

## Mythes et origines

*Digitus Dei est hic!*, c'est par ces mots que Roberto Busa<sup>1</sup>, le prêtre jésuite à l'origine de l'utilisation de l'outil informatique sur des corpus textuels, désigne la présence divine dans le texte. L'anecdote est à double sens, faisant autant référence à la pratique de l'index qui désigne, qu'à l'index en tant que technique de classement des concepts ou mots-clés présents dans un document. Cette présence digitale dans le texte s'est parfois manifestée par des symboles d'un doigt désignant justement un passage important qu'il était plus que souhaitable de lire. Cette symbolique a été transférée dans les environnements numériques pour signaler un lien hypertexte. Il s'agit d'une forme d'architexte<sup>2</sup>, un élément inclus dans un système qui se trouve dissimulé et intégré dans le code d'un logiciel informatique et qui continue de produire des effets.

Cette signalétique symboliquement tactile le devient de plus en plus concrètement avec les interfaces interactives des tablettes. Tandis que l'adjectif numérique renforce la calculabilité des données, l'adjectif digital exerce comme un rappel vers une matérialité et une « prise en main »<sup>3</sup> que les francophones distinguent mieux que les anglophones.

Mais plutôt que de s'interroger sur ce qui est préférable entre numérique et digital, il importe surtout de se demander quelles sont ces humanités auxquelles on fait référence. Le terme d'humanités s'est vu quasi ressuscité avec l'adjonction de l'adjectif numérique. Difficile aussi d'en donner une définition claire, si ce n'est que les humanités ont souvent été qualifiées d'humanités classiques. Les humanités classiques désignent une connaissance de « l'humanité », *humanitas* en latin, grâce aux connaissances et aux compétences de ceux qui se montrent capables de lire, de comprendre et d'interpréter des textes notamment en grec et en latin. Longtemps associées aux différents arts libéraux (*quadrivium et trivium*), les humanités ou études humanistes ont aussi désigné la base de l'enseignement secondaire notamment au XIXe siècle. Le mot anglais *Humanities* possède un sens quelque peu différent et renvoie à diverses disciplines dont l'histoire, la philosophie, la littérature mais aussi le droit. Ce terme d'humanités devient polysémique et finalement pratique puisqu'il permet de rassembler les tenants des « belles lettres » et d'autres disciplines réunies plus couramment sous le vocable de sciences humaines et sociales. Finalement, en dehors des enjeux possibles

---

<sup>1</sup> Roberto Busa (1913-2011) est un jésuite italien qui a convaincu Thomas J. Watson, fondateur d'IBM, de soutenir le projet d'indexation lemmatisée de l'œuvre de saint Thomas d'Aquin. Cet index est aujourd'hui disponible en ligne : <http://www.corpusthomaticum.org/wintrofr.html>;

<sup>2</sup> Gérard Genette, *Introduction à l'architexte* (Paris: Seuil, 1979).

<sup>3</sup> Voir notamment ce billet de Claire, Clivaz. «Humanités Digitales»: Mais Oui, Un Néologisme Consciemment Choisi! *Digital Humanities Blog*. Billet du 13 septembre 2012 <http://claireclivaz.hypotheses.org/114>,

de frontières disciplinaires, c'est aussi la question de l'humanisme qui est posée et sans doute à travers l'humanisme, la question de l'utilité des humanités, sans cesse menacées pour leur possible non adaptation au marché de l'emploi<sup>4</sup>, mais aussi le rapport étroit entretenu avec les techniques<sup>5</sup>.

Impossible de faire également l'impasse sur un transhumanisme qui vise à améliorer génétiquement ou par extension l'être humain. Un des plus célèbres représentant est Ray Kurzweil<sup>6</sup>, récemment engagé par Google pour travailler sur le transfert du cerveau dans son intégralité dans le numérique. Un vieux mythe et fantasme déjà porté notamment par Hans Moravec<sup>7</sup> et la L5society ainsi que par la science-fiction avec entre autres les écrits de William Gibson<sup>8</sup>.

Les *Digital Humanities* se sont positionnées différemment, même s'il s'agit aussi d'augmenter les capacités d'analyse et de partage mais peut-être à des fins plus collectives et collaboratives. Il est probable aussi qu'il faille également songer les cultures numériques actuelles et futures dans une réinterrogation des pratiques corporelles dont la culture classique avait quelque peu figé dans la position studieuse de la chaise et du bureau.

En effet, les discours des *Digital humanities* sont porteurs de « mouvement » et prônent autant le faire que le dire. L'esprit qui y règne est dans l'ordre des « makers » tels que les décrit Chris Anderson pour évoquer ces nouveaux entrepreneurs du numérique notamment autour des imprimantes 3D. Dans les *Digital Humanities*, il s'agit surtout de travailler autour d'éléments scientifiques et de recherche.

*« Digital Humanities est le terme courant qualifiant les efforts multiples et divers de l'adaptation à la culture numérique du monde savant. Il illustre, par son histoire, et surtout par son incarnation institutionnelle actuelle, aux États-Unis comme en Europe, l'évolution non seulement de l'environnement numérique dans ses effets sur l'édition académique et scientifique et sur ce qu'on appelle parfois « scholarly communication », mais surtout de l'attitude des institutions culturelles et universitaires. »*<sup>9</sup>

---

<sup>4</sup> La question agitait déjà au début du XXème siècle. Voir notamment cet ouvrage de Pierre Lasserre, (1913). La doctrine officielle de l'Université : critique du haut enseignement de l'État, défense et théorie des humanités classiques / Pierre Lasserre. "Mercure de France" (Paris). Disp. sur : <http://gallica.bnf.fr/ark:/12148/bpt6k55486248>

<sup>5</sup> [Éric Guichard, « L'internet et les épistémologies des SHS », \*Sciences/Lettres\* no 2 \(Automne 2013\). Les questions épistémologiques liées aux humanités numériques sont discutées dans cet article, qui conteste par ailleurs cette expression ou du moins la revendication de l'émergence d'une nouvelle discipline.](#)

<sup>6</sup> [Voir son dernier ouvrage traduit en Français : Ray Kurzweil, \*Humanite 2.0: la bible du changement\* \(Paris: M21 éditions, 2007\).](#)

<sup>7</sup> [Hans P Moravec, \*Mind Children: The Future of Robot and Human Intelligence\* \(Cambridge, Mass.: Harvard University Press, 1988\).](#)

<sup>8</sup> [Ces thèmes ont été plus récemment abordés dans un roman français : Aurélien Bellanger, \*La théorie de l'information: roman\* \(\[Paris\]: Gallimard, 2012\).](#)

<sup>9</sup> Milad Doueïhi, Pour un humanisme numérique

### ***Roberto Busa, le précurseur***

Le prêtre jésuite italien Roberto Busa (1913-2011) est souvent cité pour son action de précurseur dans la volonté d'utiliser les technologies informatiques auprès des textes. Il y songe dès 1946, mais c'est trois ans plus tard, lors de sa rencontre avec [Thomas J. Watson](#), le fondateur d'[IBM](#), qu'il parvient à lancer son projet d'analyse de l'œuvre de Saint Thomas D'Aquin. Le projet va se réaliser durant de nombreuses années et consiste principalement en une lemmatisation du corpus, c'est-à-dire la création d'un index organisé et structuré des termes présents dans les textes. Un des objectifs de la lemmatisation étant de réduire les mots à leur racine grammaticale, afin de repérer plus aisément les proximités, et le nombre de fois où un mot est présent dans un texte, ce qu'on appelle des occurrences. Le corpus est actuellement consultable en ligne.<sup>10</sup> La longue expérience de Busa est intéressante car elle montre l'évolution impressionnante des technologies tout au long du projet, que ce soit en puissance de calcul, ou en place nécessaire pour conserver les données. Le corpus lemmatisé était prévu initialement pour être imprimé sous la forme d'un index papier. Roberto Busa travailla longtemps avec des enregistrements de données sur cassettes, dont la longueur totale des bandes mises bout-à-bout équivalait à la distance entre Paris et Lisbonne ! On mesure ainsi mieux le cheminement technologique quand on sait que ce travail tient sur un seul DVD désormais<sup>11</sup>.

La liste des participants connus et moins connus à cette histoire pourrait être longue. Les acteurs de terrain qui ont contribué à la diffusion des outils et à la démythification sont nombreux. Il y a donc des acteurs quelque peu cachés de cette histoire. Le projet *hidden histories*<sup>12</sup> cherche justement à rassembler des témoignages de cette histoire. Ce ne sera pas notre propos ici tant il s'agirait d'effectuer une archéologie des humanités numériques.

### ***Humanities computing : les humanités assistées par l'ordinateur***

Les *digital humanities* ne sont donc pas si récentes. Elles furent précédées par un autre mouvement, celui des *Humanities Computing*, ces humanités assistées par l'ordinateur. Certains champs disciplinaires vont être particulièrement concernés par l'utilisation des outils informatiques pour pratiquer de l'analyse de corpus, notamment les études littéraires et plus particulièrement les *english studies* afin de mettre en avant des concordances<sup>13</sup>. Les *humanities computing* ont dû intégrer deux cultures différentes afin « *d'apporter la rigueur et les méthodologies procédurales et systématiques de désambiguïsation propre aux sciences dures pour répondre aux problèmes que les humanités ont traité jusque-là de façon*

---

<sup>10</sup> Corpus Thomisticum Disp. Sur : <http://www.corpusthomicum.org/>

<sup>11</sup> L'association ADHO (*Alliance of Digital Humanities Organizations*) a créé d'ailleurs à son honneur le « Prix Busa » qui récompense tous les trois ans un acteur méritant des humanités numériques. Le prix a d'ailleurs été attribué en 1998 à Roberto Busa lui-même.

<sup>12</sup> Hidden Histories. <<http://hiddenhistories.omeka.net/>>. [Ce projet se concentre toutefois sur le monde anglo-saxon.](#)

<sup>13</sup> [Pour le cas de l'histoire en France, voir le chapitre de Frédéric Clavert.](#)

*intuitive* ». <sup>14</sup> Les méthodes vont suivre les progrès dans les méthodes de traitement automatique du langage et de la fouille de texte. Mais ce sont également les évolutions du web qui vont faire également évoluer la nature des corpus mais également les méthodes de description et de balisage du texte même si pour ce cas de figure, un langage dédié avait vu le jour dans les années 1980 : la TEI (*text encoding initiative*). La TEI est un langage standard pour l'encodage de textes sous forme de balises. La TEI repose sur des logiques proches de SGML créé en 1986 et qui a inspiré le HTML, langage de base du web et le XML, langage pivot pour la structuration des données. La TEI a permis le développement d'une communauté de pratiques active. La TEI repose désormais sur le xml et demeure toujours un langage important pour décrire finement les différentes strates et interactions dans un texte afin de faciliter son étude et les requêtes précises. Elle est le produit de discussions et de travaux mettant en scène trois associations *the Association for Computers and the Humanities* (fondée en 1978), *the Association for Computational Linguistics*, and *the Association for Literary and Linguistic Computing*. Parmi ses créateurs figurent notamment Lou Bernard, de l'université d'Oxford, qui travaille actuellement en France pour le très grand équipement Adonis <sup>15</sup>.

A la suite de Roberto Busa, d'autres chercheurs se sont mis à utiliser des technologies informatiques. Joseph Raben crée d'ailleurs à cet effet en 1966 la revue scientifique *Computer and the Humanities*.

A la fin des années 1970 et au début des années 1980, un logiciel comme COCOA basé sur le langage Fortran <sup>16</sup>, connut un succès car il permettait de mesurer les occurrences et de construire des analyses de concordances. La constitution de corpus numérisé commence à prendre de l'essor et déjà la nécessité de conserver ces travaux voit la constitution d'un archivage dédié avec le projet OTA (*Oxford Text Archive*) en 1976. L'objectif est à la fois de conserver et autant que possible de rendre accessible à un grand nombre de chercheurs ces corpus. Les corpus restent toutefois parfois restreints et le besoin de constituer de plus amples collections se fait sentir. C'est le cas du projet de Théodore Brunner <sup>17</sup> qui cherche à rassembler les principaux textes de l'antiquité grecque depuis Homère au sein du projet *Thesaurus Linguae Graecae* hébergé initialement à l'université de Californie d'Irvine. Les premiers centres commencent à émerger alors et les questions sur la nécessité d'apprendre à coder se font déjà jour, débats qui se poursuivent encore aujourd'hui avec l'éternel questionnement sur les fondamentaux à maîtriser pour sa discipline.

---

<sup>14</sup> Susan Hockey. "The histories of humanities computing" in . P.3-19

<sup>15</sup> Le TGE-Adonis est une infrastructure pour conférer à la recherche une ingénierie améliorée et des services performants. Le TGE fusionne actuellement avec Corpus-IR pour devenir une nouvelle Unité Mixte de Service, la nouvelle infrastructure *Huma-Num*, qui sera une Très Grande Infrastructure de Recherche dédiée aux humanités numériques.

<sup>16</sup> Langage informatique développé dans les années 50 par IBM et principalement dédié à des projets et applications scientifiques.

<sup>17</sup> Brunner, T. F. (1993). *Classics and the Computer: The History of a Relationship*. In J. Solomon (ed.), *Accessing Antiquity: The Computerization of Classical Studies* (pp. 10–33). Tucson: University of Arizona Press.

Les années 1980 marquent aussi le tournant de l'informatique domestique qui change les méthodes et les pratiques des chercheurs. Les interfaces graphiques d'Apple sont plus séduisantes que celles de Microsoft et surtout les chercheurs peuvent utiliser le logiciel *Hypercard* qui leur permet de créer des hypertextes entre fiches. Le *perseus project*<sup>18</sup>, dont l'objectif est de constituer une bibliothèque numérique de travaux sur la Grèce Antique, sera ainsi initialement délivré à l'utilisateur final sous cette forme. Lou Bernard<sup>19</sup> considère les années 1980 comme l'avènement des gadgets qui ont permis un début de démocratisation de l'informatique, qui ont vu les institutions scolaires s'emparer du sujet pour enseigner les langages LOGO et BASIC. Incontestablement, les outils techniques vont faire évoluer les pratiques et les mentalités ainsi que les compétences associées.

Evidemment, les pratiques de messageries électroniques et l'avènement du web vont contribuer au développement des projets et faciliter les échanges. La liste *Humanist* créée en 1987 par Willard McCarthy constitue alors un « séminaire électronique »<sup>20</sup>, c'est-à-dire une communauté de pratiques active autour d'une liste de diffusion. Dans cette dynamique de partage est alors éditée l'ouvrage *Humanities Computing Yearbook (HCY)* qui est paru plusieurs années et qui a compté jusqu'à 700 pages pour faire le point sur les avancées, débats et méthodes dans le domaine des humanités assistées par ordinateur. Initialement sur papier, il s'est transformé en banque de données aisément diffusable et facilitant la recherche d'informations. La pratique du web et l'apparition du premier navigateur Mosaic pousse les chercheurs à publier sur ce nouvel espace même si le côté peu structuré du web du fait d'un langage Html trop simple est parfois sujet à critique du côté des spécialistes de la TEI qui souhaiteraient une meilleure structuration des contenus. Les premières réflexions au sujet de l'édition électronique se font jour avec des applications dont beaucoup ne sont pas passées à la postérité. Les liaisons avec l'archivistique s'opèrent alors tant le mot « archive » devient fréquemment utilisé pour définir des projets de bibliothèques numériques et de contenus en ligne. Le mot est pratique, désignant la mise à disposition de contenus sans nécessairement de réelle valeur ajoutée « numérique » comme pour *William Blake Archive*<sup>21</sup>.

L'intégration de ces nouvelles tendances se traduit par la présence d'intitulés dans les cours et les diplômes. Le *King's College* de Londres fut ainsi parmi les premiers à développer des diplômes orientés en *humanities computing*. C'est d'ailleurs sans surprise le même établissement qui a donné une forte orientation *Digital Humanities* à ces diplômes en

---

<sup>18</sup> Le projet et ses ressources sont actuellement consultables ici : <http://www.perseus.tufts.edu/hopper/#>

<sup>19</sup> Lou Bernard, *Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique*, *Read/ Write Book 2, une introduction aux humanités numériques*, Open Edition Press, 2012, p.45-58

<sup>20</sup> Cette expression continue d'être conservée pour définir la liste qui est toujours modérée par Willard McCarthy lui-même. <<http://dhumanist.org/>>

<sup>21</sup> William Blake Archive <<http://www.blakearchive.org/blake/>>

La question disciplinaire est aussi centrale, faut-il une discipline à part entière ou doit-on plutôt concevoir un espace interdisciplinaire autour de méthodes comme le décrit Willard McCarthy<sup>22</sup> ?

### **La montée en puissance des corpus numériques et des bases de données**

La constitution progressive des bases de données d'articles numériques<sup>23</sup> marque également un important tournant documentaire ainsi qu'une relation espace-temps fortement modifiée pour les chercheurs. Cette accélération de la disponibilité documentaire n'est pas sans conséquence. La réduction du délai d'obtention d'un article permet de ne pas réfréner un besoin urgent de consultation et évite des attentes pénibles voire décevantes lorsque l'article qui arrive au final ne correspond que peu au besoin d'information initial. La constitution de gisements d'articles en ligne accessible par des abonnements parfois onéreux payés par les bibliothèques s'avère déjà une étape cruciale dans le développement des humanités numériques car l'enjeu complexe de l'accessibilité représente une étape incontournable dans la réalisation des humanités numériques. La prolongation de cette accessibilité via des plateformes d'archives ouvertes l'accroît encore notamment pour un public différent des chercheurs professionnels.

La masse d'informations et d'articles disponibles en quelques clics est déjà une mutation importante. Elle permet indubitablement une accélération des processus d'écriture du fait d'une diminution des temps d'accès. De la même manière, il est plus aisé pour le chercheur d'effectuer des lectures stratégiques d'un article ou d'un ouvrage par la recherche d'occurrences d'un mot ou d'un concept. Plus les articles sont numérisés en mode texte, plus la recherche se trouve facilitée pour repérer des éléments dignes d'intérêts tandis qu'une lecture extensive aurait été beaucoup plus longue et, surtout, plus incertaine. Dès lors il ne s'agit plus de pouvoir lire de manière exhaustive un ouvrage mais de réaliser des recherches à l'intérieur du texte. On peut plus aisément repérer des concordances, des similitudes dans les usages et on peut réaliser des statistiques davantage précises. De nouvelles requêtes sont rendues possibles, le chercheur évolue dans une philosophie qu'exprime une nouvelle fois parfaitement Lou Bernard « *J'avais l'habitude de lire des textes, désormais j'apprends les outils pour jouer avec* ». <sup>24</sup> Dès lors, les outils des chercheurs en sciences humaines et sociales sont en évolution très nette depuis plusieurs décennies. Les environnements personnels de travail sont de plus en plus numériques.

---

<sup>22</sup> Willard McCarthy

<sup>23</sup> Un rôle bien démontré dans Dalbello, M. (2011). A genealogy of digital humanities. *Journal of Documentation* 67(3), 480–506.

<sup>24</sup> Citation originale « I used to read texts, but now I learn the tools to play with them ». Lou Bernard, *idem*, p. 52. [Cette citation peut être rapprochée de la provocation de Franco Moretti: "\[...\] what we really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them." \(Franco Moretti, \*Distant Reading\* \(London: Verso, 2013\), location 793.](#)

Cette production est de plus en plus numérique nativement car les chercheurs écrivent sur traitement de texte depuis bien longtemps et le développement de la documentation numérique disponible ont accru également les facilités d'échanges et de partage de documents ou tout au moins leur signalement. C'est également dans ce cadre que la pratique du blog ou carnet de recherche s'est également constituée. Longtemps marginale, elle gagne lentement en légitimité.

### **Le tournant des humanités numériques**

Le glissement progressif vers l'expression *Digital Humanities* est particulièrement visible dans l'ouvrage *A companion to digital humanities* dirigé par John Unsworth and Ray Siemens qui vont populariser le terme et par la même occasion impulser une nouvelle dynamique. C'est aussi une volonté de mieux comprendre les évolutions en cours et ne plus considérer les outils informatiques comme de simples outils permettant de mieux analyser les corpus. C'est aussi un rééquilibrage entre la raison computationnelle et la raison humaine : l'expression marque davantage une fusion qu'une juxtaposition comme c'était le cas pour *Humanities computing*. Il faut bien reconnaître que malgré quelques pionniers, la vision dominante était celle d'une délégation à l'ingénierie informatique pour tout ce qui relevait des questions de développement et de gestion des logiciels employés. Désormais, la montée en puissance du numérique touche toutes les phases du travail et de la production scientifique y compris en sciences humaines et sociales. Mais cette convergence va plus loin, tant il s'agit d'un dialogue au sein même des sciences humaines et sociales, le numérique en étant l'instrument facilitateur. Ce point est particulièrement mis en avant par le texte *A digital humanities manifesto*<sup>25</sup> et son prolongement *The Digital Humanities Manifesto 2.0*

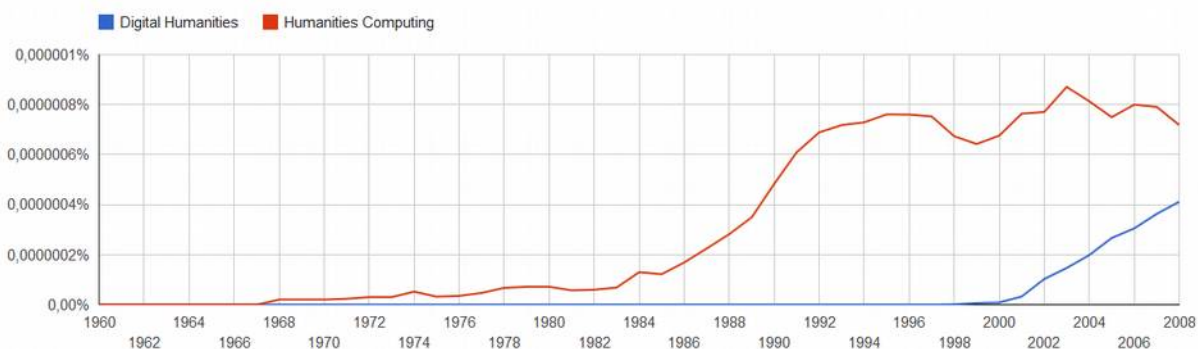
La fréquence d'utilisation de l'expression de digital humanities peut être symbolisée par une visualisation proposé par l'outil *Google Ngrams* qui effectue des requête sur les occurrences d'une expression dans les ouvrages numérisés de google livres. Si *humanities computing* reste l'expression majoritaire en 2008 (date des dernières données consultables) l'expression de *digital humanities* montr en flèche<sup>26</sup>..

---

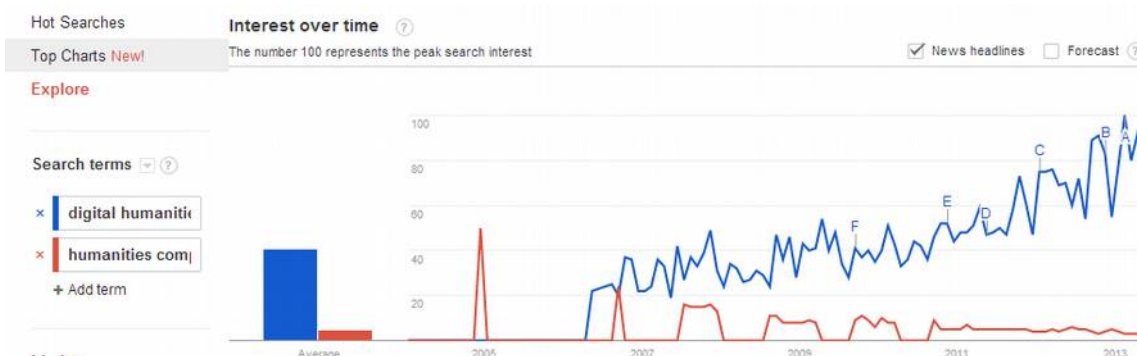
<sup>25</sup> A digital humanities manifesto, 15 décembre 2008.

<http://manifesto.humanities.ucla.edu/2008/12/15/digital-humanities-manifesto/>

<sup>26</sup> Toutefois, les imprécisions et les erreurs qui émaillent les métadonnées et le fait que Google Ngrams soit sensible à la casse obligent à une certaine distance. Nous avons ici fait le choix d'utiliser les deux expressions avec majuscule pour la première lettre de chaque mot, ce qui correspond, à notre connaissance, à l'usage le plus courant.



Finalement, c'est un autre outil de Google qui montre cette montée en puissance : Google Trends, qui est basé sur les requêtes produites par les usagers du moteur Google. Les résultats montrent une prise d'intérêt vers 2005 et une montée continue tandis que celle d'*humanities computing* tend à décliner après un pic autour de 2005. Incontestablement depuis 2005, à l'heure de l'avènement de ce que l'on appelle depuis le web 2.0, un tournant semble s'être produit.



L'influence de Google se mesure aussi à sa gigantesque entreprise de numérisation autour de Google Livres (google books). Toutefois, la numérisation reste encore imparfaite et peu riche en fonctionnalités ajoutées.

### THATcamp, les points forts des DH

Parmi les points marquants du mouvement figurent l'événementialisation autour des journées d'études, colloques et universités d'été qui se déroulent régulièrement sur les *digital humanities*. Parmi ces événements figurent les désormais fameux THATCamp. Issus d'une réflexion sur le fait que les colloques sont trop conventionnels et qu'ils peuvent engendrer de l'ennui lors des sessions officielles, l'objectif a été d'envisager une non-conférence, c'est-à-dire des temps d'expression moins formalisés et plus ouverts qui donnent à chacun des participants l'opportunité de s'exprimer. Depuis leur création, les THATCamp connaissent un succès indéniable, organisés en des lieux variés et sur des thématiques hétérogènes. Même si le THATCamp souhaite se démarquer du colloque ou de la journée d'études classiques, le dispositif n'est pas pour autant en dehors de toute institutionnalisation et d'édification de règles d'usages.

Il reste désormais à passer la barrière de la popularisation de l'expression pour toucher des publics plus larges et pour faire évoluer les DH en dehors des seules universités. Les enjeux se



portent sur l'ensemble de l'éducation et des lieux de culture et de savoir et sans doute encore au-delà.