

# Les usages de Gallica

**L**es bibliothèques électroniques sont partie constituante de la palette des sources d'informations courantes des usagers du web. Par la masse d'informations qu'elles proposent de manière délocalisée, distante, elles participent des contenus riches de l'Internet au même titre que les bases d'informations spécialisées, les quotidiens ou les portails culturels ; la question est de savoir, au-delà de cette intuition, comment les utilisateurs appréhendent ces bibliothèques, ce qu'ils en attendent, comment ils les pratiquent et, au final, comment elles s'intègrent dans leurs pratiques quotidiennes du web. Par ailleurs, le fait qu'elles soient accessibles au travers d'offres d'accès où la bande passante permet une navigation de plus en plus confortable n'est pas sans conséquence sur le mode d'utilisation.

## Catherine Lupovici

Département de la bibliothèque numérique, BnF  
catherine.lupovici@bnf.fr

## Thierry Cloarec

Département de la bibliothèque numérique, BnF  
thierry.cloarec@bnf.fr

## France de Charentenay

BnF/France Télécom R&D  
france.de-charentenay@bnf.fr

S'il ne faut pas mésestimer les questions relatives à l'interface de consultation, il est important de mesurer la fréquence des connexions, leur durée, les manipulations opérées en rapport avec le contexte de la recherche qui a amené l'utilisateur à s'intéresser à la bibliothèque numérique.

Au fond, on peut se demander si la bibliothèque numérique peut être vue comme une voie particulière d'information (représentation singulière de la bibliothèque), ou si elle est un des moments du travail des utilisateurs, au milieu d'activités variées, depuis leur table de travail personnelle ?

Après quatre années de fonctionnement de sa bibliothèque numérique sur Internet, Gallica<sup>1</sup>, la Bibliothèque nationale de France a souhaité compléter les informations sur les usagers et la fréquentation qu'elle recueille régulièrement à partir des statistiques des systèmes informatiques et de questionnaires en ligne. Ce travail a pu être réalisé dans le cadre du projet de R&D BibUsages du programme 2001 RNRT (Réseau national

de recherche en télécommunications) consacré aux études d'usages de l'Internet et conduit en partenariat avec France Télécom Recherche et Développement. Il a permis une analyse de données quantitatives et qualitatives provenant d'un recueil de l'activité effectué sur le poste « client » de l'utilisateur. Cette approche situe la consultation de Gallica dans le contexte général de l'utilisation des ressources d'Internet et plus seulement des seuls services en ligne de la bibliothèque.

## Fréquentation et usages de Gallica

Dès son ouverture à titre expérimental en octobre 1997, la fréquentation de Gallica a été surveillée afin de connaître le public et ses attentes. Ce suivi a permis de faire évoluer les contenus et l'interface utilisateur qui a été modifiée pour la mise en ligne de la totalité des collections numérisées libres de droit en janvier 2000.

Les informations recueillies comptabilisent les requêtes reçues par le serveur pour la recherche d'informa-

1. <http://gallica.bnf.fr>

Directrice de la Bibliothèque de l'Académie nationale de médecine, chef du bureau Politique documentaire à la DBMIST, puis adjointe au sous-directeur de la Documentation des services et des réseaux dans cette même direction, et responsable des activités bibliothèques chez Jouve SA, **Catherine Lupovici** est, depuis 1999, directrice du département de la Bibliothèque numérique, à la Direction des services et des réseaux de la BnF. Elle a publié des articles sur les documents électroniques et les bibliothèques numériques dans le BBF et dans Documentaliste-Sciences de l'information, et « Une bibliothèque sans frontières » dans La Bibliothèque nationale de France : collections, services, publics (Cercle de la librairie, 2001).

Adjoint à la directrice du département de la Bibliothèque numérique à la BnF,

**Thierry Cloarec**, titulaire d'une maîtrise de lettres, a été, auparavant, responsable du Pôle recherche à la Mission scientifique.

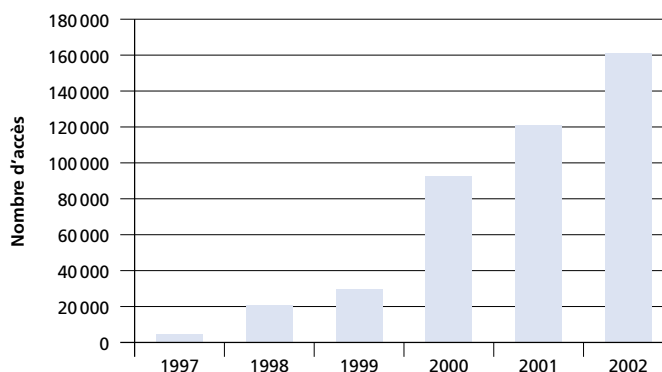
Il a coécrit « La recherche à la Bibliothèque nationale de France » (BBF, 2001, n° 4) et « Projet d'enrichissement pour le catalogue de la Bibliothèque de France » (BBF, 1993, n° 3) et a participé à la version française du Manuel Unimarc (K.G. Saur, 1991).

**France de Charentenay**, titulaire d'une maîtrise de sciences et techniques de médiation culturelle, option Musée et patrimoine, travaille au département des Cartes et Plans à la BnF, où elle collabore au Répertoire des cartographes français. Elle a été chargée d'études pour le projet BibUsages en collaboration avec France Télécom R&D et a occupé différents postes dans le secteur privé.

tion, la consultation des documents et la demande de téléchargement pour les imprimés numérisés. Par ailleurs, des questionnaires ont été mis en ligne pendant les deux premières années afin de connaître le profil du public et ses préférences en termes de contenus et de modalités de recherche et de consultation.

Dès le départ, Gallica se caractérise par un public composé aux trois quarts d'enseignants, de chercheurs et d'érudits, et localisé pour 60 % en France et 40 % à l'étranger. Plus de 85 % des consultations sont effectuées à partir de postes situés en dehors de la bibliothèque et environ 75 % des usagers entrent dans Gallica à partir d'un site non BnF (moteur de recherche, fournisseur d'accès, site offrant un lien vers Gallica). L'utilisation de Gallica connaît une augmentation constante, fonction de l'augmentation du nombre de documents

Évolution de la fréquentation de Gallica depuis l'ouverture en octobre 1997



mis en ligne (voir diagramme). La collection atteint aujourd'hui 70 000 volumes d'imprimés en mode images, soit 21 millions de pages, 1 500 volumes d'imprimés en mode texte, 80 000 images et 500 documents sonores.

Les contenus de Gallica correspondent initialement à une bibliothèque encyclopédique de culture francophone destinée aux travaux d'enseignement et d'érudition, mais aussi devant permettre la découverte de ressources culturelles par le grand public. Le suivi réalisé à partir des disciplines consultées et des enquêtes en ligne a montré une forte utilisation dans des disciplines classiques des usagers de la BnF, telles que littérature et histoire, mais aussi en histoire des sciences, bien représentée dans les collections numérisées. Un usage important et une demande forte ont aussi été constatés pour les ouvrages de référence incitant à élargir la collection vers les encyclopédies et dictionnaires, les bibliographies et la presse. Gallica s'est également enrichi de dossiers plus pointus autour de thèmes divers dictés par l'actualité culturelle, ou à l'occasion de programmes de coopération avec d'autres institutions. Un axe documentaire important s'est ainsi constitué sur l'histoire régionale française à partir du dossier *Voyages en France* et de la numérisation des publications des so-

ciétés savantes des régions Aquitaine et Lorraine.

La recherche d'information dans Gallica s'est simplifiée au cours du temps pour se rapprocher des modalités de recherche offertes sur Internet. Elle s'appuie, depuis 2001, sur des métadonnées composées des éléments des notices catalographiques, complétées par le texte intégral des tables des matières et des légendes des images, et qui sont interrogeables dans Gallica par auteur, titre, sujet et texte intégral. Toutes ces métadonnées et tous les textes de présentation sont également exposés pour l'indexation par les moteurs de recherche sur Internet, qui constituent un des modes d'accès important vers Gallica. Par ailleurs, les notices du catalogue en ligne signalent les documents numérisés et offrent un lien vers eux. Un utilisateur peut donc être conduit dans Gallica par un lien à partir du catalogue en ligne de la Bibliothèque nationale de France, par un lien lors d'une recherche sur Internet, ou parce qu'il est un habitué de Gallica qu'il vient consulter régulièrement.

C'est dans ce contexte de contenu documentaire et d'usages établis que le projet de recherche BibUsages a permis de mieux comprendre la démarche et les attentes des internautes qui utilisent Gallica et a contribué à l'un des objectifs du projet d'établis-

sement 2001-2003<sup>2</sup> d'actions destinées à permettre le *développement des services à distance* de la bibliothèque. Il s'est déroulé sur l'année 2002.

## Le projet de recherche BibUsages

Le projet BibUsages avait pour thème général l'étude des usages d'une bibliothèque électronique en ligne. Le projet a été conduit en partenariat et sous la direction de France Télécom Recherche et Développement et s'est articulé autour des objectifs suivants :

- Décrire les usages des bibliothèques électroniques en ligne en les croisant avec les caractéristiques de la population d'utilisateurs.
- Mettre en évidence la manière dont des usages émergents infléchissent et modifient des pratiques bien établies.
- Dégager une vue prospective sur les services liés aux contenus numériques sur des réseaux haut débit.

## Méthodologie

Le projet BibUsages a été découpé en trois phases :

### Phase 1 : enquête de cadrage

Questionnaire en ligne sur le site Gallica (mars 2002), qui a recueilli 2340 réponses. Il a permis de définir le profil socio-démographique des utilisateurs et de trouver des candidats volontaires pour participer au panel.

### Phase 2 : constitution du panel, recueil de trafic

Constitution du panel à partir des réponses au questionnaire : 97 volontaires, 72 retenus (mai 2002).

Recueil de trafic web (sonde Audinet, FTR&D) (de juin à décembre 2002) : 72 utilisateurs ont été observés pendant 6 mois. Les données étaient transmises à France Télécom au fur et à mesure.

2. <http://www.bnf.fr/web-bnf/connaitre/projet2001-2003.htm#2.2>

### Phase 3 : enquête qualitative sur un sous-ensemble du panel

Seize entretiens ont été réalisés (octobre 2002).

Le croisement des données recueillies a été pratiqué à plusieurs niveaux et en plusieurs phases : données quantitatives d'abord, puis données qualitatives dans un deuxième temps, elles-mêmes dédoublées en une analyse systématique du trafic, puis un entretien destiné à vérifier les hypothèses proposées par les comptes rendus d'activité.

L'analyse systématique des données recueillies par le logiciel-sonde installé sur le poste des personnes volontaires a permis de constater plus précisément les pratiques réelles et de disposer d'éléments à confronter au résultat des entretiens.

La série d'entretiens s'est déroulée sur un mois : parmi les 72 membres actifs du panel (ceux pour qui le logiciel-sonde enregistrait réellement des données), un groupe de 16 personnes a pu être délimité, où ont été retenus les plus actifs (en fonction de leur trafic Internet) et les habitants de Paris ou sa région, afin de pouvoir gérer les contacts plus aisément. De même, pour des raisons d'organisation et de gain de temps, tous les entretiens ont été menés par téléphone selon un même schéma.

Le questionnaire de mars 2002 avait permis de mieux cerner en termes de typologie socio-démographique le public de Gallica. Les entretiens avaient pour objectif de confirmer les pratiques enregistrées par la sonde et de les replacer dans leur contexte : montrer comment la pratique de Gallica s'inscrit dans une pratique générale d'Internet dans un contexte donné.

Ils se sont appuyés sur une grille d'entretien reprenant en filigrane ces objectifs :

- Une première partie sur les pratiques générales d'Internet permet de mieux cibler le profil général de l'internaute : durée, motivations, contexte de l'utilisation, modalités des recherches et traitement de l'informa-

tion. En outre, cela permet d'avoir des renseignements sur les usages hors web (*chat, e-mail, forums, p2p*...) que la sonde logicielle dans la version utilisée pour cette enquête ne détectait pas.

- La deuxième partie restreint le champ d'investigation sur la connaissance et la pratique du site de la BnF et de Gallica. Il s'agit d'en dégager les différents usages, les typologies des méthodologies de recherche et les modalités de traitement de l'information. Dans la discussion peuvent ressortir des difficultés et des souhaits de changements.

- La troisième partie sur la dualité ou la complémentarité des bibliothèques classiques et numériques permet de mieux connaître le panéliste dans ses pratiques culturelles et ouvre le dialogue sur l'apport des bibliothèques numériques.

Pour chaque entretien, on disposait d'une fiche descriptive du panéliste (à partir de ses réponses au questionnaire de Gallica et des statistiques de son trafic Internet *via* la sonde) :

- profil socio-démographique ;
- nombre de sessions<sup>4</sup> par semaine ;
- nombre de sessions par heure de début de session ;
- durée moyenne des sessions ;
- mots clés recherchés par les moteurs ;
- sites portails et services utilisés.

## Résultats et observations

Les résultats détaillés du projet pourront être consultés dans le rapport public à paraître<sup>5</sup>. Les points saillants qui intéressent plus particulièrement la BnF pour positionner son offre de bibliothèque numérique et en faire évoluer les contenus et les services associés sont :

*L'intensité d'usage du web.* Les panélistes ont été actifs sur toute

3. Peer-to-peer.

4. Une session est une période d'activité sur le web entourée de deux périodes d'inactivité supérieures à 30 min.

5. <http://www.telecom.gouv.fr/rnrt/coll/presentations/bibusages.html>

la période d'observation. Même s'il existe des écarts importants entre de très gros utilisateurs et les autres, globalement les panélistes sont de gros utilisateurs du web

*La durée des sessions.* La durée moyenne des sessions est de 30 minutes. Elle peut être comparée à la durée moyenne de 25 minutes pour un panel NetValue d'internautes français en 2000.

*L'audience des sites visités.* Les portails généralistes arrivent en tête, aux côtés des outils de recherche d'information. Ceci va dans le sens des entretiens, où beaucoup d'interviewés évoquent des pratiques de recherche d'information.

On remarque également : Gallica dans les premiers, quel que soit le tri (nombre de sessions ou nombre de panélistes), *Le Monde*, bien placé en nombre de sessions, Amazon et Fnac, dans le Top 20 en nombre de panélistes.

*L'audience des « types de portails » utilisés.* De manière générale, le panel se caractérise par une très forte fréquentation des portails « culturels ». Les bibliothèques électroniques, les sites de biens culturels (Alapage, Fnac, Amazon...) et les sites de médias occupent une place privilégiée. Les sites consacrés à la généalogie (représentative d'un centre d'intérêt personnel très fort) ont une place importante dans les usages.

*Les moteurs de recherche.* Seize moteurs de recherche différents ont été utilisés : Google est loin devant les autres. L'utilisation de méta-moteurs concerne 24 personnes, mais seulement 10 d'entre elles les utilisent de manière systématique. Les moteurs sont utilisés dans 31,2 % des sessions (2513 sur 8052), contre 20 % des sessions pour le panel NetValue.

*17 sites identifiés comme bibliothèques électroniques.*

- Gallica arrive en tête, en nombre de sessions comme de panélistes, mais l'audience des autres sites est loin d'être anecdotique : on peut penser, à la lumière des entretiens, que Gallica figure, pour les internautes,

comme une source de textes parmi d'autres. La richesse du fonds explique la fréquentation plus forte que sur les autres sites.

- La fréquentation des bibliothèques électroniques est à mettre en parallèle avec la fréquentation de sites proposant « des contenus à lire ». L'usage des portails de type « Média de presse » est révélateur à cet égard (en particulier, usage intense du site du *Monde*).

- La fréquentation des bibliothèques électroniques est à mettre en parallèle avec la fréquentation de sites de e-commerce de biens culturels.

### Les utilisateurs intensifs de Gallica sont aussi des utilisateurs intensifs du web

- L'usage des portails de biens culturels montre l'importance de la bibliophilie, avec une fréquentation importante de Chapitre.com et de Livre-rare-book.

*La place de Gallica et des sites web de la BnF* Sur les 15 500 sessions recueillies :

- 7,8 % des sessions comportent un accès à un site de la BnF (sur \*.bnf.fr) ;
- 6,2 % des sessions comportent un accès à Gallica ;
- 22,4 % des sessions sur un site de la BnF n'ont pas d'accès à Gallica.

Les utilisateurs intensifs de Gallica sont aussi des utilisateurs intensifs du web : sur la période, ils comptent 400 sessions en moyenne, pour une moyenne de 260 et 280 sessions au total pour les utilisateurs moyens ou qui utilisent peu Gallica, et 150 sessions pour ceux qui n'y sont jamais allés.

*L'utilisation de Gallica.* Les 1 063 sessions comportant un accès Gallica sont en moyenne plus longues que les autres : 1 heure 1 minute en moyenne pour les sessions avec

Gallica, 28 minutes en moyenne pour les autres sessions. Les séquences de navigation sur Gallica ont une longueur moyenne de 10 minutes, pour une médiane de 2 minutes 12 secondes.

Dans une session avec accès Gallica, le temps total passé sur Gallica est en moyenne de 24 minutes, et on peut relever les caractéristiques suivantes :

- Pluri-activité, navigation alternée : dans 52 % des cas (558 sessions), la navigation sur Gallica occupe une seule séquence, elle n'est pas alternée avec la visite d'un autre site ; pour 22 % des sessions seulement, on compte 2 séquences distinctes sur Gallica ; dans 10 % des cas seulement, on compte 5 séquences Gallica ou plus.

- Le « multi-tâche » est rarement pratiqué : la consultation de Gallica est peu ou pas (75 % des cas) alternée avec la visite d'autres sites.

- La consultation est le mode majoritaire dans les services utilisés sur Gallica. La consultation des dossiers reste marginale (entretiens) en nombre de sessions, mais concerne la moitié du panel : effet de « visite par curiosité » (confirmée dans les entretiens).

- L'usage fréquent des moteurs semble montrer l'importance de l'usage de Gallica dans un contexte de recherche, où Gallica est une source parmi d'autres pour trouver de l'information.

- L'usage des sites de biens culturels semble correspondre à un effet de catalogue inversé, ou de « test avant achat » (entretiens).

- Les sites personnels, sur-représentés dans les sessions « avec Gallica », s'imposent comme sources de données sur des sujets pointus : effet d'échange entre spécialistes.

- Les sites de type « Média-presse » sont moins présents dans les sessions Gallica que dans l'ensemble : si les utilisateurs de Gallica sont de gros consommateurs de journaux en ligne (cf. *supra*), l'accès à ces deux types de sites ne correspond pas aux mêmes pratiques, et se fait dans des sessions, des contextes différenciés.

*Les informations recueillies lors des entretiens.* Le profil socio-démographique des personnes interrogées reste dans la lignée générale du panel et de la population générale de l'enquête de Gallica : pour la plupart occupant des postes de cadre de la fonction publique ou du secteur privé, la

Si Gallica  
a été conçu comme  
une salle de lecture  
de bibliothèque,  
son utilisation est aussi  
celle d'un réservoir  
conséquent d'informations  
sur le web, utilisé  
par un nouveau public  
manipulant les outils web

moyenne d'âge se situe autour de 48 ans avec une fourchette allant de 33 à 76 ans, habitant plutôt dans des milieux urbains.

Le logiciel-sonde avait en général été installé sur leur poste à domicile. Pour certains, cet ordinateur est un ordinateur familial dont ils sont l'utilisateur majoritaire. Des entretiens, il ressort que l'utilisation par les autres membres du foyer reste occasionnelle. Les choix individuels se sont portés sur l'équipement le plus utilisé pour l'installation de la sonde. C'est pourquoi seules deux personnes ont

installé ce logiciel sur leur lieu professionnel.

Dans le cadre de l'utilisation d'Internet, on remarque qu'il s'agit d'anciens internautes, qui ont, en moyenne, déjà plus de trois ans de pratiques, un usage fréquent et une consommation importante (mesurée *via* les données de trafic envoyées par la sonde logicielle).

Sur les 16 personnes interrogées, seulement 3 fréquentent régulièrement ou très occasionnellement la bibliothèque physique.

Les conclusions du projet mettent en évidence plusieurs « portraits types » d'utilisateurs :

- *Le chercheur d'information* : Gallica joue un rôle de médiation à l'intérieur d'Internet. Intérêt pour Gallica comme « source d'information primaire ». Gallica est une source parmi d'autres, utilisée conjointement avec des moteurs de recherche. Recherches dans un cadre professionnel ou par hobby, passion.

- *Le bibliophile* : Gallica est utilisé en « pré-achat » (ou pré-emprunt), le but reste l'objet-livre. Gallica joue un rôle de médiation vers le monde « réel » (y compris la sphère marchande).

- *Le « lecteur à l'écran »* : le déplacement des pratiques de gros lecteurs du papier vers l'écran est un profil rare (1 entretien sur 16).

Gallica apparaît comme source de textes pour les lecteurs étrangers. On retrouve globalement un profil de chercheur amateur. La lecture en ligne est « rare ». On peut plutôt parler de « consultation », « navigation ». Le type de recherche s'apparente plus à une recherche d'usuel qu'à celle

d'une œuvre à proprement parler : pour l'œuvre, on passe souvent à d'autres supports, en particulier par le téléchargement de documents complets qui seront utilisés localement hors connexion.

De ce fait, Gallica acquiert une position d'intermédiation dans le monde Internet et vers la sphère marchande (bouquinistes et sites de biens culturels).

## Conclusion

Au-delà des résultats du projet BibUsages qui montrent bien l'intégration de la bibliothèque numérique dans l'information culturelle sur le web, comme complémentaire et non comme un site public gratuit et concurrent de la sphère marchande des biens culturels, les données recueillies par la sonde vont permettre une analyse plus poussée de l'utilisation des contenus et des métadonnées de Gallica.

Les modalités de recherche et de navigation dans l'interface de Gallica qui ont été tracées dans le détail par la sonde font apparaître que, si Gallica a été conçu comme une salle de lecture de bibliothèque avec son catalogue, sa classification des ressources, ses parcours thématiques organisés, son utilisation est bien aussi celle d'un réservoir conséquent d'informations sur le web, utilisé par un nouveau public manipulant les outils web dans des modalités de recherche, de navigation et de téléchargement de documents.

*Avril 2003*