



**HAL**  
open science

## Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique

Valérie Beaudouin, Jérôme Denis

► **To cite this version:**

Valérie Beaudouin, Jérôme Denis. Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique. [Rapport de recherche] BnF; Telecom ParisTech. 2014. halshs-01078530

**HAL Id: halshs-01078530**

**<https://shs.hal.science/halshs-01078530v1>**

Submitted on 29 Oct 2014

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

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



## Observer et évaluer les usages de Gallica

Réflexion épistémologique et stratégique

Rapport final

Septembre 2014

**Valérie Beaudouin**  
**Jérôme Denis**  
Télécom ParisTech



## Sommaire

<b>1. Présentation de l'étude</b> .....	<b>2</b>
<b>1.1. Méthode</b> .....	<b>2</b>
<b>1.2. Plan du rapport</b> .....	<b>2</b>
<b>2. Les usages de Gallica vus de l'intérieur</b> .....	<b>3</b>
<b>2.1. La présence multiple des usagers</b> .....	<b>3</b>
2.1.1. Supports et lieux.....	3
2.1.2. Outils et formes de représentation des usagers.....	4
2.1.3. Une pluralité d'entités.....	8
<b>2.2. Problématiques émergentes</b> .....	<b>10</b>
2.2.1. De nouvelles places pour les usagers.....	10
2.2.2. Nouvelles questions.....	10
<b>2.3. Ouverture</b> .....	<b>11</b>
<b>3. État de l'art des méthodes pour appréhender les usages numériques</b> .....	<b>12</b>
<b>3.1. La connaissance des clients dans les organisations</b> .....	<b>12</b>
3.1.1. Les études.....	13
3.1.2. L'audience.....	19
3.1.3. Du réseau à l'influence.....	20
3.1.4. La gestion de la relation client.....	23
<b>3.2. La participation des publics</b> .....	<b>23</b>
<b>3.3. Conclusion</b> .....	<b>25</b>
<b>4. Pistes méthodologiques et recommandations</b> .....	<b>26</b>
<b>4.1. Orientations générales</b> .....	<b>26</b>
4.1.1. L'articulation des outils et des formats d'investigation.....	26
4.1.2. Décentrement.....	26
<b>4.2. Propositions opérationnelles</b> .....	<b>28</b>
4.2.1. Relancer une étude globale croisant les méthodes.....	28
4.2.2. Utiliser la mesure d'audience pour étudier l'usage des documents.....	28
4.2.3. Vers un utilisateur « global ».....	29
<b>4.3. Donner plus de visibilité à la participation des publics ?</b> .....	<b>29</b>
<b>4.4. Aspects organisationnels</b> .....	<b>30</b>
<b>5. Bibliographie</b> .....	<b>31</b>



## 1. Présentation de l'étude

Cette étude s'inscrit dans le cadre de la convention de partenariat entre la Bibliothèque nationale de France et Télécom ParisTech, qui a donné naissance au *Laboratoire d'études des usages du patrimoine numérique des bibliothèques* (Bibli-Lab). Elle a pour but de consolider les fondations épistémologiques du Laboratoire en identifiant les enjeux théoriques et méthodologiques que soulève l'appréhension des usages de Gallica au regard des récents bouleversements dans les usages en ligne.

Cette étude comprend trois volets :

- un retour sur les pratiques actuelles d'observation et de mesure des usages de Gallica au sein de la BnF,
- un état de l'art international des méthodes innovantes dans l'analyse des usages en ligne,
- un ensemble de propositions et recommandations opérationnelles pour faire évoluer les processus d'étude des publics et des usages, en prenant acte des transformations actuelles liées au numérique. Ce troisième volet s'appuie sur les précédents.

Elle a été suivie par un comité de projet composé de Catherine Aurérin, Julien Barbier, Philippe Chevallier, Corinne Le Bitouzé, Aline Girard, Mélanie Leroy-Terquem, Thierry Pardé, Stéphane Pillorget.

### 1.1. Méthode

Le premier volet, qui rend compte de la manière dont la BnF appréhende ses publics et les usages de Gallica, s'appuie sur une série d'entretiens semi-directifs avec des agents de la BnF (18), menés comme des conversations ouvertes, axées sur un nombre limité de grandes thématiques. Celles-ci ont été adaptées à la marge au poste des personnes interrogées, ainsi qu'à leur plus ou moins grande proximité avec Gallica et avec le public lui-même.

Les entretiens se sont focalisés sur des aspects très concrets des pratiques professionnelles, afin d'éviter les discours trop généralisant. Dans la mesure du possible ils ont été effectués sur les lieux de travail, afin que les outils mobilisés puissent-être présentés et manipulés pendant la conversation.

Les entretiens ont duré entre 45 minutes et 2 heures. Ils ont été enregistrés et retranscrits pour en faciliter l'analyse.

Le second volet, état de l'art des méthodes employées dans la recherche et dans les institutions culturelles pour connaître leurs publics, s'appuie sur deux types de sources : d'une part, les pratiques développées par les entreprises et organisations pour connaître leurs clients ou publics, telles qu'on peut les appréhender par de la veille, par leur discours et par leurs pratiques et, d'autre part, une revue des méthodes employées dans des travaux de recherche sur les usages du numérique. Une attention particulière sera portée aux travaux portant sur les publics des institutions culturelles.

### 1.2. Plan du rapport

Le premier volet se distribue en deux parties.

La première se concentre sur les formes de présence des usagers de Gallica dans les différents départements de la BnF étudiés. Elle s'arrête sur les outils qui assurent cette présence et sur les entités à travers lesquelles les usagers apparaissent dans les préoccupations de chacun. La deuxième partie recense les principales transformations en cours sur ce sujet, tant du point de vue de la place que les

connaissances sur les usagers et les usagers eux-mêmes pourraient prendre dans les processus de travail que sur les problématiques émergentes au sein des départements.

Le second volet est organisé en deux parties.

Une première partie porte sur la [connaissance client](#). Après avoir présenté les méthodes traditionnelles en sciences sociales pour appréhender les publics, [les études](#), en montrant comment elles se voient transformées par le numérique, cette partie passera en revue trois autres champs d'exploration qui ont vu leur importance grandir avec le numérique : [la mesure d'audience](#), [la veille et l'évaluation de l'influence](#), [la gestion de la relation client](#), qui offrent toutes trois de nouvelles ressources pour la connaissance des usagers et des publics. La seconde partie portera sur le tournant que représente le passage de la connaissance des usagers à [la participation des publics](#).

Le dernier volet, pistes et recommandations, est organisé en deux temps : des pistes méthodologiques suivies de pistes organisationnelles.

## 2. Les usages de Gallica vus de l'intérieur

### 2.1. La présence multiple des usagers

Les entretiens ont permis de recenser la grande variété des modalités de présence des usagers de Gallica. Si les occasions au cours desquelles ces figures sont mobilisées sont trop nombreuses pour en faire l'inventaire, nous proposons en revanche d'interroger ces modalités sous deux angles complémentaires. Tout d'abord, nous nous arrêterons sur les différents outils qui permettent de produire des représentations des usagers. Chacun de ces outils propose des figures spécifiques qui n'ont pas le même poids ni la même force aux yeux de ceux qui les manipulent ou qui les consultent. Ensuite, nous tenterons d'organiser cette première mise à plat en donnant à voir les « entités » dans lesquelles les usagers s'incarnent. Ces entités se recoupent en partie avec les formes concrètes que les outils produisent, mais s'inscrivent également dans des questionnements plus génériques. Nous verrons que c'est la pluralité des entités d'usagers de Gallica qui fait la richesse des pratiques au sein de la BnF.

Avant d'entrer dans le vif du sujet, il est utile de revenir brièvement sur les supports et les lieux de présence des usagers.

#### 2.1.1. Supports et lieux

On n'y pense pas forcément au premier abord, mais l'un des supports de la présence des usagers en ligne est un espace physique : [la salle de lecture](#). Plusieurs personnes nous ont en effet relaté leurs expériences de ce point de vue, échangeant à propos de Gallica à l'occasion d'une rencontre en salle. Qu'il s'agisse par exemple de demandes de conseils (« Comment retrouver cela en ligne ? ») ou d'historiques de recherche (« J'ai vu que vous aviez tel document »), Gallica est bel et bien présent au sein même de la Bibliothèque dont une partie des lecteurs sont, évidemment, aussi internautes. Cette première forme de présence invite à rester prudent quant à l'automatisation d'un public « en ligne » qu'il s'agirait d'étudier et de mesurer indépendamment d'un public exclusivement présent en salles.

Les [messages](#), envoyés par mail ou par l'intermédiaire du formulaire disponible sur le site web constituent un autre support de la présence des usagers. Comme dans le cas des salles de lecture, ils donnent lieu à des échanges de personne à personne. Ils sont généralement orientés autour de deux principaux types d'objet : la plainte et, à l'inverse, l'expression d'une satisfaction. Une partie de ces messages directement adressés prennent également la forme de remarques, voire de propositions de

correction ou de complément à propos de la description d'un document. Ils témoignent à la fois de l'attachement de certains envers les collections en ligne et d'une forme d'expertise particulière. Dans le cas où ils ne peuvent faire l'objet d'un traitement immédiat par la personne qui les a reçus, ou s'ils sont susceptibles d'intéresser des équipes en particulier, certains de ces messages sont amenés à circuler dans l'institution.

Les sites de [réseaux sociaux](#) sur lesquels Gallica est présent offrent également un espace particulier de présence des usagers. Sur Facebook, Twitter et Pinterest, l'équipe dédiée à l'animation des pages entre en contact avec de nombreux usagers, sur un modèle un peu plus élargi que celui des échanges précédents. À ces discussions en ligne s'ajoute le travail de veille qui permet de repérer parmi les usagers de Gallica ceux qui produisent [sur le web](#) des documents (blogs, pages associatives, etc.) qui témoignent de leurs usages. Depuis quelques années, ces deux supports de présence étroitement liés ont pris une place importante dans l'appréhension et la compréhension des usagers. Dans une veine assez proche de la veille effectuée en ligne, il arrive par ailleurs que la [revue de presse](#) transmise par la délégation à la Communication donne à voir des usagers de Gallica mis à l'honneur dans un journal.

Les usagers apparaissent également dans les [rapports d'études](#). Ils sont alors présentés sous l'angle d'une problématique spécifique (l'usage de tel type de documents, les besoins de tels types d'usagers, etc.) travaillée avec des méthodes d'enquête stabilisées. Si ces études constituent généralement des réponses à des demandes faites par des équipes ou des personnes en particulier, les rapports ont toutefois vocation, à l'instar des revues de presse, à circuler largement dans l'institution.

La BnF dispose d'un outil de mesure des traces d'usage de ses sites web et des applications mobiles qui leur sont associées : XiTi. Cet outil permet, entre autres, d'exporter un [bilan](#) quotidien qui donne à voir certaines tendances d'usages. Ces mesures peuvent également être effectuées ponctuellement autour de questions plus précises et circulent dans ce cas dans des cercles plus restreints.

Enfin, les usagers sont présents dans les [spécifications fonctionnelles](#) qui nourrissent la conception de nouveaux services. Généralement, ces spécifications sont le résultat plus ou moins direct des nombreux autres supports précédents. Elles cristallisent des problèmes et des besoins d'usage spécifiques et offrent les moyens concrets par lesquels les services sont chargés d'y répondre.

### 2.1.2. Outils et formes de représentation des usagers

Ces différents supports permettent de dessiner une première carte « à plat » de la présence des usagers de Gallica au sein de la BnF. Pour comprendre les enjeux que représente la variété de ces représentations, il est toutefois utile de se pencher sur les principales technologies qui les produisent. Outre les particularités dans les manières dont ils donnent forme aux usagers, ces outils présentent en effet aux yeux des personnes interrogées des forces et des faiblesses qu'il est important de prendre en considération. En voici une présentation synthétique.

#### *Les statistiques du web*

Les usages en ligne laissent des traces et nombreux sont les outils statistiques qui se sont développés avec le déploiement d'internet pour enregistrer, mesurer et analyser ces traces, devenues la clef de la compétition sur le web et de sa monétisation. La BnF opère ces mesures avec XiTi, l'un des leaders du marché. Le principe de fonctionnement de XiTi est similaire à de nombreux autres outils : il repose sur l'inscription d'une marque dans le corps même du code des pages web. Le nombre et le type de ces marques varient selon les objets dont on souhaite mesurer l'accès et l'usage. Au-delà des racines de site et des pages principales, l'outil permet d'atteindre un niveau de détail important, allant jusqu'au moindre élément de chaque page.



Ce type de statistiques met en scène les usagers sous la forme d'indicateurs quantitatifs qui ont une double vocation dans l'institution : interne (en direction des équipes) et externe (en direction des autorités de tutelle, ou du grand public). Jusqu'ici, les statistiques XiTi sont surtout utilisées pour produire des chiffres de tendance, à des niveaux d'agrégation assez larges, nous y reviendrons.

Les forces des outils de ce type tiennent à leur grande modularité. On peut adopter avec XiTi des unités de mesure très variées et suivre des objets aux contours très différents. La finesse du grain de la mesure est également un des grands atouts de ces instruments qui permettent d'étudier les usages à des échelles très précises.

Les principales faiblesses de ces outils résident dans le coût que représente leur mise en œuvre. Si les statistiques de mesure sont produites automatiquement une fois le dispositif installé, la mise en place de chaque nouveau type de mesure nécessite du temps pour le déploiement du marquage et, en amont, une analyse fine des besoins, ainsi que des objets et des usages que l'on cherche à étudier. Une autre faiblesse tient au risque de faire de cet instrument une boîte noire. Comme toutes les statistiques, et contrairement à ce que laisse croire leur disponibilité apparente, les chiffres que produisent les outils de mesure d'activités en ligne ne sont pas utilisables sans un travail important. Sans mise en contexte, sans recul et sans lissage, ils ne peuvent être mobilisés comme ressources sûres. Ces chiffres nécessitent une compréhension des processus de mesure et de calcul et leur circulation à l'état « brut » est toujours délicate. Ce point, scientifiquement essentiel quant à la manipulation des statistiques, a été souligné à plusieurs reprises dans les entretiens qui insistent à la fois sur la prudence et sur l'analyse nécessaire au bon usage des traces d'usage. À cela s'ajoute la complexité des fonctionnalités disponibles. Si l'accès à l'outil est virtuellement ouvert à tous (il suffit de demander l'ouverture d'un compte), son intelligibilité n'a rien d'évident.

### *Études quantitatives*

Parmi les travaux commandités ou réalisés par la délégation à la Stratégie et à la recherche de la BnF, les études quantitatives ne sont pas les plus nombreuses, mais elles tiennent une place importante. La grande enquête confiée à un prestataire en 2011 est régulièrement citée dans les entretiens et semble avoir joué un rôle non négligeable dans les formes de représentation des usagers de Gallica au sein de la BnF. La part quantitative de ces études s'appuie sur des questionnaires (en ligne ou par téléphone) qui donnent lieu à des statistiques descriptives.

Ce type d'études présente donc les usagers à partir d'un matériau déclaratif sous la forme de chiffres agrégés qui circulent ensuite sous la forme d'indicateurs internes et externes, de chiffres « clefs » et de baromètres. Les résultats sont également présentés sous forme de catégories d'usagers à partir de variables explicatives (catégories socioprofessionnelles, sexe, âge, etc.) ou de variables d'usages, généralement croisées. Une partie des problématiques de ces études a explicitement un but d'évaluation : au-delà des usages en tant que tels, ce sont dans ce cas des avis et des jugements qui sont recueillis.

La force de cet outil réside essentiellement dans la forme de ses résultats. Inscrits dans la tradition statistique, ces derniers constituent en effet des appuis conventionnels solides. Les chiffres et les catégories mobilisés sont compatibles avec les études classiques en marketing et sont compréhensibles par un grand nombre de personnes. En outre, les statistiques assurent aux résultats une forme de représentativité qui n'est que rarement remise en question et renforce aussi bien leur stabilité que leur capacité de circulation auprès des équipes.

Les faiblesses de ces études se cristallisent autour de deux éléments. Comme toute méthode qui s'appuie sur un matériau déclaratif, elle présente d'abord des risques de surinterprétation dès qu'il s'agit des usages au sens strict du terme. La question n'est pas tant de savoir si les personnes sont

honnêtes lorsqu'elles répondent à ces questionnaires, que d'assumer qu'une partie des réponses s'appuient sur des estimations voire des ignorances, et témoignent davantage d'un certain rapport aux services en ligne que d'une description « neutre » d'une réalité univoque. Par ailleurs, précisément parce qu'elles s'inscrivent dans une tradition d'études d'usages dont la forme et la méthode sont stabilisées, ces études n'apportent pas de résultats innovants ou surprenants au sens fort du terme. Par nature, les sondages et les questionnaires sont hypothético-déductifs : ils permettent de tester des hypothèses définies à l'avance, de traiter des questions clairement délimitées. Ils ne sont pas orientés vers la découverte d'aspects inédits.

### *Études qualitatives*

Certaines études commanditées ou réalisées en interne s'appuient sur des méthodes qualitatives. Organisées autour d'une série d'entretiens approfondis, elles sont généralement menées autour de questions précises ou de domaines d'activités restreints. Elles sont davantage tournées vers la description riche d'usages et débouchent également, dans une démarche prospective, sur l'identification d'attentes de la part des usagers.

Ces études font exister les usagers sous la forme de verbatims, témoignages en personne d'usages ou d'avis. Elles permettent également d'élaborer des typologies d'usagers, parfois des scénarios d'usage.

La force des études qualitatives tient à la fois à la richesse des matériaux qu'elles permettent de glaner et à leur pertinence pratique. Parce qu'elles assument le temps long de l'entretien, elles offrent un grand degré d'approfondissement de certaines questions. Parce qu'elles laissent aux usagers le choix du vocabulaire, des questions et des catégories, elles restent fermement attachées aux contextes concrets des usages. En conséquence, les études qualitatives sont davantage tournées vers la découverte que vers la vérification d'hypothèses. Pour peu que l'on tienne la position jusqu'au bout, elles peuvent produire des résultats surprenants, susceptibles de nourrir une véritable créativité pour les équipes qui s'en inspirent.

Les faiblesses du qualitatif sont l'exact miroir de ses forces. Parce qu'ils sont ancrés et contextuels, les résultats échappent à toute forme de représentativité statistique. De ce point de vue, ils présentent un risque fort de surinterprétation et surtout d'utilisation détournée. Il est en effet tentant de traiter ces résultats sur un mode implicitement statistique en généralisant à outrance telle ou telle situation explorée en entretiens. Cette faiblesse tient donc dans la difficulté à assumer véritablement la nature des matériaux qualitatifs : singuliers et narratifs. S'y ajoute précisément la difficulté de tels matériaux à circuler au sein des équipes. Parce que leur forme n'est pas conventionnelle, qu'elle n'est jamais stabilisée, les résultats qualitatifs restent fragiles, lorsqu'ils ne sont pas purement et simplement considérés par certains comme irrecevables parce que non représentatifs.

### *Tests*

Les tests effectués à l'occasion du développement de nouveaux services en ligne ou de leur mise à jour permettent également d'offrir une certaine forme de présence aux usagers. Celle-ci est très ciblée, centrée sur un service, voire quelques fonctionnalités. Ce type de représentation est essentiel à l'organisation même des équipes chargées des outils de Gallica qui fonctionnent en méthode Agile, laquelle suppose des retours « usagers » réguliers. Les tests, notamment menés sur une plateforme dédiée (les « labs » en ligne), ont une visée d'évaluation et de correction immédiate.

Les formes de représentations sont variées dans ce cadre : elles combinent statistiques d'usages (traces que laissent les tests), évaluation plus ou moins formalisée et verbatims détaillés.

La force des tests tient au faible coût de leur mise en œuvre, à leur caractère opérationnel et à la proximité qu'ils entretiennent avec les usagers en chair et en os.

Leurs faiblesses sont de deux natures. Le recrutement des testeurs est d'abord une opération délicate qui est de plus en plus difficile au fil du temps, pour peu que l'on cherche à ne pas solliciter toujours les mêmes personnes. La population des testeurs elle-même est ensuite susceptible de produire des biais importants. À l'intérêt de passer les prototypes au crible de personnes expertes et motivées correspond le risque de laisser invisibles des problèmes qui concerneraient des usagers « lambda », moins compétents.

### *Les réseaux sociaux*

En quelques années, les sites dits de « réseaux sociaux » sont devenus des lieux incontournables dans les relations entre les entreprises, les institutions et leur public. Ils constituent, en tant qu'outil de mise en présence des usagers, des dispositifs souples, mobilisables dans des cadres variés.

Gallica est présent sur Facebook, Twitter et Pinterest. Pour ce qui est de la mise en présence des usagers, ces plateformes sont mobilisées autour de deux axes principaux. Elles permettent surtout de nourrir des *échanges* avec une partie du public en ligne, qui peuvent prendre la forme de demandes directes et d'avis sur des aspects particuliers de Gallica, mais aussi de conversations beaucoup plus larges autour des publications de l'équipe dédiée aux réseaux sociaux.

La présence sur ces sites permet aussi d'entretenir et d'organiser une veille sur le web, grâce à laquelle des usages innovants peuvent être repérés et mis en avant. Cette veille ne se réduit pas aux grandes plateformes de réseaux sociaux même si elle y est ancrée. L'utilisation de Pinterest permet notamment de garder des traces d'usages glanés ici ou là et de les publier.

Les conversations, comme la veille, permettent donc d'identifier et d'entrer directement en contact avec des personnes ressources qui peuvent être mobilisées à la fois dans la communication même de Gallica (dans le blog, la newsletter, et à terme sur le site même de Gallica) mais aussi comme relais auprès de communautés d'usagers ou de testeurs potentiels.

La force des échanges en ligne tient évidemment dans leur caractère ordinaire et « naturel ». La coprésence de Gallica et des usagers sur les plateformes de réseaux sociaux offre des occasions inédites pour tisser des liens très différents de tout ce qui a été évoqué jusqu'ici, indépendamment de toute situation « d'étude ». Associés à la veille, ces échanges sont notamment un outil précieux pour repérer des communautés d'usagers qui partagent en ligne des intérêts — voire des passions — qui impliquent des usages parfois surprenants de Gallica. La connaissance de ces communautés peut représenter une ressource importante pour les équipes de Gallica, à la fois pour étoffer les connaissances des usages, mais aussi dans la quête de relation directe avec le public.

La faiblesse de ce type de relations est du même ordre que les tests : les usagers présents sur les plateformes de réseaux sociaux, s'ils ne sont pas forcément tous compétents et spécialisés, ne constituent pas moins une niche au sein des usagers potentiels de Gallica, eux non plus ne doivent donc pas être considérés comme représentatifs d'un public global.

Accessoirement pour l'instant, ces plateformes permettent également d'obtenir des statistiques qui mesurent la place de Gallica auprès de leurs publics et en spécifient certaines caractéristiques. Comme toutes les « métriques » du web, elles sont utiles puisqu'elles mesurent des usages extérieurs au site même de Gallica. Comme nous le verrons dans le second volet de cette étude, ce point est un aspect critique des analyses des usages en ligne aujourd'hui. Ces métriques sont toutefois à manipuler avec précaution, puisqu'elles sont de véritables boîtes noires, dont on ne maîtrise pas grand-chose. Cependant, les statistiques de la page Facebook laissent entendre par exemple que la population qui la suit est sensiblement plus jeune que celle de Gallica, mesurée par l'enquête GMV.

\*

Le panorama des principaux outils qui produisent des représentations des usagers souligne la grande variété de celles-ci. Cette variété est une grande richesse, notamment parce qu'elle évite l'écueil que représente la forte domination des méthodes quantitatives que l'on observe dans de nombreuses institutions. Or, nous l'avons vu dans les usages qui peuvent être faits de ces différentes représentations : leur mobilisation s'inscrit dans des enjeux pratiques et stratégiques spécifiques. Pour y voir plus clair, nous proposons de nous détacher des outils en tant que tels et de mettre en lumière les entités sous lesquelles se rassemblent ce que nous avons volontairement appelé de manière floue jusqu'ici les « usagers ».

### 2.1.3. Une pluralité d'entités

Pour organiser des formes d'usagers, nous proposons d'adopter la perspective mise en œuvre par François Eymard-Duvernay et Emmanuelle Marchal<sup>1</sup> à propos des méthodes d'embauche dans le monde du travail en les distribuant du plus lointain (et plus abstrait) au plus proche.

#### *Le public*

L'idée de « public » est la plus lâche des catégories, et la plus abstraite. C'est une entité floue, sur le modèle de la foule, qui cristallise certaines préoccupations de manière très générique. Quasiment pas instrumentée, elle représente l'ensemble des personnes susceptibles d'être touchées par Gallica. La notion de « grand public » est sans doute la manière la plus fréquente de produire des découpages dans cette entité globale, généralement face à d'autres segments (le public de chercheurs, de spécialistes...).

#### *L'utilisateur d'interfaces*

L'utilisateur d'interfaces est une entité en décalage par rapport à toutes les autres, mais il a une importance considérable dans les préoccupations de nombreuses personnes en interne. Révélée par des tests instrumentés, ou simplement imaginée, parfois éprouvée en personne, cette figure est à la fois incarnée et abstraite. Généralement détachée des autres entités, elle s'articule autour des questions essentielles, et sans cesse évoquées au cours des entretiens, de « l'utilisabilité » des interfaces des différents services en ligne (le moteur de recherche, le lecteur, les applications mobiles...). L'utilisateur d'interfaces est essentiellement défini par des capacités cognitives génériques, des modalités de résolution de problèmes au sens large du terme, face à des fonctionnalités dont on cherche à améliorer l'ergonomie.

Cette entité est évidemment fortement mobilisée au sein des équipes techniques de la DSR, mais est aussi présente du côté des collections lorsque l'on se demande si tel ou tel document que l'on a numérisé récemment est « trouvable » sur le site, ou lorsque l'on cherche à élaborer des métadonnées qui s'inscrivent dans les logiques de recherche des usagers.

#### *L'audience*

L'audience est la forme calculable des usagers. Sur le modèle de la radio et de la télévision, elle est une entité modulable qui s'agrège, s'additionne et permet d'envisager les usagers en termes de « volume » de présence.

---

<sup>1</sup> Eymard-Duvernay, F et Marchal, E, 1997. *Façons de recruter. Le jugement des compétences sur le marché du travail*, Métailié, Paris.

### *Les sociotypes*

Les sociotypes rassemblent toutes les manières de segmenter le public ou l'audience par le biais de catégories extérieures à l'usage lui-même : lieu de résidence, âge, sexe, profession, diplômes, etc. Ils tiennent une place importante dans une problématique de démocratisation d'accès à la culture. C'est à travers ce type de segmentations, rapportées à l'ensemble de la population et comparées aux chiffres de fréquentation des salles de lecture, que peuvent être évaluées des avancées dans la diffusion en ligne des documents de la BnF.

### *Les profils d'utilisateurs*

Les profils d'utilisateurs s'appuient généralement sur des sociotypes, mais y associent des caractéristiques d'usage de Gallica. Les profils désignent des types d'utilisateurs dans leur rapport plus ou moins concret aux documents et aux fonctionnalités des services en ligne. Ces profils peuvent être statiques, assez génériques, mais ils peuvent aussi prendre des formes beaucoup plus détaillées et narratives. Dans tous les cas, cette entité permet de manipuler des êtres fictifs, nourris de connaissances quantitatives et qualitatives, qui offrent des repères solides.

### *Les communautés de pratiques*

Dans certains cas, les segmentations du public s'appuient sur des caractéristiques qui débordent les variables sociales habituelles et les usages au sens strict du terme. C'est le cas lorsque les utilisateurs apparaissent sous la forme de ce que Étienne Wenger et Jean Lave appellent des « communautés de pratiques » (Lave and Wenger, 1991). Ces dernières sont constituées de personnes qui ne se côtoient pas forcément directement, mais qui partagent un même objet (qui peut être d'ordre professionnel, mais aussi de loisir), un même vocabulaire et des problématiques similaires. Dans le milieu académique, ces communautés se confondent généralement avec les disciplines ou les spécialités. Mais on les trouve également ailleurs, les généalogistes amateurs en étant un exemple paradigmatique. Dans tous les cas, ces communautés sont supposées présenter une certaine homogénéité autant dans les usages que dans les attentes de leurs membres.

### *Les communautés en ligne*

Les communautés en ligne constituent une autre forme de collectif, plus resserré, généralement intégrées à des communautés de pratiques. Elles sont constituées de personnes qui partagent un même espace de discussion (un forum, une liste de diffusion...) et se donnent à voir comme groupe consolidé sur le web. Les membres du forum « Pages 14-18 » sont récemment apparus comme une communauté importante, par exemple. Ici aussi, la question de l'homogénéité des usages est cruciale. S'y ajoute la possibilité d'une observation directe, voire d'un contact de la part des équipes de la BnF, puisque c'est précisément leur caractère public qui définit ces communautés.

### *Les points de contact*

Enfin, les utilisateurs peuvent apparaître en chair et en os, comme personne à part entière que l'on contacte. En salle, par courrier électronique, au téléphone, au fil de conversations sur les réseaux sociaux, ces points de contacts peuvent être sollicités, ou solliciter, à propos de sujets très variés et peuvent également participer à la mobilisation d'autres entités, de l'utilisateur d'interfaces (en devenant bêta-testeur par exemple) à la communauté de pratiques (dont on peut en faire un représentant).

## 2.2. Problématiques émergentes

Une fois cette première mise à plat effectuée, qui donne à voir comment et sous quelles formes les usagers de Gallica « peuplent » les départements de la BnF, nous proposons de nous arrêter sur les transformations qui sont aujourd’hui identifiées par les personnes que nous avons interrogées. Ces transformations témoignent à la fois de premières expériences et d’appétits pour mettre en œuvre de nouvelles manières de mobiliser les usagers dans les processus de travail.

### 2.2.1. De nouvelles places pour les usagers

Deux mouvements sont aujourd’hui sensibles à la BnF, à propos des connaissances des usagers de Gallica et des nouvelles places qu’elles pourraient tenir.

#### *En amont de la numérisation*

Le premier témoigne d’une intégration, ou plutôt d’une certaine envie d’intégration, plus en amont dans la trajectoire des documents numérisés. Jusqu’ici, les usagers, quelles qu’en soient les manifestations, sont appréhendés en aval de la numérisation, plutôt interrogés sur des fonctionnalités et des services que sur les corpus en tant que tels.

En complément de ces pratiques d’évaluation au sens large, sont aujourd’hui envisagées, imaginées, des formes de présence des usagers pour compléter les processus de la numérisation en tant que telle. L’observation et la compréhension des usages permettraient de nourrir une partie de la politique documentaire elle-même, et d’outiller la production de nouvelles métadonnées.

Ce mouvement semble témoigner d’un déplacement plus large, qui marquerait la fin d’une période de numérisation dite de masse pendant laquelle le volume des documents numérisés importait avant tout. Une fois cette offre pléthorique installée, élaborée à la croisée des logiques scientifiques de corpus et des logiques techniques (conditions de numérisation, marchés, état des documents...), il semble qu’il soit possible d’envisager des processus non pas plaqués sur des attentes supposées explicites d’usagers, mais plus proches de leurs pratiques et de leurs préoccupations.

#### *Le spectre de la contribution*

L’autre mouvement s’inscrit dans des transformations qui dépassent largement le cadre de la BnF. Plusieurs indices laissent entendre que Gallica pourrait être un des points d’entrée d’un public beaucoup plus actif, ouvrant la voie à des formes de contribution et de participation sur des objets variés. Un projet de recherche est actuellement en cours sur ces questions (autour de la plateforme CORRECT) et sera l’occasion de développer un prototype autour des corrections d’OCR collaboratives. D’autres domaines de contribution sont envisagés de manière plus ou moins officielle au sein de la BnF.

Il y a sur le sujet un intérêt tout particulier à ce que ces développements ne se fassent pas indépendamment des formes d’appréhension des usagers existantes. Les études aussi bien que la veille en ligne et le travail d’animation sur les plateformes de réseaux sociaux offrent des appuis utiles pour mettre en place des formes de contribution des usagers solides et durables.

### 2.2.2. Nouvelles questions

À cette question de la place des usagers s’ajoute une marque d’intérêt pour de nouvelles problématiques associées aux usages. Pour résumer, ces problématiques invitent à compléter en les dépassant les questions habituelles centrées sur l’audience et les utilisateurs d’interfaces.



Il y a d'abord un intérêt confirmé pour explorer les [communautés en ligne](#). Nombreux sont ceux qui y voient une source riche d'inspiration et de participation. Si l'on considère les usagers en personne, on note également un intérêt partagé pour l'étude des [parcours](#). Au-delà des usages de chaque service, il y aurait par exemple une utilité à comprendre comment les usagers se débrouillent au fil de leur activité avec les différents points d'entrée de la BnF en ligne (Gallica, le catalogue général, data.bnf.fr, les bibliothèques partenaires...) au fil de leurs parcours sur le web. Plus généralement, la notion de parcours pourrait être élargie afin de comprendre la place des usages de Gallica dans des pratiques en ligne décloisonnées (qui comprennent d'autres bibliothèques en ligne, mais aussi Google Books et sans doute beaucoup d'autres sites).

Les entretiens ont également révélé l'utilité que représenterait l'[articulation des études des usages en ligne et des études des usages en salles](#). Si le plaquage d'une connaissance intuitive des usagers en salles sur celle du public en ligne est écarté par tout le monde, la possibilité de croiser les connaissances entre les salles et Gallica est envisagée comme une opportunité forte à la fois pour mesurer de manière générale un « effet » Gallica (la consultation de documents peu demandés en salle par exemple), mais aussi pour comprendre plus finement les démarches des usagers qui profitent des deux modalités de consultation.

Enfin, une analyse affinée des traces d'usages du [point de vue des documents](#) est également apparue comme un horizon imaginé à court terme. Les premiers dispositifs qui ont été mis en place dans ce sens avec XiTi (le taux de consultation par types de document) semblent avoir ouvert les appétits pour l'élaboration d'indicateurs internes détaillés et enrichis.

### 2.3. Ouverture

En guise, non pas de conclusion, mais d'ouverture, il est sans doute utile de reposer en termes plus généraux la question de la place du public ou des usagers de Gallica au sein des pratiques de la BnF. Outre les résultats exposés jusqu'ici, cette série d'entretiens a fait émerger une problématique sous-jacente qui semble traverser toutes les équipes sous une forme plutôt implicite. Pourquoi faut-il se préoccuper des usages ? Pourquoi faut-il penser au public ? C'est l'intérêt d'une méthode qualitative : ne pas imposer une thématique sous la forme d'un cadre de questionnements trop rigide pour au contraire favoriser la réflexivité. Au détour des réponses et des conversations, cette vaste interrogation a donc émergé, même si elle n'a jamais été abordée frontalement et de manière approfondie, sauf à quelques rares occasions.

Deux horizons de ces moments réflexifs se dessinent assez nettement. Penser les usagers, mesurer leur activité, observer leurs comportements, les interroger peut servir dans une perspective de compréhension d'une réalité complexe faite de pratiques hétérogènes auxquelles on souhaite adapter des services, ou dans une perspective plus focalisée de description d'attentes et de besoins d'usagers qu'il s'agit de « cibler ». Ces deux perspectives ne représentent évidemment pas des options qui s'excluent, mais des directions stratégiques possibles que l'on peut suivre en parallèle, voire que l'on peut construire en complémentarité. Elles méritent toutefois d'être présentes à l'esprit de celles et ceux qui mettent en place les outils d'observation des usages dans toute leur variété. Elles supposent en effet une place spécifique pour Gallica et les services qui lui sont attachés. Dans le premier cas, ceux-ci sont davantage pensés comme une *infrastructure* dont la qualité repose en grande partie sur sa fluidité, sa simplicité d'usage, sa capacité à s'intégrer sans heurts aux pratiques de tous. Étudier les usagers consiste ici à se donner les moyens de cette adaptation en explorant par exemple les usages les plus ordinaires mais aussi en s'inspirant d'usages innovant, en mettant en place des tests. Dans le second cas, Gallica et ses services sont organisés en *média* qui s'adressent à des publics. Étudier les

usages consiste alors à identifier des catégories d'usagers aux caractéristiques et aux intérêts spécifiques, que l'on cherche à atteindre, par exemple par la mise en avant de contenus éditorialisés.

Encore une fois, ces deux horizons ne sont pas exclusifs et les positions sont aujourd'hui également importantes dans les développements stratégiques de Gallica. Ils sont en revanche sous-jacents à la fois aux manières d'observer ou de questionner les usagers et aux conclusions qui peuvent être tirées. Leur clarification à ces deux moments clefs des études des usagers est importante.

### **3. État de l'art des méthodes pour appréhender les usages numériques**

L'objectif de ce volet est de présenter les grandes catégories de méthodes qui nous aident à appréhender les usages en général, et plus spécifiquement les usages numériques : elles sont le plus souvent héritées des sciences sociales, mais revisitées en profondeur par les développements du numérique. L'idée n'est pas de s'arrêter à un catalogue de méthodes, mais de proposer une démarche qui en fonction des problématiques que l'on cherche à aborder conduit à la définition du protocole méthodologique le plus adapté. Au vu de la complexité d'un objet comme le web, nous défendons la nécessité d'articuler des méthodologies complémentaires, de multiplier les points de vue pour arriver à une compréhension fine, de l'intérieur, des pratiques.

Après avoir exposé les quatre pôles de la connaissance des publics (études, audience, veille-influence et gestion de la relation client), nous insisterons sur le tournant que constitue le passage de la **connaissance** des publics à l'organisation de leur **participation** ou contribution.

#### **3.1. La connaissance des clients dans les organisations**

Les organisations sont soucieuses de connaître leurs publics (clients, utilisateurs, usagers, lecteurs, spectateurs...) et d'appréhender les usages de leurs offres, pour tester leur adéquation à la demande et pour identifier des attentes. Depuis les années 1950, toute une palette d'outils a été mise en place progressivement pour connaître les clients. Ces méthodes sont également utilisées en sciences sociales. Elles constituent le champ des études.

L'essor du numérique constitue un tournant important pour la connaissance client, qui remet en question la place des études.

Premièrement, la mesure de l'audience, autrefois réservée aux médias audiovisuels, devient une source de connaissance fine des usages. Pour tous les acteurs dont tout ou partie de l'activité s'est dématérialisée, le recours à la mesure d'audience devient crucial pour les usages numériques.

Deuxièmement, le numérique offre de nouvelles opportunités de visibilité aux consommateurs qui peuvent rendre publics leurs avis et opinions sur le web. En retour de cette activité croissante des publics, l'activité de veille sur ces opinions spontanées devient centrale. Comme toutes les opinions ne se valent pas, l'enjeu devient l'identification de ceux dont l'opinion est susceptible d'avoir une influence sur les autres, ce qui mobilise les méthodes d'analyse des réseaux.

Troisièmement, la gestion de la relation client a fortement évolué en lien avec le développement des technologies de l'information (développement du Customer Relationship Management, transformation de la relation client, développement du multicanal). Le déploiement des systèmes d'information permet d'avoir en temps réel le profil des clients, leur niveau de consommation et d'achat, de le mettre à jour à l'occasion de chaque contact, et d'avoir une connaissance actualisée des clients et de leurs usages.



Nous passerons en revue les méthodologies propres à ces quatre sources de la connaissance client : deux externes à l'organisation (études et veille), deux internes (gestion de la relation client et audience).

À chaque fois, nous illustrerons avec des études provenant d'enquêtes sur les pratiques culturelles.

---

**Synthèse :**

**Études :** enquêter auprès des publics pour connaître leurs attitudes et comportements. L'enjeu est ici la question de la représentativité de l'échantillon (approches quantitatives) ou de sa diversité (approches qualitatives).

**Audience :** analyser les traces des clients dans les espaces numériques. L'enjeu est de donner sens à une vision exhaustive des activités sur le site.

**Veille-Influence :** analyser ce que les clients écrivent et publient spontanément. Ce volet mobilise des méthodes de veille et d'analyse d'opinion, mais aussi des méthodes d'analyse de réseau. L'enjeu est d'identifier les clients les plus influents, ceux du premier cercle (*two-step flow* sur les industries culturelles).

**Gestion de la relation client :** constituer des historiques sur les clients, capitaliser sur la connaissance, profiter des contacts des clients avec l'organisation pour améliorer la connaissance client.

---

### 3.1.1. Les études

La connaissance des clients est devenue dès les années 1970, avec le développement du marketing et le ralentissement de la croissance, un enjeu central pour les entreprises. Si au début du siècle dernier Ford pouvait proclamer "*Any customer can have a car painted any color that he wants so long as it is black*", la concurrence et l'abondance de l'offre conduisent les entreprises à tenir compte des consommateurs pour innover. Dans ce contexte se développent tout un ensemble de méthodologies pour construire une connaissance client qui sert à comprendre comment les utilisateurs s'approprient l'offre, quels sont leurs niveaux de satisfaction, quelles sont leurs attentes. Ces études permettent d'orienter l'innovation et de segmenter l'offre selon les publics. Elles sont aujourd'hui encore très largement utilisées et trouvent leur pendant dans le numérique. Elles sont aussi mobilisées par les sciences sociales pour étudier les comportements et les attitudes.

Il y a une grande proximité dans les méthodes d'enquêtes (conception et analyse) utilisées par les institutions publiques (Insee, Ined, DEPS...), les cabinets d'études et les équipes de recherche qui s'explique par la proximité des métiers et par le rôle joué par l'instrument statistique dans les sciences sociales. En France, s'est développée une tradition des grandes enquêtes comme outil de pilotage des politiques publiques. De nombreuses institutions publiques se sont ainsi dotées de services des études : département des études, de la prospective et des statistiques du ministère de la Culture, service études et recherche de la BPI, délégation à la Stratégie et à la recherche de la BnF... mobilisant les équipes internes et externes avec des approches quantitatives et qualitatives.

Ces études centrées sur le consommateur ou l'utilisateur, prennent comme point de référence un individu isolé, soumis au poids de déterminants sociaux, qui se révèlent dans ses attitudes et pratiques. Sa place dans l'espace social en fonction de ses caractéristiques sociales et de son portefeuille de pratiques et de goûts devient un facteur explicatif de sa relation avec un produit, un bien, un service particulier.

La panoplie standard des méthodes comprend : l'enquête quantitative, les entretiens qualitatifs, les focus groups et l'observation ethnographique.

### *Enquêtes quantitatives*

L'enquête quantitative a longtemps été la méthode la plus valorisée parce qu'elle prétend rendre compte des comportements de l'ensemble de la population. Le processus d'échantillonnage (aléatoire à partir du recensement ou le plus souvent par la méthode des quotas qui garantit une équivalence de la répartition sur certaines variables comme le sexe, l'âge, la CSP, la taille de l'agglomération) garantit que la population enquêtée ressemble à la population réelle, et que l'on peut donc à partir des résultats sur un sous-échantillon (de un à plusieurs milliers d'individus ou de foyers) inférer des conclusions générales. Les enquêtes quantitatives peuvent également porter sur des échantillons représentatifs de sous-populations spécifiques comme les publics d'une institution culturelle donnée ou, au contraire, les non-publics. Ainsi des enquêtes sur la lecture ont porté sur des populations dont on savait qu'elles étaient exclues de cette activité (faible diplôme, zone rurale...) (Hersent, 2000).

Les enquêtes sont faites sous forme auto-administrée (par courrier ou en ligne), en face-à-face ou par téléphone. Les questions sont le plus souvent fermées, pour permettre des traitements statistiques, à choix unique ou multiples. Quelques rares questions ouvertes permettent d'avoir une vision plus approfondie d'une thématique. L'analyse des questions ouvertes est à l'origine du champ de la statistique textuelle ou Textmining (Lebart and Salem, 1994) : plutôt que de coder de manière fermée les réponses des individus, l'idée est de partir du discours des enquêtés tel qu'il s'exprime et essayer d'en tirer des régularités. Des typologies de réponses sont construites de manière inductive, en rapprochant les réponses qui tendent à utiliser le même type de vocabulaire. L'analyse des questions ouvertes se situe donc au carrefour entre méthodes quantitatives et qualitatives : elles prennent comme point de départ le discours des sujets pour construire des classes de réponses qui peuvent être quantifiées.

Les questionnaires comportent toujours un volet sociodémographique de caractérisation de l'individu considéré comme structurant de l'organisation sociale. Ces éléments « sociodémo » et les données sur les goûts et les pratiques vont guider les traitements et deviennent clef de lecture ou d'interprétation des comportements et attitudes.

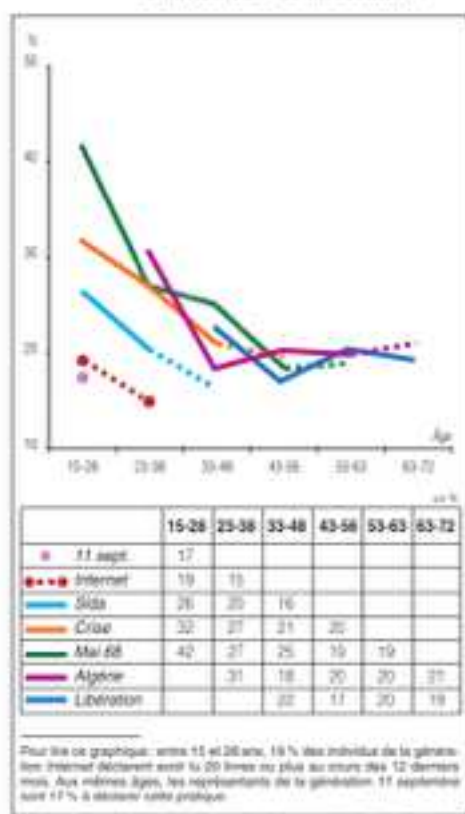
Les individus, en répondant à un questionnaire, produisent du discours sur leurs pratiques et sur les représentations qu'ils s'en font. Plus précisément, ils doivent se plier aux catégories proposées par l'enquête : la question et les types de réponses possibles, qui ne font pas forcément sens pour eux. D'où l'intérêt de faire des entretiens qualitatifs en amont pour employer un vocabulaire proche de celui des utilisateurs. Par exemple, pour la lecture, l'enquête Pratiques culturelles des Français propose comme libellé : « Combien de livres avez-vous lu dans les douze derniers mois ? » ce qui peut être interprété de manière très différente selon le rapport à la lecture.

Les enquêtes donnent lieu à des traitements statistiques : des tris à plat et des tris croisés entre variables mais aussi de l'analyse multidimensionnelle. Un type de traitement largement utilisé est la classification qui permet de segmenter la population en fonction des caractéristiques d'usage et d'éléments sociodémographiques. C'est par exemple le cas de l'enquête GMV sur les usages de Gallica qui segmente la population en fonction des usages (GMV, 2012). Ces typologies de type inductif, rapprochant les individus ayant des comportements proches, sont très répandues en France : le courant de l'analyse des données initié par J-P Benzécri (Benzécri, 1968) a connu un grand succès dans le secteur des études mais surtout en sciences économiques et sociales. Bourdieu en a par exemple été un grand utilisateur (Bourdieu, 1979) au point que l'on peut avancer que sa théorie de l'organisation de l'espace social dans un espace à deux dimensions (capital total, composition relative

du capital économique et du capital culturel) est marquée par les méthodes de l'analyse factorielle. L'approche anglo-saxonne privilégie une approche hypothético-déductive qui conduit à tester des hypothèses avec des modèles de type régression logistique.

Les enquêtes quantitatives révèlent les écarts dans les comportements et les attitudes en montrant le poids des déterminants sociodémographiques, des dispositions esthétiques et de la familiarité avec la pratique. Elles permettent aussi de suivre dans le temps les évolutions des pratiques, à condition de mettre en place des méthodes adaptées. Dans ce domaine, l'enquête Pratiques culturelles des Français constitue un exemple tout à fait unique. Elle a eu lieu pour la première fois en 1973 à l'initiative d'Augustin Girard et a été reproduite en 1981, 1989, 1997 et 2008 (Donnat, 2009). Cette enquête menée par le département des études, de la prospective et des statistiques du ministère de la Culture est centrée sur les pratiques culturelles au sens large (les loisirs) et vise à évaluer l'efficacité des politiques culturelles. Le fait que l'enquête ait gardé un noyau constant de questions sur toutes les vagues d'enquête a permis de reconstruire une étude longitudinale sur l'évolution des pratiques (Donnat, 2011) et de distinguer les effets d'âge et les effets de génération. Ainsi l'on voit dans le graphique suivant que la pratique de la lecture diminue avec l'âge (effet âge) mais que plus les générations sont récentes, moins elles lisent (effet génération). Chaque nouvelle génération lit moins que la précédente, et lira de moins en moins au cours de sa vie, ce qui ne permet pas d'envisager un avenir radieux pour la lecture.

Graphique 6 – Courbes générationnelles de la pratique « Lecture de livres »



Les limites de ces enquêtes sont évidentes : elles imposent à l'enquêté un cadre d'interprétation du réel prédéfini qui ne correspond pas forcément au sien, et il y a un décalage évident et avéré entre ce qui est déclaré et ce qui est réellement fait.

L'enquête quantitative est en partie inadaptée pour les usages du numérique pour les raisons suivantes :

- Les usages numériques sont en reconfiguration permanente (en lien avec le rythme accéléré des innovations) et il est difficile de construire des catégories avec une certaine stabilité pour appréhender les pratiques.
- Ils se traduisent par une diversité très grande des niveaux d'engagement, que montrent des distributions de type exponentiel : on a une grande masse de visiteurs qui ne laissent pas de traces et parmi ceux qui produisent (textes, notes, commentaires), une distribution de même nature : quelques individus très actifs et beaucoup faiblement actifs, voire inactifs. Ceci pose un problème pour l'échantillonnage mais aussi pour les traitements qui tendent à écraser cette diversité.
- Globalement, l'échantillonnage pour étudier les usages d'une plateforme web n'est pas simple. Les enquêtes en ligne tendent systématiquement à favoriser les gros utilisateurs, au détriment des faibles utilisateurs.
- Les usages à l'écran sont entrelacés. Une grande partie des activités se déploient sur un même écran (ordinateur, tablette ou smartphone) favorisant leur entrelacement (une séance de consultation de Gallica peut être entrecoupée de lecture de mails, d'échanges par skype, de consultation d'autres sites). Il est très difficile alors pour le sujet de déconstruire l'ensemble des activités et d'avoir une appréhension du temps passé sur chaque activité.
- Le temps passé devant les écrans ne nous informe en rien du type d'activité mené en arrière-plan. Il faut donc réinventer d'autres approches des usages.

L'utilisation d'un carnet à remplir constitue un complément intéressant à l'enquête pour appréhender les usages d'une manière plus fine, mais il devient contraignant quand les accès sont nombreux et divers. La première grande enquête sur la sociabilité, l'enquête Contact de l'Insee (Héran, 1988), s'appuyait sur un carnet où les individus rendaient compte de toutes les personnes qu'ils rencontraient et avec qui ils avaient un échange non professionnel pendant une durée d'une semaine. On voit bien qu'il serait extrêmement complexe aujourd'hui, pour étudier la sociabilité, dont une part importante se déroule sur les réseaux, de recourir à un carnet.

### *Enquêtes qualitatives*

Les **entretiens qualitatifs** constituent une démarche plus ouverte et compréhensive pour aborder pratiques et représentations. Alors que dans les enquêtes quantitatives le millier d'individus interrogés constitue un minimum, dans les enquêtes qualitatives le nombre d'entretiens réalisés sera le plus souvent inférieur à une cinquantaine d'entretiens. Il ne s'agit pas d'être représentatif, mais d'essayer de constituer un échantillon qui rende compte de la diversité des profils. L'entretien est mené par un chercheur qui s'appuie sur une grille d'entretien, assez ouverte (entretien semi-directif). Décrire les pratiques mais surtout en comprendre la logique et la signification pour le sujet sont au cœur de la méthode.

Les **focus group** sont utilisés pour tester une offre ou un prototype d'offre auprès d'un groupe d'utilisateurs. Il s'agit d'un groupe constitué d'une dizaine ou quinzaine de personnes, avec un animateur qui suit une grille de déroulement. C'est une méthode très économique en temps et qui permet pleinement de bénéficier de la dynamique collective en termes de créativité mais elle est très sensible aux effets de dynamique de groupe : la préférence pour le consensus tend à écraser les points de vue minoritaires ; des hiérarchies et spécialisation de rôles se mettent très rapidement en place (avec apparition de postures de leadership) qui tendent à se renforcer tout au long du déroulement.

L'**observation ethnographique** de l'activité constitue un volet complémentaire puisqu'il vise à observer l'activité pendant son déroulement, quitte à réserver un temps ultérieur à l'explicitation, via des

entretiens d'auto-confrontation. L'observation donne accès aux comportements, aux éléments de langage en situation naturelle, et non à du discours sur l'activité. L'approche ethnographique a été profondément renouvelée grâce aux outils du numérique, ce qui nous amène à lui consacrer un paragraphe à part.

### *L'ethnographie numérique, analyse de l'activité*

Comme nous l'avons indiqué, l'activité dans le numérique est difficile à décrire pour les sujets : comment décrire ce que l'on fait sur son ordinateur, sur son smartphone ? Ceci représente une limite importante pour les enquêtes quantitatives et qualitatives. Quels moyens pour accéder aux usages effectifs des utilisateurs ? Nous verrons que l'analyse des traces peut être une ressource, mais ces traces numériques ne nous permettent pas de reconstituer le cours de l'action, tel qu'il s'est déroulé, son insertion dans un contexte plus global et surtout la signification que lui accorde le sujet.

C'est dans cette direction que se développe un champ de recherche autour de **l'analyse de l'activité qui mobilise la vidéo** comme outil de captation de l'activité sur les écrans. Ce courant a bénéficié de la miniaturisation et simplification des outils de captation, montage et analyse vidéo.

L'analyse de l'activité est marquée par trois influences : l'ethnométhodologie et l'analyse de conversation, la cognition située et l'analyse de l'activité. (Licoppe, 2008) rend compte des différentes perspectives mobilisées pour analyser le travail.

On peut globalement distinguer **l'analyse de l'activité en situation naturelle de l'analyse de l'activité en laboratoire**. Dans le premier cas, le sujet n'est pas contraint par le chercheur dans son cours d'action, dans le second, il suit un protocole proposé par le chercheur. Voici quelques outils et méthodes utilisées pour faire de l'ethnographie.

La SenseCam est une caméra développée par Microsoft que le sujet porte à son cou et qui prend à intervalle régulier des photos. On peut avoir ainsi le film de la journée d'un individu avec des traces (carottage) de son activité par des plans fixes. L'exploration de ces bandes permet de repérer par exemple les moments de lecture, les moments de navigation sur le web et de travailler ensuite sur des sous-corpus spécifiques.

Sebe (Subjective Evidence-Based Ethnography) est un dispositif méthodologique mis au point par (Lahlou, 2011) pour analyser l'activité des sujets (le comportement et la signification qui lui est associée). Il part de deux constats : ce que décrit l'individu a peu à voir avec ce qu'il fait vraiment, d'où la nécessité de filmer l'activité ; une vue objective et externe sur l'activité ne donne pas du tout le même résultat qu'une vue subjective. Pour comprendre du *point de vue du sujet* le sens de l'activité, il faut un dispositif qui adopte son point de vue. Le protocole est en deux temps : dans un premier temps, le sujet est amené à porter au niveau des yeux une caméra (subcam, posée sur lunettes) qui enregistre toute l'activité. Dans un second temps, il revoit la bande avec l'enquêteur et cherche à reconstituer ses buts et ses motifs (théorie de l'activité russe). L'enregistrement vidéo en première personne est une ressource fondamentale pour se reprojeter dans la situation passée (accès unique à la mémoire épisodique). Ce dispositif est particulièrement adapté pour l'analyse de l'activité en situation naturelle.

Sur les usages à l'écran, certains travaux combinent plusieurs points de vue sur l'activité : captation vidéo de ce qui se passe à l'écran, enregistrement vidéo du point de vue du sujet et vidéo filmant l'ensemble du contexte (Velkovska and Zouinar, 2007 ; Licoppe, 2006). Les différents flux sont ensuite synchronisés et analysés.

Un atout de la vidéo ethnographie est que très rapidement le sujet oublie la présence du dispositif de captation et il est moins dérangé dans son activité que si un chercheur était en permanence à ses côtés.

Le Laboratoire ErgoDesign LUTIN-GOBELINS, qui prend la suite du laboratoire Lutin propose un ensemble de dispositifs pour analyser l'activité sur les écrans : des oculomètres, des EEG (électroencéphalographie), des caméras subjectives avec oculométrie. Cet environnement est particulièrement adapté pour des expériences en laboratoire. Des sujets viennent tester un site web, feuilleter une tablette. L'oculomètre mesure le parcours des regards et le résultat est une carte de chaleur qui montre les endroits les plus consultés, les zones ignorées... Cela permet de reconfigurer la présentation des contenus, la structuration d'un site web. Les sujets peuvent avoir des tâches à accomplir, comme trouver telle ou telle information.

*Focus : les enquêtes sur les usages et utilisateurs des bibliothèques*

Voyons à présent les dispositifs mis en place par les bibliothèques pour étudier leurs usages.

Carol Tenopir a fait en 2003 une recension assez exhaustive des travaux menés dans le contexte anglo-saxon sur les usages des bibliothèques numériques (Tenopir, 2003). Elle a passé en revue 200 travaux menés sur les usages des bibliothèques numériques (principalement des bibliothèques universitaires et quelques sites de revues en ligne), en accordant une place centrale à une dizaine d'études qu'elle considère de plus grande ampleur.

Les méthodes les plus utilisées sont finalement assez traditionnelles : enquêtes en ligne auprès des publics, entretiens, focus groups, tests d'utilisabilité et enfin analyse des logs de connexion (*cf.* section suivante Audience). Quelques études combinent ces différentes approches, ce qui les rend immédiatement plus intéressantes. Certains travaux, comme ceux menés sur Jstor et SuperJournal, sont vraiment centrés sur l'usage des documents et descendent donc à un grain très fin : celui de l'article consulté. Ces travaux permettent d'étudier finement quels documents sont téléchargés par qui et d'identifier des stratégies de consultation différentes. Par exemple, l'étude SuperJournal conduit à une typologie des utilisateurs en 7 groupes et montre des différences dans la manière d'accéder aux articles entre scientifiques et chercheurs en sciences humaines. L'étude Jstor a permis d'évaluer la consultation des articles en ligne sur JSTOR par rapport à la consultation des exemplaires papier dans les mêmes bibliothèques, montrant que la version numérique accroît sensiblement la visibilité des articles. Elle a aussi permis d'évaluer ce qui est fait avec les articles (lecture, impression, téléchargement, etc.).

Il existe sans doute d'autres travaux plus récents, mais dont les résultats ne sont pas publiés en ligne.

Une étude menée par la BnF mérite d'être rappelée car elle présente un dispositif méthodologique très intéressant. Il s'agit de [l'enquête BibUsages](#), menée en 2002, entre la BnF et France Télécom R&D (Assadi and Beauvisage, 2003). Le dispositif méthodologique était le suivant : une enquête en ligne avec un recrutement sur le site de Gallica, auquel ont répondu 2340 personnes sur leurs usages de Gallica. Parmi eux, 72 ont été équipés d'une sonde pour recueillir tout leur trafic sur le web et en particulier les parcours sur Gallica et d'autres sites liés à la lecture. Leur activité a été enregistrée pendant 6 mois. Des entretiens qualitatifs ont été menés auprès de 16 panélistes, avec en support leurs statistiques de trafic et leur réponse au questionnaire. Ce dispositif a permis de qualifier le profil des utilisateurs (1/3 hors de France), surreprésentation des plus de 45 ans et des hommes. L'étude montrait que près de la moitié avait une démarche de chercheur amateur (souvent liée à un intérêt pour la généalogie). L'analyse du trafic montrait la très inégale activité des internautes (un tiers des visiteurs réalisent les trois-quarts des visites). L'étude montrait aussi que les sessions avec un accès à Gallica étaient plus longues que les autres (1h au lieu de 28 minutes) et surtout que la moitié de ces sessions étaient exclusivement dédiées à Gallica. Gallica favorisait des usages concentrés sur le site, plutôt que la dispersion et la mise en concurrence avec des sites externes.



Un enseignement principal peut être tiré de ces différents travaux : pour avoir une approche fine et pertinente des usages du numérique, il est essentiel de [croiser et d'articuler des méthodes complémentaires](#) :

- Croiser un regard interne (explorer ce qui se passe sur le site Gallica et sur les espaces de communication contrôlés par l'institution) et externe (situer les usages de Gallica par rapport aux usages des autres sites).
- Adopter un point de vue centré sur l'utilisateur qui permette de retracer tous les lieux de présence et d'activité, ce qui implique analyse de traces, enquêtes, carnet, entretien, ethnographie. Essayer de reconstituer la notion d'utilisateur en articulant les usages de la bibliothèque en ligne et de la bibliothèque physique.
- Compléter par un point de vue centré sur l'usage : *i.e.* mettre en place des dispositifs pour analyser finement la consultation et l'usage des documents en ligne et sur place.

### 3.1.2. L'audience

Pour les organisations – entreprises ou institutions – qui ont vu une grande partie de leur activité se reporter sur les espaces numériques – sur les sites web –, de nouveaux indicateurs de l'activité des publics sont apparus. En même temps qu'il se construit, le web construit les affordances qui permettent d'enregistrer les traces d'usages. Très vite, les serveurs enregistrent les logs d'activité et peu à peu des outils plus sophistiqués avec marquage de page (comme Google Analytics ou XiTi) permettent d'analyser plus finement les usages du web. En parallèle, des sondes permettent d'écouter ce qui circule sur les réseaux ou sur le poste d'un utilisateur. Le domaine de la métrique du web se déploie.

Pour l'étude de l'audience de site, on distingue deux grands types d'approches : l'approche « site centric » et l'approche « user centric ». On pourra se reporter au chapitre de Beauvisage dans le Manuel d'analyse du web de (Barats, 2013).

L'approche « site centric » consiste à recueillir et analyser l'activité des visiteurs sur un site particulier. Chaque institution peut ainsi analyser la fréquentation de son propre site. Deux indicateurs centraux sont le nombre de pages vues et le nombre de visites. Ces informations donnent une vision exhaustive du trafic sur le site : nombre des pages vues, de visites, de temps passé sur le site, de son origine (localisation des adresses IP), des sources qui ont conduit le visiteur sur le site... Deux grands types de méthodes sont utilisés : soit l'analyse des logs de serveurs qui enregistrent toutes les requêtes faites au serveur, soit l'analyse des requêtes sur des pages qui ont reçu un marqueur ou un tag. Ces traces d'usage permettent de suivre l'évolution du trafic, de mesurer l'impact du redesign d'un site et de reconfigurer le site en fonction des audiences. La limite principale de cette approche est qu'elle ne parvient pas à reconstituer une notion d'utilisateur, et donc de visiteur unique, et quand bien même elle le pourrait, on n'a aucune information sur les utilisateurs eux-mêmes. Ces traces constituent de grands réservoirs de données, dont il n'est pas aisé de tirer des enseignements précis sur les utilisateurs. Une autre limite de cette approche est qu'elle ne permet pas de situer l'usage du site par rapport aux autres et en particulier par rapport aux sites concurrents.

L'approche « user centric » est comme son nom l'indique centrée sur l'utilisateur dont elle étudie l'ensemble de l'activité sur le réseau. Il s'agit en général de panels d'utilisateurs, représentatifs de la population connectée à internet. Cette approche est essentiellement développée par les sociétés de mesure d'audience (comme Médiamétrie-Nielsen-Netratings ou ComScore, qui sont les deux acteurs majeurs) sur un modèle initialement proche de celui de la mesure de l'audimat pour la télévision. La qualité des échantillons est garantie par les enjeux économiques qui y sont liés (la tarification

publicitaire en dépend). L'idée est de pouvoir inférer la fréquentation globale du web à partir de données provenant d'un échantillon. Dans cette approche, les individus du panel sont équipés d'une sonde qui enregistre toute leur activité et l'envoie au serveur. Mais la mesure peut se faire directement sur les réseaux (sans passer par les terminaux) comme c'est le cas pour les panels internet mobile. Les atouts de cette approche sont les suivants : il s'agit d'individus, qui sont décrits et identifiés à chacune de leur session internet (soit par une fenêtre d'identification, soit par des algorithmes d'apprentissage qui apprennent à différencier les utilisateurs d'une même machine). On appréhende l'ensemble de leurs usages ou des sous-ensembles définis (par exemple tous les sites liés à des pratiques culturelles), et on peut situer la place d'un usage particulier. On peut évaluer le temps de connexion de manière fine, le nombre de pages vues, les domaines visités (Beaudouin and Licoppe, 2002). On peut également appréhender les parcours des utilisateurs (Beauvisage, 2004) et identifier des types d'utilisateurs selon leur profil de navigation. Pour la recherche, ce type d'approche nécessite des partenariats avec les sociétés de mesure d'audience, soit pour faire du traitement secondaire sur les données de leur panel, soit pour constituer des panels ad hoc. C'était le cas des projets TypWeb et SensNet portés par France Télécom R&D en partenariat avec NetValue, puis Nielsen-Netratings et des équipes académiques (Beaudouin and Licoppe, 2002). Le projet BibUsages entre la BnF et France Télécom R&D (Assadi and Beauvisage, 2003) reprenait une démarche similaire en utilisant une sonde « maison » et en construisant de bout en bout l'échantillon.

Ces approches sont très sensibles aux vagues d'innovation propres au numérique : la multiplication des équipements qui permettent de donner accès à internet rend la mesure plus complexe : il faut des sondes pour ordinateur, tablette, smartphone et la capacité à reconstituer à partir de ces données hétérogènes une vision intégrée des usages. Ces méthodes indiquent qu'il y a eu accès à une page mais ne nous informent pas d'usages plus précis : comme le temps de consultation de vidéo. Ces méthodes ont du mal à rendre compte des usages des réseaux sociaux.

Une autre limite des approches « user centric » est que la taille des échantillons ne permet pas d'avoir des données fiables pour les sites web qui ne sont pas dans le haut des classements. Pour remédier à ce problème, de plus en plus de sociétés de mesure d'audience adoptent des démarches hybrides en redressant les données centrées utilisateurs par des données centrées sur les sites.

Pour finir, et à titre de transition, on peut citer le modèle Alexa qui se situe dans le domaine de la mesure d'audience, du côté de l'approche « user centric », mais sans la rigueur de constitution et de caractérisation du panel. Alexa a évolué ces dernières années, mais son atout résidait dans sa capacité à classer les sites en fonction de leur audience mais aussi en fonction du nombre de liens entrants. Il propose une approche hybride entre mesure d'audience et analyse de réseaux. Alexa s'appuie sur un panel d'individus qui acceptent d'installer une barre d'outils qui transmet des informations de navigation. Les données sont croisées avec des données de réseaux de fournisseurs d'accès. Alexa cherche à obtenir le plus d'informations sur son panel pour pouvoir qualifier plus précisément les usages. Les données d'Alexa permettent en l'absence de données de mesure d'audience d'avoir une estimation de l'audience et de la réputation d'un site. Sur une étude portant sur les écrivains sur le web, il nous a permis de voir qu'il n'y avait pas de corrélation entre l'audience et le nombre de liens entrants, que l'on pouvait avoir des sites très interconnectés avec une très faible audience.

### 3.1.3. Du réseau à l'influence

Avec le web, on assiste à un tournant important : plutôt que de regarder les utilisateurs comme des atomes indépendants, ils sont de plus en plus analysés à travers leurs réseaux de relation. La nature réticulaire de l'internet renforce évidemment la place que jouent les réseaux dans les analyses.



### *Analyse des réseaux*

Les travaux sur les réseaux sociaux remontent au début du XX<sup>e</sup> siècle. Moreno met en place une technique pour étudier la circulation des informations dans une institution de jeunes filles à Hudson. Il recueille auprès de chaque résidente la nature des relations avec toutes les autres : relation d'attraction, de répulsion ou d'indifférence. Cela lui permet de tracer un sociogramme des relations, un premier graphe de relations, pour comprendre la nature de la circulation et [diffusion de l'information](#) entre les différents pavillons de jeunes filles (Moreno, 1934).

Dans le domaine des industries culturelles, les travaux de (Katz and Lazarsfeld, 1955) tentent de montrer les [mécanismes d'influence](#) dans la réception des médias. Loin d'un modèle où les médias influenceraient directement les individus, ils montrent que le réseau de pairs joue un rôle important. Ils posent qu'il existe des [influenceurs](#) qui jouent un rôle de médiation entre les médias et les récepteurs. Ils introduisent le concept de [two-step flow](#) pour désigner ces mécanismes d'influence en deux étapes, avec un rôle joué par ces leaders d'opinion.

Les travaux de Harrison White et de Granovetter ont également cherché à comprendre la nature de la [diffusion des idées et opinions dans un réseau](#). Granovetter, dans son fameux article « La force des liens faibles » (Granovetter, 1973), rend compte d'un débat important sur la diffusion des innovations dans les années 60 et qui anime les recherches contemporaines. Rogers considère que les premiers individus à adopter une innovation sont les « marginaux » (ceux qui sont rarement choisis comme « amis ») (Rogers, 1962). Coleman, Katz et Menzel, en revanche considèrent que ce sont les individus « centraux » (les plus souvent cités) qui adoptent l'innovation avant les autres (Coleman et al., 1966). Pour Becker, l'introduction de la notion de risque perçu permet de résoudre le conflit : quand l'innovation est perçue comme risquée, l'adoption est faite par les marginaux, quand elle est non risquée elle est adoptée par les individus centraux (Becker, 1970). Ces débats se retrouvent actualisés aujourd'hui quand on cherche à identifier les influenceurs sur le web.

Le web est par sa construction même un réseau qui se prête aisément à des analyses de réseau ([social network analysis](#)), une fois que l'on sait capter les nœuds (les sites par exemple) et les liens entre nœuds (liens hypertexte). Le développement du web et des méthodes d'analyse de graphes ont permis un renouvellement de ces approches en réseaux, donnant naissance à ce que certains peuvent appeler une nouvelle science des réseaux (Watts, 2004).

Les travaux sur les graphes du web ont montré qu'ils ont des propriétés non triviales, par rapport aux graphes aléatoires : ils sont organisés dans une logique « petit monde » (distance moyenne et diamètres faibles), la distribution des degrés des sommets (nombre de liens liés à nœud) suit une loi de puissance, la densité globale est faible, alors que la densité locale est forte (Cristofoli, 2008)

Par-delà l'analyse de la [topographie du web](#), un axe de recherche important porte sur la [diffusion des documents, idées, opinions et sur le rôle des influenceurs](#) (comment identifie-t-on les influenceurs sur le web, quelle est leur capacité d'influence ?). Si certains travaux remettent en question leur rôle (Watts and Dodds, 2007), globalement domine l'idée que certains individus doivent être identifiés car ils ont une capacité d'influence plus importante que d'autres. Les notions d'autorité (individus très cités) et de centralité sont reprises dans nombre de ces travaux.

Des outils ont été développés pour permettre au chercheur en sciences sociales de construire pas à pas un réseau de sites et pour visualiser les relations entre sites. Navicrawler, développé par Matthieu Jacomy, permet d'explorer de proche en proche un réseau de sites en conservant uniquement ceux qui sont pertinents. Des outils de visualisation de graphes (Gephi ou Pajek) permettent ensuite de représenter les liens entre les sites, de repérer les sites qui ont une centralité importante et une position

d'autorité par un grand nombre de liens entrants. Ces méthodes ont été utilisées dans de nombreux projets : dans e-diaspora pour cartographier les sites web liés aux diasporas <http://www.e-diasporas.fr/>, dans le projet Anamia pour identifier les sites relatifs aux troubles alimentaires <http://www.anamia.fr/>, dans le projet Panic pour identifier la structure du réseau des écrivains actifs sur le web. Ils sont utilisés pour cartographier des controverses, des communautés de pratiques...

### *Veille et influence*

Le web a offert dès les origines des opportunités de publication tout à fait inédites pour le grand public. Sur les forums, les chats, les pages personnelles, les internautes pouvaient exprimer leur avis ou engager des conversations sur tout type de sujet. Le web 2 s'est accompagné d'un véritable passage à l'échelle en rendant plus simple encore l'autopublication (sur les blogs, les réseaux sociaux) mais surtout en proposant des plateformes dédiées à l'évaluation et à la recommandation (avis des lecteurs sur Amazon, sur les sites de biens culturels, sur les sites de vente en ligne). Tout un chacun est invité à donner son avis sous forme de note et de commentaire.

Capter et évaluer ces opinions spontanées est devenu un enjeu fort pour les organisations. Elles y découvrent ce que les consommateurs disent spontanément avec leurs mots sur leurs produits et services mais aussi sur leur marque.

Dans le secteur de la recherche, de nombreux travaux portent sur l'analyse des conversations sur le web. Des méthodes ont été mises en place pour analyser les forums de discussion (Akrich, 2012; Beaudouin et al., 2000), pour analyser les échanges sur Twitter (Cha et al., 2010; Java et al., 2007; Kwak et al., 2010). Ces approches combinent une analyse des hiérarchies des individus dans ces collectifs numériques avec une analyse des contenus.

Un nouveau champ de recherche s'est développé au début des années 2000 pour extraire des opinions à partir de ces conversations ou publications. C'est le domaine de la fouille d'opinion ([opinion mining ou sentiment analysis](#)) qui se développe comme un nouveau champ de recherche dont le domaine principal d'application est l'analyse des avis de consommateurs sur le web. Dans ce champ, l'opinion est modélisée comme un sentiment (positif, négatif ou neutre) porté par un individu sur une caractéristique (« aspect » ou « *feature* ») d'une entité (produit ou service) à une date donnée (Liu, 2012). Dans la pratique, différents types de méthodes (utilisant de l'apprentissage supervisé ou des lexiques marqués) sont utilisées pour analyser ces corpus d'opinion et en extraire une synthèse quantitative. On pourra se reporter à (Liu, 2012) pour une description précise des méthodes et à (Boullier and Lohard, 2012) pour une vision critique de ce champ.

Des sociétés se sont également spécialisées dans la veille sur ces types de contenus, comme en France : Synthesio, Digimind, Linkfluence, Temis, Lingway... Ces sociétés combinent des compétences en traitement automatique des langues mais aussi en analyse de réseaux. En effet, la plupart (comme Synthesio ou Linkfluence) partent du constat que toutes les opinions ne se valent pas, et procèdent à des formes d'échantillonnage d'un genre nouveau : l'idée va être de [sélectionner les opinions qui ont une capacité d'influence](#) ou qui ont un score d'influence supérieur à un seuil, et pour faire cela, ce sont des indicateurs relationnels (nombre de liens entrants, de followers...) qui vont être mobilisés.

Les nouvelles approches qui écoutent le web, qui analysent le discours spontané sur le web, abandonnent cette idée de représentativité fondatrice des enquêtes. Ce qui compte c'est de repérer les individus influents, ceux qui ont une structure de réseau adéquate (autrement dit, dans le langage des réseaux, des personnes qui ont une position de centralité et d'autorité). Une prise de parole, quelle que soit sa pertinence n'a de valeur que si elle est entendue. [Une hiérarchisation de l'espace des opinions sur le web transparait, qui définit les opinions « qui comptent »](#) (Kotras, 2013). Le développement du

web et plus précisément du web participatif transforme la manière de regarder le client. L'organisation découvre le pouvoir d'agir du client, de certains clients, et son pouvoir d'influence. Il importe moins d'avoir une vision représentative des consommateurs que d'identifier les postures et discours d'acteurs insérés dans des réseaux ayant une capacité d'agir sur d'autres. Et ceci représente un tournant majeur.

#### 3.1.4. La gestion de la relation client

Ces vingt dernières années les systèmes d'information et en particulier les espaces qui concernent la relation client ont connu une très forte évolution. L'entreprise investit dans le développement des points de contacts entre elle-même et ses clients, en développant l'accessibilité (extension des tranches horaires, multiplication des canaux de contact...), la flexibilité (les changements, les actions... sont pris en compte en temps réel), en enrichissant les services (Flichy & Zarifian, 2002). Chacun des contacts clients est une opportunité pour acquérir de la connaissance sur les clients. Ces **remontées clients sont utilisées pour l'amélioration des services et des processus**. Les systèmes de CRM en centre d'appel comme en boutique sont des lieux de captation des usages, des problèmes... qui rétroagissent sur l'offre. L'articulation entre centres d'appel et processus d'innovation est centrale.

Au cœur de la gestion de la relation client, il y a la fiche client, censée rassembler toutes les données sur le client, les éléments sur son identité, sur son profil client, son historique d'achat des produits et services choisis et ses usages.

Dans les évolutions actuelles, les efforts sont déployés pour **faire converger les systèmes d'information** et pour parvenir à construire une **vision unifiée du client**, centrée sur l'individu et pas sur l'offre qu'il a prise (la plupart des organisations sont organisées en fonction d'unités d'affaire déterminées par le type d'offre et non pas par le client : créer des ponts entre unités d'affaire pour arriver à une vision convergente est un enjeu fort pour beaucoup d'organisations).

L'exploitation des données issues du CRM qui correspond au champ de la Business Intelligence connaît un essor très vif. Sa limite principale tient au fait qu'elle ne travaille que sur les clients et ne prend pas en compte l'univers concurrentiel. Elle est en mesure d'évaluer l'arrivée de nouveaux clients, la disparition d'autres, mais elle reste aveugle à l'environnement global.

Quand la valeur vient de la connaissance client toute l'activité de l'entreprise est organisée autour de l'acquisition de cette connaissance comme dans beaucoup d'entreprises du web. Les dispositifs se multiplient pour capter en interne les avis et les opinions. L'internalisation des systèmes d'évaluation, permet le déploiement des systèmes de recommandations dont l'efficacité pour la vente est avérée.

Penser la stratégie du système d'information comme un lieu de captation de la connaissance et de la valeur client est une orientation que les entreprises de services, qui reposent largement sur le modèle publicitaire, ont parfaitement intégrée.

### 3.2. La participation des publics

Les approches présentées dans la section précédente correspondent à un même paradigme : les utilisateurs, consommateurs, clients ou publics sont considérés comme des entités externes à l'organisation, entreprise, institution, qui interviennent au bout de la chaîne de production, mais n'interviennent pas dans le processus de production.

Il existe une autre manière de penser le public, non pas comme simple consommateur ou utilisateur, récepteur passif à la fin de la chaîne de production, mais comme partie prenante de la production. Le consommateur devient un agent actif, contributif qui participe à la production. Le concept d'audience active force à repenser la hiérarchie entre organisation et public (Auray and Moreau, 2012; Beaudouin,

2011) en montrant le rôle que jouent les publics dans la création (pro'am) et dans la promotion (Bastard et al., 2012). De nombreux travaux ont tenté de montrer le rôle que jouent les évaluations profanes. Dominique Pasquier en fait une synthèse complète dans un récent numéro de la revue *Réseaux* (Pasquier, 2014a).

Ceci n'est pas nouveau, dès les premiers travaux sur la relation de service, on a observé que la production de services se distingue de la production des biens industriels entre autre par la participation nécessaire du client. Ainsi, Fuchs en 1968, comparant le secteur des services au secteur industriel, identifie dans le premier « l'importance du consommateur comme agent coopératif dans le processus de production » (Fuchs and Leveson, 1968). La notion de *coproduction de service* trouve ici une de ses origines et souligne la contribution du client à l'accomplissement du service dans une perspective instrumentale.

Du côté des organisations, on a pu observer ces vingt dernières années des tentatives d'ouverture dans les processus d'innovation. L'idée de base est que l'utilisateur final pourrait être bien placé pour définir des produits et services adaptés à ses besoins.

Deux tendances en résultent. La première consiste à faire « remonter » le consommateur dans l'organisation, autrement dit à le faire participer au cycle d'innovation. Ainsi au lieu d'interroger le client une fois le service fini, on l'associe en amont. Le client est censé influencer, orienter le processus d'innovation tout au long de son déroulement, de la période de foisonnement créatif, jusqu'au test du service en passant par le maquetage. Il existe de nombreux modèles de co-innovation ou d'open innovation : ceux qui consistent à favoriser les idées de personnes provenant de l'extérieur ou de l'intérieur de l'entreprise (organisation de concours à idées), ceux qui font travailler des personnes extérieures à l'entreprise sur de nouveaux concepts (barcamp, akhaton) et des modèles plus classiques où des clients sont sollicités à différentes étapes du processus.

La seconde tendance consiste à mener une veille active sur les innovations telles qu'elles sont produites par les utilisateurs eux-mêmes en dehors des organisations. Ceci repose sur l'idée que les utilisateurs sont les mieux à même de définir les services dont ils ont besoin. C'est le concept d'innovation horizontale, introduit par (Von Hippel, 2005, 2002). Des réseaux d'innovation (production, distribution et consommation) peuvent se construire de manière horizontale avec des acteurs de l'innovation qui sont de simples utilisateurs. C'est le cas du logiciel libre. Selon lui, les réseaux d'innovation par les utilisateurs peuvent fonctionner de manière complètement autonome par rapport aux entreprises quand les conditions suivantes sont réunies : 1) un noyau d'utilisateurs a une forte incitation individuelle pour innover, 2) une partie des utilisateurs sont prêts à partager, révéler leur innovation, 3) la diffusion de l'innovation portées par les utilisateurs est peu coûteuse (concurrence offres payantes) (Von Hippel, 2005). En innovant dans des réseaux, les utilisateurs fabriquent des innovations parfaitement adaptées à leurs besoins, ce qui accroît leur satisfaction. Plus généralement, de nombreuses idées de start-up naissent de l'envie de résoudre un problème rencontré par les concepteurs. Le modèle d'open innovation consiste à éviter le NIH (Not Invented Here) et à scruter et mener une veille dans l'environnement pour repérer les projets et entreprises émergentes à potentiel. L'activité d'innovation est externalisée, confiée à des utilisateurs, qui prennent en charge les risques de l'échec. L'activité de l'entreprise consiste à être à l'écoute et être prête à racheter et réinternaliser les productions externes. C'est ainsi que de nombreux services créés de manière bénévole et coopérative ont été rachetés par les grands acteurs du web (par exemple IMDB racheté par Yahoo).

Les questions qui se posent aux organisations sont les suivantes : comment organiser une veille efficace pour repérer et évaluer le potentiel des idées qui émergent, comment encourager et stimuler la production d'idées ?

Dans l'économie du numérique, il existe un modèle d'organisation qui met au cœur de sa stratégie la participation des publics : les plateformes participatives, comme Allociné, Babelio... Elles fonctionnent toutes sur le même modèle : un très faible nombre de salariés, un premier cercle de contributeurs, qui représente une faible proportion du public global, et une masse de visiteurs. Dans le cercle des contributeurs, on observe des modes d'engagements très variables avec un faible noyau très actif (Beaudouin and Pasquier, 2014). Les motivations des contributeurs ont fait l'objet de nombreux travaux : recevoir de l'attention, montrer sa compétence, se situer par rapport aux autres dans une hiérarchie de compétences, influencer les autres (Pasquier, 2014b).

Les organisations proposent des systèmes de gratification et des signes de reconnaissance qui contribuent à la motivation des contributeurs : en instaurant un Club 300, Allociné dessine un horizon d'attente pour les plus gros contributeurs, Babelio propose des labels qui apparaissent sur la fiche profil des individus.

Autrement dit, il y a un potentiel de participation des publics qui est tout à fait impressionnant, surtout pour les pratiques culturelles. Pour être durable, la participation des publics ne doit pas être perçue par les sujets comme une forme d'exploitation du travail (thème à la mode du Digital Labour). Comment proposer des dispositifs de participation auxquels les individus sont satisfaits de contribuer ? Quels systèmes de reconnaissance proposer pour que l'engagement bénévole ne soit pas perçu comme une forme moderne de l'exploitation ? Ce n'est qu'à cette condition que le crowdsourcing peut espérer avoir de belles heures devant lui.

Les grandes bibliothèques ne se sont pas encore pleinement engagées dans la logique de la participation. L'activité sur les réseaux sociaux permet de construire des échanges avec le public, de capter et de rendre visibles certaines personnes du public, mais on n'assiste pas à une présence visible et massive du public dans Gallica par exemple. Quelques pistes sont déjà envisagées (correction de l'OCR), mais d'autres expérimentations pourraient être imaginées : évaluation de la qualité des documents (lisibilité, intérêt, qualité de l'édition, participation à la politique de numérisation...).

En bref, il ne s'agit plus seulement de connaître le client, mais d'inventer les dispositifs qui vont permettre aux utilisateurs d'être partie prenante de la production, des partenaires de l'institution et pas seulement des utilisateurs externes. Dans le domaine des institutions culturelles, il y a de plus de la part des publics un désir de participation (dont atteste l'activité des publics sur les plateformes) qu'il faut parvenir à canaliser.

### 3.3. Conclusion

Deux tournants principaux dans l'évolution de la connaissance des publics doivent être soulignés : on est d'une part passé **d'une logique de la représentativité à une logique de l'influence**. Tous les avis n'ont pas la même valeur, et les organisations s'intéressent davantage aux opinions qui ont une capacité d'influence. On s'écarte fortement du modèle démocratique de la représentativité où chaque voix serait censée être équivalente. D'autre part, le second tournant est le passage **de la connaissance à la participation des publics**. Il importe moins de connaître ses clients que de les amener à contribuer à la production.

Les études (quantitatives et qualitatives) qui ont pendant longtemps été la principale source de la connaissance des publics gagnent aujourd'hui à être articulées avec des méthodes d'enquêtes propres au numérique : analyse de traces de navigation, analyse de la structure des réseaux et des opinions produites spontanément par les utilisateurs, mais aussi analyse fine des interactions avec les dispositifs techniques.

L'enjeu pour les organisations est de mettre en place des systèmes de captation des usages qui permettent une vision unifiée et qualifiée de l'utilisateur, ce qui implique un travail en profondeur sur les architectures de SI et les applications.

Offrir au public des opportunités de participation, l'associer à la production, est un moyen pour satisfaire cette inclination à la contribution que l'on observe dans certaines franges des publics tout en permettant à l'organisation d'acquérir une meilleure connaissance des usages et des attentes.

## 4. Pistes méthodologiques et recommandations

Ces propositions sont nourries par l'enquête interne et l'état de l'art. Elles ont fait l'objet de discussions lors du séminaire du 27 juin 2014 qui s'est tenu sur le site de Richelieu (22 participants : direction des Collections, direction des Services et des réseaux, délégation à la Stratégie et à la recherche, Télécom ParisTech).

### 4.1. Orientations générales

Les mutations observées dans les deux volets précédents invitent à dessiner quelques pistes succinctes sur le plan méthodologique. Nous pouvons retenir deux grandes directions sur ce plan, qui consistent essentiellement à opérer un changement de perspective par rapport aux formes de représentations des usagers privilégiées jusqu'ici (qu'il n'est pas question d'abandonner, évidemment) et à resserrer les « mailles » des modalités de saisie du public.

#### 4.1.1. L'articulation des outils et des formats d'investigation

Pour assurer une meilleure mobilisation de la variété des formats de présence des usagers et de leurs outils, et en assurer le caractère opératoire, un travail d'articulation méthodologique semble souhaitable.

Cette articulation pourrait prendre *a minima* la forme d'une intégration plus grande des études quantitatives et qualitatives. Sans imaginer une articulation systématique et procédurale qui risquerait de figer l'innovation et la créativité des études, certaines opérations pourraient mieux bénéficier des pistes et des résultats des précédentes. Les études qualitatives pourraient par exemple être mieux mises en valeur dans l'élaboration des études quantitatives, dont certains aspects pourraient permettre de confirmer ou d'explorer des éléments identifiés par les premières. Les entretiens approfondis permettent de glaner du vocabulaire et des pratiques ancrées qui pourraient par exemple être utilisés dans les questionnaires. Une population sondée qualitativement, ou observée en ligne pourrait faire l'objet d'une étude quantitative à petite échelle.

De manière plus ambitieuse, des articulations sur le long terme pourraient être envisagées, entre les études quantitatives récurrentes sur le modèle de l'étude GMV, une analyse de traces d'usages menée avec XiTi, une partie de la veille sur les réseaux sociaux et certaines études qualitatives.

#### 4.1.2. Décentrement

Quelles que soient les méthodes concernées, il est possible aujourd'hui de mettre en œuvre un décentrement des points d'entrée les plus évidents, afin de compléter les approches traditionnelles et d'enrichir la liste des entités évoquées plus haut.



### *De l'interface... au parcours*

Comment les usages Gallica s'intègrent-ils dans des cours d'action de court et moyen terme ? Pour quoi faire ? Quelles complémentarités, quelles frictions éventuelles ? Nous l'avons, vu, en complément de la figure de l'utilisateur de l'interface, ou plus généralement de l'utilisateur du bouquet de services en ligne, il y a une utilité à situer ces usages dans des parcours plus larges. Essentiellement nourrie de méthodes qualitatives, cette préoccupation pour des usages dynamiques et décloisonnés est déjà partiellement à l'œuvre. Elle pourrait être consolidée.

### *De l'utilisateur... à l'usage*

La totalité des entités décrites plus haut s'appuie sur une figure commune : celle d'un sujet, plutôt rationnel, auquel sont attachées des caractéristiques et des formes d'usage, des plus minimales (l'audience) aux plus précises (les comportements face à telle ou telle fonctionnalité). Ce sont ainsi des personnes qui forment l'unité des usages appréhendés.

Au vu des transformations repérées, une partie des études portant sur Gallica aurait tout intérêt à suivre les conseils de Howard Becker (Becker, 2002) et à se détacher d'une telle unicité. Il y a en effet une utilité scientifique et opérationnelle à favoriser une entrée par les usages plutôt que par les utilisateurs, ne serait-ce que parce que l'on fait ainsi l'économie de présupposés de cohérence, de rationalité et de constance chez les seconds qui peuvent parasiter la compréhension des premiers. La question récente des horaires de consultation de Gallica a apporté une bonne illustration de ce déplacement possible. En repérant des consultations tardives, notamment par l'intermédiaire des technologies mobiles que sont les tablettes et les téléphones, on peut assurément imaginer que l'on ne découvre pas un nouveau type d'utilisateurs, mais bien une forme d'usages, sans doute adoptée par des personnes variées qui, elles-mêmes, passent dans une même journée ou une même semaine de formes d'usages à d'autres.

Cette perspective est bien connue de la sociologie des pratiques culturelles, qui a découvert assez tard qu'un intellectuel de la classe supérieure pouvait prendre plaisir à consommer des produits culturels « populaires » et surtout passer par des états différents au fil du temps. Elle permet de se focaliser sur des choses telles que la variété des formes d'engagement dans Gallica, de rapport aux documents, aux corpus, la diversité des formes de recherche, sans que celles-ci soient rabattues sur des stéréotypes ou des profils d'utilisateurs, qui ne sont pas forcément utiles dans une institution comme celle-ci. Issus du marketing, ils permettent surtout de désigner des marchés cibles. Sur le plan opérationnel, il est par ailleurs généralement plus simple d'équiper des formes d'usage que de tenter de s'adresser à des populations particulières.

### *Du site... au document*

Enfin, comme il a été évoqué précédemment, il y a un grand intérêt à poursuivre les premières expériences autour du taux de consultation des documents et passer à des mesures plus fines sur des ensembles de documents resserrés, définis en partenariat avec les équipes des départements des collections.

Outre cette piste, certaines préoccupations révèlent le grand apport que représenterait la capacité de suivre des trajectoires de documents à travers le web. Y a-t-il moyen, à partir d'outils de ce type, de systématiser, sans que cela soit trop coûteux, ce que permet la veille à une échelle restreinte à l'occasion d'opérations spécifiques : découvrir les usages, les citations, les réutilisations de documents Gallica hors du site, sur des blogs, les plateformes de réseaux sociaux, les forums ?

## 4.2. Propositions opérationnelles

### 4.2.1. Relancer une étude globale croisant les méthodes

La BnF a confié à un prestataire en 2011 une grande enquête par questionnaire auprès des publics de Gallica. Cette enquête quantitative, complétée par une vague d'entretiens qualitatifs d'approfondissement et d'un focus sur les publics de médiateurs, donne une vision d'ensemble des usages de Gallica et de leur contexte.

Il pourrait être intéressant de relancer cette étude en la faisant évoluer dans quatre directions :

- Articulation avec les enquêtes qualitatives, pour la rédaction du questionnaire et pour approfondir quelques thématiques spécifiques ;
- Consolidation de la segmentation des utilisateurs, conçue de manière itérative, en explicitant à chaque étape les choix opérés ;
- Réflexion sur ce que pourrait apporter à la connaissance des publics de Gallica et à la place de Gallica sur le web français le recours à des échantillons prédéfinis par les sociétés de mesure d'audience ; partenariat avec l'une de ces sociétés pour pouvoir recueillir en mode « user centric » la navigation sur le site et sur les sites de même nature d'un sous-échantillon illustrant les segments d'utilisateurs identifiés ;
- Exploration en laboratoire des usages de Gallica (de sa nouvelle interface et de son nouveau moteur) avec les techniques d'eye-tracking ou bien enregistrement de l'activité en situation naturelle et autoconfrontation pour capter des usages multi-support (ordinateur, tablettes, téléphones...).

Ceci permettrait de mettre en relation le profil des individus, leurs parcours sur le web et leur manière de naviguer. L'idée serait aussi d'avoir des éléments pour faciliter l'accès à des publics non-initiés.

Ce serait l'occasion d'articuler dans un même dispositif des méthodologies complémentaires.

### 4.2.2. Utiliser la mesure d'audience pour étudier l'usage des documents

XiTi est l'outil très puissant utilisé par la BnF pour mesurer l'audience de Gallica et des autres sites. Il permet de suivre l'évolution des usages, ce qui est fondamental dans une période de croissance des usages comme celle que connaît Gallica.

L'outil est personnalisable en fonction des besoins, mais nécessite des compétences de traitement des données pour en tirer le meilleur parti. La limite principale tient à l'impossibilité de remonter de manière fiable à des individus qualifiés, voire à pouvoir suivre dans le temps une cohorte de visiteurs pour suivre l'évolution de leurs usages.

En revanche, XiTi devrait permettre de construire une vision très fine des usages du point de vue des documents, comme cela a commencé à être fait. Les données de trafic permettent une vision exhaustive, du point de vue du document.

Quel sera le coût d'un marquage fin au niveau du document qui permette de produire des indicateurs par discipline, type de documents... et qui permette aussi d'évaluer les usages qui sont faits du document : ressources demandées, consultées, téléchargées, achetées... par catégories : types, discipline... Est-ce qu'une interconnexion avec les alignements de data.bnf.fr pourrait faciliter ce travail ?

Ceci permettrait de mettre en regard l'offre numérisée avec la demande.



#### 4.2.3. Vers un utilisateur « global »

Dans les études menées par la BnF, il y a une séparation nette entre la connaissance des visiteurs telle qu'elle se construit en salle de lecture et la connaissance des usages de Gallica. Or, la construction d'une vision intégrée des usages serait en phase avec la stratégie d'intégration progressive de Gallica dans le cœur de l'activité de la bibliothèque. Elle permettrait aussi de mieux comprendre les mécanismes de complémentarité et de substitution entre les usages en salle et en ligne.

Une piste serait de proposer aux utilisateurs un espace personnel unifié (en salle et sur Gallica) qui permettrait de mesurer les usages dans les deux environnements.

Si une carte d'accès est nécessaire pour accéder aux salles de lecture de la bibliothèque, la consultation de Gallica ne nécessite pas d'identification (il existe certes un espace personnel, mais il est assez peu développé et peu utilisé). Or la construction d'une vision globale de l'usager (qui ne segmenterait plus les usages en salle des usages numériques) implique la construction d'un système d'identification unique qui permette de mettre en relation l'activité sur Gallica et la circulation dans les espaces de lecture, la consultation des documents en ligne et en salle. Pour que les utilisateurs trouvent une motivation à s'identifier sur ces espaces, il faudra qu'ils y voient un intérêt. Autrement dit, il faudra que l'offre de service ait pour eux de la valeur ajoutée (cf. point suivant sur la participation).

Une seconde piste, moins coûteuse, consisterait à trouver des modes d'appariements entre les enquêtes sur les usages en salle de lecture et les enquêtes sur Gallica en ligne, ce qui nécessite de revoir les questionnaires en amont et de tester des procédures d'appariements, pour reconstruire des usages globaux.

### 4.3. Donner plus de visibilité à la participation des publics ?

Le tournant de la participation n'a pas encore été pris par les grandes bibliothèques nationales. La BnF est très présente et active sur les réseaux sociaux, mais cette activité pionnière est peu connectée au cœur de l'activité de la BnF, qui reste encore sur le modèle de la diffusion institutionnelle. Les visiteurs ne laissent pas de traces de leurs visites, les espaces personnels sont peu développés.

Sans doute le public lui-même ne s'attend-il pas à voir les documents du patrimoine commentés, évalués... et préfère-t-il sans doute que la ligne de partage entre l'institution et le public soit bien marquée. Cependant l'appétence pour la contribution des publics peut trouver d'autres formes d'expression. Deux pistes nous paraissent avoir du potentiel :

La première piste consiste à développer un **espace personnel enrichi** qui permette au visiteur de classer dans sa **bibliothèque numérique** les documents consultés que ce soit sur Gallica ou dans la bibliothèque physique, de les annoter (pour soi et pour son réseau), de les évaluer, éventuellement d'échanger avec d'autres Gallicanutes. C'est là qu'il pourrait aussi exprimer des attentes : la recherche de tel document pas encore numérisé, par exemple.

Dans cette démarche, la **compatibilité avec les autres systèmes de bibliothèques numériques** personnelles est centrale (avec les outils de gestion bibliographiques comme Zotero, Mendeley, avec les bibliothèques personnelles de type e-book, Amazon...).

La construction d'un espace personnel est la première marque d'une personnalisation de la relation entre le site et son public, en offrant la possibilité de stocker, de mettre en scène ses usages, voire de donner corps à un collectif d'utilisateurs sur le site même.

La seconde piste vise, au vu de l'appétence du public pour la contribution (cf. la grande Collecte pour la commémoration de la Guerre et autres initiatives), consiste à inventer et à mettre en place des dispositifs pour favoriser la participation des publics :

- **Contribution à la politique de numérisation.** Comme indiqué dans la première partie du rapport, il pourrait être pertinent de solliciter les visiteurs pour les choix et la temporalité des numérisations de documents, autrement dit compléter la numérisation de masse par une numérisation personnalisée, à l'écoute des attentes des lecteurs ;
- **Contribution à la correction des documents numérisés** comme cela est expérimenté dans le cadre du projet CORRECT ;
- **Contribution à l'évaluation, qualification, annotation des documents** (également en partie expérimentée dans le cadre de CORRECT). Un dispositif qui permette aux lecteurs d'intervenir sur les documents, de signaler et corriger des erreurs, de proposer des enrichissements, de qualifier les sources, d'apporter des annotations pourrait être proposé aux lecteurs. Ces contributions viendraient enrichir son profil et permettrait de valoriser une communauté de lecteurs contributeurs actifs.

#### 4.4. Aspects organisationnels

Une grande partie des résultats de cette exploration déborde des seules questions méthodologiques et pointe vers des aspects organisationnels.

Le premier consiste à décloisonner les lieux d'observation des usages. Cela suppose d'organiser au sein de la BnF la circulation des connaissances sans doute de manière plus systématique, sous la forme de comptes-rendus approfondis des analyses de traces, des études, et de la veille, mais aussi de discussions plus spontanées et de groupes de travail qui permettraient d'outiller la politique documentaire et de nourrir la réflexion autour de la production des métadonnées. À la richesse des formes et des outils d'appréhension des usagers correspond en effet un relatif éclatement des opérations et des initiatives, parfois une méconnaissance du travail des uns et des autres. Il y a évidemment un risque à vouloir trop structurer ces aspects et à bureaucratiser des processus qui doivent pouvoir bénéficier d'une souplesse importante pour être efficaces. Il reste toutefois à imaginer un lieu partagé et une forme d'échanges autour des nombreux instruments de représentation, qui permettent de gagner en transversalité autour des problématiques liées aux usages de Gallica, voire au public en général. Cette thématique des usages pourrait aussi faire l'objet d'une réflexion collective dans le réseau mondial des bibliothèques nationales.

À cela s'ajoute l'intérêt qu'il y aurait à favoriser l'implication de certains coordinateurs des opérations de numérisation des départements de collections dans la mise en place de dispositifs de traçage des usages. Nous l'avons vu, XiTi fonctionne principalement par marquage des éléments dont on cherche à mesurer l'audience et l'usage. Dans la perspective d'un développement de mesures adaptées aux problématiques des collections (évaluer la qualité des métadonnées, comprendre les logiques d'usage des corpus...), il faudrait pouvoir monter des équipes de marquage ad hoc, chargées de définir et d'implémenter des outils de suivis statistiques sur mesure.

Ce dernier point soulève plus généralement la question de la place de XiTi au sein de la BnF. Relativement autonome aujourd'hui, et permettant de produire des bulletins réguliers ainsi que des explorations ponctuelles, on pourrait imaginer plusieurs pistes pour en favoriser l'intégration à d'autres processus internes.

Dans un premier temps, des séances de présentation de l'outil et de ses multiples possibilités auprès des équipes susceptibles d'être intéressées dans les départements de la BnF pourraient être envisagées.

Cela permettrait une meilleure visibilité de l’outil et serait l’occasion de structurer le circuit des demandes qui remontent aujourd’hui à la personne chargée des mesures. De même, les bulletins concis gagneraient sans doute à être diffusés plus largement.

Dans une perspective d’articulation telle que décrite précédemment, XiTi gagnerait également à être mobilisé, par exemple, en amont d’une étude quantitative lors de la rédaction du cahier des charges, ou au contraire en aval comme moyen de poursuivre les observations initiales.

Mais tout cela a un coût. L’accessibilité technique de l’outil pourrait laisser croire que tout le monde bénéficiant d’un compte pourrait facilement bénéficier d’une vue synthétique et utile sur les données qui l’intéressent. C’est évidemment une illusion. Non seulement l’outil et sa manipulation sont complexes, mais les données qu’ils génèrent ne sont pas intelligibles en tant que telles. Elles demandent non seulement un travail du côté de leur recueil (la définition d’une problématique), mais aussi du côté de leur lecture (comme n’importe quelles données, elles doivent être analysées). L’ampleur de la place qui peut être donnée aux mesures de traces en ligne dépend donc des ressources que la BnF peut y consacrer.

On peut, pour conclure, évoquer des pistes d’articulation avec un autre service interne qui n’a pas été mentionné jusqu’ici. L’une des difficultés dans le marquage de XiTi consiste à reconstituer des ensembles cohérents au sein de l’environnement numérique fragmenté que représente l’architecture d’un site web aussi complexe que Gallica : un document à partir de plusieurs pages, une collection à partir de document, etc. Pour marquer intelligemment, il est nécessaire de réorganiser et d’agglomérer les objets dont on souhaite mesurer l’usage. Or, les outils mis en place dans le cadre de data.bnf.fr participent précisément de cette structuration en déployant un réseau de données liées qui permettent de reconstituer des ensembles. On peut faire l’hypothèse (qui reste toutefois à confirmer) qu’une telle infrastructure pourrait être utile dans la mise en place de certains marquages.

## 5. Bibliographie

- Akrich, M., 2012. Les listes de discussion comme communautés en ligne : outils de description et méthodes d’analyse (No. n°25). Paris.
- Assadi, H., Beauvisage, T., 2003. Projet BibUsages - Rapport final.
- Auray, N., Moreau, F., 2012. Industries culturelles et Internet. Les nouveaux instruments de la notoriété, in: Réseaux. pp. 9–18. doi:10.3917/res.175.0009
- Barats, C., 2013. Manuel d’analyse du Web. Armand Colin, Paris.
- Bastard, I., Bourreau, M., Maillard, S., Moreau, F., 2012. De la visibilité à l’attention : les musiciens sur Internet. Réseaux 175.
- Beaudouin, V., 2011. Prosumer. Communications 89, 131–139.
- Beaudouin, V., Fleury, S., Velkovska, J., 2000. Études des échanges électroniques sur internet et intranet : forums et courriers électroniques, in: (eds.), M.R.& J.-C.C. (Ed.), JADT 2000 : 5es Journées Internationales d’Analyse Statistique Des Données Textuelles. EPFL, pp. 17–26.
- Beaudouin, V., Licoppe, C., 2002. Parcours sur internet. Réseaux Vol. 20.
- Beaudouin, V., Pasquier, D., 2014. Organisation et hiérarchisation des mondes de la critique amateur cinéophile. Réseaux vol. 32, p. 123–160.
- Beauvisage, T., 2004. Sémantique des parcours des utilisateurs sur le Web. PhD in Sciences du Langage, University of Paris X, .... PhD in Sciences du Langage, University of Paris X.
- Becker, H.S., 2002. Les ficelles du métier. Comment construire sa recherche en sciences sociales. La Découverte, Paris.

- Becker, M.H., 1970. Sociometric Location and Innovativeness : Reformulation and Extension of the Diffusion Model. *American Sociological Review* 35, 267–282.
- Benzécri, J.-P., 1968. La place de l'a priori, "Organum" ", in: *Encyclopedia Universalis*. pp. 11–24.
- Boullier, D., Lohard, A., 2012. *Opinion mining et sentiment analysis*. Openedition Press.
- Bourdieu, P., 1979. *La distinction. Critique sociale du jugement*. Les éditions de Minuit.
- Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P., 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM*.
- Coleman, J.S., Katz, E., Menzel, H., 1966. *Medical Innovation: A Diffusion Study*. Bobbs-Merrill Co., Inc., New York.
- Cristofoli, P., 2008. Aux sources des grands réseaux d'interactions. Retour sur quelques propriétés déterminantes des réseaux issus de corpus documentaires. *Réseaux* 26, 21–58.
- Donnat, O., 2009. *Les pratiques culturelles des Français à l'ère numérique. Enquête 2008*. La Découverte, Paris.
- Donnat, O., 2011. études. *Pratiques culturelles, 1973-2008. Dynamiques générationnelles et pesanteurs sociales*.
- Fuchs, V.R., Leveson, I.F., 1968. *The Service Economy*, October. National Bureau of Economic Research.
- GMV, 2012. *Évaluation de l'usage et de la satisfaction de la bibliothèque numérique Gallica et perspectives d'évolution*. Paris.
- Granovetter, M., 1973. The Strength of Weak Ties. *American Journal of Sociology* 78, 1360–1380.
- Héran, F., 1988. La sociabilité, une pratique culturelle. *Economie et statistique* n°216, 3–22.
- Hersent, J., 2000. *Sociologie de la lecture en France: état des lieux*, Direction du livre et de la lecture. Juin.
- Java, A., Finin, T., Song, X., Tseng, B., 2007. Why we Twitter : Understanding Microblogging usage and communities, in: *International Conference on Knowledge Discovery and Data Mining. Proceedings of the 9th WebKDD and 1st SNA-KDD*.
- Katz, E., Lazarsfeld, P.F., 1955. *Personal influence: The part played by people in the flow of mass communications*. Free Press.
- Kotras, B., 2013. *Des opinions qui (se) comptent. Influence, visibilité et hiérarchisation des opinions sur le web*.
- Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is Twitter , a Social Network or a News Media ? Categories and Subject Descriptors, in: *WWW 2010. WWW 2010*.
- Lahlou, S., 2011. How can we capture the subject's perspective? An evidence-based approach for the social scientist, *Social Science Information*. doi:10.1177/0539018411411033
- Lave, J., Wenger, E., 1991. *Situated learning : legitimate peripheral participation*. Cambridge University Press, Cambridge.
- Lebart, L., Salem, A., 1994. *Statistique textuelle*. Dunod, Paris.
- Licoppe, C., 2008. Dans le « carré de l'activité » : perspectives internationales sur le travail et l'activité. *Sociologie du Travail* 50, 287–302.
- Liu, B., 2012. *Sentiment Analysis and Opinion Mining*. Morgan&Claypool Publishers.
- Moreno, J.L., 1934. *Who Shall Survive? A New Approach to the Problem of Human Interrelations*, Vasa. Nervous and Mental Disease Publishing Co., Washington.
- Pasquier, D., 2014a. Les jugements profanes en ligne sous le regard des sciences sociales. *Réseaux* 32, p. 9–25.
- Pasquier, D., 2014b. *Les évaluations profances*. Réseaux 1.
- Rogers, E.M., 1962. *Diffusion of Innovations*, 4th Edition (Google eBook). Simon and Schuster.
- Tenopir, C., 2003. *Use and Users of Electronic Library Resources : An Overview and Analysis of Recent Research Studies*.

- Velkovska, J., Zouinar, M., 2007. Interaction visiophonique et formes d'asymétrie dans la relation de service. *Réseaux* 25, 225–264.
- Von Hippel, E., 2002. Horizontal innovation networks, by and for users.
- Von Hippel, E., 2005. *Democratizing Innovation*. MIT Press, Cambridge, Mass.
- Watts, D.J., 2004. The “New” Science of Networks. *Annual Review of Sociology* 30, 243–270. doi: 10.1146/annurev.soc.30.020404.104342
- Watts, D.J., Dodds, P.S., 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research* 34, 441–458.