



**HAL**  
open science

# Naviguer dans les savoirs à l'ère numérique Pour une ethnographie des pratiques informationnelles sur Gallica

Simon Dumas Primbault

► **To cite this version:**

Simon Dumas Primbault. Naviguer dans les savoirs à l'ère numérique Pour une ethnographie des pratiques informationnelles sur Gallica. *Études de communication - Langages, information, médiations*, 2023, 61, pp.61-89. 10.4000/edc.16108 . hal-04347149

**HAL Id: hal-04347149**

**<https://hal.science/hal-04347149>**

Submitted on 15 Dec 2023

**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 4.0 International License

# Naviguer dans les savoirs à l'ère numérique

## Pour une ethnographie des pratiques informationnelles sur Gallica

### Introduction : un chassé-croisé méthodologique

Le développement des bibliothèques numériques – entendues au sens large comme des collections organisées de documents électroniques accessibles en ligne sur des plateformes dédiées avec des outils de recherche et de consultation – a radicalement transformé les bibliothèques tant comme institutions et services que comme dispositifs sociotechniques. Du point de vue des usages, les pratiques documentaires traditionnelles de bibliothèque ont été doublées, au cours des décennies passées, d'un ensemble de pratiques numériques nécessaires aux chercheurs pour naviguer des plateformes en ligne.

Depuis leur définition dans les années 1990, notamment par Christine Borgman (2000), les bibliothèques numériques sont devenues le sujet de nombreuses études. Dans le prolongement des études sur la recherche d'information en bibliothèque dite physique, les champs de l'*information seeking* ou *information searching* aux États-Unis et au Royaume-Uni se sont penché sur les pratiques déployées par les usagers des bibliothèques numériques pour répondre à leur « besoin d'information [*information need*] » qui trouve son origine dans une inadéquation entre le savoir acquis et des objectifs individuels (Case, 2002 : 5). Les pratiques informationnelles sont alors définies de façon relativement téléologique comme des stratégies, et classifiées, selon la clarté de l'objectif final, en stratégies analytiques (*analytical strategies*) – lorsque le but est précis, la recherche formelle et dirigée – et stratégies de butinage (*browse strategies*) – si le but est flou ou le besoin mal défini, donc la recherche informelle et non dirigée (Case, 2002 : 84 ; Boubée et Tricot, 2011 : 88)<sup>1</sup>. Dans ce cadre théorique, la *directed search* est bien souvent l'apanage des usagers experts – car ils savent ce qu'ils cherchent et comment formuler leur recherche – tandis que le *browsing* est plus souvent réservé aux novices. Du point de vue du vocabulaire, les champs de l'*information seeking/searching* ont tendance à prendre pour quasi-synonymes les termes *information*, *knowledge*, *document* et *data* (Case, 2002 : 61-62), tandis que l'activité de *browsing* est réservée au contenu des bibliothèques et celui de *navigating* à leur interface.

Or, c'est justement l'intrication entre contenu et interface que souligne la polysémie du terme navigation en français qui nous intéresse : comment se saisit-on des prises d'une interface pour explorer un corpus et selon quelles modalités appréhende-t-on différemment information, savoir, document et donnée ? Par ailleurs, le terme français navigation ne charrie pas l'idée d'une lecture en diagonale ou d'un survol (comme c'est le cas pour *browsing*) ce qui permet d'y inclure effectivement la recherche dirigée comme un mode de navigation particulier. Ainsi, afin de s'extraire d'une définition téléologique des pratiques par le besoin d'information,<sup>2</sup> nous envisagerons la navigation comme pratique d'orientation qui est sa propre fin en prenant le cheminement en soit comme opération de repérage dans l'espace documentaire : il s'agira donc de s'intéresser aux *représentations* que les usagers se font de leurs cheminements en-deçà de l'interface ou de la structuration réticulaire des ensembles documentaires (grâce aux entretiens)<sup>3</sup> ainsi qu'aux *prises* que les usagers saisissent dans l'interface et l'architecture pour se repérer (grâce à l'étude de logs).

La navigation se déclinera alors en divers régimes d'engagement (Auray, 2017 ; Gilliotte, 2022) qui, au sein de Gallica, se construisent presque exclusivement par des séquences d'actions de l'ordre de la recherche dirigée (par exemple, une requête par mot-clé dans le

---

<sup>1</sup> Les stratégies de butinage sont également déclinées en divers types, selon le degré de formalité de la recherche ou de clarté de l'objectif (Case, 2002 : 87 ; Boubée et Tricot, 2011 : 89).

<sup>2</sup> Les pratiques dépendant également des contraintes temporelles et des attendus professionnels, de ce qu'une interface et une architecture informationnelle rendent possible et contraignent...

<sup>3</sup> Des représentations très souvent liées à l'action (Souchier *et al.* 2003 : 138-9).

moteur de recherche). Le régime de déambulation peut donc se cacher derrière le chaînage de recherches dirigées multiples, ce qui vient brouiller la distinction traditionnelle entre stratégies analytiques et stratégies de butinage. Par ailleurs, une telle compréhension de la navigation comme pratique en soi déclinée en régimes d'engagement permet également de revoir le partage entre experts et novices car, d'une part, même les experts naviguent, déambulent et flânent et, d'autre part, la navigation nécessite des compétences topologiques, l'absence de contraintes de performance et du temps libre qui sont inégalement répartis.

Dans la recherche francophone des années 2000, la navigation est souvent étudiée à travers tout le World Wide Web, envisagé comme un espace, qui engendrerait la fragmentation de l'unité documentaire : tout devient information (Boullier et Ghitalla, 2004). C'est ce que Ghitalla *et al.* (2003) nomment « l'outre-lecture » : la navigation envisagée dans l'espace documentaire non structuré (et hautement divers et fragmenté) du Web dans son entier se traduit en un parcours actualisé par le chaînage de fragments à travers des fenêtres multiples organisées selon des logiques variées. Pour Ghitalla (2000), qui insiste sur la rupture introduite par le document numérique, la navigation et la lecture se pratiquent en « homologie technique » avec la forme du réseau. Or, nous le verrons, non seulement un espace comme celui d'une bibliothèque numérique nationale et patrimoniale peut être considéré comme un « îlot de cohérence » au sein du Web, mais encore, l'accroissement en volume des espaces documentaires, de la qualité des documents et du confort de lecture, couplé à un accès internet en continu permettant de rester connecté sur une plateforme (pour zoomer sur des images qui ne sont pas téléchargeables en haute définition, pour feuilleter des documents...), font de la liseuse d'une bibliothèque numérique une véritable salle de lecture. Mais une salle de lecture d'un genre nouveau permettant une diversité de régimes d'engagement avec l'ensemble documentaire. Ainsi, sans présager de la nature des objets trouvés dans cet environnement, le même ensemble documentaire pourra alors être perçu, selon les pratiques, comme de l'information (*information*), du savoir (*knowledge*), une bibliographie (*documents*), une base de données (*data*)...

Plus récemment, du côté des sciences de l'information et de la communication (SIC), s'intéressant plus spécifiquement aux humanités numériques (Antonijević, 2015 ; Trace et Karadkar, 2017 ; Given et Wilson, 2018), d'autres études visent à croiser les méthodes empiriques vers une compréhension toujours plus intégrée du travail savant dans son extrême diversité de pratiques et de *media*, certains auteurs remettant au goût du jour le modèle de la cueillette (Ghitalla *et al.*, 2003 : 177-178), la distinction entre « affineurs » et « moissonneurs » (Roustan, 2013), ou bien la figure du « chasseur-cueilleur » (Le Marec et Mairesse, 2017). Le vocabulaire mécanique et psychologisant des études de la fin du XX<sup>e</sup> siècle a été remplacé au tournant des années 2020 par le « *scholarly workflow* » (Antonijević, 2020 ; Weiland, 2021) de chercheurs devenus « bricoleurs » (Antonijević et Cahoy, 2018). Ces pratiques et leurs apprentissages, les stratégies et les démarches de recherche soulèvent des questionnements importants : comprendre « comment pense un savant » (Bert, 2018) implique désormais d'élargir les objets traditionnels de l'histoire, la sociologie et l'anthropologie des savoirs.

Le présent projet s'inscrit dans cette tendance à travers le concept de *navigation* englobant de multiples pratiques – dont la recherche dirigée, à laquelle la navigation ne peut néanmoins se réduire. En effet, les pratiques de lecture et d'écriture ont beaucoup évolué au cours des dernières décennies : une analyse approfondie des journaux d'éditeurs savants menée par le groupe de recherche CIBER au Royaume-Uni a permis de décrire la lecture comme « *bouncing, flicking or skittering* » (Nicholas et Clark, 2012 : 95). La même année, une enquête mixte menée par Chérifa Boukacem-Zeghmouri sur les universitaires français en sciences, techniques et médecine a montré la « nature fondamentale de la navigation pour les chercheurs » (2012 : 89). Plus récemment, une vaste enquête internationale portant sur les pratiques informationnelles dans toutes les disciplines, a mis en évidence la prévalence écrasante de la navigation au sein de sites web uniques, notamment par chaînage de références, par rapport à la recherche par mots-clés (Tenopir *et al.*, 2019). Ainsi, bien que

traditionnellement associée à des disciplines herméneutiques comme les sciences humaines, l'accès de plus en plus fragmentaire à l'information que permettent les *media* numériques a progressivement fait de la navigation une pratique transversale.

Nous entendons ici la navigation comme une modalité d'appropriation *spatiale* d'une bibliothèque numérique par ses usagers, c'est-à-dire en tant qu'activité individuelle couplée à un support technique, permettant de comprendre l'articulation entre les « géographies subjectives de l'information » et la co-construction de « territoires communs » (Ghitalla, 2002). Après des usages immodérés avec la naissance d'Internet, les métaphores spatiales ont par la suite été maniées avec précaution, en raison de leur « naturalité » excessive soupçonnée de dénoter un territoire illusoire – le cyberspace – qui était alors remis en question (Cohen, 2007). Cependant, pour autant que l'on ne considère pas les espaces numériques comme notre environnement physique, mais que l'on s'intéresse à la manière dont ils sont construits par les pratiques et leurs représentations (Mounier, 2010), la perspective spatiale redevient pertinente. En particulier, si l'on envisage le numérique dans cette perspective, elle ouvre les trois dimensions de la bibliothèque physique à une multiplicité d'autres dimensions qui permettraient des sauts d'un « rayonnage » à un autre, autant que leur incessante reconfiguration. Dans quelle mesure peut-on filer cette métaphore et transposer par analogie les cheminements intellectuels de la bibliothèque dite physique à son avatar numérique ? Est-il pertinent de s'interroger selon ces catégories sur ce que veut dire « s'orienter dans la pensée » à l'ère numérique ?

Prenant pour étude de cas la plateforme Gallica, cette enquête pose donc la question de savoir si l'on peut reconsidérer un usage du numérique, et des bibliothèques numériques en particulier, comme l'arpentage d'un espace et selon quelles modalités. Cette recherche en cours part d'une compréhension *a priori* de la navigation pour en apprécier les effets sur l'interprétation des stratégies de recherche observées et, réciproquement, comment ces observations permettront d'épaissir *a posteriori* le concept de navigation. Une telle démarche d'ordre hypothéticodéductif entend faire preuve de l'intérêt heuristique du concept de navigation pour comprendre les pratiques d'orientation des lecteurs des bibliothèques numériques et esquisser une méthode mixte permettant de documenter celles-ci.

La pratique de la navigation apparaît comme un objet polymorphe résultant de pratiques matérielles d'appropriation d'un espace sémiotique historiquement situé. Elle appelle une diversité de perspectives et de méthodes d'enquête. Pour donner du sens aux multiples matérialités du numérique et aux pratiques savantes qui y sont liées, il est nécessaire de tisser ensemble une variété de points de vue théoriques dans un « chassé-croisé méthodologique » (Souchier *et al.*, 2003 : 29-30)<sup>4</sup> typique des méthodes mixtes, chaque perspective pouvant ainsi compléter les autres de manière réflexive et critique (Aguilera et Chevalier, 2021).

Basée sur une recherche exploratoire en cours, la présente contribution se propose donc d'ouvrir quelques pistes visant à montrer l'intérêt de documenter un type de pratique spatiale bien singulier : les pratiques de *navigation*. Le programme de recherche dont elle est tirée est structuré autour du croisement de trois perspectives : une *histoire* du développement de Gallica visant à restituer sa dimension politique ; une *étude sémiotique* de ses architectures, interfaces et codes ; une *enquête ethnographique*, à la fois qualitative et quantitative, des pratiques de navigation de ses usagers. Dans le cadre de cet article, et étant donné l'état exploratoire de cette recherche, les deux premières dimensions – historique et sémiotique – ne seront que brièvement abordées, avant de proposer un programme pour une ethnographie des pratiques de navigation sur Gallica.

---

<sup>4</sup> Les auteurs parlent également d'une « ethnosémiologique située » (Souchier *et al.*, 2003 : 26) qui refuse le « modèle fonctionnaliste "émission-réception" » (*Ibid.* : 19) – et donc la dichotomie *user-centered* / *machine-centered* à la base de la distinction entre les champs de *information seeking* et de *information retrieval*.

## 1. Gallica, brève histoire d'un espace sémiotique

Gallica est une des plateformes en ligne de la Bibliothèque nationale de France, mettant à disposition en accès ouvert plus de 10 millions de documents numérisés issus des collections du domaine public de la BnF. Ces documents de types variés – livres, manuscrits, presse, images, son... – dans 8 langues différentes sont consultables en ligne *via* le lecteur dédié (voir Figure 1) ou téléchargeables. Les publics de Gallica sont également très variés mais la plateforme est majoritairement utilisée par des généalogistes, des enseignants, des historiens amateurs, des chercheurs en sciences humaines et sociales, des archivistes et des bibliothécaires. Entre 2014 et 2018, une série d'études réalisée sous la direction de Valérie Beaudouin et Jérôme Denis (2014), croisant entretiens, vidéo-ethnographie et analyse de traces numériques, a mis en exergue la variété de ces profils usagers, des documents consultés, des pratiques de recherche, leur intrication avec des pratiques plus traditionnelles telle que la prise de note (Beaudouin *et al.*, 2016). Par le prisme de la navigation, nous nous plaçons dans le prolongement de ces études.

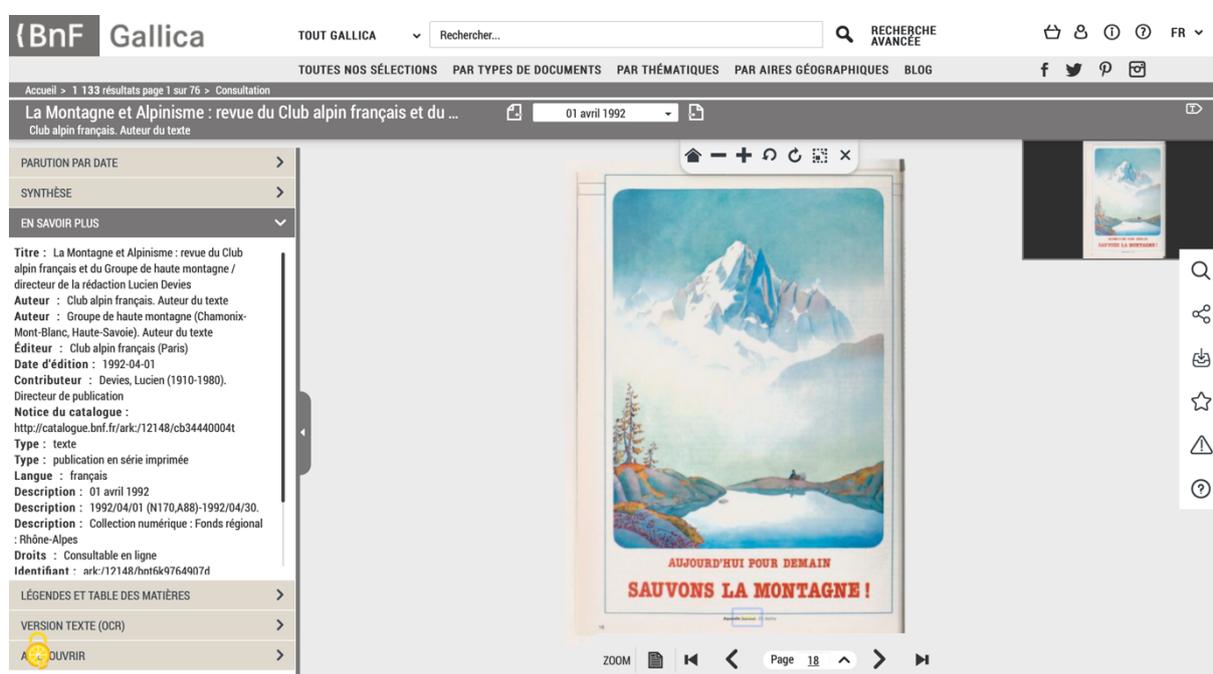


Figure 1. Capture d'écran de l'interface du lecteur de Gallica.

Créée en 1997, la plateforme Gallica prolonge dans le numérique et sur Internet la mission patrimoniale de la BnF, notamment de conservation et de mise à disposition de ses collections – bien que la notion de « collection numérique » n'apparaisse que tardivement (Bermès, 2020 : Deuxième partie). Ainsi, « la bibliothèque patrimoniale conditionne la bibliothèque numérique » (Béquet, 2015 : 16) et cela est patent pour Gallica dont l'architecture est tout orientée vers la consultation ou le téléchargement d'un ou plusieurs documents (Bermès, 2020 : 54 et *passim*). Parmi les nombreuses fonctions de la BnF physique – aire de travail, lieu de sociabilités, espace d'exposition, librairie... –, celle principalement numérisée par Gallica est la conservation et la mise à disposition de documents. En effet, lors d'une première phase de développement (1997-2003), tant pour des raisons de simplicité technique que patrimoniales – la nécessité de restituer tels quels les documents sans dissocier structure logique et structure matérielle – le choix a été fait de numériser et mettre à disposition les documents en « mode image », c'est-à-dire sous la forme de simples fac-similés numériques munis de leurs descripteurs historiques. À partir de 2004, suite à la critique élevée par le président de la BnF Jean-Noël Jeanneney à l'encontre de Google Books, il a été décidé de numériser les documents *en masse* – par simple ordre de cotes – et d'océriser systématiquement

afin de pouvoir également les indexer en plein texte et doter Gallica d'un moteur de recherche. Outre renforcer l'aspect « dépôt » du dispositif – plutôt que de constituer des collections ou développer de nouveaux descripteurs et outils, par exemple –, cette opération a permis d'ouvrir une nouvelle échelle de consultation : il est désormais possible d'extraire une information à l'échelle de la phrase ou du paragraphe.

Par opposition à l'*usager* qui s'approprie activement la plateforme, l'*utilisateur* imaginé lors de la conception et sédimenté lors des évolutions successives dans l'architecture du site est le lecteur d'une bibliothèque physique dont le comportement est envisagé sur le mode du rapport aux bibliothécaires. Si l'on considère brièvement Gallica comme un espace sémiotique dont se saisissent les usagers à travers des éléments architextuels – barre de recherche, menus déroulants, billets de blogs et sélections – qui leur donnent des prises et appuis pour se déplacer dans l'interface du site, il apparaît que, de lien en lien, tout l'espace est orienté vers les documents (voir Figure 2). Depuis l'interface du lecteur, les actions possibles sont restreintes car le document et ses métadonnées ne sont pas cliquables : il est possible de parcourir le document, de passer à un autre volume ou numéro dans le cas d'œuvres à tomes ou de la presse ou bien de quitter Gallica vers la notice du Catalogue de la BnF – à noter que, depuis peu, un module « à découvrir » et une carte géographique proposent quelques suggestions sur un mode expérimental. En conséquence, afin de passer d'un document à un autre, les usagers reviennent très souvent à la page de résultats ou bien lancent une nouvelle requête dans le moteur.

En tant que bibliothèque numérique, Gallica permet donc de formuler la demande d'une cote précise ou de (faire) réaliser une enquête bibliographique avant de consulter les documents afférents mais ne permet pas, par défaut, la libre navigation entre ses « rayonnages ». Une liberté de navigation cependant toute relative à la BnF physique puisque les magasins ne sont pas ouverts au public et, comme l'a souligné Elizabeth Losh, « bien que la bibliothèque virtuelle de la BnF reflète les idéaux libéraux d'anonymat et de libre accès, l'espace physique est l'une des bibliothèques de recherche les plus surveillées au monde » (2009 : 261). En d'autres termes, il s'agit d'un espace préformaté au sein duquel il est possible de *trouver* un contenu dont l'on présuppose l'existence, mais plus compliqué de *découvrir* des contenus inattendus dans un espace ouvert – il s'agit là d'une distinction fondamentale entre *findability* et *discoverability*. À ce titre, il est frappant qu'une enquête dise donner des formations à Gallica en présentant la plateforme comme un « Google du passé » (E7). Cette polarisation de la plateforme, due à sa nature, sa fonction et son histoire, a également été documentée quantitativement grâce à l'analyse de traces par Nouvellet *et al.* (2017), qui ont montré une écrasante majorité de « parcours d'usage » centrés sur la consultation, et plus récemment par Trabelsi (2022) dont l'analyse par fouille de processus a mis en exergue les activités de consultation ou de téléchargement comme aboutissement desdits processus.

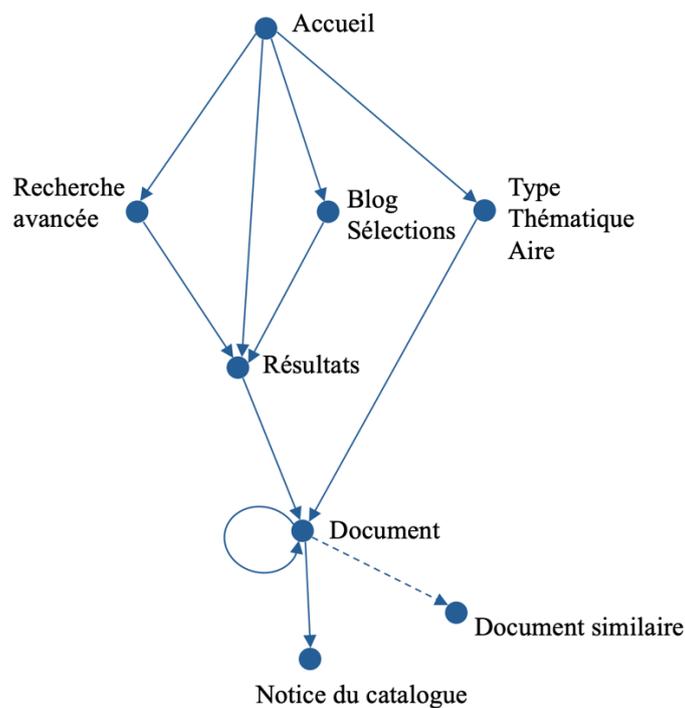


Figure 2. Schéma simplifié de l'architecture de Gallica.

Ainsi, contrairement au Web, le « milieu documentaire » qu'est Gallica est un espace structuré et balisé où, selon les régimes d'engagement, le document peut être considéré comme unité pertinente, identifié par une « notice d'autorité »<sup>5</sup>. C'est qu'au sein du Web, Gallica incarne une forme d'autorité qui en fait un « médiateur institué » (Boullier et Ghitalla, 2004 : 182) assurant la confiance et la certitude dans ce que l'on y trouve, et servant ainsi de repère pour s'orienter – d'ailleurs certains enquêtés mentionnent que leur expérience de Gallica est informée par leur pratique de la bibliothèque physique. Mais, si l'héritage historique, la mission patrimoniale, l'architecture informationnelle et l'assise institutionnelle font de Gallica un « îlot de cohérence » (Boullier et Ghitalla, 2004 : 174), il n'est néanmoins pas impossible d'intégrer l'incertitude dans son usage : nous verrons qu'un certain régime exploratoire est permis justement par (le détournement de) la structure de l'espace documentaire.

Il est important de noter cela car en s'intéressant à la navigation, nous allons tenter d'observer des pratiques qui ne sont *a priori* pas anticipées par Gallica et qui résulteront, lorsque qu'elles s'écarteront de l'extraction d'information ou de la consultation d'un corpus, de formes d'appropriation ou de détournement de la plateforme par des usagers actifs. Dans la suite de cette contribution, lorsque nous observerons ces pratiques, celles-ci ne résulteront donc pas d'un algorithme mais bien de diverses modalités d'appropriation de la plateforme par ses usagers. Contraints par la nature et l'architecture de Gallica, nous verrons que les lecteurs, à partir des mêmes actions de base – retourner à la page de résultat ou requêter à nouveau le moteur –, développent une variété de stratégies que la plateforme ne permet pas par défaut, et qui répondent à leurs besoins en fonction des phases de leur activité. Leurs témoignages ne refléteront pas une distinction entre pratique et théorie, ou entre démarche réelle et démarche fantasmée, pas plus qu'il n'y a de contradiction entre la plateforme et les discours des enquêtés. Ces discours sont à envisager plutôt comme des tentatives d'objectivation de stratégies qui résultent de la combinaison des actions de bases mais sont mal représentées par celles-ci.

<sup>5</sup> Pour cette raison, Ghitalla (2002) refuse d'envisager le Web comme une bibliothèque universelle, cette dernière étant en effet « un dispositif de gestion centralisée de l'information que l'on pourrait parcourir à l'aide d'instruments de classement omniscients » (Ghitalla, 2002 : 67).

## 2. Ethnographie des pratiques de navigation

La partie ethnographique de cette étude envisage la navigation comme un cheminement intellectuel à travers un corpus rendu accessible par une plateforme en ligne. Notre étude de cas s'intéresse aux pratiques de navigation d'utilisateurs de Gallica en combinant une approche qualitative – entretiens semi-directifs et mises en situation – avec une approche quantitative – analyse de logs serveurs. Dans un dernier temps, nous avons tenté de rapprocher les parcours mentionnés par les utilisateurs aux parcours extraits des logs.

### 2.1. Méthodologie qualitative : une ethnographie du numérique<sup>6</sup>

Une série d'entretiens a permis de porter une attention particulière aux capacités des enquêtés à objectiver leurs propres pratiques, ainsi qu'aux métaphores utilisées pour les décrire. De plus, lors de la mise en situation, l'écart entre le discours et la performance, plutôt que l'incohérence, a constitué un élément important pour évaluer la réflexivité pratique des usagers.

Les enquêtés ont été sélectionnés parmi les répondants au grand observatoire organisé par la BnF à l'automne 2020. Nous avons contacté 20 personnes en filtrant les réponses au questionnaire afin d'isoler des usagers qui pratiquent la recherche en humanités et sciences sociales comme activité principale ou professionnelle et qui ont déclaré consulter régulièrement Gallica avec des temps de connexion supérieurs à 10 mn. Nous avons reçu 7 réponses positives qui ont mené à autant d'entretiens qui ont été réalisés en ligne, enregistrés puis retranscrits. L'échantillon est très homogène car les 7 enquêtés, 4 femmes et 3 hommes, sont historiens, archivistes, conservateurs, bibliothécaires, parfois assument plusieurs de ces rôles à la fois. Ce sont donc des usagers experts de leurs domaines respectifs, ainsi que de la recherche documentaire mais ce sont majoritairement des novices des systèmes numériques, formés en autodidactes.

Les entretiens semi-directifs menés sur Zoom ont été réalisés sous la forme d'« entretiens épisodiques [*episodic interviews*] » (Mueller, 2019) visant à faire objectiver sous une forme narrative une session de recherche. Ils se sont déroulés en quatre étapes concentriques, allant progressivement de l'interface numérique vers le bureau physique : après un bref profil socio-professionnel, nous progressions depuis l'arrivée de l'utilisateur sur Gallica, à la description de séances de lecture plus longues, des liens pratiques et intellectuels faits entre les documents consultés, des phases de la recherche et de leur éventuelle morphologie (afin de provoquer l'usage de métaphores). Puis nous élargissions la discussion à d'autres plateformes en ligne et, enfin, au bureau physique et ses pratiques de papier. En dernier lieu, nous mettions les enquêtés en situation en leur demandant de partager leur écran afin de faire la démonstration d'une recherche type. Ces entretiens ont duré entre 1h et 1h30.

Quelques remarques générales :

Tous nos enquêtés travaillent « en parallèle », *i.e.* avec plusieurs onglets, à travers plusieurs plateformes et tout en prenant des notes papiers ou en consultant des ouvrages physiques<sup>i</sup>. La navigation n'est donc ni séquentielle, ni linéaire, et ressemble plutôt à des coups de sonde lancés dans plusieurs directions plus ou moins dispersées. Nous n'en saisissons donc qu'un aspect étroit en nous limitant à Gallica.

Les témoignages sont toujours extrêmement riches, notamment dans l'usage qui est fait de nombreuses métaphores spatiales<sup>ii</sup>, temporelles<sup>iii</sup>, physiques<sup>iv</sup>, renvoyant au divertissement<sup>v</sup> ou à la découverte<sup>vi</sup>, pour tenter d'objectiver des pratiques qui n'ont *a priori* pas de représentations communes. Il est important de prendre ces métaphores au sérieux. D'abord car elles ne sont ni spécifiquement numériques, ni exclusivement spatiales mais qu'elles ont en commun le mouvement, le cheminement ou la démarche qui sont des *topoi* dans l'histoire des savoirs. Il suffit de rappeler l'étymologie de *methodos* – la poursuite d'un chemin (Marion,

---

<sup>6</sup> Le protocole d'enquête et les transcriptions des entretiens ont été déposés sur la plateforme Nakala : <https://doi.org/10.34847/nkl.320eu55q>

2000) –, l'étude de la mètis comme démarche par Detienne et Vernant (1974) ou bien l'apparition à l'époque moderne du voyage comme « métaphore de l'activité de connaissance et du développement de la science empirique » (Bourguet, 2016). D'autre part car, si ces métaphores ne décrivent pas un territoire réel, elles décrivent des pratiques savantes réelles que nous souhaitons documenter. En effet, à abandonner toute idée de mouvement, de circulation, de démarche, l'on cède à l'empire de la requête (Auray, 2010), *i.e.* un extractivisme informationnel comme seule modalité d'accès au savoir, par morceaux, alors qu'il s'agit de valoriser une pluralité de pratiques savantes qui, depuis des siècles reposent sur les allées et venues, les envois et renvois, la navigation comme une ligne d'erre au sein d'une masse documentaire partagée.

Malgré cela, la plupart des enquêtés rencontrent des difficultés à objectiver leurs pratiques et cherchent donc à justifier *a posteriori* le résultat de cette opération d'objectivation<sup>7</sup>. Ces difficultés sont parfois explicitement mentionnées<sup>vii</sup>, parfois lisibles dans les tensions entre idiosyncrasie et généralité<sup>viii</sup>, entre organisation et désorganisation<sup>ix</sup>, entre doute et certitude<sup>x</sup>. D'autre part, la confrontation entre le discours et la pratique lors de la mise en situation permet de mettre en lumière des oublis ou des contradictions internes au discours<sup>xi</sup>. Ces observations soulignent la grande complexité de l'activité de recherche, qui n'est pas réductible à des comportements mécaniques et nécessite une pluralité d'approches complémentaires.

## 2.2. Méthodologie quantitative : une ethnographie numérique<sup>8</sup>

Parallèlement aux entretiens, l'étude quantitative, basée sur l'analyse de traces, a visé à identifier et modéliser des sessions moyennes et longues d'utilisateurs. Cette observation non participante induite à partir des traces de navigation générées par les utilisateurs s'est faite grâce à l'analyse de trois mois de logs serveurs de Gallica datés du printemps 2016 et formatés selon la structure détaillée en Figure 3 – il s'agit du même jeu de données qu'analysé par Nouvellet *et al.* (2017) ainsi que Trabelsi (2022). Notons que les traces étant pseudonymisées – les IP sont hashées –, il est possible de distinguer des utilisateurs uniques mais impossible de les suivre sur plusieurs sessions ou de sélectionner des profils singuliers, par exemple d'extraire les chercheurs.

En première approximation, et dans l'objectif de représenter la non-linéarité des démarches parallèles objectivées par les enquêtés, les parcours ont été modélisés sous la forme de séries de plus de 3 documents requêtés par une même IP, non consécutivement redondants, séparés de moins de 60 mn. Ces parcours ont ensuite été enrichis grâce aux métadonnées – auteur, titre, année, classe Dewey – renvoyées par l'API Gallica sur requête d'une ARK (Archival Resource Key). En restreignant ainsi les parcours aux sessions qui ne contiennent que des documents munis de classes Dewey, nous limitons donc le champ de l'enquête à la consultation d'imprimés, d'estampes et de certains manuscrits.

Bien qu'ils soient modélisés comme des séries, il ne faut cependant pas entendre les parcours comme des séquences linéaires car les deux étapes suivantes permettent d'établir des relations entre documents qui n'ont pas été requêtés consécutivement – permettant ainsi de

---

<sup>7</sup> L'entretien peut ainsi être entendu comme une « épreuve » au sens de la sociologie pragmatique.

<sup>8</sup> Cette partie de l'étude a été réalisée en partenariat avec Bayrem Kaabachi, étudiant en informatique à l'EPFL. Le code est disponible sur le Github [https://archive.softwareheritage.org/browse/origin/?origin\\_url=https://github.com/LHST-EPFL/TDA-Gallica](https://archive.softwareheritage.org/browse/origin/?origin_url=https://github.com/LHST-EPFL/TDA-Gallica). Pour une présentation détaillée du *pipeline* de traitement des logs serveurs de Gallica, voir Kaabachi et Dumas Primbault (2023).

Depuis les années 2010, les sciences des données ont concentré leurs efforts sur la seule recherche dirigée – *i.e.* le chemin le plus court et le plus direct, l'extraction d'information. Ces études sont majoritairement consacrées à l'optimisation pratique et à la résolution de problèmes – visant à améliorer le référencement des pages trouvées par la « navigation post-requête » (Singla *et al.*, 2010), à prédire la page suivante sur la base de régularités statistiques (Koopman *et al.*, 2019), ou à suggérer les « sentiers » les plus battus (White et Huang, 2010). Nombre de ces études qui prétendent anticiper le besoin des utilisateurs vont jusqu'à se débarrasser des pratiques de navigation. Ces études escamotent donc une pratique que d'autres auteurs du champ ont pourtant soulignée comme fondamentale car le « moteur de recherche parfait ne suffit jamais » et les utilisateurs naviguent toujours entre deux requêtes (Teevan, 2004), tandis que les algorithmes de suggestion existants ne favorisent pas la navigation, voire l'entravent dans certaines circonstances (Lamprecht, 2015).

restituer, par exemple, une pratique typique consistant à ouvrir un certain nombre d'onglets à partir de la liste des résultats de recherche.

Dans un premier temps, les parcours ont été transformés en chaînes de Markov de classes Dewey, d'ordre 3, *i.e.* en marches aléatoires d'une discipline Dewey à une autre. Ceci a permis de calculer les probabilités de transition d'une classe à une autre au cours d'une session. Dans un second temps, ces chaînes de classes Dewey ont permis de calculer les distances entre paires de classes afin de générer un espace dans lequel deux classes (par exemple, Psychologie et Logique) sont d'autant plus proches qu'elles ont été requêtées lors d'une même session par un grand nombre d'utilisateurs. Enfin, dans cet espace au sein duquel les utilisateurs se meuvent, les parcours ont été individuellement dessinés afin d'être regroupés en *clusters* selon leurs propriétés topologiques et repérer divers régimes de navigation correspondants à différents motifs morphologiques – des parcours ponctuels ou linéaires, en étoile ou en zigzag.

Chacune des étapes de ce *pipeline* soulève évidemment son lot de questionnements méthodologiques et épistémologiques – une présentation technique est détaillée dans Kaabachi et Dumas Primbault (2023) ; pour une analyse critique, voir Dumas Primbault (2024). Notons seulement que les entretiens jouent le rôle de garde-fou dans la manipulation et l'analyse des traces, la construction du modèle et l'interprétation des représentations visuelles.

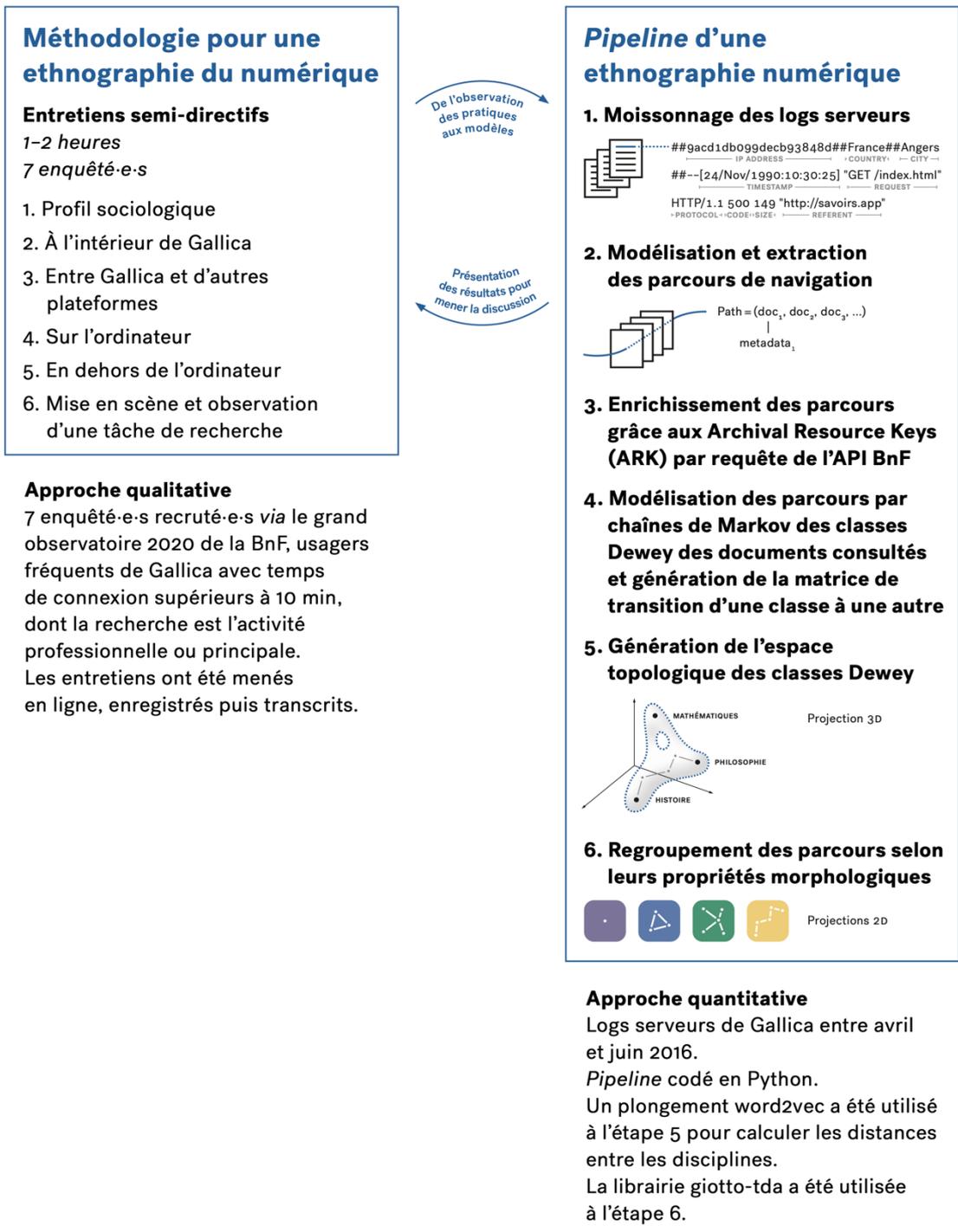


Figure 3. Schéma du chassé-croisé méthodologique entre approches ethnographiques qualitative et quantitative

### 3. Des parcours et régimes de navigation variés

#### 3.1. Classification Dewey et « littérature pivot »

La modélisation des parcours de navigation comme des marches aléatoires entre classes Dewey a permis de calculer les probabilités de naviguer d'une classe à une autre au cours d'une seule session. Ces probabilités sont illustrées par la matrice de transition de la Figure 4. À gauche sont les classes de départ – de « 000 Ouvrages Généraux » à « 990 Histoire de l'Océanie

et des pôles » –, en bas les classes d'arrivée. Plus une case est foncée, plus elle dénote une probabilité élevée de naviguer de la classe de départ à la classe d'arrivée.

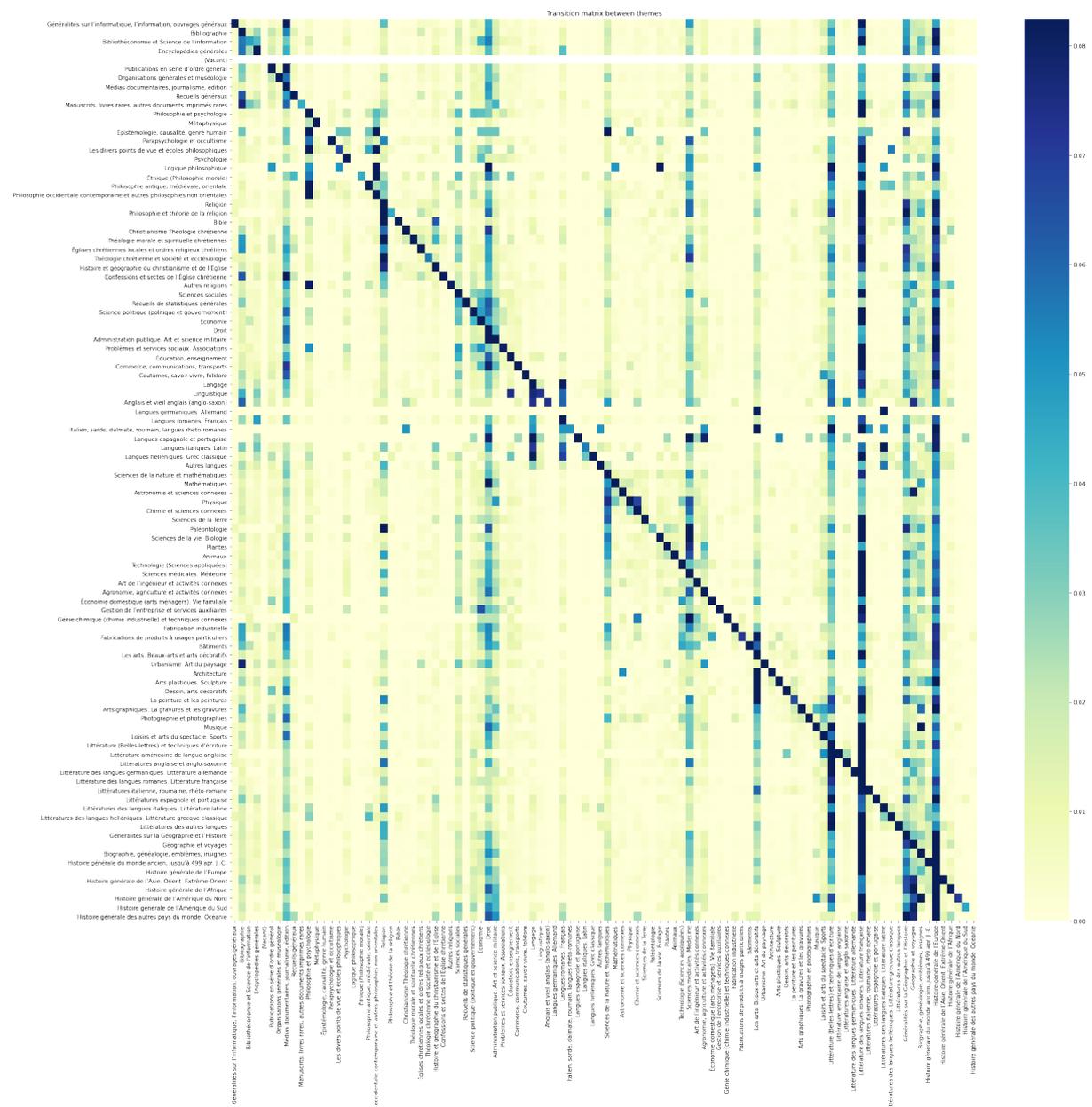


Figure 4. Matrice de transition d'une classe Dewey à une autre

La diagonale très foncée de cette matrice montre que la probabilité la plus élevée, et de façon significative, est de rester au sein d'une même classe Dewey. Au cours d'une session, les lecteurs de Gallica ont tendance à rester dans une seule classe Dewey ou, dans une moindre mesure, à s'aventurer dans les classes voisines – ce sont les « bavures » autour de cette diagonale – donc des thématiques voisines. Ainsi, par défaut, le numérique ne favorise pas plus l'interdisciplinarité et si, au sein de Gallica, la navigation est un braconnage car elle se pratique à l'ombre de l'architecture et de l'interface de la plateforme, elle connaît néanmoins des clôtures disciplinaires.

Les colonnes plus foncées sont de deux ordres. La probabilité de transition vers certaines classes comme « 340 Droit », « 610 Sciences médicales, Médecine » ou « 840 Littérature française », est très élevée quelle que soit la classe de départ, tandis que la probabilité de transition vers une autre classe depuis celles-ci est très faible. Ce sont donc des thématiques vers lesquelles les usagers convergent pour y rester. Elles dénotent surtout le déséquilibre de la

classification Dewey : ce sont toutes des sous-classes pour des disciplines pourtant importantes. Cela permet en partie de comprendre le désintérêt de nos enquêtés pour cette classification.

Pour d'autres classes comme « 900 Généralités sur la Géographie et l'Histoire », « 940 Histoire générale de l'Europe » ou « 070 Médias documentaires. Journalisme, édition », les probabilités de transition *vers* et *depuis* celles-ci sont élevées. Ces classes permettent donc aux usagers d'opérer des transitions d'une thématique vers une autre. Cette « littérature pivot » dans les parcours de navigation rassemble des généralités, la presse ou, dans une moindre mesure, des recueils et des encyclopédies.

### 3.2. Régimes de navigation

Tous nos enquêtés mentionnent diverses « phases » (E4) ou « vitesses » (E1) dans leur activité de recherche, correspondant à la mise en œuvre de pratiques documentaires diverses<sup>xii</sup>, selon des temporalités plurielles et des rapports différents aux clôtures disciplinaires. C'est ce que nous avons souhaité qualifier en identifiant des formes de la navigation au sein d'un espace défini par la proximité entre classes Dewey selon qu'elles sont régulièrement requêtées au cours d'une même session ou non. Cette étude morphologique des parcours a permis de mettre au jour quatre régimes de navigation distincts. Ces « régimes d'engagement » divers avec l'espace documentaire, qui sont aussi des régimes attentionnels différents, peuvent se succéder au cours d'une session, notamment du fait d'une rupture dans le parcours (intérêt pour un autre sujet, suggestion, résultats non pertinents...), et permettent de saisir la masse documentaire selon des modalités diverses : comme information, document, savoir, ou donnée. Plutôt que des parcours comme l'on pourrait en faire entre les rayonnages d'une bibliothèque, ces régimes idéaux-typiques qualifient et représentent l'étalement disciplinaire – des « aires » de navigation – en fonction des phases de la recherche, et la morphologie de cet étalement. Les rapprochements ici effectués entre témoignages et logs serveurs devront être consolidés par une analyse des traces générées par nos enquêtés-mêmes.

Le régime le plus évoqué par les enquêtés est celui de la recherche « dirigée » (E7), « ponctuelle » (E4), « pointue » (E1) ou « très précise » (E2) qui correspond à des « sessions courtes » (E3), dans l'objectif de trouver une information factuelle précise dans un document quelconque ou faire une lecture plus poussée (copie, notes de lecture) d'un document en particulier. Ce régime est à rapprocher du *cluster* illustré en Figure 5, qui correspond à des parcours de moins de 5 documents en moyenne, qui peuvent s'étaler sur plusieurs classes puisque c'est l'information ou le document final qui orientent la recherche. Cette recherche dirigée se fait sur le mode du « régime du plan » car elle est téléologique, elle suppose l'information existante et l'espace suffisamment bien indexé pour la retrouver (*retrieve*) et se couple donc bien, côté machine, à une ingénierie documentaire (*information retrieval*) structurée autour d'un moteur de recherche. Néanmoins, la recherche dirigée est nécessairement précédée d'un motif de découverte (Gilliotte, 2022) – bouche à oreille, référence bibliographique, éditorialisation... – rarement explicité par les enquêtés.

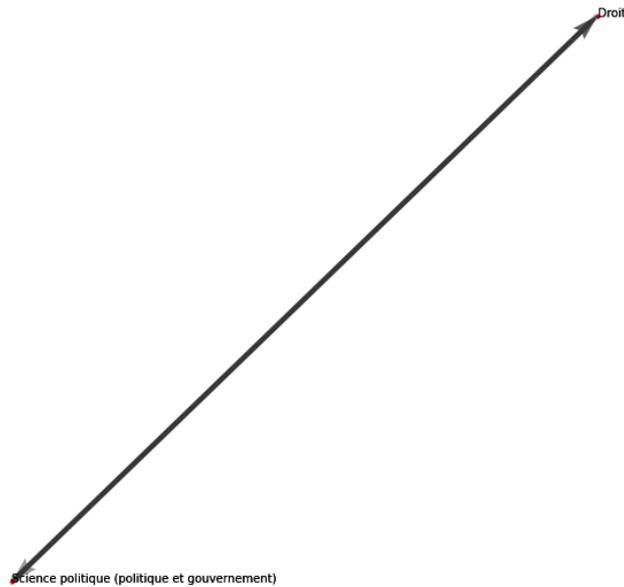


Figure 5. Exemple de recherche dirigée entre droit et science politique.

Un second régime correspond à une forme de navigation plus étendue, parfois comparative mais souvent strictement disciplinaire, dans l’objectif de « constituer » ou « consulter » tout un corpus. Il s’agit de parcours plus longs en nombre de documents requêtés mais restreints à une classe ou une sous-classe (voir Figure 6). Il s’agit d’une autre forme du « régime du plan » pour laquelle l’unité documentaire est bibliographique : il s’agit du corpus.

Langues romanes. Français

Figure 6. Exemple de constitution ou consultation d’un corpus en langues romanes.

Le régime le moins objectivé par nos enquêtés, parfois même explicitement négligé (« je ne suis pas une touriste de Gallica » (E7), « je vais pas à la pêche » (E6)<sup>9</sup>), mais qui est pourtant très souvent mentionné, comme en passant, au détour d’une phrase, est une forme de « papillonnage<sup>xiii</sup> », « vagabondage<sup>xiv</sup> », « jeu » (E5), « hasard<sup>xv</sup> », ou de « sérendipité<sup>xvi</sup> ». Les enquêtés caractérisent ce régime de recherche – souvent dénommé recherche exploratoire dans

<sup>9</sup> Ainsi, c’est aussi dans le refus d’un tel régime d’engagement que certains enquêtés se construisent une position de recherche distincte. De manière générale, les enquêtés ont valorisé les régimes d’efficience qu’il est possible de rationaliser *a posteriori*.

la littérature – par des temporalités très différentes (« temps-libre » ou « entre-temps » (E5)), souvent plus longues (des heures, voire des jours). Il s’agit d’une stratégie contextuelle de cheminement pas à pas, qui n’est ni déterministe ni dirigé (voir Figure 7), comme l’indique la métaphore du « rebond » (E4), également relevée par Beaudouin *et al.* (2016). Proche du modèle de « *berry-picking* » (Bates, 1989), cette déambulation sur le mode du « régime de la carte » peut s’envisager de manière téléologique par la redéfinition du besoin en fonction du contexte, de manière semi-téléologique comme une forme d’« attention flottante » propre à la sérendipité (Auray, 2017), ou bien comme activité en soi, afin de dessiner la carte ou comme simple source de plaisir. À noter que ce régime nécessite non seulement des « compétences topologiques » permettant de se repérer dans l’espace informationnel (Auray, 2017), mais également du temps libre et l’absence de contraintes de performance (pertinence ou efficacité) souvent imposées par un cadre professionnel (une commande, une date limite de rendu...). De plus, dans le cas de Gallica, il faut également savoir jouer avec les normes de la recherche par requête afin de détourner de son usage le moteur de recherche.

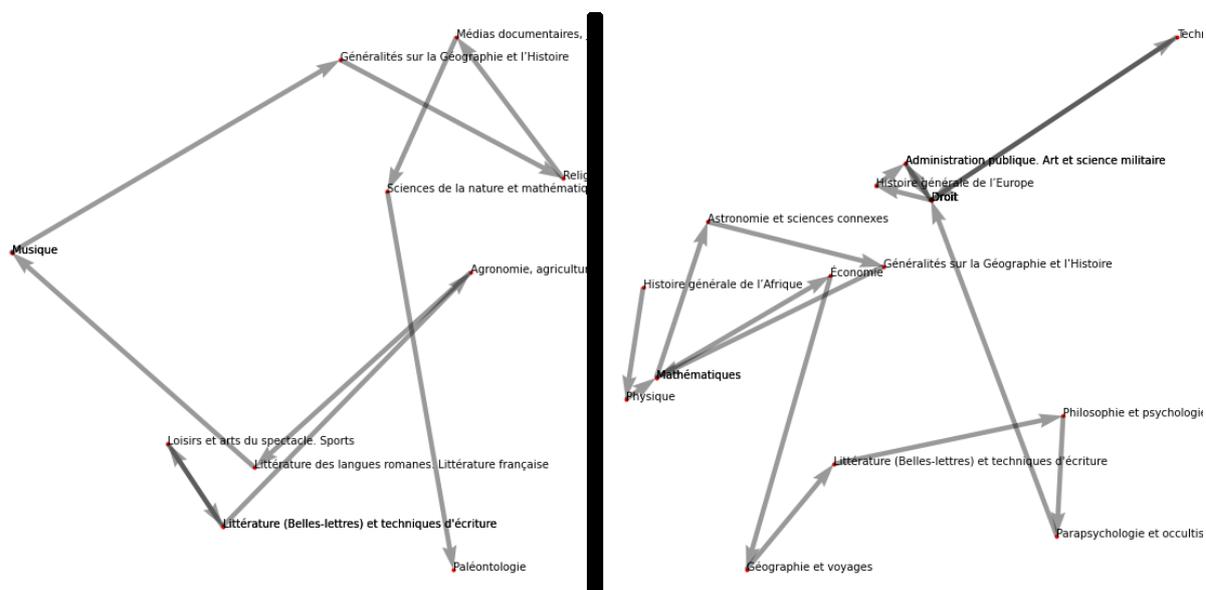


Figure 7. Exemples de recherches exploratoires.

Enfin, un régime intéressant, mentionné par un seul de nos enquêtés, pourrait correspondre à trois *clusters* rassemblant des parcours longs – en moyenne une vingtaine de documents – sur de très nombreuses disciplines – plus de dix – et structurés en forme d’étoile autour d’un ou plusieurs pôles thématiques à partir desquels les usagers rayonnent (voir Figure 8).

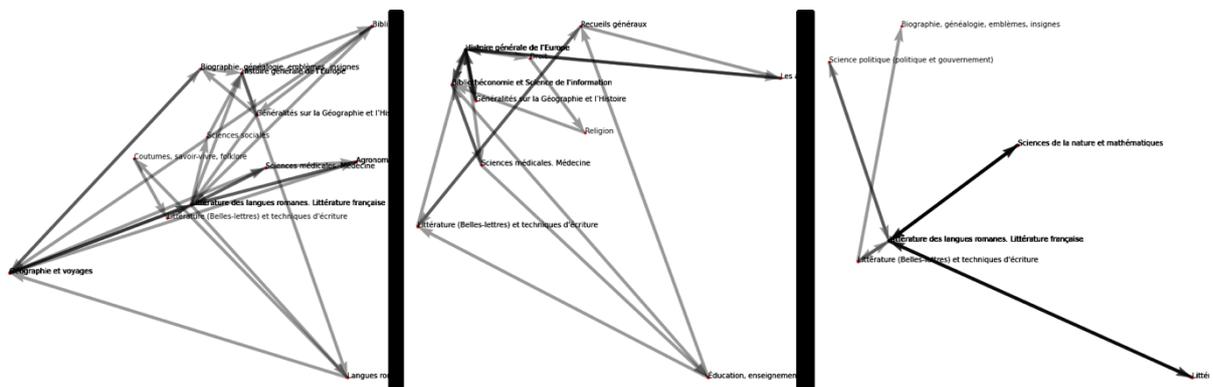


Figure 8. Exemples de navigation en étoile.

L'aboutissement, temporaire, de l'exploration est souvent évoqué comme une rupture dans le cheminement, un « hop » (E4), un « embranchement » (E3) qui produit l'effet de surprise, de trouver ce que l'on ne cherchait pas. L'objectif étant de provoquer un dépaysement qui ne souffre pas le cloisonnement disciplinaire ou thématique :

Très souvent les idées ne se suivent pas, c'est-à-dire c'est un peu le plan en étoile, ce n'est pas parce qu'on part d'un document et ensuite qu'on passe à un autre qu'il y a forcément un lien logique, si dans une recherche on sait exactement ce qu'on veut chercher et qu'on sait exactement ce qu'on va trouver, c'est que la recherche ne valait pas le coup, ça veut dire que tout est trop linéaire, quand c'est trop linéaire, pour moi très souvent c'est qu'il y a un souci, ça veut dire que soit ça a déjà été fait, soit c'est quelque chose qui reste une évidence, alors que si on le fait en mode en étoile où les idées, les recherches n'ont pas forcément un lien et qu'on arrive à trouver le liant, c'est là qu'on arrive à voir des embranchements qui peuvent être originaux. (E3)

Cette citation illustre l'articulation entre, d'une part, une plateforme comme condition de navigation et, d'autre part, les modalités d'appropriation pratique qui font advenir la navigation au sein de ladite plateforme. Cela condense l'idée d'un cheminement comme frayage dans une masse documentaire structurée, autant que l'insuffisance de la recherche dirigée seule pour provoquer des « embranchements originaux ». S'agit-il d'une recherche exploratoire méthodique ? De recherches interdisciplinaires ? De lectures croisées ? Ces *clusters* pointent vers la nécessité de raffiner notre approche afin d'obtenir des résultats plus robustes et interprétables, des idéaux-types qu'il conviendra également de remobiliser en contexte afin de retrouver l'irréductible hétérogénéité et complexité des comportements (Gilliotte, 2022 : 204).

## Conclusion

À ces quatre régimes, nous pourrions en ajouter un cinquième que nos enquêtés ne pratiquent pas et que notre approche quantitative a écarté mais qu'elle met pourtant en branle : en requêtant de manière automatique l'API de Gallica afin de récupérer en masse les métadonnées de documents, nous avons fait usage de la bibliothèque numérique comme d'une base de données. Il s'agit d'un régime spécifique, de l'ordre du plan ; un plan rigide encodé dans un algorithme et médié par un programme. L'unité documentaire est alors celle de la donnée, ce qui nécessite à nouveau des compétences spécifiques, différentes de l'expertise du domaine<sup>10</sup> ou de la recherche d'information.

Ainsi, au-delà de la tendance à la *datafication* de toute chose, les pratiques et les régimes d'engagement dont elles découlent permettent de saisir des rapports multiples au contenu documentaire d'une bibliothèque numérique. C'est dans le régime de recherche dirigé vers une information (et non un document) que le document est fragmenté et que Gallica devient un « Google du passé » : on va chercher une information, soit au niveau du paragraphe (grâce à l'OCR), soit au niveau de la page (grâce à la visualisation mosaïque). *A contrario*, pour la déambulation ou l'exploration, la structuration de l'espace numérique selon des architectures héritées de la bibliothèque (tant institutionnellement qu'individuellement) rend pertinente la notion de corpus et de parcours à travers celui-ci (plutôt qu'un démembrement total et dispersion du sens). L'espace documentaire de Gallica oscille donc, selon les pratiques et régimes d'engagement, entre espace topologique (celui du Web) où l'information est fragmentée et espace topographique (informé par celui de la bibliothèque) où le document peut être pris comme unité pertinente (Dumas Primbault, 2024).

Le concept de navigation ainsi formalisé est pertinent pour décrire les pratiques de recherche, tant des chercheurs que nous avons interrogés que d'un plus large public que nous

---

<sup>10</sup> D'ailleurs souvent rhétoriquement construite en décalage par rapport à celui-ci dans l'opposition « *domain knowledge* » / « *technical skills* » (Ribes *et al.*, 2019).

avons capté dans les logs. Les métaphores spatiales, physiques et dynamiques invoquées par nos enquêtés, qui toutes reposent sur le cheminement ou la démarche, attestent de ces pratiques de navigation et des représentations de ces pratiques, parfois informées ou formatées par l'espace physique de la bibliothèque<sup>xvii</sup>. Réciproquement, nous avons pu présenter certains parcours choisis à nos enquêtés qui ont toujours pu interpréter ceux-ci comme des démarches pertinentes. Les dynamiques intellectuelles à l'œuvre dans ces pratiques restent à documenter, certains enquêtés évoquant une approche comparative<sup>xviii</sup>, des « associations d'idées<sup>xix</sup> », le « curiosité<sup>xx</sup> », ou simplement le hasard<sup>xxi</sup>.

Cette étude exploratoire a également conduit à des résultats prometteurs sur la navigation des chercheurs au sein de la classification Dewey, sur le rôle de la « littérature pivot » permettant de passer d'une discipline à une autre, et sur une première esquisse d'une variété de « régimes de navigation » regroupés en fonction de leurs caractéristiques topologiques. Il conviendra par la suite d'étendre l'étude à d'autres plateformes et d'autres types de documents, notamment grâce à d'autres descripteurs que les classes Dewey ; d'approfondir les axes historique et sémiotique tout en liant plus étroitement ethnographies qualitatives et quantitatives – en raffinant le modèle et en analysant les données des enquêtés – ; d'élargir le recrutement à des publics plus larges – professionnels autres, militants, amateurs...

Ces premiers résultats ont néanmoins permis de commencer à documenter l'intrication des pratiques matérielles et des architectures informationnelles, ainsi que les stratégies de navigation développées par les chercheurs qui témoignent de leur volonté de trouver « des embranchements originaux ». Nous avons pu mettre en lumière la non-linéarité, l'indétermination des pratiques savantes et la nature fondamentale de la navigation comme pratique d'orientation. En tant que pratique matérielle menée à travers les bibliothèques, la navigation ne doit pas être entendue comme une errance passive et sans contrainte dans des espaces infinis mais plutôt comme un exercice actif d'orientation dans les savoirs, un arpentage contextuel qui travaille avec le cadre imposé par une ou plusieurs plateformes (Ghitalla, 2002). Il s'agit d'une pratique diachronique, mais pas simplement séquentielle, visant à trouver des relations contextuelles entre les documents en fonction de leurs types, sujets, auteurs, disciplines ou méthodes : en passant d'un document à l'autre, les chercheurs apprennent à dessiner pour elles-mêmes un paysage informationnel en regroupant les travaux rencontrés sur leur chemin en disciplines, périodes ou approches.

## Remerciements

Cette étude exploratoire a été financée par une bourse Collaborative Research On Science and Society (CROSS) 2020 attribuée par l'EPFL et l'Université de Lausanne.

La seconde phase de la recherche a débuté en octobre 2022 grâce à une bourse Mark Pigott en humanités numériques, décernée par la BnF.

## Bibliographie

Aguilera, T. et Chevalier, T. (2021). Les méthodes mixtes : vers une méthodologie 3.0 ? *Revue française de science politique*, 71, 3, 361-63. <https://doi.org/10.3917/rfsp.713.0361>

Antonijević, S. (2015). *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production*. Palgrave MacMillan.

Antonijević, S. (2020). Digital Workflow in the Humanities and Social Sciences: A Data Ethnography. Crowder, J.W., Fortun, M., Besara R. et Poirier, L. (dir.), *Anthropological Data*

*in the Digital Age: New Possibilities – New Challenges*. Palgrave MacMillan, 59-83. [https://doi.org/10.1007/978-3-030-24925-0\\_4](https://doi.org/10.1007/978-3-030-24925-0_4)

Antonijević, S. et Cahoy, E.S. (2018). Researcher as Bricoleur: Contextualizing humanists' digital workflows. *Digital Humanities Quarterly* 12, 3. <http://www.digitalhumanities.org/dhq/vol/12/3/000399/000399.html>

Auray, N. (2011). Les technologies de l'information et le régime exploratoire. Van Andel, P. et Boursier, D. (dir.), *La sérendipité. Le hasard heureux*. Hermann, 329-343. <http://ses-perso.telecom-paristech.fr/auray/2010AuraySerendipite.pdf>

Auray, N. (2017). *L'Alerte ou l'enquête. Une sociologie pragmatique du numérique*. Presses des Mines. <https://doi.org/10.4000/books.pressesmines.3657>

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13, 5, 407-424. <https://doi.org/10.1108/eb024320>

Beaudouin, V. et Denis, J. (2014). Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique. Rapport de recherche BnF, Télécom ParisTech. <https://hal.archives-ouvertes.fr/halshs-01078530>

Beaudouin, V., Garron, I. et Rollet, N. (2016). 'Je pars d'un sujet, je rebondis sur un autre' : Pratiques et usages des publics de Gallica. Rapport de recherche Télécom ParisTech, BnF, Labex Obvil. <https://hal.archives-ouvertes.fr/hal-01709238>

Béquet, G. (2015). *Trois bibliothèques européennes face à Google : aux l'origine de la bibliothèque numérique (1990-2010)*. École nationale des chartes, 2015.

Bermès, E. (2020). *Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque nationale de France (1997-2019)*. Thèse de doctorat, École nationale des chartes. <https://theses.hal.science/tel-02475991>

Bert, J.-F. (2018). *Comment pense un savant ? Un physicien des Lumières et ses cartes à jouer*. Anamosa.

Borgman, C. L. (2000). *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. MIT Press.

Boubée, N. et Tricot, A. (2011). *Qu'est-ce que rechercher de l'information ?* Presses de l'Enssib. <https://doi.org/10.4000/books.pressesensib.799>

Boukacem-Zeghmouri, C. (2012). Online Article Searching on Publisher Platforms by STM French Scholars: Findings and Analysis. *Canadian Journal of Information and Library Science* 36, 3-4, 88-105. <https://doi.org/10.1353/ils.2012.0012>

Boullier, D. et Ghitalla, F. (2004). Le Web ou l'utopie d'un espace documentaire. *Information-Interaction-Intelligence* 4, 1, 173-189. [https://archivesic.ccsd.cnrs.fr/sic\\_00001022](https://archivesic.ccsd.cnrs.fr/sic_00001022)

Bourguet, M.-N. (2016). Voyages lointains et mesure du monde : la projection du regard européen sur le monde (vers 1500-vers 1800). Hilaire-Pérez, L., Simon, F. et Thébaud-Sorger, M. (dir.). *L'Europe des sciences et des techniques. Un dialogue des savoirs, XVI<sup>e</sup>-XVII<sup>e</sup> siècle*. Presses Universitaires de Rennes, 193-201.

Case, D. O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Academic Press.

Cohen, Julie E. (2007). Cyberspace as/and Space. *Columbia Review of Law* 107, 210-256.

Detienne, M. et Vernant, J.-P. (1974). *Les ruses de l'intelligence. La mètis des Grecs*. Flammarion.

Dumas Primbault, S. (2024). Documenter la navigation en bibliothèque numérique. Retour sur un chassé-croisé méthodologique entre qualitatif et quantitatif. À paraître.

Ghitalla, F. (2000). L'espace du document numérique. *Communication et langages* 126, 74-84. <http://dx.doi.org/10.3406/colan.2000.3044>

Ghitalla, F. (dir). (2002). "La navigation". Numéro spécial *Les cahiers du numérique* 3, 3.

Ghitalla, F., Boullier, D. Gkouskou-Giannakou, P., Le Douarin, L. et Neau, A. (dir.). (2003). *L'outre-lecture. Manipuler, (s')appropriier, interpréter le Web*. Bibliothèque Centre Pompidou. <https://doi.org/10.4000/books.bibpompidou.463>

Gilliotte, Q. (2022). *L'expérience culturelle en régime numérique. Explorer, ranger, consommer*. Presses des Mines.

Given, L. M. et Wilson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69, 6, 807-819. <https://doi.org/10.1002/asi.24008>

Jeanneret, Y. (2013). Les chimères cartographiques sur l'Internet : Panoplie représentationnelle de la « traçabilité » sociale. Galinon-Mélenec, B. et Zlitni, S. (dir.). *Traces numériques : De la production à l'interprétation*. CNRS Éditions, 235-267. <https://doi.org/10.4000/books.editions-cnrs.21792>

Kaabachi, B. et Dumas Primbault S. (2023). A Topological Data Analysis of Navigation Paths within Digital Libraries. À paraître.

Koopmann, T., Dallmann, A., Hettinger, L., Niebler, T. et Hotho, A. (2019). On the right track! Analysing and Predicting Navigation Success in Wikipedia. *Proceedings of the 30th ACM Conference on Hypertext and Social Media (HT '19)*. Association for Computing Machinery, 143-152. <https://doi.org/10.1145/3342220.3343650>

Lamprecht, D., Geigl, F. et Karas, T. (2015). Improving recommender system navigability through diversification: A case study of IMDb. *i-KNOW '15*. <http://dx.doi.org/10.1145/2809563.2809603>

Le Marec, J. et Mairesse, F. (2017). *Enquête sur les pratiques savantes ordinaires. Collectionnisme numérique et environnements matériels*. Le bord de l'eau.

Losh, E. (2009). Reading Room: The Nation-State and Digital Library Initiatives. Losh, E. *Virtualpolitik: An Electronic History of Government Media-Making in a Time of War, Scandal, Disaster, Miscommunication, and Mistakes*. MIT Press, 239-281.

Marion, J.-L. (2000). *Sur l'ontologie grise de Descartes. Science cartésienne et savoir aristotélicien dans les Regulæ*. VRIN.

Mounier, P. (2010). Le livre et les trois dimensions du cyberspace. Dacos, M. (dir.). *Read/Write Book. Le livre inscriptible*. OpenEdition. <https://doi.org/10.4000/books.oep.128>

Mueller, R.A. (2019). Episodic Narrative Interview: Capturing Stories of Experience With a Methods Fusion. *International Journal of Qualitative Methods* 18. <https://doi.org/10.1177/1609406919866044>

Nicholas, D. et Clark, D. (2012). "Reading" in the Digital Environment. *Learned Publishing*, 25, 2, 93-98.

Nouvellet, A., Beaudouin, V., D'Alché-Buc, F., Prieur, C. et Roueff, F. (2017). Analyse des traces d'usage de Gallica: Une étude à partir des logs de connexions au site Gallica. Rapport de recherche Télécom ParisTech, BnF. <https://hal.archives-ouvertes.fr/hal-01709264>

Ribes, D., Hoffman, A.S., Slota, S.C. et Bowker, G. (2019). The logic of domains. *Social Studies of Science* 49, 3, 281-309. <https://doi.org/10.1177/0306312719849709>

Roustan, M. (2013). Pour un accès renouvelé aux collections. Rapport de recherche BnF. <https://hal.archives-ouvertes.fr/hal-01405341>

Souchier, E., Jeanneret, Y. et Le Marec, J. (dir.). (2003). *Lire, écrire, réécrire. Objets, signes et pratiques des médias informatisés*. Bibliothèque publique d'information. <https://doi.org/10.4000/books.bibpompidou.394>

Singla, A., White, R.W. et Huang, J. (2010). Studying trailfinding algorithms for enhanced web search. *SIGIR '10*, 443-450. <https://doi.org/10.1145/1835449.1835524>

Teevan, J., Alvarado, C., Ackerman, M. S. et Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. *CHI 2004*, 415-422. <https://doi.org/10.1145/985692.985745>

Tenopir, C., Christian, L. et Kaufman, J. (2019). Seeking, Reading, and Use of Scholarly Articles: An International Study of Perceptions and Behavior of Researchers. *Publications* 2019, 7, 18. <https://doi.org/10.3390/publications7010018>

Trabelsi, M. (2022). *Modélisation des processus utilisateurs à partir des traces d'exécution, application aux systèmes d'information faiblement structurés*. Thèse de doctorat, Université de La Rochelle.

Trace, C. B. et Karadkar, U. P. (2017). Information management in the humanities: Scholarly processes, tools, and the construction of personal collections. *Journal of the Association for Information Science and Technology*, 68, 2, 491-507. <https://doi.org/10.1002/asi.23678>

Weiland, S. (2021). *The Scholarly Workflow in the Digital Age*. Against the grain. <https://doi.org/10.3998/mpub.12119260>

White, R.W. et Huang, J. (2010). Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs. *SIGIR '10*, 587-594. <https://doi.org/10.1145/1835449.1835548>

- 
- <sup>i</sup> « je fais feu de tout bois » (Enquêté 2) ; « tous azimuts » (E6)
- <sup>ii</sup> « recherches concentriques » (E2) ; « en étoile » (E3) ; « deuxième cercle [...] comme des bretelles d'autoroute » (E3) ; « cheminement » (E4)
- <sup>iii</sup> « je pense que j'ai plusieurs vitesses » (E1)
- <sup>iv</sup> « en tache d'huile » (E3)
- <sup>v</sup> « jeu de navigation » (E4) ; « jongler » (E4) ; « trésors d'Ali Baba » (E5) ; « expérience ludique » (E5) ; « navigation touristique » (E7)
- <sup>vi</sup> « tamisage de pépites » (E2) ; « partie à la pêche » (E6)
- <sup>vii</sup> « je ne suis pas rationnelle » (E1) ; « j'ai du mal à catégoriser [...] c'est un espèce de continuum » (E2)
- <sup>viii</sup> « rien d'extraordinaire, tout le monde fait ça » v. « il n'y a que moi qui comprend » ou « mes petites combines » (E2) ; « comment savoir ce qui est idiosyncrasique ou pas, c'est difficile » (E7)
- <sup>ix</sup> E2 explicite un complexe système de prise de notes avant de constater que son ordinateur est dans un état d'« anarchie » lors de la mise en situation ; « pas d'attention particulière à la mise en page » v. « organiser du mieux que je peux » (E4) ; « méthodes de recherche un peu hasardeuses » v. « je reste assez méthodique » (E3) ; « je suis un maniaque » v. « je vous montre pas, c'est un désastre » (E5) ; « j'en fous partout » v. « j'organise ma pensée » v. « c'est un vrac, bordélique à souhait » (E6)
- <sup>x</sup> « maintenant que vous le dites, mais je pense que c'est totalement inconscient » v. « il n'y a que moi qui ai conscience de ça » (E4)
- <sup>xi</sup> E4 a des difficultés à reproduire son cheminement lors de la mise en situation ; « évidemment je vous dis que je prends rien sur des petits carnets mais si en fait j'ai des carnets A5 » puis détaille une pléthore de carnets et formats (E6) ; « qu'est-ce que je peux vous dire d'autres... ? » (E7 et récurrent)
- <sup>xii</sup> « c'est un mode de recherche mais c'est pas forcément le mode de recherche permanent [...] l'un et l'autre ne sont pas contradictoires, on est dans différentes circonstances et dans différentes approches de la recherche à un moment ou un autre [...] on n'a pas toujours la même chose à faire en matière de recherche » (E2)
- <sup>xiii</sup> « papillonner sur Gallica c'est un peu comme quand à la BnF Mitterrand [...] je parcours les rayons et il se peut que je tombe sur un ouvrage qui soudainement me donne une idée voire peut paraître totalement différent mais peut inspirer autre chose, c'est-à-dire qu'il ne faut pas se cantonner à son thème d'étude » (E3)
- <sup>xiv</sup> « parfois il m'arrive de vagabonder, effectivement, c'est-à-dire tomber sur quelque chose qui m'intéresse et qui est complètement différent de ma recherche initiale et je trouve cela aussi passionnant que perturbateur mais c'est le propre de la recherche aussi de tomber sur des choses qu'on n'avait pas prévues » (E1)
- <sup>xv</sup> « complètement au hasard, c'est quand même bien dans le sens où je m'aperçois qu'on a des pièces [...] genre de documents dont on n'a pas la référence précise ou le titre précis et donc c'est très utile de tomber dessus par une espèce de recherche au hasard [...] trouver des documents qu'on ne connaissait pas au départ, la recherche un peu au pif elle est bien quand on prépare quelque chose ou on a besoin de faire des recherches tous azimuts » (E2)
- <sup>xvi</sup> « ça fait appel au principe de sérendipité comme en bibliothèque où on consulte quelque chose et puis hop on voit que... [s'arrête] là j'ai pas le souvenir, mais est-ce que ça a pu m'arriver comme ça de jongler au fil de mes recherches, où j'étais sur quelque chose et hop j'ai vu un document qui au premier abord ne concernait pas mes recherches et de fil en aiguille je suis tombé sur quelque chose, ça a pu m'arriver [...] on n'est pas à l'abri d'un cheminement intellectuel qui ne s'était pas fait jusqu'à présent et qui puisse se faire via ce type de suggestions même si moi j'ai pas souvenir que ça ait fonctionné beaucoup comme ça sur Gallica ou bien si ça a pu le faire c'était plus par curiosité où ah tiens voilà tel document, ben tiens allons voir ce que c'est et tout [...] une manière un peu différente d'accéder à un corpus de documents sans forcément qu'il soit lié à nos recherches mais qui puisse par rebond nous amener à des choses auxquelles on n'avait pas forcément pensé ou imaginé » (E4)
- <sup>xvii</sup> « on voit le conditionnement de la biblio physique [...] on suit un clivage, on a un schéma mental qui fait qu'on a cette biblio virtuelle dans l'esprit » (E3) ; « ça fait appel au principe de sérendipité comme en bibliothèque où on consulte quelque chose et puis hop on voit que... [s'arrête] » (E4)
- <sup>xviii</sup> « volonté de comparaison » (E1)
- <sup>xix</sup> E2
- <sup>xx</sup> Par opposition aux associations d'idées chez E4.
- <sup>xxi</sup> « c'est plutôt hasardeux » (E5)