

Cours d'analyse de données en géographie

Niveau Master 1 - GEANDO

Humanités numériques

Maxime Forriez^{1,a}

¹ Institut de géographie, 191, rue Saint-Jacques, Bureau 105, 75 005 Paris,

^amaxime.forriez@sorbonne-universite.fr

4 septembre 2025

1 Introduction générale

Les sciences humaines et sociales (S.H.S.) et les outils informatiques ont une histoire qui est déjà relativement fournie. Comme dans la plupart des domaines, l'informatique a apporté une **transformation**, ou plutôt une révolution dans le cas des S.H.S. dans son sens premier de retour en arrière [Abitoul et Hachez-Leroy, 2015].

Jusqu'au début du XIX^e siècle, la connaissance ne distingue guère la séparation classique qui paraît normale, naturelle, de nos jours, à savoir la distinction entre les sciences et les lettres. Celle-ci date, en France, de l'université napoléonienne. À l'époque de l'industrialisation, il était clair que l'on opérait cette séparation dans la connaissance entre ce qui était utile au projet de la modernité, les sciences, et ce qui lui était inutile, les lettres.

Avec le développement de techniques permettant la gestion automatique de données textuelles, jusque là très loin du monde industriel, la réintégration des lettres dans la connaissance utile devient une évidence [Abitoul et Hachez-Leroy, 2015]. En Amérique du Nord, les lettres sont plutôt appelées **humanités**, et ce terme possède un sens beaucoup plus large que son sens en français. Les humanités dont il est question, sont : l'histoire, la géographie, la linguistique, la littérature, les arts, l'économie, la sociologie, le droit, la théologie, *etc.* [Abitoul et Hachez-Leroy, 2015].

De son côté, le **numérique** est entendu en tant qu'outil de communication, instrument de recherche et objet de recherche [Abitoul et Hachez-Leroy, 2015]. On pourrait ajouter que cela permet d'effectuer des calculs improbables au niveau de la rédaction, comme compter le nombre de mots d'un texte, le nombre d'espaces, le nombre de paragraphe, *etc.* Tout ce qui était difficile semble devenir plus simple avec l'usage d'un ordinateur.

Dans un monde où l'ordinateur n'est plus un simple outil de calculs, l'articulation entre le monde numérique et science est au cœur de l'informatique [Abitoul et Hachez-Leroy, 2015].

Certains vont même jusqu'à avancer que cela renouvelle le vieux projet **humaniste** du XVI^e siècle. En effet, l'accumulation des savoirs et des connaissances tout au long des XVIII^e, XIX^e et XX^e siècles a fragmenté, spécialisé la connaissance scientifique. Contrairement au projet humaniste de l'homme universel, il était devenu normal d'être spécialisé, d'avoir une expertise précise. Il n'était également pas rare d'avoir un sujet de niche. Avec l'accès à l'informatique, cette segmentation pourrait s'achever, et aboutir à l'invention d'un nouvel humanisme. En cela, le choix du terme « humanités » est d'autant plus pertinent que le terme « lettres ».

Que l'on soit en sciences ou en lettres, bases de données et moteurs de recherche utilisent tout aussi bien des données numériques (dans le sens restreint de nombre) que des données textuelles. Par ailleurs, il existe de nombreux algorithmes communs en sciences et en lettres. La frontière, si elle existe toujours dans de nombreux esprits, a disparu depuis longtemps. La connaissance, qu'on le veuille ou non, a été transformée par la pensée informatique, **pensée computationnelle**, ce qui conduit naturellement à enseigner différemment [Abitoul et Hachez-Leroy, 2015].

Concernant les humanités, la pensée computationnelle a conduit à davantage de rigueur, dans le sens où les bases de données et les algorithmiques qui leur sont associés, permettent une vérification plus rapide, une recherche de l'information nettement plus performante, *etc.* Toutefois, les ordinateurs sont également un moyen puissant de communication dans lequel le sacro-saint écrit perd de sa puissance. Désormais, la connaissance n'est plus excluante. Elle se diffuse grâce à différents supports : le son, la vidéo, l'image, et ce à partir d'un texte auquel on peut greffer des liens hypertextuels. Écrire au XXI^e siècle n'est plus la même chose qu'au XX^e siècle. Un enseignant, par exemple, peut donner en temps réel accès à ses sources à ses apprenants qui peuvent se les approprier à leur tour. Un chercheur peut faire de même. Correctement utilisé, l'informatique permet un meilleur apprentissage des humanités.

2 L'évolution des rapports entre humanités et informatique : des métadonnées aux humanités numériques

Dès les années 1970, l'informatique fit son entrée dans les humanités autour de deux importants piliers : les métadonnées et les bases de données. Il s'agit de décrire un certain nombre de sources grâce à une fiche contenant des métadonnées d'un livre par exemple (son auteur, son titre, son éditeur, sa date, *etc.*). Parallèlement, on commençait à constituer d'immenses bases de données, dont les plus célèbres en géographie sont les systèmes d'information géographique (S.I.G.).

S'est mis en marche un immense projet de numériser de tout ce qui pouvait l'être : cartes, livres, tableaux, *etc.* À cela s'est couplé le développement d'internet et des liens hypertextuels.

Des travaux furent lancés en linguistique autour du *Literary and Linguistic Computing* (L.L.C.). En soi, c'était logique, développer des bases de données contenant le dictionnaire d'une langue était simple, et quoi de plus simple de les diffuser par internet. Néanmoins, le projet initial était beaucoup vaste. Le dictionnaire enregistré numériquement, il devenait facile de proposer des analyses algorithmiques autour de textes 1. en repérant les concordances, 2. en calculant des fréquences de mots, d'expression, 3. en traitant automatiquement des langues et

4. en créant des corpus [Dacos et Mounier, 2015]. Assez rapidement, le projet évolua vers le **web sémantique**.

Le web sémantique étudie le sens des textes analysés par un algorithme. Il permet un réel traitement de l'information permettant de la transformer en connaissance. C'est le moteur qui alimente les intelligences artificielles (I.A.) telles que ChatGPT. Les I.A. sont en mesure d'écrire des textes élaborés grâce à ses travaux. Toutefois, il reste nécessaire de les enrichir régulièrement avec des informations de qualité et non ambiguës. Les I.A. permettraient de réaliser une synthèse de la connaissance accumulée des humanités, donc de manière utopique de tout synthétiser dans une base de données universelle. Néanmoins, dans ce projet humaniste, les scientifiques constituent un garde-fou. Il est nécessaire d'utiliser leurs textes, car l'I.A. pourrait apprendre des « vérités » erronées. Il suffit de penser à l'I.A. Tay, devenue raciste, antisémite et xénophobe en moins de 24 h. C'est pour cela que le web sémantique utilise le terme « ontologie » dans sa formalisation des connaissances.

Les bases de données ont évolué, quant à elle, vers une structure relationnelle. Techniquement, par un système de clé-valeur, il est possible de relier entre elles les bases de données. Cela permet de limiter la quantité d'information stockée, tout en rendant performant les liens existants entre les bases de données.

Si l'ordinateur effectue techniquement toujours des calculs (même avec des données textuelles), les **humanités assistées par ordinateur** construisent des liens. On est passé progressivement d'un outil de calcul à un moyen de communication.

Par ailleurs, les métadonnées finirent par être standardisées par un langage informatique, le XML. C'est le *Text Encoding Initiative* (T.E.I.), datant de 1987, qui porta le projet. C'est cet événement qui permit à Willard McCarty de parler de *Humanities Computing* (H.C.) [McCarty, 2005].

Cette phase s'intéresse à six processus, ou plutôt les accélère considérablement [Abitoul et Hachez-Leroy, 2015].

1. Il faut opérer à la numérisation de tout ce qui peut l'être dans le but de créer des bases de données, des liens hypertextuels et une diffusion sur internet.
2. Les réseaux numériques favorisent l'interaction et la communication. Les humanités disposent de bibliothèques en ligne par exemple. Il est de moins en moins nécessaire de se déplacer physiquement en bibliothèque, surtout celles contenant des documents fragiles, voire uniques.
3. Le développement du web sémantique ne se limite plus aux textes, mais à tout le **patri-moine culturel** grâce au *Concept Reference Model*.
4. La collecte et l'analyse de données peuvent alimenter des algorithmes comme ceux de la fouille de texte et de son analyse, de l'approche statistique des langues avec notamment la traduction automatique, et de la recherche d'information afin de produire de nouvelles connaissances.
5. L'informatique permet la modélisation et la simulation par des automates cellulaires (A.C.) ou des systèmes multi-agents (S.M.A.).
6. L'archivage va au-delà des bibliothèques en lignes. La production numérique engendre des contenus non physiques. Faut-il les conserver uniquement sous forme numérique ou

doit-on les imprimer pour avoir une copie physique ? Loin d'être anodine, cette question est très complexe. De manière sous-jacente, comment préserver la qualité de l'information ? Comment conserver massivement l'information ? Que conserver ? livres numérisés ? tweets ? *etc.* ? Ces questions sont déjà classiques pour l'archivage physique, mais sont amplifiées avec les données numériques. De plus, si un événement empêche l'accès aux informations virtuelles définitivement, comment retrouver les informations, savoirs et connaissances perdus ?

Ces six processus permettent la **mise en réseau** de toute chose, le lien hypertextuel assurant le lien entre les informations (les métadonnées). C'est ce que David Berry appelle le tournant computationnel¹ [Berry, 2011] et Patrik Svensson, la zone d'échange² [Svensson, 2012].

Cette mise en réseau massive et ultra-rapide (moins d'un décennie) a fait basculer les humanités assistées par ordinateur vers une nouvelle dimension que l'on appelle **humanités numériques**³, c'est-à-dire faire des recherches autrement. « Les humanités numériques recouvrent un ensemble de pratiques de recherche à l'intersection des technologies numériques et des différentes disciplines des sciences humaines » [Dacos et Mounier, 2015, p. 7]. Dans son blog, Melissa Terras parle de *Big Tent* [Terras, 2011]. elle utilise une métaphore autour d'un chapiteau pour expliquer l'aspect transdisciplinaire des humanités numériques, qui rassemblent les méthodes et les outils liés au numérique afin de produire des contenus, des recherches, en S.H.S.

Cette phase dans laquelle les humanités sont actuellement, repose plus que jamais sur l'usage d'outils de communication informatiques qui ont bouleversé le monde académique : les médias sociaux, les *chats* et les courriers électroniques, les blogs, les multimédias, les wikis, les données en *open access*, *open data*, *open peer review*, *open commentary*, les *data visualisation*, *etc.* Jamais les humanités n'ont autant partagé qu'avec l'avènement des sciences de l'information et de la communication (S.I.C.). En parallèle, les publications académiques ont un certain nombre d'indicateurs qui n'existeraient pas avant. Il est également possible de débusquer plus facilement un plagiat. Les acteurs des humanités numériques gagnent en crédibilité et en visibilité.

Le juriste Lawrence Lessing constata une inadaptation des lois sur le copyright par rapport à l'émergence du web. Pour lui, il s'agit d'un frein à la diffusion de la culture. Par ailleurs, il distingue la propriété intellectuelle de la propriété physique. Il créa en 2002 l'organisation *Creative Commons*. Elle poursuit deux objectifs principaux : 1. faciliter la diffusion et le partage des œuvres ; 2. faciliter la création de contenus numériques. Elle propose l'institution de sept licences dont les buts sont : 1. de protéger les droits d'auteur d'une œuvre artistique ; 2. d'encourager la libre circulation des contenus culturels ; 3. de contribuer à un patrimoine d'œuvres accessibles **librement** à tous. Les licences sont :

Lawrence
Lessing
(né en
1961)

1. la licence zéro (CC0) qui est un renoncement total aux droits patrimoniaux de l'auteur d'une œuvre numérique ;
2. l'attribution (BY) ;
3. le partage (SA) ;

1. En anglais : *Computational Turn*

2. En anglais : *Trading Zone*

3. En anglais : *Digital Humanities*

4. le contenu qui est modifiable ou non modifiable (ND) ;
5. le contenu qui est commercial ou non commercial (NC) ;
6. la licence « Tous droits réservés » qui n'offre aucun droit de copie et de partage.

Il est à remarquer que les licences sont restrictives. Si, par exemple, on ne choisit pas NC, alors le contenu est commercialisable. Les créateurs de contenu que peuvent être les chercheurs, doivent être attention à la licence utilisée par lui-même ou par le site qui publie son contenu. Par défaut, tout contenu est CC0.

Néanmoins cette nouvelle dimension de la recherche autour des humanités a ses détracteurs, tels que Hubert Guillaud en 2022. « En tant que non-spécialiste, on a du mal à voir, saisir l'apport de cette nouvelle forme de science, au-delà de ce qu'elle apporte pour le chercheur. On constate bien dans le *Manifeste des Digital Humanities* [Dacos, 2011] qu'il y a d'abord cette volonté de faire communauté autour de l'intégration de la culture numérique dans les pratiques de recherche d'aujourd'hui. [...] On est beaucoup dans la structuration d'une discipline, alors qu'on souhaiterait surtout comprendre ce que ces outils apportent concrètement » [Guillaud, 2010]. Sa critique est utilitariste. L'auteur n'a pas conscience que le support numérique qu'il utilise, *La Feuille*, est l'enfant des humanités numériques. C'est toute la beauté des S.H.S. Leur utilité est inexistante dans un monde techno-industriel, mais elles sont pourtant essentielles pour comprendre le monde tout court. Toutefois, il est clair que « les humanités sont le vecteur d'une véritable « utopie numérique » qui a d'ailleurs fait l'objet de nombreuses études » [Dacos et Mounier, 2015, p. 23].

Les humanités numériques souffrent par ailleurs d'une faiblesse structurelle. Il existe peu de centres de recherche, peu de projets, *etc.* Néanmoins, cela est dû à sa nature transdisciplinaire. Il y a quelques années. C'était la « complexité » qui fédérait les actions transdisciplinaires. Pourtant, aucun centre de recherche s'appelant simplement « complexité » n'a vu le jour. C'est ça la difficulté majeure des humanités numériques. Si les outils sont universels, les disciplines académiques se maintiennent et structurent l'enseignement et la recherche.

Le dernier aspect des humanités numériques concerne les propriétés de l'objet numérisé. Comment garantir les éléments fondamentaux permettant de mettre en avant son esprit critique ? L'analyse de documents se fonde, d'une part, sur une critique interne, et, d'autre part, sur une critique externe. La critique interne se subdivise en trois éléments : 1. la critique d'interprétation ; 2. la critique d'exactitude ; 3. la critique de sincérité. La critique d'exactitude a désormais le soutien du web sémantique, puisqu'elle a pour objectif de déterminer la signification d'un document par sa sémantique et ses données chiffrées. Toutefois, ce sont souvent les non-dits d'un document, le non-visible, tout ce qui est masqué, caché, volontairement ou inconsciemment par son auteur, qui l'explique. Est-ce possible automatiquement ? Le web sémantique peut-il fournir aux humanités numériques des moyens pour élaborer une représentation des données qui soit plus orientée vers l'expression du sens ? [Paquien-séguy et Pélissier, 2021] Si la réponse était affirmative et tranchée, on se dirigerait vers un web herméneutique. La critique d'exactitude est rendue beaucoup facile avec le web sémantique, puisqu'elle cherche à déterminer si l'auteur se trompe (en toute bonne foi), par manque d'informations par exemple. La critique de sincérité, au contraire, tente d'établir si l'auteur a fait exprès d'oublier telle ou telle information afin de tromper ses lecteurs, auditeurs, *etc.* À l'époque des *fake news*, elle est de plus en plus essentielle. Toutefois, une I.A. est-elle capable de la repérer ? Si on s'en

tient aux dernières nouvelles, ce n'est pas certain. La critique externe se subdivise également en trois éléments : 1. la critique d'identité ; 2. la critique de véracité ; 3. la critique de restitution. La critique d'identité est garantie numériquement par les métadonnées. C'est par elles que tout a commencé, car elles sont essentielles. Elles établissent qui est l'auteur, sa biographie, son degré de popularité. Elles permettent d'enregistrer une date et un lieu de rédaction. La critique de véracité est certainement la plus subtile à programmer. Un acte juridique est-il vrai ou faux ? Cette question tente de déterminer si cet acte est attesté par une autorité publique. Cette critique n'a rien à voir avec l'établissement de l'authenticité du document. En effet, de nombreux documents officiels sont remplis de contre-vérités. Un chercheur essaye de comprendre les motifs de celles-ci. La critique de restitution est largement facilitée par les algorithmes. Il s'agit d'établir l'évolution d'un document. Ce dernier est-il original ou n'est-il qu'une copie, voire une copie de copie ? Pour les historiens et les linguistiques, c'est une information essentielle lorsqu'il s'agit de documents anciens, car les copistes peuvent avoir ajouté ou retiré des éléments du texte original. À l'ère du numérique, les copies de copie ne semblent plus poser de problèmes. Toutefois, comment assurer l'intégrité et la pérennité d'un objet numérique ? Si on prend une image numérisée ou numérique, le degré de compression peut fortement altérer l'information initiale. Comment gérer la diversité des contenus ? Désormais, la conversation ne concerne plus les seuls textes ou images, mais également des vidéos, des documents hypertextuels, des flux comme les tweets, *etc.* Ainsi, de nouvelles techniques comme la reconnaissance de formes et d'analyse d'images deviennent essentielles, mais mal maîtrisées par les chercheurs en S.H.S. L'informatique a fortement, voire trop, diversifié les sources d'information. Comment distinguer l'essentiel, l'important du superflu ? Toutes ces critiques peuvent toutefois bénéficier d'un grand soutien des procédures de traitement automatique des grands corpus. Il en découle une question essentielle. Les humanités numériques peuvent-elles introduire, avec l'analyse des données, la méthode expérimentale au sein de ses disciplines ?

À la différence des critiques énoncées, un élément a davantage d'importance à l'ère numérique : l'authentification tant des utilisateurs que des objets. Il faut préciser que, de manière informatique, la virtualité des utilisateurs en fait des objets. De fait, comment authentifier un objet numérique ? Une garantie juridique concerne, par exemple, la signature numérique, garantie par des algorithmes cryptographiques. Celle-ci garantit que l'objet est bien un utilisateur connu et reconnu, mais ces algorithmes aussi puissants soient-ils, ont montré, montre et montreront toujours des lacunes, des faiblesses.

3 Le projet des humanités numériques

Il y eut des précurseurs avec l'approche médiologique de Régis Debray [Debray, 1991] [Debray, 2000], ou l'approche des *Digital Studies* de Bernard Stiegler [Stiegler, 2014].

Avec la médiologie, Régis Debray pose le postulat que le **medium**, c'est-à-dire l'outil, précède l'idée. Dit autrement, la **technè** précède la pensée. Historiquement, les exemples ne manquent pas. Le développement du christianisme sont liés à celui du *volumen*, celui du protestantisme, à celui de l'imprimerie, celui du socialisme, à celui de l'industrie, et le médialisme, à celui du web. Plus profondément, la médiologie permet de comprendre et d'étudier

Régis
Debray
(né en
1940)

le phénomène d'acculturation tant au niveau de la logosphère que de la vidéosphère, de l'audiosphère, *etc.* D'un point de vue historique, la transmission des connaissances s'effectue en trois temps : 1. le temps de la traduction ; 2. le temps de la diffusion ; 3. le temps de l'adoption. Concernant les humanités numériques, et plus largement le web, le temps de la traduction se situe entre les années 1950 et 1970. Même si l'informatique impose l'anglais, sa traduction est nécessaire afin de former des ingénieurs. Le temps de la diffusion peut être daté entre les années 1970 et 2010. Le web se développe dans tout l'Occident. Depuis les années 2010, il semble que la troisième phase se soit amorcée. Elle est plus délicate, car, contrairement, aux livres, à l'imprimerie et à l'industrie, l'usage du web nécessite un apprentissage et un investissement personnel très important et constant. Une « société connectée » n'est pas forcément quelque chose de désirable. Pour l'instant, seule une élite (politique et industrielle) a un intérêt à sa création, mais on est loin d'un projet collectif aussi porteur que les grands bouleversements techniques évoqués.

Les *Digital Studies* de Bernard Stiegler s'inscrivent dans les humanités numériques. Sa thèse propose la distinction entre deux camps : 1. ceux qui étudient les effets du numérique d'un point de vue sociologique, anthropologique, culturelle, *etc.* ; 2. ceux qui étudient les techniques. Son analyse est difficilement contestable. Les chercheurs en S.H.S. restent peu à l'aise avec l'usage des ordinateurs. Certains font même un rejet total en dehors des logiciels de traitement de texte. Bien entendu, la situation est loin d'être homogène. Chacun individu réagit différemment, mais, encore de nos jours, il est fréquent que les étudiants, hormis l'usage des réseaux sociaux, n'ont aucune culture numérique des traitements de texte, des tableurs, des bases de données, même parfois des moteurs de recherche. Ce constat ne peut que faire douter sur le projet de société du tout numérique. Pourtant, cette méconnaissance n'est pas un frein à l'étude du phénomène numérique. Aucune des deux positions n'est valable en soi. Il faut être, en tant que chercheur en S.H.S., être dans un entre-deux acceptable.

Bernard Stiegler (1952-2020)

Le projet des humanités numériques se déclinent en cinq éléments [Paquienséguy et Péliissier, 2021] :

1. la transformation des pratiques infocommunicationnelles des chercheurs au niveau de (a) la source, (b) la collecte, (c) le traitement, (d) la visualisation, et (e) la valorisation des données ;
2. la revitalisation des filières pédagogiques aux effectifs et à l'attractivité en déclin ;
3. les nouvelles formes de dialogue interdisciplinaire autour de programmes fédérateurs ;
4. les possibilités de renouvellement de l'écriture et de la diffusion de la recherche à partir des outils développés par la science ouverte et collaborative ;
5. les opportunités de renforcer la scientificité des méthodes en permettant une meilleure complémentarité entre recherches qualitatives et quantitatives.

4 Les enjeux des humanités numériques

Les enjeux des humanités numériques s'organisent autour de six axes [Paquienséguy et Péliissier, 2021] :

1. inscrire les humanités numériques dans la modernité des « humanités » telles qu’elles se présentent au XXI^e siècle ;
2. placer au premier plan de l’agenda scientifique l’étude des transformations de la société sous l’effet des technologies numériques ;
3. assumer la charge de travail en recherche pour identifier et analyser les spécificités du numérique dans chaque secteur de la société contre les simplifications globalisantes dans l’interprétation de la dimension numérique du monde et pour améliorer à terme, par cumul de recherches spécialisées, la compréhension de cette dimension ;
4. assumer la charge de travail d’étude des différenciations de conceptions et d’usages du numérique d’un champ linguistique à l’autre ;
5. retrouver dans les humanités numériques le pluralisme axiologique théorique et paradigmatique qui a fait la richesse des humanités modernes en favorisant l’esprit critique de chaque point de vue vis-à-vis des autres, donc la controverse vertueuse, par ses effets heuristiques ;
6. favoriser la diversité des formes techniques, cadres organisationnels et modèles économiques de l’édition numérique.

5 Les risques des humanités numériques

Le projet des humanités numériques n’est pas sans risques. Ils se déclinent en cinq éléments [Paquienséguy et Pélissier, 2021] :

1. la sous-estimation du contexte social et économique de la production des données ;
2. la priorisation des approches quantitatives au détriment d’approches plus qualitatives ;
3. l’hégémonie accrue de l’informatique sur les sciences humaines avec une mise en compétition des chercheurs en fonction de leurs usages du numériques ;
4. l’illusion scientifique quant au rôle des statistiques dans la compréhension des faits sociaux ;
5. le risque de privilégier le traitement des données à leur interprétation.

6 Les devoirs des humanités numériques

Afin de ne pas prendre le risque d’une « physique sociale », telle que se l’imaginent bons nombres d’acteurs scientifiques ou non, il faut rappeler les trois devoirs des humanités numériques [Paquienséguy et Pélissier, 2021] :

1. un devoir de pratique ;
2. un devoir de critique ;
3. un devoir de formation.

Les acteurs des humanités numériques doivent savoir comprendre les informations, c’est-à-dire les données et leur sens évident, ou caché, crypté.

7 Le questionnement éthique autour des humanités numériques

Le chercheur en humanités se doit [Paquienséguy et Pélissier, 2021] de :

1. définir les étapes de sa recherche en contexte numérique ;
2. définir ses questions de recherche ;
3. opérer et justifier ses choix méthodologiques ;
4. définir son terrain autour des données de sa recherche ;
5. publier numériquement ses résultats.

Toutefois, cela pose trois problèmes [Paquienséguy et Pélissier, 2021] :

1. la délégation logicielle de l'interprétation ;
2. l'industrialisation de la logistique scientifique ;
3. les risques d'une marchandisation accrue de la recherche.

8 Les méthodes des humanités numériques

D'un point de vue technique, les humanités numériques ont besoin qu'un chercheur développe trois compétences [Paquienséguy et Pélissier, 2021] :

1. des compétences techniques ;
2. des compétences analytiques ;
3. des compétences interprétatives et critiques.

Pour y parvenir, il existe trois éléments essentiels [Paquienséguy et Pélissier, 2021].

1. Les principes du codage et de l'algorithmique sont essentiels.
2. Les principes et les techniques des écritures et de l'édition numérique sont au cœur de l'écriture hypertextuelle, qui possède de nombreux usages que n'a pas l'écriture classique.
3. Il faut maintenir une veille dans le traitement et l'usage des données.

Il faut penser, se former ou former, transformer les humanités classiques en humanités numériques.

8.1 Les grandes méthodes

Il existe cinq étapes caractérisant la méthode apportée les humanités numériques [Van Hoo-land et al., 2016] :

1. trouver l'information ;
2. modéliser les données ;
3. numériser les sources ;
4. analyser le contenu ;
5. valoriser les résultats.

8.1.1 Trouver l'information

Trouver l'information nécessite de :

- avoir la capacité à définir une question de recherche ;
- comprendre le fonctionnement d'internet et du web ;
- utiliser des moteurs de recherche ;
- maîtriser les opérateurs booléens, par exemple, pour Google ou Google Scholar :
 - l'opérateur « ET » qui correspond à un espace ;
 - l'opérateur « OU » qui correspond à un tilde ~ ;
 - l'opérateur « NOT » qui correspond à un tiret -. Il permet d'exclure des résultats d'une recherche.

8.1.2 Modéliser les données

Modéliser les données s'effectue :

- par l'application des principes des modèles conceptuels de données (données tabulaires, bases de données relationnelles, langage de balisage (XML), web sémantique) ;
- par l'implémentation des modèles de données à l'aide de formats et d'outils concrets ;
- par l'impact du partage des données et leur évolution dans le temps ;
- par l'interaction entre la recherche structurée et non structurée.

8.1.3 Numériser les sources

Numériser les sources a pour triple objectif de : 1. les conserver, 2. les diffuser et 3. les visualiser.

Numériser les sources implique de :

- comprendre les enjeux liés à la numérisation du patrimoine culturel :
 - déjouer le mythe de l'accès numérique du patrimoine culturel ;
 - gérer le patrimoine qui est source de tensions entre passé et présent.
- survoler les différentes étapes d'un projet de numérisation :
 1. préparer la numérisation ;
 2. numériser ;
 3. créer les métadonnées ;
 4. utiliser un logiciel de gestion des collections.
- savoir créer des métadonnées ;
- avoir un panorama clair et exhaustif des logiciels de gestion de collections numériques.

Numériser les sources pose des problèmes :

- d'ordre financier, car numériser coûte cher ;
- d'ordre logistique. Faut-il numériser en interne ou en externe ?

8.1.4 Analyser le contenu

Analyser le contenu se décompose en cinq éléments :

- comprendre les enjeux du traitement automatique des langues (T.A.L.) :
 - l'ère des données massives⁴ ;
 - l'analyse lexicale avec un logiciel comme UNITEX.
- poser les possibilités et les limites du *Distant Reading* ;
- maîtriser quelques outils simples de fouille de textes ;
- être capable de créer un corpus web, de le nettoyer et de l'exploiter (*scrapping*, *work-flow*) ;
- dompter des techniques de base de visualisation de l'information.

8.1.5 Valoriser les résultats

Valoriser les résultats obtenus par des méthodes issues des humanités numériques s'effectue :

- en écrivant des textes scientifiques à l'aide de LaTeX et de Markdown ;
- en rédigeant des contenus web en HTML et avec ou sans l'aide d'un *Content Management System* (C.M.S.) ;
- en comprenant le fonctionnement et les enjeux autour du *Cloud Computing* ;
- en étant au courant des dérives liées à l'usage des cookies ;
- en pouvant gérer le droit d'auteur en publiant ses travaux de recherche en ligne, ce qui comprend :
 - le droit d'auteur ;
 - le droit à l'image ;
 - la protection de la vie privée ;
 - l'*open source*.

8.2 Les courants existants

Les humanités computationnelles⁵ étudient les aspects culturels des réseaux sociaux.

L'informatique sociale⁶ étudie les impacts sociaux de la culture numérique.

Les méthodes numériques⁷ ont été proposées par Richard Rogers. Il oppose les méthodes exclusivement numériques aux données numérisées [Rogers, 2013] [Rogers, 2019] [Rogers, 2015]. Les méthodes numériques ne concernent que le premier élément.

4. En anglais : *Big Data*

5. En anglais : *Cultural Analytics*

6. En anglais : *Social Computing*

7. En anglais : *Digital Methods*

Le *Computational Communication Science* s'effectue dans le contexte de la science « communication » [Hilbert et al., 2019].

Le *Distant Reading* permet l'exploration de grands volumes de données textuelles d'une manière automatique.

Le web sémantique est un *Resource Description Framework* (R.D.F.). Il repose sur une ontologie répondant au modèle logique suivant : un sujet, un prédicat et un objet.

9 Conclusion générale : problématiques à résoudre

Il s'agit d'exposer les problèmes philosophiques, et non techniques.

1. Que sont les humanités numériques ?
2. En quoi constituent-elles une évolution importante pour les recherches en S.H.S. ?
3. Comment le champ de recherche des humanités numériques est-il structuré ?
4. Quels sont les acteurs les plus importants ?
5. Comment passer des données aux connaissances ? Comment mutualiser les savoirs ?
6. Comment ne pas limiter les humanités numériques aux équations ou aux algorithmes ? Cette question correspond à la limite de l'objectivité [Abitoul et Hachez-Leroy, 2015] afin d'éviter de créer une « physique sociale » [Paquenséguy et Pélissier, 2021].
7. Comment les humanités numériques peuvent participer à l'invention d'un nouvel humanisme ?
8. À quoi servent les humanités numériques ? Il s'agit de la limite utilitariste.
9. Qu'apportent les humanités modernes sur le tournant numérique de la société ?
10. Que peut-on faire avec les outils numériques dans les humanités numériques ?
11. Comment garantir les propriétés d'un objet numérique ? son identité ? son authenticité ? son intégrité ? sa pérennité ?
12. Comment les humanités sont-elles transformées par le numérique ?
13. Quelle est la géopolitique des humanités numériques ? Quelles sont les stratégies de recherche ? Quels en sont les acteurs principaux ?

Références

- [Abitoul et Hachez-Leroy, 2015] ABITOUL, S. et HACHEZ-LEROY, F. (2015). Humanités numériques, pages 43–57.
- [Berry, 2011] BERRY, D. M. (2011). The computational turn : Thinking about the digital humanities. Culture Machine, 12:1–22.
- [Dacos, 2011] DACOS, M. (2011). Manifeste des Digital Humanities.
- [Dacos et Mounier, 2015] DACOS, M. et MOUNIER, P. (2015). Humanité numérique. État des lieux et positionnement de la recherche française dans le contexte international.
- [Debray, 1991] DEBRAY, R. (1991). Cours de médiologie générale. Bibliothèques des idées. Gallimard, Paris.
- [Debray, 2000] DEBRAY, R. (2000). Introduction à la médiologie. Collection Premier Cycle. P.U.F., Paris.
- [Flechy, 2001] FLECHY, P. (2001). L’imaginaire d’internet. La Découverte, Paris.
- [Gold, 2012] GOLD, M., éditeur (2012). Debates in the Digit Humanities, pages 1–532. University of Minnesota Press, Minneapolis.
- [Guillaud, 2010] GUILLAUD, H. (2010). Qu’apportent les digital humanities ? La Feuille.
- [Hilbert et al., 2019] HILBERT, M., BARNETT, G., BLUMENSTOCK, J., CONTRACTOR, N., DIESNER, J., FREY, S., GONZÁLEZ-BAILÓN, S., LAMBERSO, P. J. et PAN, J. (2019). Computational communication science : A methodological catalyzer for a maturing discipline. International Journal of Communication, 13:3912–3934.
- [McCarty, 2002] MCCARTY, W. (2002). Humanities computing : Essential problems, experimental practice. Literary and Linguistic Computing, 17(1):103–125.
- [McCarty, 2005] MCCARTY, W. (2005). Humanities Computing. Palgrave-Macmillan, London.
- [Paquienséguy et Péliissier, 2021] PAQUIENSÉGUY, F. et PÉLISSIER, N., éditeurs (2021). Questionner les humanités numériques, pages 1–298. Société française des sciences de l’information et de la communication & Conférence permanente des directeurs de laboratoires en sciences de l’information et de la communication, Paris.
- [Rogers, 2013] ROGERS, R. A. (2013). Digital Methods for Web Research. The M.I.T. Press, Cambridge (M. A.).
- [Rogers, 2015] ROGERS, R. A. (2015). Digital methods for web research. Emerging Trends in the Social and Behavioral Sciences, pages 1–22.
- [Rogers, 2019] ROGERS, R. A. (2019). Doing Digital Methods, pages XVIII–234. Sage, London.
- [Stiegler, 2014] STIEGLER, B. (2014). Digital Studies : organologie des services et technologies de la connaissance. FYP, Limoges.
- [Svensson, 2012] SVENSSON, P. (2012). Envisionning the digital humanities. Digital Humanities Quarterly, 6(1).

- [Terras, 2011] TERRAS, M. (2011). Peering inside the big tent : Digital humanities and the crisis of inclusion. Melissa Terras' Blog. <https://melissaterras.blogspot.com/2011/07/peering-inside-big-tent-digital.html>.
- [Turner, 2012] TURNER, F. (2012). Aux sources de l'utopie numérique de la contre-culture à la cyberculture : Steward Brand, un homme d'influence. C. & F., Caen.
- [Van Hooland et al., 2016] VAN HOOLAND, S., GILLET, F., HENGCHEN, S. et DE WILDE, M. (2016). Introduction aux humanités numériques : méthodes et pratiques. Méthodes en sciences humaines. De Boeck, Louvain-la-Neuve.

Table des matières

1	Introduction générale	1
2	L'évolution des rapports entre humanités et informatique : des métadonnées aux humanités numériques	2
3	Le projet des humanités numériques	6
4	Les enjeux des humanités numériques	7
5	Les risques des humanités numériques	8
6	Les devoirs des humanités numériques	8
7	Le questionnement éthique autour des humanités numériques	9
8	Les méthodes des humanités numériques	9
8.1	Les grandes méthodes	9
8.1.1	Trouver l'information	10
8.1.2	Modéliser les données	10
8.1.3	Numériser les sources	10
8.1.4	Analyser le contenu	11
8.1.5	Valoriser les résultats	11
8.2	Les courants existants	11
9	Conclusion générale : problématiques à résoudre	12
	Références	13