

## « Les Mardis des sciences de l'homme »

Mardi 15 février 2005

« Les bibliothèques virtuelles, le savoir à portée de tous? »

Une réflexion sur les perspectives ouvertes par la numérisation du savoir.

Les invités:

- Frédérique Joannic-Seta : Bibliothèque Nationale de France, responsable scientifique de *Gallica*, responsable des pôles associés (réseau de bibliothèques en France).
- Laurent Romary : Chargé de mission auprès du directeur du CNRS pour l'information scientifique et technique, directeur de recherche au LORIA/INRIA (Institut National de Recherche en Informatique et Automatique).
- Alain Giffard : Président de l'association Alphabétville, ancien directeur de la Mission Interministérielle pour l'Accès Public à la micro-informatique, à l'Internet et au multimédia, ancien directeur informatique de la Bibliothèque Nationale de France.

Les intervenants de la MSH Paris Nord :

- Michel Porchet, coordonnateur du thème 4 (Esthétiques, Arts et Industries) de l'axe 1 (Industries de la culture et Arts), ingénieur en information, professeur de l'Ecole Polytechnique Fédérale de Lausanne.
- Henri Hudrisier (thème 4), professeur en Documentation à l'université Paris 8, membre des laboratoires LEDEN et Paragraphe, pilote du projet MSH Paris Nord « Fabula Meda » (Text Encoding Initiative et information structurée).
- Ouahmi Ould Braham, chef de projet Multimédia Berbère, directeur de la revue *Etudes et Documents Berbères*.
- Gabriel Popovici, responsable du centre de documentation de la MSH Paris Nord.

Compte-rendu:

Mardi 15 février 2005, la Maison des Sciences de l'Homme Paris Nord organisait une conférence-débat sur les bibliothèques virtuelles. Sujet d'actualité puisque quelques jours auparavant, Jean-Noël Jeanneney, directeur de la Bibliothèque Nationale de France, avait attiré l'attention du grand public, dans les colonnes du journal *Le Monde*, sur les dangers économiques et culturels que pouvait représenter le projet de bibliothèque virtuelle de « Google » (*Le Monde*, 23 janvier et 5 mars 2005).

L'objet de cette conférence était de réfléchir sur le décalage qui peut exister entre le mythe et la réalité: les bibliothèques virtuelles réaliseront-elles l'utopie d'un savoir universel accessible à tous de manière immédiate? Ou bien seront-elles un instrument de domination culturelle, sélectionnant le savoir mis à disposition selon des critères qui n'auront rien de transparent? L'enjeu serait alors de veiller à ce que certains documents ne deviennent pas, comme le disait Michel Porchet, « encore plus invisibles » qu'ils ne sont déjà.

L'initiative de cette rencontre avait été prise par trois chercheurs du thème 4 (Esthétiques, Arts et Industries) de l'axe 1 (Industries de la culture et Arts) de la Maison des Sciences de l'Homme Paris Nord: Michel Porchet, Henri Hudrisier et Ouahmi Ould Braham – ce dernier travaillant lui-même à l'élaboration d'une bibliothèque virtuelle sur la langue et la culture berbère. Le thème 4 est en effet l'occasion, pour ces chercheurs, de mener une réflexion commune sur la mise à disposition du savoir, en particulier dans le cadre de la communication entre chercheurs ainsi qu'entre enseignants et étudiants (projet « Fabula Meda »). A cette équipe s'était joint Gabriel Popovici, responsable du centre de documentation de la MSH Paris Nord.

Dans son introduction, Michel Porchet a précisé les enjeux de cette réflexion sur les bibliothèques virtuelles. Il s'agit d'attirer l'attention du public sur l'importance des techniques dans les industries culturelles et les arts. Les techniques en effet ont une influence directe sur les activités des chercheurs, qui ne produisent pas du savoir dans une tour d'ivoire coupée du monde, mais sont largement tributaires des instruments qu'ils utilisent pour produire et diffuser ce savoir, qu'il s'agisse, dans leur travail personnel, des conditions d'accès aux documents, de la communication dans les travaux collectifs – traditionnellement assez rares en sciences humaines – ou des conditions de diffusion et de validation scientifique de leurs travaux. C'est le but du projet « Fabula Meda », piloté à la MSH Paris Nord par Henri Hudrisier, que d'expérimenter dans le milieu des chercheurs de la MSH les effets de ces technologies.

Le problème central est celui de la standardisation, c'est-à-dire de la mise au point de normes collectives qui permettent l'archivage et la diffusion en ligne de manière efficace. L'établissement de ces normes influence en retour la manière dont le savoir est produit à l'origine, le chercheur étant tributaire de la mise en forme standardisée de son travail. Les conditions techniques d'archivage et de réception déterminent les conditions de production du savoir. Les bibliothèques virtuelles s'inscrivent dans un processus d'industrialisation de la production

intellectuelle. Mais, comme pour la production industrielle, les nouvelles technologies permettent de fusionner des activités disjointes, en l'occurrence, la chaîne d'autorité, la chaîne éditoriale, la chaîne documentaire: avec la numérisation, un texte est dans un même geste mis en forme et mis en ligne. Un certain nombre de redondances sont ainsi éliminées, ce qui, d'un point de vue économique, peut paraître un progrès. Mais c'est méconnaître l'utilité que peuvent avoir les redondances dans la production, et dans la production intellectuelle en particulier. Ainsi, il y a une différence entre un texte numérisé que l'on trouve en cherchant sur « Google », et un texte issu d'une équipe de recherche, validé par un conseil scientifique, publié par un éditeur reconnu, par exemple, dans un certain milieu professionnel, et indexé par une bibliothèque elle-même identifiée comme une bibliothèque de référence. Il ne s'agit donc pas d'un simple transfert de support, en passant par scanner du papier au numérique. Il y a une illusion du scanner. Car il s'agit en fait de passer d'une chaîne globale écriture/édition/documentation, vers une autre chaîne globale, fondée sur un autre support, avec une autre histoire, héritière, en particulier, de la tradition d'indexation de la nomenclature industrielle du *Computer Integrated Manufacturing*: les bibliothèques virtuelles se situent ainsi au carrefour entre la tradition industrielle et la tradition documentaire.

Les trois exposés des invités ont permis d'avoir des points de vue variés et complémentaires sur ces questions.

Frédérique Joannic-Seta d'abord, responsable scientifique de *Gallica*, a proposé une présentation générale de la bibliothèque virtuelle de la Bibliothèque Nationale de France. Son exposé illustrait particulièrement clairement l'idée que le but de *Gallica* était moins de prétendre à une exhaustivité illusoire, ou à une simple volonté de préserver les documents fragiles, que d'offrir aux chercheurs des ensembles cohérents sur des « dossiers » précis (par exemple 250 titres sur les journaux des tranchées, ou la numérisation de la presse du 19<sup>e</sup> siècle), réunissant une masse critique de documents qui auront donné lieu à un travail de balisage et d'indexation en amont de leur numérisation, réalisé par des chercheurs spécialisés dans les domaines concernés (par exemple le Centre de Recherche sur la Renaissance de Tours, ou le réseau Math-Doc, ou encore le projet Multimédia berbère de Ouahmi Ould-Braham à la MSH Paris Nord). Il ne s'agit pas seulement d'un choix économique – puisque, au fond, ce n'est pas le plus économique – mais d'une volonté d'offrir aux utilisateurs des documents numériques de qualité (qui ont vocation, à terme, à être accessibles aussi en mode texte, et pas seulement en mode image), permettant une navigation qui fournisse à la fois des visions globales des documents (« comme dans les rayonnages d'une bibliothèque ») et des investigations en profondeur dans les documents.

Laurent Romary, nouvellement nommé chargé de mission auprès du directeur du CNRS pour l'information scientifique et technique, a ensuite entraîné les auditeurs dans des considérations plus techniques (« après le pavillon-témoin, la maçonnerie ») concernant les normes informatiques mises en oeuvre pour numériser les documents. Où l'on comprenait à quel point ces questions ne sont pas neutres, car ces normes déterminent les moyens d'accès aux documents et les types de recherche auxquels ils peuvent donner lieu. Il ne s'agit pas seulement de partager des ressources, mais aussi de partager les outils qui permettent l'accès, la transcription, l'indexation, la manipulation de ces ressources, et de partager les pratiques professionnelles à l'occasion desquelles ces outils ont été mis au point, testés et affinés. D'où la nécessité de bases de normalisation internationales, permettant de stabiliser ces compétences technologiques. On peut distinguer ainsi deux étapes dans ce processus: celle de la « standardisation », au cours de laquelle des chercheurs définissent entre eux une charte commune; et celle de la « normalisation », par laquelle ces standards sont officialisés et pérennisés, via des organismes internationaux dont c'est la fonction. Ainsi par exemple l'ISO, qui réunit des instances nationales (dont l'Afnor est le représentant français), qui permet de garantir la pérennité de la norme, son caractère consensuel, et qui oeuvre pour diffuser ces résultats.

Cette normalisation se déroule à son tour selon deux axes: un axe horizontal (par exemple celui des normes XML) qui permet la traduction uniforme des informations, quelle que soit leur contenu. Et un axe vertical, correspondant à la gestion des informations le long d'une application particulière.

L'ISO voisine avec d'autres instances comme le W3C ou la TEI (Text Encoding Initiative) ou l'ATIF de Nancy.

A l'horizon de ce travail de normalisation, il n'y a pas seulement la production d'un jeu de balises, mais une véritable description d'objets sémantiques qui permettrait de faire correspondre par exemple le concept d'« auteur » de la BNF à celui d'« author » de la Bibliothèque du Congrès. C'est-à-dire établir un véritable registre de métadonnées qui constitue un point de référence international des ontologies sémantiques – une sorte d'« alphabet des pensées humaines », sou de « caractéristique universelle », selon le rêve de Leibniz.

Il ne s'agit pas que tous les chercheurs réinventent les normes, ou deviennent spécialistes de ces questions, mais aucun chercheur aujourd'hui ne devrait être ignorant des technologies qui permettent à son domaine de recherche de vivre, d'autant qu'en dernière instance, ce sont les chercheurs eux-mêmes qui produisent les normes propres de leurs domaines.

Alain Giffard, pour finir, a élargi le questionnement, en réfléchissant sur la dépendance des conditions de production des documents par rapport à la question de la lecture. « Une bibliothèque, dit-il, a une dépendance par rapport au support, mais sa dépendance principale, c'est par rapport à la lecture. A un moment donné, une certaine forme de lecture est constituées historiquement, culturellement, et elle s'applique sur des supports qui doivent s'adapter. Si ce sont les supports d'avant, on va les lire différemment, et ça ne collera pas très bien. Donc on aura des supports qui seront différents. Et puis si les supports ne collent pas, la bibliothèque, elle-même, va faire les supports qui collent, et si elle n'a pas les livres, elle va les faire. A Alexandrie, ce ne sont pas les textes qu'on va chercher et qu'on

*met dans la bibliothèque. On cherche des textes, on les édite. Et toujours la bibliothèque a eu un rapport autonome, visant le texte, par rapport au support. La Bibliothèque Nationale a toujours fait des microfiches, de la numérisation, elle a toujours eu sa propre manière matérielle de traiter cette question. (...) La bibliothèque est un ordre de livres, et il faut comprendre comment celui-ci correspond à un ordre de la lecture. »*

L'idée est que le problème n'est pas seulement celui de l'évolution de la technologie, mais celui de l'usage. On pourrait ainsi distinguer, suggère Alain Giffard, la bibliothèque « numérique » et la bibliothèque « virtuelle ». La première consiste en l'informatisation de la bibliothèque classique, accompagnée de la numérisation des textes. La seconde rajoute à ce dispositif le réseau et l'appropriation individuelle (dans le cadre, en particulier du « poste de lecture assisté par ordinateur »). La question posée par le projet de « Google » est celle du croisement de ces deux types de bibliothèque, dont Giffard propose une analyse selon quatre paramètres (l'industrie, la référence, les pratiques, le sujet), à propos desquels il convient de renverser les perspectives habituelles.

On comprendra ainsi, en suivant Giffard, que le lecteur joue à la fois un rôle dans la production de nouveaux objets de lecture (les listes de liens qu'il crée à mesure de sa lecture-écriture), mais aussi un rôle dans la création de réseaux de lecteurs. Et c'est pour cela que, si une alliance entre « Google » et la BNF fait courir des risques à cette dernière, celle-ci devrait orienter sa stratégie vers les lecteurs, car tôt ou tard, « *il faudra reconnaître la collectivité des lecteurs numériques* ». C'est parce que le lecteur est premier que la bibliothèque virtuelle qui aura su reconnaître ses besoins sera la véritable bibliothèque virtuelle.

[Le texte préparatoire de l'exposé d'Alain Giffard est lisible en ligne sur: <http://alaingiffard.blogs.com/culture> ]

Compte-rendu établi par Emmanuel Pasquier, février 2005.

L'enregistrement intégral de ces communications et de la discussion qui leur a fait suite est disponible sur CD, sur simple demande.

Contact/organisation des Mardis des Sciences de l'Homme

Emmanuel Pasquier : [epasquier@mshparisnord.org](mailto:epasquier@mshparisnord.org)

01.55.93.93.06

---

### **Bibliothèques virtuelles : quelques liens utiles:**

- Gallica: <http://www.gallica.bnf.fr> la bibliothèque virtuelle de la Bibliothèque Nationale de France.
- Text Encoding Initiative: <http://www.tei-c.org>
- En français une introduction sur le site des cahiers gutenbergs: <http://www.gutenberg.eu.org/publications/autres/TEILITE>
- Open Archives Initiative: <http://www.openarchives.org> Un système ouvert d'archivage de projets.
- Portail CNRS, bibliothèque numérique sur la recherche en Technologies de l'Information et de la Communication: <http://www.tematice.fr>
- Sciences humaines: [http://www.uqac.quebec.ca/zone30/Classiques\\_des\\_sciences\\_sociales](http://www.uqac.quebec.ca/zone30/Classiques_des_sciences_sociales)
- <http://rtp-doc.enssib.fr> site participatif du CNRS sur le thème « documents et contenu: création, indexation, navigation ».
- LORIA, Laboratoire lorrain de recherche en informatique et en automatique: <http://www.loria.fr>
- Sur les normes: l'ISO (Organisation Internationale de Normalisation): <http://www.iso/tc37/sc4>
- Alphabétville et blog d'Alain Giffard: <http://alaingiffard.blogs.com>
- Bibliothèque virtuelle de langue et civilisation berbère: <http://bereberemultimedia.fr>