



Bibliothèque nationale de France

Usages des bibliothèques numériques en sciences et techniques

Rapport d'étude

Bibliothèque nationale de France
délégation à la Stratégie et à la Recherche
département de la Coopération

Auteurs : Muriel Amar et Cécile Touitou
Contacts : muriel.amar@bnf.fr
cecile.touitou@bnf.fr

Résumé exécutif

Les bibliothèques numériques spécialisées en sciences et techniques (nommées ci-dessous SCT) sont une dizaine dans le domaine francophone avec des collections conséquentes tant en volume qu'en qualité. Une étude qualitative a été menée entre fin 2010 et début 2011 afin de mieux comprendre comment s'articulent les usages de ces bibliothèques avec ceux de Gallica, bibliothèque à visée encyclopédique. Quinze entretiens semi-directifs auprès d'utilisateurs intensifs de quatre bibliothèques SCT partenaires de la BnF (Medica, le Cnum, la Jubilotheque et Numdam) ont apporté des éléments de réponse aux questions suivantes : du côté des bibliothèques SCT, quelle opportunité de continuer à développer des bibliothèques de niche dans un contexte de numérisation de masse ? ; du côté de la BnF : quelle visibilité est donnée à l'offre de service de Gallica auprès des communautés scientifiques ?

Les chercheurs interrogés ici ont, pour la majeure partie d'entre eux, engagé leur carrière scientifique avant l'existence des bibliothèques numériques (l'âge moyen est de 47 ans) : leur habitus reste celui propre au travail sur l'imprimé. À cet égard, les résultats méritent toujours d'être contextualisés.

Les bibliothèques de « niche » sont appréciées avant tout parce qu'elles prennent en compte les besoins documentaires des chercheurs : face à la masse attractive des documents en ligne (Google Books), la bibliothèque de niche représente le segment rassurant de l'information organisée dans lequel on sait ce que l'on va trouver et la forme sous laquelle on va l'y trouver ; ceci d'autant plus que les chercheurs enquêtés sont, le plus souvent, associés à la sélection documentaire. Les fonctionnalités (recherche, visualisation, etc.) offertes par les bibliothèques SCT comptent beaucoup moins que les fonds proposés, même en mode image. Cette bonne acceptation du mode image révèle une défiance vis-à-vis d'un mode texte qui fragmente et dissémine les contenus. Les fonctionnalités techniques propres aux bibliothèques numériques sont d'autant moins attendues que le moteur de recherche de Google, presque systématiquement utilisé pour explorer les collections numériques, donne toute satisfaction en matière d'accès aux contenus des textes numérisés.

Les postures et pratiques identifiées pour les bibliothèques de niche se retrouvent dans les usages de **Gallica** : la principale motivation de visite et le principal motif de satisfaction tiennent à la qualité (richesse et volumétrie) des corpus numérisés ; les fonctionnalités techniques intéressent peu ce public de chercheurs qui recourt massivement au moteur de recherche de Google pour identifier pages, documents ou corpus de travail. Par contre, les principes de sélection et de mise en ligne de Gallica sont jugés insuffisamment transparents, contrairement aux bibliothèques de niche, faites pour et surtout avec les chercheurs. La possibilité de faire une recherche en mode texte, considérée comme l'un des points forts de Gallica, est en même temps l'objet de suspicions quant à la pertinence des résultats. Ceci révèle une difficulté à lâcher le modèle maîtrisable de la bibliothèque de niche – interrogeable, le plus souvent, par mots de la notice bibliographique (et non en texte intégral) et feuilletable par tables des matières – pour le « non-modèle » bruyant du plein-texte. Si la recherche en texte intégral est le seul moyen de faire des découvertes et d'aller au-delà des corpus connus, comment valider ces nouveaux corpus fortuitement construits sur la base d'une similitude de chaîne de caractères ? On aime que Gallica nous permette de faire résonner nos termes de recherche dans tous les textes disponibles mais on n'aime pas perdre l'épaisseur documentaire qui contextualise les contenus.

Les deux types de dispositifs – Gallica et les bibliothèques de niche – sont donc complémentaires sur plus d'un point : en termes d'offre documentaire, bien sûr, mais aussi en termes de modalités d'accès à l'information : interface de recherche et traitement documentaire différenciés.

Au-delà de ces questions de positionnement, l'étude permet de dresser le portrait **d'usagers intensifs de corpus numérisés** et de fournir un aperçu des pratiques actuelles de recherche documentaire, de lecture et d'étude à l'écran. Les usagers intensifs des bibliothèques numériques ne manifestent qu'un intérêt très limité pour les technologies de l'information et de la communication. Ce conservatisme dans leurs pratiques du web cohabite avec le constat d'un impact important du numérique dans leur



travail scientifique, le corpus numérique pouvant influencer des directions de recherche et renouveler des méthodes d'exploration des contenus. Les racines d'un travail à l'ancienne se signalent cependant encore en matière de lecture approfondie des textes, qui privilégie toujours le support imprimé. Ce profil influence les attentes exprimées en matière de bibliothèque numérique idéale : l'intérêt des chercheurs est beaucoup plus tourné vers les contenus (encore plus de sources, de tout type, rares ou difficiles à trouver de préférence) que vers les dispositifs techniques d'accès ou d'exploration des documents numériques.

Nous remercions Françoise Gaudet, responsable du service Études et recherche de la Bibliothèque publique d'information, pour sa lecture et les précieuses remarques qu'elle nous a communiquées.

Sommaire

1- Introduction

- 1.1- Contexte général de l'étude
- 1.2- Objectif et conduite de l'étude
- 1.3- Méthodologie et profil des enquêtés
- 1.4- Présentation des bibliothèques numériques scientifiques soumises à l'enquête

2- Pratiques et évaluation de quatre bibliothèques numériques scientifiques

- 2.1- Le Conservatoire numérique
- 2.2- La Jubilothèque
- 2.3- Numdam
- 2.4- Medic@
- 2.5- Conclusions intermédiaires

3- Pratiques et évaluation de Gallica

4- Conclusions intermédiaires : *d'une bibliothèque à l'autre, points communs et différences ; faits saillants ; retour sur l'hypothèse initiale : complémentarités ou/et concurrences entre bibliothèque spécialisée et bibliothèque encyclopédique*

5- Caractéristiques des usagers intensifs des bibliothèques numériques scientifiques

- 5.1- Aperçu de leurs pratiques en ligne
- 5.2- Aperçu de leurs comportements en matière de recherche d'information et de pratiques culturelles

6- Conclusions intermédiaires : *d'une bibliothèque, sur place, à l'autre, en ligne : points communs et différences de pratiques ; retour sur l'hypothèse initiale : spécificités de l'ancrage disciplinaire sur les pratiques de recherche.*

7- Conclusions

Annexes

Annexe 1- Grille d'entretien

Annexe 2- Profil des usagers interrogés

Annexe 3- Compléments statistiques (Jubilothèque)

Annexe 4- Compléments statistiques et données chiffrées (Gallica)

1. Introduction

1.1 Contexte général de l'étude

Cette étude s'inscrit en complément d'une enquête d'usage de grande ampleur sur Gallica pilotée par la délégation à la Stratégie et à la recherche (DSG) et réalisée par la société GMV Conseil.

Menée en interne sous l'impulsion du département de la Coopération (DCP), cette étude aborde les usages de Gallica en privilégiant deux dimensions spécifiques : d'une part, la dimension coopérative de la constitution des collections numériques et, d'autre part, le segment des publics en sciences et techniques.

Rien pour autant de marginal dans ces choix : les bibliothèques numériques en sciences et techniques, partenaires de Gallica et de la BnF, sont nombreuses dans le domaine francophone (on en dénombre une dizaine), existent pour la plupart d'entre elles de longue date (dix ans pour la bibliothèque numérique de la Bibliothèque interuniversitaire de Santé et du Conservatoire national des Arts et métiers) et proposent des collections conséquentes tant en volume qu'en qualité. La plupart de ces bibliothèques ont par ailleurs mis à disposition leurs notices pour un référencement dans Gallica (*via* le protocole OAI-Pmh) : comment s'articulent, pour les publics, le passage d'une bibliothèque scientifique¹ spécialisée à la grande bibliothèque encyclopédique Gallica ?

Sur le plan des usages, en outre, on connaît l'appétence de la communauté en sciences exactes pour les ressources contemporaines en ligne, mais consulte-t-elle également le patrimoine numérisé ? Quelles sont les spécificités des recherches en histoire des sciences, sont-elles proches de celles des historiens ou proches de celles des scientifiques ?

Ce sont en particulier ces deux ensembles de questions -

- articulation bibliothèque de niche / bibliothèque encyclopédique
 - spécificité disciplinaire et/ou exploration toujours historique du patrimoine numérisé -
- qui ont été à l'origine de cette étude qui d'emblée souhaitait associer, outre les différents services de la BnF concernés, le réseau des pôles associés susceptibles d'être

¹ . Dans ce document, le terme « scientifique » désigne, de façon restrictive, les sciences exactes et appliquées.

intéressés. Cette étude s'est donc construite aussi en fonction de leurs besoins et de leurs intérêts.

1.2 Objectifs et conduite de l'étude

A- Objectifs

Le recueil des pratiques développées à partir des bibliothèques numériques spécialisées en sciences et techniques devait permettre de répondre à un double questionnement :

- du côté des bibliothèques partenaires de la BnF : nombre de partenaires s'interrogeait sur l'opportunité de continuer à développer des bibliothèques de niche dans un contexte de numérisation de masse ;
- du côté de la BnF : aussi bien le département de la Coopération que celui des collections en Sciences et techniques s'interrogeaient sur la visibilité des services de la BnF auprès des communautés scientifiques, historiens des sciences compris.

L'étude des usages des bibliothèques spécialisées en sciences et techniques s'est donc donné pour objectif de dégager d'une part une typologie des pratiques effectives et d'autre part un corpus de représentations relatives à Gallica et à la BnF dans le paysage de la coopération.

Ces objectifs ont conduit à établir le profil suivant de recrutement : des utilisateurs « intensifs » de bibliothèques scientifiques numériques partenaires de la BnF, et pas nécessairement de Gallica. Précisément, cet éventuel écart de pratiques devait permettre d'examiner les conditions de passage (ou non) d'usages de bibliothèques spécialisées aux usages spécialisés d'une bibliothèque encyclopédique comme Gallica.

B- Conduite du projet

Le projet a associé la délégation à la Stratégie et à la recherche, le département Sciences et techniques et les bibliothèques partenaires intéressées : la Bium pour Medica, le Cnam pour le Cnum, la BUPMC pour la Jubilothèque, MathDoc pour Numdam. Ces quatre bibliothèques sont présentées ci-dessous (§ 1.4).

Le recrutement des personnes à interroger était à la charge des bibliothèques partenaires (quatre utilisateurs intensifs maximum par bibliothèque numérique) ; la BnF validait le recrutement, organisait et pilotait les rencontres.

Les entretiens ont été menés par la BnF² dans le lieu retenu par l'utilisateur (en présence du responsable de la bibliothèque numérique associée s'il le souhaitait), ils ont été enregistrés et retranscrits. Ils ont été analysés conjointement par la BnF et par les bibliothèques partenaires, sur la base d'une grille de classement des verbatim élaborée collectivement.

1.3 Méthodologie et profil des enquêtés

La méthodologie retenue repose sur une série de quinze entretiens semi-directifs avec des usagers des bibliothèques numériques en sciences et techniques (SCT) pour cerner les pratiques installées. La grille d'entretien a été élaborée conjointement par la BnF et les bibliothèques partenaires participantes.

La grille d'entretien est présentée en annexe 1. Elle a bien fonctionné et a permis d'échanger avec les enquêtés pendant une heure et demie (deux heures maximum) sur l'ensemble des aspects qui avaient été retenus. Le choix était laissé à l'utilisateur de recourir ou non à un poste informatique pour illustrer son propos : 9 personnes sur 15 ont fait ce choix. De façon générale, il reste bien sûr très difficile d'obtenir d'eux la relation précise de leur session de recherche... ce sont plus leurs motivations, leurs routines d'usage, leurs impressions, leurs façons de faire personnelles et leurs connaissances de l'environnement numérique qui ont pu être approfondies.

² . Muriel Amar, responsable du projet au département de la Coopération, a mené tous les entretiens ; Cécile Touitou, responsable du projet à la délégation à la Stratégie et à la recherche, a participé à plus de la moitié des échanges ; Louis Jaubertie, correspondant du projet au département Sciences et techniques, a assisté au tiers des entretiens. Sauf mention contraire, le terme « BnF » désigne, dans ce document, ces trois porteurs du projet.

Le profil détaillé des utilisateurs interviewés est présenté en annexe 2. Le tableau suivant résume les données signalétiques succinctes.

Sexe		Activité professionnelle	
Femmes	7	Recherche-enseignement	9
Hommes	8	Patrimoine-documentation	4
Age		Formation initiale	
Moins de de 40 ans	3	Sciences fondamentales	8
Entre 40 ans et moins de 50 ans	4	Histoire	3
Entre 50 ans et moins de 60 ans	6	Autres	4
Plus de 60 ans	2		

Grille de classement des verbatim :

La grille de classement a été établie collectivement sur la base de la relecture de l'ensemble des retranscriptions d'entretien. Son objectif était de préparer le travail d'analyse interprétative ultérieure en dégagant quatre problématiques :

- Usages des bibliothèques numériques en sciences et techniques. Cet axe est composé de douze thèmes : la moitié environ porte sur des modalités d'usage (type de recherche, mode de lecture, etc.), l'autre moitié invite à une évaluation du dispositif (points forts/points faibles ; facilités/difficultés, etc.). Son analyse a été assurée par chacun des responsables des bibliothèques numériques concernées.
- Gallica : usages, connaissances, appréciations... dix thèmes sont identifiés sous cet angle. De façon générale, Gallica est très utilisée par notre groupe d'utilisateurs en dépit d'un niveau de satisfaction déclaré modeste. L'analyse a été conjointement menée par le DCP et la DSG.
- Pratiques en ligne : cet axe, composé de vingt thèmes, permet de regrouper tous les verbatim relatifs non aux contenus particuliers d'une bibliothèque numérique mais aux comportements plus généraux qui motivent le recours massif aux réservoirs

numériques. Cet ensemble a également été analysé conjointement par le DCP et la DSG.

- Comportements généraux (en matière de pratiques documentaires et de pratiques culturelles). Axe composé de huit thèmes, permettant d'isoler notamment l'usage des bibliothèques physiques (très important pour notre groupe), leur culture bibliothéconomique (en général, très forte aussi), leurs pratiques de la lecture-plaisir. L'étude de cet axe a été confiée à notre correspondant du projet au département Sciences et techniques.

Ce sont ces sept analyses qui sont présentées dans les pages qui suivent.

1.4 Présentation des bibliothèques numériques scientifiques soumises à l'enquête

Chacun des responsables des quatre bibliothèques numériques soumises à l'enquête présente les caractéristiques documentaires et techniques de son corpus en ligne en précisant les publics visés.

Grâce au protocole OAI-Pmh, Gallica donne accès aux notices de trois des réservoirs sur quatre : Cnum, Jubilothèque et Medic@ sont accessibles dans Gallica ; Numdam ne l'était pas au moment de l'enquête.

A- Le Conservatoire numérique, par Mireille le Van Ho, directrice de la bibliothèque centrale du CNAM

Le Conservatoire numérique des Arts et Métiers (CNUM) est une bibliothèque numérique consacrée à l'histoire des sciences et des techniques, constituée à partir du fonds ancien de la bibliothèque du CNAM.

Le CNUM est un projet commun de la Bibliothèque, du Centre d'histoire des techniques et de l'environnement (CDHTE) et du Centre d'études et de recherche en informatique (CEDRIC) du CNAM, mis en place en 2000. Une collection d'ouvrages et de périodiques francophones, significatifs du fonds patrimonial de la bibliothèque, traitant des sciences et des techniques, de l'économie et de la sociologie appliquées, est consultable sur Internet (cnum.cnam.fr). Le Conservatoire numérique s'adresse aux chercheurs et aux enseignants en histoire des sciences et des techniques, en épistémologie, en didactique, en offrant à la fois des textes et une documentation iconographique spécifiques. Le CNUM est aussi un outil de vulgarisation scientifique qui remplit l'une des missions fondamentales du CNAM : la diffusion du savoir et la reconnaissance du patrimoine scientifique et technique francophone.

Riche de plus de 1200 volumes dont la plupart déchargeables en PDF, et de 630 000 pages en ligne³, cette bibliothèque numérique présente les publications éditées à

³ . Mode image, saisie manuelle des tables des matières, des tables des illustrations et des légendes, le cas échéant. Le mode image seul ne permet pas la recherche dans le texte des documents. Si les documents en mode image sont soumis à une conversion (à travers un logiciel dit d'OCR, reconnaissance optique de caractères), ils deviennent interrogeables par leurs chaînes de caractères : cette conversion produit des documents « en mode texte ».

l'occasion des expositions universelles du XIX^e siècle, des monographies scientifiques et techniques comme la *Description des arts et métiers* ou des catalogues de constructeurs et des revues comme *La Nature* (1873-1905), *Les Annales du Cnam* (1861-1933), *La Lumière électrique* (1879-1894 ; 1908-1916), le *Bulletin de la société des ingénieurs civils* (1848-1900), *Le Génie industriel* (1848-1965) ou le *Bulletin de la Société d'encouragement pour l'industrie nationale* (1802-1943).

En 2005, la Bibliothèque nationale de France et le CNAM ont signé une convention de coopération documentaire dans le cadre du réseau des pôles associés. Les deux établissements ont établi en commun un corpus de numérisation, à partir de leurs collections respectives, sur le thème des expositions universelles du XIX^e siècle. Les documents officiels produits à l'occasion de ces manifestations restent difficiles d'accès : rapports officiels des jurys internationaux, rapports publiés par la France sur les nouvelles inventions, congrès, documents préparatoires aux expositions, revues, guides.... Les fonds de la bibliothèque du CNAM sont particulièrement riches sur ce sujet qui constitue l'une des principales sources d'information sur l'innovation industrielle et technique, la diffusion et la médiatisation de la culture scientifique et technique. La convention a également permis de développer une passerelle de navigation d'après le protocole Open Archives Initiative (OAI) entre Gallica et le CNUM et, depuis 2009, entre le Cnum et Gallica.

B- La Jubilothèque, par Frédérique Flamerie de la Chapelle, responsable de la bibliothèque numérique à la bibliothèque universitaire de l'Université Pierre et Marie Curie (Bupmc).

La Jubilothèque propose, depuis 2007, en accès libre sur internet des documents numérisés libres de droits issus des collections de ses bibliothèques de recherche et sélectionnés en partenariat avec la communauté scientifique. Les principaux fonds patrimoniaux de la BUPMC sont le fonds Charcot, constitué autour de la bibliothèque personnelle de Jean-Martin Charcot, et les collections du CADIST en Géosciences. De fait, ce sont les deux domaines les plus développés dans la Jubilothèque, qui compte en juin 2011 plus de 1100 documents en ligne (soit un peu plus de 330 000 pages), organisés en collections thématiques.

Le public cible de la Jubilothèque est avant tout la communauté scientifique au sens large, chercheurs, enseignants et étudiants, dans tous les domaines couverts par la Jubilothèque, ainsi qu'en histoire des sciences. Certaines collections attirent toutefois un public plus atypique : géologues amateurs pour l'histoire de la géologie et la géologie régionale, chercheurs en sciences humaines et sociales, iconographes ou journalistes pour le fonds Charcot (l'iconographie photographique de la Salpêtrière principalement).

En l'absence d'un outil statistique, il est toutefois difficile d'appréhender finement le public réel de la Jubilothèque, la façon dont il y accède (Gallica *via* l'interopérabilité, moteurs de recherche, accès direct) ainsi que les documents qu'il consulte. Le fonds Charcot, compte tenu des demandes de reproduction ou de prêt reçues par ailleurs, semble le plus sollicité.

Mis à part les manuscrits et une partie du fonds Charcot numérisée au début des années 2000, l'ensemble des documents est numérisé en mode texte (OCR non corrigé, OCR corrigé seulement pour les tables des matières). Les fonctionnalités offertes sont simples, qu'il s'agisse de la recherche dans la bibliothèque ou de la consultation d'un document. Des lots de vignettes et la table des matières facilitent la navigation au sein d'un document. Une visionneuse dédiée permet de consulter les images en haute définition (300 dpi), de zoomer, d'imprimer ou de télécharger tout ou partie du document. Une migration de l'application début 2011 (soit juste après l'enquête) de Pleade 2 vers Pleade 3 a permis un enrichissement des fonctionnalités, notamment du point de vue de la recherche, avec la possibilité de recourir aux index titre, auteur et sujet, de lancer une recherche dans plusieurs collections, et de dépouiller plus facilement les résultats de recherche, contextualisés au sein du document et de la collection thématique auxquels ils appartiennent.

C- Numdam, par Hélène Falavard, ingénieur documentaliste à la Cellule de Coordination Documentaire Nationale pour les Mathématiques (MathDoc).

NUMDAM est la bibliothèque numérique dédiée aux mathématiques et domaines connexes, conçue et portée par la cellule MathDoc (Cellule de Coordination Documentaire Nationale pour les Mathématiques). Mise en ligne fin 2002, elle s'adresse essentiellement aux chercheurs en mathématiques mais aussi à ceux dont le sujet de recherche est connexe aux mathématiques comme par exemple les sciences humaines, la physique ou les sciences économiques.

Le corpus est composé tout à la fois de revues françaises et européennes, académiques ou non et d'actes de séminaires. A ce jour, plus de 43 000 articles téléchargeables pour la grande majorité sont mis à disposition des chercheurs qui ont ainsi la possibilité de consulter des articles fondateurs issus de revues mondialement reconnues⁴.

Dès le lancement du programme en 2001, des choix techniques forts et novateurs ont été faits pour la numérisation et le catalogage des métadonnées. De façon à permettre une restitution la plus fidèle possible des textes, les pages sont numérisées à 600 dpi noir et blanc et 300 dpi niveaux de gris ou couleur lorsque cela se justifie (planches, photos, ...). Le texte des articles est ocrisé pour pouvoir être intégré dans le moteur de recherche. Les métadonnées sont cataloguées de manière fine : non seulement les informations concernant l'article lui-même sont relevées (auteurs, titre, pagination, années de publication) mais également des informations supplémentaires telles que les références bibliographiques, les résumés, les mots clés ou les formules de mathématiques dans les titres. Lorsqu'elles existent, les relations entre articles (suite de, partie de, ...) sont également cataloguées. Les articles sont proposés à la fois au format PDF (avec le texte intégral protégé pour empêcher la copie illégale) et au

⁴ Notons l'article de Vladimir Arnold "[Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits](#)" paru en 1966 dans les *Annales de l'Institut Fourier* et qui contribue à fonder la théorie de l'"[effet papillon](#)". Notons également l'article de Jacques Hadamard "[Sur la distribution des zéros de la fonction \$\zeta\(s\)\$ et ses conséquences arithmétiques](#)", paru en 1896 dans le *Bulletin de la société mathématique de France*, qui contient la démonstration du théorème des nombres premiers. Enfin, signalons la présence, parmi les articles de Numdam, des travaux d'Alexander Grothendieck, fondateur de la théorie des schémas ([Publications mathématiques de l'IHES](#), [séminaire Bourbaki](#), etc.).

format DjVu (format qui propose un meilleur taux de compression pour des fonctionnalités et une qualité d'image identiques à celles du PDF).

D- Medic@, par Jean-François Vincent, responsable du service d'Histoire de la médecine de la bibliothèque interuniversitaire de Santé (Paris).

Lancée à l'automne 2000, la bibliothèque Medic@ est le projet de numérisation patrimoniale et d'édition scientifique mené par le Service d'histoire de la BIU Santé, complémentaire des autres produits élaborés par le Service (banque d'images, base bio-bibliographique des médecins, expositions virtuelles, renseignements à distance...) Elle compte à ce jour 8200 documents (2,3 M pages) et en signale 31000 issus de diverses sources⁵. Elle s'accroît de 300 à 400 000 pages par an.

Sa politique documentaire est décrite en ligne⁶.

Quelques points concernant Medic@ doivent être rappelés pour la suite de cette synthèse :

- Medic@ est organisée en séries (ex. : Corpus des médecins de l'Antiquité) et subdivisée en dossiers (ex. : Galien, Hippocrate). Ces dossiers peuvent être feuilletés depuis la page d'accueil.
- Les dossiers sont souvent réalisés en partenariat avec des chercheurs ou des équipes de chercheurs, qui participent à la sélection et écrivent des présentations des dossiers.
- La numérisation est faite en mode image et est associée à des tables des matières.
- Une recherche par mots clés est possible sur les notices. La recherche par mots clés sur les tables des matières n'est pour l'instant possible que sur les sous-ensembles des dictionnaires et des périodiques (sauf par l'intermédiaire de Google qui indexe la totalité des tables des matières.)

E- Gallica, par Muriel Amar à partir de l'article d'Arnaud Dhermy⁷ (coordinateurs scientifiques au service Pôles associés/Gallica de la Bibliothèque nationale de France).

Gallica est la bibliothèque numérique hébergée et développée par la Bibliothèque nationale de France depuis sa création en 1994, accessible sur le web depuis octobre 1997.

Initié en 1991, en même temps que le projet présidentiel d'une bibliothèque « d'un genre nouveau », Gallica est d'abord conçu comme une station de travail accessible à partir des futures salles de lecture du site de Tolbiac : le premier corpus de documents numérisés privilégie les textes de référence (libres et sous droits), les grandes revues françaises et surtout les principaux instruments de recherche (encyclopédies, bibliographies, dictionnaires). Les exemplaires numérisés proviennent non pas des collections du dépôt légal mais soit d'acquisitions spécifiques soit de prélèvements ciblés dans les fonds de bibliothèques spécialisées partenaires (en sciences,

⁵ http://www.biusante.parisdescartes.fr/histmed/medica/oai_pmh.htm

⁶ http://www.biusante.parisdescartes.fr/histmed/medica/poldoc_medica_fra.pdf

⁷ . DHERMY Arnaud, 2010 : « Gallica : construction et stratégie », *La Pensée*, 2010, Janvier – Mars, n°361 (Dossier : De la numérisation), p.51-63, 193-194.

essentiellement des bibliothèques d'Orsay, de Sainte-Anne, de l'Observatoire de Paris, de l'Ecole des Ponts et chaussées et de l'Ecole polytechnique).

Le déploiement de Gallica sur le web en 1997 engage une nouvelle politique documentaire qui privilégie les besoins d'un public distant auquel sont proposés une sélection de documents représentatifs du patrimoine national - rares, épuisés ou difficiles d'accès - et un ensemble encyclopédique d'« essentiels » de nature à amorcer ou poursuivre une recherche⁸. Là encore, la numérisation de ces ensembles, tous libres de droits dans cette deuxième tranche de sélection, s'appuie pour une bonne part sur les fonds des bibliothèques partenaires.

La riposte à Google Books que propose Jean-Noël Jeanneney se traduit à partir de 2006 par un engagement fort de la BnF dans le développement de Gallica, dont la progression est spectaculaire : de 35 000 volumes et 10 000 images en 1997 (ensemble déjà unique à cette époque) à 90 000 documents en 2006, à environ 926 000 en 2009 et plus d'un million en 2010. Sur cette troisième tranche, ce sont principalement les collections BnF qui sont sollicitées, et de façon particulièrement intense (100 000 volumes imprimés par an à partir de 2007), grâce au soutien financier du Centre national du livre (CNL). Cette numérisation de masse permet d'identifier des ensembles de cotes cohérentes (pour les sciences, médecine et thérapeutique, sciences naturelles notamment) mais pas de cibler, en deçà, des segments documentaires thématiques ou des listes de document par disciplines.

La collection numérisée relevant des disciplines scientifiques est actuellement estimée à 10% environ de l'ensemble des documents mis en ligne dans Gallica : ce taux a oscillé entre 8% et 12% au cours des trois principales strates de constitution documentaire de Gallica.

À partir de 2006 également, la numérisation en mode image des documents se double, quasi systématiquement, d'une conversion en mode texte, tandis qu'un traitement rétrospectif des documents déjà mis en ligne est entrepris pour permettre aux Gallicanautes de disposer, autant que faire se peut, du texte intégral des documents, en plus de leur version image. Il reste actuellement à océriser environ 20% des documents de Gallica (susceptibles de l'être ; sont exclues les publications antérieures à 1750).

L'offre s'élargit encore en mars 2008 avec l'expérimentation d'une offre payante de livres sous droits.

La version actuelle du site Gallica propose un accès unique à cette diversité de collections : le produit de la numérisation à grande échelle des collections BnF, les documents sous droits provenant de l'édition contemporaine et les ensembles issus des fonds des bibliothèques partenaires, soit par prélèvement et numérisation opérés par la BnF, soit par interopérabilité des notices entre réservoirs numériques. Gallica se donne ainsi à voir comme un portail agrégeant toujours plus de ressources et comme

⁸ . La politique documentaire en sciences et en histoire des sciences est décrite par Natalie Pigeard : "Science et histoire des Sciences dans Gallica", *High Energy Physics Libraries Webzine*, issue 11, August 2005 - URL: <http://library.web.cern.ch/library/Webzine/11/papers/3/>.

un produit en évolution constante : une nouvelle version de Gallica est chargée environ tous les deux mois depuis 2008, apportant des améliorations, des correctifs ou de nouvelles fonctionnalités sur les différents aspects de la bibliothèque numérique ou du site Gallica (modes de consultation, référencement des documents, ergonomie des écrans, etc.). Le moteur de recherche actuellement utilisé devrait être remplacé dans le courant 2012 pour permettre de réaliser, de façon plus efficace, les fonctions de recherche, simple et avancée, à la fois dans les notices bibliographiques et dans le plein texte des tables des matières et des documents traités par la BnF. Parmi les autres développements en cours, la conception d'une interface dédiée *mgallica.bnf.fr* ainsi que des applications pour Iphone et Ipad permettront d'adapter la bibliothèque numérique aux usages mobiles.

Deux initiatives en faveur du partenariat numérique méritent enfin d'être signalées :

- sur le plan technique : le réservoir des documents numérisés et mis en ligne dans Gallica est ouvert au moissonnage (*via* le protocole OAI-PMH) depuis 2005 : le Cnum moissonne tout Gallica, Medic@ le sous-ensemble « médecine » ; inversement, Gallica moissonne les catalogues des bibliothèques numériques partenaires de notre étude, le Cnum, Medica@, la Jubilothèque et Numdam dans leur intégralité (moissonnage de Numdam non effectif au moment de l'enquête) ;
- sur le plan politique : les modalités de coopération avec les bibliothèques pôles associées se sont élargies, à partir de 2008, aux actions de numérisation (soutien méthodologique, technique et financier à hauteur de 50% du montant des projets) grâce aux subventions du Service du livre et de la lecture du Ministère de la Culture (trois des bibliothèques scientifiques étudiées bénéficient de ce soutien financier). À ce volet de financement et d'accompagnement des bibliothèques numériques partenaires s'ajoute celui de l'intégration, dans les dispositifs de numérisation BnF, des fonds des bibliothèques partenaires à hauteur de 20 000 volumes imprimés par an (sur les quatre années de l'actuel marché 2011-2014).

L'annexe 4 présente les principales statistiques et données chiffrées relatives à Gallica.

2. Pratiques et évaluation de quatre bibliothèques numériques scientifiques

Le tableau de l'annexe 2 indique le codage utilisé pour contextualiser les verbatim cités dans la suite de ce rapport.

2.1 Le Conservatoire numérique, synthèse par Mireille le Van Ho, directrice de la bibliothèque du Cnam

Dans le cadre de cette enquête, quatre personnes ont été interrogées sur les usages et pratiques qu'elles ont du Conservatoire numérique : une chercheuse au Centre d'histoire des techniques et de l'environnement du Cnam, deux chargées de collections du Musée des arts et métiers qui poursuivent aussi des recherches et une bibliothécaire du Musée des arts et métiers, voir annexe 2.

Du fait de leur ancrage institutionnel, leur connaissance du Cnum est ancienne, le plus souvent depuis sa création avec un usage fréquent à quotidien depuis quatre à cinq ans, correspondant à une montée en charge de la base qui s'est enrichie. Les chercheurs du CDHTE et chargées de collections du Musée ont été directement associés à la constitution des corpus de numérisation d'où leur intérêt pour la bibliothèque et l'utilisation importante qu'ils en font. Dans leur pratique professionnelle, l'utilisation des bibliothèques numériques est un « réflexe » : pour Cnum1, « [j'utilise le Cnum] presque quotidiennement. », « J'ai une utilisation quasi obsessionnelle du Cnum », comme pour Cnum4 : « A partir du moment où on fait une recherche, si on n'a pas la source ici sur place, on se tourne tout de suite vers les bibliothèques numériques. En premier le Cnum, en second Gallica ». Le recours à d'autres bases est également très fréquent : Google Books, Gallica bien sûr mais aussi Eleonore, Scientifica, Architectura... Dans les motivations d'usage, il s'agit très souvent de retrouver des ouvrages déjà connus, même si l'exploration de plus vastes corpus n'est pas exclue mais souvent à la marge de la recherche, parce que dévoreuse de temps. S'agissant du Conservatoire numérique, la littérature officielle sur les expositions universelles du XIX^e siècle et les revues d'histoire des techniques (*Mémoires et compte-rendu de la Société des ingénieurs civils*, *Bulletin de la Société d'encouragement pour l'industrie nationale*, *la Nature*...) sont les deux thématiques majeures exploitées par les chercheurs.

Le mode de recherche passe soit par le catalogue de la base numérique, soit par le catalogue de la bibliothèque qui renvoie sur les ouvrages numérisés, même si la possibilité de ce lien n'a pas été identifiée par toutes les utilisatrices. La recherche sur Internet (via Google) est souvent un préalable pour trouver du texte OCRisé, option actuellement non proposée dans le Cnum, déplorée par deux des interviewées (« *Le défaut principal du Cnum, c'est de ne pas OCRiser* », Cnum3) et rejetée par les deux autres par crainte du bruit généré (« *je pense que ça ferait pas mal de bruit. C'est un peu à double tranchant. On a toujours la tentation d'en demander plus, mais ça ne serait pas forcément très*

sage », Cnum2). L'entrée biographique sur Google est un mode de recherche préalable souvent sollicité qui permet ensuite d'accéder à des ouvrages précis en mode texte ou image. S'agissant du Cnum, le mode d'accès par mots des tables des matières qui permet d'accéder rapidement, sans génération de bruit, aux informations recherchées est d'emblée retenu.

La masse documentaire offerte par la numérisation des revues dans le Cnum est un champ d'exploration et de découvertes apprécié des chercheurs dont le corpus mis à disposition détermine souvent la recherche. Pour autant, le corpus est encore jugé insuffisant (brochures, petits imprimés, catalogues de vente...) surtout pour tout ce qui concerne les éditions successives, les rééditions et la transmission. Le recoupement de corpus dans différentes bibliothèques numériques est alors un biais pour pallier les manques ou faire des recherches à l'articulation de plusieurs champs disciplinaires (histoire de l'art/histoire des techniques par exemple). Si les ressources numériques grâce à leur masse rendue aisément accessible permettent de vérifier et d'approfondir des hypothèses, elles ne permettent pas d'en ébaucher de nouvelles, selon les chercheurs interrogés (*« Une hypothèse est basée sur des connaissances qu'on a déjà, sur des questions qu'on pense qu'on va pouvoir résoudre. Je ne pense pas que les ressources numériques permettent d'avoir de nouvelles hypothèses. Mais en revanche, elles permettent de vérifier des hypothèses plus facilement, et donc parfois de les rendre plus ambitieuses, d'aller un peu plus loin dans ce qu'on peut trouver »*, Cnum3).

L'accès dédié aux images (notamment dans le cas de l'illustration technique, qui pourrait être extraite des volumes) est une demande récurrente et l'illustration technique grâce aux corpus numérisés pourrait alors devenir un thème de recherche en histoire des techniques : *« les corpus numériques permettent de travailler sur des corpus plus vastes, et nous conduisent à faire une recherche différente, puisque le corpus détermine la recherche. Ça renouvelle les corpus et les méthodes parce que, par le Net, il y a des méthodes qu'on ne pouvait pas avoir, comme des recherches par mot, par image... On espère que ça va se développer. Je pense que très certainement, si les images sont plus accessibles sur le Net, on verra l'importance du dessin, de l'image devenir un thème des historiens des techniques, plus qu'il ne l'était »*, Cnum3.

En revanche, **la navigation à l'intérieur de l'ouvrage**, le positionnement dans l'ouvrage en cours de lecture est jugé difficile à établir, la lecture linéaire d'un ouvrage numérisé étant exclue. L'affichage et le mode de visualisation des planches sont appréciés et reconnus comme simples et pratiques, la qualité de la numérisation étant toujours mise en avant. Les difficultés d'impression sont souvent pointées sans que ne soit curieusement jamais évoquée la possibilité du téléchargement en .pdf de la plupart des ouvrages qui permet un affichage et une impression à la demande.

Les nouveautés de la bibliothèque numérique sont envoyées sur des listes de diffusion propres aux chercheurs de la discipline, aux collectionneurs, aux collègues des musées nationaux et internationaux. **Les outils de valorisation autour de la bibliothèque** font l'objet d'avis mitigés : la présentation détaillée d'ouvrages est souhaitée par un

«Les ressources numériques permettent de vérifier des hypothèses plus facilement, et donc parfois de les rendre plus ambitieuses, d'aller un peu plus loin dans ce qu'on peut trouver»

chercheur (Cnum3⁹), les expositions virtuelles rejetées par le motif qu'il ne peut s'agir que « *d'un périphérique d'une exposition in situ* », Cnum1. Les « services » (glossaire, historiques, enrichissement des notices...) liés à la bibliothèque numérique posent une vraie question, car on constate, dans les logs de connexion, qu'ils sont très peu utilisés par les internautes : inutilité ou mauvaise adéquation des contenus ?

Cette enquête a confirmé pour les acteurs du Cnum la nécessité de :

- continuer à travailler en partenariat étroit avec les chercheurs pour la constitution de corpus pertinents en lien avec les axes de la recherche en train de se faire ;
- refondre l'interface de consultation autour de l'amélioration de la recherche et de la visualisation des illustrations techniques qui deviennent un objet de la recherche à part entière. Cette refonte est en cours.

La question de l'OCR, vivement souhaitée par certains chercheurs, reste coûteuse sur des corpus antérieurs à 1800, dès lors qu'une certaine qualité est requise. Le Cnum attend beaucoup des développements en cours au Centre d'Études Supérieures de la Renaissance de Tours (Agora) ou du projet européen Impact ((IMProving ACcess to Text) entre autres.

2.2 La Jubilothèque, synthèse réalisée par Frédérique Flamerie de la Chapelle, responsable de la bibliothèque numérique à la bibliothèque interuniversitaire de l'Université Pierre et Marie Curie.

Les utilisateurs de la Jubilothèque interrogés dans le cadre de la présente enquête ont été sollicités soit du fait de leur lien avec la bibliothèque (membre de commission d'acquisition, participation au travail éditorial de la Jubilothèque), soit parce qu'ils s'étaient manifestés lors d'une panne conséquente de la Jubilothèque en 2010. Trois des utilisateurs sont des chercheurs ou enseignants-chercheurs rattachés à des établissements parisiens, le quatrième réside en province et a concilié son activité de recherche avec une activité professionnelle, voir annexe 2.

Ces utilisateurs soulignent **l'accès rapide au document et la simplicité de l'interface** de la Jubilothèque. La multiplication des fenêtres ainsi qu'un champ URL inutile dans la notice bibliographique ont été déplorés. La Jubilothèque a aussi pâti de ses plantages fréquents avant le changement d'application : « *Pour télécharger les documents, je ne sais pas comment ils se débrouillent, mais ça ne marche régulièrement pas* », Jubil3. Les images revêtant une importance particulière pour l'un des chercheurs, ce dernier a apprécié la **navigation par lot de vignettes**, ainsi que celle par la table des matières : « *Dans Jubilothèque, vous tapez page et vous avez toutes les pages en miniature. Par exemple, pour chercher de la photographie, vous repérez facilement les pages où il y a des*

⁹ . Cnum3 : « Sinon, de manière générale, je trouve que ce serait bien d'avoir des présentations d'ouvrages. Quand on a, comme dans le Cnum, un corpus d'ouvrages qui n'est pas énorme, je pense que la partie outils pourrait être améliorée, en faisant systématiquement, comme dans le site Architectura, une notice sur l'ouvrage pour savoir si c'est la 1^{re} édition, la 4^e édition ».

photos. [...] *Plus la table des matières. Donc, on peut très rapidement aller à la page qui nous intéresse plus particulièrement* », Jubil4. **Toutefois, tous les utilisateurs donnent la priorité à la qualité de la numérisation et au nombre de documents numérisés sur les fonctionnalités diverses, qui semblent peu exploitées (le téléchargement de l'ouvrage entier est souvent privilégié).**

Du fait du nombre relativement restreint de documents qu'elle propose et de leur classement thématique, la Jubilothèque demeure une bibliothèque numérique qui peut s'appréhender « intuitivement », surtout si l'on est intéressé par un domaine seulement en particulier : « *j'ai quand même une idée relativement précise de ce qu'il y a dans la Jubilothèque. Il y a des choses que je n'irais pas y chercher, ou pas en premier parce que je vais me dire que ce n'est pas la peine parce que je sais qu'il y a peu de chance que ça y soit* », Jubil2. Des commentaires du même ordre sont formulés concernant d'autres bibliothèques numériques : « *Je ne connais pas 100% des choses, mais si je cherche le Journal de mathématiques pures et appliquées, c'est sur Gallica et pas sur Numdam* », Jubil1. L'utilisation d'un moteur de recherche n'est pas citée majoritairement comme mode d'accès, ou en tout cas s'il est utilisé c'est en sachant ce qui était cherché : « *Essentiellement, je suis passé par Google pour retrouver les ouvrages. Mais je connaissais ce que je cherchais* », Jubil1. L'OCR est considéré comme important mais également comme générateur de bruit, si les termes ne sont pas suffisamment spécifiques. La nécessité qu'il y a de connaître les documents préalablement à leur recherche et à leur consultation en ligne est un sujet récurrent, évoqué par une grande partie du groupe, au-delà des seuls utilisateurs de la Jubilothèque. Se dessine au travers des entretiens l'image de bibliothèques numériques, réservoirs de documents que l'on peut plus ou moins facilement consulter, à condition de les connaître préalablement. Une inquiétude pointe : comment les utilisateurs néophytes pourront naviguer dans cet océan de documents ?

«*Essentiellement, je suis passé par Google pour retrouver les ouvrages. Mais je connaissais ce que je cherchais* »

Il ressort enfin que les chercheurs font un usage intensif des bibliothèques numériques (consultation quotidienne dans deux cas), pour **retrouver des documents déjà connus** ou vérifier des références, ou encore découvrir d'autres documents, comme dans une bibliothèque traditionnelle : « *Ce sont des questions spécifiques, des recherches de documents particuliers, bien spécifiques justement, en espérant qu'ils seront disponibles en ligne. Il y a aussi un peu de curiosité de voir ce qui est disponible. Souvent, je suppose que c'est le cas, aussi bien quand on utilise une bibliothèque papier qu'une bibliothèque numérique, on trouve des choses un peu par hasard, on va les voir et ça part en cascade et on découvre des choses qu'on n'attendait pas au départ quand on a commencé les recherches* », Jubil2.

Au total, ces entretiens confortent l'usage pressenti de la Jubilothèque et se recoupent en partie avec l'analyse des statistiques de recherche de la nouvelle application (*i. e.* internautes ayant utilisé l'interface de la Jubilothèque entre février et fin mai 2011). Ces utilisateurs de la nouvelle Jubilothèque utilisent pour moitié la recherche « à la Google » (barre de recherche dans le texte intégral des documents de la page d'accueil) et pour moitié les formulaires de recherche avancée¹⁰.

¹⁰ Voir annexe 3.

Les critères de recherche utilisés le plus souvent (combinés ou non à d'autres) sont « Tous les mots de la notice » (39% des recherches) et « Auteur » (27%). On observe ensuite des spécificités en adéquation avec les particularités de chaque collection (prédominance du critère de date pour les publications de la Société géologique de France par exemple) et une utilisation des index relativement marginale.

Cette enquête a également confirmé l'importance pour la Jubilothèque de travailler autant que possible avec la communauté scientifique pour la constitution des corpus et de maintenir un niveau de numérisation de haute qualité, surtout pour les illustrations. Le peu d'attention porté aux fonctionnalités conforte le choix, renouvelé pour la nouvelle interface, de fonctionnalités simples.

2.3 Numdam, synthèse rédigée par Hélène Falavard, ingénieur documentaliste à la Cellule de Coordination Documentaire Nationale pour les Mathématiques (MathDoc).

Dans le cadre de cette enquête, quatre personnes ont été interrogées : trois chercheurs (un jeune maître de conférence, une directrice de recherche au CNRS et une professeure des universités) et une bibliothécaire d'une grande bibliothèque de mathématiques, voir annexe 2.

Une première constatation est **l'utilisation régulière, voire systématique**, de Numdam par les chercheurs dans leur travail quotidien : « *Je l'utilise presque tous les jours* » (Numdam2), « *Tous les jours : je suis complètement accro !* » (Numdam4). Que ce soit pour retrouver un article (« *C'était pour regarder un article paru aux publications de l'IHES. Je l'ai récupéré. C'était un article que j'étais censé avoir mais je ne l'avais pas sur mon ordinateur* », Numdam1), ou comme source primaire (« *je me sers des textes de maths disponibles du passé comme source primaire pour travailler* », Numdam2) ou encore pour consulter des documents difficiles à trouver (« *et il y a aussi des séminaires. Ce sont des choses difficiles à trouver, qui étaient dactylographiées, mal tapées. Certaines bibliothèques n'ont pas forcément la collection entière* », Numdam1), l'utilisation de Numdam semble aujourd'hui incontournable et bien ancrée dans les usages : « *Ça remplace la bibliothèque* », Numdam1.

«Tous les jours : je suis complètement accro ! »

Les deux modes de consultation des collections proposés sont utilisés par les utilisateurs interrogés : le feuilletage et la recherche. Pour ce qui est du feuilletage, l'accent est mis sur la possibilité qui est donnée aux utilisateurs de **consulter une revue ou un séminaire dans son intégralité et suivant l'organisation connue de l'imprimé** : « *Numdam fait une interface beaucoup plus agréable parce qu'ils ont des tables des matières avec les années, et on peut cliquer directement. Et on peut en même temps avoir les articles. Ça restitue quand même les logiques de recherche qu'on peut avoir en tant que personne* » (Numdam2). Pour chaque revue ou acte de séminaire, les volumes et fascicules sont classés par année de publication et cliquables. Sont ensuite listés l'ensemble des articles publiés dans le volume sous forme de notice abrégée. L'utilisateur a enfin la possibilité de cliquer pour afficher la notice complète qui lui

propose les références bibliographiques de l'article, un résumé, des mots clés ou codes MSC (Mathematical Subject Classification) lorsque ces informations sont disponibles. Une des fonctionnalités les plus avancées dans Numdam est le lien qui est proposé, pour les articles ou les références bibliographiques, vers la recension correspondante dans les bases de données *Mathscinet* ou *Zentralblatt Math* ou vers l'article lui-même s'il est disponible en ligne. Cependant, cette possibilité offerte aux chercheurs ne semble pas ou peu utilisée : elle n'est mentionnée qu'une seule fois par Numdam4 (documentaliste) : « [le] lien avec *MathScinet* [est] *fondamental* ». Il serait probablement opportun de communiquer plus largement autour de cet aspect de la bibliothèque numérique : cette fonctionnalité permet de lier les ressources importantes pour les chercheurs et d'élargir l'éventail des documents consultables en ligne.

Les fonctionnalités de recherche dites avancées sont appréciées pour leur capacité à restreindre le bruit éventuel engendré par les résultats de recherche : « *Par exemple, chercher un auteur au sein d'un journal, je trouve que c'est extrêmement précieux. Souvent, on sait ce qu'on a en tête, on sait l'auteur, on sait que probablement, on l'a vu dans telle revue. De temps en temps, on a une période de temps. Les recherches qui fonctionnent avec ce genre de données, c'est très rare* », Numdam2.

Il apparaît très nettement **qu'un des points d'accès naturel à Numdam est le moteur de recherche Google** : « *J'utilise beaucoup le moteur de recherche dans Google, parce que ça va plus vite. Et on tombe tout de suite sur les trucs de Numdam, même par le contenu* » (Numdam4). Ce constat soulève la question de la qualité du référencement de la bibliothèque numérique Numdam par Google. Actuellement, le référencement est fait directement sur les fichiers PDF (*archive.numdam.org/article/SCC_1958-1959__4__A2_0.pdf*) avec affichage, dans la liste des résultats, des premiers termes du texte intégral. Ce type de référencement, aujourd'hui peu optimal, devrait être amélioré. De nouvelles métadonnées (Dublin core, CoIns) vont être ajoutées au code HTML de façon à améliorer le référencement des notices de Numdam par Google et les autres moteurs de recherche.

«J'utilise beaucoup le moteur de recherche dans Google, parce que ça va plus vite. Et on tombe tout de suite sur les trucs de Numdam, même par le contenu».

Parallèlement à l'utilisation de Google comme point d'entrée à Numdam, les chercheurs interviewés indiquent utiliser de manière complémentaire les deux grandes bases de données de recensions en mathématiques que sont *Mathscinet* et *Zentralblatt Math*. Ces deux bases de données constituent généralement la première étape dans une recherche documentaire. Pour autant, comme signalé précédemment, les liens entre ces deux bases de données et Numdam ne sont pas suffisamment exploités par les chercheurs.

Un des principaux reproches formulés à l'encontre de Numdam est l'écart grandissant qu'il y a entre les numérisations et les mises en ligne : « *Le problème que j'ai principalement avec ça maintenant, c'est le retard terrible de Numdam par rapport aux mises en ligne réelles* » (Numdam2). D'autre part, un des chercheurs (Numdam1) mentionne le fait que certaines données ne sont pas mises en ligne dès que le créneau mobile¹¹

¹¹ . « Les revues dont les archives figurent sur [Numdam] ont accepté le principe d'un *créneau mobile*. Il s'agit de l'intervalle de temps qui sépare la publication d'un tome de la revue (sous la forme qui est

bouge : « *Le deuxième truc, c'est qu'il y a certains murs qui sont censés bouger mais qui ne bougent pas. Et s'ils n'ont plus d'accord, j'aurais aimé qu'ils le disent pour savoir à quoi s'en tenir* » (Numdam1). Enfin, le fait que les PDF des articles soient protégés et que l'accès au texte intégral soit bloqué¹² pose problème dans la mesure où cela rend impossible la copie de ce même texte : « *Le document était protégé. J'ai trouvé ça assez dommage, limite mesquin parce qu'il a fallu que je le retape à la main* » (Numdam1).

Malgré ces points négatifs, les entretiens ont confirmé que Numdam est un outil utile et utilisé : « *Numdam est un site extrêmement pratique* », « *Je suis vraiment très content de ce site* », Numdam1. La concertation avec les chercheurs sur les collections à numériser comme sur la façon de les mettre en ligne est particulièrement appréciée : « *Avec Numdam, ça se passe bien. Ils ont toujours tenu compte de ce qu'on leur a dit et ils nous ont donné des outils vraiment extraordinaires. Je leur suis très reconnaissante de ça. Ils vont mettre en ligne les thèses et on va essayer d'accompagner* » (Numdam1) ou « *c'est peut-être parce qu'il y a des mathématiciens derrière, mais ce qu'ils font correspond vraiment à ce dont nous, mathématiciens, avons besoin* », Numdam4. Pour conclure, voici ce qu'a dit un chercheur interviewé à propos de Numdam : « *Numdam, c'est comme avoir une étagère sur internet* », Numdam1.

« C'est peut-être parce qu'il y a des mathématiciens derrière, mais ce qu'ils font correspond vraiment à ce dont nous, mathématiciens, avons besoin ».

2.4 Medic@, synthèse réalisée par Jean-François Vincent, responsable du service d'Histoire de la médecine de la bibliothèque interuniversitaire de Santé (Paris)

Les trois usagers proposés dans le cadre de cette enquête ont été invités en raison de leur diversité. Ils représentent trois générations : un médecin ayant réalisé une thèse prosopographique après sa retraite (doyen du groupe interrogé), un chercheur en histoire des sciences de 43 ans en pleine carrière, une jeune chercheuse philologue ayant soutenu sa thèse il y a peu.

Les besoins documentaires de ces trois personnes sont très différents et leurs attentes à l'égard des bibliothèques numériques également. Le premier a réalisé un dépouillement extensif de la littérature d'une époque, le second s'attache à analyser des concepts (avec un besoin crucial du mode texte), la troisième travaille sur un corpus, limité par nature, de textes grecs et latins.

Pour donner une idée du contexte dans lequel se situe pour eux l'usage des bibliothèques numériques, il est intéressant de noter la **faible fréquentation des bibliothèques physiques par les deux chercheurs professionnels**. Elle est même plus faible que ce qu'ils ont déclaré lors de l'enquête. Medic@2 nous a indiqué que sa carte de la BnF n'était plus à jour. Même chose pour sa carte de la BIU Santé « *La dernière fois que vous êtes allé en bibliothèque, c'était où ? - A la Bium, mais ça fait un an que je n'y suis pas allé.* » Vérification faite, sa carte n'a pas été mise à jour depuis 2008 ! Il en va de même pour Medic@3 « *On a quand même une très bonne bibliothèque dans mon*

proposée aux lecteurs) de la disponibilité du texte intégral sur le serveur numdam.org. », <http://bit.ly/pTYkrL>.

¹². Ce choix a été fait pour garantir le respect du droit d'auteur (dans le cas des revues sous droit).

laboratoire. Ce qu'on n'a pas chez nous, je vais le chercher ailleurs, essentiellement ici [i.e. à la BIU Santé], à la BnF aussi. » « Maintenant que j'ai un laboratoire avec une très bonne bibliothèque, j'y vais peut-être un ou deux jours par mois [dans les autres bibliothèques]. » En réalité, sa carte de la BIU Santé n'a pas été renouvelée depuis 2008 également. Ce décalage entre discours et réalité est plus qu'anecdotique. S'il montre peut-être un peu de difficulté à avouer une fréquentation faible de lieux académiques valorisés, on peut faire l'hypothèse que **les bibliothèques numériques, prises globalement, évitent désormais nombre de déplacements** – au-delà même de ce que croient savoir les usagers eux-mêmes. Medic@2 indique que **les corpus disponibles en ligne déterminent désormais pour une part ses sujets de recherche** : dans son cas, les bibliothèques physiques sont donc devenues des ressources marginales. Un seuil, manifestement, a été franchi. *« Il faut bien dire que tout le monde utilise Google Books. Maintenant, ça devient la chose qu'on utilise le plus. C'est très récent qu'on l'utilise aussi massivement. - Vous le dateriez de quand ? - Moins d'un an, peut-être un an. Dans mon domaine, il y a eu une phase où on ne trouvait pas grand-chose, et tout d'un coup, on trouvait tout. Quand je dis tout, il y a des choses qu'on trouve sur Gallica et Medic@, qui ne sont pas dans Google Books. Mais de plus en plus, on trouve tout, y compris des livres 2010, qui sont en aperçu, mais parfois suffisamment pour pouvoir travailler. »* Medic@2

« Il faut bien dire que tout le monde utilise Google Books. Maintenant, ça devient la chose qu'on utilise le plus. C'est très récent qu'on l'utilise aussi massivement ».

Au contraire, pour Medic@1, il y a une continuité (jusqu'à la confusion dans les termes) entre la bibliothèque numérique et la bibliothèque physique : *« Est-ce que des échanges ont conduit à faire connaître auprès de vous Medic@ ? - Je vante le fonds allemand de la Bibliothèque interuniversitaire de médecine à Paris. »*

Les personnes interrogées ont des niveaux variables de connaissance des outils de la recherche documentaire en général. Tous les trois les exploitent abondamment et apparemment avec fruit, mais on perçoit que la connaissance théorique des outils est très contrastée – d'un usage totalement empirique (Medic@1), construit en tâtonnant au fil du travail de thèse, à une connaissance méthodique (Medic@2, qui a par exemple relevé les discordances dans les échanges de notices OAI entre Gallica et Medic@ et qui comprend de quoi il s'agit). Signalons que Medic@2 est particulièrement engagé dans Medic@ dont il participe à la sélection documentaire.

Le besoin de formation à Medic@ n'est formulé par aucun des usagers.

A- Une bibliothèque numérique qui fait partie du paysage

Medic@ fait partie depuis ses débuts de l'écosystème de nos deux chercheurs professionnels : Medic@3 l'a toujours connue depuis qu'elle est chercheuse (une dizaine d'années). Au-delà de son cas, selon elle : *« ...pour les étudiants en histoire de la médecine, c'est une évidence. L'histoire de la médecine grecque en l'occurrence. »* « C'est très connu dans mon domaine d'étude. » Il en va de même pour Medic@2 : *« Tout le monde est sans arrêt sur Gallica ou Medic@. On y est tout le temps. Dès qu'il y a une nouveauté, les gens sont au courant. »*

Réponses à nuancer dans le cas de Medic@2, qui précise à un autre moment (et avec un peu d'embarras face aux bibliothécaires français qui l'interrogent) que Medic@ vient en troisième lieu dans ses recherches : d'abord Google Books, ensuite Gallica,

enfin Medic@. Il ne semble pas qu'il y ait d'autres bibliothèques numériques qu'il utilise régulièrement (le nom d'Internet Archive par exemple ne lui dit rien).

B- Niveau global de satisfaction élevé

Le niveau de satisfaction exprimé est très grand. « *C'est un site qui fait référence en histoire de la médecine.* » (Medic@3) « *Le sens de ma question, c'est quelle est votre bibliothèque idéale en ligne ? - C'est encore celle-ci [Medic@]* » selon Medic@1. La numérisation est jugée de bonne qualité : néanmoins : « *Sur Medic@, elle est quand même un peu moins bonne que sur Gallica, mais franchement, ce n'est pas gênant.* » Medic@2. La rapidité de consultation est appréciée. La sélection paraît satisfaire les usagers, même si, bien entendu, une quantité plus grande serait souhaitée (« *On ne trouve pas toujours tout ce qu'on aurait souhaité.* » Medic@1) L'organisation en dossier est très appréciée par Medic@3, pour qui l'accès par auteur est primordial. Medic@1 la critique, lui, de façon modérée : « *Il n'y a qu'un truc, c'est qu'il y a des chapitres qui ne parlent pas tellement et qui méritent d'être étudiés de façon systématique pour retrouver des choses qu'on ne pensait pas y trouver.* ». La qualité des tables des matières est soulignée.

A part l'absence du mode texte de la part de Medic@2 seulement, les retours négatifs ont été très peu nombreux. Medic@1 reproche à Medic@ de contenir des pages blanches (« *C'est très gênant parce qu'on perd du temps.* ») Il repère aussi le défaut difficilement corrigé lors de la recherche dans l'indexation des périodiques et dictionnaires : « *...vous m'auriez donné la première page plutôt que la dernière, ça aurait été plus vite...* »¹³.

Le contexte des entretiens doit conduire à nuancer les appréciations louangeuses qui en ressortent : non seulement le responsable de Medic@ y assistait mais encore c'est précisément parce que ces usagers étaient contents qu'ils ont accepté de participer à l'enquête. Néanmoins, la bibliothèque numérique qui leur est proposée semble assez bien adaptée à leur attente.

C- Attentes des usagers

- **Faible niveau d'attente technique**

Les entretiens concernant Medic@, comme les entretiens qui ont concerné les autres bibliothèques numériques, ont montré que **les chercheurs n'attendent pas de fonctions avancées** (annotations, système de partage, intégration dans des réseaux sociaux...) **Ils cherchent avant tout des contenus, dans une relation somme toute très traditionnelle avec les bibliothèques.** Aucune demande n'a émergé pour une recherche avancée par exemple (la recherche par mots clés, dans Medic@, est une recherche globale sur les notices sans possibilité d'affinage.)

On doit garder à l'esprit que cela va de pair avec un faible niveau de connaissance technique. Medic@1, par exemple, ignorait qu'il pouvait imprimer des parties de documents, parce que l'icône indiquant la présence de PDF ne lui était pas compréhensible.

- **Signalement réciproque entre bibliothèques numériques**

Les trois chercheurs connaissent et apprécient l'existence de renvois entre Medic@ et Gallica : « *Parfois, on trouve dans Medic@ des choses qui sont sur Gallica et qu'on n'aurait pas trouvées sur Gallica.* » Medic@3.

Seul Medic@2 semble avoir également repéré le manque de fiabilité de ces renvois. : « Je sais qu'il y a des documents sur Medic@ qui n'étaient pas visibles depuis Gallica. J'ai toujours ça en tête. Je ne sais pas si ça a été mis à jour. Donc, j'utilise aussi le mode de consultation spécifique de Medic@. ».

- **Le mode texte**

Seul Medic@2 déplore, mais avec insistance, l'absence de mode texte dans Medic@. Les deux autres usagers n'en éprouvent pas le besoin.

Tout dépend de la nature des recherches effectuées bien sûr : une philologue (Medic@1) qui travaille sur des éditions grecques anciennes n'attend pas de la bibliothèque numérique ce qui lui paraît tout simplement impossible à obtenir. En revanche, l'historien des idées (Medic@2) qui travaille particulièrement sur les XIX^e et XX^e siècles, et qui a besoin de repérer des occurrences, voit à juste titre dans le mode texte un outil indispensable.

- **Les présentations des dossiers**

Les présentations des dossiers suscitent un intérêt très modéré. Elles sont approuvées, éventuellement parcourues. Mais il est clair qu'elles ne sont pas perçues comme d'un intérêt essentiel pour les chercheurs (« *Ça peut être intéressant pour le grand public, pour la pédagogie.* » Medic@2).

- **Une remarque marginale : la citation des sources**

Un détail curieux est revenu dans ces entretiens, concernant les modalités de citation des documents numériques, voir aussi §4.1 *infra*. On y voit que **les usages académiques n'ont pas encore suivi l'évolution de la documentation**. Medic@1, s'il cite un document consulté sous forme numérique, donne la cote de l'exemplaire imprimé. Medic@2, quant à lui, indique qu'il ne donne l'URL d'un document que s'il est rare en bibliothèque (« *Est-ce que vous localisez ou pas, quand vous avez à citer un document ? - Ce n'est pas systématique. La plupart du temps, on ne met rien, quand ce sont des ouvrages qu'on aurait pu consulter facilement en bibliothèque, on ne met rien.* »).

Est-ce le signe qu'il est délicat d'attribuer aux sources numériques la même validité qu'aux sources imprimées ? Les chercheurs pensent-ils que l'original imprimé reste plus crédible que sa « photocopie » numérique ? Et plus largement pourquoi cite-t-on : pour permettre aux lecteurs de consulter les documents utilisés ? pour « communiquer », au-delà d'un contenu, les preuves de sa solidité scientifique ? Le système de preuves change-t-il avec la mise en ligne des documents en ligne ?

¹³ En fait, dans ce cas, le moteur de recherche renvoyait deux liens, l'un vers la première, l'autre vers la dernière page des articles pertinents. Notre lecteur n'a pas cliqué sur le bon. Ce défaut est réellement perturbant mais

D- Stratégie d'avenir

Comment Medic@ doit-il se poursuivre ?

Sur la base de ces entretiens, et avec la limite très importante du mode image, il peut sembler que Medic@ n'a pas fait fausse route jusqu'ici : c'est un constat satisfaisant bien entendu.

Face aux projets de numérisation de masse, l'interaction avec les chercheurs semble essentielle : *« Pour Gallica et Medic@, vous pouvez vous différencier de Google Books en interagissant avec les chercheurs et par la pertinence des corpus. Eux [Google Books], je ne sais pas comment ils sélectionnent, mais à mon avis, c'est à la fois aléatoire et systématique. Ils peuvent passer à côté de certaines choses »* Medic@2.

« Pour Gallica et Medic@, vous pouvez vous différencier de Google Books en interagissant avec les chercheurs et par la pertinence des corpus. »

2.5 Conclusions intermédiaires

Ces quatre synthèses, issues de rédacteurs différents examinant des retours d'usages dans des contextes documentaires eux aussi fort différents, identifient cependant un ensemble de pratiques identiques que l'on rappelle brièvement ici :

- Les bibliothèques de « niche » sont appréciées avant tout parce qu'elles prennent en compte les besoins documentaires des chercheurs : face à la masse certes attractive des documents en ligne (l'attraction de Google Books est réelle), la bibliothèque de niche représente le segment rassurant de l'information organisée dans lequel on sait ce que l'on va trouver et sous quelle forme on va l'y trouver (d'autant que les chercheurs enquêtés ont été associés à la sélection).
- Dans le même esprit, les aspects techniques des bibliothèques numériques (leurs fonctionnalités de recherche, de visualisation, etc.) comptent beaucoup moins, aux yeux de ce groupe d'utilisateurs intensifs que la collection, les fonds, les corpus, même en mode image¹⁴, régulièrement observée, peut aussi être vue comme une défense de la « bibliothèque patrimoniale » elle-même, sous des atours qu'elle peut paraître perdre avec d'autres modes d'exposition de ses documents (le mode texte fragmente et dissémine les contenus). Il faut enfin noter que les fonctionnalités techniques propres aux bibliothèques numériques sont d'autant moins attendues que Google, presque systématiquement utilisé pour explorer les collections numériques, donne toute satisfaction en matière de recherche documentaire...
- Mais comment savoir si cette défense de la bibliothèque résulte d'un besoin réel, d'un biais propre à l'enquête (menée par des bibliothécaires à des fins d'évaluation de bibliothèques), ou encore de la symbolique forte de la bibliothèque ancrée au sein de cette population particulière que sont les chercheurs...

on n'en trouve pas le remède.

¹⁴ . Voir aussi § 4, *infra*.

Il n'en reste pas moins que, à la lecture de ces quatre synthèses, la spécificité et la raison d'être des bibliothèques numériques de niche se situent dans cette concertation étroite avec une communauté de chercheurs : c'est ce point qui distingue crucialement la politique de numérisation de Gallica ; c'est aussi sur ce point que les actions de coopération de la BnF avec les bibliothèques partenaires pourraient s'élargir pour soutenir non plus la seule numérisation mais aussi la valorisation documentaire et scientifique des collections numérisées.

3. Pratiques et évaluation de Gallica

Rappels :

- Le guide d'entretien semi-directif soumis au groupe d'utilisateurs interrogés comportait un volet de dix items, concernant l'usage et l'appréciation de Gallica.
- Sur les 15 personnes interrogées, seules 2 connaissent très mal Gallica, qu'elles utilisent peu¹⁵ ; les autres sont des utilisateurs assidus de Gallica, qu'ils pratiquent pour la plupart de longue date.

Ce sont les verbatim recueillis auprès de ces treize personnes qui sont ici synthétisés.

3.1 Pratiques de Gallica

A- La recherche dans Gallica

De façon générale, le moteur de recherche de Gallica est déclaré « facile » à trouver, même si la page d'accueil peut sembler à certains (Jubil2 en particulier) inutilement chargée et complexe.

L'interface de recherche simple est d'emblée privilégiée, autant par « facilité » que par « principe » : s'exprime en effet l'attente (voire l'exigence) de disposer – comme avec Google – d'une interface « simple », synonyme d'efficacité, autant pour la saisie des requêtes que pour le choix des résultats. Cette recherche « simple » est qualifiée d'« empirique » (Jubil1), permettant de « voir ce qui va tomber » (Cnum1) : la démarche générale consiste à faire « résonner » un terme, une chaîne de caractères dans le plein texte de Gallica. Même les bibliothécaires se laissent séduire par les facilités de la recherche simplifiée : « J'utilise la recherche avancée si je ne trouve pas ce que je veux. Mais, d'emblée, je suis devenue comme tout le monde, je fais tous les mots », Cnum4.

Cette « googlisation » - assumée - des pratiques de recherche en ligne conduit en outre nos usagers à **utiliser d'abord Google pour interroger Gallica**. Sur ce point, Numdam4 exprime de façon explicite ce que beaucoup d'autres interviewés ont laissé entendre : « Google devient presque un accès à Gallica. Je dis ça avec un peu d'ironie ! ». La plupart des chercheurs interrogés utilisent en effet cette possibilité pour accéder aux documents dans Gallica : les résultats sont jugés satisfaisants...

Il faut dire que le moteur Gallica suscite quelques insatisfactions.

Les performances du moteur Gallica peuvent être sévèrement jugées : de « pas très bonnes » (Medica2) à « catastrophiques » (Numdam4) ou « frustrantes » (Numdam2¹⁶).

Plus gênant est le sentiment de ne pas comprendre les résultats générés par le système : « Un collègue me faisait remarquer que parfois, même en mettant entre guillemets un bloc de mots, on a des réponses complètement à côté. Et c'est perturbant », (Cnum4 et Jubil3).

« J'utilise la recherche avancée si je ne trouve pas ce que je veux. Mais, d'emblée, je suis devenue comme tout le monde, je fais tous les mots. »

« L'accès à tout est merveilleux, sauf qu'on n'a accès à rien. J'exagère, mais parfois, on a ce sentiment. C'est extraordinairement frustrant. »

¹⁵ . Il s'agit du sujet le plus âgé (Medica1, 73 ans) et du sujet le plus jeune (Numdam4, 28 ans) !

¹⁶ . « L'accès à tout est merveilleux, sauf qu'on n'a accès à rien. J'exagère, mais parfois, on a ce sentiment. C'est extraordinairement frustrant », Numdam2.

Le surlignage décalé dans le texte est également « *perturbant* », « *très gênant* » (Medica2, Cnum4).

Enfin, tout aussi déstabilisante est l'expérience de ne pas retrouver dans Gallica ce qu'on y a précédemment trouvé... : « *Il y a des choses dont on sait qu'elles se trouvent dans Gallica et en mettant tous les mots-clés, on ne trouve pas* », Medica2, Jubil3, Medica3¹⁷.

Pour certains qui utilisent Gallica depuis les débuts, c'est la version actuelle de Gallica qui est en cause : « *il y a des livres qu'on trouvait avant sur Gallica et qu'on ne trouve plus sur le nouveau. Il m'a fallu six mois pour me rendre compte que ce n'est pas parce qu'ils avaient disparu, mais parce qu'ils étaient devenus introuvables (...) Il n'y avait aucun problème avec la précédente interface* », Numdam2, mais aussi Cnum4¹⁸ et Numdam3¹⁹. Pour d'autres, en revanche (Jubil4), il y a eu « *améliorations* ».

De façon générale, pour le groupe, savoir bien utiliser Gallica prend finalement beaucoup de temps et **c'est encore le catalogue général qui reste le meilleur moyen pour accéder aux documents...**

- « *Il faut toujours utiliser les opérateurs, sinon on a 3 000 réponses* » (Jubil3).
- « *Je sais qu'il faut mettre "publié par le secrétaire perpétuel", sinon on a 36 trucs qui arrivent avant. Quand on a l'habitude, ça va* », Numdam4, pour une recherche sur les Comptes rendus de l'Académie des sciences.
- « *Sur Gallica, on peut passer beaucoup de temps à chercher. On sait que quelque chose s'y trouve et on n'arrive pas à le trouver. Je commence toujours par utiliser le catalogue* », Medica3.
- « *Si je les cherche par le catalogue, je vais mettre la main dessus* », Numdam2.
- « *Parfois, je considère que c'est plus facile de passer par le catalogue de la BnF [...] Pour auteur-titre, je passe plutôt par le catalogue* », Jubil1.

Selon le groupe d'utilisateurs interrogés, pour bien interroger Gallica, mieux vaut éviter le moteur de recherche de l'application et privilégier soit Google soit le catalogue général de la BnF, autrement dit deux dispositifs d'accès à l'excellence peu égalée, l'un sur le plein texte, l'autre sur les données bibliographiques...

B- La consultation des documents dans Gallica

La principale difficulté ergonomique que signalent les utilisateurs est liée à l'absence d'un espace de l'écran dédié à la table des matières (ou à tout autre outil d'orientation dans le document). Cnum3 et Jubil1 évoquent un « *effet de désorientation* », lié à « *l'absence [...] d'outils [...] qui nous permettent d'aller et venir dans le texte, d'aller aux illustrations, de revenir...* ». Medica3 note, pour sa part que, « *par exemple, dans Gallica, on sait qu'on peut aller à la page 235, mais on ne sait pas ce qu'il y a* ». On est très vite

17. « Mais parfois, il y a des choses qui m'exaspèrent parce que je les ai trouvées une fois, et la fois d'après, je n'arrive plus à les retrouver, alors que je sais qu'elles y sont », Medica3.

18. « A chaque nouveau changement, je me disais que c'était mieux avant (...) ça a été, pas un traumatisme, n'exagérons pas, mais on ne s'y retrouve plus, c'était mieux avant. C'est peut-être très français comme réaction », Cnum4.

« perdu » : « Si on se dit : la table des chapitres, où est-ce que je voulais aller, déjà ?, on ne sait plus trop où aller, on perd la page », Cnum3.

Ce qui est attendu est une interface de recherche avec « deux volets » (Jubil1) dont l'un serait dédié à la navigation « raisonnée » dans le document et l'autre à la seule consultation. Faute d'un tel dispositif, l'exploration du document est perçue comme particulièrement « lente » (Medica2) et fastidieuse : « ça prend vraiment beaucoup plus de temps que de lire un livre qu'on a sous la main », Numdam3. Ces difficultés de lecture en ligne expliquent en partie une pratique intensive du téléchargement de fichiers (Jubil1, Medica2).

On le voit : **ce qui n'est pas satisfaisant, c'est la mise à plat des deux dimensions du livre** – le contenu et le repérage dans le contenu – tout est numérisé et restitué au même niveau et de la même manière. De ce point de vue, le mode Flash, très mal connu de notre groupe d'utilisateurs (une personne interrogée sur quinze), n'apporte pas le gain attendu : « Ça ressemble à un livre. Ça m'amuse parce ce qu'on vous propose, c'est quelque chose qui se rapproche de plus en plus de la forme livre, c'est-à-dire qui n'apporte pas la nouvelle pensée profonde de la forme numérisée », Numdam2.

Si l'absence de dispositif de navigation adaptée au « livre numérique » est globalement déplorée par l'ensemble du groupe d'utilisateurs interrogés, **les avis sont plus contrastés sur le sujet de la qualité de numérisation** : tout dépend de la discipline.

La qualité du mode image dans Gallica est jugée insuffisante pour les mathématiques : « vous pouvez reconstituer des mots manquants quand il manque une lettre. Mais s'il manque un bout de formule, c'est inutilisable. Dans certains cas, il y a des fractions de cette taille avec beaucoup de chiffres, à 300 points par pouce, ça ne se lit pas », Numdam2 (et Numdam4, notamment pour le *Journal de mathématiques pures et appliquées*).

Les premières numérisations de Gallica sont également difficilement exploitables en paléontologie (Jubil2) et en philologie (Medica3), tandis que la qualité disponible semble convenir pour des travaux en histoire « générale » des sciences (Medica2 : « Pour la visualisation, la qualité de l'image me convient »).

C- Les produits dérivés de Gallica et la dimension collaborative

Autant **la coopération numérique entre bibliothèques est bien perçue, connue et appréciée**, autant la participation de Gallica aux espaces collaboratifs hors bibliothèques ne suscite que peu d'intérêt de la part des personnes interrogées. L'interopérabilité de Gallica avec d'autres bibliothèques numériques²⁰ est effectivement connue de tout le groupe : reste qu'un seul usager (Medica2) privilégie Gallica comme moteur « fédéré » à l'ensemble des sources moissonnées ; les autres utilisateurs ont tendance à interroger spécifiquement chacune des bibliothèques numériques qui les intéressent.

19 . « J'étais plus à l'aise avec Gallica première mouture qu'avec ce qui avait pris la suite. Maintenant, je me retrouve avec une nouvelle interface que je n'ai pas expérimentée », Numdam3.

²⁰ . Au moment de l'enquête (décembre 2010-janvier 2011), Numdam n'était pas encore moissonné dans Gallica.

De façon générale, les usagers interrogés ne s'intéressent pas au compte Facebook et au fil Twitter Gallica²¹ ; il faut dire, qu'à deux exceptions près, « *Gallica, c'est le travail* » (Medica3). De la même façon, le blog est peu fréquenté (1 personne sur 15, Jubil4) et les dossiers peu utilisés (1 personne sur 15, Jubil1).

Dans l'ensemble, **les usagers interrogés ne manifestent que peu d'intérêt pour des produits de scénarisation ou d'éditorialisation des contenus** : ils privilégient des listes ciblées de documents mis en ligne, transmis par mail *via* leur liste de diffusion préférée (type liste Theuth²²).

Les pratiques développées sur Gallica par notre groupe d'usagers dessinent une **communauté d'utilisateurs très assidus depuis fort longtemps** : leur haut niveau d'appropriation à la fois des contenus et des techniques proposées se révèlent par la façon dont ils évitent les écueils de la recherche dans Gallica et déjouent les limites de la consultation en ligne. Les innovations proposées à travers et dans Gallica sont globalement bien perçues sans pour autant susciter un quelconque intérêt : leurs attentes restent basiquement celles de chercheurs envers une bibliothèque.

3.2 Evaluation de Gallica : points faibles, points forts et suggestions

Centré sur des bibliothèques numériques spécialisées en sciences et techniques, l'entretien semi-directif proposé à notre groupe d'usagers prévoyait également une évaluation, synthétique, des points forts, des points faibles et des améliorations envisageables de Gallica.

A- Points faibles

Contrairement à d'autres bibliothèques numériques, Gallica propose un traitement des documents en mode texte mais « *tout n'est pas ocrisé* », ce que déplorent Cnum1 et Jubil1. Ce qu'ils regrettent en particulier, c'est le manque d'information sur ces aspects, cruciaux pour la recherche : « *Est-ce qu'on fait des recherches OCR, sur 15% du contenu ou 85% du contenu ? Ça change tout. Est-ce que les revues sont comprises ou pas ? Plein de choses comme ça. Ça change la façon dont on interprète les résultats de manière importante* » (Jubil1) ; à noter cependant que l'exigence en matière d'informations de cette nature reste modeste : « *on ne peut pas et il serait illusoire de penser avoir des résultats bétonnés là-dessus. Mais ces aspects-là sont utiles pour un chercheur* », Jubil1.

Dans le même ordre d'idée, l'adressage des expressions trouvées en mode texte, très fréquemment défectueux ou décalé, peut créer déception et incompréhension : « *Quand le décalage est toujours le même, on sait qu'il est un peu trop en haut et à droite par rapport au mot, on arrive à retrouver facilement le mot. Mais quand le décalage est un peu aléatoire, on s'arrache un peu les cheveux* », Medica2 (et Cnum3).

« Est-ce qu'on fait des recherches OCR, sur 15 ou 85 % du contenu ? Ça change tout. Est-ce que les revues sont comprises ou pas ? Plein de choses comme ça. Ça change la façon dont on interprète les résultats de manière importante. »

²¹ . Voir aussi, sur ce point, la partie de cette étude consacrée aux pratiques en ligne

²² . Liste francophone de référence dans les domaines de l'épistémologie et de l'histoire des sciences, existe depuis 1996 : <http://theuth.univ-rennes1.fr/theuth-liste.html>.

Incompréhension encore quelques fois quand manque, de façon « *aléatoire* », une table des matières (Cnum1 et Jubil1). De manière générale, la qualité n'est pas suffisante : **les tables des matières devraient être beaucoup plus soignées dans Gallica, voire reprises à la main** (Medica2) ; une attention devrait être portée en particulier aux tables des matières des périodiques : « *Dans Gallica, il y a beaucoup de périodiques où on fait des liens aux tables des matières qui ne sont pas les tables des matières des volumes mais des numéros. C'est difficile, il faut tout vérifier* », Jubil1.

De façon générale, **ce sont les périodiques qui donnent le moins de satisfaction, en raison de leur taux important d'incomplétude** : « *Si une collection de périodiques n'est pas exhaustive ou aussi exhaustive que possible, elle perd beaucoup de son utilité* », Numdam2 et Jubil1.

De façon plus marginale, sont également signalés un classement des résultats « *moins bon que sur GoogleBooks* » (Jubil1) et une recherche des images « *particulièrement difficile* » dans Gallica (Jubil4).

On le voit, **les deux principaux points faibles de Gallica énoncés par notre groupe viennent d'un constat d'instabilité ou d'absence de systématique du dispositif**, ce qui tend à affaiblir le « contrat de lecture » et le « contrat de confiance » entre Gallica et ses publics spécialisés et fidèles, le constat serait certainement différent auprès d'autres segments de publics de Gallica²³.

B- Points forts

En même temps que l'instabilité du mode texte est dénoncée, le **mode texte lui-même constitue, pour presque tout le groupe, l'un des points forts indéniables de Gallica, notamment par comparaison avec les services offerts par les autres bibliothèques numérisées spécialisées** de l'enquête (deux sur quatre choisissent la numérisation en mode image seul) : « *Pour Gallica, la qualité, c'est la recherche en mode texte dans des corpus aussi bien sélectionnés et très diversifiés. C'est un gros point fort* », Medica1 (Jubil1, Medica2).

La qualité et la variété des sélections documentaires sont effectivement unanimement reconnues comme un autre point très fort de Gallica et en font une référence incontournable pour le groupe interrogé : « *Dans le domaine des sciences, Gallica n'a pas à rougir de ce qu'il y a. C'est une collection très importante, où il y a énormément d'ouvrages classiques. Il y en a peut-être qui manquent, mais vraiment, pour chercher des ouvrages importants en histoire des sciences, on a beaucoup de choses sur Gallica* », Jubil1 (Numdam2 et Medica2).

La variété des sources, sur fond d'un corpus scientifique solide, est particulièrement appréciée : « *Je tiens à souligner que Gallica a changé ma vision de l'histoire des sciences, même en étant formé en histoire des sciences. Et je pense même que, de façon plus large, Gallica va changer notre rapport à l'histoire. On a un accès très rapide à une énorme variété de sources, auxquelles on n'avait pas accès auparavant, ou difficilement. On se rend compte que le paysage culturel, littéraire, scientifique, est très dense, très riche. Il y a plein d'auteurs*

« *Si une collection de périodiques n'est pas exhaustive ou aussi exhaustive que possible, elle perd beaucoup de son utilité.* »

« *Pour Gallica, la qualité, c'est la recherche en mode texte dans des corpus aussi bien sélectionnés et très diversifiés. C'est un gros point fort.* »

« *Gallica a changé ma vision de l'histoire des sciences [...] Ça apporte beaucoup de nuances et de subtilités aux études que l'on peut faire.* »

mineurs auxquels on prêtait peu attention auparavant qui, tout d'un coup, redeviennent plus visibles [...] Ça apporte beaucoup de nuances et de subtilités aux études que l'on peut faire. », Jubil1 (Medica1).

C- Suggestions d'amélioration

En écho à l'un des points faibles précédemment énoncés, une suggestion d'amélioration concerne **l'exposé des critères de sélection des corpus mis en ligne sur Gallica**, qui garantit une utilisation scientifique : « *Il faut qu'on ait une vision assez précise de la manière dont ça fonctionne, des rouages. Pas besoin de connaître les aspects très techniques, mais quand même [...] je n'ai jamais vu de documents un peu politiques sur comment est-ce qu'on numérise les choses. Les choix ne sont pas neutres* », Jubil1²⁴.

Une mise en perspective non seulement des corpus mais aussi des documents mis en ligne serait également appréciée, de façon à **restituer le contexte de production d'une publication** : « *Pour moi, il faut garder toutes les classifications [documentaires et scientifiques] de toutes les époques. Il faut qu'elles soient disponibles. Et je trouve totalement inepte les tentatives de balayer les classifications antérieures, de les faire disparaître [...] sous prétexte que les mathématiciens, à une époque donnée, ont une certaine vision du domaine. A mon avis, c'est la meilleure façon d'engloutir les possibilités de réactivation dont je parle* », Numdam2 : cette chercheuse en mathématiques explique que c'est cette recontextualisation dans un système de savoirs historiquement situé qui permet aux sources de rester « *actives* » et « *réactivables tout le temps [...] par exemple, cette source a été réactivée 5-6 fois, pour des problématiques complètement différentes* ». Cnum4 déplore, de son côté, « *une mise à plat* », tandis que Medica3 suggère de lier aux documents numérisés les fiches bibliographiques des manuscrits (quand elles existent).

Sur le même plan qualitatif, **un lien plus étroit avec les communautés d'utilisateurs paraîtrait particulièrement adapté** : « *Ce qui a été fait sur Gallica sur les périodiques de math ça ne correspond pas vraiment à ce dont on a besoin* » contrairement à Numdam « *C'est peut-être parce qu'il y a des mathématiciens derrière [Numdam], mais ce qu'ils font correspond vraiment à ce dont nous, mathématiciens, avons besoin* », Numdam4²⁵.

Quelques (rares) suggestions plus techniques sont énoncées : on peut relever le souhait d'un système de visualisation et de sélection d'images aussi simple que Picasa (Jubil4), le confort qu'offrirait une liste de résultats récapitulant les contextes d'occurrences d'une recherche : « *Ce qui serait bien, comme dans Google Books, ça serait la présentation en quelques lignes de toutes les occurrences* », Medica2. Pour les rares personnes interrogées qui utilisent un logiciel de gestion de références bibliographiques, un

²³ 34% des 3828 Gallicanautes interrogés au cours de l'enquête qui s'est déroulée au printemps 2011 déclarent utiliser Gallica depuis moins de deux ans (dont 4,3% depuis moins d'un mois).

²⁴ . Jubil1 rappelle qu'il était utile de disposer de listes de documents : « *Quand il y avait des listes de journaux, de quotidiens, de périodiques en mathématiques, en physique, en chimie, c'est utile. Juste avoir la liste des quotidiens* ».

²⁵ . L'exemple typique est celui de la numérisation du *Journal de mathématiques pures et appliquées*, revue française de référence en mathématiques : mise en ligne par volume dans Gallica (<http://bit.ly/nKHS13>), la revue numérisée par la BnF fait aussi l'objet d'un dépouillement à l'article dans Gallica-Math (Numdam-Mathdoc : <http://math-doc.ujf-grenoble.fr/JMPA/>) : on peut également interpréter ce double traitement comme un exemple réussi de coopération numérique. Accès à Gallica-Math : <http://portail.mathdoc.fr/GALLICA/>.

déchargement compatible avec leur système serait apprécié : Jubil3 (EndNote), BibTex (Numdam4 et la communauté des mathématiciens en général).

3.3 Conclusions

Le niveau de satisfaction exprimé au sujet de Gallica, globalement moyen, doit être analysé à la lumière de deux caractéristiques de l'enquête :

- le profil des enquêtés, d'une part : usagers « captifs » de Gallica par leurs orientations de recherche, les personnes interrogées sont d'autant plus exigeantes envers le « système » Gallica (recherche et exploration en ligne) que le fonds documentaire leur est précieux et utile. Les insatisfactions au sujet des modes d'accès ou de la lecture en ligne sont régulièrement exprimées et déplorées... sans pour autant se traduire par un abandon de consultation de Gallica, tout simplement inenvisageable dans leur cas ;
- d'autre part, l'enquête portait, de façon centrale, sur leur usage intensif d'une bibliothèque numérique spécialisée, « de niche » : de façon générale, les personnes interrogées défendent avec beaucoup de chaleur ces réservoirs dédiés, cette défense pouvant conduire à grossir le trait critique envers Gallica (défendre le mode image par l'argument que le mode texte de Gallica est vraiment trop bruyant, par exemple).

Ces réserves faites, il importe de noter la réelle difficulté exprimée à mener une recherche fructueuse dans Gallica, difficulté d'autant plus crédible que le groupe interrogé est doté d'une réelle expertise en matière de recherche documentaire. De la même façon, on doit être attentif à améliorer la qualité d'exploration des documents : les outils d'orientation dans les contenus (tables des matières, index, tables des illustrations, etc.) sont régulièrement évalués comme déficients dans Gallica.

On notera enfin la modestie des suggestions d'amélioration et on retiendra la nécessité de mettre en ligne « *un document politique* » explicitant les sélections documentaires, « *jamais neutres* ».

4. Conclusions intermédiaires : *d'une bibliothèque à l'autre, points communs et différences ; faits saillants ; retour sur l'hypothèse initiale : complémentarités ou/et concurrences entre bibliothèque spécialisée et bibliothèque encyclopédique.*

On retrouve, dans les verbatim consacrés à Gallica, des postures et des pratiques déjà relevées à propos de l'usage des bibliothèques numériques de niche : la principale motivation de visite et le principal motif de satisfaction tiennent à la qualité des corpus numérisés ; les fonctionnalités techniques intéressent peu ce public de chercheurs qui recourt massivement à Google pour identifier pages, documents ou corpus de travail.

Fait nouveau sur ce point : on découvre, avec ce volet consacré à Gallica, le catalogue bibliographique (quand celui-ci est accolé à la bibliothèque numérique) comme complément inattendu à Google. On en déduit que l'actuel système de recherche d'information de Gallica – hybride (un catalogue bibliographique sans toutes les qualités du catalogue général BnF, une recherche plein texte de qualité bien inférieure à Google) – souffre terriblement de la concurrence des autres dispositifs d'accès existants : les critiques sont particulièrement sévères et massives²⁶.

A ce titre, on doit également interroger le sens des critiques adressées à Gallica par notre groupe d'utilisateurs interviewés... par des enquêteurs BnF.

En effet, tous les chercheurs interrogés sont de gros utilisateurs de Gallica, depuis fort longtemps ; pour beaucoup d'entre eux, leur usage est quotidien et bien plus assidu que celui de leur bibliothèque numérique de niche préférée. Cette satisfaction, de fait, s'exprime insuffisamment dans l'ensemble des entretiens : quel message nous est-il adressé ?

Plusieurs analyses peuvent être avancées :

- Gallica est clairement identifié par notre groupe d'utilisateurs comme une production BnF, relativement opaque dans ses principes de sélection et de mise en ligne : outre l'exigence (légitime ?) attendue à l'encontre de la BnF, le groupe interrogé regrette-t-il, en creux, de ne pas retrouver dans Gallica la chaleur de l'« entre-soi » propre aux bibliothèques de niche, faites pour et surtout avec les chercheurs ?

La question pour la BnF est, dans ce cas, celle de son positionnement explicite envers cette communauté de chercheurs qui a besoin d'elle mais qui constate que la BnF, elle, ne semble pas avoir besoin d'eux...

Un axe concret de travail avec les communautés scientifiques autour des fonds numérisés pourrait s'appuyer sur les réseaux déjà constitués par les bibliothèques partenaires - projet de valorisation scientifique des collections par exemple.

- Un autre ensemble de critiques adressées à Gallica concerne le mode texte : or parallèlement, le mode texte est plébiscité comme point fort de Gallica alors même

²⁶ . Les interfaces de recherche des bibliothèques de niche présentent, elles, d'autres intérêts (voir ci-dessous), qui semblent les soustraire à la critique.

que le mode image des bibliothèques numériques de niche n'est jamais, dans de fortes proportions, déploré. Ces positions paradoxales révèlent bien plus que la seule affection pour « sa » bibliothèque numérique sur-mesure ; elles révèlent une réelle difficulté à lâcher le modèle maîtrisé et maîtrisable de la bibliothèque - interrogeable par mots de la notice et feuilletable par tables des matières - pour le « non-modèle » bruyant du plein-texte. Certes, la recherche en texte intégral est bien le seul moyen de faire des découvertes et d'aller au-delà des corpus connus, mais comment précisément valider ces corpus fortuitement construits sur la base d'une similitude de chaîne de caractères ? On aime que Gallica nous permette de faire « résonner » nos termes de recherche dans tous les textes disponibles mais on n'aime pas perdre l'« épaisseur documentaire » qui contextualise les contenus, on n'aime pas non plus ignorer les raisons du choix de Gallica pour tel corpus ou tel auteur.

La question qui se pose, dans ce cas, à la BnF, est celle des dispositifs de médiation que les bibliothèques numériques peuvent mettre à disposition des chercheurs : organisation des savoirs adéquate aux époques comme aux disciplines (classifications spécifiques), information régulière sur les corpus mis en ligne et les modalités de sélection documentaire par exemple, etc.. Sur le premier volet, le savoir-faire des bibliothèques partenaires paraît adapté : l'intérêt de leurs interfaces de recherche tient, le plus souvent, à la finesse et la spécialité d'un plan de classement qui donne aux pièces numérisées toute leur intelligibilité. Sur ce point encore, la coopération numérique entre la BnF et ses bibliothèques partenaires pourrait être enrichie, au-delà de la numérisation, par ce volet de valorisation documentaire.

Comme on le voit, les deux types de dispositifs – Gallica et les bibliothèques de niche – sont, aux yeux des usagers, complémentaires sur plus d'un point : en termes d'offre documentaire, bien sûr, mais aussi en termes de modalités d'accès à l'information : interface de recherche différenciée, comme on l'a vu, mais aussi traitement documentaire différencié. L'exemple, devenu classique de Gallica-Math²⁷, incarne cette complémentarité des compétences et des moyens : à la BnF revient la numérisation de masse, à MathDoc le dépouillement fin des volumes massivement numérisés. Et c'est ce dernier accès que les chercheurs privilégient.

De telles complémentarités, au-delà de la seule numérisation, constituent vraisemblablement des exemples à suivre pour élargir la coopération numérique à des opérations conjointes de valorisations documentaire et scientifique des collections numérisées : les bibliothèques de niche sont à même de reconstituer des ensembles documentaires intelligibles et cohérents pour une communauté ciblée, ce qui ne constitue pas la mission première de Gallica.

Cette étude d'usage, si elle ne permet pas de pointer à proprement parler des usages spécialisés de Gallica, permet, en tout cas, de les imaginer, notamment sur la base des quelques exemples disponibles.

²⁷ . cf. note 25.

5. Caractéristiques des usagers intensifs des bibliothèques numériques scientifiques

5.1- Aperçu de leurs pratiques en ligne

Le guide d'entretien semi-directif consacrait deux ensembles importants de questions d'une part, aux motifs d'un recours massif aux corpus en ligne et, d'autre part, aux pratiques du web développées par un groupe d'assidus des bibliothèques numériques patrimoniales en sciences et techniques.

Classés en vingt items, les verbatim recueillis permettent d'identifier en particulier :

- les bénéfices que tirent ces utilisateurs d'un usage intensif des corpus numériques, tant d'un point de vue matériel qu'intellectuel,
- les modalités de lecture des documents numériques,
- le niveau de technophilie et d'intérêt pour les nouvelles technologies de l'information et de la communication manifesté par ce groupe spécifique d'usagers de documents numériques.

A- Impact des corpus numériques sur le travail scientifique

Fréquentant intensivement les collections imprimées comme les corpus numériques, **les usagers interrogés formulent les avantages (et les limites) de disposer d'une documentation en ligne essentiellement par contraste avec le travail qu'ils mènent habituellement en bibliothèque sur place.** Cette comparaison fait apparaître trois types d'impact positif des corpus numériques sur le travail scientifique : **l'amélioration des conditions de recherche, l'élargissement de la couverture documentaire explorée et de nouvelles modalités d'exploration textuelle des documents.** Ces avantages du support numérique ne disqualifient cependant pas le support imprimé : le travail simultané sur les deux supports reste largement plébiscité par de nombreux chercheurs. C'est la complémentarité plus que la substitution qui est le plus souvent évoquée.

A.1- Impact sur les conditions de travail

Ce sont d'abord les conditions de travail du chercheur qui sont nettement améliorées, désormais affranchies des contraintes de déplacement, d'horaires, de disponibilité et de nombre de volumes consultables, propres au travail sur place en bibliothèque, (Numdam2, Cnum1, Jubil2, Medica1). Le temps auparavant consacré à la « préparation » du séjour en salle d'étude (anticipation des demandes de communication par exemple) peut désormais être consacré à des travaux qui paraissent plus centraux.

Cette absence de contraintes matérielles donne le sentiment de recherches plus « rapides »²⁸ et plus « efficaces », avec un va-et-vient plus fluide entre phases de

« Toutes les recherches en ligne permettent de faire dix fois plus de découvertes que d'aller en bibliothèque. »

²⁸ . Medica3 : « Si j'ai besoin de travailler le dimanche, les bibliothèques sont fermées. Si j'ai besoin de voir quatre éditions en même temps pour les comparer, il y en a une qui est à la Bium, l'autre à la bibliothèque de

recherche et phases de réflexion : « *on ne fait pas d'une part la recherche, et ensuite le cours. On fait les deux en même temps* », Jubil1.

Plus rapide et plus efficace²⁹ qu'en bibliothèque sur place, le travail à distance permet de consulter plus de documents : « *C'est l'efficacité des recherches. C'est pour ça que toutes les recherches en ligne permettent de faire dix fois plus de découvertes que d'aller en bibliothèque* », Medica2 (Jubil3³⁰).

A.2- Impact sur l'exploration documentaire

A ce gain matériel et quantitatif s'ajoute la perception d'un gain qualitatif (Jubil3). La facilité et la rapidité d'accès encouragent en effet la « *rigueur* » scientifique³¹ et la « *curiosité* » intellectuelle. De façon quasi unanime, le groupe interrogé souligne **l'impact positif de la « découverte hasardeuse » que permet la recherche sur les corpus numériques** : « *On découvre des choses auxquelles on n'aurait pas pensé, un peu par hasard* » (Jubil2), « *on découvre des trucs qu'on n'avait pas forcément imaginé* » (Jubil4), « *ça ouvre des pistes qu'on ne connaît pas* » (Cnum3) ; on peut « *chercher des choses qu'on ne connaît pas forcément* » (Jubil2), se risquer à des recherches (« *c'est une façon très rapide de savoir si le mot dinosaure apparaît quelque part dans le volume, même si je ne suis pas du tout sûr qu'il y aura quelque chose* », Jubil2), élargir son champ d'intérêt : « *Ça permet aussi de s'intéresser à des choses auxquelles on ne se serait pas intéressé* » (Medica3). Bref, on trouve ce que l'on ne cherche pas, sans avoir le sentiment de perdre son temps (Medica1). Ce phénomène régulièrement observé et désigné par le terme de sérendipité est-il lié, dans le cas présent, à la disponibilité du mode texte ? Il y a en effet toujours à lire avec les bibliothèques numériques (les réponses donnent toujours accès aux documents primaires), ce qui n'est pas forcément le cas de la consultation du seul catalogue, dont les notices bibliographiques peuvent susciter quelques frustrations.

L'exploration documentaire large que facilite la mise à disposition libre de documents numérisés souligne le double rôle du corpus dans les pratiques scientifiques développées par notre groupe : à la fois point d'arrivée, bien sûr, de la démarche de recherche documentaire mais aussi point de relais d'une réflexion en train de se construire, qui rebondit et s'enrichit au rythme des découvertes fortuites. Les bibliothèques numériques participent, de ce point de vue, à ce sentiment grisant, propre au web, que la pensée peut progresser au rythme du clic sur les documents...³²

« On découvre des choses auxquelles on n'aurait pas pensé, un peu par hasard. »

l'Académie de médecine : je suis embêtée. Là, j'ai mes pdf, je peux les comparer. C'est dans ce sens que je parle d'un gain de temps ».

²⁹ . Cnum3 : « *Ça va beaucoup plus vite - vérifier tout de suite si la chose est intéressante ou pas - avant, il fallait perdre une journée à aller dans une bibliothèque, pour finalement découvrir que ça n'était pas intéressant* » ; Medica2 : « *Travailler sur des textes imprimés, maintenant, ça paraît tellement lointain (...) ce sont des après-midi entières à photocopier, des journées entières même. On n'est pas du tout dans la même fenêtre de temps* ».

³⁰ . Jubil3 : « *Quand on a un ouvrage très long et qu'on veut savoir si l'auteur parle de telle chose, avoir accès à la ressource texte, c'est fantastique. Avant, on était obligé de lire tout le livre pour arriver à la énième page où se trouve le mot-clé qui nous intéresse* ».

³¹ . Medica3 : « *Ça nous permet d'être plus rigoureux, de mieux vérifier nos informations, de mieux retourner à l'original, éviter de faire des citations de citations...* ».

³² . Numdam2 rappelle, opportunément, que le « *hasard heureux* » s'éprouvait aussi avec les collections imprimées : « *Selon la classification de la bibliothèque, vous voyez le livre à côté de celui que vous cherchez. Et on est censé avoir perdu ça dans le numérique. Et il y en a d'autres, par exemple des homonymes qui parfois*

Sur ce point, nos chercheurs ne perdent cependant pas de vue qu'il leur revient de mettre du « relief » là où la numérisation tendrait à « une espèce d'aplatissement de toutes les informations » reliées les unes aux autres. C'est encore ce que souligne Jubil 1. « On a un accès très rapide à une énorme variété de sources, auxquelles on n'avait pas accès auparavant, ou difficilement. On se rend compte que le paysage culturel, littéraire, scientifique, est très dense, très riche. Il y a plein d'auteurs mineurs auxquels on prêtait peu attention auparavant qui, tout d'un coup, redeviennent plus visibles. Le danger de l'érudition est d'aplatir tout ça. [...]. Le travail de l'historien reste d'essayer de mettre du relief dans ce paysage-là. Mais le paysage s'est énormément densifié. Ça apporte beaucoup de nuances et de subtilités aux études que l'on peut faire. »

« Le travail de l'historien reste d'essayer de mettre du relief dans ce paysage-là. Mais le paysage s'est énormément densifié. »

A.3- Impacts sur les méthodes d'exploration textuelle

Les bibliothèques numériques ont un impact encore plus décisif pour certains chercheurs. Numdam4 et Jubil4 confient que la direction de certains de leurs travaux est intimement liée à la disponibilité des corpus en ligne ; Medica2 renchérit en disant que « le choix [des] sujets est fortement influencé par ce qui est numérisé et ce qui ne l'est pas. (...) Il y a des sujets auxquels vous renoncez parce que les éléments ne sont pas en ligne » ; la nécessité de connaître les critères de constitution des corpus en ligne est d'autant plus importante (Jubil1). De façon un peu plus générale, Cnum3 remarque que les corpus numériques « conduisent à faire une recherche différente » qui « renouvelle les corpus et les méthodes (...) il y a des méthodes qu'on ne pouvait pas avoir, comme des recherches par mot ». En effet, ce type de recherche, particulièrement délicate à conduire sur des corpus imprimés, s'impose avec les capacités qu'offre le plein texte, et ce quels que soient la discipline ou encore l'objet de recherche (Jubil4 sur la notion d'« idiot », Medica2 sur celle de « neurone »). Cette **clé lexicale de sondage** « [qui] permet de lier tous les livres de la planète ensemble » (Cnum3) ouvre en outre des perspectives pluri- ou trans-disciplinaires, identifiées comme « riches » par une partie du groupe.

Mais tous les usagers de notre groupe ne sont pas convaincus par les bénéfices de cette recherche par mots, perçue comme « aléatoire » pour Cnum1, beaucoup trop bruyante pour Numdam2³³, dérisoire pour Numdam1 qui rappelle que la seule façon de savoir si on a appris ou compris quelque chose, c'est « quand on est capable de le refaire tout seul » ; en mathématiques en l'occurrence, la lecture « savante » consiste à pouvoir refaire « seul » la démonstration que l'on vient de lire. Le corpus, aussi abondant soit-il, ne suffit évidemment pas à faire une bonne recherche (en mathématiques *a fortiori*).

Les plus enthousiastes signalent, de leur côté, le risque de mirage du plein texte : « Je n'irais pas jusqu'à dire qu'on ne découvre que ce qu'on a imaginé, mais c'est un peu ça. On définit tellement le type de recherche que finalement, peut-être qu'on restreint ce qu'on est amené à découvrir », Medica2.

Autant réjouis des « découvertes » que méfiants envers les « hasards » du plein texte, les chercheurs du groupe interrogé ne se départissent ni du doute ni de la vigilance

« Je n'irais pas jusqu'à dire qu'on découvre que ce qu'on a imaginé, mais c'est un peu ça. On définit tellement le type de recherche que finalement, peut-être qu'on restreint ce qu'on est amené à découvrir. »

sont liés, des auteurs qui sont frères, et on commence à découvrir tout un réseau familial. Ça peut être intéressant. Des groupements de sujets inattendus, qu'on a repérés par le numérique plus facilement. »

³³ . Numdam2 : « La création plein texte sur un corpus d'autant d'occurrences, c'est une ineptie. C'est ingérable pour le moment ».

propres au travail scientifique, n'hésitant pas à expérimenter, jusque dans leurs ultimes limites, les possibilités du mode texte, comme Jubil³⁴.

Certains, enfin, récusent complètement le bien-fondé de la recherche « chaîne de caractères » dans le tout-venant du plein texte (Cnum1 : « *Je n'agis pas par curiosité pour découvrir quelque chose* »), recourant, y compris pour la recherche dans une bibliothèque numérique, au travail préalable de dépouillement bibliographique³⁵, le seul qui garantisse la « découverte » (Jubil3 : « *Alors que si on fait vraiment de manière bien ordonnée, on couvre des zones d'ombre dans la bibliographie qui permettent de découvrir des choses qui changent la vision qu'on a du problème* »).

A.4- Impact sur la lecture et l'exploitation des documents ?

L'impact des corpus numériques peut être d'autant plus décisif que leur consultation est couplée avec celle de l'imprimé, **ce cumul des pratiques**³⁶ **est notable dans le groupe interrogé** (Numdam3, Cnum1, Numdam4), ne serait-ce que « *parce qu'on s'aperçoit qu'on n'utilise pas du tout un document numérique comme on utilise un document papier* » (Cnum4) ; Numdam2 estime de son côté que « *les services en ligne ne sont absolument pas suffisants pour qu'on supprime les bibliothèques papier* », tandis que Cnum3 relève même que la prise de connaissance de l'architecture d'ensemble d'un ouvrage est un « *travail particulièrement long* » quand il est numérisé « *alors qu'en réalité, ça ne l'est pas* ».

Sur un certain nombre de points, **l'imprimé reste inégalé** (Numdam2, Cnum1) : plus « *lisible* » (Medica3), « *plus facile à manipuler* » (Cnum3, Medica3), il reste en outre le support privilégié de la lecture de loisir, indépendamment de l'âge (Numdam1, 28 ans : « *Pour le reste, honnêtement, je préfère aller acheter un bouquin en librairie et le lire, plutôt que de me tuer les yeux sur un écran à mettre deux heures pour tourner chaque page* »)³⁷.

L'imprimé reste en outre, le plus souvent, le support privilégié de la lecture approfondie : « *Quand j'ai besoin de lire vraiment, j'ai besoin du support physique. Je ne lis jamais en ligne. La lecture est un mode d'appropriation* », Medica2 (de même Medica1 et Jubil3).

« Quand j'ai besoin de lire vraiment, j'ai besoin du support physique. Je ne lis jamais en ligne. La lecture est un mode d'appropriation. »

³⁴. Jubil1 : « J'ai écrit un article que j'ai appelé mon article Google Books, parce que je pense que je n'aurais pas pu l'écrire autrement. (...) J'ai [ensuite] étudié de manière assez systématique toutes les références qui m'étaient proposées. Et même celles qui n'étaient pas accessibles en ligne, parce que de toute façon, je pouvais récupérer ces ouvrages autre part, en notant les numéros de pages et les références. C'est ce que j'ai fait pour un certain nombre d'entre eux. Évidemment, il y a beaucoup de redites, mais on peut aussi isoler les grands corpus. Et on se fait rapidement une idée de comment les choses s'organisent les unes par rapport aux autres. Et on peut ensuite écrire cet article, faire cette étude ».

³⁵. Pour certains chercheurs, notamment Jubil3 et Cnum1, la découverte tient à l'exploration systématique des instruments bibliographiques existants et pertinents pour la recherche et non à l'exploration systématique des occurrences trouvées à l'issue d'une requête dans une bibliothèque numérique. Pour eux, la bibliothèque numérique serait plutôt l'équivalent d'une étagère en ligne d'où ils tirent une partie des documents recherchés, rien de plus.

³⁶. Numdam4 : « Ce qui est bien, c'est de cumuler les deux. Très souvent, je suis à la bibliothèque avec mon ordinateur. J'utilise le wifi de la bibliothèque. Avec mon ordinateur, je cherche où c'est et après, je vais lire le volume. Pour moi, c'est la grande richesse. Ça accélère la recherche. Pour lire un article, c'est beaucoup plus confortable de pouvoir tourner les pages, revenir. (...) J'utilise autant la bibliothèque papier qu'avant, mais je l'utilise de façon beaucoup plus efficace parce que j'ai ça ».

³⁷. Voir aussi sur ce point le paragraphe 5.2 de cette étude.

La lecture en ligne (cf. *infra*) répond effectivement à d'autres objectifs : lire, certes, mais aussi étudier des occurrences, retrouver des passages spécifiques, etc.

B- Modalités de lecture des documents numériques

Certaines pratiques cumulent des modalités de lecture différentes des mêmes documents numériques : de la lecture en ligne en temps réel à la lecture différée de la version imprimée, la coexistence de l'électronique et de l'imprimé ne cesse d'entretenir une tension jusqu'à venir troubler les pratiques de citation des documents numériques.

B.1- Lecture en ligne

La lecture en ligne pratiquée par notre groupe est une lecture savante, c'est-à-dire une lecture d'étude : « *j'ai tendance à dire qu'on ne lit pas. On étudie un texte, mais on ne le lit pas* », Medica2.

« J'ai tendance à dire qu'on ne lit pas. On étudie un texte, mais on ne le lit pas. »

Cette lecture d'étude est généralement une lecture ciblée sur des passages précis et souvent déjà connus (Jubil2 ; Numdam1 : « *il y a cette lecture où on va prendre des petits bouts* »), tendue vers la recherche d'une information spécifique, (Jubil2 et Medica2), ne nécessitant pas de lecture linéaire et complète : « *Je peux passer dix minutes sur trois lignes et lire en diagonale trente pages en cinq secondes* », Medica2.

Ce peut aussi être **une lecture de repérage de structure** « *pour voir comment ça s'organise* » (Medica3), « *ce qui est fait et comment c'est fait* » (Numdam1).

Dans tous les cas, cette lecture est perçue comme rapide (Jubil1 : « *Ce que je veux, c'est pouvoir retrouver l'information le plus rapidement possible* », Medica3 : « *en général, je les lis assez rapidement. Je regarde le contenu* ») : elle est soit relecture (je me remets en mémoire un passage que j'ai déjà étudié, comme Jubil1³⁸) soit lecture à venir (d'un document que je vais alors imprimer ou télécharger). A ce titre, **les fonctions dites de lecture instrumentent de façon adéquate cette prise de reconnaissance et/ou de décision** : « *J'utilise toutes ces fonctions (les fonctions de lecture en ligne, comme le zoom, la possibilité d'avoir deux pages ou une page, la possibilité de défiler de haut en bas et de droite à gauche) qui sont bien* », Jubil4.

B.2- Lecture des documents numériques imprimés

Le groupe interrogé présente tous les types de posture : de ceux qui déclarent ne jamais imprimer (Cnum2, Cnum3) à ceux qui disent tout imprimer (Medica1 pour qui l'impression est une méthode de travail), en passant par ceux qui impriment parfois : **le passage à l'impression vaut alors pour validation d'un contenu, qui mérite une lecture approfondie** (Jubil1), différée (Jubil2) ou répétée : « *les articles vraiment intéressants, clés pour ma recherche, on peut y revenir dix fois* », Jubil3. L'impression peut aussi être motivée par la mobilité (Numdam2, Jubil2 : l'imprimé est plus commode dans les transports).

³⁸ . Jubil1 : « Internet peut aussi être utile pour me rappeler des livres que j'ai lus, dont je ne me rappelle pas en détail. (...) plutôt que de retrouver les notes que j'ai pu garder dans mon ordinateur ou ailleurs. Dans un certain sens, ça fait partie de mes dossiers. J'utilise la bibliothèque numérique comme si c'étaient des dossiers que j'avais quelque part ».

Certains notent **un recours à l'impression en baisse, liée aux évolutions techniques** (annotations possibles des PDF pour Medica2) comme à une nouvelle ergonomie du poste de travail (« *Je travaille beaucoup plus en ligne (...) parfois avec deux écrans : un écran sur lequel j'écris et un autre sur lequel je consulte les ouvrages* », Jubil1).

B.3- Lecture des documents numériques téléchargés

La plupart des usagers de notre groupe déclarent **décharger beaucoup de documents**, parfois plus qu'ils n'en lisent (Jubil1, Jubil2, Jubil4). L'objectif recherché est souvent la **constitution d'une bibliothèque numérique personnelle** : « *Il y a des auteurs avec lesquels je travaille quotidiennement. Ceux-là, je les ai téléchargés. J'ai aussi certains journaux entiers que j'ai téléchargés. J'ai une bibliothèque numérique un peu classée par ordre alphabétique (...) Je m'en sers à peu près comme de ma bibliothèque non numérique* », Numdam2 (motivation identique pour Medica3, Numdam1). Dans ce dernier cas, ce sont des usagers qui enrichissent leur bibliothèque numérique au-delà du seul téléchargement, en numérisant ou en photographiant eux-mêmes les documents imprimés qu'ils utilisent pour leur recherche (Jubil4, Numam1, Numdam3).

« Il y a des auteurs avec lesquels je travaille quotidiennement. Ceux-là, je les ai téléchargés. J'ai aussi certains journaux entiers que j'ai téléchargés. »

Certains recourent aux téléchargements dans des cas particuliers (mobilité pour Cnum3 et exploitation approfondie de PDF cherchables pour Medica2).

Ce sont ces mêmes documents téléchargés qui peuvent être imprimés ou travaillés par saisies successives : « *En ce moment, je fais un fichier doc de tous les passages qui m'intéressent. Ça m'arrive souvent de faire des petites saisies du PDF que je réincorpore dans un document Word. Je fais une espèce de résumé de tout ce qui m'intéresse dans l'ouvrage* » (Medica2).

B.4- Support de lecture mobile

Si l'impression et le téléchargement des documents numériques peuvent être privilégiés par les chercheurs en situation de mobilité, **les usagers interrogés ne plébiscitent pas, pour autant, un support de lecture mobile** de type Ipad : « *On peut être tenté de penser qu'un nouvel appareil, une nouvelle technologie va régler des choses, et généralement, ça régénère des contraintes* » (Cnum2).

Aucun d'entre eux n'en possède ; son intérêt n'est pas évident : « *Quand je travaille, j'ai toujours dix fichiers ouverts en même temps, des fichiers Word. Je fais plein de choses en même temps. J'ai besoin vraiment d'un ordinateur. Si j'avais besoin d'autre chose, j'achèterais un écran qui ferait 1 m de large et un ordinateur très puissant, qui aurait plus de mémoire. C'est plutôt ça dont j'aurais besoin que d'une tablette* », Medica2 (et aussi Numdam2 et Numdam4).

B.5- Citations des documents lus et utilisés

Massivement, **le groupe interrogé cite l'édition imprimée du document, même s'il est consulté sous forme numérique** : « *on fait comme si on allait en bibliothèque* », Medica2. Certains nuancent : si le document est difficile à trouver, y compris en ligne, on indique la source, mais plus volontiers en colloque à l'oral que dans un article (Medica3, Numdam4).

Très peu utilisent un logiciel de gestion bibliographique (Jubil3, Numdam1) et aucun Zotero ; la plupart font des copier-coller (Medica2, Jubil1, Medica3) ou récupèrent des notices BnF (Numdam2) ; reste que le lien vers les documents Gallica n'est pas toujours enregistré, même si certains font désormais des sauvegardes (Cnum3).

C- Pratiques numériques

Les quinze personnes interrogées ont été amenées à s'exprimer sur leurs pratiques du web. Deux types d'outils en particulier leur ont été soumis : d'une part, les réservoirs de contenus et outils de recherche courants (Google Books, Wikipedia, Google), d'autre part, les outils de communication et d'échange scientifiques.

C.1- Réservoirs de contenus et outils de recherche

- **Google Books**

Tous les usagers interrogés ont, au moins une fois, utilisé Google Books (y compris le doyen du groupe, Medica1 - 73 ans, mais sans le savoir). La plupart d'entre eux l'utilisent régulièrement (Cnum1, Cnum2, Cnum4, Numdam3 et Numda4), voire systématiquement, Jubil1 et Medica2 : « *Il faut bien dire que tout le monde utilise Google Books. Maintenant, ça devient la chose qu'on utilise le plus. C'est très récent qu'on l'utilise aussi massivement. Moins d'un an, peut-être un an* ».

« Il faut bien dire que tout le monde utilise Google Books. Maintenant, ça devient la chose qu'on utilise le plus. »

L'évaluation du service est contrastée. Les mauvais points concernent essentiellement la qualité de numérisation, jugée médiocre (Cnum1 et Cnum2) et le signalement des périodiques, « *catastrophique* » pour Jubil1, « *grave* » pour Medica2 : « *quand on a trouvé l'occurrence d'un mot dans une revue, la date qui est retenue est celle du début de la revue et non pas la date du volume où on a trouvé l'occurrence du terme. Ça pollue toute la recherche* » (Medica2). Jubil1 note également un « *problème de corpus* » dans Google Books qui impose au chercheur la plus grande vigilance. L'absence de politique documentaire et l'accès « *fragmenté* » aux documents sous droits en découragent certains (Jubil2 et Numdam2) qui tendent à délaisser Google Books, alors que d'autres, au contraire, ne sont pas gênés par cet aspect hétérogène des résultats qui leur permet d'identifier des « *nouveautés* » (Numdam4).

Les points forts isolent la qualité du moteur : pertinence de la recherche qui permet « *des découvertes inattendues* » et du classement des résultats ; la consultation est également jugée « *assez pratique* » (Jubil1).

- **Wikipedia**

Si tous les usagers interrogés utilisent Wikipedia, certains sont très critiques (pour Cnum1, « *c'est un outil extrêmement pervers* »), d'autres sont, au contraire, presque éblouis, comme Cnum2 : « *Maintenant, j'ai l'impression qu'en tapant n'importe quel nom, on arrive souvent à avoir une réponse. C'est complètement étonnant. Je suis toujours étonnée par ça, un peu comme l'arrivée du printemps. Ça paraît fou* ». Entre ces deux positions se trouvent des pratiques raisonnées de Wikipedia, comme celle exprimée par Numdam1 : « *J'aime bien regarder comment telle notion est introduite dans Wikipédia. Vu que ça a été fait par des collègues et que c'est à but pédagogique* ».

Un tiers de notre groupe contribue, alors même que « *ce type de contribution ne compte pas pour une publication scientifique* », comme le rappelle Jubil2 qui justifie ainsi son absence de participation. La plupart contribue de façon ponctuelle et pas nécessairement sur leur domaine scientifique : « *Je contribue sur des centres d'intérêt que j'ai, qui ne sont pas professionnels. Mais sur des choses professionnelles, il y a d'autres gens qui le font. Wikipédia est plutôt fait par des amateurs qui se passionnent pour certains sujets. En tant que professionnel, on ne contribue pas* », Medica2.

Une seule personne a contribué de façon soutenue dans son domaine de recherche et témoigne justement de la faiblesse de l'implication des scientifiques français : « *J'ai découvert avec stupeur que la communauté est très petite sur des sujets donnés. Vous avez une quinzaine de personnes qui contribuent aux articles de maths. (...) c'est vraiment une communauté minuscule. C'est tout à fait ahurissant* », Numdam2.

- **Google**

L'ensemble du groupe utilise massivement Google, de façon généralement rudimentaire, à deux exceptions près (Medica2 et Cnum4 connaissent des fonctionnalités avancées). Certains déplorent cette uniformité des pratiques (Medica3), regrettent de ne rien connaître de différent (Cnum1). D'autres, au contraire, défendent leur position : Google, c'est « *pratique* » (Cnum3), « *ça permet de ressortir ce que justement tout le monde sait* » (Jubil3) et surtout c'est très performant : « *y compris sur Gallica. C'est plus facile parfois (...) C'est ça qu'ils ont réussi à faire, qui est magique. On cherche quelque chose ou quelqu'un, et ils ont l'air d'identifier ce pourquoi on pourrait s'intéresser à ce truc* », Numdam2. Sur ce point, un usager note cependant une baisse de performance, notamment pour la recherche de documents numérisés : « *Google, c'est de moins en moins pertinent pour ce type de recherche. Avant, on arrivait à tomber sur des trucs un peu plus pertinents. Maintenant, ça ne marche plus* », Numdam1.

C.2- Outils de communication et de gestion des flux

- **Facebook**

Facebook ne se positionne pas du tout, pour notre groupe, comme un dispositif d'information et de communication scientifiques : seules quatre personnes interrogées sur quinze disposent d'un compte, utilisé uniquement à des fins personnelles. Globalement, le groupe interrogé ne se sent pas concerné (Numdam4 : « *je fais déjà assez de trucs, il m'arrive de recevoir 80 mails par jour* ») ; certains sont particulièrement hostiles (Cnum1, Numdam1 et Numdam2 par exemple : « *c'est un truc de gens qui ont envie de bavarder. Je n'ai pas spécialement envie de bavarder avec des inconnus* »).

- **Les flux RSS**

Seules les bibliothécaires du groupe interrogé connaissent et utilisent les flux RSS. De façon générale, **les chercheurs se déclarent peu intéressés par la notion de veille sur les nouveautés** (Jubil1, Jubil3 : « *S'il y a des choses qui viennent en alerte, ça arrivera quand on est en train de penser à autre chose...* »). Quand ils le sont, ils privilégient le mail (la liste de diffusion en particulier, Cnum1, Cnum3), les contacts informels

(Numdam4 : « *Je parle avec les gens* ») ou encore les rubriques dédiées aux actualités sur les sites web (Medica1, Numdam2).

- **Modalités de partage des informations et des publications scientifiques**

Le mail et les rencontres physiques restent les outils privilégiés de communication et de diffusion scientifique. Les publications se font uniquement dans le cadre d'édition imprimée soumises à comité de lecture, elles sont très peu déposées dans Hal (Medica2), plus souvent mises à disposition sur un site web personnel (Jubil1, Medica2), facilement transmises par mail à des collègues (Cnum1, Cnum2, Numdam1, Numdam4). Dans leur ensemble, les usagers interrogés ne se sentent pas concernés par des dispositifs en ligne de partage (qu'il s'agisse de signets, de documents ou encore d'annotations) : de ce point de vue, il sera intéressant d'observer les pratiques que développeront les chercheurs qui auront grandi avec YouTube, Facebook et Flickr.

Conclusions

De façon générale, le groupe retenu d'usagers intensifs des bibliothèques numériques ne manifeste qu'un intérêt très limité pour les technologies de l'information et de la communication : usage faible d'outils de gestion bibliographique, inexistant des flux RSS et des supports de lecture mobile, connaissance standard de Google, de Wikipedia et de Facebook, faible intérêt pour les plates-formes collaboratives...

Ce relatif « *conservatisme des pratiques et des conceptions* » du web cohabite avec le constat d'un « *impact important du numérique* »³⁹ sur le travail scientifique, le corpus numérique pouvant influencer des directions de recherche et renouveler des méthodes d'exploration scientifique des contenus. **Les racines d'un travail à l'ancienne se signalent encore en matière de lecture approfondie des textes, qui privilégie toujours le support imprimé.**

La suprématie de l'imprimé est encore marquée dans les choix des modalités de citation et des supports de publication, contribuant à créer ce paradoxe que si les chercheurs passent l'essentiel de leur temps d'étude en ligne, la majeure partie de leurs pratiques scientifiques se déploient, elles, hors ligne, que ce soit en amont pour les pratiques de veille (réseaux informels de proximité) ou en aval pour la diffusion scientifique (peu de dépôt en archive ouverte et de citation de documents numériques).

Ce profil influence les attentes exprimées par notre groupe en matière de bibliothèque numérique idéale : **la majeure partie d'entre eux est intéressée beaucoup plus par les contenus** (encore plus de sources, de tout type, rares ou difficiles à trouver de préférence) **que par les dispositifs techniques d'accès ou d'exploration des documents numériques.** Certains évoquent cependant la qualité d'océrisation (et le

³⁹ . Analyse proposée par l'Urfist de Nice dans le cadre d'une enquête sur les pratiques informationnelles des chercheurs (2010) : <http://urfistinfo.hypotheses.org/1901>.

cas particulier de celle des formules mathématiques, Numdam1 et Numdam2), la nécessité de disposer d'instruments de navigation performants au sein d'un document et d'un corpus (Jubil1 en particulier). Deux propositions innovantes sont toutefois à signaler : d'une part, la restitution de la matérialité de l'objet livre à travers des dispositifs de réalité augmentée (Cnum4) ; d'autre part, la possibilité d'afficher de façon concomitante l'iconographie citée dans un document : « *Imaginez le Laugier qui aurait un feuilletage comme ça et qui nous permettrait de voir le bâtiment dont il parle ! C'est la première fois où je me suis dit : il y a quelque chose dans cet outil qui va apporter beaucoup plus et qui rendra obsolète le livre* », Cnum3.

Pour conclure, il faut rappeler que les analyses livrées ici sont issues de verbatim recueillis auprès d'un groupe de chercheurs⁴⁰, qui ont, pour la majeure partie d'entre eux, engagé leur carrière scientifique bien avant l'existence des bibliothèques numériques : leur habitus reste celui propre au travail sur l'imprimé et l'on ne peut que constater l'extrême difficulté, éprouvée par tous, à développer de nouvelles pratiques conformes ou adaptées à de nouveaux supports : les pratiques semblent être conformes d'abord aux praticiens eux-mêmes (!), à leurs habitudes, leurs préférences, leurs envies et leurs critères d'efficience. A cet égard, les jugements de satisfaction (ou d'insatisfaction) méritent toujours d'être contextualisés et ne peuvent suffire à l'analyse : ainsi de notre usager Medical, médecin à la retraite, qui se déclare très satisfait de Medic@ alors qu'il vient de découvrir le format PDF dédié à l'impression, qui aurait pu alléger la peine qu'il s'est donné d'imprimer, trois ans durant, page-écran après page-écran l'ensemble des documents qu'il a utilisés pour sa thèse...

5.2-Aperçu de leurs comportements en matière de recherche d'information et de pratiques culturelles

Le guide d'entretien semi-directif consacrait un ensemble de questions aux usages des nouvelles technologies dans les pratiques professionnelles et personnelles des chercheurs d'une part, à leurs pratiques culturelles – fréquentation des bibliothèques en particulier – d'autre part.

Classés en huit items, les verbatim recueillis permettent d'identifier en particulier :

- le niveau de technophilie de ces usagers et leur fréquence de connexion à Internet ;
- leur fréquentation et leur usage des bibliothèques physiques ;
- leurs pratiques culturelles personnelles.

A- Des usagers « 1.0 »

S'ils utilisent de manière approfondie les collections numérisées, les usagers interrogés ne sont pas pour autant des usagers « 2.0 » et montrent peu de curiosité pour les outils ou les sites d'échange et de partage avec d'autres internautes. Quelques uns

⁴⁰. L'âge moyen du groupe, composé de quinze personnes âgées de 28 ans à 73 ans, est de 47 ans, âge moyen des répondants à l'enquête menée par ailleurs en 2011 sur Gallica par GMV Conseil.

continuent même d'affirmer leur nette préférence pour le papier, ceci allant jusqu'à la revendication quasi militante de rester « *très livre* », Cnum1.

La plupart d'entre eux possèdent un ordinateur portable et utilisent les logiciels de bureautique courants (Word, Excel). A l'inverse, ils ne connaissent pas ou n'utilisent pas d'outils 2.0 dans leur travail : fils RSS, logiciels de synchronisation des dossiers etc. Un seul usager déclare avoir commencé à utiliser « *des logiciels de synchronisation qui sont très bien*.⁴¹ »

Certains utilisent des appareils photo numériques, notamment pour numériser des ouvrages (Jubil4), mais très peu possèdent des smartphones. Aucun d'entre eux n'a encore acheté de tablette. Les usagers intéressés soulignent la trop petite taille de l'écran et le choix encore limité du catalogue proposé, mais estiment que les tablettes seront probablement un de leurs futurs outils de travail⁴².

L'utilisation d'Internet est massivement professionnelle. Si l'ensemble du groupe interrogé a une connexion Internet au travail, plusieurs usagers déclarent ne pas avoir de connexion Internet chez eux. Un seul déclare être connecté « *plus ou moins en permanence* » (Numdam4), et un seul crée des pages web pour le site de l'université.

B- Fréquentation et usages des bibliothèques

Les usagers interrogés se déclarent pour la plupart utilisateurs réguliers des bibliothèques, même si leur fréquentation marque le pas. Ils se rendent le plus souvent dans les bibliothèques rattachées à leur lieu de travail : BIUM, bibliothèque du CNAM, bibliothèques de recherche de Jussieu, bibliothèques universitaires de Grenoble. **Une majorité, en particulier les usagers parisiens, fréquente aussi la BnF.**

L'usage quotidien de la bibliothèque comme lieu de travail semble cependant en perte de vitesse. Il n'est souvent pas évoqué dans les entretiens conduits. Seuls deux usagers, tous deux mathématiciens, évoquent cette pratique, l'un en parlant au passé⁴³, l'autre parlant de la fréquentation quotidienne des bibliothèques comme fait de résistance⁴⁴.

L'usage majeur des bibliothèques physiques est un usage spécialisé, qui vient en complément de celui des bibliothèques numériques. Il s'agit alors de consulter des ouvrages imprimés pas encore disponibles en ligne, car « *actuellement, les services en ligne ne sont absolument pas suffisants pour qu'on supprime les bibliothèques papier*. », Numdam2. Cet usage est très souvent associé aux bibliothèques spécialisées

⁴¹. Medica2 : « J'utilise maintenant des logiciels de synchronisation qui sont très bien. C'est un peu compliqué. Je me suis fais récemment un tableau de toutes les synchronisations entre mes ordinateurs, sachant que je synchronise les titres de dossier. C'est assez compliqué. Mais c'est assez formidable. Ça me permet de travailler comme si je n'avais qu'un ordinateur ».

⁴². Numdam1 : « D'ici 5 ans, on aura un truc léger, style Kindle ou Sony reader, avec un écran assez gros, suffisamment rapide, et on n'aura plus besoin d'imprimer. Ce sera nickel ».

⁴³. Numdam1 : « A Grenoble, ça représentait 80 %, je travaillais beaucoup à la bibliothèque. Je m'installais à la bibliothèque. La bibliothèque de Grenoble est géniale. Ailleurs, c'est moins enthousiasmant ».

⁴⁴. Numdam2 : « [Parlant de la fréquentation des bibliothèques] Je résiste, parce qu'il y a une question forte derrière. En maths, on se bagarre énormément pour que les bibliothèques restent ».

qui ne sont pas à proximité immédiate du lieu de travail, qui sont fréquentées quand on sait « *ce que l'on va faire* », Jubil4.

La BnF occupe une place particulière. C'est la bibliothèque où l'on « trouve tout », Cnum3, où « *il y a des ouvrages [que l'on] ne trouve pas ailleurs*.⁴⁵ » C'est une bibliothèque fréquentée par la grande majorité des chercheurs interrogés, qui connaissent parfaitement le lieu. Ils fréquentent en majorité le Rez-de-jardin – salles de sciences en particulier –, et soulignent les difficultés d'accès aux collections anciennes⁴⁶.

Les usagers interrogés ont bien souvent des connaissances approfondies sur les instruments de recherche à leur disposition en bibliothèque : catalogues, bases de données etc. Si les catalogues de bibliothèques sont utilisés par l'ensemble des chercheurs interrogés, le recours aux bases de données est bien plus spécifique et varie suivant le champ de recherche du chercheur interrogé : bases en histoire des sciences et des techniques, bases en géologie, en chimie, en mathématiques etc.⁴⁷ Ces bases de données servent la plupart du temps à établir une bibliographie pour « *une recherche bien précise* », Jubil3. La plupart des chercheurs interrogés n'ont pas manifesté d'intérêt pour être formés à de nouveaux instruments de recherche. La majorité des répondants estiment qu'ils maîtrisent suffisamment les outils de recherche dont ils ont besoin⁴⁸. Seule concession, l'envoi d'un « *bulletin d'information* » par mail, qui pourrait être utile pour signaler la mise en ligne de nouvelles ressources (Medica2).

C- Pratiques culturelles

Si les lectures professionnelles passent essentiellement par l'électronique pour les chercheurs interrogés, ce n'est pas le cas pour les lectures personnelles. La lecture-plaisir privilégie le support imprimé : « Je préfère aller acheter un bouquin en librairie et le lire », Numdam1. La plupart se déclarent lecteurs réguliers de la presse quotidienne, abonnés à une ou plusieurs revues et disent acheter en moyenne « *deux, trois livres par mois*.⁴⁹ » Une très faible minorité (deux seulement) déclare emprunter en bibliothèque.

⁴⁵. Jubil1 : « En général, je vais à la BnF parce qu'il y a des ouvrages que je ne trouve pas ailleurs. C'est une bibliothèque que j'utilise parce que je la trouve pratique, proche d'ici, agréable, efficace. On a les choses assez rapidement. On a un choix incomparable de documents ».

⁴⁶. Jubil3 : « Les fonds anciens en sciences, il y en a 1/5e hors d'usage, d'après mes statistiques ».

⁴⁷. Jubil3 : « Pour la Jubilothèque, les deux ressources essentielles sont l'accès à des revues en ligne et l'accès à des catalogues type Georef, Inspec. Sachant qu'à la Bibliothèque nationale, vous êtes bien mieux fourni qu'eux [Bibliothèque Pierre et Marie Curie]. En particulier, il n'ont pas renouvelé l'abonnement à Inspec, qui est très handicapant pour moi. Donc, je viens souvent faire cette recherche ici [à la BnF] ».

⁴⁸. Jubil1 : « Je ne pense pas que j'ai besoin de formation. Mais peut-être que je me trompe. Je suis assez satisfait. Il y a peut-être des choses qui m'échappent encore. Parfois, c'est difficile de trouver certaines sources, ça peut arriver. Peut-être qu'il y aurait moyen d'avoir une meilleure connaissance de tout ça. Mais je ne sais pas comment on pourrait être formé à ça. L'important, c'est de diversifier les accès aux sources. Ça s'apprend par la pratique ».

⁴⁹. Jubil3 : « Des quotidiens, oui. Des quotidiens d'information générale. J'achète des livres, je lis quelques revues, un journal par jour. [...] Peut-être 2-3 livres par mois, revues comprises. Pour les revues par exemple, je suis abonnée à 2 ou 3 annuaires qui paraissent une fois par an sous la forme d'un livre ».

6. Conclusions intermédiaires : d'une bibliothèque, sur place, à l'autre, en ligne : points communs et différences de pratiques ; retour sur l'hypothèse initiale : spécificités de l'ancrage disciplinaire sur les pratiques de recherche en ligne.

Ce volet de l'étude consacré aux pratiques de recherche en ligne et en bibliothèque ainsi qu'aux pratiques culturelles de notre groupe d'utilisateurs permet d'apporter une contribution à la question – classique – du rapport – de rupture ou de continuité – entre bibliothèques numériques et bibliothèques physiques.

L'analyse des verbatim sur cette question indique qu'il y a à la fois continuité et rupture :

- S'il est certain que la fréquentation des bibliothèques physiques a été remplacée – dans une proportion que l'étude ne peut mesurer – par la consultation de la bibliothèque numérique, il ressort clairement des verbatim que l'on consulte d'autant plus et mieux les collections en ligne que l'on a déjà bien et beaucoup travaillé sur les collections imprimées : ce constat vaut pour notre groupe d'utilisateurs, dont la moyenne d'âge de 47 ans reste, encore en 2011, représentative de l'âge du Gallicanaute-type. Que feront les générations qui auront directement rencontré les corpus en ligne (sans du tout ou peu connaître les originaux imprimés) ?⁵⁰

Cette continuité des pratiques se lit aussi dans leur posture, paradoxale, vis-à-vis des modes image, peu critiqué, et texte, beaucoup plus polémique. On a avancé l'hypothèse que le mode image ne pouvait être franchement critiqué au risque de critiquer aussi la valeur elle-même de l'intégrité du document patrimonial ; le mode texte, lui, pouvait être plus soumis à question et à exigence puisqu'il engage la bibliothèque à devenir simple base de données, avec les avantages et les inconvénients que l'on sait, mais sans (encore) mise au point véritable de nouvelles méthodologies de création de corpus⁵¹ : là encore, le suivi des générations ayant découvert directement le patrimoine en ligne devrait s'avérer riche d'enseignements.

- La rupture peut s'identifier à travers les critiques portant sur les modalités de visualisation du livre numérisé : les modalités actuellement proposées s'attachent à copier, le moins mal possible, les manipulations autorisées par le livre imprimé (feuilleter, regarder de près un détail, utiliser un sommaire ou une table des matières). Reste que ces simulations ne convainquent pas vraiment : la manipulation du livre électronique homothétique reste lente, inadaptée à l'écran et à la souris et c'est toujours la version imprimée qui assure le meilleur confort de lecture (en tout cas pour notre groupe d'utilisateurs).

⁵⁰ . On peut s'en faire une idée grâce à l'étude commanditée par la British Library et le JISC (Joint Information Systems Committee) : *Information Behaviour of the Researcher of the Future* ('Google Generation' project), 2008 : <http://www.ucl.ac.uk/infostudies/research/ciber/downloads/ggexecutive.pdf>.

⁵¹ . Rappelons que les verbatim sur ce point se situent essentiellement dans le registre émotionnel de l'étonnement (devant la rapidité, la facilité), voire de l'émerveillement (d'avoir tant de textes « Cette interrogation qu'on peut faire en mode texte permet de lier tous les livres de la planète ensemble (...) C'est ça qui est extraordinaire. », Cnum3).

De façon là aussi paradoxale ou en tout cas inattendue, les entretiens laissent poindre une véritable attente pour un livre numérique de nouvelle facture, ou en tout cas de facture suffisamment adaptée à l'écran pour que l'orientation dans les textes numériques soit au moins aussi efficace que celle permise par l'imprimé.

Ces éléments d'analyse restent intimement liés aux spécificités de notre groupe d'utilisateurs dont il apparaît, en fin d'étude, qu'elles tiennent moins à l'ancrage disciplinaire scientifique qu'à la pratique d'historien que ces scientifiques développent.

À une exception près (Numdam1), tous nos utilisateurs, quelle que soit leur formation initiale scientifique, se livrent, avec Gallica, à des travaux de nature historique. Ces travaux les amènent à interroger Gallica non pas comme une base de données mais comme un réservoir – ou mieux une étagère (voire plusieurs !) – grâce à laquelle ils vont pouvoir constituer un corpus le plus cohérent et exhaustif possible. De ce point de vue, les usages analysés dans cette étude restent exemplaires du travail sur le document patrimonial plus que du travail sur des corpus scientifiques : est-ce à dire qu'il n'y aurait pas, véritablement, de spécificités au travail sur le patrimoine scientifique ?

7. Conclusions

Cette étude s'est donné pour objectif d'analyser la place des bibliothèques spécialisées dans le dispositif Gallica en donnant la parole à ses usagers, pas à tous les usagers : un groupe de quinze utilisateurs intensifs de bibliothèques numériques scientifiques.

Chaque participant du projet a recueilli dans les entretiens menés des éléments susceptibles de l'intéresser :

- les responsables de bibliothèques numériques de niche ont analysé les pratiques majeures d'un petit échantillon de leur public et estimé sa satisfaction,
- la BnF s'est, elle aussi, intéressée aux usages et évaluations de Gallica mais également aux pratiques numériques, documentaires et culturelles du groupe interrogé, avec une attention particulièrement portée à l'incidence de la spécialité disciplinaire sur les usages.

Il reviendra aux responsables de bibliothèque numérique d'évaluer l'apport de cette étude à leurs propres interrogations.

Du côté de la BnF, cette étude a apporté des réponses aux deux questions initialement posées et a ouvert des perspectives d'approfondissement.

Rappel de la question 1 : comment s'articule pour les publics le passage d'usages de bibliothèques spécialisées aux usages spécialisés d'une bibliothèque encyclopédique comme Gallica ?

Les entretiens ont mis en valeur ce trait constant de l'adaptation des publics à un dispositif d'organisation des connaissances⁵². Autrement dit, chaque type de bibliothèque numérique est sollicité sur la base de son offre explicite de collections : Gallica est utilisé pour interroger les collections numérisées par la BnF et les bibliothèques de niche pour consulter leurs collections spécifiques. À une exception près, Gallica est peu utilisé comme agrégateur de contenus de bibliothèques numériques spécifiques : il faut préciser que, dans ce cas, le moteur de recherche Gallica effectue une requête de nature différente selon qu'il interroge les collections moissonnées des partenaires (notices bibliographiques uniquement) ou les collections numérisées par la BnF (tous les champs textuels disponibles). Cette interrogation mixte du plein-texte des collections BnF et des champs notices des autres collections moissonnées permet difficilement de constituer des corpus scientifiquement viables.

⁵² . Par exemple, dans les *Espaces du livre : perception et usages de la classification et du classement en bibliothèque*, Eliséo Véron note que l'essentiel, pour les usagers des collections en libre accès, est de disposer d'un système de classement... quel qu'il soit : « le rapport des usagers à la classification implique tout simplement la nécessité d'un système de repérage par rapport auquel une stratégie puisse se constituer. La classification garantit l'existence d'une convention régulière et l'absence d'arbitraire. Autrement dit : une classification est indispensable, ne serait-ce que comme élément contre lequel organiser une stratégie. Elle ne nécessite cependant pas de perfectionnements particuliers, et l'on peut soupçonner que n'importe quelle classification, pourvu qu'elle soit stable et régulière, ferait l'affaire », *Espaces du livre : perception et usages de la classification et du classement en bibliothèque*. Paris : Bibliothèque publique d'information, Centre Georges Pompidou, 1990.

En revanche, ce mode de recherche fédérée est très utile pour les phases exploratoires de repérage des segments documentaires disponibles en ligne.

Les usagers s'adaptent donc plutôt aux dispositifs disponibles et développent des stratégies et des attentes en fonction de l'offre documentaire et de l'interface mises à disposition (d'où les exigences d'une meilleure recherche plein texte dans Gallica alors que le mode image est à peine critiqué par exemple). S'il y a des usages spécialisés des collections Gallica, ils se développent moins sur le site Gallica que dans des espaces documentaires spécifiques où les contenus de Gallica ont été retravaillés et mis à disposition dans un contexte dédié (cas typique de Gallica-Math). Nous n'avons pas pu, à proprement parler, identifier des usages spécialisés de Gallica à travers l'enquête.

Du point de vue du département de la Coopération, il importe de savoir que le dispositif Gallica, bibliothèque numérique de référence certes mais avec sa galaxie de bibliothèques numériques autonomes, a un véritable sens du point de vue des usages.

Rappel de la question 2 : Quelles sont les spécificités des recherches en histoire des sciences, sont-elles proches de celles des historiens ou proches de celles des scientifiques ?

En première analyse, il semble que les recherches menées par notre groupe s'apparentent très nettement aux recherches des historiens ou, en tout cas, l'ancrage scientifique n'a pu être rendu visible à travers notre enquête. Le travail en mathématiques doit être distingué, quoique l'enquête isolait déjà elle-même un groupe homogène de mathématiciens (là où les autres groupes d'utilisateurs étaient d'emblée multidisciplinaires).

Rien d'étonnant cependant dans ce résultat : on sait les efforts des historiens des sciences pour introduire des mises en perspective historique dans les cursus scientifiques... complètement dépourvus de ce type d'approche. Quand un scientifique développe un intérêt pour ce type d'approche, c'est déjà que sa carrière est bien engagée et qu'il a le loisir de développer deux pistes de travaux bien distinctes, ou alors qu'il devient historien des sciences, abandonnant le travail scientifique dans sa discipline d'origine... bref, le portrait du scientifique en quête de patrimoine n'est sans doute qu'une figure marginale ou rêvée.

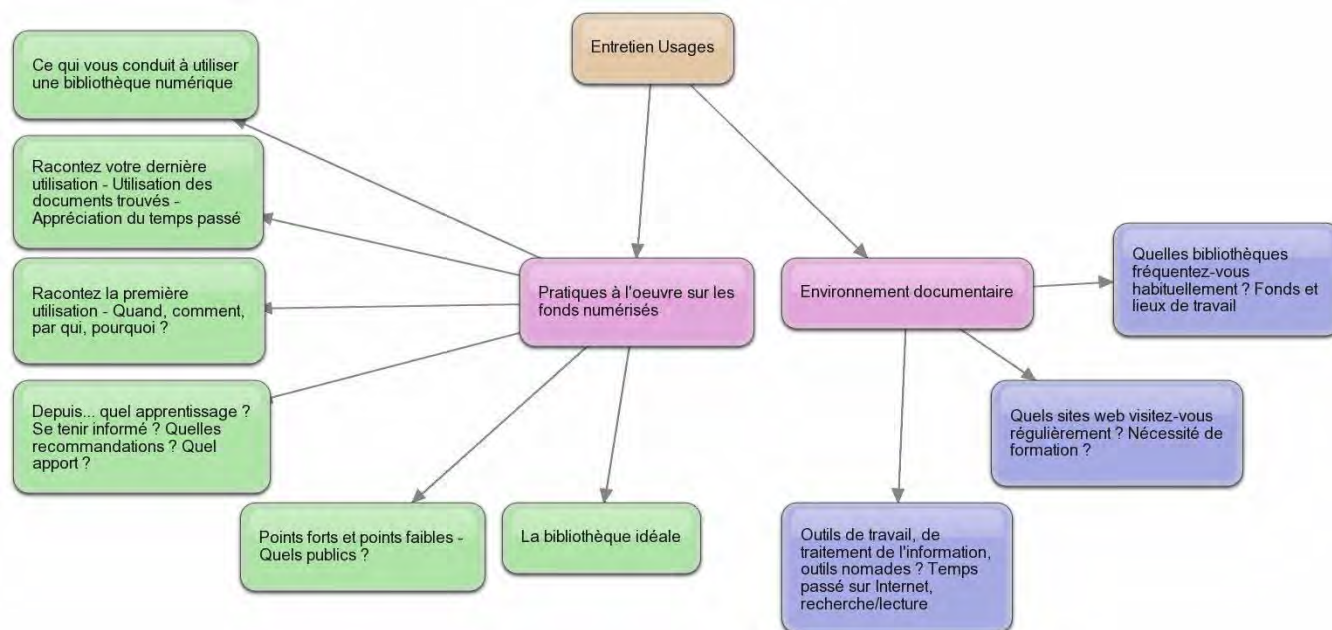
Cécile TOUITOU

Muriel AMAR

Délégation à la stratégie et à la recherche

Service Pôles associés - Gallica

Annexe 1- Grille d'entretien (version courte sous forme de carte)



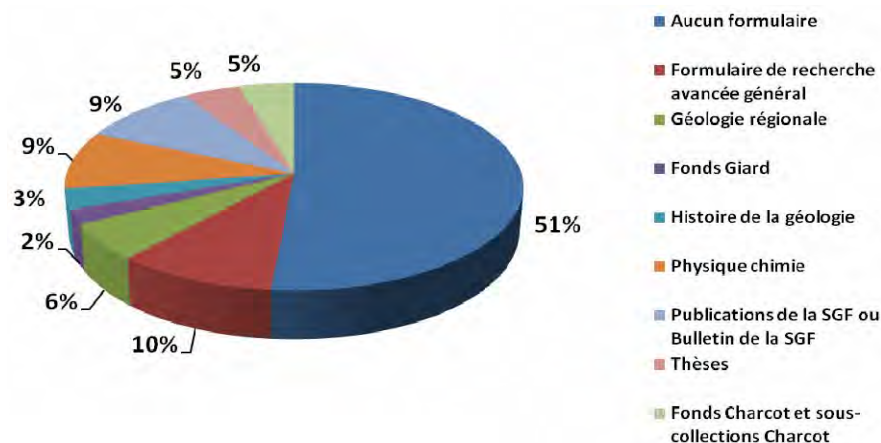
Annexe 2- Profil des utilisateurs interrogés

Profession	Usages	Ancrage géographique	Age	Divers	Code
Professeur des universités - Histoire des sciences mathématiques	Cnum, Numdam, Gallica, Google	Paris - études aux USA	41 ans	Formation en maths et en physique	Jubil1
Professeur des universités	NumDam, Linum	Strasbourg	50-55 ans	Oulipo - Travail en math et très récemment en histoire des maths	Numdam4
Maître de conférences en histoire des neurosciences, chargé de recherche CNRS, HDR en cours	Medic@	Paris	43 ans	Formation scientifique (thèse pharmacologie) puis réorientation en histoire des sciences (deuxième thèse)	Numdam2
Professeur d'EPS dans un institut médico-éducatif, docteur en STAPS	Fonds Charcot / Gallica	Le Mans	52 ans	Autodidacte	Jubil4
Maître de conférences en mathématiques	numdam, linum	Paris et Lille	28 ans		Numdam1
Paléontologue – Directeur de recherche CNRS	Jubilothèque	Paris	61 ans	Docteur ès sciences naturelles	Jubil2
Chargée de collection (Musée) - Ingénieur de recherche	Cnum	Paris	40 ans	Thèse histoire des techniques (verre)	Cnum2

Profession	Usages	Ancrage géographique	Age	Divers	Code
Chargée de collection (Musée)	Cnum	Paris	39 ans	Thèse en histoire	Cnum1
Chargée de recherche CNRS - IRHT	Medic@, profil Philologue	Paris	33 ans	Doctorat section des sciences historiques et philologiques (EPHE)	Medica3
Pétrologue – Maître de conférences en géologie	Jubilothèque	Paris et Italie	55 ans		Jubil3
Directrice de recherche CNRS	NumDam, Linum	Paris	52 ans	Travaille et publie en maths et en histoire des maths	Numdam2
Ingénieur de recherche, responsable d'une bibliothèque de recherche	numdam, linum, gallica	Paris	50-55 ans	Formation en maths- Auparavant à l'Inria	Numdam3
Médecin à la retraite - Doctorat en sciences historiques et philologiques	Medic@	Paris	73 ans	Médecine générale, médecine interne, expertise médico-légale	Medica1
Architecte DPLG ; professeur associée au Cnam	Cnum	Paris	51 ans	Doctorat urbanisme et aménagement	Cnum3
Bibliothécaire (centre de documentation d'un Musée)	Cnum	Paris	40-43 ans	5 ans au Centre de doc. du Musée ; formation généraliste	Cnum4

Annexe 3 - Utilisation des formulaires de recherche de la Jubilothèque

Utilisation des formulaires de recherche de la Jubilothèque – 5 600 requêtes



Annexe 4- Statistiques et données chiffrées (Gallica) Etat des collections dans Gallica (15/09/2011)



Etat des partenariats dans Gallica (15/09/2011)

a) Interopérabilité via le protocole OAI-PMH



b) Intégration des collections des partenaires dans les circuits de numérisation BnF



Fréquentation (données 2010)

- 30 000 visiteurs par jour en moyenne.
- 20 pages vues en moyenne par visite
- 15 minutes en moyenne la durée de visite

A noter : audience soutenue le week-end et en dehors des heures de bureau.