human Resource Management Review*, 20(2), 115-131. <https://doi.org/10.1016/j.hrmr.2009.10.001>

WANG, S., NOE, R. A., et WANG, Z.-M. (2005). *An exploratory examination of the determinants of knowledge sharing*. Society for Industrial/Organizational Psychology.

ANNEXES

Lien vers la base DUMAS prétraitée et anonymisée : <https://zenodo.org/records/12710484>

Figure 5. Nombre de pages des manuscrits de DUMAS en fonction de leur discipline de rattachement. MHEP : Médecine Humaine Et Pathologies

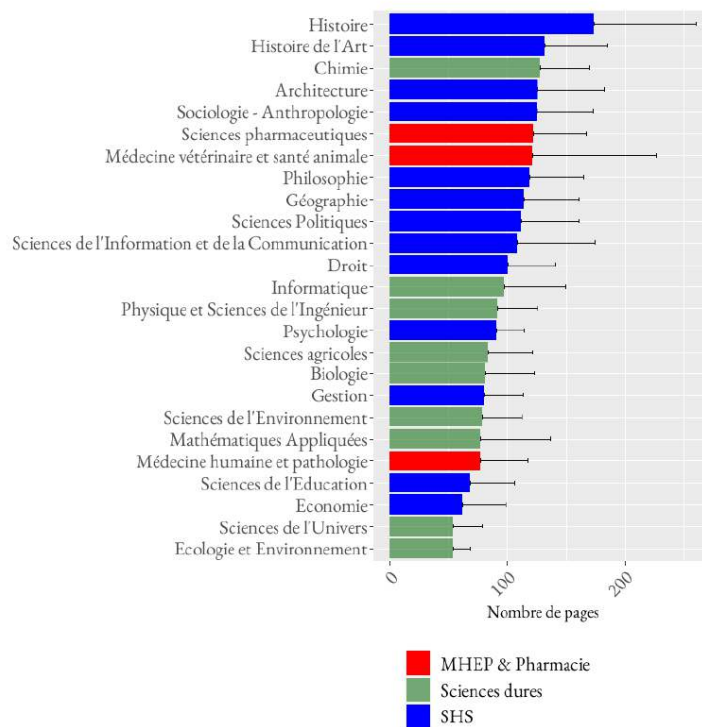
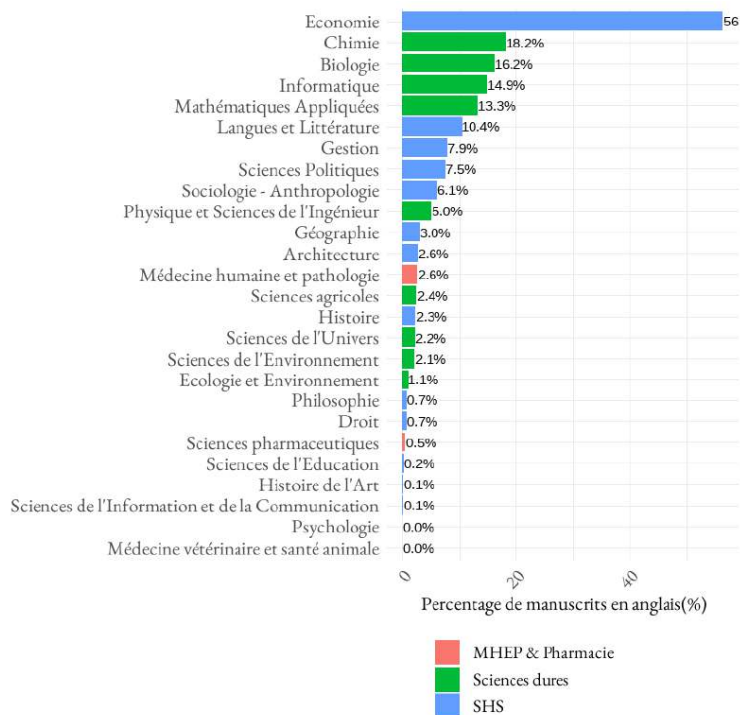


Figure 6. Pourcentage des manuscrits de DUMAS rédigés en anglais en fonction de leur discipline de rattachement. MHEP : Médecine Humaine Et Pathologies



NOTES

1. <https://dumas.ccsd.cnrs.fr/>
 2. <https://agorha.inha.fr/ark:/54721/58>
-

RÉSUMÉS

S'ils ne sont généralement pas considérés comme des écrits scientifiques de premier plan, les mémoires de Master de bonne qualité peuvent néanmoins se révéler utiles à la recherche, notamment car ils traitent souvent de sujets d'actualité et peuvent ce faisant accélérer le travail de chercheurs plus confirmés. Encore faut-il qu'il existe des leviers de motivations susceptibles d'inciter les étudiants à mettre à disposition leur travail. Nous nous proposons ici d'explorer au prisme de la théorie de l'échange social l'un des mécanismes qui poussent les étudiant.es à mettre en ligne leur mémoire dans des archives dédiées : la visibilité du travail partagé, mesurée sur la base du nombre de téléchargements du document. DUMAS, site français qui avait pris son essor à la fin des années 2000, centralisait en 2023 les manuscrits de plus de 50 000 étudiant.es. Via une technique de web scraping, nous avons collecté les métadonnées de l'ensemble ces manuscrits, pour mieux identifier les éléments qui déterminent la visibilité en ligne de ces travaux, tout en portant la focale sur les disciplines. Nous montrons que les travaux relevant des sciences de gestion, bien que minoritaires dans la base, sont ceux qui attirent le plus l'attention des internautes au fil des ans, qui, au sein du répertoire, semblent rechercher avant tout des écrits à visée praxéologique. Nous analysons au prisme de la théorie de l'activité d'Engeström les contradictions qui peuvent émerger de la possibilité pour les étudiants d'acquérir via DUMAS de la visibilité au sein du milieu académique comme en dehors de celui-ci.

While they are usually not considered as important for the advancement of science, well-written electronic dissertations of Master students can prove useful for researchers, since they often cover hot topics and can foster the work of more advanced scholars. However, there must exist incentives to share one's document. We hereby propose to explore through the social exchange theory one of the mechanisms that is likely to drive students to share their dissertation : its visibility, measured through the number of downloads. DUMAS, a French archive, gained momentum in the late 2000s. In 2023, it hosted more than 50 000 manuscripts. We scraped the metadata associated with the entire database, in order to better assess which disciplines were the more likely to attract attention. We show that dissertations from the management sciences scholarly consistently get the highest number of downloads, as the archive users seem to look for documents that are likely to be used in an applied manner. We analyze through the prism of Engeström's activity theory the contradictions that can emerge from the possibility for students to acquire visibility via DUMAS within the academic environment as well as outside it.

INDEX

Keywords : open science, dissertation, web scraping, altmetrics

Mots-clés : science ouverte, mémoire, web scraping, métriques alternatives

AUTEURS

MATTHIEU CISEL

Docteur en sciences de l'éducation et de la formation, l'essentiel de mes recherches portent sur la question des mouvements *open* dans l'enseignement supérieur et la formation d'adultes, open science (à travers les archives scientifiques), open education (avec les ressources éducatives libres, les MOOC), et plus généralement sur les dynamiques de partage au sein de la formation initiale ou continue. Laboratoire AGORA, équipe IDHN. Courriel : [matthieu.cisel\[at\]cyu.fr](mailto:matthieu.cisel[at]cyu.fr)

NICOLAS LAUDIER

Ingénieur en Data Science, Nicolas Laudier a réalisé au cours de ses études une série de stages de recherche durant lesquels il a mis à l'épreuve ses compétences de programmation. L'un de ces stages a spécifiquement porté sur les archives theses.fr et DUMAS.