

## Les sources nativement numériques pour les sciences humaines et sociales

Camille Paloque-Berges

Il est parfois difficile de cerner le sens du terme « numérique ». Adjectif à l'origine, il renvoie aux nouvelles méthodes technoscientifiques<sup>1</sup> de traitement de l'information par le codage et le calcul à travers les programmes informatiques, et se différencie des techniques dites analogiques. Par extension, et avec la massification des applications et des usages de l'informatique, il en est venu à qualifier les pratiques sociales et culturelles médiées par les ordinateurs et les réseaux. Substantivé, « le numérique » signifierait l'existence d'une nouvelle culture dont les sous-basements technologiques deviendraient faits et structures de société ; il participe au champ sémantique de l'innovation, depuis son sens économique le plus prégnant jusqu'à son sens philosophique<sup>2</sup>. Pour les historiens de l'innovation, l'apparition conjointe du numérique à celle de l'informatique il y a plus d'un demi-siècle, ainsi que son expansion favorisée par le développement des micro-ordinateurs depuis la fin des années 1970, constituent une rupture et une nouvelle étape dans l'évolution des grands systèmes techniques<sup>3</sup> ; pour les anthropologues de l'écriture, il en est de même concernant l'évolution des technologies de l'intelligence, à penser cependant dans une continuité avec les systèmes de signes et d'inscription (Goody, 1979 ; Herrenschildt, 2007). Entre rupture et continuité, le sujet est forcément attrayant pour les sciences humaines et sociales (SHS), en particulier en termes de fabrication et de manipulation des sources conditionnées, voire produites, dans l'usage des nouvelles technologies. Afin d'aborder les problématiques pratiques, méthodologiques et épistémologiques posées aux SHS par les sources « nativement numériques » (ou SNN<sup>4</sup>) – c'est-à-dire les traces, données, documents, et artefacts issus directement du contexte numérique<sup>5</sup>, et pouvant acquérir une valeur de source pour le chercheur en SHS, nous avons organisé une journée d'étude<sup>6</sup> ; cet article en

---

<sup>1</sup> Selon Gilbert Hottois, les technosciences s'appuient sur les dimensions opératoires et techniques des sciences : « la théorie elle-même est conçue comme un outil ou un instrument pour l'action et la transformation de l'expérience [...] est réel ce qui est (re)productible, manipulable, transformable et non plus le visible, l'intelligible ou le compréhensible » (Hottois, 1984, p. 61-62).

<sup>2</sup> En termes philosophiques, Bruno Bachimont définit le numérique comme « l'expression la plus pure de la technique, en tout cas de son essence » (Bachimont, 2010 : 19). La définition économique relève de l'usage courant du terme « le numérique » en référence aux transformations technologiques récentes du secteur des télécommunications et de l'audiovisuel.

<sup>3</sup> La notion de système technique qualifie un ensemble de technologies à une époque donnée liées par une interdépendance, une cohérence et une synergie, ensemble corrélatif à un système social (Gille, 1978).

<sup>4</sup> Expression traduite et adaptée de l'anglais « *born digital* » forgée pour répondre au besoin de nommer les nouvelles productions numériques candidates à la patrimonialisation (UNESCO, 2003).

<sup>5</sup> Nous distinguons ces sources numériques natives (SNN) des sources issues d'une étape de numérisation ou d'encodage par le numérique de documents existant au préalable.

<sup>6</sup> Cette journée, intitulée « Les sources nativement numériques pour la recherche en histoire et patrimoine : pratiques et méthodes » s'est déroulée au Cnam le 09/12/2013 [<http://technique->

fait la synthèse tout en proposant un retour d'expérience empirique de l'auteure sur ces matériaux doublé d'un état de l'art analytique de la pratique des SNN (et non de la notion de « SNN », encore jamais traitée en tant que telle).

Cette journée d'étude a été placée sous le signe des sources pour l'histoire à dessein : il s'agissait d'attirer l'attention d'une discipline pour laquelle la notion de source est centrale, mais qui pourtant est encore éloignée de la question du « nativement numérique<sup>7</sup> ». L'historiographie contemporaine, se penchant explicitement sur la nécessité d'intégrer aux corpus des sources variées au moins depuis le tournant de l'école des Annales, est depuis longtemps armée pour faire face à leur hétérogénéité. Si les sciences auxiliaires de l'histoire prennent en charge les sources anciennes d'inscriptions, de documents et de matériaux aux formes diverses, un nouveau chantier les attend dans les formalismes et les calculs composant les environnements numériques et relevant du code informatique, « troisième écriture » après « la langue » et « le nombre » selon Clarisse Herrens Schmidt (2007)<sup>8</sup>. L'« histoire numérique » prend en compte les nouvelles opportunités que l'informatique offre en termes de stockage, d'encodage et de traitement analytique des données, mais aussi de reconfiguration de l'épistémologie et de la diffusion des savoirs historiens alors que les environnements sociaux et économiques sont marqués par un tournant non seulement computationnel, mais plus généralement numérique (Lamassé et Rygiel, 2014 ; Clavert et Noiret, 2013 ; Delalande et Vincent, 2011). Elle dialogue avec les autres disciplines rassemblées au sein des humanités numériques<sup>9</sup>, et certaines des réflexions présentées ici leur sont transversales : la difficulté à faire face à la

---

[societe.cnam.fr/reseaux-/les-sources-nativement-numeriques-pour-la-recherche-en-histoire-et-patrimoine-pratiques-et-methodes--616161.kjsp](http://societe.cnam.fr/reseaux-/les-sources-nativement-numeriques-pour-la-recherche-en-histoire-et-patrimoine-pratiques-et-methodes--616161.kjsp). Elle a été organisée avec le soutien du Labex HASTEC (Hésam Université) au sein duquel l'auteure de cet article a effectué sa recherche post-doctorale, ainsi que le projet PATRIMONIUM (Peps CNRS), les laboratoires DICEN et HT2S du Cnam. Dans la suite, chaque mention d'un nom sans référence bibliographique désigne un intervenant de la journée selon le programme suivant : panel « Faire confiance aux sources numériques ? Archives et documents entre mémoire, vérité et preuve » animé par Louise Merzeau, avec Marie-Anne Chabin (« Analyse diplomatique des effets du numérique sur la production des documents de preuve à vocation historique »), Sophie Derrot (« Le dépôt légal du Web : vers un patrimoine numérique »), Emmanuel Ruzé (« Quelques approches récentes des archives communautaires numériques ») ; panel « Des artefacts techniques comme sources : nouveaux documents pour l'histoire des technologies en société », animé par Valérie Schafer, avec Alexandre Hocquet et Camille Paloque-Berges (« Collectifs scientifiques et techniciens étudiés à partir de leurs échanges en ligne »), Francesca Musiani (« Le virtuel entre matérialité et invisibilité : étudier les infrastructures numériques »), Benjamin Thierry (« "Le village sans sauvage" : une approche ethno-historique du concept d'interactivité ») ; panel « Disciplines et numérique », avec Éric Guichard (« Les humanités résistent-elles à l'histoire de l'écriture ? »), Frédéric Clavert (« Mise en données des sciences humaines et sociales : l'exemple de l'histoire ») et Évelyne Broudoux (« L'apport des sciences de l'information à l'étude des sources »).

<sup>7</sup> Les disciplines s'étant les premières intéressées aux matériaux numériques issus des productions et usages informatiques et en réseau, en particulier les sciences de l'information et la sociologie des usages, n'ont que récemment pris la mesure des problématiques temporelles posées par l'analyse des SNN (Ertzscheid, Gallezot et Simonnot, dans Barats, 2013 : 53-68 ; Chevalier et Herscovici, 2012).

<sup>8</sup> Le paléographe Marc Smith a donné une conférence intitulée « La véridique histoire de l'arobase », signe d'un intérêt récent pour les sciences auxiliaires de l'histoire (séance du 29 janvier 2013 du cycle de conférences « Du rare à l'unique », École des Chartes [<http://www.enc.sorbonne.fr/actualite/vie-de-l-ecole/conference-la-veridique-histoire-de-l-arobase-par-marc-smith>]).

<sup>9</sup> Les humanités numériques (ou *digital humanities*) sont une appellation qualifiant une série de pratiques savantes hétérogènes, en sciences humaines en priorité mais aussi en sciences sociales, qui non seulement systématisent l'utilisation d'outils informatiques dans leurs méthodes d'analyse (d'abord sur des matériaux documentaires numérisés), mais se développent aussi et surtout à la faveur des environnements numériques (en termes de documentation, communication, publication scientifique), tout en étant en général tournées vers l'ouverture et le partage des données de la science.

complexité et à l'instabilité documentaires des sources numériques ou numérisées, à en situer les cadres d'autorité, à retrouver la preuve dans un système technique où la notion d'original est bouleversée<sup>10</sup>. Cependant, là où les humanités et l'histoire numériques tendent à traiter des matériaux non nativement numériques, nous envisageons le numérique comme matériau de recherche à part entière, à la fois dans sa technicité et dans l'analyse des contenus qui sont produits dans son contexte socio-technique.

Dans cet article, en raison de la complexité des SNN, on prend d'abord le temps de les décrire à quatre niveaux de matérialité. Une deuxième partie entre dans le vif des propos tenus pendant la journée, et traite de la transformation numérique du support qui traditionnellement donne accès à la source, l'archive. Dans une troisième partie, on rend compte des jeux d'échelles d'observation et d'analyse de ces sources tels que dégagés par les intervenants. Enfin, on montrera comment ils ont pu mettre en lumière les apports réflexifs du travail sur les SNN par rapport à leur identité et leur épistémologie de chercheurs en SHS<sup>11</sup>.

## Typologie des sources numériques

Une typologie des SNN semble nécessaire en ce qu'elles sont non seulement complexes, mais surtout qu'elles peuvent être invisibles à l'œil de l'utilisateur profane d'un ordinateur. Celle que nous proposons repose sur un état de l'art des recherches en SHS ayant réfléchi, sur les plans empirique, théorique mais aussi méthodologique sur la matérialité de leurs objets d'étude dans le domaine de l'informatique et du numérique. On peut d'ores et déjà annoncer que ces sources, si elles sont inédites en termes techniques, prolongent cependant la désacralisation de la notion d'authenticité attribuée aux sources par les sciences auxiliaires de l'histoire, dans les rapports entre original, copie et faux (Canfora, 2012 ; Chabin, 2012). En contexte numérique, en effet, « la possibilité de manipulation est co-originale à la nature même du support numérique [...]. Ainsi, l'essence du numérique [...] est-elle : "ça a été manipulé" » (Bachimont, 2010 : 158).

### Les données numériques

Les données sont assimilées à de l'information brute qui a besoin d'être « traitée » pour devenir lisible. Cependant, il n'existe pas de données brutes, puisqu'elles sont prises immédiatement dans des logiques de catégorisation (« *raw data* » is an *oxymoron* » : Bowker, 2008 ; Gitelman, 2013). On s'intéresse ici à un niveau où les données ne relèvent plus des signaux électroniques, mais sont déjà interprétées dans des codes symboliques sous la forme de chiffres ou de lettres, produites par l'utilisateur ou générées automatiquement par les logiciels dont il fait l'usage. Les données d'usage laissées implicitement ou explicitement par les utilisateurs des machines numériques (par exemple, l'heure ou la localisation de la connexion à un réseau, ou encore l'adresse d'un site consulté) sont décrites comme des « traces numériques » (Ertscheid, 2013). On les assimile au même registre sémiotique que les empreintes et les signatures (Merzeau, 2009). Les métadonnées sont des données qui donnent

---

<sup>10</sup> Selon les réflexions en cours du groupe de travail « Cultures savantes numériques » du Labex HASTEC auquel a participé l'auteure.

<sup>11</sup> Je voudrais remercier Loïc Petitgirard et Samuel Hayat pour leur relecture attentive et leurs conseils bibliographiques.

des renseignements à l'ordinateur ou à l'humain sur des données ; par exemple, les métadonnées d'expédition, de réception, etc. qui accompagnent la circulation d'un message électronique.

Les données sont aussi précieuses dans le cadre de l'étude des bases de données, structure paradigmatique des usages numériques (Manovich, 2013), ou plus généralement forme organisationnelle du savoir dans ses aspects informationnels et anthropologiques (Beltrame et Jungen, 2013), sociologiques (Flichy et Parasio, 2013) et historiques (Grad et Bergin, 2009). Elles sont au cœur des récentes réflexions sur la *data* à différentes échelles (Ford, 2014) et sur la construction de corpus et d'archives (Habert et Huc, 2010).

### **Le code informatique**

Avant d'être des « sources » pour la recherche, les SNN sont composées de « sources informatiques ». En effet, on appelle « code source » la série d'instructions textuelles qui prévaut au traitement des données dans les programmes et les logiciels, de l'exécution d'une action sur les données (par des algorithmes) à l'affichage d'un contenu (par des instructions appelées « balisage ») sur les médias informatisés. Les études sur la communication et les médias ont pris la mesure de l'importance du code : on le compare à une loi qui rivalise avec le régime juridique légitime pour déterminer ce que l'on peut faire ou non avec les objets culturels du réseau (Lessig, 1999 et 2006). En France, la théorie sémiotique des médias informatisés utilise la notion d'« architexte » pour analyser le rôle du code qui conditionne les modalités techniques, sémiotiques et d'usage des textes à l'écran (Souchier, Jeanneret et Le Marec, 2003). Le code joue un rôle social, par analogie avec les codes sociaux d'un environnement culturel. L'exemple souvent cité aujourd'hui est celui de l'espace des opinions publiques redessiné par l'encodage des relations sociales sur les réseaux sociaux numériques : les visions du monde que cet espace offre sont déterminées, en partie, par les comportements informationnels des internautes encodés et calculés par les plateformes numériques en réseau. Les *software studies* et les *code studies* proposent une lecture critique du code et des programmes informatiques pour comprendre comment ils encadrent les comportements sociaux, culturels et politiques dans les environnements numériques (Manovich 2013 ; Paloque-Berges, 2009 ; Fuller, 2008 ; Marino, 2006 ; Cramer, 2001).

Dans cet article, on ne fera qu'aborder la question, pourtant vaste et importante, des nouvelles modalités de la preuve s'appuyant sur la modélisation et la prédiction numériques, qui laisse entrevoir un nouveau type non seulement de source, mais surtout d'épistémologie de la probation. Dans le cadre des nouvelles méthodes du *Big Data* appliquées aux données numériques, cette question se pose dans les SHS, mais aussi dans les sciences cognitives, naturelles, physiques et mathématiques – par exemple pour construire des modèles d'anticipation qui font preuve en regardant vers le futur, et non plus en s'appuyant sur des documents issus du passé (Varenne, 2001).

### **Les documents numériques**

En sciences de l'information, un document est « vu » (une forme et son support), « lu » (un contenu), et « su » (des normes qui accompagnent sa reconnaissance) (Salaün, 2012). Le document numérique déstabilise ces trois dimensions, dans la mesure où il ne bénéficie pas de la fixation (même relative) du support papier : ses formes et supports se modifient à la vitesse de l'évolution des techniques informatiques et de leur appropriation par les usages ; son contenu peut être

composé, augmenté, modifié par un ou plusieurs auteurs ; ses normes... sont loin d'être normalisées, bien que des standards existent en bibliothéconomie. Les théoriciens du document numérique parlent de « documentarisation » pour comprendre comment un document numérique passe par plusieurs cycles de vie documentaire tributaires des environnements logiciels et des usages associés, et de « redocumentarisation » pour qualifier le paradigme numérique du traitement documentaire, structuré, statistique, calculé sous le régime du code et des langages informatiques de programmation et de description (Pédaque, 2006, Salaün, 2007). On pourra retenir que le document numérique, celui qui se compose en réseau en particulier, est souvent dynamique, créé dans l'interaction, et composite<sup>12</sup>, une « archive en train de se constituer » (Marcocchia, 2001).

Sur le plan de la preuve juridique, il faut prendre en compte également les méthodes de l'« investigation numérique », qui visent à localiser les documents dans un système ou à comprendre comment celui-ci se comporte afin de reconstituer l'histoire (interne) d'un ordinateur<sup>13</sup>.

### Les artefacts matériels

Les artefacts matériels servant de supports à la conception et à l'usage des technologies numériques, les équipements et les instruments informatiques eux-mêmes participent largement des SNN. Deux approches principales existent : une approche épistémologique et historique descriptive et une approche matérialiste critique. La première est celle des historiens des sciences et des techniques se penchant sur l'histoire des ordinateurs et des environnements informatiques : elle s'intéresse aux mécanismes ou à la physique de la machine (Ramunni, 1989) ; elle met en relation la machine avec l'histoire des institutions, des entreprises et des industries (Mounier-Kuhn, 2010). La seconde est une extension des *software studies* au domaine des matériaux structurels des environnements informatiques, comme les réseaux, leurs câbles, protocoles, etc. (*infrastructure studies*) ; elle est investie par des interrogations issues de la sociologie de l'innovation (Musiani, 2015), des sciences de l'information (Blanchette, 2011), des sciences du judiciaire (« *forensic materialism* », Kirschenbaum, 2008), des *media studies* revues à la lumière de l'histoire et des études culturelles critiques appliquées aux technologies (*Media archeology*, Parikka et Huhtamo 2012).

Si cette typologie présente avant tout des postures théoriques, il s'agit dans ce qui suit d'être au plus près des préoccupations pratiques et méthodologiques des chercheurs se penchant sur les SNN : les possibilités et les limites de la reconnaissance des SNN dans des cadres d'autorité (en premier lieu des archives) ; l'exploration de méthodes prenant en compte la complexité des SNN en termes de matérialité et de temporalité ; enfin, la manière dont les SNN peuvent engager le chercheur à une réflexivité critique par rapport à ses propres pratiques et celles de son domaine, voire de son champ.

---

<sup>12</sup> Par exemple les documents que sont les forums, les pages des médias sociaux, les espaces de commentaires... tous les espaces en lignes que l'on qualifie de « contenu généré par les utilisateurs » (*user-generated content*).

<sup>13</sup> Voir la notice Wikipédia pour « *Digital Forensic Process* » qui donne une liste de références sur le sujet [[http://en.wikipedia.org/wiki/Digital\\_forensic\\_process](http://en.wikipedia.org/wiki/Digital_forensic_process)].

## Les SNN confiées aux archives

La formalisation des systèmes archivistiques pour les documents numériques nativement à des buts historiques et patrimoniaux est engagée, même si ces archives se présentent encore comme un « fatras », comme en témoigne Clément Oury, ancien responsable du dépôt légal du Web à la Bibliothèque nationale de France (BnF)<sup>14</sup>. Face à la quantité exponentielle des documents et des données numériques, ainsi qu'à la complexité des dispositifs informatiques qui les accueille, on peut se demander comment s'assurer que les sources que l'on étudie témoignent « fidèlement » d'une situation de discours ou de pratique. Comment les reconnaît-on et comment leur accorde-t-on notre confiance ?

### Autorité et crédibilité des sources en archivistique

L'écrit électronique est reconnu officiellement par le code civil depuis la loi du 13 mars 2000 au même titre que l'écrit sur papier ; il peut être authentifié et contrôlé par des algorithmes de signature. Cette reconnaissance juridique très formelle n'existe dans les faits que pour des documents administratifs particuliers (tels que les actes notariés) ; la reconnaissance de l'autorité attachée au document numérique n'est pas toujours si bien définie sur le plan technique ou symbolique. Marie-Anne Chabin rappelle que dans un contexte juridique ou historiographique, le document ne fait pas preuve par lui-même ; c'est le juge ou l'historien qui en décide grâce un ensemble d'éléments contextuels. Marie-Anne Chabin convoque la méthode de la diplomatique comme moyen de s'assurer de la crédibilité contextuelle d'un document source. Méthode inventée par Mabillon au XVII<sup>e</sup> siècle pour authentifier les diplômes mérovingiens de l'abbaye de Saint-Germain-des-Prés, elle sert à analyser la tradition, la forme, la genèse des actes écrits pour faire leur critique, juger de leur sincérité, de leur valeur, des éléments à exploiter par l'historien, et enfin les dater et les éditer. Elle porte en pratique sur l'identification de caractères externes et internes du document : les éléments visuels de validation (signature, cachet, sceau) et de transmission (enregistrement, tampon). Cette valeur de probation par l'identification contextuelle ne change pas avec la diplomatique numérique, sauf au niveau du type de traces permettant d'authentifier le document original. Dans le cas du numérique, on ne peut pas conserver l'original car le document ou les données sont distincts (mais pas indépendants) du support matériel : on conserve donc un agrégat composite d'éléments assurant le plus de stabilité possible à l'identification, comme par exemple les conditions symboliques ou techniques de mise en œuvre (les clauses et les dispositifs logiciels). Selon les normes internationales de l'archivistique, décontextualiser équivaut à désinformer. Dans le cas du numérique, comme l'original échappe, l'attention aux conditions de mise en œuvre doit procéder par reconstruction de conditions virtuelles et matérielles, dans une logique : on informe à nouveau. Il est donc accepté que l'on recrée entièrement une trace signifiante numérique (complète, précise, et intelligible). Le document fait preuve s'il peut être identifié par des données prises ensemble (de complétude, de fixation, de validation, de datation) et conservé grâce à d'autres jeux de données (métadonnées,

---

<sup>14</sup> Lors de la conférence « Mémoire numérique : une amnésie programmée ? » 16 mai 2013 [<http://www.arts-et-metiers.net/musee/memoire-numerique-une-amnesie-programmee>]. Voir le compte rendu sur le site de veille *Webarchivists* [<http://www.webarchivists.org/2013/05/memoire-numerique-une-amnesie-programmee-compte-rendu>].

horodatages/scellement, contrôle de l'intégrité). Pour le document numérique plus que jamais, conclut Marie-Anne Chabin, c'est le cycle de vie du document qui doit être documenté ; sa véritable nouveauté réside dans le fait que ce cycle traverse de très large ensembles d'environnements numériques, et que l'on peut le reconstituer grâce à des données de traçabilité (ce document a été enregistré à telle date, à telle heure, par tel auteur... pour prendre un exemple concret et quotidien de l'écriture numérique)<sup>15</sup>.

L'un des aspects pratiques de cette nécessaire recontextualisation est le travail de documentation mené par les nouveaux archivistes du numérique. Sophie Derrot rappelle que la loi française du dépôt légal du Web de 2006 assure aux archives une indépendance formelle par rapport à des supports technologiques spécifiques. Le contenu numérique vivant (publié en ligne, sur une page Web et consultable à travers un navigateur) est donc différent fonctionnellement et légalement de son pendant archivé, stocké sur les serveurs de l'institution qui le préserve, consultable sur des ordinateurs et à travers des logiciels dédiés. Depuis 2008, l'équipe du dépôt légal du Web à la BnF se porte ainsi garante (ou « tiers de confiance ») des archives du Web français grâce à un processus de conservation en trois temps détaillés par Sophie Derrot :

- la collecte des pages Web à archiver : opérée par un programme automatisé qui parcourt le Web français, un robot se comportant à la manière d'un internaute (il consulte et indexe les pages une à une) ;
- le figement d'un ensemble documentaire : dans une logique de reconstitution la plus fidèle possible sont enregistrés les contenus et les liens hypertexte de la page, son plan de navigation, tout ceci à un instant précis ;
- l'accès aux pages archivées : il se fait sur place à la BnF, pour éviter la décontextualisation liée à l'usage de logiciels différents pour consulter les données (les contenus pouvant prendre différentes formes selon le logiciel de navigation avec lequel on les consulte) ; mais il est à noter qu'il est difficile de faire en sorte que les pages s'interfacent toujours sans problème. C'est d'ailleurs le même problème que rencontre la collection gigantesque en ligne d'archives Web de la fondation « Internet Archive »<sup>16</sup>.

Ces préconisations d'archivistique numérique par les professionnels représentés ici par Chabin et Derrot sont de leur propre aveu des idéaux qui rencontrent bien des obstacles face à des collections de documents et artefacts hétérogènes et évolutifs, et donc difficiles à documenter et à classer.

### **Les archives numériques au prisme de l'autoritativité**

Le développement et l'accessibilité exponentiels des systèmes de gestion documentaire informatisée mettent entre les mains des usagers la capacité à créer et à maintenir des archives de la matière numérique. Il s'agit là d'une capacité « autoritative<sup>17</sup> », c'est-à-dire le fait que les technologies de publication et

---

<sup>15</sup> On retrouve les propositions diplomatiques de Marie-Anne Chabin dans son article « Archives » (2013-2014).

<sup>16</sup> Cette collection est accessible à travers un moteur de recherche, la « *wayback Machine* » (« machine à remonter le temps ») qui à partir de l'URL d'un site Web (vivant ou défunt) affiche les versions archivées par la fondation sur ses serveurs [<http://archive.org/web/>].

<sup>17</sup> L'autoritativité est un concept proposé par Évelyne Broudoux pour définir la production et la publication de texte dans le cadre de l'auto-édition sur le Web « sans passer par l'assentiment d'institutions de référence référées à l'ordre imprimé » (Broudoux, 2003).

d'organisation des documents et des données numériques sont, dans une certaine mesure, à la portée de tous, y compris dans le choix des formes et des modalités de publication éditoriales. La prise en compte des « auteurs » est donc primordiale pour toute analyse et ingénierie d'archives de données et de documents nativement numériques. La publication Web repose sur des manipulations techniques associées à des compétences, et sur des normes légales relatives à la fois au droit d'auteur et au droit des données. Ces dimensions sont inscrites dans la matérialité des sources et doivent être présentes à l'esprit de celui qui les analyse. Si l'institution a ainsi une responsabilité face aux archives et aux auteurs, producteurs et utilisateurs des documents et données qu'elles abritent, la responsabilité des services du Web et de leurs usagers est également engagée.

Sophie Derrot attire l'attention sur la dimension pédagogique d'une structure institutionnelle comme le dépôt légal du Web à la BNF, c'est-à-dire sur la manière dont les auteurs initiaux des pages Web sont informés de la pratique d'archivage. La collecte n'est opérée que sur des pages Web publiques, c'est-à-dire accessibles à tous les usagers du Web sans verrou (sans mot de passe par exemple) ; la sphère privée est hors d'atteinte<sup>18</sup>. Un formulaire de demande de collecte des sites Web français est mis à disposition, permettant aussi de requérir que la page ne soit pas collectée. Quand la collecte se fait de manière automatique<sup>19</sup>, le robot laisse des traces de son passage dans les « logs » informatiques (les comptes rendus d'activité logicielle), décline son identité et renvoie à la documentation associée, afin que l'auteur des pages soit mis au courant. Pour prolonger le propos de Sophie Derrot, il faut préciser que la démarche devient compliquée dès que l'auteur n'est pas celui qui techniquement publie ou héberge le contenu, comme c'est souvent le cas au sein de l'environnement des services Web<sup>20</sup> ; elle se complique d'autant plus quand on considère la fragmentation des contributions quotidiennes sur les réseaux numériques (courriels, commentaires, etc.). La responsabilité face aux auteurs des pages collectées est un des grands problèmes posés par des initiatives d'archivage du Web ; celle d'Internetarchive.org, pionnière en la matière depuis 1996, est par exemple critiquée car elle ne propose pas d'encadrement légal ou pédagogique.

L'archivage des données et documents nativement numériques n'est pas l'apanage des institutions : de nombreuses collections d'archives courantes et intermédiaires sont rassemblées et maintenues sur le Web ou d'autres réseaux d'Internet<sup>21</sup> par des individus ou des collectifs. Emmanuel Ruzé insiste sur l'importance de ces archives en contexte vivant, en particulier celles générées et maintenues par les communautés en ligne (par exemples les archives de Wikipédia, ou celles des utilisateurs d'outils de blogging comme Wordpress). Elles deviennent pour lui des sources permettant de

---

<sup>18</sup> La question de la publicité et la privauté des données de la personne sur Internet ne se réduit pas à cette différence technique marquée par le verrouillage de la page Web.

<sup>19</sup> Des robots logiciels naviguent sur les pages pour les indexer, selon la technique du *crawling*.

<sup>20</sup> Par exemple les plateformes de contenus gérés par les utilisateurs (expression dérivée de *user generated content sites*), comme Youtube ou Tumblr, ou les réseaux sociaux comme Facebook ou Twitter.

<sup>21</sup> Une technologie historique pour la transmission de documents par Internet est le protocole FTP (*File Transfer Protocol*), qui permet d'envoyer des fichiers à partir de ses archives informatiques personnelles. Beaucoup des premières collections de documents numériques ont été constituées pour le FTP, stimulant l'échange documentaire sur Internet et devenant l'équivalent d'une bibliothèque numérique en réseau. Les technologies du pair à pair (*bit torrent* par exemple) et celles, plus récentes du *cloud* (serveurs virtuels) en sont les descendantes, et peuvent également donner accès à des archives vernaculaires de documents numériques par les utilisateurs



retracer l'histoire d'un collectif et de ses usages socio-techniques des réseaux, grâce à leur contenu, mais aussi grâce à leur usage par les collectifs eux-mêmes. Emmanuel Ruzé observe dans les archives communautaires en ligne la manière dont est inscrit le savoir dans le temps et dont émergent des routines d'organisation (structuration de contenu, prises de décision). Il en conclut que ces archives viabilisent le collectif, lui donnent une identité grâce à l'accès à la mémoire des connaissances et des activités qu'elle porte et reproduit. Elles lui renvoient une image de lui-même, sont un support de mémoire didactique et culturelle car elles permettent aux nouveaux entrants de prendre connaissance des manières de faire et de s'acclimater aux us et coutumes de la communauté (comme par exemple les archives de listes de discussion, le plus souvent ouvertes aux membres de la liste) et de faire des choix pour son futur (Ruzé, 2011). On retrouve cette réflexion chez Madeleine Pastinelli (2009), qui montre que la gestion des archives des conversations instantanées (ou « tchat ») d'une communauté donne la possibilité à celle-ci de réguler symboliquement ses propres usages (l'archive étant une preuve permettant de résoudre des conflits éventuels). Les travaux de Ruzé et Pastinelli signalent aussi les problèmes d'intégrité des archives communautaires numériques. Si certains systèmes d'information en ligne, comme sur Wikipédia, préservent automatiquement chaque version et chaque trace de modification des pages archivées, une communauté peut également modifier ses archives pour des raisons diverses (de la manipulation à la perte). Les sources issues de ces archives communautaires peuvent être pensées comme une « mémoire vivante de la communauté, de ses choix et de ses erreurs » (Ruzé, 2011 : 6). L'analyse des SNN provenant de fonds d'archives communautaires doit donc impliquer une réflexion sur la manière dont leur sélection, leur gestion et leur transmission engagent le rapport du collectif à lui-même, à ses savoirs et pratiques, mais aussi au regard des publics qui peuvent être amenés à les consulter (dans ou hors des communautés).

### **Archive (figement) ou *archai* (commencement) ?**

L'archivage de matières numériques est paradoxal car, comme le suggère Louise Merzeau<sup>22</sup>, il porte sur la fixation de flux d'informations. Comment faire confiance à un document qui ne cesse d'être en mouvement ? C'est un paradoxe pour le monde des archives, qui est celui du figement (selon son étymologie, rappelée par Marie-Anne Chabin), de l'institutionnalisation synchronique et diachronique (comme le souligne Éric Guichard). La source numérique native nous pousse à repenser le rapport entre archives institutionnelles (logique monumentale, archives figées) et archives courantes (logique communautaire, archives vivantes).

Éric Guichard propose de revoir le sens de l'archive numérique dans une logique non plus de figement mais de processus – logique corrélée selon lui à l'écriture historique à l'ère numérique. Il se demande : à qui donne-t-on la parole, (et donc) la possibilité de l'archiver ? Cette problématique est marquée par le sens foucaldien de l'archive : l'*archai* (commencement) pose la question des systèmes d'énonciation aux fondements des régimes de vérité adoptés par une époque. Éric Guichard soutient qu'il est vain de parler d'archive numérique, et lui préfère la notion de systèmes d'écriture en transformation (binaire, documentaire, réticulé, automatisé) au sein desquels les SNN émergent. Ces dernières gardant la nature du mille-feuille dynamique qui les constituent depuis le matériel jusqu'au logiciel, en passant par les

<sup>22</sup> Louise Merzeau développe cette question dans son mémoire d'HDR (2011).

formules algorithmiques, les bases de données, les systèmes d'information qui les interprètent et les rendent lisibles. La culture numérique, pour Éric Guichard, est un ensemble de processus d'écriture non stabilisés<sup>23</sup>.

Ceci n'est pas sans poser de grands problèmes pratiques, par exemple dans l'archivage du Web, comme le rappelle Sophie Derrot. Paradoxe vivant, l'archive Web telle qu'elle est constituée dans le dépôt légal, c'est-à-dire une expression de l'institution, doit être préservée par des standards<sup>24</sup> qui ne sont ni respectés par les auteurs et producteurs de contenu sur le Web, ni stabilisés formellement par les professionnels. La préservation devient complexe quand on ne peut pas vraiment faire confiance à ce que déclarent les sites Web sur leurs propres contenus (à travers les métadonnées qui les décrivent), ou à ce que les normes en vigueur promettent. Sophie Derrot, insistant sur le fait qu'« il faut tout revérifier constamment », suggère que l'archiviste entre dans une logique de flux et de recommencement. À ce titre, son équipe à la BnF tente de proposer des normes stables au niveau de la communauté professionnelle des archivistes du numérique (Derrot, Fauduet, Oury, et Peyrard, 2012).

Les SNN sont ainsi au croisement de deux définitions de l'archive formulées par Louise Merzeau : la prise en charge d'un document qui témoigne des indices d'une activité et qui est catégorisé dans un ensemble documentaire (définition classique de l'archive), et d'un matériau qui est un média en soi et qui est donc éditorialisé dans un ensemble médiatique (définition nouvelle de l'archive numérique).

## **Méthodologies de fouille de la mémoire vivante du numérique**

On vient de voir que la confiance accordée aux SNN est tributaire de dispositifs archivistiques. Cependant, ces dispositifs ne sont pas seulement le fait d'institutions de confiance : l'ordinateur produit sans cesse de l'archive courante et intermédiaire, grâce à des systèmes de gestion documentaire (*records management systems*) intégrés. Ils conservent, organisent et permettent de consulter les documents et les données pour des raisons avant tout fonctionnelles ; ils établissent des versions documentaires successives, qui font partie de l'histoire de la SNN. Pour Bruno Bachimont, ce fonctionnement de la « mémoire numérique » (au sens informatique du terme) décline trois processus documentaires : transformation, inscription et organisation (Bachimont, 2004). Pour Marie-Anne Paveau (2013), on doit aussi parler de « mémoire technodiscursive » pour mieux qualifier les processus temporel de traitement de l'information en (au moins) deux temps : les informations visibles à l'interface homme-machine ; les informations invisibles des programmes qui traitent les informations.

En ceci, les méthodologies déployées par les chercheurs étudiant des SNN doivent prendre en compte ces jeux d'échelle en termes d'espace et de temps. La mémoire des SNN est feuilletée, et le regard du chercheur doit ajuster sa focale, selon l'objet de recherche qu'il construit et la matérialité des sources qu'il utilise pour l'analyse.

---

<sup>23</sup> Voir aussi Éric Guichard, 2011.

<sup>24</sup> Les formats déclarés à l'IANA, l'organisme qui gère notamment les ressources requises pour la communication sur Internet.

## Rendre visibles les systèmes d'information

La SNN doit donc être pensée dans le contexte du système d'information où elle a été produite ; le chercheur est ainsi invité à effectuer un travail de fouille des systèmes documentaires, voire des données numériques<sup>25</sup>. Évelyne Broudoux rappelle que les usages du numérique se font dans le cadre de « systèmes tournés vers les utilisateurs » (Morville et Rosenfeld, 2006) et qu'il faut ainsi prêter attention à l'architecture de l'information qui soutient et conditionne la production (Broudoux, Chartron et Chaudiron, 2013). En ceci, une source (les données qu'elle contient) n'est pas séparée de ses métadonnées, des informations sur les informations qui sont des normes permettant de trouver leur place et d'être retrouvées dans un système de classification documentaire. Ces métadonnées sont générées dans la production du document, ou au fil de ses « redocumentarisations » successives : à chaque fois que le document passe par une étape documentaire, de nouveaux codes remplacent, modifient ou s'ajoutent aux codes précédents (Pédauque, 2006). Celles-ci sont cruciales : elles permettent de décrire le droit, les relations, les formats, les dates... et donc d'organiser la gestion et l'insertion du document dans un écosystème documentaire numérique. Évelyne Broudoux les décrit comme les chevilles ouvrières des systèmes de gestion de contenu. Elles peuvent dire beaucoup, mais aussi taire beaucoup et introduire des biais selon les normes qu'elles suivent.

Dans un système d'information, la lisibilité (le « lu » de Salaün) est conditionnée par les données et les métadonnées. Emmanuel Ruzé explique comment l'analyse des sources numériques natives peut se faire à l'échelle du système d'information lui-même, au niveau du conditionnement architectural de la lisibilité, brisant ainsi l'unité historiographique du document qui caractérise généralement la source. Il propose une démarche micro-historique, centrée sur la manière dont le « lu » est construit dans le temps par les utilisateurs. Cette approche implique un regard qualitatif aussi bien que quantitatif<sup>26</sup>. Dans le premier cas, en effet, il s'agit d'accompagner le codage thématique traditionnel des sources par une contextualisation historique, sociale et culturelle des usagers impliqués dans la conception et la gestion du système (et donc leurs mécanismes socio-techniques de contribution, d'interaction, d'appropriation). Cela est particulièrement important dans l'étude de cas de collectifs dont l'identité est construite dans la relation aux outils, comme les communautés réunies autour de la pratique de l'*open source*<sup>27</sup>. Dans le deuxième cas, on se penche sur les traces laissées par les utilisateurs au cours de leur usage, auxquelles on peut appliquer un travail de calcul et de décompte statistique (économétrie et analyse de réseau) pour observer les tendances dans le temps (évolution des populations d'utilisateurs et fréquence d'utilisation, réseaux de sociabilité...). On reviendra plus bas sur cette approche quantitative, après avoir proposé quelques exemples d'analyses qualitatives des pratiques numériques.

---

<sup>25</sup> Le terme « fouille » est ainsi approprié dans les nouvelles méthodes de traitement des données (fouille de données ou *data mining*).

<sup>26</sup> Le rapprochement entre les propositions de Ruzé et le concept de « lu » chez Salaün est opéré par l'auteur de l'article.

<sup>27</sup> L'*open source* caractérise une économie de codes informatiques ouverts à l'appropriation et à la modification, différents des programmes logiciels dits propriétaires.

## Rendre visibles les interactions par les pratiques et les structures

Une analyse des pratiques numériques peut s'intéresser aux « coulisses » des interactions en contexte de communication médiée par l'informatique, par exemple en utilisant les concepts de l'analyse goffmanienne. Les *infrastructure studies*, dans le domaine des études de science et technologie (*science and technology studies*) observent ainsi ce qui se trame en matière de codage informatique, de programmation des systèmes, de régulation des flux de réseau, de protocoles et de standards. Se penchant sur des sources qui sont des documents de travail et des sources au sens informatique du terme (le « code source »), ces études s'accompagnent d'une analyse ethnographique des acteurs pris dans ce travail des « ouvriers invisibles », comme l'explique Francesca Musiani. Elle souligne que le « travail du code » implique non seulement des choix professionnels, mais aussi politiques qui sont intégrés dans la fabrique des infrastructures – avec des conséquences sur les usages. L'originalité et la difficulté d'approche des sources numériques natives résultant de ce travail résident dans trois aspects majeurs : leur production collaborative ; leur circulation à l'échelle du réseau, et donc l'existence de versions différentes, intermédiaires, augmentées, appropriées, modifiées... ; leur interprétation liée au matériel (*hardware*) dans lequel les codages et spécificités techniques auxquelles elles donnent accès sont implémentés.

Deux exemples illustrent la manière dont cela peut s'étudier à plusieurs niveaux, et à partir de sources différentes.

Tout d'abord, comme en témoignent Alexandre Hocquet et Camille Paloque-Berges, certains documents produits collectivement sur les réseaux numériques sont particulièrement aptes à dévoiler ce travail de coulisses. Il s'agit des correspondances des acteurs des champs scientifiques et techniques par le biais des outils du numérique : emails, listes de diffusion, groupes de discussion et autres moyens de communication par les réseaux informatiques. Ces communications sont gérées par des systèmes qui créent automatiquement des archives courantes, et de nombreux gisements existent, conservés par les administrateurs des systèmes ou par les coordinateurs des projets. À travers l'analyse de discours des contenus des conversations, mais aussi des interactions entre acteurs et des relations entretenues entre les messages échangés, ils ont observé le travail technique en train de se faire. Les messages numériques entre acteurs n'abritent pas seulement le contenu d'une discussion médiée par les ordinateurs en réseau. Camille Paloque-Berges montre que l'on trouve, dans les métadonnées qui accompagnent le corps des messages, des informations techniques précieuses pouvant intéresser l'analyse : routes numériques prises par les messages (permettant de retracer le chemin du message dans un réseau de réseaux), précision sur les lieux, dates et personnes impliquées dans la transmission du message, formats des messages au fil de ses « redocumentarisations », autant de signes de l'activité qui peuvent être lus à travers une perspective socio-technique. Alexandre Hocquet souligne que l'intérêt des messages électroniques, notamment grâce aux métadonnées, réside dans le fait qu'ils peuvent être codés facilement avec des outils d'analyse de réseau, afin d'étudier par exemple les controverses entre acteurs<sup>28</sup>. Ces objets ont ainsi plusieurs facettes que l'on peut décider d'étudier conjointement à travers le prisme analytique socio-technique : données, contenus, significations, pratiques, organisation...

---

<sup>28</sup> Voir aussi Paloque-Berges et Kembellec, 2014 et Hocquet, 2013.

Dans le domaine de l'étude des usages, l'analyse de l'interaction homme-machine, c'est-à-dire la relation dynamique entretenue par les utilisateurs aux artefacts informatiques à travers leurs interfaces, peut s'appuyer sur l'observation *in situ* ; mais qu'en est-il quand l'étude porte sur des interactions passées ? Benjamin Thierry, dans ses travaux sur l'évolution des interfaces utilisateurs<sup>29</sup>, compare cette situation d'observation impossible à un village déserté à l'arrivée des ethnographes. Le recours à des sources textuelles et iconographiques (par exemple les photographies et schémas publiés dans les magazines d'informatique ou les instructions des manuels d'utilisation) ne donne qu'une vision figée des interfaces. L'une des solutions à ce problème méthodologique passe par la reconstitution des parcours d'utilisateurs grâce à l'émulation, à savoir la possibilité d'exécuter des logiciels anciens et obsolètes dans des environnements informatiques récents<sup>30</sup>. L'émulation rend visible, de manière plus approfondie et plus dynamique que ce qu'en dirait un manuel d'utilisation, la façon dont les systèmes informatiques du passé structurent leur interface et conditionnent les usages ; en exécutant à nouveau le programme, elle produit de l'information sur l'interaction et l'utilisateur qui y est engagé. La labilité de la source numérique native devient palpable, car l'émulation produit une nouvelle catégorie de source au second degré, rendant caduque la possibilité de considérer l'archive numérique comme un document original.

Valérie Schafer souligne l'importance, pour des analyses qui peuvent porter sur des niveaux micro des pratiques, de leur mise en perspective dans une approche globale des phénomènes numériques, en convoquant une analyse des stratégies économiques et politiques et des pratiques sociales – comme le font l'histoire de l'innovation et les *infrastructure studies* par exemple, en particulier en ce qui concerne l'histoire des réseaux informatiques (Schafer, 2012).

### **Rendre visibles les relations par les données**

Les humanités numériques ont pris acte du changement d'échelle impliqué par le numérique, au-delà des SNN qui ne constituent que très partiellement leur objet : face à la quantité exponentielle de données auxquelles donnent accès les sources numériques, rappelle Louise Merzeau, l'analyste doit réajuster ses méthodes d'interprétation. Les masses de données permettent de faire émerger des hypothèses *a posteriori*, selon la manière dont on organise leur observation : selon les variations dans le temps (approche diachronique), selon les relations cartographiées dans l'espace (approche synchronique), ou encore dans des agrégats de données hétérogènes rassemblées en corpus (approche comparative). Le rôle de l'archiviste s'en trouve accru et évolue dans la mesure où il participe à l'agrégation des données qui peuvent être éparpillées dans le temps et dans l'espace des environnements logiciels.

Frédéric Clavert précise la manière dont les humanités numériques prennent en charge la source numérique native : à travers la « mise en données » des sources, elles transforment des phénomènes en objets quantifiables. Cela consiste à croiser des données pour faire ressortir des éléments non visibles à la lecture humaine. L'analyse des données est dépendante des outils informatiques pour exécuter ces analyses selon différents types de calcul :

---

<sup>29</sup> Voir aussi Thierry, 2008 et 2012.

<sup>30</sup> On parle aussi de « virtualisation », car les machines d'origines sont reconstituées virtuellement sur des machines récentes.

- calcul temporel : avec par exemple l'application NGram qui fouille les données du corpus d'ouvrages numérisés de Google (Google Books) pour révéler quand des termes émergent et leur fréquence d'utilisation dans l'histoire des textes ;
- calcul spatial : avec par exemple les technologies de géolocalisation comme les GPS, qui produisent des données en même temps qu'elles en cartographient ;
- calcul de relations : par exemple les méthodes d'analyse de réseau appliquées aux données des réseaux socio-numériques.

Les données remplacent-elles les documents, jusqu'alors prééminents dans le domaine d'investigation des sources ? La question n'est pas nouvelle, précise Clavert, et se pose depuis les débuts de l'histoire sociale<sup>31</sup>. Plutôt qu'une innovation radicale, les humanités numériques proposeraient une autre échelle d'observation et d'analyse des sources – Frédéric Clavert prenant comme exemple la méthode de la lecture distante proposée par Franco Moretti. Plus encore, continue Frédéric Clavert, les nouvelles méthodes dites du *big data*, à savoir l'analyse corrélative de très grands ensembles de données numériques, trouvent sur les réseaux socionumériques des sources abondantes pour mettre au jour les positionnements d'acteurs dans un champ informationnel ou communicationnel, à travers l'étude des relations entre individus et groupes et de leurs agrégats relationnels. On parle notamment de « science du Web » : fondée sur l'analyse de réseau (héritée des méthodes statistiques et mathématiques de la sociométrie et utilisée prioritairement en sociologie ou en économie), elle applique ses modèles aux données du Web, en particulier les liens entre unités d'information (page Web, contenus multimédia gérés par les applications, profils utilisateurs...). Face aux critiques formulées à l'encontre des méthodes quantitativistes (qui ne permettraient pas de vérifier les hypothèses car elles ne rentrent pas dans l'analyse des contenus), Frédéric Clavert rappelle qu'en effet elles ne font que permettre d'observer des tendances et doivent, pour entrer dans la finesse des usages, être complétées par le regard qualitatif.

Enfin, les *infrastructure studies*, comme le rappelle Francesca Musiani, sont particulièrement inclusives : elles révèlent des formes de sociabilité, des valeurs inscrites dans la programmation et l'usage des technologies. Elles combinent des approches qualitatives avec des analyses quantitatives pour mieux cerner le déroulement des choix politiques, éthiques et sociaux dans des dimensions chronologiques et multimodales (grâce aux méthodes de l'analyse de réseaux sociaux aussi bien que d'analyse de données et de discours).

## Réflexivité des pratiques de chercheurs du numérique

Un dernier grand problème posé par les sources numériques natives réside dans la relation étroite entre objets et outils de recherche faits de la même matière numérique. Sur le plan logiciel, par exemple, un système analytique de visualisation des sources développé pour la recherche en SHS peut ainsi n'être que le dernier d'une série de systèmes d'information par lesquels sont passés documents et données numériques. Chacun des systèmes les rend visibles de manière différente, voire les modifie ; cela peut poser problème si l'on prélève les sources dans l'un en particulier et non à la racine d'un ensemble d'origine souvent inaccessible. Il faut donc

<sup>31</sup> Comme en témoigne par exemple l'article de François Furet et Adeline Daumard sur les « Méthodes de l'histoire sociale : les Archives notariales et la Mécanographie » paru en 1959 dans *Les Annales*.

constamment ajuster ses cadres et niveaux d'interprétation. Le chercheur est un usager de l'information numérique lui-même, impliqué dans l'utilisation d'outils mnémotechniques impliqués par la production de tout document numérique (Le Deuff, 2010). Si les méthodes d'analyse des SNN montrent leur pouvoir heuristique à plusieurs échelles, elles mettent au jour également le rapport épistémologique réflexif que le chercheur entretient à sa propre pratique.

### **De l'usage informatique à l'usage numérique : changement de paradigme ?**

Malgré le succès récent et grandissant des humanités numériques, qui introduisent selon Louise Merzeau un regard plus réflexif des outils informatiques dans l'usage par les chercheurs, les SHS ont déployé des méthodes de codage des données et d'analyse des sources par l'informatique depuis le « tournant computationnel » (Barberousse et Imbert, 2011), ou du moins depuis que les ordinateurs leurs sont accessibles. Évelyne Broudoux rappelle ainsi que les littéraires et les linguistes sont parmi les premiers à en faire usage pour l'analyse de contenu, les outils de bibliométrie, de lexicométrie, dans un mouvement d'humanités numériques avant l'heure<sup>32</sup>. Cependant, cet usage d'outils informatiques restait sommaire et d'accès limité avant la généralisation des équipements numériques dans les espaces de travail des chercheurs, et ce changement d'intensité d'usage s'explique aussi par l'émergence des humanités numériques dans les années 2000<sup>33</sup>.

Dans cette lignée, les nouvelles méthodes d'analyse de grands ensembles de données (autour du *big data*) seraient au fondement d'une nouvelle révolution paradigmatique, en modifiant les relations entre théorie, hypothèses et données. La tentation est grande, face à la richesse des sources numériques natives, de postuler que certaines méthodes comme celles du *big data* permettent d'atteindre un niveau de connaissance de l'expérience humaine englobant toutes les échelles. À ce titre, Frédéric Clavert donne l'exemple du projet *World Historical Dataverse*, une agrégation de bases de données qui sous-tend la création d'une « archive » historique unifiée (qui s'apparente en fait à une banque de données). Il précise qu'il reste cependant des ambiguïtés non résolues, par exemple le développement des théories à des échelles globales sur plusieurs siècles, sans savoir quoi anticiper sur les connaissances futures. Selon les aficionados du *big data*, la massivité des données numériques permettrait de faire remonter des hypothèses *a posteriori*, sans problématique au préalable. Les grands ensembles de données ne peuvent être pré-analysés, appréhendés dans leur exhaustivité, et ceci par définition. Difficile, donc, de proposer des hypothèses sur des matériaux qui ne se laissent plus circonscrire dans des corpus. Ce serait aux outils de fouille de données de prendre le relais : leur analyse permettrait de faire remonter des corrélations impossibles à voir à l'œil nu, et d'emblée prouvées par l'analyse quantitative et/ou relationnelle. Ce « progrès » dans l'instrumentation de l'analyse, largement inspiré par les pratiques des sciences expérimentales, est critiqué : il ne ferait que souscrire une fois de plus au mythe positiviste de l'objectivité des données (Boyd et Crawford, 2012). C'est un débat

---

<sup>32</sup> Parmi les pionniers listés par Broudoux, on trouve des collectifs de chercheurs tels que : Association for Literary and Linguistic Computing, 1978, Association for Computers and the Humanities, 1973, Société canadienne des humanités numériques, 1986, Consortium pour ordinateurs en sciences humaines, 1986.

<sup>33</sup> Voir Franziska Heimbürger et Émilien Ruiz, « Faire de l'histoire à l'ère numérique : retours d'expériences », dans Delalande et Vincent, 2011 : 71.

récurrent en SHS, à ceci près que la numérisation exponentielle d'une grande quantité de la vie sociale fait espérer que les ambitions universalistes des démarches positivistes trouveront dans les environnements numériques de quoi s'incarner de manière non spéculative.

Cette réflexivité instrumentale et épistémologique n'est pas déparée d'une réflexion sur ce que signifie le numérique dans l'espace cognitif du chercheur, lui-même usager. Éric Guichard avance que le numérique d'une part introduit des ruptures dans les formations discursives, et d'autre part fait le lien entre le terrain, le concept, l'expérience et la théorie. C'est en termes épistémologiques que le numérique permet de repenser les techniques intellectuelles, d'où l'idée selon laquelle la technologie numérique est une écriture. Éric Guichard propose une approche anthropologique : les chercheurs doivent d'abord être des empiristes des nouvelles écritures du numérique (programmes, systèmes et interfaces) pour mieux faire émerger les questions épistémologiques nouvelles. En effet, la question de l'écriture est sous-jacente à la notion même de source, comme on l'a évoqué plus haut : elle pose la question du rapport du contenu au code, du discours au document, des données au support, des langages naturels aux langages formels. Le déchiffrement des écritures est commun aux historiens de l'Antiquité et aux analystes de l'informatique – et ce même au-delà de l'alliance de l'ingénieur et du savant prônée par les humanités numériques. L'écriture est donc pour Éric Guichard la clef pour passer de l'analyse des techniques à la réflexion épistémologique.

### **Un miroir des nouvelles médiations et politiques de la recherche**

Il faut enfin considérer comment l'objet « nativement numérique » reflète l'évolution des pratiques des chercheurs eux-mêmes dans un environnement où la documentation et l'écriture tendent à être de plus en plus médiés par le numérique.

Alors que l'historiographie a intégré des formes diverses de documentation intermédiaire, notamment au plan de l'écriture scientifique (Brian, 2001, Lefebvre, 2013), toute trace laissée numériquement par le chercheur, si elle est préservée, devient un réservoir de sources potentiellement analysables. Cela est inscrit, en quelque sorte, dans l'ADN des technologies informatiques, et particulièrement des réseaux de communication distribués. Camille Paloque-Berges rappelle que dans l'histoire d'Internet, les premières archives de documents nativement numériques concernent les documents et les conversations conservées des chercheurs et des ingénieurs de la communauté scientifique, première conceptrice et usagère d'Internet. Les doubles mouvements *science studies* et *science, technology and society studies* embrassent ainsi dans leur démarche d'analyse à la fois des communautés scientifiques et de leurs pratiques techniques (en l'occurrence, numériques) afin d'établir la sociologie ou l'histoire matérielle, sociale et intellectuelle des domaines scientifiques et techniques. Dans cette lignée, la sociologie de l'innovation réinterprète la notion de médiation à l'aune des médias informatisés pour qualifier les types de relations entre acteurs dans les réseaux socio-techniques (Akrich, Méadel, et Paravel, 2000). La médiation par les réseaux informatique a précocement compté dans la manière dont la communication scientifique a pu évoluer ces dernières décennies, du plus informel (des emails entre membres d'un projet scientifique collectifs) au plus formel (évaluation, publication et réputation scientifiques). Pour l'historien, le sociologue, l'anthropologue des sciences et des techniques, ce sont des trésors à transformer en sources et corpus ; pour tout autre chercheur, réfléchir à sa manière d'écrire et de communiquer à partir des traces



qu'il enregistre ou laisse derrière lui devient crucial. Alexandre Hocquet et Camille Paloque-Berges, quand ils présentent leurs corpus d'emails issus des échanges électroniques collectifs au sein de communautés scientifiques, soulignent qu'ils y ont trouvé matière à analyser la science et la technologie en train de se faire, les controverses qui leur sont liées, ainsi que les méthodes d'organisation et de résolution des problèmes des acteurs de la recherche. Francesca Musiani explique que les *infrastructure studies*, dans une perspective d'épistémologie sociale, s'attachent à mettre en lumière « les travailleurs du savoir » (Vann et Bowker, 2001) des communautés scientifiques et techniques collaborant sur les réseaux. Ces différentes approches ont en commun de ne pas séparer le social d'un côté et l'infrastructure de l'autre : les sciences sociales, insiste Francesca Musiani, doivent apprendre à naviguer entre les couches des codifications numériques pour mieux comprendre leur propre travail.

Si les SNN sont des matériaux pour faire l'étude des sciences, on peut également regarder comment l'usage des SNN est également l'occasion d'une réflexivité du chercheur sur les institutions qui accueillent et conditionnent sa pratique (Le Marec, 2010), et sur la manière dont ses choix peuvent être le moteur d'une évolution institutionnelle des disciplines. À l'international, les premiers travaux sur les SNN prennent racine dans le mouvement de la « *cyberanthropology* », qui se déploie dans les années 1990 à partir de l'exploration des environnements numériques en réseau. C'est en développant la critique de ce mouvement pionnier que se construisent les *Internet studies*, qui consolide les pratiques et les méthodes d'analyse des SNN (Wellman, 2001). Leur pendant français se trouve en sciences de l'information et de la communication et de la sociologie qui se rencontrent sur l'étude des techniques et des médias (Barats, 2013). C'est la communauté de linguistes qui a la première commencé une série de recherches sur la transformation de données de langue prélevées directement dans les contextes des médias numériques en archives formalisées et exploitables par la communauté scientifique, ceci étant explicable par la proximité relative, en termes de méthodes, avec les sciences informatiques<sup>34</sup>. Une avancée décisive est le fruit du rapprochement d'équipes de diverses institutions, depuis la collaboration informelle à l'infrastructure de recherche, pour la mise en commun d'outils et de méthodes, l'interopérabilité des collections et l'internationalisation de la recherche en SHS, qui fonde le projet des humanités numériques dans les années 2000 (Berra, 2012).

Le métier du chercheur serait amené à évoluer face aux SNN. Les archives Web sont récemment devenues aussi un objet de réflexion et de pratique à part entière en termes d'ingénierie de la recherche, pour lesquelles on développe des logiciels pertinents dans le cadre de dispositifs de consultation et d'aide à l'analyse, comme en témoigne une série de lieux français où travaillent ensemble des chercheurs en SHS, des ingénieurs en informatique, des documentalistes – par exemple au sein de l'équipe du dépôt légal du Web de la BnF ou son pendant à l'INA<sup>35</sup>. On l'a vu, les garants des archives officielles du Web français admettent volontiers que les cadres

---

<sup>34</sup> Par exemple à travers le groupe CoMeRe (Communication Média par Ordinateur) du consortium Corpus Écrits (labellisé par la TGIR HumaNum en 2011-2015).

<sup>35</sup> On citera également, parmi les structures françaises dédiées à ces activités collaboratives et interdisciplinaires de cet ordre, l'Institut de Recherche et d'Innovation (IRI) du Centre Pompidou, ou encore le Media Lab de Sciences Po Paris, mais on retrouve ces logiques dans de nombreux laboratoires de SHS, notamment en linguistique, sciences de l'information et de la communication, histoire, géographie, sociologie...

du dépôt légal peuvent être un obstacle à un travail d'analyse de sources par le chercheur, d'où les tentatives de déposer des projets communs dans lesquels les archivistes peuvent jouer un rôle pivot<sup>36</sup>. À côté de la place complexe des archivistes, la place du chercheur lui-même au sein des institutions se trouve réinterrogée. Frédéric Clavert montre que « l'histoire numérique » déploie le rapport réflexif du chercheur à son outil, dans la perspective plus générale de la reconnaissance de ce qu'on pourrait appeler le « chercheur outillé » dans sa pratique sociale et institutionnelle de chercheur. À travers l'exemple des échanges Twitter entre historiens depuis plusieurs années, qu'il a retracé grâce à l'étiquette numérique (ou *hashtag*) #twitterhistorians, il voit émerger des préoccupations spécifiques de ces acteurs de la recherche pris dans le numérique : trouver des emplois qui peuvent accueillir les nouvelles expertises hybrides de l'ingénierie informatique et de la recherche, expérimenter de nouveaux dispositifs pour parcourir la diffusion de l'information historique, etc. On peut souligner à sa suite que les analyses de réseaux et de données numériques proposent une perspective informationnelle (systèmes d'information et traitement algorithmique des données) pour répondre à un problème de communication stratégique à un moment où la globalisation de la science est marquée fortement par l'injonction à l'excellence déployée par les nouvelles gouvernances de la recherche : qui (se) communique le mieux dans les environnements socionumériques et comment. Ces investigations peuvent être alors le socle d'une approche réflexive socio-politique du monde de la recherche – ainsi que dans d'autres domaines. Cependant, des considérations pragmatiques sont importantes à rappeler pour ne pas se perdre dans les SNN, qui non seulement présentent une grande complexité technique, mais peuvent également servir de discours d'accompagnement à la rhétorique de la recherche innovante et sur projet, derrière laquelle se posent des problèmes en termes d'économie et de politique des sciences. Benjamin Thierry rappelle qu'il est ainsi difficile pour un historien, par exemple, de devenir spécialiste du droit, de l'économie, en contexte numérique, ou même de porter la double expertise de programmeur et chercheur en SHS, malgré les prophéties de Le Roy Ladurie sur le futur historien-programmeur<sup>37</sup>. Ici encore, des questions sociales relatives au métier de chercheur se posent : quels moyens lui sont-ils donnés pour faire face à un monde technologique changeant, où l'innovation est une injonction parfois contradictoire ? Doit-on mieux faire converger les compétences dans quelques profils de type « mouton à cinq pattes », ou plutôt mieux les distribuer et les coordonner au sein et aux frontières du monde académique ?

Les sciences humaines et sociales ont depuis quelques années vu s'accroître le regard réflexif des chercheurs sur leurs terrains et corpus, leurs gestes, leurs outils et instruments, en particulier dans le rapport aux technologies de l'information et de la communication en contexte de prolifération des technologies et usages numériques. Le tour d'horizon présenté ici ne se limite pas à une discipline ou à une école théorique précise – même si l'on peut arguer du fait que l'intérêt pour la source

---

<sup>36</sup> Par exemple le projet OTMédia, Observatoire Transmédia (financé par l'ANR en 2010-2013).

<sup>37</sup> Voir ce billet d'Émilien Ruiz, « Les historiens seront-ils finalement programmeurs ? » qui fait le point sur la question, publié sur son blog *La boîte à outils des historiens* le 22/09/2011 [En ligne, URL : <http://www.boiteaoutils.info/2011/09/les-historiens-seront-ils-finalement/>].

numérique native s'ancre dans un « tournant pratique » qui insiste notamment sur l'importance de l'incorporation des activités et leur médiation par des artefacts<sup>38</sup>. Si les humanités numériques ont montré les apports du traitement informatique pour des fonds de documents numérisés, de nouveaux matériaux sont produits dans l'usage même des techniques et environnements du numérique, et ce depuis que les ordinateurs sont nés il y a plus d'un demi-siècle. Il était temps de se pencher sur les problèmes pratiques, méthodologiques et épistémologiques posés par ces données et documents nativement numériques.

Parmi les questions générales posées par cette journée, celle des rapports entre vérité historique et mémoire documentaire dans le contexte numérique ne trouve encore que des réponses non définitives, soumises à l'instabilité des normes dans un contexte de diversification des techniques et des capacités archivistiques, dans les institutions abritant des usages spécialisés aussi bien que dans les environnements d'usages courants comme les réseaux numériques. Si le document numérique est instable, marqué par une obsolescence technique rapide (évolution des supports, formats, codes) et infiniment réinscriptible à différents niveaux de données (du code au texte), quelle valeur de vérité lui accorder ? Quel est le régime de vérité des documents numériques dans une société où l'accès à la technique est généralisé ? La réflexion sur la confiance que le chercheur place en la SNN (via les cadres techniques et institutionnels qui la certifient, la valident, etc.) est liée à la fiabilité des conditions de transmission. Mais au-delà, elle interroge les processus de légitimation et d'autorisation des corpus mémoriels qui sont les inscriptions privilégiées d'une historiographie en devenir. L'intensité des pratiques d'archivage vernaculaires des documents et données numériques, à l'extérieur ou dans les marges des institutions officielles de la mémoire, nous pousse à analyser non plus les nouvelles formes, mais les nouveaux sens d'un patrimoine scientifique et technique en train de se constituer.

Les sources documentaires et de données numériques sont marquées par l'hétérogénéité, et la multiplicité des couches d'information entre interface et machine. Elles composent un mille-feuille numérique, avec des niveaux de temporalité, de spatialité et de granularité qu'il faut savoir repérer à différentes échelles des couches de code et des données pour la bonne conduite de l'analyse, depuis l'interface des écrans jusqu'aux infrastructures de support de la matière et de l'activité numériques. Ces sources ne sont que très difficilement lisibles et compréhensibles en dehors de dispositifs qui conditionnent leur appréhension. Ces dispositifs sont de plusieurs ordres. Ils concernent le stockage, sous la forme d'archives numériques (avec ses modalités de consultation, d'accès, d'analyse, d'enrichissement...). Ils permettent la constitution de collections par des acteurs agréant les sources, les sélectionnant, et éventuellement permettant leur versement dans des corpus (institutions, communautés amateurs, chercheurs...). Ils traitent d'organisation, de classification, et de médiation de la mémoire documentaire par des outils et cadres techniques. Enfin, ils touchent aux écritures du numérique, qui engagent une mise en forme de la pensée à plusieurs niveaux, du format de description des textes jusqu'à l'hypertextualité, la visualisation de données, et la programmation d'actions (interrogation des moteurs de recherche ou des bases de données). La SNN ne dit pas tout, ou peut dire mal : malgré la tendance objectiviste

---

<sup>38</sup> À noter, une frange « post-humaniste » du *practice turn* défend l'idée que les pratiques incluent l'activité de « non-humains » comme les machines (Schatzki, dans Schatzki, Knorr Cetina et Von Savigny, 2001, p. 11).

et positiviste des méthodes informatiques quantitatives, elle n'est pas objective par nature, et n'est pas réductible à la rationalité des processus techniques. Elle constitue un nouveau défi pour ceux qui considèrent, avec Marc Bloch, que « si l'outil ne fait pas la science [...] une société qui prétend respecter les sciences ne devrait pas se désintéresser de leurs outils » (Bloch, 1999).

## Bibliographie

- Madeleine Akrich, Cécile Méadel, et Verena Paravel, « Le temps du mail : écrit instantané et oral médiat », *Sociologie et sociétés*, vol. 32, n° 2, 2000, pp. 153-170.
- Bruno Bachimont, *Le sens de la technique : le numérique et le calcul*, Paris, Les Belles Lettres, 2010.
- Bruno Bachimont, « Arts et sciences du numérique : Ingénierie des connaissances et critique de la raison computationnelle », mémoire d'HDR soutenu à l'Université technologique de Compiègne, janvier 2004.
- Christine Barats (dir.), *Manuel d'analyse du Web*, Paris, Armand Colin, 2013.
- Tiziana Beltrame et Christine Jungen (2013), « Cataloguer, indexer, encoder. Ou comment les données prennent vie », *Revue d'anthropologie des connaissances*, 7 (4), dossier « Les textures matérielles de l'accumulation », p. 747-59.
- Aurélien Berra, « Faire des humanités numériques », dans Pierre Mounier (dir.), *Read/Write Book 2*, Marseille, OpenEdition Press, p. 25-43.
- Jean-François Blanchette, « A Material History of Bits », *Journal of the American Society for Information Science and Technology*, Volume 62, Issue 6, p. 1042-1057, June 2011.
- Marc Bloch, *Apologie pour l'histoire ou Métier d'historien*, Paris, Colin, 1999.
- Danah Boyd et Kate Crawford « Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication & Society*, Volume 15, Issue 5 (Special Issue : A Decade in Internet Time : the Dynamics of the Internet and Society), 2012, p. 662-679.
- Geoffrey Bowker, *Memory Practice in the Sciences*, Cambridge, Mass., MIT Press, 2008
- Éric Brian, « Archives et mémoire des sciences : enjeux historiographiques », *Revue d'histoire moderne et contemporaine*, 2001/5, n° 48-4bis, p. 44-48.
- Évelyne Broudoux, Ghislaine Chartron et Stéphane Chaudiron, « L'architecture de l'information : quelle réalité conceptuelle ? », *Études de communication*, 41, 2013, p. 13-30.
- Évelyne Broudoux, « Autoritativité, support informatique, mémoire », *Archive ouverte en sciences de l'information et de la communication*, @rchivesic 2003 [En ligne, URL : [http://archivesic.ccsd.cnrs.fr/sic\\_00001137.html](http://archivesic.ccsd.cnrs.fr/sic_00001137.html)].
- Luciano Canfora, *Le copiste comme auteur*, Anarchasis, 2012.
- Marie-Anne Chabin, « Archives », *Médium*, 2013/4-2014/1 (n° 37-38), p. 102-117.
- Marie-Anne Chabin, « L'ère numérique du faux », *Médium*, 2012/2 (n° 31), p. 46-66.
- Y. Chevalier et L.-Z. Herscovici (dir.) (2012), *ESSACHESS*, vol. 5, n° 2(10), « Communication and/of memory / Communication et/de la mémoire ».

- Frédéric Clavert et Serge Noiret (dir.), *Contemporary History in the Digital Age / L'histoire contemporaine à l'ère numérique*, Bruxelles, Peter Lang, 2013.
- Florian Cramer, « Digital Code and Literary Text », *Beehive*, 4, 3, 2001 [En ligne, URL : [http://beehive.temporalimage.com/content\\_apps43/cramer/oo.html](http://beehive.temporalimage.com/content_apps43/cramer/oo.html)].
- Nicolas Delalande et Julien Vincent (dir.), *Revue d'histoire moderne et contemporaine*, 2011/5 (n° 58-4 bis), dossier « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? ».
- Sophie Derrot, Louise Fauduet, Clément Oury, Sébastien Peyrard (2012), « Preservation Is Knowledge : A community-driven preservation approach », 9<sup>th</sup> International Conference on Preservation of Digital Objects (iPRES), Canada.
- Mélanie Dulong de Rosnay et Francesca Musiani, « The Preservation of Digital Heritage : Epistemological and Legal Reflections », *ESSACHESS - Journal for Communication Studies*, 2012, vol. 5, no 2(10), « Communication and Memory », p. 81-94.
- Patrice Flichy et Sylvain Parasio, « Sociologie des bases de données : présentation », *Réseaux*, 178-179 (2), 2013, p. 9-19.
- Heather Ford, « Big Data and Small : Collaborations between Ethnographers and Data Scientists », *Big Data & Society*, 1 (2), 2014, p. 1-3.
- Matthew Fuller, *Software Studies : A Lexicon*, Cambridge, MIT Press, 2008.
- Bertrand Gilles, *Histoire des techniques*, Paris, Gallimard/La Pléiade, 1978.
- Jack Goody, *La raison graphique*, Paris, Minuit, 1979.
- Burton Grad et Thomas J. Bergin, « History of Database Management Systems », *IEEE Annals of the History of Computing*, 31 (4), 2009, p. 3-5.
- Éric Guichard, « L'internet et les épistémologies des SHS », *Revue Sciences/Lettres*, n° 2, automne 2014 [En ligne, URL : <http://rsl.revues.org/389>].
- Benoit Habert, et Claude Huc (2010), « Building Together Digital Archives for Research in Social Sciences and Humanities », *Social Science Information sur les sciences sociales*, 49 (3) : 415-443.
- Clarisse Herrenschmidt, *Les trois écritures. Langue, nombre, code*, Paris, Gallimard, 2007.
- Alexandre Hocquet, « RTFM ! La chimie computationnelle : une communauté scientifique loin du libre », dans Christophe Masutti et Camille Paloque-Berges (dir.), *Histoires et cultures du Libre. Des logiciels partagés aux licences échangées*, Paris, Framasoft, 2013.
- Gilbert Hottois, *Le signe et la technique*, Paris, Aubier-Flammarion, 1984.
- Anouk Barberousse et Cyrille Imbert, « Le tournant computationnel et l'innovation théorique », dans Soazig Le Bihan (dir.), *La philosophie de la physique : d'aujourd'hui à demain*, Paris, Vuibert, 2013.
- Matthew Kirschenbaum, *Mechanisms : New Media and the Forensic Imagination*, Cambridge, MIT Press, 2008.
- Stéphane Lamassé et Philippe Rygiel, « Nouvelles frontières de l'historien », *Revue Sciences/Lettres*, 2/2014 « Épistémologies digitales des sciences humaines et sociales » [En ligne, URL : <http://rsl.revues.org/411>].

- Joëlle Le Marec (dir.), *Les études de sciences : pour une réflexivité institutionnelle*, Paris : Archives contemporaines, 2010.
- Muriel Lefebvre, *Approche patrimoniale de la communication scientifique. Les écritures (infra)- ordinaires de la recherche*, HDR en sciences de l'information et de la communication, Toulouse Le Mirail, 2013.
- Lawrence Lessig, *Code and other Laws of Cyberspace*, New York, Basic Books, 1999.
- Lawrence Lessig, *Code 2.0*, New York, Basic Books, 2006.
- Olivier Le Deuff, « Quelles mnémotechniques pour l'Internet ? », *MEI*, n° 32, « Mémoires et Internet », déc. 2010, p. 41-51.
- Marcoccia Michel, « L'animation d'un espace numérique de discussion : l'exemple des forums Usenet », *Document numérique*, 2001/3, vol. 5, 2001, p. 11-26.
- Lev Manovich, *Software takes command*, New York, Bloomsbury Academic, 2013.
- Mark C. Marino, « Critical Code Studies », *Electronic book review*, 2006 [En ligne, URL : <http://www.electronicbookreview.com/thread/electropoetics/codology>].
- Louise Merzeau, *Pour une médiologie de la mémoire*, mémoire de HDR soutenu en Sciences de l'information et de la communication à l'Université Paris X-Nanterre en 2011.
- Pierre Mounier-Khun, *L'Informatique en France de la Seconde Guerre mondiale au Plan Calcul. L'Émergence d'une science*, Paris, Presses de l'Université Paris-Sorbonne, 2010.
- Louis Morville et Peter Rosenfeld, *Information Architecture for the World Wide Web*, Sebastopol (California), O'Reilly, 2006.
- Francesca Musiani, *Nains sans géants. Architecture décentralisée et services Internet*, Paris, Presses des Mines, 2015.
- Roger T. Pédaque (dir.), *Le document à la lumière du numérique*, Caen, C&F édition, 2006.
- Camille Paloque-Berges et Gérald Kembellec, « Nouvelles sources numériques et logiques d'open corpus : l'intérêt d'archiver et de partager des courriers électroniques », *Cahiers de la Sfsic*, n° 9, 2014, p. 239-244.
- Camille Paloque-Berges, *Poétique des codes sur les réseaux informatiques*, Paris, Archives contemporaines, 2009.
- Jussi Parikka et Erkki Huhtamo (dir.), *Media Archaeology, Approaches, Applications, and Implications*, University of California Press, 2012.
- Madeleine Pastinelli, « La mémoire et l'oubli dans l'univers de l'archive totale », *EspacesTemps.net*, 2009 [En ligne, URL : <http://www.espacestemp.net/articles/la-memoire-et-lrsquooubli-dans-lrsquounivers- de- lrsquoarchive-totale/>].
- Marie-Anne Paveau, « Discours et mémoire 7. « La mémoire numérique. Réflexivité et technodiscursivité », billet publié sur le carnet de recherche *Pensée du discours* le 5/10/2013 [En ligne, URL : <http://penseedudiscours.hypotheses.org/8204>].
- Girolamo Ramunni, *La Physique du calcul. Histoire de l'ordinateur*, Paris, Hachette, 1989.
- Emmanuel Ruzé, « Traiter les archives de la Toile. Une histoire d'un système d'information dans une communauté, WordPress (2003-2008) », *Entreprises et Histoire*, 55, 2009, p. 74-89.

- Jean-Michel Salaün, *Vu, lu, su. Les architectes de l'information face à l'oligopole du Web*, Paris, La Découverte, 2012.
- Jean-Michel Salaün, « La redocumentarisation, un défi pour les sciences de l'information », *Études de communication*, 30, 2007, pp. 13-23.
- Valérie Schafer, *La France en réseaux (1960-1980)*, Paris, Nuvis, 2012.
- Theodore Schatzki, Karin Knorr Cetina et Eike Von Savigny (eds.), *The Practice Turn in Contemporary Theory*, London and New York, Routledge, 2001.
- Emmanüel Souchier, Yves Jeanneret, et Joëlle Le Marec, *Lire, écrire, récrire. Objets, signes, pratiques des médias informatisés*, Paris, BPI Centre Pompidou, 2003.
- Benjamin Thierry, « "Révolution 0.1." Utilisateurs et communautés d'utilisateurs au premier âge de l'informatique personnelle et des réseaux grand public (1978-1990) », *Le Temps des médias*, 2012/1 (n° 18), pp. 54-54.
- Benjamin Thierry, « La communication hommes-machines et le développement de l'informatique », *Hermès*, n° 50, 2008/1, pp. 91-98.
- Katie Vann et Geoffrey Bowker, « Instrumentalizing the Truth of Practice », *Social Epistemology*, Vol. 15, n° 3, 2001, pp. 247-262.
- Franck Varenne, « What Does a Computer Simulation Prove? The Case of Plant Modeling at CIRCAD (France) », dans N. Giambiasi and C. Frydman, *Simulation in Industry, Proceedings of the 13<sup>th</sup> European Simulation Symposium*, Ghent, SCS Europe Bvba, 2001, pp. 549-554.
- Barry Wellman, « Studying Internet Studies Through the Ages », dans Mia Consalvo and Charles Ess (eds) *Blackwell Handbook of Internet Studies*, West Sussex, Wiley-Blackwell, 2011.

## L'auteur

Ingénieure de recherche au laboratoire Histoire des Techno-Sciences en Société (HT2S, Conservatoire national des arts et métiers), et chercheuse associée au pôle Trajectoires du numérique de l'Institut des Sciences de la Communication du CNRS (ISSC), Camille Paloque-Berges est docteure en Sciences de l'information et de la communication, spécialisée dans l'histoire et les cultures des réseaux informatiques. Après des premières recherches sur la créativité dans la programmation informatique (*Poétique des codes*, 2009, Archives contemporaines), elle a conduit une thèse intitulée *Histoire vernaculaire d'Internet. Emergence et médiations d'un folklore de réseau* (2011, Université Paris 8) et mené des travaux sur les méthodologies associées à l'analyse des documents, archives, sources et patrimoine nativement numériques. Ses recherches actuelles concernent le développement socio-technique d'Internet en Europe et le rôle des associations d'utilisateurs qui participent à faire un objet de débat public en France dans les années 1980 et 1990.

Parmi ses dernières publications, on peut trouver : « Usenet as a Web Archive: Multi-layered Archives of Computer-Mediated-Communication » (dans N. Brugger (ed.), *Web 25: Histories from the First 25 Years of the World Wide Web*, Peter Lang Publishing, janvier 2017). « Vers des lieux de mémoire réticulaires ? Construire un patrimoine de la communication des sciences et des techniques du numérique » (RESET, n°6, novembre 2016), « L'imaginaire du "grand public" au tournant du Web (1993-1997) » (*Revue française des sciences de l'information et de la communication*, n°7, 2015).. Ses derniers travaux sur l'arrivée des réseaux Internet en France sont en cours de publication

(« Mapping a French Internet Experience. A Decade of Unix Networks Cooperation (1983-1993) », dans *Routledge Companion to Global Internet histories*).

### Résumé

Cet article vient proposer une synthèse et une extension d'une journée d'étude prenant pour problématique les questions méthodologiques et épistémologiques associées aux sources numériques natives (SNN), c'est-à-dire produites originellement dans le contexte d'usage des outils et environnements numériques. On s'attache d'abord à décrire le type de sources dont il s'agit du point de vue de leur matérialité, afin de les faire apparaître derrière le mille-feuille des artefacts informatiques. Données, code, documents, équipement : autant de niveaux allant du logiciel au matériel informatique dans lesquels s'inscrivent les gestes et mots du présent. Ces sources sont ensuite considérées à travers le processus de mise en archive, censée leur donner une valeur patrimoniale stable – et soulevant les problématiques de l'originalité, de l'intégrité et de l'authenticité de ces sources. Enfin, un échantillon de méthodologies associées à l'analyse des SNN est présenté à travers une observation à plusieurs échelles, avant de conclure ce tour d'horizon par les questionnements épistémologiques entraînés par la réflexivité du chercheur qui les manipule.

**Mots clés :** numérique, sources, archives, réseaux, méthodologie, épistémologie

### Abstract

This article offers an overview and builds upon a workshop concerning the methodological and epistemological issues associated with native digital sources (NDS) – that is, sources originally produced in the context of the use of digital tools and environments. It first seeks to describe the type of sources at issue from the point of view of their materiality in order to distinguish them from the many other types of digital artifact. Data, code, documents, equipment: from software to hardware, the gestures and words of the present are inscribed at all levels. The article then examines these sources from the perspective of the archival process that is supposed to assign them a stable value as heritage while at the same time raising issues concerning the sources' originality, integrity and authenticity. Finally, the article surveys the methodologies associated with NDS analysis at several levels and concludes with a consideration of the epistemological issues that are entailed by the researcher's reflexivity.

**Key words :** digital, sources, archives, networks, methodology, epistemology

**Pour citer cet article :** Camille Paloque-Berges, « Les sources nativement numériques pour les sciences humaines et sociales », *Histoire@Politique*, [en ligne], n° 30, septembre-décembre 2016, [www.histoire-politique.fr](http://www.histoire-politique.fr)