

PRÉFACE

Systèmes d'organisation des connaissances et humanités numériques

Le tournant des humanités numériques (HN) valait bien qu'un colloque de l'ISKO y soit consacré. Le dixième colloque de l'ISKO-France a eu lieu à l'université de Strasbourg les 5-6 novembre 2015. Les contributions réunies dans ce volume donnent accès aux contenus présentés lors du colloque intitulé « Systèmes d'organisation des connaissances et humanités numériques ». Le choix de cette thématique des humanités numériques est aussi opportune que pertinente : opportune car les humanités numériques en tant qu'interdiscipline connaissent un développement important et constant, pertinent car elle répond aux préoccupations d'un chapitre national d'une grande société savante comme ISKO (*International Society for Knowledge Organization*).

Les humanités numériques sont une thématique scientifique dynamique, florissante et plurielle ; elles réunissent des thématiques disciplinaires allant de l'informatique au patrimoine en passant par les sciences humaines et sociales. Elles impulsent un déplacement de frontières comme diraient Mounier et Darcos : frontières entre les disciplines, dont le recours commun aux instruments informatiques peut faciliter le rapprochement ; et frontières entre chercheurs en sciences humaines et ingénieurs, puisqu'un dialogue fécond se noue, entre les enjeux scientifiques et les possibilités techniques. Les humanités numériques se trouvent ainsi à l'intersection des sciences et technologies informatiques et des sciences humaines et sociales. La large disponibilité des ressources d'information numérique et des installations informatiques en réseau, couplée avec des systèmes de communication, de partage collaboratifs a conduit au développement de ce nouveau paradigme de recherche généralement décrit comme *e-science*, *e-recherche* ou encore *e-humanités*. Il prône une recherche collaborative, entre chercheurs de différentes institutions et qui est largement répandue géographiquement. Les normes

et standards de description de métadonnées, leur stockage, les formats d'échange de données y apparaissent comme essentiels pour permettre le partage des ressources et des données, tout comme l'accès ouvert à ces données. Les systèmes qui prennent en charge ces formes de travail sont parfois appelés «°*co-laboratories*°» ou «°*environnements virtuels de recherche*°» ou infrastructures de recherche ou cyber-infrastructures.

Le terme « humanités numériques », censé être l'équivalent du terme anglais «°*digital humanities*°», est attribué à John Unsworth, Susan Schreibman et Ray Siemens avec l'édition de *A Companion to Digital Humanities* en 2004 désignant à la fois l'utilisation des SHS pour étudier les technologies numériques et inversement les pratiques numériques et l'utilisation de ces technologies pour étudier les objets culturels et sociaux. Ce tournant numérique (*computational turn*) remonte cependant au milieu du XX^e siècle lorsque les sciences humaines (*humanities computing*), en particulier les linguistes, les historiens et les sociologues s'emparent de l'informatique pour traiter les données dans leurs travaux de recherche. Simultanément au développement de la sociologie quantitative, ce sont surtout les chercheurs en linguistique et les spécialistes de l'analyse des textes qui ont été parmi les premiers à utiliser les potentialités de l'informatique pour mener leurs recherches. Le Père Roberto Busa s.j. (1913-2011) est ainsi considéré comme un des précurseurs des humanités numériques lorsqu'il a lancé l'*Index Thomisticus* à la fin des années 1940. Alors qu'il travaillait à l'édition des œuvres complètes de saint Thomas d'Aquin, le père Busa entreprend avec l'aide des équipes d'IBM l'élaboration d'un index réunissant tous les termes du corpus, soit quelques 11 millions de mots. Concernant les objectifs de l'organisation des connaissances ou l'organisation de l'information le travail du père Busa comme, l'avait indiqué Julianne Nyhan ici-même, pourrait être considéré comme un travail précurseur. Un lien intéressant a été établi dans sa conférence entre l'index du père Busa censé aider à comprendre l'œuvre de Saint Thomas d'Aquin et la définition même des concepts d'organisation de l'information ou l'organisation de l'information analysé par Bawden et Robinson. Le sujet est appelé soit «°organisation de l'information°» ou «°organisation de la connaissance°». Mais cette définition nous rappelle que l'objectif peut être soit la compréhension de la structure de la connaissance elle-même soit le but pragmatique de disposer de documents; ces documents, initialement physiques sur des étagères, sont devenus récemment des documents numériques dans un espace virtuel.

Les humanités numériques ont eu une histoire assez intéressante. Elles ont commencé par ce qu'on identifiait dès les années 60-90 comme «°l'informatique pour les sciences humaines» (*Humanities Computing*). Ces pratiques de l'outil informatique ont été souvent considérées comme un support technique aux travaux de «°vrais°» chercheurs en sciences humaines. Cela impliquait l'application du calcul informatique à l'étude et à l'analyse des contenus disciplinaires relevant des

sciences humaines et surtout la linguistique, ce qui a donné naissance plus tard à l'informatique linguistique. Un autre moment fort dans l'histoire des HN est l'exploitation des technologies informatiques pour la constitution de vastes corpus de textes. Le *Brown Corpus*, lancé dans les années 1960 par Henry Kurcea et Nelson Francis à l'université de Brown aux États-Unis, ont ainsi proposé un corpus de référence pour la langue américaine. Ce mouvement est connu dans les années 90 sous le label de *linguistique de corpus*.

Une analyse de la littérature montre comment l'introduction de la technologie informatique dans les recherches en sciences humaines est en train de changer la façon dont nous pouvons transmettre, comprendre, questionner et imaginer tous les aspects de la connaissance humaine. L'une des transformations les plus profondes liée au numérique est la modification des conditions d'exercice et de production du savoir. C'est un nouveau territoire qui favorise la collaboration, l'ouverture et la valorisation des connaissances mais qui est aussi l'objet d'appropriations privées notamment des métadonnées qui sont constituées de rivalités entre institutions et entreprises.

L'essor des nouvelles technologies a également eu un profond impact sur le plan scientifique. Les conditions mêmes de la production et de la diffusion de la recherche scientifique en sont considérablement modifiées. Il a, en effet, permis de transformer les modalités de construction des savoirs, la définition des objets de recherche et les questionnements, les méthodes d'analyse, qu'elles soient qualitatives ou quantitatives, ainsi que les procédés et les pratiques de la recherche scientifique et de la diffusion de ses résultats.

La thématique du colloque, mérite qu'on revienne sur les liens entre systèmes d'organisation des connaissances et HN. Je voudrais commencer par le lien le plus ancien, c'est-à-dire le rapport de ce domaine émergent aux infrastructures de recherche pour les humanités. Moulin fait remonter la naissance des HN modernes à l'histoire des infrastructures de recherche. Dès le 3^{ème} siècle avant JC, l'impératif de recueillir, d'organiser et de conserver des savoirs pour le service de l'avancement des connaissances a donné naissance au premier «°Centre d'information°» sous la forme du *Mouseion*, un centre culturel, universitaire et une bibliothèque fondée à Alexandrie (Egypte) par les successeurs d'Alexandre le Grand. Les conséquences positives de cette entreprise ambitieuse sont vite devenues évidentes, par la production de diverses infrastructures de recherche sous forme de grammaires et de lexiques, ancêtres des infrastructures actuelles pour les ressources linguistiques telles que CLARIN, élément de la galaxie de DARIAH EU (*Digital Research Infrastructure for the Arts and Humanities*). Comme l'avait noté O. Olesen-Bagneux, on en sait très peu à propos de l'ancienne bibliothèque d'Alexandrie. Certaines sources indiquent qu'elle aurait pu contenir entre 40 et 700 000 rouleaux de manuscrits. La bibliothèque faisait partie d'une institution religieuse, le

Mouseion, et les savants étaient en fait des esclaves qualifiés qui y avaient été emprisonnés. Toute tentative d'évasion pouvait être punie de mort et une partie de la poésie écrite par ces chercheurs sous les verrous a été réalisée au cours de cérémonies religieuses.

Ce succès précoce des infrastructures pour la recherche en sciences humaines (si l'on ose s'exprimer ainsi) ne fut pas perdu dans la période qui suivit la fin de l'Empire romain ; il a plutôt inspiré une série d'activités qui devaient favoriser un système émergent de connaissances et faire remarquer l'importance de leur collecte, de leur organisation et de leur conservation au service du développement scientifique ultérieure, par exemple lors de la création des bibliothèques médiévales et, plus tard, lors de la formation des collections d'art. Ces «°bases de données°» précoces ont fourni le matériau de base pour les phases ultérieures des infrastructures de recherche pour les sciences humaines.

Il convient de rappeler un lien bien connu entre les HN et les sciences de l'information, discipline qui sous-tend l'organisation des connaissances et par conséquent des bibliothèques numériques. Ces dernières jouent un grand rôle dans les HN, surtout dans le monde anglo-saxon ; elles sont bien plus que de simples substituts numériques aux bibliothèques classiques déjà existantes. Elles sont considérées par Svensson comme une partie importante de l'infrastructure des HN. Elles ont le potentiel d'être des environnements complexes virtuels pour la recherche scientifique ou des laboratoires virtuels (co-laboratories), c'est-à-dire un système en ligne aidant les chercheurs à travailler en mode collaboratif. Les caractéristiques de ce type d'environnement comprennent habituellement le support technique, les outils numériques permettant le travail collaboratif (forums, wikis, blogs, etc.), les documents d'hébergement et des outils spécifiques à la discipline, pour l'analyse des données, la visualisation ou la gestion de la simulation, la gestion des publications et des outils pédagogiques. Les environnements virtuels de recherche, équivalents de *virtual research environments*, sont devenus importants dans les domaines où la recherche est principalement effectuée par des équipes qui mobilisent plusieurs institutions et pays. La possibilité de partager facilement les données et les résultats de la recherche est précieuse.

Un dernier élément essentiel pour comprendre l'émergence et l'institutionnalisation des HN est de les comparer à celles des Sciences de l'information, qui sous-tendent les fondements théoriques et méthodologiques de l'organisation des connaissances et par conséquent celles des bibliothèques numériques. Les Sciences de l'information ont importé connaissances et méthodes d'autres disciplines, puis exporté d'autres idées vers différents domaines tels que l'informatique et la gestion. Les HN sont, à cet égard, similaires en ce qu'elles se sont développées aux côtés de la linguistique de corpus, à la frontière entre les sciences humaines, sciences sociales et sciences appliquées. La première

manifestation moderne des HN a été l'adoption précoce du traitement automatisé de données textuelles et langagières et du corpus connue aujourd'hui sous les vocables de *linguistique computationnelle* et de *linguistique de corpus*. Les Sciences de l'information peuvent certainement encore contribuer au développement des HN de diverses manières. Certains chercheurs préconisent que les Sciences de l'information jouent un rôle essentiel dans la gestion des grands ensembles de données numériques nécessaires pour l'e-recherche, expliquant qu'il s'agit d'une instanciation moderne de notre préoccupation traditionnelle pour la bibliographie. Les Sciences de l'information devraient également permettre la mise en place de co-laboratoires pour soutenir la recherche et la pratique dans les sciences de l'information elles-mêmes. Le passage à un environnement reposant en grande partie sur l'information numérique et les outils associés, exige un réexamen du rôle des bibliothèques et des centres d'information en tant qu'« espaces d'information » (*Information spaces*). Nous devons penser à de tels espaces d'information et aux produits d'information dans leur double aspect physique et virtuelle, au sein desquels les utilisateurs individuels d'information fonctionneraient et interagiraient.

Les nombreuses contributions de cet ouvrage qui reprend les communications du 10^e colloque de l'ISKO-France témoignent de l'actualité de la thématique des HN et illustrent la synergie potentielle entre ces deux domaines, celui de l'organisation des connaissances et celui des HN.

WIDAD MUSTAFA EL HADI
Professeur des Universités
Présidente ISKO-France