RESET
Recherches en sciences sociales sur Internet

6 | 2017
Patrimoine et patrimonialisation numériques

Vers des lieux de mémoire réticulaires ?

Construire un patrimoine de la communication des sciences et des techniques du numérique

Towards reticular memory spaces? Building heritage for the communication of digital science and technology

Camille Paloque-Berges

Édition électronique
URL : http://journals.openedition.org/reset/839
DOI : 10.4000/reset.839
ISSN : 2264-6221

Éditeur
Association Recherches en sciences sociales sur Internet

Référence électronique
Camille Paloque-Berges, « Vers des lieux de mémoire réticulaires ? », *RESET* [En ligne], 6 | 2017, mis en ligne le 30 octobre 2016, consulté le 01 mai 2019. URL : http://journals.openedition.org/reset/839 ; DOI : 10.4000/reset.839

Vers des lieux de mémoire réticulaires ?

Construire un patrimoine de la communication des sciences et des techniques du numérique

Towards reticular memory spaces? Building heritage for the communication of digital science and technology

Camille Paloque-Berges

Introduction

1 Le patrimoine *du* numérique, si jeune soit-il, constitue un élément non négligeable de la construction d'un patrimoine numérique plus général, repensé par le domaine des humanités numériques, dont il nourrit la culture historique et épistémique (Bachimont, 2010 ; Mounier, 2010 ; Vanhoutte, Nyhan, & Terras, 2013 ; Guichard, 2014). Pris en charge par le patrimoine contemporain des sciences et techniques, le numérique se confond souvent avec l'informatique. Son histoire scientifique et technique s'inscrit alors dans des logiciels et des machines, sa mémoire dans les témoignages des acteurs (Ballé, Cuenca & Thoulouze, 2010). Mais ce programme comprend rarement les usages numériques de ses acteurs, par exemple les courriers électroniques, pourtant explicitement porteurs d'une mémoire sociotechnique (Paravel, 2007). Or, comme nous l'avons défendu ailleurs, l'idée d'un patrimoine de la communication émerge aujourd'hui comme l'un des éléments clefs des rapports entre mémoire et numérique (Paloque-Berges & Schafer, 2015).

2 Cet article propose un retour réflexif sur la construction, à l'occasion d'une étude historique[1], d'un corpus de communications numériques échangées entre chercheurs et ingénieurs ayant joué un rôle dans le développement d'Internet en France entre 1983 et 1993. Des listes et groupes de discussion électroniques ont été rassemblés à partir de fonds d'archives privées maintenus sur le Web[2]. La manipulation de ce corpus d'archives à but d'analyse historique a fourni l'occasion de réfléchir en pratique à la patrimonialisation de ce type de documents. Qu'est-ce que la construction d'un corpus de

recherche sur la base de documents nativement numériques peut nous apprendre de la mise en patrimoine de ces nouveaux matériaux ?

3 Le premier objectif de cet article est de confronter les premières formes communautaires de valorisation patrimoniale de la communication médiée par les réseaux (CMC[3]) aux cadres et normes du patrimoine institutionnel. Une première forme de réflexivité est en jeu, relevant d'une méthodologie éprouvée par l'étude des sciences en société (Le Marec, 2010) : par l'étude d'un terrain (ici la mise en patrimoine des histoires et mémoires scientifiques et techniques d'Internet dans des communautés), on observe des pratiques qui ont anticipé et accompagné l'évolution des normes de patrimonialisation institutionnelle du numérique. Signe d'un intérêt institutionnel, l'UNESCO propose depuis 2003 de se pencher sur le patrimoine numérique et en particulier les ressources « nativement numériques » (*born digital*). Si ces dernières tendent souvent à être réduites à leur fonction de support (Abdelaziz, 2004 ; Collectif, 2012), notre matériau relève aussi du « patrimoine immatériel » du numérique, qui inclut les « pratiques, représentations, expressions, connaissances, savoir-faire », ainsi que les « instruments, objets, artefacts et espaces culturels » (Desvallées, 2004). On verra ainsi comment la conception même de ce qui fait patrimoine peut évoluer à la croisée de ces définitions et des pratiques patrimoniales institutionnelles initiées sur ce type de matériaux.

4 Le second objectif est d'expérimenter en pratique cette évolution, à travers la construction d'un corpus test. En une deuxième forme de réflexivité, il s'agit de tester les méthodes formelles et normes institutionnelles de la mise en patrimoine numérique à l'occasion de l'étude de ce corpus, en composant avec les méthodologies d'analyse qui ont pu être développées. Celles-ci reposent sur une double perspective historique et communicationnelle qui s'appuie aussi sur les apports des *Science and Technology Studies* pour l'étude de la communication numérique en réseau (Paravel, 2007). En ceci, j'adopte la perspective d'une documentarisation - à savoir la manipulation et la gestion des documents, tenant compte de l'évolution technologique des outils et environnements documentaires (ici numériques), ainsi que de leurs normes associées, dans le but de mieux les consulter et les exploiter (Pédauque, 2006 ; Zacklad, 2007).

5 En somme, cet article se propose d'étudier la préparation et la construction d'un corpus de textes numériques pris comme objet de recherche[4], tout en expliquant les qualités de matériaux nativement numériques encore mal connus – les courriers électroniques, prototypes de la CMC sur Internet. Poursuivant l'idée selon laquelle une mémoire sociale, technique et culturelle se transmet dans ces objets de CMC archivés puis redocumentarisés à des fins de recherche historique ou patrimoniale, l'enjeu de cette étude est de montrer, à travers la prise en mains de ces nouveaux matériaux, les évolutions des « lieux de mémoire » à l'épreuve d'une mémoire réticulaire[5].

Communications scientifiques et techniques sur les réseaux numériques : une patrimonialisation en cours

6 Sujet « non noble » en histoire, la communication est un objet de recherche récent de l'historiographie (Schafer & Thierry, 2013). Matérialisée dans les réseaux informatiques ainsi que les données et documents numériques générés à travers leurs usages, elle est aussi objet d'étude des mémoires construites et médiées sur Internet (Pastinelli, 2009, Ruzé, 2009). En pratique, les acteurs de la jeune histoire de l'informatique et du

numérique participent activement à construire la valeur patrimoniale de cette mémoire et de cette histoire (Paloque-Berges, 2014). On s'attache dans cette partie à mieux cerner un patrimoine de la communication numérique en voie d'institutionnalisation.

Paternités et légitimations de ce patrimoine

7 Le patrimoine des sciences et des techniques porte essentiellement sur les inventions et les innovations, symboles positifs du progrès technoscientifique[6], et le cas de la communication numérique ne fait pas exception. La patrimonialisation d'Internet, calquée sur la jeune historiographie du réseau des réseaux, porte ainsi la mémoire épidictique des « pères fondateurs » d'une technologie victorieuse, qui s'est imposée comme un standard (Russell, 2014). Cette mise en patrimoine procède d'abord des acteurs majeurs de l'histoire technique et économique d'Internet eux-mêmes, qui participent activement de l'institutionnalisation de leurs innovations technologiques.

8 Par exemple, l'*Internet Society* (ISOC), fondée en 1992 afin d'accompagner la définition technique de standards et normes, et plus généralement la gouvernance d'Internet, est active en termes de préservation – et de promotion – de l'histoire d'une technologie où la plupart des membres d'ISOC sont des acteurs. Son *Internet Hall Of Fame* célèbre la mémoire des « grands hommes » d'Internet (ceux ayant développé ses propriétés techniques fondamentales) et déroule l'histoire officielle. Il prépare aussi son histoire future, en se positionnant de manière prospective sur ses développements techniques, et en adoptant une perspective de conseil et d'accompagnement de l'innovation.

9 Autre exemple, dans le domaine de l'économie numérique de l'information : la préservation et la gestion depuis 2001 des archives d'Usenet (service de CMC et ancêtre illustre des médias socio-numériques du Web) au sein du service de forum Google Groups ont accompagné les efforts de légitimation de l'entreprise auprès des publics d'utilisateurs, à une époque où Google était en phase de développement et de diversification de ses activités. Google, alors en train de gagner la guerre de moteurs de recherche, s'est érigé par ce geste en protecteur du passé du réseau, ainsi qu'en candidat à sa propre reconnaissance au sein de cette histoire (Paloque-Berges, 2017). C'est une problématique classique de la mise en patrimoine que ce caractère stratégique et ambivalent. Loin donc d'être inédite, cette problématique permet toutefois de comprendre comment de nouvelles organisations associatives ou à but lucratif travaillent à leur propre légitimation en se donnant comme garantes de la mémoire de la technologie au cœur de leur activité. De fait, la question du patrimoine des réseaux numériques trouve sa place dans la réflexion sur le renouvellement des institutions (Dulong de Rosnay & Musiani, 2012). Mais de manière plus originale, Internet est aussi le médium de cette double dynamique d'institutionnalisation-patrimonialisation, déployant une « mémoire du réseau [...] à l'image du réseau lui-même » (Schafer, 2012), qui peut entraîner quelques confusions quant au niveau d'interprétation de la mission patrimoniale affichée.

10 Au-delà de la mémoire positive et stratégique, la multitude d'archives conservées – due au caractère ubiquitaire du document numérique, infiniment copiable – révèle la complexité sociotechnique de l'histoire des réseaux numériques. Un exemple éclairant réside dans l'archivage des *Requests for Comments* (RFC), documents de travail spécifiant les standards infrastructurels d'Internet, témoignant de la redéfinition de la littérature grise dans la communication numérique : leur forme finale est le résultat de discussions

et de délibérations d'ingénieurs et techniciens via Internet. Préservés par l'*Internet Engineering Task Force*, à qui échoie cette mission de standardisation depuis 1986, les RFC portent une mémoire du développement d'Internet qui passe par la multitude de ses acteurs, de leurs accords à leurs désaccords. Ce patrimoine n'est pas seulement celui d'une innovation finalisée mais aussi celui de sa construction, de sa circulation et de sa mise en question sur les plans sociaux, culturels, voire politiques et économiques ; non seulement celui des « pères fondateurs », mais aussi celui des communautés, des collectifs et réseaux sociaux qui constituent leur environnement.

11 En ceci, les acteurs des réseaux numériques ont précédé dans leurs pratiques les chercheurs et professionnels du patrimoine. Ces derniers ont quant à eux ouvert les cadres méthodologiques et théoriques permettant la reconnaissance non plus seulement des « grands hommes » et des innovations finalisées, mais aussi des collectifs et des étapes intermédiaires de leur réalisation, notamment dans le champ de l'histoire des sciences (Brian, 2001) et du patrimoine contemporain des sciences et des techniques (Ballé, Cuenca & Thoulouze, 2014). Toutefois, malgré l'intérêt maintenant institutionnalisé des politiques culturelles pour l'organisation sociale des sciences et des techniques (Bergeron & al., 2014), hérité notamment du regard sociologique posé sur les équipements, les infrastructures et les réseaux (Callon, 1989 ; Latour & Woolgar, 2006), la patrimonialisation publique des réseaux numériques reste succincte, limitée à l'exposition de quelques machines dans les musées de science et technologie, ou aux politiques de dépôt légal des bibliothèques nationales (Chevallier & Illien, 2011).

La valeur patrimoniale d'un objet « ordinaire »

12 Les courriers électroniques font partie des archives « que l'on fabrique tous les jours [...] un problème très contemporain de préservation de la mémoire » (Caillet & Maitte, in Bergeron & al., 2014 : 212). Leur archivage, qu'il relève d'une initiative personnelle, informelle ou fasse partie d'une politique de sauvegarde documentaire au sein des systèmes d'information, porte la marque d'un travail relevant d'abord des « pratiques ordinaires engageant de manière personnelle les mains des chercheurs » (Dalle-Nazébi & Aguéra, 2013). Version contemporaine des matériaux ordinaires, comme les correspondances, reconnus par l'historiographie depuis l'école des Annales, ils restent peu étudiés hors des analyses de l'anthropologie historique, telles que l'ouvrage *Lieux de savoir*, qui rassemble des analyses sur l'épistolaire papier aussi bien que numérique (Jacob, 2011).

13 Les courriers électroniques sont les premières formes de ce genre médiatique nouveau qu'est la CMC. Déployée dès les années 1970 sur les réseaux informatiques, sa valeur historique en tant qu'accompagnement sociotechnique du développement d'Internet est attestée en histoire sociale des technologies (Abbate, 2000) comme en sociologie des médias, des sciences et des techniques (Flichy, 2001 ; Paravel, 2007). Support de la communication entre scientifiques et ingénieurs participant à la construction de ces réseaux, elle est dès le départ un moteur de développement récursif : c'est pour mieux communiquer entre pairs que l'on développe les réseaux, c'est pour mieux développer le réseau que l'on communique entre pairs par son moyen. Je me concentre en particulier sur les formes collectives et asynchrones de la CMC : les listes et les groupes de discussion en ligne. Médiums sociotechniques de communication distribuée, elles constituent l'un des lieux de la construction, de la communication et de la mise en débat des savoirs

savants contemporains dans les réseaux numériques. Elles dessinent le contour d'une « république des informaticiens » (Flichy, 2001) à l'origine des réseaux, dont il est possible d'analyser les rituels énonciatifs ordinaires, entre langue technique et langue courante (Mourlhon-Dallies & Colin, 2004, Hert, 1998), ainsi que les formes de la démocratie technique expérimentées en ligne (Paravel, 2007).

14 La valeur patrimoniale des listes et groupes électroniques s'inscrit dans le cadre d'un patrimoine de « l'écriture ordinaire de la recherche », composé des écritures intermédiaires précédant l'écrit finalisé et révélant la science en train de se faire dans son rapport au social (Lefebvre, 2013). Ils constituent des « bassin[s] référentiel[s] [...] lieux d'une histoire où s'entreposent, strates après strates, les gisements des connaissances collectives » (Paravel, 2007 : 1108), et auxquels les collectifs eux-mêmes, en particulier s'ils sont impliqués dans la maintenance de leurs archives, renvoient volontiers comme mémoire de leur activité passée, à des fins de régulation des usages ou d'appropriation d'une culture commune (Pastinelli, 2009 ; Ruzé, 2009). Leur archivage automatique par les systèmes d'information numériques relève d'un enregistrement de l'acte de communication qui comprend aussi bien le texte des communications que leur paratexte – se matérialisant dans les métadonnées des messages. Cet enregistrement s'accompagne des fonctions relatives au traitement numérique des messages : distribution et diffusion, tri, classement, format, information sur les routes que prennent les messages, etc. Listes et groupes sont en cela une extension du bureau du chercheur dans la sphère communicationnelle et les supports de la construction et de la structuration de communautés savantes. La ténacité des écrits des correspondances électroniques en fait ainsi de « puissants instruments de mémoire » (*ibid.* : 1100).

15 Comme pour les RFC précédemment évoquées, communautés et organisations associatives ont tendance à héberger les archives de leur CMC sur leurs sites Web. C'est le cas par exemple de l'organisation RIPE NCC (Réseaux IP Européens – *Network Coordination Centre*), qui gère l'enregistrement des adresses Internet (IP – *Internet Protocol*). Son site offre l'accès aux archives des nombreuses listes qu'elle administre pour ses membres ou maintient pour des « raisons historiques » (dixit le site de RIPE), qui viennent compléter des raisons d'ordre informationnel et documentaire. En France, le réseau académique de la recherche française RENATER offre une « listes des listes » (ou « universalistes »), proposant la gestion et les archives de listes universitaires créées pour les besoins d'un projet de recherche, pédagogique ou administratif, et dont la durée de vie est variable. Ces archives ont un but d'abord fonctionnel : permettre à leurs utilisateurs de garder la trace de leurs activités de travail. C'est à partir d'une exploration de ces archives de listes que j'ai débuté le travail de mise en corpus qui sert de base documentaire à l'expérience décrite ci-après.

16 En définitive, on peut considérer ce matériau comme un patrimoine intermédiaire, au sens propre – parce qu'il est en cours de définition – et au sens figuré, car sujet à l'ouverture sociale et culturelle de mémoires scientifiques et techniques.

Aspects pratiques et précautions d'usage d'un corpus à visée patrimoniale

17 Mon travail, portant sur l'histoire et la mémoire des collectifs d'utilisateurs précoces d'Internet, s'attache à évaluer les aspects pratiques de ces deux domaines d'étude : les

analyses de données ont été donc accompagnées par une analyse expérimentale des méthodes et méthodologies pour manipuler les documents nativement numériques abritant ces données (Paloque-Berges, 2016). Les initiatives de patrimonialisation décrites ci-dessus ont été déployées en dehors d'un cadre légal de protection patrimoniale ; les normes qu'elles suivent sont essentiellement techniques, relatives à la nature documentaire des courriers électroniques et des systèmes d'information qui leur offrent un cadre utilitaire. Je me suis demandé comment ces matériaux en voie de patrimonialisation pouvaient être pensés dans un cadre patrimonial plus normatif, en prévoyant les contraintes institutionnelles attachées à leur exploitation.

18 Deux cadres normatifs patrimoniaux se présentent : celui des collections du patrimoine culturel dans les musées et les bibliothèques ; celui des corpus de sources et de données préservés au titre du patrimoine de la recherche. C'est le deuxième qui servira de cadre principal à mon travail. L'objectif était de préparer le corpus aux standards académiques recommandés par les humanités numériques – qui travaillent à définir de nouvelles normes d'ingénierie scientifique dans le contexte du traitement et de la diffusion numériques des données et documents. Idéalement, cette préparation devait aboutir à rendre le corpus interopérable, c'est-à-dire mettre en œuvre les conditions d'accessibilité numérique des données et documents de la recherche, ces conditions impliquant une identification et un formatage du corpus qui soient standardisés et pérennes (Pouyllau, 2010).

Définitions et collectes de corpus à partir d'archives vivantes

19 La problématique guidant la mise en corpus était plurielle. D'une part, le questionnement historique : qui sont les collectifs pionniers de l'Internet en France (1983-1993[7]) et comment communiquent-ils dans une logique sociotechnique de réseau à des buts de coordination ? D'autre part le questionnement documentaire, relatif au parti pris de travailler en priorité sur des sources nativement numériques : comment ces activités ont-elles été documentées et ces documents préservés – et comment les prendre en charge dans le cadre d'un travail à visée patrimoniale ? Si l'on souhaite ici surtout répondre au second questionnement, la problématique historique doit être rappelée, car elle suppose une connaissance du contexte permettant d'avancer l'hypothèse suivante : les courriers électroniques, en particulier utilisés dans les échanges collectifs, sont un vecteur de cette communication à but de coordination et ont produit des documents que l'on peut envisager comme sources. La collecte des sources documentaires est primordiale pour des raisons de conservation – en particulier si elles sont hébergées sur les réseaux numériques, et donc propices à disparaître, mais aussi dans le cas où l'on souhaiterait effectuer une analyse appareillée (avec l'aide d'un logiciel) d'ordre qualitatif et/ou quantitatif.

20 J'ai choisi deux formats de communication asynchrone à distance, ceux les plus couramment utilisés dans le contexte des réseaux pré-Web pour s'informer et discuter sur Internet entre spécialistes :

- les listes (ou *mailing-lists*), développées dans la lignée de l'invention du courrier électronique (*email*) au tournant des années 1970 : des messages envoyés collectivement avec pour but la diffusion d'information et/ou la discussion (*diffusion / discussion list*), avec gestion à la main ou automatisée de leur administration (dont l'archivage) ;

- les groupes (ou *newsgroups*), aux fonctions similaires, mais sur un réseau dédié : le service Usenet créé en 1979, un réseau de groupes de discussion thématiques, initié par des informaticiens travaillant sur les systèmes ouverts et les réseaux, dont le protocole fait partie de la famille Internet depuis 1986, devenu l'un des premiers et plus larges réseaux sociaux médiés par les réseaux numériques.

21 Ma recherche a ciblé des listes et groupes ayant pour thématique principale Internet et les technologies informatiques de réseau, au carrefour des acteurs du monde académique français (laboratoires de recherche ou services de support informatique), leurs interlocuteurs dans les laboratoires de R&D et services informatiques d'entreprise, et des amateurs impliqués. Deux difficultés initiales sont apparues.

22 A. La première concerne la localisation de gisements d'archives pertinents pour la période. Cette difficulté traduit le peu d'attention accordée à l'archivage systématique des courriers électroniques avant les années 1990. Or, l'analyse contextuelle montre que les collectifs recherchés sont familiers avec les messageries électroniques depuis au moins la fin des années 1970, au gré d'échanges scientifiques transatlantiques et de l'expérimentation sur des réseaux locaux.

23 La première piste pour cette localisation est la grande archive en ligne de l'ancêtre des forums Web, Usenet, maintenue par Google et accessible librement sur le Web[8]. Usenet, service de discussion sur réseau informatique décentralisé et géré par ses utilisateurs, existe depuis 1979 et propose d'innombrables groupes thématiques. Dès 1989 s'ouvre une branche dédiée à FNET, pivot essentiel de l'introduction d'Internet en France[9]. À partir de 1993, des branches linguistiques sont développées (fr. pour les usagers francophones), au moment même où Usenet voit le nombre de ses utilisateurs s'accroître considérablement. Alors que les technologies Web contribuent à l'intéressement de publics plus larges à Internet, 1993 reste une date symbolique de la « culture Internet » pour les utilisateurs précoces (Schafer, Paloque-Berges & Georges, 2015), et permet de borner mon corpus. Afin que le corpus puisse faire l'objet d'un traitement documentaire exhaustif sur un échantillon, je me suis focalisée sur des groupes spécifiques : fnet.general, fr.network.divers et fr.comp.infosystemes, où se concentrent à l'époque les professionnels et amateurs de l'informatique de réseau. Les conversations y portent sur la transition de l'Internet des spécialistes à un Web d'accès général, évoquant les derniers logiciels intéressants pour se connecter au Web, les nouveaux fournisseurs d'accès à Internet, les représentations d'Internet dans les médias, les nouveaux usagers, etc. (Paloque-Berges, 2015).

24 La seconde piste est l'archive de listes maintenue par le réseau RENATER, en activité depuis 1993 mais abritant des listes plus anciennes. Deux listes répondent à mes critères de sélection (avoir été créées par des acteurs académiques de la normalisation des réseaux Internet, en France, entre 1983 et 1993) : la liste IP, un outil de communication du « Groupe Réseaux »/GERET, collectif travaillant sur les techniques des protocoles d'Internet, initiée en 1989 ; la liste DNS, dédiée aux travaux sur les noms de domaines Internet (*Domain Name Systems*), initiée en 1993. Ces listes correspondent aussi au critère d'exhaustivité du corpus, dans la mesure où c'est la totalité des échanges (dans une période donnée) qui vont être intéressants pour l'analyse.

25 B. La deuxième difficulté concerne la collecte des données, qui pose des problèmes à plusieurs niveaux. En effet, les systèmes qui gèrent ces archives en ligne (de manière largement automatisée) n'ont pas de visée autre qu'informationnelle et technique : leur

maintenance est parfois défaillante, peu ou pas accompagnée de fonctions de médiation ou de possibilité de communication avec l'administration[10]. Ensuite, ces services bloquent la récupération automatique des archives par des logiciels d'aspiration des données ou des scripts permettant leur moissonnage. Enfin, ces archives sont des objets documentaires qui superposent plusieurs couches d'information. En soi, ce sont des objets simples : les courriers électroniques peuvent être récupérés sous la forme de fichiers textes « simples », ou « brut » – *plain text*[11]. Cependant, comme tout document numérique, ils sont destinés à être traités par différents systèmes d'information au cours de leur circulation, au premier chef les logiciels de messagerie, qui ont chacun des propriétés et paramètres différents. Ce traitement laisse des traces aussi bien dans les données de messages que dans les métadonnées (les données décrivant les données à transmettre et précisant comment faire). C'est le cas par exemple pour l'interprétation des accents : l'encodage des signes diacritiques[12] peut ne pas être reconnu par le logiciel qui aura alors recours à des codes spéciaux pour les indiquer. Nous avons donc affaire à un matériau qui a subi de nombreuses re-documentarisations préalables.

26 Des bricolages techniques peuvent pallier ce problème. Par exemple, le copier-coller est une alternative intéressante pour des petits corpus. Cependant, il peut entraîner d'autres problèmes : copier-coller depuis une interface Web implique que des métadonnées invisibles sont copiées en sus des données ciblées. Ces métadonnées invisibles sont relatives à la structure de l'interface Web[13]. Une fois l'extrait collé dans un fichier, elles peuvent venir s'insérer dans les données d'origine et brouiller la lecture (le processus de re-documentarisation devient alors visible). L'accès à des archives « brutes », c'est-à-dire dans un état antérieur à leur re-documentarisation sur un site Web d'archivage, est recommandé, par exemple grâce au contact avec un administrateur ou propriétaire de la liste. Cette solution a l'avantage de fournir des documents en texte brut, une couche précieuse parce qu'elle donne accès aux métadonnées du message, permettant notamment de retracer son historique de documentarisation. En illustration (figures 1 et 2[14]), deux versions d'un même message : la première interfacée à travers le système de consultation des messages archivés et publiés sur le Web au sein du système Google Groups ; la seconde, en texte brut avec les métadonnées d'origine.

Figure 1 : version interfacée du courrier électronique – redocumentarisée par l'interface de groupes de discussion du service Google (Google Groups)

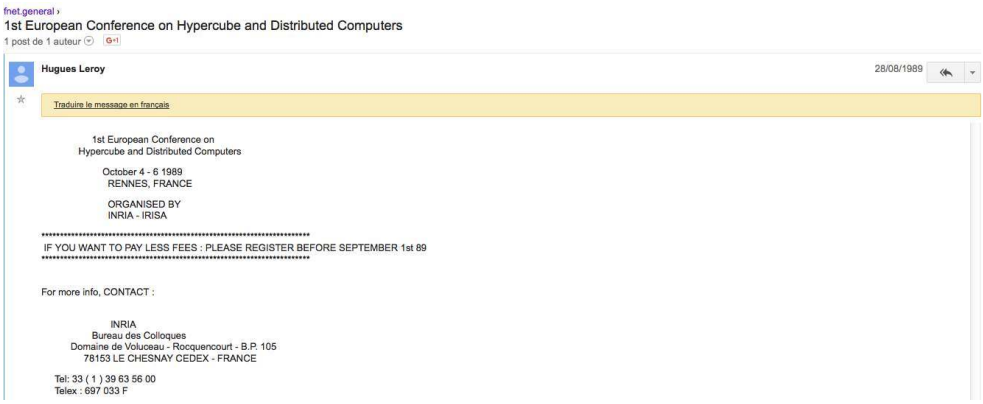
fnet.general ›
1st European Conference on Hypercube and Distributed Computers
1 post de 1 auteur
Hugues Leroy 28/08/1989
Traduire le message en français
1st European Conference on
Hypercube and Distributed Computers
October 4 - 6 1989
RENNES, FRANCE
ORGANISED BY
INRIA - IRISA
****
IF YOU WANT TO PAY LESS FEES : PLEASE REGISTER BEFORE SEPTEMBER 1st 89
****
For more info, CONTACT :
INRIA
Bureau des Colloques
Domaine de Voluceau - Rocquencourt - B.P. 105
78153 LE CHESNAY CEDEX - FRANCE
Tel: 33 (1) 39 63 56 00
Telex : 697 033 F

Figure 2 : version complète du courrier électronique avec les données de contenu (message) et les métadonnées de description et d'instruction (en-tête).

```
Path: gmdzi!unido!mcsun!inria!irisa!hleroy
From: hle...@irisa.irisa.fr (Hugues Leroy)
Newsgroups: eunet.general,fnet.general
Subject: 1st European Conference on Hypercube and Distributed Computers
Message-ID: <1404@irisa.irisa.fr>
Date: 28 Aug 89 08:36:50 GMT
Distribution: eunet
Organization: IRISA, Rennes (Fr)
Lines: 250
Xref: gmdzi eunet.general:954
Posted: Mon Aug 28 09:36:50 1989


                1st European Conference on
            Hypercube and Distributed Computers

                    October 4 - 6 1989
                      RENNES, FRANCE

                      ORGANISED BY
                      INRIA - IRISA

**************************************************************************
 IF YOU WANT TO PAY LESS FEES : PLEASE REGISTER BEFORE SEPTEMBER 1st 89
**************************************************************************


For more info, CONTACT :


                         INRIA
                   Bureau des Colloques
          Domaine de Voluceau - Rocquencourt - B.P. 105
               78153 LE CHESNAY CEDEX - FRANCE

     Tel: 33 ( 1 ) 39 63 56 00
     Telex : 697 033 F
```

Le positionnement juridique et éthique face aux données récoltées

27 Aussi bien le chercheur que le professionnel du patrimoine, dans leur usage des documents nativement numériques produits sur les réseaux, doivent se demander dans quelles conditions ils peuvent exploiter et publier les données récoltées. Les données publiées relèvent en général du droit des personnes et leur réutilisation et republication peuvent poser problème. C'est une des raisons pour laquelle le dépôt légal du Web à la BnF restreint la consultation de ses collections patrimoniales à la lecture seule (sans possibilité de modifier – par exemple en annotant – ou de récupérer les documents et données) et sur place.

28 Le rapport éthique aux données récoltées sur le Web est peu formalisé en France en sciences humaines et sociales, en tout cas peu explicite, du moins jusqu'à la publication de préconisations par des chercheurs canadiens dans l'un des premiers ouvrages de méthodologie d'analyse des données Web (Latzko-Toth & Proulx *in* Barats, 2013). Deux niveaux d'encadrement y sont mis au jour : un cadre réglementaire fixant l'ensemble des contraintes formelles posées par un contexte institutionnel, à travers des comités de réglementation vérifiant le respect de critères juridiques et éthiques du chercheur ; un cadre normatif définissant les principes consensuels internes à une discipline ou un courant de recherche émettant, par le biais de sociétés savantes, des codes de bonne conduite et de bonnes pratiques (compétence professionnelle, intégrité, responsabilité scientifique et sociale, respect pour les droits, la dignité et la diversité des personnes)[15]. Ce positionnement réflexif doit être mesuré à l'aune des propriétés des médias

numériques. Il faut ainsi, selon Latzko-Toth et Proulx (*ibid.*), prendre en compte les qualités documentaires de l'information en réseau :

> la « recherchabilité » : l'indexation de l'information lui permet d'être trouvée par les moteurs de recherche Web, ce qui détermine l'accès à des données qui autrement ne seraient pas ou peu visibles ;
> l'ubiquité : la visibilité de l'information n'est pas limitée à un seul lieu puisqu'elle est copiable et diffusable ;
> la persistance : l'information possède une rémanence concrétisée par des traces (métadonnées) laissées au cours de leur cycle documentaire ;
> la mutabilité : malgré sa persistance, l'information est instable et labile, pouvant disparaître et se modifier totalement ou partiellement ;
> l'invérifiabilité : l'identification des acteurs et témoins interrogés est complexe, en particulier dans le cas de l'utilisation de pseudonymes ou de données massives, notamment pour obtenir leur accord pour la republication des données.

29 Dans le cas des emails, la question des données personnelles est cruciale : nom, prénom, emploi et/ou affiliation, adresse postale, électronique et coordonnées téléphoniques professionnelles et/ou personnelles sont des données d'usage courant. On les trouve à plusieurs niveaux : dans les métadonnées, dont elles sont la substance (adresse email, nom de l'expéditeur, du destinataire, fichier de signature électronique, etc.). Mais on les observe bien évidemment aussi dans le corps des messages. Si l'on envisage que le corpus soit réutilisé par d'autres chercheurs, et *a fortiori* s'il fait l'objet d'une publication numérique pour le rendre accessible, le travail de dés-identification des auteurs d'email est quasi impossible, voire non souhaitable car on peut perdre des données précieuses pour l'analyse (on y reviendra).

30 Cette logique devient encore plus problématique lorsque les données sont recherchées, fouillées et analysées de manière automatisée. Les données peuvent n'avoir aucune importance si elles sont trouvées dans un email au hasard d'une recherche sur les archives Google de Usenet. Cependant, et cela illustre les aspects de la « *recherchabilité* » et de la *persistance*, la recherche systématique d'un nom à l'échelle de toute l'archive Usenet peut laisser entrevoir un parcours de vie ou une prise de position qui n'auraient pas été visibles autrement, une évolution de la situation professionnelle au fil du temps par exemple. Quand Google a acheté et rendu accessibles sur le Web les archives de Usenet en 2001, on a pu apprécier à cet égard les premières manifestations de ce qui sera appelé plus tard le « droit à l'oubli ». Nombre d'anciens usagers de Usenet n'apprécient pas de voir leurs interventions passées ressurgir sur les forums grâce à un moteur de recherche, Usenet étant connu notamment pour ses disputes pléthoriques et hyperboliques, et ils souhaitent les voir disparaître des archives publiées sur le Web (Paloque-Berges, 2017). La récolte systématique de ces données peut alimenter des bases de données qui, grâce à des croisements, peuvent servir à établir des profils d'utilisateurs dans un but de sollicitation commerciale : les logiques d'*ubiquité* et de *mutabilité* s'illustrent ici alors que des données peuvent réapparaître dans des lieux et pour des usages divers. Enfin, les problématiques liées à l'anonymat et au pseudonymat sont bien sûr très présentes dans les corpus d'emails, en lien avec les problématiques de l'*invérifiabilité*. En sus, si les écrits restent, les coordonnées s'envolent… En effet, l'obsolescence numérique touche aussi au renouvellement rapide des moyens de communiquer – des outils certes, mais aussi des identités numériques qui leurs sont associées ; les coordonnées numériques, qui composent en partie ces identités (une adresse email par exemple) sont loin d'être pérennes. Ayant établi un certain nombre de contacts avec des acteurs récurrents dans

mon corpus afin de procéder à une douzaine d'entretiens, la moitié d'entre eux reste injoignable pour des raisons techniques ou personnelles.

31 Les professionnels français du patrimoine doivent travailler selon des cadres réglementaires stricts. Dans le cas des données de la recherche pour la patrimonialisation scientifique et technique, les régimes de propriété des instruments s'appliquent aussi bien au matériel qu'à la documentation et s'adaptent en fonction du régime de protection propre aux collections *in situ*. Parmi les « œuvres scientifiques ou techniques » faisant l'objet d'une protection du droit d'auteur, la compilation d'information est la plus proche de nos objets (Rainette, Cornu & Wallaert, 2008). Cependant, comme le montre l'étude de Lefebvre (2013) sur le patrimoine des écritures ordinaires de la recherche, les écrits intermédiaires ne font pas (encore) l'objet de formalisation juridique en raison de leur nature informelle. Mais si le cadre juridique du patrimoine scientifique et technique s'applique mal à nos objets, il concerne toutefois nombre d'artefacts (matériel, logiciels, bases de données, brevets ...) qui font partie du contexte des correspondances électroniques entre informaticiens. La question juridique se repose au niveau transnational autour de la notion de bien commun informationnel ou numérique qui postule des modalités d'accès ouvert (Dulong de Rosnay & Musiani, 2012). Des institutions garantes du dépôt légal peuvent également archiver les sites Web d'archives de listes et de groupes, comme l'avait déjà fait la BnF avec les services de liste de Renater et du CNRS sur lesquels a été trouvé la moitié des archives du corpus. Cependant, ces documents ont été collectés selon les normes de la politique de préservation du Web français, qui ne fait pas de différence de traitement pour les courriers électroniques mais sauvegarde plutôt la page où ils sont archivés.

32 On notera que les emails possèdent un identifiant universel dans leurs métadonnées. Dans un contexte de forte ubiquité des documents numériques, alors que la distinction entre original et copie devient obsolète, cela permet de donner au message une « numérotation unique » pouvant identifier un objet documentaire précis à la fois dans son contexte d'archivage informatique natif et dans un contexte d'archivage institutionnel ou professionnel ou, encore dans le cas d'une situation de probation juridique (Le Crosnier, 2006).

Partager, c'est standardiser

33 Un corpus patrimonial institutionnel doit être standardisé, c'est-à-dire répondre à des normes définies par une politique locale (relative à un établissement ou une organisation publique), nationale ou internationale du patrimoine. Dans le cas des corpus numériques, cette standardisation intervient dans la construction du corpus, en priorité dans sa description, mais aussi dans sa structuration. C'est la dernière étape de construction du corpus numérique, qui devient alors un fichier où sont codés les documents sous la forme de données.

34 La description se fait au niveau des métadonnées de corpus (que je distingue des métadonnées d'origine, générées lors de l'envoi du courrier) : un ensemble d'informations qui décrit ce qu'est et comment est construit le corpus, généralement placé dans l'en-tête du fichier. Selon les recommandations des humanités numériques, les efforts de standardisation doivent être tendus vers l'interopérabilité (Pouyllau, 2010). Concrètement, cela veut dire qu'une fois publié sur le Web (sous une forme intégrale, ou partielle, limitée à des descripteurs), le corpus doit répondre à la demande des futurs

utilisateurs des données du patrimoine de la recherche : il doit être accessible, dans un format qui lui permet d'être lu, voire d'être à nouveau analysé. Cet accès et cette lisibilité sont conditionnés au fait que des machines peuvent également être « utilisatrices » du corpus. Par exemple, un service d'agrégation de métadonnées des productions scientifiques comme Isidore, devrait pouvoir trouver dans les métadonnées du corpus en question de quoi indexer et rapatrier automatiquement les informations de description dans son espace de recherche. Autre exemple : si le corpus est structuré pour être analysé de manière appareillée (par un logiciel d'analyse), les choix de structuration privilégieront un langage standard qui permette de nouvelles analyses par d'autres logiciels utilisant ce standard. Dans le cas de mon corpus, j'ai fondé mes choix de description sur les recommandations d'un groupe de travail dédié aux corpus de communications numériques à l'intérieur du consortium ECRITS, patronné par l'infrastructure de recherche française HUMANUM. Ce groupe, intitulé CoMeRe (voir note 3), rassemble des linguistes qui offrent des procédures pour le code des échanges écrits électroniques (Chanier et al., 2014). Ils recommandent les standards de l'*Open Language Archive Community* (OLAC), un partenariat international d'institutions cherchant à créer une bibliothèque virtuelle mondiale de ressources linguistiques. OLAC fonctionne également comme un dépôt de métadonnées de corpus : si le corpus est bien décrit selon leurs standards, on peut déclarer le corpus auprès de l'organisation qui ensuite viendra le référencer – cette opération d'interopérabilité valant pour garantie de dépôt auprès d'un tiers de confiance.

35 La structuration se fait au niveau des données. En soi, la structuration d'un corpus n'est pas synonyme de standardisation pour un usage patrimonial : elle intervient dans le cas où l'on souhaite appareiller le corpus, c'est-à-dire le préparer et le coder pour qu'il puisse être traité par des processus d'analyse automatisés. J'ai choisi le langage XML, un standard pour la construction de corpus numériques[16], choix détaillé ci-après. Dans notre cas, la structuration avait pour contrainte la prise en compte du cadre futur de sa diffusion : échantillon issu d'une collecte représentative, ajout des métadonnées pour l'interopérabilité, respect des cadres éthiques et juridiques.

36 Cette standardisation est un cadre idéal permettant de construire le corpus en pensant à sa consultation, voire à sa réutilisation par la communauté académique. Cependant, la bonne réalisation de la description et de la structuration dépend du matériau et des limites qu'il pose à la standardisation. Dans le cas du courrier électronique, au moins deux écueils sont soulevés lors de ces étapes. Tout d'abord, on rencontre avec ce document nativement numérique un objet documentaire déjà structuré en données et métadonnées d'origine, déjà accompagné d'un code et d'un format. La re-documentarisation de ces objets implique donc une nouvelle couche de structuration, pouvant entraîner des confusions entre les niveaux – un point que je développerai dans la dernière partie. Ensuite, pour qu'un corpus soit publiable et partageable « publiquement », du moins selon les standards d'OLAC, il doit être épuré de données personnelles. Or, les courriers électroniques, non contents de faire figurer des noms et coordonnées électroniques dans les métadonnées, en comportent également dans ses données de contenu - et en particulier dans un espace intermédiaire relevant à la fois de contenu et de métadonnées, l'espace de la signature électronique.

37 À visée de test, deux méthodes ont été adoptées pour la structuration. D'une part, à la main, sur de petits échantillons (100 messages environ pour chaque liste ou groupe électronique choisi), afin de bien percevoir les problèmes à chaque étape. D'autre part, de

manière automatisée, afin de travailler sur un plus large échantillon de messages, à savoir l'intégralité d'une liste : la liste DNS (1993-2013, environ 2500 messages)[17]. Une procédure de structuration automatique étant requise pour pouvoir standardiser des corpus de moyenne et grande ampleur, nous considérions que l'expérience aurait été un succès si cette automatisation permettait de résoudre les problèmes de données personnelles et la confusion possible entre deux niveaux métadonnées (d'origine et secondaires, produites par la structuration et la description). Cela n'a pas été le cas : le corpus n'a pu être finalisé dans sa version patrimoniale, c'est-à-dire, selon le cadre normatif institutionnel décrit plus haut, déposable et référençable par un tiers de confiance tel qu'OLAC. Cependant, nous pouvons considérer les problématiques méthodologiques qui se sont dégagées comme un résultat en soi, ce qui fera l'objet de la dernière partie.

Encoder et décoder : problèmes croisés de structuration et de méthodologie

38 Un patrimoine numérique natif doit tenir compte des expériences elles-mêmes natives de la mémoire en contexte numérique. Le projet de développement d'un outil d'archivage Web pour consulter et chercher dans les listes de discussions de l'organisation IETF, proposé par la RFC 6778 en octobre 2012 et réalisé en janvier 2014, donne une idée des besoins, mais aussi des enjeux liés à cette préservation : le volume des données, la temporalité des échanges, la thématisation des discussions. Si les corpus de listes et groupes de discussion ne sont pas assez massifs pour entrer dans la définition du *big data*, ils sont suffisamment riches pour qu'une analyse « à la main » soit parfois insuffisante pour traiter certains de ses aspects (notamment l'évolution dans le temps). Il a fallu choisir un logiciel d'analyse. La suite d'application Calico, développée par des didacticiens [18] a retenu mon attention car, portant sur le genre des échanges médiés par les réseaux de type forum, elle permet d'analyser l'émergence de thématiques et les interactions entre participants dans le temps et dans l'espace de la discussion : analyses de fréquence d'apparition des occurrences, de leur regroupement en clusters thématiques, de leur visualisation diachronique à plusieurs échelles, de la participation des interlocuteurs, et comparaisons systématiques selon différents critères.

39 Dans cette partie finale, je souhaite montrer comment la structuration du corpus permet non pas seulement de le préparer pour opérer des traitements analytiques appareillés, mais surtout pour mieux comprendre et affiner la méthodologie d'analyse elle-même. L'expérience de structuration a été une expérimentation méthodologique réflexive. Je propose l'idée que la méthodologie d'approche de ces objets comporte un aspect paléographique non négligeable, à savoir le déchiffrement, la compréhension et l'interprétation d'écritures – ici liées au format numérique. Les résultats proposés ici relèvent moins de l'interprétation du contenu que de celle des formes de ces écritures. Il nous semble que c'est à ce niveau que cette expérience de corpus soulève des enjeux de méthodologie.

Une méthodologie portée sur les interactions médiées par les réseaux numériques

40 La communication numérique est un objet fait de langages, aussi bien humain qu'informatique. Mon modèle initial est celui des corpus épistolaires aussi bien au niveau

de l'unité textuelle minimale (la lettre/l'email) qu'à celui des relations intertextuelles entre chaque unité (liens d'un email avec les autres emails de la liste ou des groupes). Les perspectives littéraire et linguistique fournissent donc ici un postulat méthodologique : on s'intéressera à la manière dont les messages sont adressés et distribués aux interlocuteurs, ainsi qu'aux cadres participatif et normatif de la correspondance (Leriche, 2006 ; Siess, 2007). Il reste cependant à confronter ces aspects à leur cadre technique, celui du courrier et de la liste électronique. Quelques travaux ont initié cela : si la perspective en sémiotique des langages et des discours est pionnière (Mourlhon-Dallies & Colin, 1995 ; Labbe & Marcoccia, 2005), c'est davantage la perspective en sociologie des sciences et des techniques (STS) qui nous inspire, car elle prend en compte les problématiques de médiation sociotechnique qui me préoccupent (Paravel, 2007). Cependant, aucune approche ne prend en compte l'intégralité de l'objet documentaire : c'est l'aspect essentiel du questionnement méthodologique que je défends ici.

41 La figure 2 (§ 2.1) montre un courrier électronique dans sa forme complète. Il était souhaitable de conserver ces métadonnées d'origine, que je considère comme partie intégrante du matériau à analyser. En effet, dans une étude des communications numériques, prendre en compte la couche infrastructurelle de ces communications est un parti-pris méthodologique. Un email est « décrit » par ses métadonnées d'origine : auteur/expéditeur, destinataire, date, sujet, sont celles qui apparaissent de manière classique dans les logiciels de messagerie. Or, le balisage XML sert à insérer des métadonnées qui renseignent sur les données, des informations qui transforment les données elles-mêmes en informations (pour préparer à l'interopérabilité, à l'analyse automatisée, etc). L'application de l'XML à un courrier électronique est donc à la fois facilitée et compliquée par la structure originelle du document, entraînant des risques de confusion. Dans la structuration XML, un choix hybride a été réalisé : les métadonnées d'origine classiques ont été réutilisées comme des balises de structuration – des métadonnées de re-documentarisation : le champ « expéditeur » supporte ainsi le codage des auteurs sources des messages, le champ « destinataire » celui de leurs récepteurs, le champ « sujet » celui du thème assigné des messages, le champ « date » celui de la temporalité des échanges. Lors d'une analyse logicielle automatique, ainsi, le traitement d'un type de données est donc facilité par la structure d'origine. Les métadonnées moins classiques ont été regroupées et traitées comme des données de message, mais l'on peut imaginer une extension du principe de structuration à partir des métadonnées d'origine. Je donne quelques exemples de ces possibilités et limites dans ce qui suit.

42 L'analyse des interactions se fonde sur les mêmes métadonnées d'origine classiques : qui a pris la parole, qui a répondu, sur quel thème, et à quel moment. Mais par défaut, chaque message est adressé à l'ensemble des abonnés de la liste ; il faut donc aller voir dans le corps des messages pour obtenir une vision plus fine des interactions. Les jeux d'interactions y sont rendus visibles à la fois dans le tissu du discours, mais aussi dans les aspects les plus formels du texte, et déterminent une « poly-énonciation épistolaire » (Leriche, 2006) : jeu des citations et des réponses formalisés grâce à des signes distinctifs – par exemples les chevrons ou les guillemets ajoutés automatiquement dans l'usage de la messagerie. On a précédemment évoqué que le modèle d'analyse des échanges épistolaire était valide pour les courriers électroniques, mais il s'agit de le mettre à jour à l'aune des spécificités de format des textes et paratextes des courriers électroniques. Le texte brut utilise les caractères de l'alphabet ASCII pour recréer des effets de structuration, comme les chevrons de citation, les tirets, les astérisques pour signaler une mise en valeur, etc.[19]

L'encodage des citations et autres signes de ce type reste un défi de la structuration en XML. En effet, ces signes sont extrêmement diversifiés selon le logiciel de messagerie utilisé par l'auteur de l'email (par exemple, pour les citations : « », ‘, ‘’, >, >>, « X a dit : », etc.), et donc un traitement automatique de toutes les occurrences est quasiment impossible. Par ailleurs, certains signes comme les chevrons « > » se mélangent avec la syntaxe XML. En outre, le codage thématique des messages par le biais du champ « sujet » est non seulement tributaire du choix de l'auteur du message, mais relatif à sa reprise, sa réécriture ou son abandon par les interlocuteurs. Structurer les interactions en suivant la reprise exacte d'un sujet dans la réponse est souvent compliqué par ces variations : l'insertion de signes de réponse (« Re : »), de transfert (« Fwd : »), ou la reformulation partielle ou totale du terme définissant le sujet de départ. Cela achève de faire de l'automatisation du codage et de la structuration de ce type de corpus une tâche quasi impossible à finaliser, sauf en reprenant le codage à la main.

Coder la distribution dans le temps et dans l'espace

43 Les métadonnées d'origine sont riches pour analyser les logiques de distribution du courrier. Dans un espace-temps qui est à la fois social et technique, comment les messages sont-ils distribués, c'est-à-dire tissent-ils des relations entre des expéditeurs et des destinataires ? L'espace d'une liste ou d'un groupe de discussion est par défaut celui borné par les adresses électroniques des membres abonnés, mais selon les logiques de distribution, il peut se déformer.

44 Le champ « auteur » ou « expéditeur » est une métadonnée d'origine pouvant servir aisément à renseigner une balise XML et donc automatiser l'analyse des sources des messages. On voit ainsi se profiler des auteurs prolifiques, qui possèdent une autorité technique (ils sont les administrateurs de la liste ou du groupe) et/ou intellectuelle (ils prennent beaucoup la parole). Les données permettant d'identifier et qualifier cette autorité sont elles-mêmes distribuées dans l'espace du message. Aussi bien l'adresse électronique que les signatures renseignent sur l'expéditeur : son nom et prénom, position et coordonnées professionnelles, mais aussi ses goûts et valeurs, si, par exemple, il a inséré dans la signature une citation, un bon mot, ou encore un dessin avec des caractères ASCII. Une analyse automatisée de fréquence d'émission de messages par auteur et dans le temps, comme le permet Calico, permet de faire ressortir des profils quantitatifs que l'on peut ensuite enrichir grâce aux données personnelles et sociologiques recueillies.

45 Les champs « destinataire » et « copie à » servent pareillement de support à une balise XML. On y observe les logiques d'extension de l'espace-temps de la liste ou du groupe, si dans ces champs figurent des destinataires secondaires ou parallèles. Sur Usenet, par exemple, les messages sont souvent distribués dans plusieurs groupes, dans le cas d'annonces, d'appels, et autres processus d'adressage requérant une plus large diffusion. Par ce biais, on peut aussi identifier des expéditeurs qui participent à plusieurs groupes – et jouent, non seulement par la récurrence, mais aussi par la diversité de lieux d'intervention, un rôle de pivot. J'ai été amenée à interviewer de telles « personnes pivots » au rôle très actif dans l'histoire étudiée. Dans le cas des listes, médium de communication plus ciblé sur des groupes de professionnels, on peut facilement identifier les collaborations effectives ou potentielles tissées entre les individus et groupes. Les listes analysées révèlent ainsi un tissu de collaboration académique ayant participé aux

premières standardisations techniques *sur* et *par* Internet. En allant plus loin, les métadonnées de routage d'origine peuvent servir à une analyse relationnelle (quelles sont les relations entre les messages ou entre leurs auteurs) non seulement au niveau des expéditeurs et récepteurs humains, mais au niveau de leurs équivalents machiniques – si l'on s'intéresse à l'infrastructure des réseaux informatiques, dans une approche matérielle des sources nativement numériques (Paloque-Berges, 2016). Ainsi, dans les groupes Usenet, les spécialistes d'Internet regardent et commentent les routes qu'ont pris leurs correspondances pour arriver à destination au tournant des années 1990, alors que la cartographie des réseaux Internet est loin d'être stabilisée et relève du bricolage de la part des premiers fournisseurs d'accès à Internet, qu'ils soient académiques, associatifs, ou entament une logique de commercialisation tournée vers le grand public. Autour de 1993, alors que des programmes politiques et économiques commencent à se mettre en place pour simplifier l'usage des réseaux auprès de publics plus larges, les mêmes acteurs s'amusent de voir que les routes sont toujours aussi complexes (messages émis de France passant par les États-Unis pour revenir en France).

46 Enfin, les données de contenu renseignent sur l'extension de l'espace-temps de la discussion dans le « hors-ligne », important à prendre en compte. Dans mon corpus, l'annonce de conférences, journées, séminaires, et autres événements de la recherche en informatique, ou encore de l'actualité commentée par les participants, non seulement inscrit les interlocutions dans un contexte plus large que celui de la liste ou du groupe, mais montre également les formes d'organisation et de mobilisation qui peuvent précéder, accompagner ou prolonger cette communication. À titre d'exemple, on a pu constater l'importance des « Journées Réseaux » des années 1990 pour structurer une communauté plus large que celle qui se montre dans les listes IP et DNS, et pour accueillir les discussions techniques qui finalisent concrètement la mise en place des réseaux Internet en France.

47 En définitive, j'ai tenté de faire parler non seulement les acteurs (dont le contenu des correspondances devient témoignage) mais aussi les artefacts numériques qui supportent cette parole (matériels, logiciels, formats, protocoles). Si l'on s'autorise un rapprochement sémantique, les sources historiques rencontrent les sources informatiques (les codes de programmation). En ceci, j'ai pu travailler à la transformation de « documents-traces [...], conjonction entre une activité humaine et une technique d'écriture » structurée par des codes, en « documents-sources [...] source d'enseignement » dans un processus de manipulation (Chabin, 2004).

Conclusion

48 Les archives non standardisées du Web héritent de la succession d'états transitoires et informels des cycles documentaires numériques, succession due au mode même de développement d'une documentation « sauvage » sur Internet qui précède les efforts des archivistes professionnels. Les réseaux Internet et le Web ont ainsi pu nourrir le fantasme documentaire de beaucoup (et, de fait, les sources primaires abondent, sous des formes éparses ou dans des fonds d'archives privés), mais ne présentent en fait qu'une illusion d'archive loin des modèles formels requis par le patrimoine institutionnel ou institutionnalisé. L'importance accordée aux aspects pratiques, méthodologiques mais aussi éthiques et juridiques de la collecte révèle la situation complexe des archives en ligne. En ceci, les formes de la communication médiée par les réseaux pré-Web

préfigurent les espaces sociaux du Web, en particulier les lieux d'expression et de communication que sont les blogs, « bric à brac du chercheur », textes fourre-tout et informels mais au potentiel relationnel important pour la communication des sciences (Dacos & Mounier, 2011). Cependant, pour produire une réflexion patrimoniale légitime du point de vue des institutions de préservation de l'héritage et de la mémoire, il faut formaliser, afin de prévoir la sauvegarde certifiée et pérenne. Par le codage du corpus, je me suis donné pour tâches : de structurer les matériaux afin de pouvoir analyser finement les données et métadonnées, en donnant accès à des sous-ensembles catégorisés selon des critères divers (date, auteur, sujet, circulation et interactions du message) ; de standardiser cette structure afin que la lecture soit possible par le biais d'instruments d'analyse (des logiciels de traitement de texte spécialisés dans les interactions langagières numériques en réseau appareillant le corpus). En définitive, les corpus étudiés, s'ils ne sont pas prêts, pour des raisons techniques et légales, à entrer dans une logique de patrimonialisation publique, sont cependant des sources d'enseignement pour le futur du patrimoine numérique natif.

BIBLIOGRAPHIE

Abbate Janet (2000). *Inventing the Internet*, Cambridge, The MIT Press.

Abdelaziz Abid (2004). "Preserving Our Digital Heritage : A UNESCO Perspective", in Tolla Vittoria & Castellani Cecilia (dir.), *The Future of Digital Memory and Cultural Heritage (Conference Proceedings)*, ICCU, Florence, pp. 61-84.

Bergeron Andrée, Blemus Nicolas & Tallec Marie-Pierre (2014). « Les Archives de La CSTI », in Caillet Elisabeth, Guillet Philippe, Guiraudon Jean-Claude, Maitte Bernard, Morand Olivier & Van-Praët Michel (dir.), *Hier pour demain : une mémoire de la culture scientifique, technique et industrielle*, Orléans, L'Harmattan/Muséologies, pp. 199–215.

Bachimont Bruno (2010). *Le sens de la technique : le numérique et le calcul*, Paris, Les Belles Lettres.

Ballé Catherine, Cuenca Catherine & Thoulouze Daniel (2010). *Patrimoine scientifique et technique. Un projet contemporain,* Paris, La Documentation française.

Baym Nancy K. (1995). « The emergence of community in computer-mediated communication », in Jones Steve (dir.), *Cybersociety : Computer-Mediated Communication and Community*, Londres, Thousand Oaks, pp. 138-164.

Chevallier Philippe & Illien Gildas (2011). « Les Archives de l'Internet : une étude prospective sur les représentations et les attentes des utilisateurs potentiels », Enquête réalisée en 2010-2011 [en ligne], consulté 05.07.2016.URL : http://www.bnf.fr/documents/enquete_archives_web.pdf

Brian Éric (2001). « Archives et mémoire des sciences : enjeux historiographiques », *Revue d'histoire moderne et contemporaine*, n° 48-4bis/5, pp. 44-48.

CALLON Michel (1989). *La science et ses réseaux. Genèse et circulation des faits scientifiques*, Paris, La Découverte.

CHABIN Marie-Anne (2004). « Document trace et document source. La technologie numérique change-t-elle la notion de document ? », *Information-Interaction-Intelligence*, n° 1, vol. 4, pp. 141-157.

CHANIER Thierry, POUDAT Céline, SAGOT Benoit, ANTONIADIS Georges, WIGHAM Ciara R., HRIBA Linda, LONGHI Julien & SEDDAH Djamé (2014). « The CoMeRe Corpus for French : Structuring and Annotating Heterogeneous CMC Genres », *JLCL - Journal for Language Technology and Computational Linguistics*, 29 (2), 1–30.

COLLECTIF (2012). « UNESCO/UBC Vancouver Declaration », in Duranti Luciana & Schaffer Elizabeth (dir.), *The Memory of the World in the Digital Age : Digitization and Preservation. An international conference on permanent access to digital documentary heritage*, Vancouver, UNESCO, pp. 1452-1456.

DACOS Marin & MOUNIER Pierre (2011). « Les carnets de recherche en ligne, espace d'une conversation scientifique décentrée », in Jacob Christian (dir.), *Lieux de Savoir (vol. 2). Les mains de l'intellect*, Paris, Albin Michel.

DALLE-NAZEBI Sophie & AGUERA Dimitri (2013). « Logique de production et de partage d'écrits de travail. Pratiques d'informaticiens », *Sciences de la société*, no. 89, pp. 96-111.

DESVALLÉES André (2004). « La muséologie et les catégories de patrimoine immatériel. Questions de terminologie, à propos de patrimoine immatériel et patrimoine intangible », *ISS33 Supplement, Complete Edition of the Papers*, Seoul, ICOFOM Study Series.

DULONG DE ROSNAY Mélanie & MUSIANI Francesca (2012). « The Preservation of Digital Heritage : Epistemological and Legal Reflections », *ESSACHESS – Journal for Communication Studies* 5 (2(10)), pp. 81–94.

FLICHY Patrice (2001). *L'imaginaire d'Internet*, Paris, La Découverte.

GUICHARD Éric (2014). « L'internet et les épistémologies des sciences humaines et sociales », *Revue Sciences/Lettres* [en ligne], 2 | 2014, mis en ligne le 07 octobre 2013, consulté le 07 juillet 2016. URL : http://rsl.revues.org/389 ; DOI : 10.4000/rsl.389

HERT Philippe (1998). *Jeux, écritures, espaces d'énonciations. Contribution à une étude anthropologique de l'usage d'Internet en milieu scientifique*, Thèse de Doctorat en Sciences de l'Information et de la Communication, Université de Strasbourg.

JACOB Christian (dir.) (2011). *Lieux de Savoir. Les Mains de L'intellect*, Paris, Albin Michel.

LABBE Hélène & MARCOCCIA Michel (2005). « Communication numérique et continuité des genres : l'exemple du courrier électronique », in *Texto !* [en ligne], consulté 05.07.2016. URL : http://www.revue-texto.net/Inedits/Labbe-Marcoccia.html

LATOUR Bruno & WOOLGAR Steve (2005). *La vie de laboratoire : La production des faits scientifiques*, Paris, La Découverte.

LATZKO-TOTH Guillaume & PROULX Serge (2013). « Enjeux éthiques de la recherche en ligne », in Barats Christine (dir.), *Manuel d'analyse du Web en sciences humaines et sociales*, Paris, Armand Colin.

LE CROSNIER Hervé (2006). « Architecture informatique et économie du document numérique : deux questions étroitement liées », in Chartron Ghislaine & Broudoux Evelyne (dir.), *Document numérique et société*, Paris, ADBS, pp. 29-41.

LEFEBVRE Muriel (2013). « L'infra-ordinaire de la recherche. Écritures scientifiques personnelles, archives et mémoire de la recherche », *Sciences de la société*, no. 89, pp. 3-17.

LERICHE Françoise (2006). « Quel balisage pour les corpus épistolaires numériques ? De l'annotation traditionnelle du 'Document' à une analyse générique et pragmaticienne », in Rastier François & Ballabriga Michel (dir.), *Corpus en lettres et sciences sociales : des documents numériques à l'interprétation*, Paris, Texto, pp. 262-270.

MARCOCCIA Michel (2001). « L'animation d'un espace numérique de discussion : l'exemple des forums Usenet », *Document numérique*, 2001/3, vol. 5, pp. 11-26.

MARKHAM Annette N. & BAYM Nancy K. (2009). *Internet Inquiry. Conversations About Method*, Thousand Oaks, Sage publications.

MOUNIER Pierre (dir.) (2012). *Read/Write Book 2 : Une Introduction Aux Humanités Numériques*, Marseille, OpenEdition Press.

MOURLHON-DALLIES Florence & COLIN Jean-Yves (1995). « Les rituels énonciatifs des réseaux informatiques entre scientifiques » *Les Carnets du Cediscor. Publication du Centre de recherches sur la didacticité des discours ordinaires*, no. 3, pp. 161-172.

PALOQUE-BERGES Camille (2014). « Le rôle des communautés patrimoniales d'Internet dans la constitution d'un patrimoine numérique : des mobilisations diverses autour de l'auto-médiation », in Saou-Dufrêne Bernadette (dir.), *Heritage and Digital Humanities*, Berlin, Lit Verlag, pp. 277-290.

PALOQUE-BERGES Camille (2015). « L'imaginaire du « grand public » au tournant du Web (1993-1997) », *Revue française des sciences de l'information et de la communication* [En ligne], 7 | 2015, mis en ligne le 05 octobre 2015, consulté le 07 juillet 2016. URL : http://rfsic.revues.org/1478

PALOQUE-BERGES Camille (2016). « Les sources nativement numériques pour les sciences humaines et sociales ». *Histoire@Politique*, mai-août 2016. n° 29 [en ligne], consulté 05.07.2016. URL : https://halshs.archives-ouvertes.fr/halshs-01239053/document

PALOQUE-BERGES Camille (2017). « Usenet as a Web Archive : Multi-layered Archives of Computer-Mediated-Communication », in Brügger Niels (dir.), *Web 25 : Histories from the First 25 Years of the World Wide Web*, Peter Lang (en cours de publication).

PALOQUE-BERGES Camille & KEMBELLEC Gérald (2014). « Nouvelles sources numériques et logiques d'open corpus : l'intérêt d'archiver et de partager des courriers électroniques », *Cahiers de la Sfsic*, n° 9, pp. 239-244.

PALOQUE-BERGES Camille & SCHAFER Valérie (2015). « Quand la communication devient patrimoine », *Hermès*, no. 71, pp. 157-63.

PASTINELLI Madeleine (2009). « La mémoire et l'oubli dans l'univers de l'archive totale », *Revue électronique des sciences humaines et sociales* [en ligne]. URL: http://ecole-ident-num.sciencesconf.org/conference/ecole-ident-num/pages/Pastinelli_espace_temps_2009.pdf

PARAVEL Véréna (2007). « De la plume d'oie à la souris : la recherche en réseaux », in Jacob Christian (dir.), *Lieux de Savoir (vol. 1). Espaces et Communautés*, Paris, Albin Michel, pp. 1095-1118.

PÉDAUQUE Roger T. (dir.) (2006). *Le document à la lumière du numérique*, Caen, C&F Editions.

POUYLLAU Stéphane (2010). « Construire le Web de données pour les sciences humaines et sociales. Fonds documentaires scientifiques de la recherche en SHS », Note informationnelle du Centre national pour la numérisation de sources visuelles/ CN2SV-TGE Adonis.

RAINETTE Caroline, CORNU Marie & WALLAERT Catherine (2008). *Guide juridique à l'usage des professionnels du patrimoine scientifique et technique*, Paris, L'Harmattan.

RUZÉ Emmanuel (2009). « Traiter les archives de la Toile. Une histoire d'un système d'information dans une communauté, WordPress (2003-2008) », *Entreprises et Histoire,* vol. 55, pp. 74-89.

SIESS Jurgen (2007). « "Les missives sont écrites pour inventer le réel" : l'épistolaire dans la perspective de l'analyse du discours », *Filol. Lingüist. Port.*, n° 9, pp. 369-386.

SCHAFER Valérie (2012). « Internet, un objet patrimonial et muséographique », in *Actes du colloque MINF*, Cnam, Paris [en ligne], consulté 05.07.2016. URL : http://minf.cnam.fr/Papiers-Verifies/7.3_internet_objet_patrimonial_Schafer.pdf

SCHAFER Valérie & THIERRY Benjamin (2013). « Internet ou la fin de la communauté scientifique idéale. Pour une approche historique de la science en réseaux », Communication au colloque international « Formes et enjeux contemporains de la communication et de la culture scientifiques et techniques » (Institut de la Communication et des Médias), Grenoble.

SCHAFER Valérie, PALOQUE-BERGES Camille & GEORGES Fanny (2015). « La culture Internet au risque du Web », *Cahiers du CIRCAV* 24, pp. 15-30.

VANHOUTTE Edward, NYHAN Julianne & TERRAS Melissa (2013). *Defining Digital Humanities : A Reader,* Farnham, Ashgate Publishing.

WELLMAN Barry (2001). « Computer Networks As Social Networks », *Science*, 293 (5537), pp. 2031-2034.

ZACKLAD Manuel (2007). « Réseaux et communautés d'imaginaire documédiatisées », in Skare Roswita, Lund, Niels Windfeld et Varheim Andreas (dir.), *A Document (Re)turn*, Frankfurt am Main, Peter Lang, pp. 279-297.

Webographie

Tous les sites Web ont été visités pour la dernière fois en février 2016.

ASSOCIATION OF INTERNET RESEARCHERS (AoIR), « Ethics Guide » [http://ethics.aoir.org/].

BORTZMEYER, Stéphane (2012). « RFC 6778 : IETF Email List Archiving, Web-based Browsing and Search Tool Requirements », billet publié en octobre 2012 et mis à jour en janvier 2014 [http://www.bortzmeyer.org/6778.html].

ERTÉ CALICO, Suite logicielle Calico [http://woops.crashdump.net/calicorss2/index.php].

GROUPE « COMERE » (Communication Médiée par les Réseaux) du consortium Ecrits de l'IR CORPUS/HumaNum [https://corpuscomere.wordpress.com/].

IETF, « RFC Archive » [http://www.rfc-archive.org/].

IETF, « IETF Mail Archive » [https://mailarchive.ietf.org/arch/].

IETF, « RFC 6778 : IETF Email List Archiving, Web-based Browsing and Search Tool » (Auteur : R. Sparks) [https://tools.ietf.org/html/rfc6778].

ISOC, « Internet Hall of Fame » [http://www.internethalloffame.org/].

RIPE, « Inactive Lists » [https://www.ripe.net/participate/mail/inactive-lists].

NOTES

1. Ce travail a été initié à l'occasion d'un contrat post-doctoral du Labex HASTEC (2012-2013). Cet article se concentrant sur le traitement documentaire associé au travail, les méthodologies et résultats historiques ne seront convoqués qu'à titre illustratif, faisant l'objet de publications distinctes. Je remercie ici chaleureusement Gérald Kembellec pour sa contribution aux aspects techniques et documentaires de la construction numérique du corpus.

2. Les archives privées (produites par des individus, associations ou entreprises) se distinguent des archives publiques (produites dans un cadre légal par des administrations publiques). Dans notre cas, la première moitié du corpus est issue d'une préservation par une administration d'archives privées de listes de discussion de personnels de l'enseignement supérieur et de la recherche (service d'archives de listes de Renater [https://groupes.renater.fr/sympa] et de la DSI du CNRS [https://listes.services.cnrs.fr/wws], deux entités d'administration des outils numériques pour le milieu académique). La deuxième est issue d'une préservation par une organisation d'archives privées d'échanges électroniques de la branche francophone du service de forum Usenet (les archives Usenet des forums Google Groups [https://support.google.com/groups/answer/6003482]). Les deux fonds sont accessibles publiquement sur le Web.

3. La « *Computer Mediated Communication* » (CMC) est un objet de recherche associé à l'analyse des communications sur réseaux informatiques, défini à la croisée des recherches en informatique, en sciences du langage, et reprise dans le domaine interdisciplinaire des *Internet Studies* (Baym, 1995 ; Wellman, 2001). L'expression française « communication médiée par les réseaux » est proposée par les linguistes français spécialistes des corpus de CMC (groupe CoMeRe du consortium Écrits de l'IR HumaNum, anciennement appelée IR Corpus).

4. Je reprends le titre d'un colloque de 2016 en humanités numériques de l'École normale supérieure de Lyon dont l'argumentaire présente le même type d'enjeux, mais, comme souvent, sans proposer de travaux sur les matériaux nativement numériques http://www.ciera.fr/ciera/les-corpus-de-textes-numeriques.

5. Le terme de « lieux de mémoire » fait référence à son usage par l'historien Pierre Nora dans sa description du patrimoine monumental de la mémoire nationale (1997). Il s'agit ici de montrer que cette acception des lieux de mémoire a bien entendu évolué et qu'il est temps de prendre en compte le rôle des réseaux numériques dans cette évolution.

6. À l'exception notable du patrimoine industriel, qui peut être révélateur de crises.

7. 1983 est la date de l'implémentation officielle du protocole définissant Internet, TCP/IP, sur son réseau prototype, ARPANET, le raccordant ainsi de manière standardisée à d'autres réseaux. En France, c'est la date de la première connexion au réseau UUCP/Usenet, qui trois ans plus tard sera redéfini pour être compatible avec les standards d'Internet. Les acteurs de la mise en place du réseau UUCP/Usenet appartiennent au même collectif qui implémentera TCP/IP sur les réseaux en France. En 1993 les technologies du Web, système greffé sur TCP/IP, entrent dans le domaine public, entraînant une popularisation des usages d'Internet.

8. Lors de mes repérages de corpus en 2012, c'est le service Google Forums déjà évoqué qui détenait la plus grande archive de Usenet. Depuis, la fondation Internet Archive a considérablement accru ses archives de Usenet. Les deux fonds posent néanmoins toujours de nombreux problèmes d'accessibilité et d'exploitation (Paloque-Berges, à paraître en 2017), dont certains sont décrits plus bas dans le cas de l'archive Google.

9. FNET fonctionne entre 1983 et 1992 comme le réseau associatif non-officiel du monde académique français (essentiellement celui des sciences informatiques). Le récit de son développement constitue un de nos résultats historiographiques principaux en cours de publication.

10. Dans le cas des Google Groups et de leurs archives Usenet, la portée patrimoniale du service a été et continue à être vivement critiquée par les communautés parties prenantes (Paloque-Berges, à paraître en 2017).
11. Les fichiers en texte brut, ou *plain text*, sont encodés dans le format *American Standard Code for Information Interchange* (ASCII), depuis 1963. C'est un format par défaut, compatible et donc affichable sur tous les ordinateurs, résultat des travaux de l'Organisation internationale de normalisation (*International Organization for Standardization* ou ISO) depuis 1960.
12. Les signes diacritiques sont ceux qui modifient une lettre de l'alphabet par un signe supplémentaire. Ils sont très présents dans la langue française sous la forme, par exemple, des accents, qui ne sont pas reconnus dans l'ASCII.
13. Ou, pour reprendre le vocabulaire de la sémiotique des médias informatiques, son architexte (Jeanneret, Le Marec et Souchier, 2003).
14. Deux versions du premier message de « Fnet.general » conservé et publié sur les archives Usenet de Google [https://groups.google.com/forum/#!topic/fnet.general/s304HJZBdbM/discussion].
15. Pour les recherches sur Internet, c'est la charte de l'*Association of Internet Researchers* (AoIR), pionnière en matière de réflexions et propositions épistémologiques et méthodologiques sur les recherches portant sur Internet, qui fait autorité depuis une quinzaine d'années auprès de la communauté internationale interdisciplinaire des chercheurs se rassemblant annuellement lors de ses congrès (Markham & Baym, 2009).
16. L'XML (*eXtensible Mark up Language*) est connu pour sa facilité de prise en main, son adaptation (*extensible*) à différents types de structuration (ou balisage, en anglais *mark up*), mais aussi sa compatibilité avec de nombreux logiciels d'analyse de données et avec des standards d'interopérabilité. Il est préconisé, entre autres, par l'OLAC ainsi que l'initiative OAI PMH (*Open Archives Initiative Protocol for Metadata Harvesting*).
17. Le cahier des charges et la réalisation technique des aspects de programmation et d'automatisation de l'encodage ont été réalisés en co-pilotage avec Gérald Kembellec, avec l'aide de Claire Scopsi et d'un groupe d'étudiants du Master Professionnel Sciences humaines et sociales, Mention Information et Communication, Spécialité Documents Electroniques et Flux d'Informations (DEFI), Université Paris Ouest-Nanterre (Paloque-Berges & Kembellec, 2014).
18. Dans le cadre de l'ERTé CALICO (Communautés d'apprentissage en ligne, instrumentation, collaboration). L'un des membres du projet, François Blondel (STEF, ENS Cachan) m'a particulièrement accompagnée dans mes tests et je l'en remercie ici.
19. La structuration est intégrée avec la généralisation du texte enrichi dans les messageries électroniques à partir des années 1990 et y amène la richesse des logiciels de traitement de texte. À ce jour, la plupart des messageries laissent le choix à l'utilisateur d'écrire en texte brut (plus léger, davantage lisible à travers différents standards de systèmes de lecture) ou en texte enrichi.

RÉSUMÉS

À l'occasion d'une étude historique sur les débuts de l'Internet en France (1983-1993), cet article revient sur les problématiques associées à la reconnaissance et à la construction d'un patrimoine de la communication entre acteurs des sciences et techniques des réseaux numériques, à partir d'archives nativement numériques. La notion de lieux de mémoire réticulaires est analysée à

travers deux conceptions du patrimoine : l'une communautaire, l'autre institutionnelle. Je présente d'abord l'émergence de la valorisation pratique, dans et par les communautés parties prenantes, d'un patrimoine de la communication médiée par ordinateur. Ensuite, je reviens de manière réflexive sur la construction d'un corpus rassemblant des archives de listes et groupes de discussion, dans le but d'expérimenter l'institutionnalisation de ce patrimoine. Ce corpus est encadré par les recommandations normatives des humanités numériques à but de manipuler les données de la recherche en ligne, ce qui me donne l'occasion de mettre en perspective les contraintes éthiques et documentaires avec les problématiques méthodologiques associées à cette expérience.

Within the frame of a historical study on the beginnings of the Internet in France (1983-1993), this paper sheds light on the issues raised by acknowledging and building heritage for the communication of digital networks early participants, based on born-digital archives. The idea of reticular memory space is analyzed through two conceptions of heritage : community and institution. First, I present the emergence of a community-based promotion of computer-mediated communication. Then, I reflect on the building of a corpus gathering list and group discussions archives for institutional heritage purposes. This corpus is framed with the normative recommendations of digital humanities for manipulating heritage research data online, with a series of documentary and ethical constraints that are put in perspective with methodological issues.

INDEX

Mots-clés : archives nativement numériques, corpus, humanités numériques, histoire d'Internet, réseaux numériques, documentarisation
Keywords : born-digital archives, corpus, digital humanities, Internet history, digital networks, documentarization

AUTEUR

CAMILLE PALOQUE-BERGES

Cnam, HT2S