

## **Les humanités numériques et les spécialistes en information documentaire : compte-rendu de la 1re Digital Humanities Summer School, 26-29 juin 2013, Berne**

Mathilde Panes

[mathilde.panes@hesge.ch](mailto:mathilde.panes@hesge.ch)

Haute Ecole de Gestion, Genève

Igor Milhit

[igor.milhit@hesge.ch](mailto:igor.milhit@hesge.ch)

Haute Ecole de Gestion, Genève

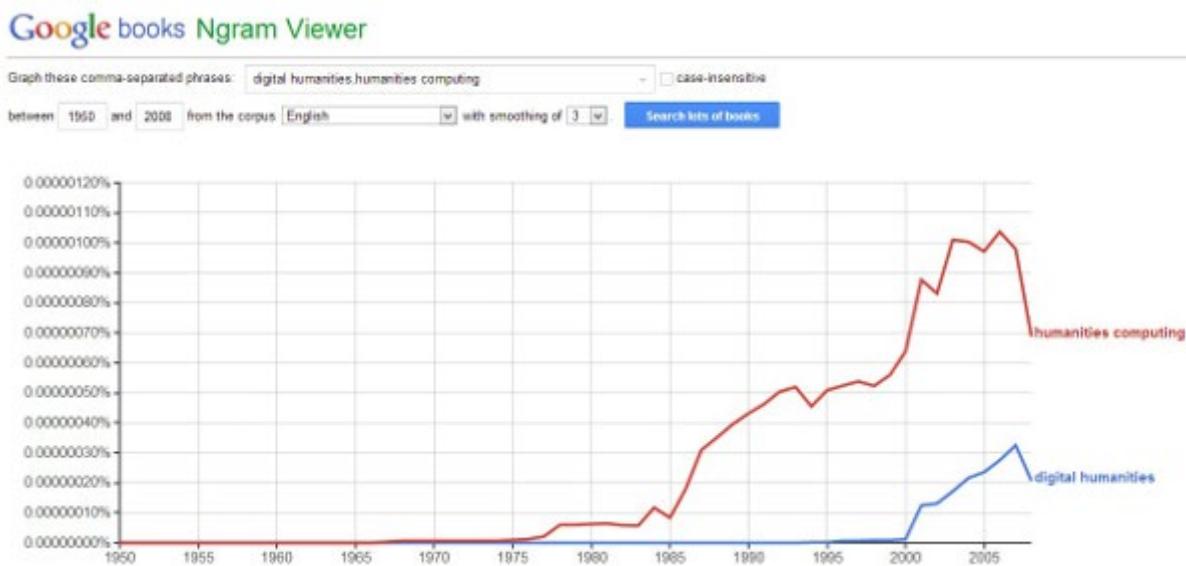
### *Mots-clés*

*humanités numériques, sciences humaines, analyse des données, visualisation, TEI*

Les *Digital Humanities*, un terme en vogue actuellement, suscitent la curiosité d'une partie des professionnels des sciences humaines et de l'information, font l'objet de nombreux articles et ont eu droit à leur première école d'été en Suisse. Organisé par infoclio.ch (Infoclio 2013), cet événement a eu lieu du 26 au 29 juin 2013 à l'Université de Berne (Natale 2013a).

La jeunesse de cette discipline induit une certaine liberté de définition. En effet, les acteurs des *Digital Humanities* n'ont pas encore formulé une définition établie de ce domaine, mais plutôt des esquisses. Lors de son intervention durant l'école d'été, Susan Schreibman, professeur au Département des Humanités numériques du Trinity College de Dublin (Trinity College Dublin 2011), interroge l'audience : les humanités numériques sont-elles un domaine, un outil, une discipline ? Tout cela à la fois ? Ou encore autre chose ? La question reste en suspens. *Wikipédia*, l'encyclopédie libre en ligne, présente les *digital humanities* de la manière suivante : elles « sont un domaine de recherche, d'enseignement et de création concernés par le croisement entre l'informatique et les disciplines des sciences humaines. » (*Humanités numériques* 2013). L'expression *digital humanities* se traduit en français par « humanités numériques » ou encore « humanités digitales » (Clivaz 2012). Nous utiliserons dans ce texte ces trois expressions indifféremment.

Si la plupart des chercheurs travaillent dans un environnement numérique, l'utilisation des nouvelles technologies pour élargir et mettre en valeur les sciences humaines en est à ses balbutiements, ou alors, selon le point de vue, à sa redécouverte. Comme le mettent en évidence les deux graphiques ci-dessous, l'expression *Digital Humanities* semble apparaître au début des années 2000. Le terme "*Humanities computing*" est son prédecesseur, ce qui souligne que les sciences humaines, comme la plupart des sciences, se sont intéressées à l'ordinateur il y a quelques décennies déjà. L'outil *Google Trends* met en avant l'émergence du terme *Digital humanities* à partir de 2006.



**Graphique 1 - Comparaison des termes "Humanities computing" et "digital humanities", selon leurs occurrences dans un corpus de l'outil Ngram Viewer de Google - <http://goo.gl/RJ7Ycb>**



**Graphique 2 - Evolution des recherches effectuées sur Google utilisant le terme "digital humanities", avec l'outil Google Trends - <http://goo.gl/6Fmyr0>**

Dans cet article, nous esquissons les ponts qui rapprochent le domaine des humanités numériques, où beaucoup reste à explorer, et celui de l'information documentaire, à la recherche de convergences prometteuses.

Le programme, serré mais bien équilibré, de ces trois jours d'école d'été a permis d'aborder le sujet de manière conceptuelle et pratique. Les deuxième et troisième journées ont été rythmées par une alternance entre cours et ateliers, la dernière demi-journée ayant été consacrée à des mini *unconferences* : dans un mode horizontal de collaboration, les participants ont pu proposer des sujets de discussions, dont une partie a été sélectionnée par l'assemblée par une procédure de vote. Ces sujets ont ensuite été approfondis de manière informelle par tous les participants intéressés. La collaboration et la communication ont été facilitées par l'usage d'un outil de prise de notes collectives et par l'échange de tweets grâce au hashtag [#dhch](#).

## 1. Des mécanos pour l'humaniste

Parmi les concepts et les outils abordés lors de l'école d'été, nous avons choisi une sélection susceptible d'intéresser, voire d'impliquer les métiers de l'information documentaire.

### 1.1. Analyse textuelle avec des outils en ligne

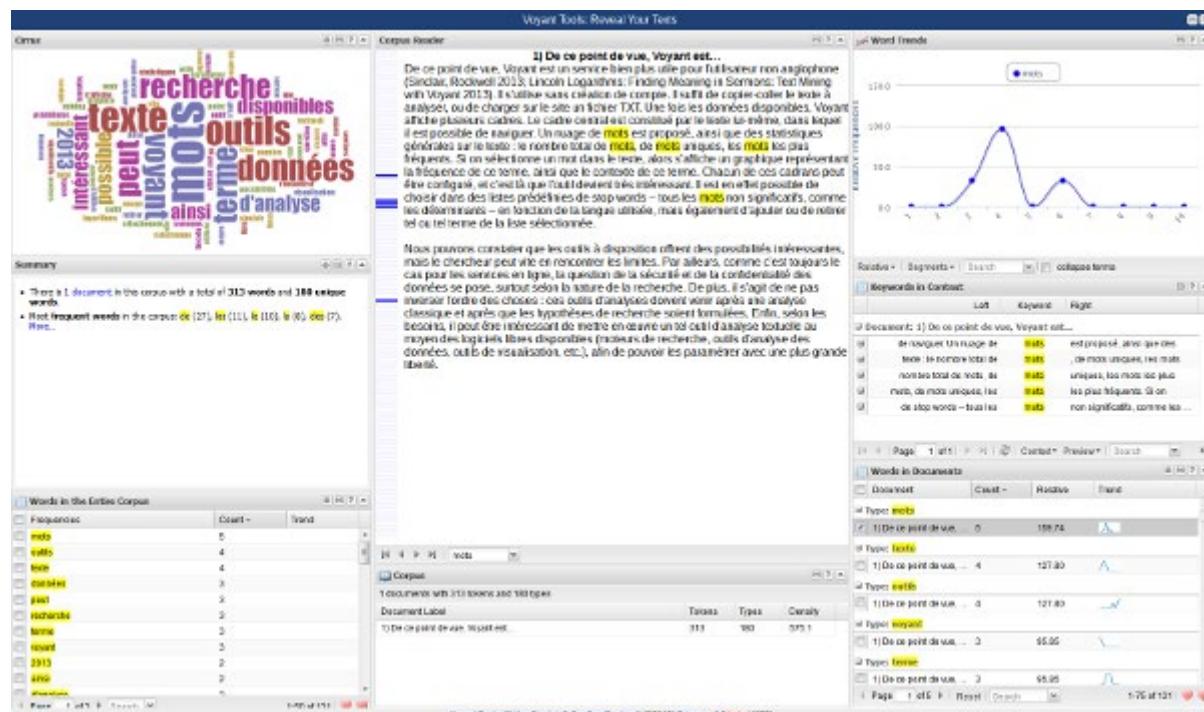
Le cours de Susan Schreibman a offert un regard rétrospectif sur les *digital humanities* et avancé quelques définitions. Il a mis en évidence que l'analyse textuelle et l'édition critique a, dès les années 50, rencontré l'informatique (*Roberto Busa* 2013). Schreibman a par ailleurs animé un atelier (Natale 2013b), qui a permis de découvrir un certain nombre d'outils en ligne pour accompagner l'analyse de texte, des services web qui ont l'avantage d'être relativement simples à utiliser. Simples, parce qu'ils ne nécessitent pas de prendre en charge l'installation de logiciels sur un ordinateur, ni l'apprentissage d'un langage de programmation ou de manipulation des données. Nous sommes d'avis que l'analyse textuelle est un domaine que peuvent s'approprier les spécialistes de l'information, et ce grâce à leurs connaissances des mécanismes de recherche d'information à travers les moteurs de recherche. En effet, les moteurs de recherche font subir plusieurs traitements synchrones et asynchrones aux requêtes formulées et au corpus indexés pour pouvoir être le plus efficaces possibles. Nous pensons qu'un des nouveaux rôles des bibliothèques académiques est d'initier les chercheurs

intéressés à ces outils. Aussi, les compétences spécifiques à ces enseignements doivent être représentées au sein de l'équipe.

Nous constatons que certains des exemples présentés par Schreibman ne sont pas inconnus de nos professions. Le fameux *Google Ngram Viewer*, cité plus haut, est un des outils de recherche, mis à disposition par Google, que les spécialistes de la recherche connaissent. Il exprime graphiquement le nombre d'occurrences d'un terme ou d'une expression dans le corpus de livres numérisés par Google. Celui-ci est certainement l'un des plus conséquents disponible à ce jour. Néanmoins, il est important de garder à l'esprit qu'il est circonscrit aux collections qui ont été numérisées, et qu'il ne représente pas la "littérature mondiale". L'univers anglo-saxon y est par exemple surreprésenté. De plus, l'algorithme qu'utilise Google n'est pas public, ce qui empêche d'utiliser *Ngram Viewer* de manière scientifique, en reproduisant la méthodologie utilisée. Lors de l'atelier, il a été remarqué qu'il n'est pas toujours évident de comprendre comment fonctionne le lissage (*smoothing*) de la courbe. *Google Ngram Viewer* ne permet pas d'ajouter des données pour les analyser, l'utilisateur est contraint de se servir du corpus proposé, même si celui-ci est impressionnant. Il est toutefois possible de sélectionner une langue particulière ou de définir une échelle temporelle.

IBM offre également un outil en ligne très intéressant pour la visualisation de données : *Many Eyes* (IBM 2007). Il est possible de charger ses propres données, par exemple grâce à l'export d'un tableur en CSV, ou tout simplement une œuvre littéraire sous la forme d'un fichier TXT. Ce fichier peut être le résultat de l'OCR d'un livre numérisé, ou un fichier trouvé dans une bibliothèque numérique comme *Projet Gutenberg*, *Internet Archive* ou encore *Wikisources*. Différents types de visualisations sont possibles : créer un nuage de mots tout en sélectionnant une racine particulière (*stemming*), ou observer une structure du contexte d'un terme particulier, ou le réseau qui peut exister entre deux termes mis en relation de diverses manières. Par contre, *Many Eyes* ne s'adapte pas très bien aux textes francophones : l'outil d'IBM ne reconnaît pas les caractères accentués et les signes diacritiques.

De ce point de vue, *Voyant* est un service bien plus utile pour l'utilisateur non anglophone (Sinclair, Rockwell 2013; *Lincoln Logarithms: Finding Meaning in Sermons : Text Mining with Voyant* 2013). Il s'utilise sans création de compte. Il suffit de copier-coller le texte à analyser, ou de charger sur le site un fichier TXT. Une fois les données disponibles, *Voyant* affiche plusieurs cadres. Le cadre central est constitué par le texte lui-même, dans lequel il est possible de naviguer. Un nuage de mots est proposé, ainsi que des statistiques générales sur le texte : le nombre total de mots, de mots uniques, les mots les plus fréquents. Si on sélectionne un mot dans le texte, alors s'affiche un graphique représentant la fréquence de ce terme, ainsi que le contexte de ce terme. Chacun de ces cadres peut être configuré, et c'est là que l'outil devient très intéressant. Il est en effet possible de choisir dans des listes prédéfinies de *stop words* – tous les mots non significatifs, comme les déterminants – en fonction de la langue utilisée, mais également d'ajouter ou de retirer tel ou tel terme de la liste sélectionnée.



**Figure 1 - Interface de Voyant**

Nous pouvons constater que les outils à disposition offrent des possibilités intéressantes, mais le chercheur peut vite en rencontrer les limites. Par ailleurs, comme c'est toujours le cas pour les services en ligne, la question de la sécurité et de la confidentialité des données se pose, surtout selon la nature de la recherche. De plus, ces outils d'analyses ne se substituent pas à une analyse classique, mais viennent la compléter. Enfin, selon les besoins, il peut être intéressant de mettre en œuvre un tel outil d'analyse textuelle au moyen des logiciels libres disponibles (moteurs de recherche, outils d'analyse des données, outils de visualisation, etc.), afin de pouvoir les paramétrier avec une plus grande liberté.

## 1.2. Encodage de manuscrits

Avec Elena Pierazzo (Pierazzo 2012), nous nous sommes penchés sur l'édition de textes (Natale 2013c), et plus particulièrement sur l'édition de textes à partir de manuscrits. Cette activité relève d'une tradition ancienne et marquée par une exigence d'exactitude, condition nécessaire à la pratique de l'analyse littéraire, de l'étude philosophique ou des sciences historiques. Quel est l'apport du numérique dans ce domaine ? Représenter un texte sur un écran revient-il au même que de le faire sur du papier ?

Au-delà des difficultés qui ne peuvent être résolues que par des compétences précises en littérature, histoire, paléographie, etc., retranscrire un manuscrit consiste à rendre compte de la « matérialité » de l'écrit, de sa disposition sur un support. De sa temporalité également : par rapport à un texte imprimé, un manuscrit reflète l'activité de la pensée qui n'est pas aussi linéaire que ce que l'imprimé suggère trop facilement (Grandjean 2013). Du point de vue de ces particularités de l'édition du manuscrit, il s'agit à la fois d'être en mesure de transcrire cette matérialité sur un écran, et de pouvoir utiliser la puissance de calcul de l'ordinateur pour retranscrire le côté dynamique de la pensée. Le « Proust Prototype » est un exemple pertinent de cette démarche (Pierazzo, André 2012). La retranscription du texte se superpose au fac-similé numérique et le lecteur peut soit suivre la chronologie de l'écriture, soit celle de la lecture.

Le codage de textes imprimés pour un support numérique, se fait au moyen d'un standard particulier, mis au point par la [Text Encoding Initiative](#) (TEI). Il s'agit d'un ensemble de directives constituant un format XML, dont l'objectif est de transcrire un texte imprimé pour que la machine puisse le traiter, en encodant à la fois la structure et les métadonnées. Aussi le TEI, jusqu'à récemment, ne permettait pas de retranscrire un document, avec ses aspects physiques, comme la page, la disposition du texte sur la page ou les ratures. C'est dans le but de pouvoir le faire que s'est constitué le *TEI Manuscript Special Interest Group* (Schreibman, Pierazzo, Vanhoutte 2013). Le groupe s'est attelé au développement d'un nouveau module dédié à l'édition critique de manuscrits et à la critique génétique (*Génétique des textes* 2013). L'atelier de l'après-midi a été l'occasion de pratiquer cette méthode d'encodage, à partir d'un court extrait d'un manuscrit de Jane Austen, dont on peut admirer, non pas les résultats de l'atelier, mais l'édition officielle à l'URL suivante :

<http://www.janeausten.ac.uk/manuscripts/pmwats/b1-4.html>.

Cette évolution de la TEI est certainement en mesure d'interpeller les professionnels de l'information documentaire, dans la mesure où, historiquement, elle est apparue au début de l'aventure des bibliothèques numériques. D'autant que cette technique offre la possibilité de pousser encore plus loin la mise en valeur des collections de manuscrits numérisés, ajoutant à l'image scannée une transcription du texte qui respecte la matérialité du document.

### 1.3. RDF pour un retour dans le futur

Frédéric Kaplan (EPFL 2013) a présenté le projet *Venice Time Machine*, qui a pour but d'offrir aux chercheurs et au public des outils de visualisation de l'histoire de Venise. Il s'agit donc de récolter des données et de leur faire porter du sens grâce à des modèles sémantiques (Natale 2013d, 2013e).

Pour la première étape, l'acquisition des données, il est question de numériser la totalité des archives de Venise : 80 km de documents d'archives, sur une surface temporelle de plus de 1000 ans. Kaplan pense qu'en optimisant les techniques de numérisation, il est possible de parvenir à traiter 450 volumes par jour, afin de numériser le tout en dix ans. Quant à l'OCR, c'est un fait connu qu'actuellement le taux de réussite de la reconnaissance optique des caractères appliquée sur des manuscrits atteint, au mieux, 80 %. Ce taux peut sembler élevé, or il signifie que le cinquième de chaque volume doit être corrigé à la main, par des personnes capables de lire de telles écritures. Ce qui rend le projet de numérisation impossible. Kaplan pense que l'on aborde la question de l'OCR des manuscrits par un angle qui n'est pas des plus pertinents. La recherche actuelle s'attelle à produire des algorithmes capables de reconnaître des écritures manuelles quels que soient les contextes. Il se trouve que si l'on tient compte de ce contexte – est-ce un livre de droit, écrit à Venise au XVI<sup>e</sup> siècle, une chronique du XI<sup>e</sup> ? –, alors il est possible de réduire considérablement le type de langues, de phrases, de mots, voire de formes de caractères, qui sont vraisemblablement utilisés dans un corpus précis. Aussi faudrait-il produire autant d'algorithmes que de corpus particuliers, et, selon Kaplan, par cette méthode il est possible de parvenir à un bien meilleur taux de réussite.

En imaginant que le point de la numérisation, y compris de l'OCR, soit résolu, reste à donner du sens à cette énorme masse de données, et que ce sens soit accessible au traitement automatique par les machines. C'était le propos de l'atelier *Semantic Modelling for Humanities* (Natale 2013e). Celui-ci a abordé les différentes méthodes connues jusqu'ici pour enregistrer et structurer des données. La méthode la plus simple est d'utiliser un tableau (Microsoft Excel,

LibreOffice Calc, etc.). On dispose ainsi d'éléments sémantiques : l'intitulé des colonnes. Bien entendu, avec une telle méthode les requêtes de recherche possibles sont relativement limitées, et les logiciels actuels ne permettent pas de dépasser un nombre fixe de colonnes.

L'étape suivante consiste à utiliser un système de bases de données relationnelles (par exemple SQL). Le problème de la limite des colonnes se contourne par la création de nouvelles tables qui relient les informations entre elles. Et dans un tel contexte, la complexité de requêtes est bien plus élevée et donne lieu à des résultats plus approfondis. En revanche, alors que la quantité des données évolue, il sera peut-être nécessaire de migrer d'un schéma de base de données à un autre, ce qui peut vite devenir complexe et pénible.

Pour se libérer de ces contraintes, il est souhaitable d'utiliser non plus seulement les données, mais les métadonnées. Plus s'ajoutent des données, plus s'enrichissent les métadonnées, qui sont une manière de structurer les données, et qui offrent des possibilités de requêtes très développées. C'est lors de cette étape qu'intervient le RDF (*Resource Description Framework* 2013) : il s'agit de relier des données au moyen de métadonnées. Avec cette solution, les éléments sémantiques sont nombreux, puisque les relations entre les données donnent justement du sens à celles-ci. Structurer revient à définir, et inversement. De plus, cette structure n'est ni figée ni limitée par un logiciel (les limites d'un tableur comme Excel) ou par un schéma (le schéma d'une base de données), mais évolue et se développe simplement en ajoutant des données et des relations entre elles.

Ce type de projet rencontre un certain scepticisme à la fois des chercheurs en sciences humaines et des spécialistes de l'information documentaire, principalement concernant le volet de la numérisation et de « l'OCRisation » de telles masses de documents d'archives, le plus souvent manuscrits. Pourtant, comme cela a souvent été le cas dans l'histoire récente, c'est du monde des ingénieurs que des évolutions significatives sont venues modifier de manière fondamentale les méthodes à la fois des sciences humaines et des sciences de l'information. Mais, au-delà de ce scepticisme bien compréhensible, il est indéniable que les spécialistes de l'information documentaire ont des compétences à faire valoir, que ce soit en termes de projets de numérisation, de traitement de textes anciens et de gestion des métadonnées, voire de leur mise à disposition justement dans le but de construire de grands ensembles sémantiques de données (voir par exemple Hügi, Prongué 2013).

## 2. Un nouveau souffle pour les ID

Au-delà de l'intégration des spécialistes de l'information documentaire (ID) dans le développement des humanités numériques, l'émergence de ce nouveau domaine peut être une source d'inspiration pour les professionnels de l'ID. En effet, nous pensons tout d'abord que le rapport entre le passé, la tradition et la transition vers le monde numérique est un processus similaire, qu'il se déroule dans les sciences humaines ou dans l'information documentaire. L'adoption des nouvelles technologies est inéluctable. Les humanistes et les professionnels de l'information documentaire doivent identifier les méthodes et le savoir-faire propres à leurs disciplines et les utiliser dans un contexte numérique. Dans les deux cas, cela nécessite d'importantes capacités d'introspection et d'adaptation.

Par ailleurs, le poids de la tradition joue un rôle que nous percevons comme potentiellement contre-productif. L'image d'une profession ou d'une discipline véhiculée auprès des autres communautés professionnelles, du grand public et des détenteurs des cordons de la bourse,

ainsi qu'auprès des acteurs du domaine eux-mêmes, peut freiner l'évolution vers la modernité. Assurément, un déficit en termes d'image entraîne une crédibilité réduite lorsqu'il s'agit de débuter des projets audacieux.

Lors de la participation à l'école d'été des humanités numériques, nous avons pu saisir l'enthousiasme bouillonnant qui va de pair avec la naissance des nouvelles idées et de nouveaux concepts, soutenus par les nouvelles technologies. Cet enthousiasme se traduit par une utilisation accrue d'outils informatiques, parfois imparfaits ou non-standards, ainsi que par la collaboration à la création de nouveaux instruments. Le site [infoclio.ch](http://infoclio.ch) en est un exemple : il rassemble de nombreux outils (informationnels, techniques, méthodologiques) à destination des professionnels des sciences historiques.

Sans oublier les fondations de leurs disciplines respectives, ni mettre de côté un regard critique et questionnant, les participants à la *Summer School* ont naturellement adopté une posture favorisant les échanges, et donc l'innovation. Comme nous l'avons évoqué en introduction, la création d'un [espace collaboratif](#) pour la prise de notes (*DHCH 2013*) a permis à l'audience de s'impliquer. Les *unconferences* (non-conférences), qui ont eu lieu lors de la dernière demi-journée, ont été un autre indicateur d'une communication transversale qui ne se formalise pas du statut de chacun. Dans un contexte si dynamique, la prise de parole par chacun est encouragée. Ce contexte précis pourrait bénéficier au domaine de l'information documentaire pour la construction de concepts novateurs et de nouveaux paradigmes. Aujourd'hui, l'effervescence qui a eu cours dans le milieu ID lors du début de l'informatisation des catalogues est retombée, et une association avec les humanistes est l'occasion de donner un nouveau souffle à l'engouement pour l'innovation technologique en matière de traitement de l'information.

La combinaison de l'ancien et du neuf donne lieu à des associations inédites : l'histoire et la visualisation de données, la philologie et l'encodage de texte ou l'information documentaire et la curation de données. Pour ce faire, il est nécessaire de définir les compétences métiers intrinsèques à une discipline, puis de mener une réflexion concernant les outils qui serviront le mieux cette discipline. Une fois les premiers pas effectués, une communication adéquate sur les projets est nécessaire. En cela, il est important que les chercheurs s'entendent sur une définition concertée des humanités numériques. Tout comme les professionnels de l'information tentent constamment de définir « l'information documentaire ».

### 3. Une place à prendre

Dans ce contexte, quels peuvent être les rôles des professionnels de l'information documentaire dans les humanités numériques ? En préambule à ce point, il n'est peut-être pas inutile de rappeler que si les humanités numériques sont nées de la confrontation des sciences humaines avec l'informatisation généralisée, la science de l'information est souvent située à l'intersection des sciences de l'ingénieur et des sciences humaines. Ici, la question de la double ou triple compétence est centrale. En effet, dans les humanités numériques, l'équation est la suivante : sujet (littérature du XVI<sup>e</sup> siècle, manuscrits médiévaux, etc.) ET technique (encodage de textes, visualisation des données, fouille de données textuelles, etc.) ET utilisation des outils documentaires informatisés (gestion des références bibliographiques avec, par exemple, Zotero, système de gestion de contenu, wiki, etc.). Dans ce cadre, il y a donc toujours un sujet, plus ou moins spécifique, avec ses particularités et ses concepts fondamentaux, ainsi qu'une couche technique. Les professionnels de l'information

documentaire sont plus ou moins familiers avec cette partie technique, car, avec l'évolution de leurs métiers, ils sont souvent versés dans l'utilisation d'outils informatiques.

Traditionnellement, les professionnels ID sont en mesure d'appréhender des sujets pour lesquels ils n'ont pas forcément d'expertise. Par exemple, lors de l'indexation, le bibliothécaire sélectionne les mots-matières adéquats sans comprendre le contenu du livre dans son intégralité. Il est simplement capable d'en extraire assez d'informations pour pouvoir rendre le livre accessible via une recherche. Si nous transposons cette pratique dans le cadre des humanités digitales, le professionnel ID devra appliquer les techniques de gestion de l'informatique numérique à des sujets relatifs aux sciences humaines. Il pourra apporter une aide concrète, plutôt technique, sans devoir être un spécialiste, par exemple, de la correspondance de tel écrivain du XVIII<sup>e</sup> siècle.

Pour saisir les enjeux des humanités numériques, un spécialiste de l'information devrait comprendre les méthodologies et les compétences spécifiques au domaine des humanistes, en faisant appel à ses capacités d'adaptation et d'apprentissage ou en privilégiant l'échange de connaissances. Le niveau d'expertise du spécialiste dépend principalement des exigences d'un employeur qui pourrait souhaiter qu'une double ou triple compétence soit déjà acquise par la personne qu'il emploie. En partant du principe que les humanités numériques ne sont pas encore circonscrites, les emplois et les compétences qui s'y rapportent ne sont pas encore clairement définis. En clair, les humanistes n'ont pas encore déterminé quels étaient leurs besoins pour une transition vers le numérique, les professionnels ID peuvent être intégrés au sein de ces terrains mouvants.

Si la *summer school* de 2013 a permis de mettre en évidence qu'une partie des chercheurs en sciences humaines ont acquis des compétences certaines à la fois dans les domaines de l'informatique (outil de visualisation basé sur la maîtrise de langages) et des sciences de l'information (Zotero, RDF), ce n'est de loin pas le cas de toute la communauté. Or, ces mêmes chercheurs vont de plus en plus se tourner vers ce type d'outils. Et si les professionnels de l'information documentaire leur apportent déjà une aide pour s'emparer des outils de recherches existants (OPAC, banques de données commerciales, recherche sur le Web, logiciel de gestion de références bibliographiques), ils peuvent certainement étendre ce rôle. Notamment parce que les projets de recherche impliquent aujourd'hui des équipes d'une certaine ampleur, des productions de documents importantes, ainsi que la gestion de volumes de données non négligeables. De ce point de vue, les compétences ID sont d'évidence des ressources utiles dans l'accompagnement des usagers, la structuration de l'information ainsi que le développement d'interfaces et d'outils de recherche permettant de soutenir les efforts des chercheurs et l'exploitation future des données.

À cela s'ajoute que la position particulière des professionnels de l'information documentaire est tout à fait intéressante. Nos métiers, nous l'avons déjà mentionné, ont une longue tradition de services offerts aux chercheurs, professeurs et étudiants en sciences humaines. Mais, depuis quelques décennies au moins, nous avons été confrontés avec le monde de l'informatique. Comme le travail des chercheurs, même en sciences humaines, implique de gérer de l'information, avec les moyens et les contraintes de l'informatique, les professionnels de l'information documentaire doivent développer les compétences nécessaires pour soutenir les chercheurs dans leur travail. Nous pensons, par exemple, aux exigences pointues qui sont de plus en plus imposées aux chercheurs en matière d'archivage électronique des données de la recherche.

Dans ce sens, nous pensons qu'il est essentiel que les spécialistes ID puissent, d'une part, intervenir très tôt dans le cursus universitaire, comme c'est d'ailleurs la tendance, non seulement pour former les étudiants à la recherche d'information, à la citation et la rédaction de bibliographies, mais également pour introduire des notions de gestion de l'information, ne serait-ce que la maîtrise des sauvegardes, ou le choix de formats offrant de meilleures assurances d'archivage. D'autre part, nos métiers doivent accompagner les chercheurs, et ce dès la phase de conception d'un projet de recherche, afin d'apporter toute l'aide nécessaire à la *digital curation* des données de la recherche, ou à la manipulation de certains outils d'analyse (comment préparer un corpus de texte pour que des outils d'analyse puissent les traiter, par exemple).

## BIBLIOGRAPHIE

- CLIVAZ, Claire, 2012. «Humanités Digitales»: mais oui, un néologisme consciemment choisi! Suite. *Digital Humanities Blog* [en ligne]. 13 septembre 2012. [Consulté le 16 décembre 2013]. Disponible à l'adresse : <http://claireclivaz.hypotheses.org/114>
- DHCH, 2013. *Framapad Lite* [en ligne]. [Consulté le 16 décembre 2013]. Disponible à l'adresse : <http://lite.framapad.org/p/DHCH>
- EPFL, 2013. Frédéric Kaplan. *EPFL* [en ligne]. 2013. [Consulté le 16 décembre 2013]. Disponible à l'adresse : <http://personnes.epfl.ch/frederic.kaplan>
- Génétique des textes, 2013. *Wikipédia* [en ligne]. [Consulté le 16 décembre 2013]. Disponible à l'adresse : [https://fr.wikipedia.org/wiki/Génétique\\_des\\_textes](https://fr.wikipedia.org/wiki/Génétique_des_textes)
- GRANDJEAN, Martin, 2013. #dhch Manuscript edition « Sometimes there is no text! » @epierazzo (non linear examples) pic.twitter.com/cCEoy3h1oc. *@GrandjeanMartin* [en ligne]. 10 AM - -06-27 2013. [Consulté le 25 novembre 2013]. Disponible à l'adresse : <https://twitter.com/GrandjeanMartin/status/350164196094255104>
- HÜGI, Jasmin et PRONGUÉ, Nicolas, 2013. Marc contre Élodie, ou les avantages des Linked Open Data en bibliothèque. *Recherche d'ID : carnet de recherche des étudiants du master en information documentaire de la Haute école de gestion de Genève* [en ligne]. 10 décembre 2013. [Consulté le 16 décembre 2013]. Disponible à l'adresse : <http://recherchemid.wordpress.com/2013/12/10/marc-contre-elodie/>
- Humanités numériques, 2013. *Wikipédia* [en ligne]. [Consulté le 2 décembre 2013]. Disponible à l'adresse : [https://fr.wikipedia.org/wiki/Humanités\\_numériques](https://fr.wikipedia.org/wiki/Humanités_numériques)
- IBM, 2007. Many Eyes. *Many Eyes* [en ligne]. 2007. [Consulté le 8 décembre 2013]. Disponible à l'adresse : <http://www-958.ibm.com/software/data/cognos/maneyes/>
- INFOCLO, 2013. infoclio.ch. *infoclio.ch : le portail professionnel des sciences historiques en Suisse* [en ligne]. 2013. [Consulté le 2 décembre 2013]. Disponible à l'adresse : <http://www.infoclio.ch/>
- Lincoln Logarithms: Finding Meaning in Sermons: Text Mining with Voyant, 2013. *Emory Libraries : Digital Scholarship Commons* [en ligne]. [Consulté le 8 décembre 2013]. Disponible à l'adresse : <http://disc.library.emory.edu/lincoln/voyant/>
- NATALE, Enrico, 2013a. First DH Summer School : Switzerland, June 26-29, 2013, University of Bern. [en ligne]. 2013. [Consulté le 2 décembre 2013]. Disponible à l'adresse : <http://www.dhsummerschool.ch/>
- NATALE, Enrico, 2013b. Text Analysis with online Tools. *dhsummerschool* [en ligne]. 4 juillet 2013. [Consulté le 16 décembre 2013]. Disponible à l'adresse : [http://www.dhsummerschool.ch/?page\\_id=358](http://www.dhsummerschool.ch/?page_id=358)
- NATALE, Enrico, 2013c. Course: Pierazzo. *dhsummerschool* [en ligne]. 4 juillet 2013. [Consulté le 8 décembre 2013]. Disponible à l'adresse : [http://www.dhsummerschool.ch/?page\\_id=283](http://www.dhsummerschool.ch/?page_id=283)

- NATALE, Enrico, 2013d. Course: Kaplan. *dhsummerschool* [en ligne]. 4 juillet 2013. [Consulté le 16 décembre 2013]. Disponible à l'adresse : [http://www.dhsummerschool.ch/?page\\_id=335](http://www.dhsummerschool.ch/?page_id=335)
- NATALE, Enrico, 2013e. Semantic Modelling for the Humanities. *dhsummerschool* [en ligne]. 4 juillet 2013. [Consulté le 16 décembre 2013]. Disponible à l'adresse : [http://www.dhsummerschool.ch/?page\\_id=364](http://www.dhsummerschool.ch/?page_id=364)
- PIERAZZO, Elena et ANDRÉ, Julie, 2012. Autour d'une séquence et des notes du Cahier 46 : enjeu du codage dans les brouillons de Proust. *Proust Prototype* [en ligne]. 2012. [Consulté le 8 décembre 2013]. Disponible à l'adresse : [http://research.cch.kcl.ac.uk/proust\\_prototype/](http://research.cch.kcl.ac.uk/proust_prototype/)
- PIERAZZO, Elena, 2012. Elena Pierazzo. *Elena Pierazzo* [en ligne]. 17 novembre 2012. [Consulté le 8 décembre 2013]. Disponible à l'adresse : <http://www.elenapierazzo.org/>
- Resource Description Framework, 2013. *Wikipédia* [en ligne]. [Consulté le 16 décembre 2013]. Disponible à l'adresse : [https://fr.wikipedia.org/wiki/Resource\\_Description\\_Framework](https://fr.wikipedia.org/wiki/Resource_Description_Framework)
- Roberto Busa, 2013. *Wikipedia, the free encyclopedia* [en ligne]. [Consulté le 2 décembre 2013]. Disponible à l'adresse : [https://en.wikipedia.org/wiki/Roberto\\_Busa](https://en.wikipedia.org/wiki/Roberto_Busa)
- SCHREIBMAN, Susan, PIERAZZO, Elena et VANHOUTTE, Edward, 2013. TEI: Manuscripts SIG. *TEI : Text Encoding Initiative* [en ligne]. 7 mars 2013. [Consulté le 8 décembre 2013]. Disponible à l'adresse : <http://www.tei-c.org/Activities/SIG/Manuscript/>
- SINCLAIR, Stéfan et ROCKWELL, Geoffrey, 2013. Voyant Tools: Reveal Your Texts. [en ligne]. 2013. [Consulté le 8 décembre 2013]. Disponible à l'adresse : <http://voyant-tools.org/>
- TRINITY COLLEGE DUBLIN, 2011. School of English : Trinity College Dublin, The University of Dublin, Ireland. [en ligne]. 2011. [Consulté le 2 décembre 2013]. Disponible à l'adresse : <http://www.tcd.ie/English/staff/academic-staff/susan-schriebman.php>