



HAL
open science

Les bibliothèques et la visibilité des données culturelles sur le web : réalité d'un défi

Kaouther Azouz

► **To cite this version:**

Kaouther Azouz. Les bibliothèques et la visibilité des données culturelles sur le web : réalité d'un défi. Colloque Spécialisé en Sciences de l'Information (COSSI), Jun 2014, Poitiers, France. hal-01393087

HAL Id: hal-01393087

<https://hal.science/hal-01393087>

Submitted on 18 Nov 2016

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.

Kaouther AZOUZ

Université Charles de Gaulle – Lille 3 (France)

Laboratoire de recherche (Sciences de l'Information et de la Communication) :

GERiiCO

Résumé : La visibilité des données culturelles sur le web est l'une des préoccupations majeures des bibliothèques à l'ère du web des données. L'émergence de nouveaux usages, le développement des technologies du web sémantique, la richesse et la dissémination des données culturelles dans les silos du web invisible incitent les bibliothèques à mettre en place des stratégies de valorisation de leurs données sur le web. L'absence partielle des *OPACS* des premières *SERP* (*Search Engine Pages*), le recours des internautes aux moteurs de recherche, la concurrence de la donnée culturelle par d'autres types de données accentuent la mise en place de cette initiative. Faciliter l'accès aux ressources rares des bibliothèques, qui ne parviennent pas à être indexées par les agents du web, et dont l'intérêt est majeur pour l'internaute, demeure l'une des principales visées de cette démarche. L'atteinte d'un public dont elles ne maîtrisent pas les usages serait pour les bibliothèques le plus grand défi à réaliser. Le recours aux technologies du web sémantique et l'intégration de *Schema.org* permettront-ils l'aboutissement à cette finalité ?

Mots-clés : Bibliothèques, Web des données, Schema.org, Knowledge Graph, Référencement, Visibilité, Données culturelles.

1. Introduction

1.1. Étude du contexte d'émergence

Les bibliothèques sont au cœur des débats actuels des professionnels de l'information, de la documentation et du web. Principaux pivots de la connaissance, ces institutions sont en train de connaître des évolutions révolutionnant leurs processus de travail et leurs méthodes de traitement et d'analyse de l'information. Renommées pour la préservation du patrimoine scientifique et culturel, la structuration des données, la normalisation des méthodes de traitement et d'analyse de l'information, la gestion et l'organisation des connaissances à travers un ensemble d'outils et de dispositifs cognitifs, les bibliothèques ont toujours réussi à mettre en place des stratégies d'intermédiation info-communicationnelles auprès de leur lectorat. Le contact physique avec leur public leur facilite le ciblage et la gestion de ses besoins. De ce fait, les bibliothèques « physiques : traditionnelles ou hybrides » ont toujours su

répondre aux exigences de leurs publics, mais quelle serait leur attitude vis-à-vis de l'émergence d'un nouveau public dont elles ignorent les besoins et les usages ? Comment se comporteraient-elles face à un usager habitué à trouver l'information en un temps réel et sans avoir à passer par le biais des portails ou des sites des bibliothèques ?

Avant d'évoquer le comportement des bibliothèques vis-à-vis des besoins émergents de l'internaute, il serait opportun de comprendre leur existence sur le web. L'intégration de cet espace instable et mouvant, par ses technologies et usages, les contraint à revoir leurs stratégies de valorisation et de mise à disposition de leurs *OPACS*⁶ et de l'ensemble de leurs ressources sur le web. L'accès à l'information et sa disponibilité ne semblent plus être les seuls critères pour l'utilisateur. Les institutions culturelles dont les bibliothèques se rendent de plus en plus compte de la baisse constante de la consultation de leurs catalogues. Enfouies dans les silos du web invisible, les données bibliographiques ne remontent que très peu dans la liste des résultats générés par les moteurs de recherche.

L'émergence des technologies du web sémantique, du web des données, des données ouvertes, des données liées et des données liées et ouvertes (*Linked Open Data*) sont-ils en train de remédier à cette situation. Des vocabulaires permettant une granularité très fine de la donnée, des normes compatibles avec les standards et les protocoles du web, des outils de référencement « sémantiques » et des stratégies de médiation empruntées parfois au marketing du web sont en train de se mettre en bibliothèques. L'objectif est d'assurer une meilleure visibilité, améliorer le référencement, parvenir à « garder » l'internaute sur leur site et le rediriger vers des ressources institutionnelles de confiance sans avoir à recourir aux moteurs de recherche.

A l'ère du web des données et des technologies du web sémantique, les bibliothèques sont de plus en plus impliquées dans la question de visibilité et de valorisation de leurs ressources sur le web. L'enjeu est stratégique, elles doivent atteindre leur lectorat mais aussi l'internaute. Des statistiques concernant les visites ponctuelles de ce type d'utilisateurs pour les pages web des bibliothèques sont collectées et mises à jour. Elles visent à déceler les éléments permettant la construction du « parcours critique » de l'internaute lors de son processus de recherche sur le web. La notion du « parcours » est très importante. Elle ne cesse de susciter l'intérêt des professionnels des bibliothèques car elle reflète le raisonnement et le cheminement qu'emprunte ce dernier dans sa recherche de l'information, depuis l'élaboration de sa requête jusqu'à la récupération du document final ou de l'information qui l'intéresse.

⁶ *Online Public Access Catalog*

Outre l'analyse des facteurs de non-visibilité des données culturelles dans les premières pages de résultats des moteurs de recherche et l'étude des comportements des internautes dans leur processus de recherche de l'information sur le web, on mettra l'accent sur la question du référencement. L'analyse des algorithmes et des stratégies de collecte, de moissonnage et de valorisation de la donnée des bibliothèques et celle des institutions culturelles par les moteurs de recherche sera évoquée tout au long de cette recherche. Elle permettra une meilleure compréhension des mécanismes de fonctionnement du référencement. De plus en plus évoqué par les institutions culturelles, le référencement suscite l'intérêt des moteurs de recherche. Le « marché de la donnée culturelle » et celui du « web culturel » s'ancrent progressivement chez les géants du web. Cette nouvelle tendance est décelée chez les grands moteurs de recherche : *Google, Yahoo, Bing et Yandex* à travers « *Schema.org* ». L'objectif de ce schéma est de faciliter le repérage et l'indexation des données exprimées en *RDF (Resource Description Framework)*, principal vocabulaire de publication du web sémantique.

Le dernier volet de cette recherche est dédié à l'étude des stratégies mises en place par les bibliothèques et les institutions culturelles pour remédier à l'invisibilité partielle de leurs données. Le recours aux liens ou aux *URI (Uniform Resource Identifier)* serait l'alternative idéale pour les bibliothèques pour faire face au processus de *crawling* actuel des moteurs de recherche pour la donnée culturelle.

1.2. Problématique et hypothèse

L'implémentation des technologies du web sémantique et la mise en place des stratégies de liens sont-elles les nouvelles mesures d'amélioration de la visibilité des données et des contenus culturelles des bibliothèques sur le web ? Les bibliothèques parviendraient-elles à relever le défi du référencement et le *crawling* de leurs données par les moteurs de recherche ? Réussiraient-elles à découvrir les secrets de l'internaute dans la construction de son parcours « critique » ? *Schema.org* est-il la solution parfaite pour que les bibliothèques et les moteurs de recherche coopèrent ensemble pour la valorisation de la donnée culturelle ?

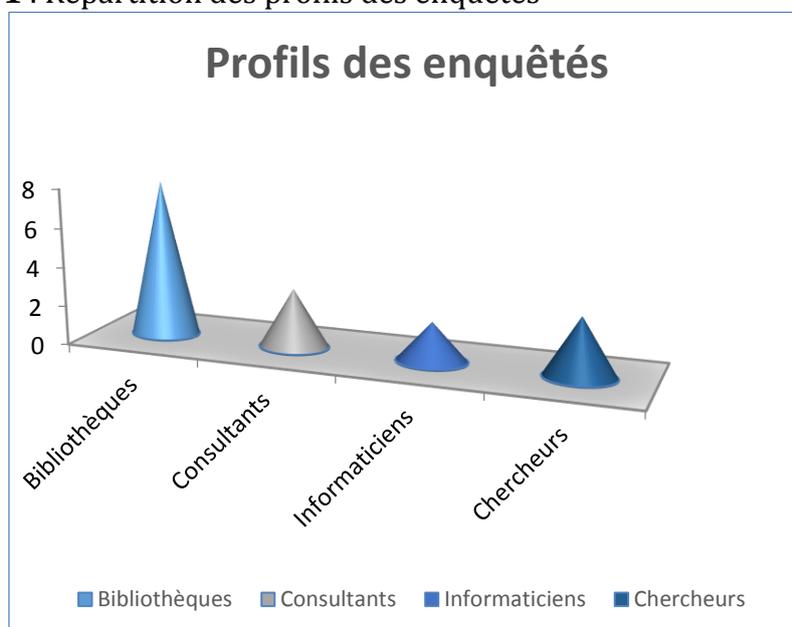
Si les bibliothèques choisissent d'intégrer les technologies du web sémantique et *Schema.org*, elles parviendront à la visibilité et assureront un meilleur référencement auprès des moteurs de recherche et des usagers.

2. Méthode

Pour la réalisation de cette recherche nous avons mis en place une enquête préliminaire permettant de déceler les différentes approches des professionnels de l'information et des bibliothèques. Des entretiens ont été menés avec quatre grandes catégories réparties de la manière suivante :

- Les professionnels des bibliothèques
- Les chercheurs universitaires (spécialisés dans la question du web sémantique)
- Les informaticiens développeurs et concepteurs de systèmes d'information)
- Les consultants (ayant une expertise pour les questions d'implémentation des technologies du web sémantique)

Figure 1 : Répartition des profils des enquêtés



2.1. Les professionnels des bibliothèques

Huit professionnels des bibliothèques ont été sollicités pour la réalisation de cette enquête préliminaire. Les quatre premiers appartiennent à la Bibliothèque nationale de France, particulièrement à l'équipe « *data.bnf.fr* ». Les deux autres se répartissent de la manière suivante : une responsable du réseau régional du *SUDOC* et un conservateur de la « *Bibliothèque de Rennes Métropole* ». La principale caractéristique de ces profils c'est qu'ils appartiennent tous à la même communauté et se concertent pour attester de l'invisibilité partielle de leurs données sur le web. Là où ils divergent c'est sur la question des publics. Si la majorité des professionnels ciblent l'internaute de manière générale, d'autres expriment leur volonté d'améliorer la qualité des services à leurs propres

lecteurs et publics. Ils considèrent que ces derniers sont plus concernés par la question de visibilité que l'internaute.

2.2. Les consultants

Les entretiens avec les consultants nous ont permis une meilleure compréhension des mécanismes de fonctionnement des *crawlers* dans leurs processus de moissonnage et d'indexation des pages web. N'ayant pu avoir des entretiens avec les experts du *SEO* (*Search Engine Optimization*) nous avons recouru aux connaissances des consultants en matière de référencement. Leur approche est intéressante car elle complète celle des bibliothèques. Confirmant à leur tour l'absence de la donnée culturelle des premières pages de résultats des moteurs de recherche et sa non détection par les *crawlers*, les consultants nous livrent une série de recommandations et de stratégies « prospectives » pouvant remédier à l'amélioration de cette visibilité.

Les liens sémantiques semblent être l'une des principales alternatives pour l'atteinte des premières *SERP* (*Search Engine Pages*). Ils sont même présentés par certains comme étant un nouveau service de navigation pour l'utilisateur. La question des liens suscite beaucoup d'intérêt car elle fascine à la fois les bibliothèques, les consultants, les informaticiens et les chercheurs.

2.3. Informaticiens (développeurs et concepteurs de systèmes d'information)

Outre les deux premières catégories, nous avons aussi contacté deux développeurs et concepteurs de systèmes d'information fondant leur offre de service sur les technologies du web sémantique. Ces derniers ont mis l'accent sur *Schema.org* et son impact sur le référencement des données et des institutions culturelles. Ils expliquent les raisons pour lesquelles *Google* privilégie la donnée commerciale à la culturelle. Ils évoquent aussi la question des liens et l'intérêt qu'elle peut représenter pour les bibliothèques afin d'assurer leur visibilité sur le web.

2.4. Chercheurs universitaires spécialisés dans le web sémantique

Le monde de la recherche scientifique est impacté par les problématiques du web des données et des technologies du web sémantique. Des études et des travaux confirment cette tendance. Tout au long de notre recherche, nous avons pu mettre en place des entretiens avec trois figures emblématiques du domaine. Leurs réflexions et travaux autour de la question du web sémantique et de la donnée culturelle nous ont bien aidés à cerner les enjeux de la valorisation de ce

type de donnée par les bibliothèques et les moteurs de recherche à travers des angles d'analyse divergents.

3. Analyse

Le déploiement des réponses a permis le dégagement des axes suivants : la question de visibilité de la donnée culturelle, *Schema.org* et le référencement sémantique de la donnée culturelle : quels usages, quels apports et quelles limites ? Le *Page Rank* et l'affichage des résultats des données des bibliothèques, les stratégies de visibilité élaborées par les bibliothèques (intégration du *Schema. Org* comme c'est le cas pour l'*OCLC (Online Computer Library Center)* et l'*ABES (Agence Bibliographique de l'enseignement Supérieur)* ou adoption de la stratégie des liens sémantiques tout comme le fait actuellement la *BnF*, les publics ciblés par la visibilité, l'intérêt de cette visibilité, les limites du référencement sémantique.

3.1. Motifs d'amélioration de la visibilité de la donnée culturelle en bibliothèques

Les bibliothèques sont de plus en plus conscientes de l'invisibilité partielle de leurs données. Elles sont confrontées à la concurrence des moteurs de recherche, des encyclopédies en ligne (*Wikipédia*) et des plateformes de services (*Amazon*) permettant l'accès rapide à l'information. « [...] *la bibliothèque n'est plus un passage obligé pour accéder aux documents* » [*Bermès et al., 2010*]. Ce constat est confirmé par les professionnels des bibliothèques, de l'information, de la documentation des consultants, des chercheurs et des informaticiens. Quelles sont les raisons d'une telle situation et quelles seraient les stratégies et solutions mises en place pour remédier à cette invisibilité ? Quel type de public cible-t-on par une telle démarche ?

Selon les professionnels des bibliothèques et l'ensemble des enquêtés, la question des formats de production de la donnée culturelle est primordiale à l'explication de cette invisibilité. La non homogénéisation des formats (*Marc, InterMarc, Marc21, Unimarc, Dublin Core, EAD, etc.*), leur non formulation et publication dans des vocabulaires compatibles avec les standards du web et compréhensibles par les *crawlers* demeurent les principales raisons pour lesquelles les données ne peuvent ni être moissonnées ni remontées des silos du web invisible.

On relève aussi l'émergence de la notion de « longue traîne » en bibliothèques. C'est un concept emprunté à *Chris Anderson* et à son étude exploratoire pour les statistiques de ventes des sites web de commerce électronique : « [...] (*la tête de la traîne*) représente les *best-sellers* : peu de produits très populaires représentant

*de très nombreuses ventes pour chaque référence...(la « queue » de la longue traîne) est représentative de produits peu vendus...mais en très grand nombre de références différentes »*⁷. Dans le cadre des bibliothèques et du référencement des données culturelles, les professionnels précisent que cette « longue traîne » ne touche que les ressources rares des bibliothèques. L'internaute est de plus en plus intéressé par la consultation et la récupération de données spécifiques qui n'existent nulle part ailleurs. « *Ce que veulent les bibliothèques...c'est que leurs données soient présentes sur le web mais pas n'importe où, plutôt là où c'est pertinent...Ce que vise la BnF c'est de publier des ressources peu connus qui seront consultées par des internautes »*⁸.

Ce nouvel usage génère un autre : la précision et la notion du détail dans la recherche d'information de l'internaute. Le choix d'une donnée ou d'une information précise dans un bloc de données, est de plus en plus prisé par l'utilisateur du web. Les professionnels constatent que ce dernier ne se limite plus à l'information généraliste des moteurs de recherche et tente des requêtes complémentaires avec des rebonds vers d'autres sites pour parvenir au document final. Ce comportement est connu sous le nom de « Parcours critique » ou « Parcours de l'utilisateur » dont l'intérêt est de déceler et comprendre le comportement de l'utilisateur dans son processus de recherche de l'information.

3.2. Les publics cibles et la notion du « parcours critique »

La notion du « Parcours Critique » reflète l'acheminement et le raisonnement logique de l'internaute dans son processus de recherche de l'information. C'est un comportement qui lui permet d'accéder par « hasard » ou de manière « réfléchi » aux pages web et aux ressources des bibliothèques. Selon les professionnels de la BnF, des visites ponctuelles des pages web de *data.bnf.fr* ont été enregistrées. Ce qu'ils ignorent c'est la manière à travers laquelle l'internaute a pu accéder à ces pages. « *Pour le moment il est difficile d'estimer ce que tapent les gens dans les moteurs de recherche pour arriver à la BnF de manière systématique et quelles sont les pages qu'ils préfèrent consulter ?* »⁹. Ce questionnement dénote un constat d'envergure : la non maîtrise des bibliothèques des mots-clés de l'utilisateur. Pour y remédier, les bibliothèques mettent en place des stratégies de visibilité permettant de croiser leurs ressources aux parcours de recherche de l'internaute. « *Ce que demande l'utilisateur et les lecteurs aujourd'hui c'est des ressources directes. Les processus de recherches éparses les faisant rebondir d'un site à un autre pour trouver le document final ne les intéressent plus. Il faudrait trouver un dispositif facilitant ce processus »*¹⁰.

⁷ Andrieu, O. Réussir son référencement web. Paris: EYROLLES. (2010)

⁸ Entretien avec l'équipe « data.bn.fr », 17/02/2014

⁹ *Ibid*

¹⁰ *Ibid*

3.3. Les liens sémantiques : nouvelle stratégie de visibilité ou nouveau service de navigation pour l'utilisateur ?

Le principal dispositif que voudraient mettre en place les bibliothèques pour améliorer leur visibilité sur le web et atteindre les différents types d'utilisateurs sont les liens par identifiants pérennes (*Uniform Resource Identifier*). L'intérêt de cette stratégie c'est qu'elle fournit à l'utilisateur une nouvelle manière de naviguer sur le web. Ce dernier ne sera plus contraint à passer par un moteur de recherche généraliste dans son processus de recherche. Il sera reconduit à d'autres sites institutionnels ou partenaires des bibliothèques à travers les liens de rebonds des catalogues en ligne. En mettant en place cette stratégie, les bibliothèques ne cherchent en aucun cas la concurrence avec les moteurs de recherche. Elles voudraient simplement affirmer leur existence dans l'écosystème du web.

La question des liens est importante. Elle améliore, selon les professionnels de l'information, la visibilité des bibliothèques mais ne parvient pas à leur faire éviter l'impact des moteurs de recherche. En effet, si les bibliothèques arrivent à garder l'internaute sur leurs pages web, elles parviendront à être plus citées par les autres sites en tant que référence de confiance et mieux référencées par les moteurs de recherche. A long terme, cette situation améliorera leur référencement.

3.4. Les moteurs de recherche et le référencement des données culturelles

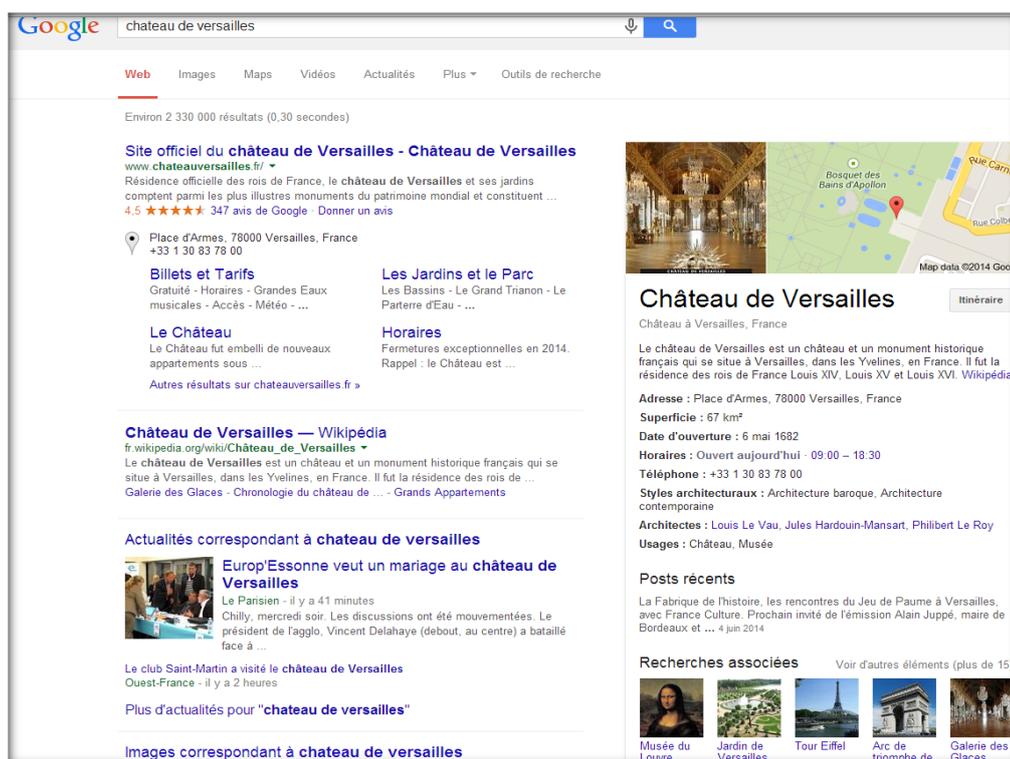
Les moteurs de recherche sont de plus en plus conscients de la nécessité de repérer et d'indexer les données culturelles, d'où la mise en place de différents types de stratégies. *Schema.org* et le *Knowledge Graph* sont les principaux fondements de collecte de la donnée structurée sur le web. *Schema.org* est particulièrement utilisé comme instance d'annotation dans les déclarations des pages web des bibliothèques. C'est un schéma de référencement très polémique. Il a été créé pour remédier à la question de visibilité des données sur le web et leur repérage par les agents du web. Il reste, par ailleurs, très pauvre en propriétés et ne parvient pas à décrire toute la granularité et la richesse des métadonnées des bibliothèques. Sa critique par les professionnels des bibliothèques infirme le discours prônant son efficacité à faire remonter les données vers les moteurs de recherche et de leurs *crawlers*. La réalité est toute autre, la question du référencement est plus complexe qu'elle ne semble l'être. La donnée culturelle n'est pour les moteurs de recherche qu'un critère parmi tant d'autres. Or pour parvenir à être visible dans les premières *SERP* (*Search Engine Pages*) il est primordial de répondre à tous les critères du *Page Rank*. La logique des moteurs de recherche est fondée sur la notion du profit et du marché. Elle est de ce fait différente de celle de l'institution culturelle qui ne cherche que la

diffusion gratuite de la connaissance auprès d'un grand nombre d'internautes. Toutefois, l'intérêt des moteurs de recherche pour la donnée culturelle commence à se faire sentir à travers une notion qui s'ancre de plus en plus dans les usages des internautes c'est le « Panachage » et la présentation de résultats de « suggestion ». En quoi consistent ces deux concepts ?

Le panachage est une pratique générée par les moteurs de recherche. Elle permet à l'internaute d'accéder à un résultat et de l'affiner à travers des liens de rebonds. On passe d'un résultat précis à un autre encore plus précis via un moteur de recherche. La question est comment parvient-on à obtenir cette précision à travers des moteurs de recherche généralistes ?. C'est ce que cherchent à comprendre les bibliothèques. Outre la précision des résultats, le panachage permet à l'utilisateur de rebondir vers des sites institutionnels de confiance qui lui fourniront des ressources et des informations bien documentés et avec des référentiels de qualité lui permettant de retrouver une information très pointue et qui ne pourra lui être fournie par aucun moteur de recherche.

La seconde notion est celle des « résultats par suggestion ou recherches associées ». Cette tendance est proposée par le *Knowledge Graph* dans son encart de droite qui suggère des liens et des résultats annexes susceptibles d'intéresser l'utilisateur [voir Fig.2]. C'est une manière pour l'internaute d'affiner sa recherche et de rebondir sur des sites intéressants.

Figure 2 : Exemple de recherche par suggestion ou « associées » du *Knowledge Graph*



Cet ensemble d'usages nous permet de conclure que la structure du web est en train de changer, elle n'est plus en adéquation avec les besoins de l'utilisateur. Ce dernier ne consulte plus de manière spontanée les catalogues des bibliothèques mais effectue de plus en plus ses requêtes via les moteurs de recherche.

4. Discussion

La question des données culturelles et leur exploitation par l'utilisateur sur le web interpelle les professionnels des bibliothèques, de l'information et les moteurs de recherche. Les premiers constatent que les données de leurs catalogues sont cachées dans les silos du web invisible et ne parviennent pas à être indexées par les moteurs de recherche. Les seconds, prennent conscience de l'intérêt de recenser la donnée culturelle dans leurs résultats de recherche. Cette réalité est exprimée par certains chercheurs en sciences de l'information et de la communication tels que *Manuel Zacklad* dans son ouvrage « *Vu, Lu, Su : les architectes de l'information face à l'oligopole du web* ». Elle est aussi évoquée par des professionnels des bibliothèques à l'instar d'*Emmanuelle Bermès* et son ouvrage « *Le web sémantique en bibliothèques* ». Ce sont les rares ouvrages qui posent explicitement la problématique de visibilité des données culturelles sur le web. Concernant l'évolution des moteurs de recherche vers un référencement plus structuré et une recherche plus sémantique, nous avons consulté

l'incontournable ouvrage de *David Amerland* : « *Google Semantic Search* ». Cette nouvelle tendance est confirmée par le lancement d'un référentiel de référencement « sémantique » : *Schema.org* et l'application *Knowledge Graph* de *Google*, dont l'objectif est de relier les entités du monde entier entre elles.

Par ailleurs et pour bien comprendre les mécanismes de fonctionnement et processus de référencement des moteurs de recherche, nous avons consulté l'ouvrage d'*Olivier Andrieu* « *Réussir son référencement web* » et de *Jean-Pierre F.Eskenazi* « *Référencement : comment référencer son site web* » et « *Référencement : comment se rendre visible sur internet* ». Ces deux experts du SEO (*Search Engine Optimization*) orientent leur analyse vers l'étude des critères et conditions de réussite d'un référencement naturel ou automatique. La question du « référencement sémantique » et de la visibilité des données culturelles n'est guère évoquée dans leurs ouvrages. Des mémoires de recherche français et étrangers ont à leur tour abordé la problématique de visibilité des données culturelles¹¹. Des blogs de professionnels dont celui de *Richard Wallis* : « *Liberate Data* », le « *Rapport final du groupe d'incubation : Bibliothèques et web de données* » au W3C et le site officiel du *Schema.org* ont fait partie de notre corpus. Le recours à cette documentation nous a permis de mieux cerner la question et ses différents enjeux.

L'analyse des réponses a dévoilé une concertation générale des interviewés sur la nécessité d'améliorer la visibilité des bibliothèques. Consultants, experts du web sémantique et professionnels des bibliothèques sont d'accord pour dire que les bibliothèques ratent leur objectif d'accès à l'information si on ne retrouve pas facilement leurs ressources et qu'on doit atteindre leurs par le biais d'un moteur de recherche. En ce qui concerne l'hypothèse du départ, nous pouvons confirmer que l'implémentation du web sémantique par les bibliothèques et les moteurs de recherche améliore la visibilité de la donnée culturelle. Certaines ressources parviennent à figurer en première page de résultats. Cela concerne surtout les ressources peu connues par les moteurs de recherche. Une requête sur « *Alexandre-Olivier Exquemelin* » permet d'afficher les ressources du *data.bnf.fr* en première page. Contrairement à une autre sur *Théophile Gautier* qui reste rivalisée par une masse informationnelle de données commerciales mises en avant par le *Knowledge Graph*. Cette situation nous mène à dire que la visibilité n'est pas totale mais partielle comme nous l'expliquent certains professionnels de bibliothèques. Par ailleurs nous infirmons l'hypothèse concernant le référencement sémantique de la donnée culturelle par les moteurs de recherche via *Schema.org*. Ce que nous confirment les professionnels c'est que quand bien

¹¹ Hügi Jasmin, P. N. (2014). *Les bibliothèques face au Linked Open Data : de nouvelles applications web et de nouvelles compétences professionnelles*. Genève: Haute École de Gestion de Genève.

même les bibliothèques intègrent *Schema.org* ce n'est pas pour autant qu'elles amélioreront leur référencement. La donnée n'est qu'un critère parmi tant d'autres qui aide à améliorer la visibilité et ne change pas le processus du référencement. Le défi de visibilité est relevé par les bibliothèques puisqu'on retrouve plus facilement leurs données sur le web. Leur absence partielle des premières pages de résultats n'infirmes pas pour autant leur évolution.

Dans notre processus d'analyse et de comparaison des entretiens aux sources étudiées, nous relevons des réflexions et des conclusions intéressantes. En ce qui concerne la question du référencement des données culturelles et des stratégies mises en place pour atteindre cette finalité, *Emmanuelle Bermès*¹² confirme l'emprunt de cette stratégie aux pratiques du *SEO* (*Search Engine Optimization*). Elle déclare que la finalité même du référencement c'est de faire « [...] remonter les services de la bibliothèque (documents, portail) dans le classement par pertinence des moteurs de recherche généralistes »¹³. Elle mentionne aussi d'autres stratégies pouvant améliorer la visibilité des bibliothèques sans recourir au référencement. Le recours aux réseaux sociaux du web lui semble être une excellente alternative. Les bibliothèques peuvent très bien valoriser leurs données à travers ces outils d'échange. Cette suggestion nous est confirmée par l'émergence d'un groupe de travail institutionnel visant la valorisation de la donnée culturelle à travers les technologies du web sémantique dans les réseaux sociaux. Il semble ainsi que le référencement et la stratégie de liens que voudraient mettre en place les bibliothèques ne sont en réalité que des pratiques empruntées au *SEO*. Elles nous sont bien expliquées par *Olivier Andrieu*, et *Jean-Pierre F.Eskenazi*. La stratégie des liens, de plus en plus évoquée par les bibliothèques comme principal fondement de visibilité et de substitution aux moteurs de recherche, est qualifiée par *Jean-Pierre F.Eskenazi*. comme étant du « marketing Indirect » :

« *Le marketing indirect : l'établissement de liens hypertextes étend la diffusion d'un site ; les internautes ont pour habitude de « surfer sur la vague ».* Ils se déplacent rapidement d'un site à l'autre via ces liens » [*Jean-Pierre F.Eskenazi*.1995]. D'autres emprunts tels que la notion du Graal (les mots-clés), les fautes de frappe et d'orthographe de l'internaute sont en train de s'ancrer dans les stratégies de visibilité des bibliothèques qui voudraient découvrir les mots-clés les plus utilisés par l'internaute pour atteindre leurs pages web. La notion du « Push » ou le fait de se mettre sur le chemin de l'internaute pour qu'il retrouve les ressources des bibliothèques, est une notion du *SEO*. Nous constatons ainsi un changement d'outillage et de méthodes de valorisation de la

¹² Emmanuelle Bermès, G. P. (2013). *Le web sémantique en bibliothèques*. Paris: Editions du Cercle de la Librairie.

¹³ *Ibid*

donnée culturelle fondés sur l'implémentation des technologies du web sémantique et l'usage des stratégies *SEO* des moteurs de recherche.

5. Conclusion

La visibilité des données des bibliothèques sont les plus grands enjeux et objectifs actuels des professionnels de l'information et des bibliothèques. Elle est primordiale pour la valorisation des ressources culturelles et principalement celles des bibliothèques. Sa mise en place génère un ensemble de problématiques aux quelles les bibliothèques n'étaient pas vraiment préparées. C'est une question qui reste en relation étroite avec les usages des internautes. Ce sont ces usagers du web qui sont en train de changer la donne de la recherche d'information. Ne parvenant pas à avoir des résultats satisfaisants, les internautes et les lecteurs des bibliothèques s'orientent de plus en plus vers des recherches éparées pour atteindre le document ou l'information de leur intérêt. Les usages de recherche, et les usages sociaux du web sont à l'origine des réflexions et débats sur la visibilité des données culturelles. De ce fait, on peut considérer les technologies du web sémantique comme un moyen technique à l'aboutissement de cette visibilité. Les données structurées, les liens sémantiques, la pérennisation de la donnée, le choix des jeux de données, leurs alignements avec de nouveaux jeux favoriseront la valorisation, l'interopérabilité, la sérendipité et la visibilité de la donnée culturelle sur le web.

Par ailleurs, les bibliothèques ne sont pas les seules bénéficiaires de cette visibilité. Les moteurs de recherche s'y intéressent aussi. Conscients de l'ampleur du « marché de la donnée et du web culturel », les géants du web commencent à mettre en place des dispositifs permettant le repérage et l'exploitation de cette donnée émergente. Ils préservent pour autant leur politique de référencement et privilégient toujours la donnée commerciale à celle de la culture. Toutefois, ils commencent à la prendre en compte dans leurs algorithmes de référencement. La politique de panachage du *Knowledge Graph* et sa volonté d'intégrer la donnée culturelle dans ses résultats de suggestion dénotent l'ampleur que commence à avoir l'institution culturelle dans l'écosystème du web.

Outre les aspects techniques et sociaux, la visibilité des données culturelles génère des rapports de force entre les grands acteurs du web comme *Google* et *OCLC*. Ces deux géants se confrontent pour avoir le monopole du marché de la donnée culturelle. Le moteur de recherche tente par son *Schema.org* de mettre la main sur la donnée des bibliothèques et des institutions culturelles. Il doit pour cela faire face à *OCLC* qui à son tour est en train de créer des modèles propriétaires de données visant leur diffusion dans l'ensemble des bibliothèques, L'objectif est clair : mettre la main sur le marché international de la donnée culturelle.

Bibliographie

Amerland, D. *Google Semantic Search*. USA: QUE, 2013.

Andrieu, O. *Réussir son référencement web*. Paris : Eyrolles, 2010.

Bermès, E. et al. *Le web sémantique en bibliothèques*. Paris : Collection Bibliothèques, éditions du Cercle de la Librairie, 2013.

Breton, Ph. *L'utopie de la communication : le mythe du « village planétaire »*. Paris : Editions la Découverte, 1995.

Breton, Ph. et al. *L'explosion de la communication : introduction aux théories et aux pratiques de la communication*. Montréal : Editions la Découverte, 2002.

Colliot, P-J. et al. *Référencement et visibilité web : de la stratégie à l'efficacité*. Paris : Collection Lire Agir, éditions Vuibert, 2012.

Eskenazi, J. *Référencement : comment référencer son site web*. Versailles : Webédition, 1999.

Eskenazi, J. *Référencement : comment se rendre visible sur internet*. Versailles : Webédition, 2001.

Proulx, S. et al. *Internet, une utopie limitée*. Québec, Canada : Les presses de l'université de Laval, 2005.

Salaün, J. *Vu Lu Su : les architectes de l'information face à l'oligopole du web*, Paris : La Découverte, 2012.