



HAL
open science

Observatoire des usages COMMONS

Ioanna Faïta

► **To cite this version:**

| Ioanna Faïta. Observatoire des usages COMMONS. 2024, 10.58079/1209h . hal-04753388

HAL Id: hal-04753388

<https://hal.science/hal-04753388v1>

Submitted on 21 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1. Contexte de la recherche

Ce travail de thèse se concentre sur les pratiques de navigation au sein de l'environnement de la science ouverte en SHS dans le périmètre du consortium COMMONS. C'est un projet de recherche qui s'inscrit au premier axe d'OpenEdition Lab et s'aligne avec l'objectif de l'Observatoire des usages qui est de mieux connaître les usages et les usagers des plateformes de publications scientifiques en accès ouvert en SHS. Il rejoint les axes de recherche d'Elico "Science et société" qui accueille la thèse. Mené en tandem avec une autre recherche doctorale en informatique et science des données, cette recherche a un double objectif : opérationnel et réflexif. D'une part, la conception d'un protocole de recension et de documentation des usagers et des usages des services, et d'autre part l'élaboration d'une réflexion critique. Comment la science ouverte affecte-t-elle le panorama des SHS ? *Quid* de son influence sur les contenus, usages, publics académiques et non académiques tout comme sur le savoir en général ? Dans ce billet je traiterai le cas d'OpenEdition, plus précisément de deux plateformes, OpenEdition Books et Journals, sous l'angle d'une première étude de terrain à caractère exploratoire démarrée en mars 2023 et qui aboutira fin 2024.

2. Maillage disciplinaire : quels concepts mobiliser ?

Avant de rentrer dans le vif du sujet, il me paraît important de partager les questionnements théoriques suscités par ce terrain de recherche particulièrement riche, éléments qui guident mes choix méthodologiques et ont influencé l'angle adopté. En effet, les usages étant au noyau de mon approche, cela nécessite un fin maillage disciplinaire et la définition d'un cadre conceptuel composite qui puisse s'adapter au caractère interdisciplinaire du sujet.

Dans l'objectif d'explorer les usages et les interactions avec les outils numériques, je fais donc un double focus, à la fois sur l'usage, comme relation et comme activité, et sur les outils. Au carrefour des deux, le concept d'*affordance* ressort. Établi par le psychologue J.J. Gibson pendant les années 70, les affordances décrivent des possibilités d'action dont un objet peut être porteur. Repris par Norman (1988, 1993) qui fait une distinction entre affordances réelles et donc indépendantes de la perception de l'agent, et affordances perçues et donc liées à la subjectivité de la personne, on peut s'interroger sur les pratiques liées à un outil. Quelles sont les pratiques prescrites par l'objet en lui-même (Paveau, 2012) et inversement quels autres usages peuvent exister, quels détournements ? Par "détournement", j'entends tous les emplois d'un objet ou d'un dispositif, non-attendus ou subversifs mais toujours liés aux propriétés d'un outil, leur matérialité ; le potentiel qui se niche derrière le *design* mais qui n'est pas mis en avant dans son utilisation canonique, prescrite. Ces usages obliques, usages *autres* (Ahmed, 2019), visent à adapter l'outil au contexte précis de l'usage. Par là, ce sont les usages qui réussissent à situer l'outil, à le localiser, à l'ancrer dans un ici et un maintenant. Ce que je souhaite explorer sont donc les appropriations et les détournements des outils opérés par les usages : comment on pourra *hijack* l'utilisation.

La nature hybride de l'objet d'étude peut seulement s'éclaircir par une approche combinée, ce qui est assez récurrent dans la sociologie du web qui a pour objet d'étude des phénomènes sociaux numériques complexes. En effet, pour capturer cette complexité, il est indispensable d'articuler plusieurs grilles d'analyse et mobiliser un panel de méthodes complémentaires. Il s'agit de recourir à plusieurs méthodes afin d'obtenir une compréhension holistique de l'objet et de confronter les résultats de plusieurs sources, validation croisée, (in)validation, contraste, complémentarité, complexification. Cela est possible seulement grâce à une flexibilité méthodologique des éléments contextualisés (Pope et Mays, 1995 ; Bloor et Wood 2006) - ce sont des étapes préalables si l'on veut se pencher sur la question des usages.

La notion d'usage est saisie par différentes disciplines et l'approche ici présentée est plus proche du champ des sciences de l'information et de la communication (SIC). J'aborde les outils numériques en tant qu'objets, c'est cela qui crée la place pour une étude empirique ; *partir des outils pour faire de la sociologie*. Leur désignation en tant qu'objets, dans le contexte d'une étude qui porte sur une infrastructure qui déploie et maintient des plateformes vouées à la communication scientifique en SHS, fait appel à la fois aux outils numériques, aux algorithmes et aux APIs – éléments à la fois visibles et invisibles mais toujours opérants sur les plateformes d'OpenEdition.

OpenEdition Books, OpenEdition Journals et Calenda opèrent grâce à un gestionnaire de contenu *ad hoc* développé par ses propres équipes : Lodel¹. Il s'agit d'un logiciel d'édition électronique spécialisé dans l'édition de textes dans un environnement éditorial structuré. Le déploiement de ce système permet la mise en ligne et la gestion des livres et des revues dans le domaine des SHS. Le portail Hypothèses, dédié au blogging scientifique, est pour sa part basé sur un autre CMS, Wordpress. Lodel fonctionne tout autant comme environnement de lecture, de recherche et de citation – environnement de circulation des ressources numériques. Ces ressources sont des *objets informationnels* de diverses natures : texte, image, carte, billet de blog, annonce d'événement scientifique. Par objet informationnel on entend une collection d'informations qui possède une structure, qui forme un ensemble, qui est interprétable et qui n'est pas obligatoirement numérique. Si ces informations sont aussi numériques, elles sont par extension computationnelles : une collection de données numériques est un objet computationnel, c'est-à-dire, selon certaines analyses un objet sur lequel nous pouvons opérer des calculs sans aller jusqu'à accorder une sorte de réalité essentielle à l'objet computationnel. Dans le cas d'OpenEdition les objets sont informationnels, numériques et computationnels. Ils fonctionnent comme dispositifs, à la fois de communication et de médiation (Jeanneret, 2014) entre usager-es et plateforme.

OpenEdition comme un ensemble englobant les portails Books, Journals, Calenda et Hypothèses, est une infrastructure qui produit et maintient quatre plateformes. Les usager-es, en fonction de leur rôle et leur niveau d'intervention, font face à l'infrastructure ou

¹ Sur la base d'un Content Management System (CMS), modèle de publication préétabli, très répandu à l'ère du Web 2.0 « qui pousse encore plus en profondeur la technique informatique » (Boullier, 2016).

à la plateforme, où aux deux. Pour mieux illustrer ce point, prenons comme exemple une personne qui fait partie de l'équipe éditoriale d'une revue. Cette personne a des identifiants uniques pour accéder au *back-office* Lodel de la revue afin de mettre en ligne le dernier numéro de la revue mais cette même personne peut aussi visiter la plateforme Books, en tant que lecteur·ice, sans connexion préalable, et consulter un ouvrage d'un éditeur de livres. Il s'agit d'un environnement multicouches partagé entre usager·es et outils techniques. Cela nécessite la mobilisation des analyses portant à la fois sur l'infrastructure et l'ethnographie de l'infrastructure (Star, 1999 ; Star et Ruhleder, 2010) mais aussi une approche critique mettant en perspective les plateformes et leur fonctionnement.

Or, cela peut également déclencher une réflexion plus ample sur l'artefact technique et son rapport à la norme et la normativité des outils. Où se place la norme ? Est-elle intérieure à l'usage ? Ne se manifeste-t-elle qu'au cours de l'usage ? Ou est-elle inhérente à l'usage ou est-elle inhérente à l'artefact technique (Pencolé, 2018) qu'est le dispositif ? Mon angle d'approche n'est pas compatible avec une interprétation des outils et des algorithmes comme sujets rationalisés (Verbeek, 2014) mais plutôt comme médiations et moyens de gouvernance. Dans ce sens, je m'appuie sur les épistémologies féministes et la théorie queer pour traiter des questions d'objectivité et de neutralité des outils et m'orienter vers leurs utilisations situées et quotidiennes. « La présence et la circulation d'une représentation n'indiquent nullement ce qu'elle est pour ces utilisateurs » (De Certeau, 1980), alors je souhaite étudier la zone grise entre représentation instruite et appropriation ancrée. Lire l'usage comme expérience vécue par et à travers un corps vécu, permet de mobiliser les outils de la phénoménologie. Toutefois cela seulement pour parler du rapport développé avec les outils et pas dans le but de présenter le numérique comme phénoménologiquement immatériel, doté d'ubiquité et d'instantanéité (Dyer-Witheford, 2015 ; Terranova, 2000).

En outre, dans ce contexte numérique abordé par le prisme des usages ancrés, une entrée particulièrement riche est le concept de "données". Qu'est-ce que la donnée et qu'est-ce qui est encodé en elle ? Alors, quels sont les différents types que la donnée peut en avoir ? La multitude des réponses possibles manifeste l'importance accrue que la question prend en lien avec le développement exponentiel des technologies.

Commençons par distinguer deux définitions des données : en un sens très général mais aussi bien adapté au contexte des SHS, tout ce qui est mobilisé comme matériau, source, trace, observation pour fabriquer de l'information et/ou de la connaissance peut être considéré comme "données". Selon une définition numérique plus stricte, les données sont des mesures souvent quantitatives, réalisées par des appareils/instruments selon certains standards. C'est grâce à leur structure que l'on peut y opérer des calculs.

Ce que l'on peut entendre donc par donnée dans un contexte numérique est très varié (Da Sylva 2017) et rassemble des caractéristiques des deux définitions mobilisées: **données ouvertes** ou *open data* (Dickner 2017 ; Comité OGGO 2014 ; Peugeot 2014 ; Janssen, Charalabidis & Zuidewijk 2012 ; Mercier 2011), **mégadonnées** ou *big data* (Gandomi & Haider 2015 ; Chen & Zhang 2014 ; Kim, Jeong & Kim 2014 ; Boyd & Crawford 2012), **données liées** ou *linked data* (Harth, Hose & Schenkel 2016 ; Bizer, Heath & Berners-Lee 2011), **métadonnées** c'est à dire des données qui décrivent d'autres données, ou encore **données de recherche** en d'autres termes des données produites dans un contexte de recherche. C'est ici que le paramètre disciplinaire s'insère : *quid* des données de recherche

numériques dans le domaine précis des sciences humaines et sociales ? Cela fait appel aux données générées lors de l'observation, comme mentionné auparavant, ou bien lors de la recherche dans des archives, mais cela peut également représenter les retranscriptions d'entretiens ou les notes prises en marge dans un journal de bord, mais qui peuvent être aussi soumis à des calculs pour cause de leur nature numérique.

Si l'on emprunte la voie des sciences de l'information et de la communication, le terme donnée vient différencier un élément brut, d'une part d'une information et, d'autre part, d'une connaissance (Rowley, 2007 ; Zins, 2007). Une donnée c'est une trace qui n'a pas une valeur en soi si elle est découplée de son contexte de production. Une donnée peut donc être indicative de quelque chose mais hormis son contexte elle n'est guère interprétable. De plus, une donnée doit être sémantisée avant qu'elle ne se transforme en information. Cette définition est valable dans le champ disciplinaire des SHS mais la question de la donnée peut aussi se transposer autrement dans le terrain numérique d'OpenEdition. C'est-à-dire, dans cet environnement les données sont au centre des activités des internautes : les données qui sont traitées, partagées et mises en circulation, mais qui sont aussi le reflet de nos activités en ligne. Elles sont aussi les empreintes que nous laissons sur les serveurs lors de notre passage dans le réseau et si correctement sémantisées elles peuvent devenir une des sources principales pour documenter les pratiques dans le cadre de cette recherche. Toutefois, sur OpenEdition tout n'est pas *donnée*. Cela dépend des formats, des outils, des usages et des contextes d'usage. Un exemple : pour une personne qui utilise la plateforme OEB, option lecture, pour lire le chapitre d'un ouvrage et prendre des notes à côté, ce contexte d'usage précis fait de l'objet chapitre, un document, une ressource électronique. En revanche, pour une personne qui récupère systématiquement les métadonnées des articles d'un journal dans un contexte d'usage type veille informationnelle sur les nouveaux articles parus sur telle thématique à telle année pour constituer, l'objet article a plutôt le rôle de la donnée.

Par là, il est aussi pertinent de s'interroger sur les transformations survenues au milieu de la recherche scientifique comme conséquence du rôle changeant des données. La mise à disposition d'une masse critique de données, ouvre un panorama de possibilités d'actions. Cela se manifeste à travers un vocabulaire très riche qui décortique et décrit les différents niveaux de contact potentiel aux données (*i.e.* accès, navigation, appropriation, diffusion, exploitation, exposition, indexation, valorisation, partage, réutilisation, restitution etc.) et nous permet/amène devant la nécessité de mobiliser des nouvelles méthodes d'analyse, comme la fouille de données (*data mining*) et la fouille de textes (*text mining*), des outils et des techniques qui permettent le classement de sources, l'extraction des informations à l'aide de la statistique ou de l'intelligence artificielle. Ces évolutions relèvent des défis sur le plan technique, épistémologique, théorique ou encore économique.

La place des données est centrale dans cette recherche tout comme elle l'est dans l'économie des plateformes à l'ère du capitalisme de plateformes. Il est néanmoins crucial d'éviter un amalgame et donc nécessaire de distinguer entre les plateformes marchandes et les plateformes d'OpenEdition qui ne le sont pas. En tout cas les plateformes d'OpenEdition ne sont pas à part entière marchandes ; cependant l'existence d'une offre de commercialisation, à savoir OpenEdition Freemium, qui est explicitée plus loin, brouille cette distinction absolue même si OpenEdition ne devient pas une plateforme marchande puisque les revenus de ses offres économiques sont redistribués dans la communauté scientifique

d'édition. Par ce prisme là, je souhaite me positionner vis-à-vis de lectures sur le capitalisme de plateforme qui ne peuvent pas se transposer entièrement à ce contexte d'étude sans écraser les spécificités de la communication scientifique numérique, mais qui ont néanmoins beaucoup inspiré ma compréhension du terrain. Des ouvrages comme la *Sociologie du numérique* de Boullier et maintes autres analyses qui traitent des questions liées au capitalisme de plateformes (Srnicek, 2016) à l'économie numérique et au *digital labor* (Broca, 2017), montrent comment la donnée devient une source d'extraction de plus-value économique et renforcent de formes de travail gratuit ou invisible. De plus, un autre rapprochement identifié, c'est le rôle que la proposition des services vient occuper au sein du fonctionnement des plateformes. Même si dans le cas des plateformes d'OpenEdition ce ne sont pas des services à but lucratif ou en tout cas suivant une logique de marchandisation des contenus, la proposition des services est présente. Il s'agit de proposer des services supplémentaires : de statistiques, de suggestions, de référencement dans les catalogues collectifs et de moissonnage par les outils utilisés par les bibliothèques. Malgré le fait que le modèle économique d'OpenEdition ne repose pas sur la monétisation des données et comme mentionné auparavant, tout n'est pas donnée sur OpenEdition en tout cas de manière définitive, tout de même l'infrastructure d'OpenEdition se manifeste par quatre plateformes et a par conséquent certaines caractéristiques communes avec les plateformes marchandes et on peut y identifier des effets de plateforme. C'est un axe qui mérite d'être analysé par la suite.

2.1. Quelles données ?

Pour resserrer la focale sur la donnée dans l'analyse des situations d'utilisation dans le cadre de cette recherche, on revient avant tout à la donnée sous sa forme de trace unique mais pas seulement. En effet, nous disposons de différentes sources à traiter : ce qu'un usager laisse après son passage, ses *traces numériques* (Merzeau, 2013), log serveur et statistiques de fréquentation. Cela peut nous aider à identifier des modèles à grande échelle. De plus, toutes les informations liées à l'indexation des contenus sont disponibles et peuvent être croisées aux observations afin de mieux comprendre le lien entre navigation et contenu. Afin de construire une stratégie qui pourra dans un deuxième temps s'appliquer à l'ensemble des plateformes, j'envisage une analyse préalable et parallèle de tout ce qui est lié au système. Par là j'entends tout ce qui est de l'ordre du dispositif et qui fait l'interface avec l'utilisateur, c'est-à-dire la matérialité de cette communication : analyse sémiotique des plateformes et de l'architecture technique des plateformes.

Diversité des données

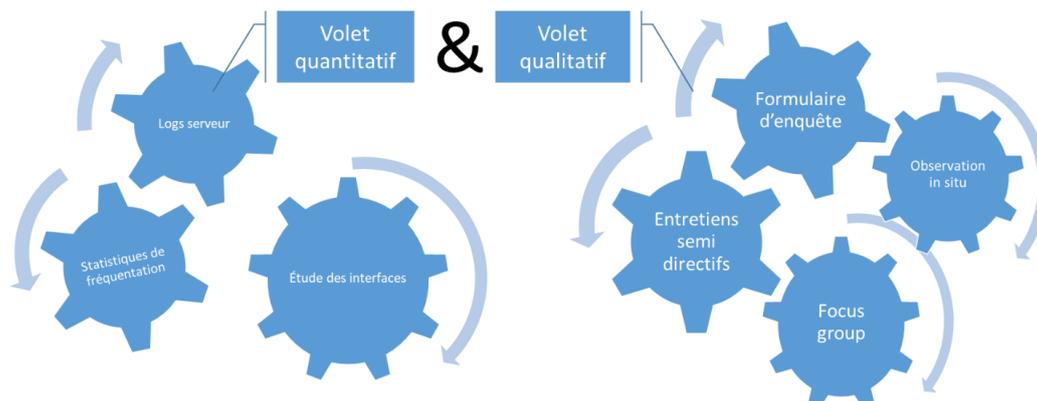


Schéma représentant les différents types de données disponibles pour l'analyse de l'étude de cas généré par Faïta I. pour les diapositives explicatives lors des rencontres avec les équipes éditoriales, 2024

La diversité des données à disposition justifie la définition d'une méthode combinant différentes approches, en articulant le computationnel, au quantitatif et au qualitatif. Nous cherchons à approfondir notre compréhension de la relation dynamique, de l'agence des acteurs humains et non humains, entre les utilisateurs et les outils numériques au sein d'une infrastructure d'hébergement et de diffusion de connaissances. L'approche quantitative n'est pas mobilisée dans un but de mesurer. Au contraire, elle vise à mieux comprendre tenant compte du contexte et de l'environnement culturel.

Pour illustrer la nécessité d'articuler les méthodes, je prends appui sur les travaux menés au sein du laboratoire Ciber. Plus précisément, je m'inspire des recherches de Carol Tenopir qui est plutôt orienté recherche qualitative et David Nicholas dont l'approche est quantitative : ensemble, les deux ont travaillé sur l'articulation des deux approches dans le cadre de nombreux projets de recherche et groupes de travail. Un de ces groupes de travail interdisciplinaire et à visée internationale qui a tout de suite attiré mon intérêt s'intitule "Beyond Downloads" et a profité d'un financement par Elsevier en 2014. Ces travaux sont centrés sur les pratiques des lectorats universitaires et extra-universitaires des revues de publication scientifique en ligne. Ce qui est souligné au sujet des usages sont les différents degrés d'utilisation possibles, c'est-à-dire que malgré le fait de disposer des statistiques de téléchargement, éléments quantifiables, cependant cela ne peut pas se traduire directement en termes de pratique. Ainsi, quelles sont véritablement les utilisations « secondaires » et comment cerner le comportement post-téléchargement ? Cela souligne la nécessité de dépasser certains réflexes ou habitudes méthodologiques qui peuvent être plus enfermants que révélateurs, de penser à d'autres indicateurs et méthodes croisées, dans le but de mesurer l'utilisation au-delà des téléchargements certes mais surtout afin de comprendre les usages. En d'autres termes, il s'agit d'aller au-delà des statistiques COUNTER 5, mesures standardisées pour les téléchargements d'articles proposées par les plateformes de communication scientifique y compris OpenEdition. Ce qui est proposé à la place, c'est un croisement de ces mesures avec une observation des pratiques de partage, académique formelle et informelle, à travers l'analyse qualitative et la recherche ethnographique.

Il est aussi nécessaire de distinguer ici quantitatif et computationnel. Le volet quantitatif de cette recherche comprend les enquêtes, les métriques COUNTER 5 et l'exploitation des données Matomo. Or, ce que nous entreprendrons en collaboration avec la recherche informatique sera d'analyser en plus de ces données les logs serveur.

Qu'est-ce qu'un *fichier log* ? Il s'agit d'un fichier texte, qui permet de stocker l'historique des événements survenus sur un serveur afin de maintenir l'infrastructure numérique, tracer et résoudre des erreurs, documenter les interactions entre serveur et la machine du client (communication machine-machine). Ce fichier trace, liste et horodate toutes ces interactions. Ce sont des événements de consultation. Pour transformer ces données en informations sur les usages, il faudra croiser ces éléments à des hypothèses : combiner les *traces* numériques après avoir d'une part constitué une taxonomie d'usages possibles sur les plateformes, et d'autre part produit une cartographie d'acteurs. C'est ainsi que l'on pourra mieux comprendre les usages et les traduire en pratiques réelles.

Quelques analyses tirées de la littérature déjà existante sur ce que l'on peut faire avec les logs vont dans le même sens que l'approche adoptée: « [...] *log describe the usage data of a system or application* » (Agosti and al., 2012) mais « *only log analysis can reveal the usage status* » (Fu, 2021). Donner du sens au descriptif ; non pas seulement s'appuyer sur les logs pour produire des métriques, dans une approche traditionnelle, mais mettre en place un examen critique. Selon les travaux de Oliphant, qui développe un regard qui s'inscrit dans le courant des *critical data studies*, l'usage du terme de *big data*, qui représente des milliers de lignes de log dans notre terrain d'OpenEdition, n'a pas pour vocation à renoncer à l'exploitation de ces données, au contraire, c'est pour souligner la nécessité de les enrichir avec des "*small data*" (Kitchin and Lauriault, 2014; Thatcher and Burns, 2013) ou des "*rich data*".

Pour mieux comprendre comment cela pourra s'appliquer sur notre terrain : nous avons décidé d'utiliser les fichiers log comme source et y appliquer des critères spécifiques afin d'extraire des sessions de navigation. Il s'agit d'un détournement d'usage puisque ces fichiers textes descriptifs d'actions n'ont pas vocation à nous renseigner sur les pratiques humaines. Pour ce faire, nous avons dû nous baser sur des éléments multiparamètres comme le changement de *user-agent* et du *referer*, combinés au laps de temps écoulé entre deux requêtes consécutives. Les sites des revues et des éditeurs de livres ont été traduits par des points dans un espace bidimensionnel, où les points rapprochés illustrent des groupements (*clusters*) de sites qui ont été consultés lors d'une même session de navigation. C'est ainsi que l'on peut visualiser des rapprochements entre les sites, distances créées par les consultations, autrement dit, par les usages. Il est aussi intéressant d'introduire la notion du mouvement et par là essayer de représenter dans une carte l'idée de passer d'un contenu à un autre, mettre en lumière des potentiels motifs (*patterns*) de navigation par exemple. Réussir à créer des *clusters* dans une cartographie de sites n'est pas néanmoins une information exploitable en tant que telle. Ces distances doivent être sémantisées : quelles sont les raisons qui ont conduit à la formation des proximités et des distances ? Afin d'y répondre, on peut d'abord explorer les contenus de chaque site et leurs particularités pour expliquer la formation de ces *communautés d'usages* : disciplinaires, langagières, ou bien liées au modèle de diffusion, la périodicité ou l'ère géographique de circulation. Seul bémol : le manque de granularité fine de cette approche, qui met sur un

même plan le sites entier d'une revue et le site d'un éditeur de livres sans prendre en compte la spécificité d'une collection de livres, ou bien d'un numéro thématique ou d'une rubrique alimentée au fil de l'eau. Ainsi, sans pouvoir préciser les liens exacts entre tel article, tel compte-rendu d'un livre et tel chapitre d'un ouvrage d'un livre par exemple, nous perdons en précision.

La notion de *communauté d'usages* qui s'introduit ici n'est pas sans importance. Pour approfondir la question de ce qui fait communauté, il est davantage important d'articuler ces observations et hypothèses préliminaires à l'approche qualitative. C'est dans l'optique de localiser les traces dans un espace, de les incarner et les lier à la mémoire, que l'on se fixe l'objectif de tirer des fils entre les pratiques identifiées par les traces numériques et les pratiques, telles qu'observées ou telles que décrites par les usagers et usagères. De plus, l'idée qu'exercer une certaine pratique, dans l'univers de l'information scientifique et plus particulièrement au sujet des revues, est lié à l'objet revue scientifique dans son historicité, est aussi à explorer.

3. La nécessité d'une première étude de cas

Afin de réduire dans un premier temps le terrain des quatre plateformes d'OE nous prévoyons une première étude sur deux plateformes, OpenEdition Books et OpenEdition Journals. C'est un travail circonscrit à deux cas de revues et une collection d'une maison d'édition qui ont subi un changement de modalité d'accès.

Quels sont les usages sur les deux plateformes, cela couvrant les usages de lecture mais pas seulement ; au même titre je m'interroge sur la façon dont les équipes éditoriales s'approprient des outils et des services ainsi que le *back-office*, sur leurs pratiques de lecture quand ils et elles sont lecteur-ices, dans ce cas on peut parler d'un double statut d'utilisation, et sur le rôle des intermédiaires.

Le caractère exploratoire de cette première campagne permet d'éprouver et mieux articuler les concepts et tester la méthodologie pour ensuite monter en échelle et concevoir un protocole plus générique, répliquable, couvrant l'ensemble des plateformes et capable de suivre les usages en temps réel. Il s'agit d'une procédure doublement itérative, entre les méthodes dans chaque étape du projet et entre les étapes dans le temps long de la recherche doctorale.

Échéance de ce premier cycle : fin 2024

4. Quels terrains ?

Deux revues :

1. La *Revue de la Bibliothèque Nationale et Universitaire de Strasbourg* (RBNU) qui a subi un changement de modalité (FOA vers OA Freemium) à partir de janvier 2023. Cela demande donc de mener un travail comparatif entre les années 2022 et 2023.
2. La Revue *Hommes & migrations* qui a abandonné sa barrière mobile (BM) à partir du 1er janvier 2024 et jusqu'en 2026 car elle est bénéficiaire du plan stratégique

national du Ministère de l'enseignement supérieur et de la recherche (PSN MESR) Couperin/ABES version Cairn. Cela demande donc de porter un regard comparatif entre les années 2023 et 2024.

Un éditeur de livres :

1. Trois collections des éditions Université Grenoble Alpes (UGA éditions) qui ont subi un changement de modalité (format exclusif² vers OA freemium) à partir de 2023 : à savoir les collections *Ecotopiques* et *Langues, gestes, paroles* en juillet 2023, *Carrefours des idées* en septembre 2023. Il est intéressant de mener un comparatif entre les années 2022 et 2023.

Quelques éléments à prendre en considération :

- L'aspect temporel, alias l'étendue temporelle (pour comparer, pour comprendre quelle usage est ancrée sur quelle temporalité)
- Les spécificités de chaque revue et des collections de livres (champ disciplinaire, public ciblé)

Dans les trois cas nous nous confrontons à un différent type de changement quant à la modalité d'accès. Afin de mieux comprendre ce que cela implique, voici quelques éléments factuels concernant le modèle Freemium et quelques éléments de vocabulaire pour éclaircir la transition et établir une base commune de compréhension.

4.1 Programme Freemium

OpenEdition Freemium est un programme pour le développement de l'édition scientifique en libre accès dans le domaine des sciences humaines et sociales. Ce partenariat, proposé exclusivement aux institutions (bibliothèques, campus, centres de recherche), vise à construire un modèle économique durable. Ce n'est pas donc une offre de commercialisation s'adressant aux lecteur·ices individuel·les. La totalité des revenus engendrés par ce programme est réinvestie dans le développement de l'édition électronique scientifique en libre accès.

- + Les contenus (revues, livres, carnets de recherche, annonces scientifiques) sont diffusés sur les plateformes en libre accès.
- + Ces derniers sont complétés par des services (i.e. statistiques d'usages au format COUNTER, service permettant aux institutions partenaires de bénéficier de statistiques personnalisées) et formats premium (PDF, ePub) mis à disposition exclusivement pour les institutions et leurs usagers.
- + Des services pour faciliter l'intégration dans les catalogues et des outils de découverte des bibliothèques.

par exemple : Pour les contenus qui font partie du programme Freemium, les textes sont accessibles en libre accès au format HTML pour tout internaute, et ils sont

² Accès exclusif : modalité d'accès sur OEB. C'est le statut le plus fermé qui puisse exister sur OpenEdition. Cela signifie que seul un extrait ou résumé de l'ouvrage est disponible librement en HTML ; la version complète, que ce soit le HTML ou les PDF, ePub, est réservée aux abonné·es.

téléchargeables aux formats PDF et ePub uniquement pour les utilisateur·ices des institutions partenaires. Aucun DRM ni quota de téléchargement ne sont appliqués.

Le programme est composé de deux offres complémentaires : Freemium pour les revues et Freemium pour les livres.

4.1.1 Freemium pour les revues (OpenEdition Freemium for Journals)

L'offre OpenEdition Freemium for Journals, dédiée aux revues. Il s'agit d'un abonnement annuel au bouquet de revues Freemium, bouquet qui évolue au 1er janvier de chaque année, intégrant de nouvelles revues au bouquet existant. En 2024, le bouquet est constitué de 185 revues en sciences humaines et sociales, publiées sur la plateforme OpenEdition Journals. À cela s'ajoutent également 12 titres bonus issus du deuxième plan de soutien aux revues, dispositif conçu pour accélérer l'ouverture d'accès aux revues.

4.1.2 Freemium pour les livres

L'offre OpenEdition Freemium for Books, dédiée aux livres. Plusieurs modes d'acquisition sont proposés (titre à titre, par bouquet, commande permanente sur critères ou par collection via l'abonnement au service Nouveautés). Types de licence : l'acquisition pérenne, la location avec ou sans option d'achat.

5. Protocole d'enquête

Un seul protocole est prévu pour les trois cas. C'est un protocole consistant en un formulaire commun, adressé aux lectorats des trois entités éditoriales, suivi par un cycle d'entretiens semi-directifs dans un deuxième temps, afin de faire parler les acteur·ices. Donc pour chaque projet ce formulaire sera multiplié par trois.

Le formulaire/questionnaire a une double fonction, à la fois contribuer à la récolte des informations préliminaires sur les pratiques de lecture et la conception de la revue/éditeur mais aussi au recrutement des enquêté·es pour le premier type d'entretiens.

Pour chaque revue et éditeur, trois types d'entretiens sont prévus et par conséquent trois trames différentes :

1. Une trame pour les lecteur·ices
2. Une trame pour l'équipe d'édition
3. Une trame pour l'intermédiaire chargée de la diffusion en bibliothèques, librairies : il s'agit du personnel interne d'OpenEdition, les correspondants et correspondantes du programme de commercialisation Freemium

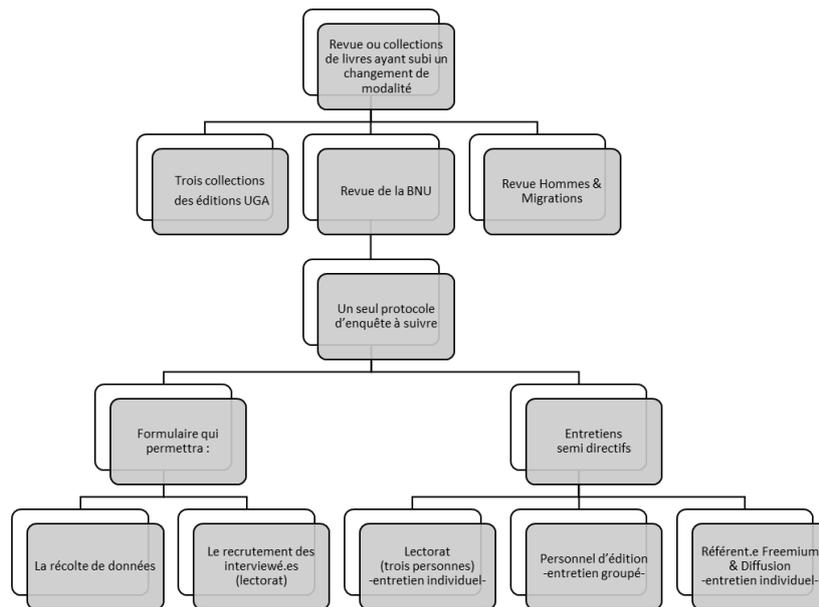


Schéma représentant les différents phases prévues pour l'étude de cas, généré par Faïta I. pour les diapositives explicatives lors des rencontres avec les équipes éditoriales, 2024

Bibliographie

Agosti, M. *et al.* 2012. Web log analysis: A review of a decade of studies about information acquisition, inspection and interpretation of user interaction, *Data Mining and Knowledge Discovery* 24, pp. 663–696. doi : <https://doi.org/10.1007/s10618-011-0228-8>

Ahmed, S. 2019. *What's the Use?: On the Uses of Use*. Duke University Press.

Bizer, C. Heath, T. & Berners-Lee, T. 2011. Linked Data : The Story So Far. In *Semantic Services, Interoperability and Web Applications : Emerging Concepts : Emerging Concepts*, par Sheth Amit. IGI Global.

Bloor, M. & Wood, F. 2006. *Keywords in qualitative methods. A vocabulary of research concepts*, Thousand Oaks, Sage.

Boullier, D. 2016. *Sociologie du numérique*. Armand Colin.

Boyd, D. & Crawford, K. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15 (5), pp. 662-79.
doi : <https://doi.org/10.1080/1369118X.2012.678878>

Broca, S. 2017. Le digital labour : l'hybridation des marxismes et ses fragilités théoriques. *Poli – Politique de l'Image*, 13, pp.26-38.

Chen, C.L. & Chun-Yang, Z. 2014. Data-intensive applications, challenges, techniques and

technologies : A survey on Big Data-ScienceDirect. *Information Sciences* 275, pp. 314-347.

Comité OGGO. 2014. *Les données ouvertes : la voie du futur. Rapport du comité permanent des opérations gouvernementales et des prévisions budgétaires*. Rapport d'un comité parlementaire (Canada) No 5-OGGO (41-2). Ottawa : Chambre des communes du Canada.

Da Sylva, L. 2017. Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information. *Documentation et bibliothèques*, 63(4), 5–34. doi : <https://doi.org/10.7202/1042308ar>

De Certeau, M. 1990. *L'Invention du quotidien*, t. I, Arts de faire [1980], édition établie et présentée par Luce Giard, Paris, Gallimard.

Dickner, N. 2017. *Comprendre et manipuler les données ouvertes de l'administration publique. La situation au Gouvernement du Québec et à la Ville de Montréal*. (Mémoire de maîtrise, Université de Montréal, Montréal).

Dyer-Witheford, N. 2015. *Cyberproletariat: Global Labour in the Digital Vortex*, Toronto, Pluto Press.

Fu, Y. Lomas, E. Inskip, C. 2021. Library log analysis and its implications for studying online information seeking behavior of cultural groups, *The Journal of Academic Librarianship*, Volume 47, Issue 5, doi: <https://doi.org/10.1016/j.acalib.2021.102421>.

Gandomi, A. & Murtaza H. 2015. Beyond the Hype : Big Data Concepts, Methods, and Analytics. *International Journal of Information Management* 35 (2), pp. 137-44. doi : <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Gibson, J.J. 1977. « The Theory of Affordances » in R. Shaw and J. Brandsford (eds), *Perceiving, Acting, and Knowing : Toward an Ecological Psychology*, pp. 62-82.

Granovetter, M. 1985. « Economic-Action and Social-Structure – The Problem of Embeddedness ». *American Journal Of Sociology* 91(3), pp. 481-510.

Harth, A. Hose, K. & Schenkel, R. 2016. *Linked Data Management*. Boca Raton, FL : CRC Press, Taylor & Francis Group.

Janssen, M., Charalabidis, Y. & Zuiderwijk A. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29 (4), pp. 258-68. doi : <https://doi.org/10.1080/10580530.2012.716740>

Jeanneret, Y. 2014. *Critique de la trivialité. Les médiations de la communication, enjeu de pouvoir*, Paris, Éditions Non Standard

Kim, W. Jeong O.-R. & Chulyun K. 2014. A Holistic View of Big Data. *International Journal of Data Warehousing and Mining* 10 (3), pp. 59-69.

- Kitchin, R. 2013 "Big Data and Human Geography: Opportunities, Challenges, and Risks." *Dialogues in Human Geography* 3, no. 3, pp. 262–267.
- Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage Publications, 2014.
- Kitchin, R. & Lauriault, T. P. 2014 "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." *Social Science Research Network: Tomorrow's Research Today*. ("The Programmable City Working Paper 2"; pre-print version of chapter to be published in Eckert, J., Shears, A. and Thatcher, J. (eds.), *Geoweb and Big Data*. University of Nebraska Press. Forthcoming.)
- Mays, N. & Pope, C. 1995. « Qualitative Research: Rigour and qualitative research », *BMJ*, 311(6997).
- Mercier, S. 2011. Open data et bibliothèques. *Documentaliste-Sciences de l'Information* 48 (3), pp. 8-13.
- Merzeau, L. 2013. L'intelligence des traces. In: *Intellectica. Revue de l'Association pour la Recherche Cognitive*, n°59, 2013/1. De la trace à la connaissance à l'ère du Web. pp. 115-135. doi : <https://doi.org/10.3406/intel.2013.1088>
- Norman, D.A. 1988. *The Design of Everyday Things*. New York : Doubleday.
- Norman, D.A. 1993. « Les artefacts cognitifs » in Conein, B. et al. (1993) (dir.). *Les objets dans l'action*.
- Paveau, M.-A. 2012. Ce que disent les objets. Sens, affordance, cognition. *Synergies Pays Riverains de la Baltique* 9, pp. 53-65.
- Pencolé, M. 2018. Nos algorithmes peuvent-ils être plus justes que nous ?. *Revue française d'éthique appliquée*, 5, pp. 67-80.
- Peugeot, V. 2014. Données publiques ouvertes : or du 21^e siècle ou biens communs ? *Documentaliste-Sciences de l'Information* 50 (4) , pp. 48-63.
- Rowley, J.. 2007. The Wisdom Hierarchy : Representations of the DIKW Hierarchy. *Journal of Information Science* 33 (2), pp. 163-80. doi : <https://doi.org/10.1177/0165551506070706>
- Srnicek, N. 2016. *Platform Capitalism*, Polity.
- Star, S. L. 1999. The Ethnography of Infrastructure. *American Behavioral Scientist*, 43(3), pp. 377-391.
- Star, S. L. & Ruhleder, K. 2010. Vers une écologie de l'infrastructure: Conception et accès aux grands espaces d'information. *Revue d'anthropologie des connaissances*, 4,1, pp. 114-161.

Tenopir, C. Hughes, G. Christian, L. Allard, S. Nicholas, D. Watkinson, A. Woodward, H. Shepherd, P. & Anderson, R. 2014. "To Boldly Go Beyond Downloads: How Are Journal Articles Shared and Used?". Proceedings of the Charleston Library Conference.
<http://dx.doi.org/10.5703/1288284315614>

Terranova, T. 2000. Free Labor: Producing Culture for the Digital Economy. *Social Text* 18(2), pp. 33-58.

Verbeek, P.P. 2014. Some misunderstanding about the moral significance of technology. In: Kroes P, Verbeek P-P (eds) *The Moral Status of Technical Artefacts*. Dordrecht: Springer, pp. 75–88.

Zins, C. 2007. Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the American Society for Information Science and Technology* 58 (4), pp. 479-93.
doi : <https://doi.org/10.1002/asi.20508>