

Les instruments de recherche sont-ils adaptés à la demande des publics ? Le cas du public universitaire

Pierre Vernus

Citer ce document / Cite this document :

Vernus Pierre. Les instruments de recherche sont-ils adaptés à la demande des publics ? Le cas du public universitaire. In: La Gazette des archives, n°220, 2010. Les instruments de recherche : évolutions, publics et stratégies. pp. 133-141; http://www.persee.fr/doc/gazar_0016-5522_2010_num_220_4_4747

Document généré le 15/03/2017

Les instruments de recherche sont-ils adaptés à la demande des publics ? Le cas du public universitaire

Pierre VERNUS

Le thème de réflexion qui nous a été proposé est vaste tant en raison de la diversité des instruments de recherche – niveau de description, forme, support, diffusion, etc. – que de celle du public universitaire – enseignant-chercheurs de disciplines différentes, étudiants de la licence au doctorat, etc. Dans l'espace limité de cette contribution, il nous a donc semblé préférable de centrer notre propos sur la dynamique actuelle liée à l'essor des technologies de l'information et de la communication. Les effets de cet essor se font sentir quotidiennement dans le travail des chercheurs en histoire dont ils transforment peu à peu les pratiques, comme dans celui des archivistes ainsi que l'illustrent, entre autres manifestations, la multiplication des services disposant d'un site Web et parallèlement celle des instruments de recherche mis en ligne.

En effet, si elle ne modifie pas les exigences traditionnelles des chercheurs à l'égard des instruments de recherche, la diffusion des technologies de l'information et de la communication, à travers la généralisation de l'usage de l'ordinateur et du recours au Web, bouleverse l'environnement dans lequel s'exerce leur travail et encourage l'émergence de nouvelles pratiques susceptibles de modifier les modes de consultation des instruments de recherche, de donner naissance à de nouvelles attentes mais également d'ouvrir la voie à de nouvelles collaborations entre recherche universitaire et archives.

L'historien, l'ordinateur et le Web

Aujourd'hui, le chercheur dispose d'outils informatiques puissants qui, en contrepartie d'un minimum d'investissement technique pour les maîtriser, lui offrent des capacités de stockage, de traitement de l'information mais également de production de connaissances inconnues jusqu'alors.

L'essor de l'Internet représente un second aspect des mutations en cours, indissociablement lié au précédent et qui le renforce. Il offre aux chercheurs un accès à des quantités toujours croissantes de ressources digitales de toute nature : documents, catalogues et autres instruments de recherche, ouvrages, etc. ; il facilite les échanges scientifiques et favorise le travail collaboratif ; il constitue, enfin, un nouveau médium de présentation de leurs travaux¹.

L'apparition de nouvelles expressions telles que *digital history*, *e-history* ou *historical information science*² témoigne, au-delà des discussions sur la définition de ces notions (qui renvoient à des visions différentes des usages des technologies de l'information et de la communication et des rapports entre l'histoire et les sciences de l'information), que des communautés de chercheurs s'interrogent sur les effets des technologies de l'information et de la communication sur la manière dont ils mènent leurs recherches et en diffusent les résultats auprès de leurs pairs, de leurs étudiants ou d'un public plus large. Les mutations en cours ne peuvent donc être réduites à une simple adaptation des comportements ou des objectifs anciens des chercheurs à un nouvel environnement technologique. En effet, si certains usages des technologies de l'information et de la communication permettent d'améliorer des méthodes, des pratiques ou des techniques déjà en place, d'autres donnent naissance à des applications et à des recherches que seul un environnement numérique rend possibles. Il ne fait donc guère de doute que les mutations en cours et à venir modifieront les stratégies de recherche et de traitement de l'information, de production et d'utilisation des connaissances mises en œuvre par les chercheurs. On le voit actuellement avec l'édition électronique de sources manuscrites ou imprimées

¹ COHEN (Daniel J.) et ROSENZWEIG (Roy), *Digital History. A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Philadelphia, University of Pennsylvania Press, 2005 ; COHEN (Daniel J.), « History and the Second Decade on the Web », *Rethinking History*, 8-2, juin 2004, p. 293-301 ; SCHREIBMAN (Susan) *et alii*, *Companion to Digital Humanities*, Oxford, Blackwell, coll. « Blackwell Companions to Literature and Culture », 2004 ; BOONSTRA (Onno) *et alii*, *Past, present and future of historical information science*, Data Archiving and Networked Services, Den Haag, 2006.

² Voir McCRAW (Lawrence J.), *Historical information science*, Medford, New Jersey, Information Today Inc., 2001 ; BOONSTRA (Onno), *op. cit.*, p. 11-23.

ou la mise en place d'outils pédagogiques novateurs, actuellement deux des principales manifestations de la *digital history*.

D'ailleurs, comment l'essor des technologies de l'information et de la communication pourrait-il rester sans incidence sur le travail des historiens, dont une des activités fondamentales est, pour le dire vite, d'interroger des sources de natures variées pour en tirer des informations scientifiquement validées, c'est-à-dire dont l'élaboration répond à un certain nombre de critères reconnus par la communauté des pairs ? En attendant, il est évident que, dans sa majorité, la communauté des historiens ne tire pas encore pleinement partie des potentialités offertes par ces technologies. L'usage de certains outils (messagerie, navigateur Web, traitement de texte, tableur, système de gestion de bases de données, etc.) s'est généralisé mais souvent la réticence à fournir l'investissement intellectuel minimum nécessaire à un usage raisonné et raisonnable des outils disponibles conduit à des effets dommageables, quand ce n'est pas à des régressions méthodologiques. Il suffit d'évoquer le nombre de bases de données superficiellement exploitées ou auxquelles ne peuvent être appliqués d'autres questionnements que ceux initialement envisagés, faute d'avoir été structurées selon les règles fondamentales de la modélisation ou d'être accompagnées des métadonnées indispensables. Néanmoins, malgré ces limites, des pratiques différentes émergent. Quelles nouvelles attentes peuvent-elles engendrer à l'égard des instruments de recherche ?

Nouveau médium, nouvelles stratégies de consultation ?

Les services d'archives contribuent activement à l'accroissement des ressources numériques disponibles sur le Web. En partie pour des raisons compréhensibles liées à un souci de préservation des fonds dont l'intégrité matérielle était menacée par la vogue des recherches généalogiques, la priorité a été donnée à la mise en ligne de documents. En revanche, la numérisation des instruments de recherche est actuellement beaucoup moins avancée.

Le chercheur a les mêmes demandes fondamentales envers les instruments de recherche dématérialisés que celles qu'il adresse aux instruments imprimés. Il désire repérer rapidement quelles sont les ressources documentaires qu'il peut mobiliser pour mener la recherche qu'il entreprend, connaître leur importance quantitative et la nature des informations qu'il est susceptible d'y trouver. Il

attend également des renseignements éclairant le processus d'élaboration et de transmission des documents. De la mise en ligne des instruments de recherche, il espère d'importants progrès quantitatifs (gain de temps) et qualitatifs dans ses activités de repérage des sources utiles à ses travaux.

Les instruments de recherche mis en ligne se présentent, comme ceux sous format papier, sous des formats hétérogènes : textes, tableaux, images d'instruments de recherche imprimés, bases de données relationnelles notamment. Les publications d'instruments de recherche encodés en EAD sont encore rares. La modalité d'interrogation dominante est la recherche, générale ou avancée, plein texte ou par saisie de critères dans un ensemble de champs proposés. Elle offre une aide utile mais qui peut être également trompeuse puisque le résultat de la requête est dépendant des termes sélectionnés par l'archiviste ou présents dans les lexiques et thésaurus. Il convient donc que le lecteur puisse avoir accès à ces listes de termes afin de savoir, par exemple, si l'absence de réponse à sa requête résulte de l'absence d'articles concernant l'objet de sa recherche ou bien tout simplement parce que le ou les termes utilisés ne figuraient pas dans les thésaurus ou les listes utilisés pour l'indexation. Par ailleurs, il conviendrait d'indiquer systématiquement et clairement sur quels instruments de recherche disponibles sur le site s'effectue l'interrogation. Ces éléments d'information évitent que le processus d'interrogation se transforme en une sorte de boîte noire dont l'utilisateur n'est pas à même d'estimer les limites. Ces remarques valent avant tout pour les thésaurus matières ou thématiques dont le choix des termes n'est pas indépendant de la demande sociale ou de l'état de la recherche au moment où il a été confectionné. Le chercheur universitaire ne peut donc se contenter de ce type d'interrogation : il doit la compléter par la navigation dans la hiérarchie des différents niveaux de description des fonds susceptibles de l'intéresser, afin d'y déployer une stratégie de recherche restant proche de celle qu'il met en œuvre avec les instruments de recherche papier.

Une interrogation portant sur les fonctions ou compétences des producteurs constituerait une avancée intéressante puisqu'elle permettrait notamment de regrouper les descriptions de documents concernant une même attribution mais dispersées entre différents fonds ou différentes séries. En effet, à l'orée d'une recherche, l'historien doit se poser la question fondamentale : « quels producteurs d'archives possédaient les compétences dont relève l'objet de mon étude ? ». Cette lacune pourrait être assez rapidement comblée avec la mise en œuvre de la norme ICA-ISDF.

L'application des normes archivistiques internationales, la diffusion de la pratique du balisage des instruments de recherche selon la DTD-EAD, le schéma EAC-CPF et, espérons-le bientôt, celui dédié aux fonctions ainsi que la conversion progressive au format XML-EAD des inventaires issus de traitements de texte, de tableurs ou en PDF permettront également de tirer le meilleur profit des fonctionnalités de plus en plus puissantes des applications Web¹. On peut également espérer qu'elles offriront rapidement au chercheur la possibilité d'alléger son travail de collecte et de copie des références documentaires. Il existe déjà des outils permettant la gestion des références bibliographiques rassemblées et leur échange grâce à des formats standards (bibTeX, Dublin Core, MODS par exemple) mais ils ne concernent pas encore les références archivistiques. Parmi ces outils, citons Zotero² qui connaît un grand succès parmi les chercheurs en sciences humaines mais également au-delà. Logiciel libre et open-source, développé par le *Center for History and New Media de la George Mason University* à Washington, il permet de collecter automatiquement des références documentaires en leur associant éventuellement des documents (un fichier pdf par exemple), de les organiser, de les classer, de les marquer avec des mots clés, de les annoter en leur attachant des notes et, enfin, de les intégrer directement dans un texte au moment de la rédaction sous Open Office ou Word. Pourquoi ne pas imaginer la création de filtres autorisant le moissonnage des cotes d'archives et permettant au chercheur de réaliser rapidement un état des sources disponibles en relation avec son thème de recherche ? Notons que la dernière version du logiciel permet de partager ses bibliographies afin de favoriser le travail collaboratif et les contacts entre chercheurs intéressés par les mêmes thèmes.

Mise en ligne, mise en réseau et partenariats

L'élaboration d'instruments de recherche normalisés et structurés sur les bases du formalisme XML devrait permettre aux historiens et aux archivistes de tirer profit des grandes potentialités du Web dans le domaine de l'interopérabilité et

¹ Voir les réflexions de Yann Nicolas à partir de l'expérience de Calames, catalogue en ligne des archives et manuscrits de l'enseignement supérieur : « Calames, et après ? », *Bulletin des bibliothèques de France*, t. 53, n° 6, p. 29-33.

² <<http://www.zotero.org>> ; voir aussi le blog francophone qui lui est consacré : <<http://zotero.hypotheses.org>> (ces sites ont été consultés en février 2010).

des développements du Web sémantique. Parallèlement, les nouvelles technologies de l'information et de la communication ouvrent de nouvelles voies de collaboration entre le monde des archives et celui de la recherche en histoire.

Aujourd'hui, le stockage et l'accès effectif à des données en quantité toujours croissante ne posent plus de problèmes insolubles : globalement, le défi technique des années 1960-1970, celui de la maîtrise des systèmes d'information, a été surmonté. Le nouveau défi concerne l'organisation et la maîtrise des masses de données disponibles, l'extraction de connaissances de bases de données complexes (c'est-à-dire souvent non structurées car stockées dans leur format d'acquisition natif et plus ou moins indexées), distribuées (stockées dans des bases de données éloignées, accessibles en temps réel) et hétérogènes (images, textes, fichiers audio ou vidéo, tableaux de chiffres, courbes, etc.). Les recherches dans le domaine des entrepôts de données ainsi que de la fouille ou extraction de données sont actives.

Parallèlement, on assiste à l'essor des technologies tournant autour du Web sémantique ou encore Web 3.0. Prolongement du Web actuel, il vise à rendre le contenu des ressources disponibles en ligne accessible et utilisable par les machines grâce à un système de métadonnées exprimées en utilisant notamment la famille des formalismes développés par le W3C (par exemple RDF¹). En d'autres termes, il s'agit de décrire le contenu d'une ressource présente sur le Web ou plus largement tous les objets ou individus pouvant y être identifiés et de rendre compréhensible cette description par les machines informatiques, afin qu'elles puissent faire une grande partie du travail de recherche, de partage et de regroupement de l'information.

Des avancées rapides sont à prévoir dans ces domaines, en relation notamment avec l'usage croissant du langage XML, d'autant que les recherches sont encouragées par les implications qu'elles pourraient avoir pour les entreprises dans les domaines décisionnel ou commercial.

Ces nouvelles technologies, permettant déjà à des serveurs éloignés de répondre ensemble à des requêtes, rendent envisageable la mise en place d'un portail à partir duquel, à terme, il serait possible d'interroger d'un même mouvement l'ensemble des instruments de recherche élaborés par le réseau des archives publiques, auquel pourraient être associés d'autres centres. Parallèlement, ce portail pourrait donner un accès centralisé à un certain nombre de ressources dont l'élaboration, vue de l'extérieur, paraît générer une

¹ <<http://www.w3.org/TR/rdf-primer/>> (consulté en janvier 2010).

duplication des efforts : les aides à la recherche concernant certains types de documents, comme le cadastre ou l'enregistrement, pourraient être mutualisées, chaque service proposant pour sa part un complément précisant les spécificités des fonds qu'il détient. Le chercheur engageant une recherche sur un sujet de portée nationale économiserait ainsi un temps considérable. Cet outil constituerait un progrès dans le sens d'une homogénéisation des modules d'interrogation dont la grande diversité d'affichage déroute quelque peu le chercheur que ses travaux amènent à passer d'un site à l'autre.

Mais la mise en ligne des instruments de recherche impose de réfléchir à leur connexion avec des ressources extérieures. Pour ce faire, il faut rappeler que les instruments de recherche ne sont pas seulement des descriptions du contenu des fonds mais également des bases de connaissances, plus ou moins riches selon leur nature et leur niveau de description. Ils contiennent dans la zone du contenu, dans celle du contexte de production ou dans les annexes, un ensemble d'affirmations telles que : une entreprise X a déposé telle année un acte au tribunal de commerce de Y ; un mouvement de grève a touché l'établissement Z à telle date ; tel service a été créé à telle date, supprimé à telle autre date, etc. Chacune de ces affirmations représente une information historique atomisée relative à une ou plusieurs entités historiques. Les instruments de recherche évoquent donc aussi des personnes, des institutions, des concepts, des lieux, des événements, etc. D'autres informations sur ces entités peuvent figurer dans plusieurs instruments de recherche d'un même service mais également dans ceux d'autres services et au-delà dans d'autres ressources (catalogues de bibliothèque, base de données d'un laboratoire de recherche, etc.)¹.

Ce constat peut être rapproché des réflexions menées par certains historiens sur la modélisation des données historiques. S'il n'existe pas encore de modélisation générique des informations historiques², différentes propositions tendent à converger vers une position caractérisant l'information historique par un libellé, une date de début, une date de fin (dates éventuellement approximatives ou incertaines). Elle met en relation des entités (personnes, acteurs collectifs, concepts, lieux, objets matériels, etc.). Pour être valide, elle doit être reliée à des références archivistiques ou bibliographiques. Enfin, elle doit pouvoir être localisée. Dans un contexte de base de données relationnelle, les différentes informations correspondent aux lignes d'une table et sont liées aux entités concernées par une table de jonction. Ainsi décomposées et

¹ Voir NICOLAS (Yann), *art. cit.*

² BOONSTRA (Onno) et *alii*, *op. cit.*, p. 101-102.

insérées dans un réseau de relations, les informations historiques concernant un individu, un événement, un lieu, etc., particulier peuvent être agrégées et organisées chronologiquement par des requêtes en fonction des besoins. De la sorte, il est possible de « reconstruire » la fiche biographique d'un individu ou l'histoire d'une institution. Le Système Modulaire de Gestion de l'Information Historique (SyMoGIH), développé au sein du pôle méthodes du laboratoire de recherche historique Rhône-Alpes et actuellement utilisé par environ vingt-cinq chercheurs ou étudiants, représente un exemple de mise en œuvre de ces principes¹. Dans le cadre de ce système, un ensemble de dictionnaires communs a été mis en place après modélisation des entités concernées (acteurs individuels, acteurs collectifs, lieux, caractères sociaux, etc.). Chaque instance d'un dictionnaire (par exemple un individu, pour le dictionnaire des acteurs individuels ou une entreprise pour celui des acteurs collectifs) est affectée d'un identifiant unique. Elle est désignée par un nom standard mais un nombre non limité de formes autorisées peut également lui être associé. Les recherches se font sur l'ensemble de ces formes. Cet identifiant est utilisé dans une opération de codage permettant de lier une information historique avec les entités qu'elle concerne. Les liens entre les dictionnaires communs et d'autres référentiels sont prévus.

On constate qu'une telle démarche se rapproche étroitement de celle conduite par les archivistes désireux de rendre possible une interrogation combinée dans un ensemble d'instruments de recherche. Plutôt que de voir historiens et archivistes mener leurs réflexions de manière dispersée et d'établir des référentiels de portée limitée, ne serait-il pas envisageable de faire converger leurs efforts, de développer des projets de rapprochement des dictionnaires communs, qui autoriseraient à terme la juxtaposition d'informations dispersées concernant une même entité historique ? Les historiens pourraient ainsi bénéficier de l'agrégation des informations historiques dispersées dans les instruments de recherche et actuellement sous-utilisées ou ignorées et, de leur

¹ Voir : <http://larhra.ish-lyon.cnrs.fr/Pole_Methodes/SyMoGIH_fr.php> (mars 2010) ; SyMoGIH est un système de stockage de l'information historique ouvert et indépendant de la problématique. Il permet d'inscrire les recherches, y compris individuelles, dans un cadre collectif permettant un travail cumulatif. Ce stockage permet ensuite l'extraction de données en vue de leur traitement avec différents outils d'analyse (analyse des réseaux, cartographie, statistiques, etc.). Le projet Système d'information sur les patrons et le patronat français (SIPPAF) financé par l'ANR s'appuie sur les principes de SyMoGIH. Une partie des données est publiée sur le site Patrons de France (<http://sippaf.ish-lyon.cnrs.fr/index_fr.php>).

Une modélisation proche a été proposée par le *Historical Event Markup and Linking Project* (HEML) : <<http://heml.mta.ca/samples/blocks/heml/>> (consulté en mars 2010).

côté, les archivistes bénéficieraient des apports de la recherche quand, lors des opérations de classement ou de rédaction des instruments de recherche, ils doivent identifier des acteurs ou des institutions ou dater des événements. Notons au passage que la conception traditionnelle de la notice d'autorité nécessite d'être reconsidérée dans une perspective différente de celle qui prévaut actuellement dans le monde des bibliothèques. Elle devrait plutôt être conçue comme une agrégation d'informations autour d'un identifiant, susceptible d'être constamment enrichie et complétée.

Une collaboration du type de celle qui vient d'être envisagée nécessiterait évidemment d'être précédée d'une réflexion sur les modalités concrètes de sa mise en œuvre, mais elle est compatible avec les ressources technologiques actuelles. En favorisant un accroissement collectif et cumulatif de l'information historique scientifiquement validée et un enrichissement continu des dictionnaires communs, elle constituerait un progrès considérable car, dans ce domaine, le résultat est supérieur à la simple somme des parties.

La dématérialisation des instruments de recherche, leur normalisation et leur mise en ligne sont donc susceptibles de créer de nouvelles attentes de la part des chercheurs. Mais, par ailleurs, les technologies de l'information et de la communication, la normalisation et la mise en œuvre de formalismes, notamment XML, dont l'usage est bien avancé du côté des archivistes et, on l'espère, le sera de plus en plus du côté des historiens, ouvrent de nouvelles voies de partenariats, à condition cependant que les instruments de recherche ne soient plus perçus dans une perspective autarcique pour les premiers ou consumériste pour les seconds, mais plutôt comme de riches sources d'informations scientifiquement validées et insérées dans le vaste ensemble des ressources disponibles sur le Web.

Pierre VERNUS
Maître de conférences en histoire contemporaine
Université de Lyon
LARHRA UMR 5190
pierre.vernus@ish-lyon.cnrs.fr