

Où va la philologie numérique ?

Jean-Baptiste Camps

► **To cite this version:**

Jean-Baptiste Camps. Où va la philologie numérique ?. à paraître en 2018 dans Fabula-LHT, 20, Le Moyen Âge pour laboratoire, dir. Florent Coste et Aman.. 2017. <halshs-01674953>

HAL Id: halshs-01674953

<https://halshs.archives-ouvertes.fr/halshs-01674953>

Submitted on 3 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OÙ VA LA PHILOLOGIE NUMÉRIQUE ?

Jean-Baptiste Camps

Jean-Baptiste.Camps@enc-sorbonne.fr

École nationale des chartes – Centre Jean-Mabillon
Paris Sciences & Lettres

Abstract

If the first philologists to use computers, since the 1950's, considered it mostly an instrument to which they could delegate repetitive or quantitatively important tasks, I argue here that computational methods, or, better said, data-oriented philology can be a powerful vector of transformation of our disciplines, inside the framework provided by the emerging fourth paradigm of a unified scientific method. Indeed, if the digital publication of editions often has, until now, captured the attention of textual scholars, the core of the potential transformation lies instead in a data-centric approach: upstream, this means using computational methods to produce data in quantities or granularities that could not be considered until now; downstream, it has the ability of reversing the very way in which we formulate statements or build knowledge, by moving from the traditional hypothesis-led approach to a procedure in which a formalisation is drawn from the data themselves, allowing us to be freer from the more deeply-rooted postulates or from "common-sense" assumptions. As we register new sets of facts selected from our sources, new enquiries and analyses become possible, as long as we don't stay prisoners of interfaces, that are in essence limited and whose sustainability is questionable. If "Data is the Important Long-Term Outcome", they should be shared without restrictions in order to allow for accumulative progress of knowledge, as well as to conform to the standards of scientific research: reproducibility, falsifiability. For this kind of open science to be enacted, we should be less indulgent towards copyfraud and absurd claims of ownership on ancient texts.

Keywords : philology, digital scholarly editing, humanities computing, data science

Résumé

Si, pour les premiers philologues à l'utiliser dès les années 1950, l'ordinateur était avant tout un outil auquel déléguer des tâches répétitives ou très vastes d'un point de vue quantitatif, je défends dans cet article l'idée que les méthodes computationnelles, ou, pour mieux dire, la philologie tournée vers les données peut être un vecteur puissant de transformation pour nos disciplines, au sein du cadre fourni par le quatrième paradigme, émergent, d'une méthode scientifique unifiée. En effet, si la publication électronique d'éditions de texte a souvent, jusqu'à présent, retenu l'attention des chercheurs en sciences du texte, le cœur des transformations potentielles réside plutôt dans une approche qui mette les données au centre : en amont, cela signifie tirer profit des méthodes computationnelles pour produire des données dans des quantités ou granularités jusque là inenvisageables ; en aval, cette approche est à même de renverser la manière même dont nous formulons nos énoncés ou construisons la connaissance, en passant d'une approche fondée sur des hypothèses préexistantes à un processus dans lequel on fasse émerger une formalisation à partir des données elles-mêmes, nous émancipant ainsi un peu plus de nos postulats les plus ancrés ou des présupposés du sens commun. Tandis que nous enregistrons de nouveaux ensembles de faits sélectionnés dans nos sources, de nouvelles enquêtes ou analyses deviennent possible, pour autant que nous ne restions pas prisonniers des interfaces, limitées par nature et dont la pérennité est douteuse. Si « les données sont le résultat important sur le long terme », elles devraient être partagées sans restrictions, tant pour permettre le progrès cumulatif des connaissances, que pour se conformer aux exigences de la recherche scientifique : reproductibilité, réfutabilité. Pour que ce type de science ouverte soit réalisée, nous devrions être moins indulgents envers les revendications abusives de droits patrimoniaux ou de propriété des textes anciens.

Mots-clés : philologie, édition électronique, humanités computationnelles, science des données

Les philologues, notamment médiévistes, et parmi eux des jésuites ou bénédictins, sont assez largement aux origines de ce que l'on est venu plus tard à nommer les humanités numériques¹. Quoiqu'ils aient réalisé parmi les premiers les possibilités fournies à leur science par l'informatique naissante, préfigurées par des tentatives de mathématisation de la critique textuelle dès l'entre-deux-guerres², ces pionniers considéraient l'ordinateur avant tout comme un outil susceptible d'automatiser des tâches trop fastidieuses ou irréalisables sans son soutien, sans que cette approche pratique ne remette nécessairement en cause les implicites théoriques et les conceptions du texte et de l'auteur qui sous-tendent la démarche philologique³. Des années 1960, où l'emploi de l'ordinateur par les philologues est jugé « courant » par Dom Froger⁴, il faudra attendre les travaux de Bernard Cerquiglini pour que les développements de l'informatique soient vus comme à même de remettre en cause les idéaux de fixité du texte et d'autorité, issus de la civilisation de l'imprimé, et lui restituer sa « variance » ou « mouvance » médiévale, par la multiplication des textes, des lectures, la fluidité ou la non-linéarité de ses accès, grâce à l'« ordinateur [...] en avance sur l'imagination philologique et linguistique »⁵.

On ne peut certes nier le rôle joué par ce medium dans l'émergence, au sein de la critique textuelle, d'approches ecdotiques tentant une synthèse entre respect d'un témoin ou respect de l'intention de l'auteur⁶, ou dans le développement de nouveaux types d'édition, qui remettent fondamentalement en cause la traditionnelle (et assez largement implicite) sélection d'information telle qu'elle était présentée par les éditions critiques imprimées, en nous amenant à expliciter et modéliser les choix que nous opérons dans l'ensemble des faits observables dans nos sources⁷. Pour ma part, l'objet de cet article ne sera pas de revenir longuement sur ces questions, mais d'examiner les transformations de la « vie de laboratoire » du philologue, autrement dit les transformations susceptibles de toucher la manière même dont nous formulons des énoncés ou établissons des connaissances sur nos objets d'étude, et d'envisager la place de la philologie dans la révolution scientifique que nous vivons depuis les années 2000, comme d'interroger la possibilité qu'elle participe d'une unité retrouvée de la méthode scientifique, rangée derrière un nouveau paradigme.

Cela ne signifie pas pour autant que ces transformations n'influencent pas également notre rapport aux textes médiévaux ou nos procédés de lecture, ni le statut de l'éditeur critique, comme on le verra.

-
- 1 Les travaux de Dom Roberto Busa sur l'*Index thomisticum* sont souvent considérés comme étant aux origines de la linguistique computationnelle et des « *Humanities computing* », ayant existé à l'état de projet depuis les années 1940 (pour un historique du projet par son auteur, R.A. Busa, « Foreword: Perspectives on the digital humanities », *A companion to digital humanities* (2004). URL : [http://www.digitalhumanities.org/companion/..](http://www.digitalhumanities.org/companion/) Consulté le 24 août 2017.). Il n'était néanmoins pas le seul, car, comme le rappelle Dom Froger, « Depuis vingt ans [en 1968], exactement depuis qu'en 1946 A.D. Booth entreprit de construire une traductrice automatique, les philologues utilisent de plus en plus fréquemment les ordinateurs » (J. Froger, *La critique des textes et son automatisation*, Paris, 1968., « Préface »).
 - 2 Dans le champ de la critique textuelle, ce sont bien évidemment les travaux de Dom Quentin qui sont les plus fameux pour cette période (H. Quentin, *Essais de critique textuelle (ecdotique)*, Paris, 1926.).
 - 3 Comme le relève Dom Froger, sur 120 projets mettant, en 1968, l'ordinateur au service de la philologie, « la plupart (...) sont, en dernière analyse, des *index verborum*, c'est-à-dire de simples mises en fiches de tous les mots d'un texte, travail matériel que les ordinateurs exécutent quasiment à la perfection et qui ne demande qu'une préparation manuelle assez rapide et facile », ajoutant qu'« on peut considérer l'ordinateur comme l'instrument par excellence de tous les travaux qui relèvent plus ou moins directement de la lexicographie » ; J. Froger, *La critique des textes et son automatisation*, *op. cit.* (« préface »).
 - 4 « Il est maintenant courant d'« automatiser » les recherches philologiques ou littéraires, et d'utiliser la machine électronique ou mécanographique pour des études lexicographiques ou stylistiques ; ces procédés ont déjà été employés pour aider à la critique conjecturale, et par exemple combler les lacunes des manuscrits de la mer Morte » ; *Ibid.*, « Avant-propos ».
 - 5 B. Cerquiglini, « Éloge de la variante », *Langages*, vol. 17, n° 69 (1983), p. 35.
 - 6 On songera notamment au concept d'édition « orientée vers la tradition » (G. Palumbo et P. Rinoldi, « Prolégomènes à l'édition du corpus français de la 'Chanson d'Aspremont' », dans *Epic Connections / Rencontres épiques Proceedings of the Nineteenth International Conference of the Société Rencesvals, Oxford, 13–17 August 2012*, M.J. Ailes, P.E. Bennett, A.E. Cobby (éd.), Édimbourg, [s.n.], 2015.), ou de « quatrième voie » (N. Altschul, « The Genealogy of Scribal Versions: A 'Fourth Way' in Medieval Editorial Theory », *Textual Cultures: Texts, Contexts, Interpretation*, vol. 1, n° 2 (2006).), unissant méthode critique et intérêt pour les versions scribales, pouvant aller jusqu'à l'édition du texte de tous les témoins et la reconstitution du texte de tous les intermédiaires perdus ou archétypes ; pour une brève synthèse, on pourra consulter J.-B. Camps, « Copie, authenticité, originalité dans la philologie et son histoire », *Questes: revue pluridisciplinaire d'études médiévales*, vol. 29 (janvier 2015).
 - 7 Sur ce point, on pourra se reporter à C.M. Sperberg-McQueen, « How to teach your edition how to swim », *Literary and Linguistic Computing*, vol. 24, n° 1 (2009)., ou à la synthèse que je propose, concernant plus spécifiquement les textes français, dans J.-B. Camps, *La 'Chanson d'Otinél': édition complète du corpus manuscrit et prolégomènes à l'édition critique*, thèse de doctorat, dir. Dominique Boutet, Paris-Sorbonne, Paris, 3 décembre 2016. URL : <http://www.theses.fr/2016PA040173>., sect. 2.1.1, « Qu'est-ce qu'une transcription ? », p. cxcv-cc.

1 De l'édition électronique à une philologie des données ?

1.1 Édition électronique : une définition

Qu'est-ce qu'une « édition électronique » ? L'heure est passée où l'on pouvait faire de l'existence sous forme numérique d'un objet un critère essentiel de sa définition. Aujourd'hui les éditions, même les plus traditionnelles, vivent nativement dans un format numérique. L'édition « papier », ou, pour mieux dire, imprimée, n'est guère plus qu'un dérivé du fichier dans lequel l'éditeur a réalisé son travail philologique d'établissement du texte. Le médium de diffusion n'est pas non plus suffisant : la simple mise en ligne d'un fichier contenant la numérisation des pages d'une édition imprimée (ou pensée en tant que telle), comme on les trouve, innombrables, dans les bibliothèques virtuelles, ne saurait constituer une édition électronique de plein exercice.

Pour surmonter cette difficulté de définition, Patrick Sahle propose de différencier édition « numérisée » et « numérique ». Fort efficace, sa définition repose sur les potentialités propres au numérique : est numérique une édition qui ne peut être retranscrite sous forme imprimée sans perte d'information ou de fonctionnalité.

Digital scholarly editions are not just scholarly editions in digital media. I distinguish between digital and digitized. A digitized print edition is not a « digital edition » in the strict sense used here. A digital edition can not be printed without a loss of information and/or functionality. The digital edition is guided by a different paradigm. If the paradigm of an edition is limited to the two-dimensional space of the « page » and to typographic means of information representation, than it's not a digital edition⁸

Si l'on souhaite envisager la philologie numérique dans sa globalité, il importe néanmoins de s'abstraire un instant de la question de l'édition électronique, qui concentre depuis quelques décennies l'attention, en tant que publication résultant d'un travail philologique, qui peut, lui, demeurer tout à fait traditionnel, pour envisager les méthodes computationnelles comme participant de l'établissement du texte et, plus généralement, des modes de production de nouveaux savoirs. Un travail philologique tout à fait traditionnel peut déboucher sur une édition électronique, de la même manière qu'un travail de philologie numérique peut avoir comme aboutissement une publication imprimée, comme le rappelle Tara Andrews⁹.

La philologie numérique peut alors être définie comme une transformation dans les méthodes d'établissement du texte et d'analyse, par l'intégration d'outils computationnels, dès que ceux-ci peuvent ap-

8 P. Sahle, *Catalog of: Digital Scholarly Editions*, Cologne, 2008. URL : <http://www.digitale-edition.de/>. « Les éditions savantes numériques ne sont pas simplement des éditions savantes dans un médium numérique. Je distingue les éditions numériques et numérisées. Une édition numérisée n'est pas une « édition numérique » dans le sens restreint utilisé ici. Une édition numérique ne peut être imprimée sans perte d'information et/ou de fonctionnalité. L'édition numérique est guidée par un paradigme différent. Si le paradigme d'une édition est limité à l'espace bidimensionnel de la « page » et à une représentation de l'information par des moyens typographiques, alors ce n'est pas une édition numérique » (je traduis). Pour un stade ultérieur de la réflexion de cet auteur, on se reportera à P. Sahle, *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Norderstedt, 2013., tome 2, « Befunde, Theorie und Methodik », part. sect. 2 « Zur Definition und Begriffsbildung « Digitale Edition », p. 125-155 ; il y exprime notamment l'idée que l'édition imprimée peut être vue comme un produit dérivé, une vue unique parmi un large panel, d'une édition numérique. On consultera également la discussion des définitions de P. Sahle par E. Pierazzo, *Digital scholarly editing: theories, models and methods*, Farnham, 2015. URL : dx.doi.org/10.4324/9781315577227. (notamment l'« Introduction »).

9 T. Andrews, « The third way: philology and critical edition for a digital age », *Variants: the Journal of the European Society for Textual Scholarship*, vol. 10 (2012), p. 2. URL : <http://boris.unibe.ch/43071/>, « Part of the difficulty in defining what a digital edition might be is that the term refers simultaneously to two things. There is the eventual digital publication of a text edition, which may have been prepared using state-of-the-art digital tools for the purpose, specialist software (...), in a spreadsheet, in a word processor, or even on paper and then transcribed into electronic form. Alternatively, there is what we might call “digital philology”, an approach to textual editing that welcomes the aid of technology wherever possible and which will usually, but not necessarily, result in a digital publication » ; « Une partie de la difficulté à définir ce que peut être une édition numérique est que le terme se réfère simultanément à deux choses. Il y a l'éventuelle publication numérique d'une édition de texte, qui peut avoir été préparée en utilisant des outils numériques de pointe pour ces usages, un logiciel spécialisé (...) un tableur, un traitement de texte, ou même faite sur papier et ensuite transcrite sous forme numérique. À l'inverse, il y a ce que l'on pourrait nommer la « philologie numérique », une approche de l'édition de texte qui accueille volontiers l'aide de la technologie dès que cela est possible, et dont le résultat sera en général, mais pas obligatoirement, une édition numérique » (je traduis). E. Pierazzo, *Digital scholarly editing, op. cit.*, sect. « 1.2 Digital Editing, Digital Editions », intègre également à sa définition de l'édition numérique le critère méthodologique : « when not only the working methods are digital, but also the outcome (or one of the outcomes) of such a work, then we can normally talk of a digital edition », « quand non seulement les méthodes, mais aussi le résultat (ou un des résultats) d'un travail de ce type sont numériques, alors on peut normalement parler d'une édition numérique » (je traduis).

porter des gains dans le processus d'établissement du texte (gains de temps, de finesse, de granularité dans la transcription, la collation, ...), sa représentation (enrichissement par la représentation de phénomènes graphiques, linguistiques, sémantiques,... et enregistrement des opérations éditoriales dans l'édition même) ou son analyse scientifique. Le contraste avec la philologie traditionnelle en sort renforcé :

The difference between the traditional approach to philology, whether "old" or "new", and the digital approach lies in their respective willingness to divide labour between human and artificial intelligence; where the former tends to be reluctant to embrace digital possibilities, the latter favours a more efficient division of labour and encourages the production of new methods of presenting texts. The method of production, rather than the published form that the resulting editions take, is the practice wherein lies most of the promised revolution within textual scholarship, but it has attracted considerably less attention than the question of digital publication¹⁰

Ce glissement du regard, du numérique comme medium de diffusion et de lecture d'une philologie traditionnelle, au computationnel comme vecteur d'une transformation méthodologique profonde, mène naturellement vers une philologie qui place en son centre les données (et leurs modèles), par la redéfinition de leur amont (production), comme de leur aval (analyse).

1.2 Vers une philologie intensive en données

La distinction entre humanités « numériques » et « computationnelles » est dans l'air. Au-delà d'un pur choix terminologique distinctif ou d'un retour aux *humanities computing* du XX^e siècle¹¹, la revendication d'une dimension computationnelle rend compte d'un basculement, à mon sens éminemment souhaitable, d'une perspective tournée vers la diffusion et la publication électronique, à une emphase mise sur les données et leur exploitation pour la création de nouveaux savoirs scientifiques. Ce basculement s'intègre pleinement dans le tournant de la science des données (*data science*, ou *data-driven*, *data-intensive*), qui marque actuellement les champs de la connaissance, à des rythmes et des degrés divers, et va de pair avec l'explosion des quantités de données disponibles et exploitables (le « déluge de données »), ainsi que le renouveau de l'intelligence artificielle ou l'omniprésence des méthodes d'analyse statistique et de visualisation. En ce sens, il serait d'ailleurs plus juste de parler de *Data-driven Humanities*¹².

Ce changement terminologique, de manière finalement assez similaire à celui des *Humanities computing* aux *Digital humanities*, se veut aussi révélateur d'un changement dans la considération du rôle du numérique : comme le nouvel outil transforme progressivement l'art de celui qui le manipule¹³, le computa-

10 T. Andrews, « The third way », *op. cit.*, « La différence entre l'approche traditionnelle de la philologie, qu'elle soit « vieille » ou « nouvelle », et l'approche numérique réside dans leurs consentements respectifs à la répartition du travail entre l'intelligence humaine et artificielle ; tandis que l'une tend à être réticente à s'emparer des possibilités fournies par le numérique, l'autre favorise une répartition efficace du travail et encourage la conception de nouvelles méthodes de présentation des textes. La méthode de production, plutôt que la forme publiée des éditions qui en résultent, est la pratique dans laquelle réside l'essentiel de la révolution promise des sciences du texte, mais elle a attiré considérablement moins d'attention que la question de la publication numérique » (je traduis).

11 Les expressions de « computing in the humanities » ou « humanities computing » ont prévalu jusque dans les années 2000. Selon R. Busa, « *Humanities computing is precisely the automation of every possible analysis of human expression* », « les humanités computationnelles sont précisément l'automatisation de toute analyse possible de l'expression humaine » (je traduis) ; R.A. Busa, « Foreword », *op. cit.*, p. xvi. Le terme de « *Digital Humanities* » aurait, selon le témoignage de Stephen Ramsay (rapporté par N.K. Hayles, « How we think: Transforming power and digital technologies », dans *Understanding digital humanities*, 2012, p. 43. URL : http://link.springer.com/chapter/10.1057/9780230371934_3.. Consulté le 24 août 2017.), été créé par Johanna Drucker, John Unsworth et Jerome McGann, à la fin des années 1990, pour remplacer le terme d'« *Humanities Computing* », « *too closely associated with computing support services* » (« trop fortement associé avec les infrastructures de calcul », je traduis). Ce changement devait signifier que ce champ « *had emerged from the low-prestige status of a support service into a genuinely intellectual endeavour with its own professional practices, rigorous standards, and exciting theoretical explorations* » (« avait émergé du statut peu prestigieux d'une infrastructure de soutien à la recherche vers celui d'une démarche authentiquement intellectuelle, avec ses propres pratiques professionnelles, ses critères rigoureux, et des investigations théoriques enthousiasmantes », je traduis).

12 Du moins si l'on souhaite respecter la distinction, proposée par Jim Gray en 2007, entre les quatre paradigmes de la recherche : expérimentale, théorique (depuis Newton et Kepler), computationnelle (*i.e.*, fondée sur des simulations par ordinateur), et enfin « data-intensive » ; J. Gray, « Jim Gray on eScience: A transformed scientific method », dans *The fourth paradigm: Data-intensive scientific discovery*, T. Hey, S. Tansley, K. Tolle (éd.), Washington, 2009.; G. Bell, T. Hey et A. Szalay, « Beyond the data deluge », *Science*, vol. 323, n° 5919 (2009).

13 W. McCarty, « What is humanities computing? Toward a definition of the field », *Centre for Computing in the Humanities (King's College)* (1998). URL : <http://www.mccarty.org.uk/essays/McCarty,%20What%20is%20humanities%20computing.pdf>.. Consulté le 25 août 2017..

tionnel est progressivement passé d'une dimension ancillaire (« *slave labour* » dit McCarty dans sa leçon inaugurale de la chaire de *Humanities Computing* du King's College¹⁴), à une composante de la démarche scientifique, indispensable pour répondre aux questions que sa disponibilité permet de se poser¹⁵.

En amont, le travail philologique se voit ainsi renouvelé par des possibilités nouvelles en termes d'acquisition du texte, avec la reconnaissance du texte manuscrit (couramment, *Handwritten Text Recognition* ou *HTR*), la collation assistée par ordinateur, ainsi que la possibilité d'entraîner des modèles d'apprentissage machine (notamment, des réseaux de neurones) pour réaliser toutes sortes de tâches, qu'il s'agisse d'annotation linguistique (lemmatisation, annotation morpho-syntaxique ou syntaxique), sémantique (des entités nommées aux formules, motifs narratifs, figures rhétoriques, ...), ou stylistique, métrique, de catégorisation des lieux variants, etc.

En aval, ce qui s'offre au chercheur est la possibilité d'allier finesse, granularité descriptive et quantité importante de données dans l'analyse, dans des proportions jusqu'ici inenvisageables et permettant notamment de discerner des régularités difficilement perceptibles à l'œil humain, qu'il s'agisse par exemple de paléographie quantitative, scriptométrie, stémmatologie, stylométrie ou « lecture distante ».

Ce changement en aval n'est d'ailleurs pas uniquement quantitatif, mais a le potentiel de transformer également notre démarche scientifique, en nous permettant de passer d'une approche qui cherche dans des données délicates à manier la validation d'hypothèses préexistantes, à une approche qui fasse émerger une formalisation à partir des données¹⁶.

C'est là, selon toute vraisemblance, que se nichent les progrès à venir de la philologie et de notre connaissance du passé, en permettant des questionnements jusque là impossibles et en nous permettant de lever ou d'interroger certains des postulats les plus ancrés de notre heuristique¹⁷, comme les « présupposés du sens commun qui obstruent la connaissance de ces époques » passées, selon la formule d'Alain Guerreau¹⁸.

Néanmoins, ce changement, comme tous les changements de paradigme scientifique, voire de « matrice disciplinaire », dans la terminologie de Thomas Kuhn¹⁹, possède aussi une dimension sociologique et ne peut être entièrement porté par un sous-ensemble de « philologues computationnels », mais demande une transformation dans les valeurs et pratiques admises par la communauté savante dans son ensemble, comme le relève Tara Andrews au sujet de l'édition électronique :

The onus cannot fall entirely on the producers of critical editions, however; production is almost always driven by demand. Until those who might use our editions, beyond printing out a PDF copy of the critical text and citing the page number of an associated printed version, present themselves, our digital editions will continue to offer a convenient PDF version of the critical text with apparatus, and they will continue to have associated print publications to which most of the effort is devoted. It is the practice of deep and/or large-scale text analysis, rather than that of textual criticism itself, which must drive the development of digital editions in all their potential²⁰

14 L'expression est reprise d'un compte-rendu fait par Charles Geoffrey Vickers en 1971 ; W. McCarty, *Attending from and to the machine*, 2 février 2010, p. 8.

15 Pour reprendre la formulation de D.M. Berry, « The computational turn: Thinking about the digital humanities », *Culture Machine*, vol. 12 (2011), p. 2. URL : <http://www.culturemachine.net/index.php/cm/article/download/440/470>., « *computational technology has become the very condition of possibility required in order to think about many of the questions raised in the humanities today* », « les méthodes computationnelles sont devenues la condition même de possibilité de la réflexion sur bon nombre des questions soulevées aujourd'hui dans les sciences humaines » (je traduis).

16 En la matière, il faut insister sur le rôle pionnier du traitement automatique de la langue et de la linguistique de corpus, ou « outillée », comme le souligne F. Duval, « Transcrire le français médiéval: de l' 'instruction' de Paul Meyer à la description linguistique contemporaine », *Bibliothèque de l'École des chartes*, vol. 170 (2012), p. 339-340. R.A. Busa, « Foreword », *op. cit.*, p. xviii. parle pour cela d'« *Hermeneutic informatics* ». Loin d'être propre à la linguistique ou aux sciences humaines, cette transformation s'observe également dans les sciences du vivant ou de l'univers. Comme le rappellent Bell *et al.*, « *in contrast to the traditional hypothesis-led approach to biology, Venter and others have argued that a data-intensive inductive approach to genomics (...) is necessary to address large-scale ecosystem questions* » (« en contraste à l'approche traditionnelle, prédéterminée par les hypothèses, de la biologie, Venter et d'autres ont plaidé qu'une approche à la génomique inductive et intensive en données [...] est nécessaire pour se confronter aux questions d'écosystème à grande échelle », je traduis) ; G. Bell *et al.*, « Beyond the data deluge », *op. cit.*, p. 1297.

17 T. Andrews, « The third way », *op. cit.*, p. 8.

18 A. Guerreau, « Pour un corpus de textes latins en ligne », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* (avril 2011), paragr. 1. URL : <http://cem.revues.org/11787>. Consulté le 25 août 2017.

19 T.S. Kuhn, *The Structure of scientific revolutions*, 4e éd., Chicago, 2012., notamment p. 10-12 et 181-186.

20 T. Andrews, « The third way », *op. cit.*, p. 6., « Le poids ne peut retomber uniquement sur les producteurs d'éditions critiques, cependant ; la production est presque toujours déterminée par la demande. En attendant que surviennent ceux qui

2 Des données, pour quoi faire?

2.1 Nouvelles données, nouveaux questionnements

Avec le développement de l'édition électronique et la fin des contraintes imposées par le format imprimé, des éditeurs de texte n'ont pas tardé à s'emparer des possibilités nouvelles, permettant la démultiplication des textes et de leurs représentations ou l'enregistrement d'une large gamme de phénomènes. Puisqu'il est devenu possible de rendre compte de l'intégralité du texte de tous les témoins en sus de textes critiques, comme de représenter, dans les données, le physique (feuilles, colonnes, lignes) à côté du structural (chapitres ou strophes, vers), l'allographétique (variantes de formes des lettres, abréviations, segmentation ou ponctuation ancienne) à côté du graphématique, l'observé à côté de son interprétation, on a progressivement vu apparaître des éditions se détachant de la sélection et de la normalisation uniques traditionnellement présentées par les éditions imprimées. Dans le même temps, la spécificité de chaque texte comme de chaque approche interprétative, de pair avec l'individualité et la créativité de chaque éditeur, ont favorisé un écosystème dans lequel chaque édition répond à son modèle propre, quand bien même celui-ci s'inscrit généralement dans un cadre, large et permissif, tel que celui proposé par la *Text Encoding Initiative*²¹.

Soucieux de rendre apparente la richesse des données de leurs éditions et de laisser entrevoir la variété des analyses qu'elles permettent, certains projets ont donné lieu au développement d'interfaces de consultation novatrices²². Si, dans certains cas, ces interfaces permettent une valorisation légitime du travail éditorial, il est très regrettable que, dans d'autres, elles soient venues se substituer tout à fait à un accès direct aux données, parfois inaccessibles en tant que telles²³. Les interfaces, souvent le résultat de développements *ad hoc*, avec un financement limité dans le temps, et sans que l'on prenne nécessairement en compte le besoin de les maintenir dans la durée, sont de toute façon sujettes à une obsolescence rapide ; leur utilité est en partie limitée dans le temps et certainement limitée face à la richesse des réutilisations possibles des données, qui sont, elles, en tant que production scientifique, pérennes et susceptibles d'une vie beaucoup plus longue et de multiples renaissances.

Quelles que soient les raisons de l'obsession des interfaces – besoin légitime des lecteurs, simple réponse à la demande d'une communauté encore peu encline à se livrer à une manipulation directe des données, manque d'implication des chercheurs dans des aspects jugés « techniques », ou bien résultat de restrictions sur la diffusion des données, imposées par les éditeurs commerciaux ou la possessivité de certains savants vis-à-vis de leur travail – cette pratique tend à limiter la pérennité des éditions, comme les usages nouveaux des données, la réalisation d'éditions critiques prenant en compte de vastes traditions ou la constitution de corpus de grande envergure. Or, dans l'édition, même électronique, ce qui mobilise des compétences rares et engendre la plus-value scientifique majeure, et qui est à la fois le plus gourmand en temps, en moyens et en expertise, c'est le travail éditorial en lui-même. « *Data is the Important Long-term Outcome* » pour reprendre l'expression de Magdalena Turska, James Cummings et Sebastian Rahtz²⁴.

utiliseront nos éditions – au-delà de l'impression d'une copie en PDF du texte critique et de la citation d'un numéro de page d'une version imprimée associée – nos éditions numériques continueront d'offrir une version PDF maniable du texte critique avec l'apparat, et elles continueront d'avoir des publications imprimées associées, auxquelles l'essentiel de l'effort sera dévoué. C'est la pratique de l'analyse de profondeur et/ou à grande échelle, plutôt que celle de la critique textuelle elle-même, qui doivent entraîner le développement des éditions numériques dans tout leur potentiel » (je traduis).

21 TEI Consortium, « TEI P5: Guidelines for Electronic Text Encoding and Interchange », 2015. URL : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>.. Consulté le 9 mai 2015.

22 Parmi les prototypes particulièrement aboutis, on citera *Queste del saint Graal: Édition numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77)*, C. Marchello-Nizia (éd.), version 0.8.1, Lyon, 2012. URL : <http://portal-textometrie.org/txm/>. ; Thomasin von Zerklare, *Welscher Gast digital*, J. Šimek (éd.), Heidelberg, [s.d.]. URL : <http://digi.ub.uni-heidelberg.de/wgd>.. Consulté le 19 juillet 2016.

23 Pour une discussion de ce point et une défense de la prise en compte des interfaces comme un élément central de la démarche d'édition électronique, E. Pierazzo, *Digital scholarly editing, op. cit.*, sect. « 7.7 Are User Interfaces Necessary? ». Bien sûr, la possibilité d'une exploitation des données n'éteint pas le besoin de donner un ou des textes à lire ; au sujet de la tension et complémentarité entre le chercheur comme lecteur d'un texte ou comme utilisateur de données, on lira F. Duval, « Pour des éditions numériques critiques », *Médiévales*, à paraître. sect. « 1. Textes ou données ? Lecteurs ou utilisateurs ? ».

24 M. Turska, J. Cummings et S. Rahtz, « Challenging the Myth of Presentation in Digital Editions », *Journal of the Text Encoding Initiative*, Issue 9 (janvier 2016). URL : <http://jtei.revues.org/1453>.. Consulté le 3 septembre 2017. (c'est le titre de la sect. 2) ; les auteurs suggèrent que « *the encoding policy design and the later application of said policy to annotate a text are the most important acts that make all further research and long-term preservation of editors' wealth of knowledge [...] possible. Therefore in digital editions the encoded texts themselves are the most important long-term outcome of the*

Ne nous méprenons pas : même avec les gains de temps ou d'efficacités que permettent les outils numériques, l'édition électronique reste un processus difficile, chronophage et coûteux, notamment parce que nous préférons profiter des possibilités nouvelles pour enrichir et approfondir nos données et nos questionnements plutôt que de nous contenter d'une production standardisée et en série (il n'est pas interdit de s'en réjouir !). Il nous faut en revanche sortir du fantasme selon lequel une édition pourrait rendre compte de tous les faits d'un manuscrit ou d'un texte : acte interprétatif et sélectif par essence, elle n'en présente qu'un sous-ensemble²⁵. Si, armé des potentialités infinies de l'outil numérique, le philologue doit faire face à la question brûlante « *Where to stop ?* »²⁶, les données qu'il produit ne se retrouvent, elles, pas nécessairement condamnées à une seule représentation, à un usage limité, à un modèle ou un questionnaire spécifique : le partage et la diffusion des données leur permettent d'être reprises, transformées, incluses dans un nouveau corpus ou enrichies de nouvelles interprétations, de sélections de nouveaux faits. Ce n'est que ce faisant qu'on se laisse la possibilité que surviennent des usages et des découvertes que l'éditeur n'aurait jamais envisagés. Qui peut prévoir toutes les questions futures de la science ?

Un consensus fort existe selon lequel le partage et la libre diffusion des données n'a, pour la science, que des avantages. Les quatre justifications principales en sont, dans les termes de Christine Borgman,

*(1) to reproduce or to verify research, (2) to make results of publicly funded research available to the public, (3) to enable others to ask new questions of extant data, and (4) to advance the state of research and innovation*²⁷

Pour la philologie numérique, comme pour d'autres sciences, des méthodes prometteuses existent, qui ne pourront guère progresser, si nous nous acharnons à reproduire toujours les mêmes opérations, plutôt que de profiter pleinement de la dimension cumulative que facilitent les outils et formats numériques.

La production de corpus a beaucoup à gagner des méthodes fondées sur l'apprentissage machine, employées notamment pour la reconnaissance des écritures manuscrites, la lemmatisation et l'annotation linguistique, ou une variété d'autres formes d'enrichissement des données. Le point commun de ces méthodes est d'être très fortement dépendantes de la quantité et de la qualité des données d'entraînement disponibles, seules à même de retarder le surentraînement (*overtraining*) qui en limite l'efficacité et la généralité. Ainsi, si nous voulons pouvoir, à l'avenir, parvenir à des modèles plus performants et plus aisément applicables à de nouveaux textes, la constitution de jeux de données de grandes dimensions est un prérequis.

Du point de vue de l'analyse de ces données, beaucoup d'études se trouvent encore circonscrites à des descriptions isolées ou peinent à dégager à la fois la spécificité des textes, genres, zones ou périodes étudiés vis-à-vis du plus large contexte de l'écrit médiéval autant que les tendances générales qui traversent celui-ci²⁸. Comment identifier les caractéristiques paléographiques ou de *scripta* propres à un manuscrit de chanson de geste, sans corpus de manuscrits épiques auxquels le confronter ? les caractéristiques des manuscrits épiques sans un corpus d'oïl généraliste ? les caractéristiques de l'écrit d'oïl sans corpus occitan ou latin ?

project, while their initial presentation within a particular application should be considered only a single perspective on the data » (« la conception des règles d'encodage et l'application ultérieure desdites règles pour annoter un texte sont les actes les plus importants pour rendre possible de nouvelles recherches, ainsi que la préservation sur le long terme de la richesse du savoir des éditeurs. En conséquent, dans les éditions électroniques, ce sont les textes encodés eux-mêmes qui sont le résultat le plus important d'un projet sur le long terme, tandis que leur présentation initiale dans une application donnée ne devrait être considérée que comme une des vues sur les données », je traduis). Reconnaisant que les éditeurs scientifiques demeurent très préoccupés par la présentation de leur édition, ils proposent, pour leur permettre de s'en abstraire et de se concentrer sur la qualité des données, de créer un cadre général pour la définition de modèles de traitement et d'affichage, aussi simple et accessible que possible aux éditeurs.

25 C.M. Sperber-McQueen, « How to teach your edition how to swim », *op. cit.*, p. 31. Je renvoie également, sur ce point, à ma thèse de doctorat, J.-B. Camps, *La Chanson d'Otinel*, *op. cit.*, sect. 2.1, « Transcription et formalisation des données », p. cxcv-ccxvi.

26 E. Pierazzo, « A rationale of digital documentary editions », *Literary and Linguistic Computing*, vol. 26, n° 4 (2011).

27 « (1) pour reproduire ou vérifier une recherche, (2) pour rendre les résultats de la recherche financée par des fonds publics accessibles au public, (3) pour permettre à d'autres de poser de nouvelles questions à des données existantes, et (4) pour faire progresser l'état de la recherche et de l'innovation » (je traduis) ; C.L. Borgman, « The conundrum of sharing research data », *Journal of the American Society for Information Science and Technology*, vol. 63, n° 6 (juin 2012)..

28 En disant cela, je pense, parmi d'autres, aux miennes ; j'ai fait l'expérience, de première main, de ces difficultés, notamment en essayant d'identifier les particularités du corpus épique français et, en son sein, des manuscrits de la *Chanson d'Otinel*, ce qui m'a conduit à compiler des données de sources diverses – mais celles-ci sont encore rares, notamment pour ce qui a trait au système graphique et à l'allographie – pour disposer de points de comparaison (J.-B. Camps, *La Chanson d'Otinel*, *op. cit.*). Il s'agit, bien sûr, d'un phénomène qui touche encore assez largement nos domaines et n'est pas unique à mes travaux.

En d'autres cas, l'existence de jeux de données communs de référence est l'occasion d'éprouver, sur un même terrain, et de comparer objectivement, les résultats de différentes approches. La manifestation la plus visible – mais certainement pas la seule – en est formée par les défis organisés à l'occasion de congrès. Si ce type de compétitions scientifiques n'est pas nouveau, leur développement dans le champ des études textuelles – en dehors du traitement automatique de la langue, où elles sont monnaie courante – est plus récent. Elles touchent néanmoins désormais aux champs de la paléographie ou de la stemmatologie²⁹.

2.2 Des humanités reproductibles ?

L'intérêt de libre diffusion et de la pérennisation des données de la recherche ne se limite pas à une dimension cumulative ou comparative, mais touche également d'autres pans de la démarche scientifique. Au fur et à mesure que l'analyse de données prend de l'importance dans la constitution de nouveaux savoirs, le besoin se fait plus criant de vérifier l'intégrité des données, de reproduire les expériences, de vérifier ou infirmer les énoncés qui en découlent. Les sciences humaines n'ont, en effet, guère de justification pour se dispenser des critères de reproductibilité et de réfutabilité qui, depuis au moins Popper, forment les piliers de l'épistémologie³⁰. Il est souhaitable, au contraire, qu'elles s'inscrivent pleinement dans le mouvement pour la science ouverte, qui tire son origine de la vaste « Crise de la reproductibilité » sur laquelle, dans les années 2000, John Ioannidis et d'autres ont attiré l'attention³¹ : une vaste majorité des découvertes de la science seraient fausses (*i.e.* biaisées et non-reproductibles), voire souvent de simples mesures des biais induits par les chercheurs eux-mêmes. C'est sur ce constat que s'est construit le mouvement pour la science ouverte et la reproductibilité, impulsé notamment par Brian Nosek avec le *Reproducibility Project* (2011) – confirmant dans les grandes lignes les craintes qui s'étaient exprimées – puis le *Center for Open Science* (2013), et qui a déjà commencé à se propager de la psychologie et de la médecine vers les sciences humaines³².

29 Mentionnons le défi de stemmatologie organisé par une équipe de l'université de Helsinki, T. Roos, T. Heikkilä et P. Myllymäki, « Computer-Assisted Stemmology Challenge », 2007-2006. URL : [http://www.cs.helsinki.fi/u/ttonteri/casc/..](http://www.cs.helsinki.fi/u/ttonteri/casc/) Consulté le 11 janvier 2012. ou la compétition de classification des écritures menée par D. Stutzmann, *ICFHR2016: Competition on the Classification of Medieval Handwritings in Latin Script*, Paris, 2016. URL : <http://icfhr2016-clamm.irht.cnrs.fr/>.

30 Pour répondre au problème de l'induction (comment généraliser une observation, ou un ensemble d'observations individuelles, pour en tirer des énoncés de valeur générale, des connaissances ?) et éviter d'être privé de toute méthode d'évaluation de la vérité d'une théorie (une théorie ne pouvant être vérifiée par des observations, à partir du moment où d'autres observations la contredisant pourraient toujours survenir), Popper opère une forme de renversement : plutôt que de considérer qu'une théorie est vraie à partir du moment où elle est appuyée par un ensemble d'observations, on peut l'accepter temporairement si elle soutient l'épreuve des observations et tant qu'aucune observation contraire n'a été faite. Pour que cela soit possible, encore faut-il qu'un énoncé ait la capacité d'être réfuté (par ex., « tous les cygnes sont blancs » ou « la forme *soignor* n'apparaît qu'en Lorraine », « toutes les chansons de geste sont en décasyllabes »). S'il n'en dispose pas, il s'agit alors, non d'un énoncé scientifique, mais métaphysique (« les chansons de geste sont composées ou non de décasyllabes », vérifiable mais non réfutable ; « la forme *soignor* est susceptible d'apparaître dans n'importe quelle région », peut-être vérifiable, mais non réfutable ; « le style d'un auteur lui est toujours propre et unique, mais ne peut être mesuré ou décrit », dans cette formulation, ni vérifiable, ni réfutable). Ces énoncés métaphysiques peuvent être utiles, puisque, après tout, la manière dont naissent les idées répondrait en partie, d'un point de vue psychologique, à des processus tels que l'intuition créative, l'inspiration, les convictions ou les sentiments qui habitent le chercheur – mais sont par définition en dehors du champ de la connaissance scientifique ; il en allait ainsi d'ailleurs à l'époque de la publication de l'ouvrage de Popper de certains éléments de la théorie de la relativité. Cette proposition apporte également une solution partielle au problème de la subjectivité du chercheur : si un énoncé peut-être testé, si ces tests peuvent être reproduits par d'autres ou de nouveaux tests conçus, l'énoncé en devient inter-subjectivement testable (et donc réfutable) – ce qui introduit également la nécessité de la reproductibilité. Voir K.R. Popper, *The logic of scientific discovery*, London, 2002..., part. sect. 1.1 « A survey of some fundamental problems », p. 3-26 et 2.9, « Some observations on quantum theory », p. 209-247.

31 J.P. Ioannidis, « Why most published research findings are false », *PLoS medicine*, vol. 2, n° 8 (2005).

32 Le premier projet a porté sur le champ de la psychologie (Open Science Collaboration, « Estimating the reproducibility of psychological science », *Science*, vol. 349, n° 6251 (2015).), démontrant que, sur un échantillon de 100 études, entre un et deux tiers seulement des résultats répliqués restaient significatifs ; les résultats d'un second projet du même type, concernant les études sur le cancer, sont en train d'être publiés (B.A. Nosek et T.M. Errington, « Reproducibility in Cancer Biology: Making sense of replications », *eLife*, vol. 6 (janvier 2017).). À ma connaissance, les chercheurs en sciences humaines, numériques ou non, ne se sont pas encore massivement inscrits dans cette démarche. Les sociétés savantes et revues gagneraient néanmoins à s'inscrire dans une démarche telle que celle préconisée par B.A. Nosek, G. Alter, G. Banks et al., « Transparency and Openness Promotion (TOP) Guidelines » (2016). URL : [https://osf.io/preprints/vj54c/..](https://osf.io/preprints/vj54c/) Consulté le 1 septembre 2017. Au dernier congrès *DHBenelux*, en 2017, un poster a été présenté par Yvonne Colijn, Marten Jan Bok et Harm Nijboer avec le titre « Artists in late medieval Ghent: a digital replication study ». Les auteurs, qui s'inscrivent dans la lignée des travaux de Nosek, y répliquent, avec des outils numériques, les travaux prosopographiques menés par Els Cornelis en 1987 et 1988 sur 248 artistes de Gand au Moyen Âge tardif, pour en vérifier, remettre en cause ou augmenter

Pour être réalisée, cette exigence de reproductibilité implique de rendre aussi transparents que possible les jeux de données utilisés, les protocoles d'analyse et la manière dont les résultats obtenus s'articulent avec les hypothèses formulées dans les publications. Concrètement, cela demande de rendre compte des données et logiciels dans la version utilisée, des manipulations, transformations, requêtes ou scripts, ainsi que des résultats, de préférence de manière plus large que la sélection qui franchit le seuil de la publication, la partie émergée de l'iceberg de données, pour paraphraser Jim Gray³³.

En ce qui concerne plus précisément la philologie, nous avons besoin de standards pour citer exactement une portion donnée d'une édition particulière d'un texte, dans la version qui a servi à l'analyse³⁴ ; d'entrepôts pérennes permettant d'accéder aux données dans le temps, avec des éléments indispensables de contextualisation de celles-ci, à l'instar du service fourni par l'*University of Oxford Text Archive*³⁵ ; de modèles et d'outils interopérables et transparents³⁶. Plus généralement, nous avons besoin d'infrastructures et d'un écosystème au sein desquels les données et leurs interprétations circulent librement, et où le chercheur peut aller facilement de l'un à l'autre – et en cela, nous pouvons bénéficier des précédents de l'*eScience*³⁷. Mais nous avons aussi besoin d'être convaincus, en tant que chercheurs, de la nécessité et de

les conclusions, grâce aux données qui ont été collectées lors de la réalisation de la base *ECARTICO* : *Linking cultural industries in the early modern Low Countries, ca. 1475 - ca. 1725*, dir. Marten Jan Bok et Harm Nijboer, Amsterdam : Centre for the Study of the Golden Age, 2011-..., en ligne : <http://www.vondel.humanities.uva.nl/ecartico>. Les données utilisées sont accessibles en ligne, sous une licence libre permettant la réutilisation (CC BY SA). Je remercie les auteurs d'avoir bien voulu me communiquer une copie de leur poster.

- 33 « *The published literature is just the tip of the data iceberg. By this I mean that people collect a lot of data and then reduce this down to some number of column inches in Science or Nature [...]. So what I mean by data iceberg is that there is a lot of data that is collected but not curated or published in any systematic way* » (« Les publications scientifiques ne sont que la partie émergée de l'iceberg des données. Par cela, je veux dire que les gens recueillent de grandes quantités de données, et les réduisent à des colonnes de quelques centimètres dans *Science* ou *Nature* [...]. Donc, ce que j'entends par iceberg de données, c'est qu'il existe beaucoup de données qui sont recueillies, mais sans être conservées ou publiées d'une manière systématique », je traduis) ; J. Gray, « Jim Gray on eScience », *op. cit.*, p. xvii. Selon G. Bodard et J. Garcés, « Open source critical editions: a rationale », *Text editing, print and the digital world* (2009), p. 86., « *without full access to the raw code, the documentation and the methodological statement that makes an experiment or a solution reproducible, a given publication is a dead end; it cannot be built upon* » (« sans un accès complet au code brut, la documentation et l'énoncé de la méthodologie qui rendent une expérience ou une solution reproductibles ne sont qu'un cul de sac ; on ne peut construire sur eux », je traduis).
- 34 Des solutions dans ce domaine se développent ces dernières années, par exemple C. Blackwell, C. Roughan et N. Smith, « Citation and Alignment: Scholarship Outside and Inside the Codex », *Manuscript Studies*, vol. 1, n° 1 (juin 2017). URL : http://repository.upenn.edu/mss_sims/vol1/iss1/2. ; T. Clérice, « CapiTainS, l'édition numérique de texte et son exploitation », *Médiévales*, à paraître.
- 35 Ce service exemplaire vise à fournir tous les jeux de données textuels produits par l'Université d'Oxford, en précisant notamment le format, la licence, et en l'accompagnant parfois de liens vers des outils d'analyse ; certaines des données qui y sont fournies ont entamé leur vie sur des cartes perforées dans les années 1970 (voir par exemple le n° 0004, « keypunched June 1971 »). Les services généralistes visant à la pérennisation et citabilité des données de recherche se sont démultipliés ces dernières années. En Europe, la plateforme Zenodo a été lancée par le CERN et le projet OpenAIRE en 2013 avec le soutien de l'Union Européenne (OpenAIRE et CERN, « Zenodo », 2013. URL : <http://zenodo.org/>), et, en France, l'infrastructure de recherche Huma-Num maintient depuis 2014 un service dédié aux sciences humaines, Nakala (TGIR Huma-Num, « Nakala », 2014. URL : <http://www.nakala.fr/>).
- 36 Pour répondre à ces enjeux, G. Bodard et J. Garcés, « Open source critical editions », *op. cit.*, p. 84-85. ont proposé, en 2009, une vue de ce que doivent être des « *Open Source Critical Editions* », à savoir « *more than merely the final representations of finished work; in their essence they involve the distribution of raw data, of scholarly tradition, of decision-making processes, and of the tools and applications that were used in reaching these conclusions. The protocols and technologies for this manner of publication need to be made available and comprehensible to all textual scholars if the unique advantages and opportunities afforded by digital scholarship are to become entrenched in Classical and other philological disciplines* » (« plus que les simples résultats finaux d'un travail fini ; dans leur essence, elles impliquent la diffusion des données brutes, de la tradition savante, du processus de décision, et des outils et applications qui ont été utilisés pour parvenir à ces conclusions. Les protocoles et technologies pour ce type de publication doivent être rendus accessibles et compréhensibles par tous les chercheurs spécialistes du texte, si l'on souhaite que les avantages et les opportunités incomparables de la recherche numérique deviennent fermement établis au sein de la philologie classique et des autres disciplines philologiques », je traduis).
- 37 Il s'agit d'une des dimensions constitutives de l'*eScience* telle que définie par J. Gray, « Jim Gray on eScience », *op. cit.*, p. xxv., « *the Internet can do more than just make available the full text of research papers. In principle, it can unify all the scientific data with all the literature to create a world in which the data and the literature interoperate with each other. You can be reading a paper by someone and then go off and look at their original data. You can even redo their analysis. Or you can be looking at some data and then go off and find out all the literature about this data. Such a capability will increase the "information velocity" of the sciences and will improve the scientific productivity of researchers* » (« L'Internet peut faire plus que de simplement rendre disponible le plein-texte des articles de recherche. Sur le principe, il peut rassembler toutes les données de la science avec toutes les publications pour créer un monde dans lesquels données et publications interagissent entre elles. Vous pouvez lire l'article de quelqu'un et rebondir pour regarder leurs données originales.

l'intérêt qu'il y a au partage de nos données, encore plus souvent prôné que réellement mis en œuvre³⁸.

3 Libérez les données !

Même si l'on goûte fort peu les anglicismes, il en est un qui a le mérite de nommer un phénomène jusque là peu appréhendé, le *copyfraud*. Désignant l'escroquerie au droit d'auteur, ou, plus exactement, la revendication fallacieuse de droits sur des œuvres du domaine public, ce terme a été inventé en 2006 par Jason Mazzone pour décrire une forme omniprésente, mais peu remarquée, d'usurpation du domaine public :

These false copyright claims, which are often accompanied by threatened litigation for reproducing a work without the putative "owner's" permission, result in users seeking licenses and paying fees to reproduce works that are free for everyone to use, or altering their creative projects to excise the uncopyrighted material. [...] False assertions of copyright are everywhere. In general, copyright belongs to the author of a published work and expires seventy years after the author's death. Yet copyright notices appear on modern reprints of Shakespeare's plays, on Beethoven's piano scores, and on greeting card versions of Monet's Water Lilies. Archives claim blanket copyright in everything in their collections, including historical works as to which copyright, which likely never belonged to the archive in the first place, has long expired. [...] As a result, publishers and the owners of physical copies of works plaster copyright notices on everything. These publishers and owners also restrict copying and extract payment from individuals who do not know better or find it preferable not to risk a lawsuit. These circumstances have produced fraud on an untold scale, with millions of works in the public domain deemed copyrighted and countless dollars paid out every year in licensing fees to make copies that could be made for free³⁹

Mazzone répartit les cas de *copyfraud* en trois catégories, dont chacun d'entre nous a probablement déjà fait l'expérience : les mentions de *copyright* en tête des éditions d'œuvres du domaine public ou de nu-

Vous pouvez même refaire leur analyse. Ou, vous pouvez être en train de regarder les données de quelqu'un, puis rebondir et trouver toutes les publications qui les concernent. L'aptitude à faire cela augmentera la « vitesse de l'information » au sein des sciences et améliorera la productivité scientifique des chercheurs », je traduis). C.L. Borgman, « The digital future is now: A call to action for the humanities », *Digital humanities quarterly*, vol. 3, n° 4 (2009). appelait en 2009 les sciences humaines à une évolution similaire. Dans le domaine de la philologie, ce besoin a en partie été évoqué dès 2007 par G. Crane, D. Bamman, A. Babeu et al., « EPhilology: when the books talk to their readers », *A companion to digital literary studies* (2007), part. « Building the Infrastructure for ePhilology ».

38 C'est le constat, pour les sciences économiques, de P. Andreoli-Versbach et F. Mueller-Langer, « Open access to data: An ideal professed but not practised », *Research Policy*, vol. 43, n° 9 (novembre 2014). Sur 488 chercheurs étudiés, 80 % environ ne partagent jamais leurs données et 2,5 % seulement le font de manière claire et systématique. Parmi les raisons, on trouve, tout simplement, la volonté de se garder l'exclusivité de données longues et coûteuses à produire, pour d'éventuelles recherches ultérieures ; le coût en temps nécessaire pour les documenter et rendre disponibles, ainsi que le manque de contreparties (la publication de données étant souvent encore peu prise en compte dans l'attribution de financements ou l'avancement des carrières). Ces facteurs expliquent probablement la corrélation positive, identifiée par les auteurs, entre partage des données et stature du chercheur (avancement de la carrière, prestige de l'institution, nombre de citations, etc.). Le non-partage des données peut pourtant avoir des conséquences graves, facilitant les fraudes, collusions, ou simples erreurs (l'erreur de programmation qui a mené à l'établissement d'une corrélation négative entre dette et croissance, et servi de justification à des politiques d'austérité, est fameuse, mais ce n'est pas la seule).

39 J. Mazzone, *Copyfraud*, Rochester, NY, 25 août 2005, p. 1028-1030. URL : <https://papers.ssrn.com/abstract=787244>. Consulté le 28 août 2017., « Ces fausses revendications de droits patrimoniaux, qui sont souvent accompagnées de menaces de poursuite en cas de reproduction d'une œuvre sans l'accord de son « propriétaire » supposé, ont pour résultat que des utilisateurs recherchent des licences ou payent des droits pour reproduire des œuvres qui sont libres d'utilisation pour tout un chacun, ou transforment leur projet de création pour en retirer des matériaux qui ne sont pas sous droits. [...] Ces fausses affirmations de droits sont partout. En général, les droits d'auteurs patrimoniaux appartiennent à l'auteur d'une œuvre publiée et expirent soixante-dix ans après sa mort. Pourtant, des mentions de droits apparaissent sur des réimpressions modernes des pièces de Shakespeare, sur les partitions de piano de Beethoven, et sur les *Nymphéas* de Monet au format carte de vœux. Des archives revendiquent des droits englobant sur tout ce qui se trouve dans leurs collections, y compris des œuvres historiques dont les droits – qui ne leur ont pour commencer, selon toute vraisemblance, jamais appartenu – ont depuis longtemps expiré. [...] En conséquence, des éditeurs commerciaux et des possesseurs restreignent aussi la copie et extorquent des paiements d'individus non avertis ou qui préfèrent ne pas risquer de poursuites judiciaires. Ces circonstances ont engendré des fraudes sur une échelle inouïe, où des millions d'œuvres dans le domaine public sont considérées comme sous droits et des dollars innombrables payés chaque année en frais de licence pour faire des copies qui auraient pu être faites gratuitement » (je traduis).

mérisations ; les restrictions d'utilisation (citation, copie, usages académiques ou fouille de données⁴⁰) qui vont au-delà du droit ; les revendications de droits sous prétexte de la possession d'une copie matérielle de l'œuvre, telle que les pratiquent notamment archives, musées ou bibliothèques⁴¹.

Si cette usurpation du domaine public est si répandue, et touche jusqu'aux peintures rupestres des grottes de Lascaux⁴², c'est que les causes qui la favorisent sont multiples : la méconnaissance du droit (notamment par les milieux universitaires), l'impunité quasi-généralisée et le succès (financier) qui couronne ces fraudes, et, comme le souligne Mazzone, la désinformation à laquelle se livrent certains éditeurs⁴³, comme le détail des législations elles-mêmes, beaucoup plus attentives au respect du droit d'auteur qu'à celui du domaine public, surtout lorsqu'il s'agit de punir les contrevenants⁴⁴.

Pourtant, pour le médiéviste, la situation est simple : comme il n'existe, par définition, aucun auteur médiéval mort il y a moins de 70 ans, tous les textes médiévaux sont du domaine public ; transcrire et établir un texte ne crée pas de droit nouveau sur le texte en lui-même, pas plus que de s'en faire l'éditeur commercial, de le numériser ou d'en posséder un manuscrit⁴⁵. Des jugements récents sont venus confirmer cette protection des textes médiévaux vis-à-vis de revendications de propriété⁴⁶. Les variantes aussi, « elles-mêmes résultats de modifications fort anciennes »⁴⁷, appartiennent au domaine public.

Si les textes et leurs variantes sont, donc, du domaine public et libres de droit, ce n'est pas nécessairement le cas des introductions, notes et éventuelles traductions, ce qui signifie concrètement que tout un chacun peut récupérer, utiliser et rediffuser un texte, mais dépouillé des matériaux critiques fournis par l'éditeur scientifique, pourtant à même de le contextualiser et d'en éclairer le sens. Pourquoi priver la

40 On notera qu'en France, depuis la loi du 7 octobre 2016 « Pour une République numérique », la fouille de données est permise même, en tant qu'exception, pour des œuvres qui sont effectivement sous droits, sans parler donc des textes anciens qui ne le sont pas ; sont ainsi légalisées les « copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale » ; A.-L. Stérin, « Le chercheur « fouille » dans des textes ou dans des données », Billet, [s.d.]. URL : <http://ethiquedroit.hypotheses.org/1528>.. Consulté le 29 août 2017.

41 J. Mazzone, *Copyfraud, op. cit.*, p. 1038-1058.

42 Défense de rire ; et ce n'est rien moins que le Département de la Dordogne qui revendique des droits sur ces peintures vieilles de 17 000 ans, quoi qu'ayant déjà été débouté par le tribunal de commerce de Périgueux en 2012. Dans cette même série surréaliste, les découvreurs de la grotte « Chauvet » ont tenté de faire valoir un droit d'auteur au titre de la publication d'œuvres posthumes (inédites), et ont été déboutés « car des vestiges [préhistoriques] attestent d'une présence humaine postérieure à la réalisation des peintures » ! Voir L. Maurel, « Ces étranges « biens communs de la Nation » cachés dans la loi sur la Création... », - *S.I.Lex* -, 21 juillet 2015. URL : <https://scinfolex.com/2015/07/21/ces-etranges-biens-communs-de-la-nation-caches-dans-la-loi-sur-la-creation/>.. Consulté le 29 août 2017.

43 Pour des exemples américains, voir J. Mazzone, *Copyfraud, op. cit.*, p. 1051.

44 En France aussi, les pouvoirs publics se montrent réticents à renforcer les sanctions contre le *copyfraud*, en dépit d'une proposition de loi, en 2013, qui prévoyait notamment « l'intégration [au domaine public] des œuvres produites par des agents publics dans l'exercice de leurs fonctions », confirmait que « lorsqu'une base de données contient des œuvres de l'esprit appartenant au domaine public, sous quelque forme que ce soit, le producteur de la base ne peut ni interdire ni s'opposer à leur extraction ou à leur réutilisation », tout en punissant « d'un an d'emprisonnement et de 100 000 euros d'amende le fait de porter atteinte à l'intégrité du domaine public en faisant obstacle ou en tentant de faire obstacle à la libre réutilisation d'une œuvre qui s'y rattache ou en revendiquant abusivement des droits sur celle-ci » ; I. Attard, *Proposition de loi visant à consacrer le domaine public, à élargir son périmètre et à garantir son intégrité*, 21 novembre 2013. URL : <http://www.assemblee-nationale.fr/14/propositions/pion1573.asp>.. Consulté le 5 septembre 2017.. En 2016, lors de l'examen de la loi *Pour une République numérique*, les députés ont même voté un amendement pour permettre à « toute association régulièrement déclarée depuis au moins deux ans et se proposant, par ses statuts, de « protéger la propriété intellectuelle, de défendre le domaine public ou de promouvoir la diffusion des savoirs » de saisir le tribunal de grande instance « afin de faire cesser tout obstacle à la libre réutilisation d'une œuvre entrée dans le domaine public » (A. Isabelle, *Assemblée nationale - RÉPUBLIQUE NUMÉRIQUE(no 3399) - Amendement no 613*, [s.d.]. URL : <http://www.assemblee-nationale.fr/14/amendements/3399/AN/613.asp>.. Consulté le 16 novembre 2017.), contre l'avis du gouvernement, mais il n'a pas été repris dans la loi n°2016-1321 telle qu'elle a été promulguée le 7 octobre 2016.

45 Pour un résumé, point à point, de la situation, on se reportera à A. Guerreau, « Pour un corpus de textes latins en ligne », *op. cit.*, sect. « Le cadre fondamental (en France) : le droit d'auteur dans sa version française ». La seule exception possible pourrait être constituée par les œuvres « posthumes », c'est-à-dire rendues publiques après la mort de leur auteur, pour lesquelles les ayant-droits disposeraient d'un délai d'exploitation exclusive de 25 ans, mais, comme le rappelle Alain Guerreau, « qui connaît des héritiers d'auteurs médiévaux ? ».

46 Le jugement de l'affaire Droz/Garnier en 2014 par le tribunal de grande instance de Paris est désormais de notoriété publique chez les médiévistes, et a été confirmé en appel en 2017 ; pour un résumé du premier jugement, L. Maurel, « Une victoire pour le domaine public : un cas de copyfraud reconnu par un juge français », - *S.I.Lex* -, [s.d.]. URL : <http://scinfolex.com/2014/04/13/une-victoire-pour-le-domaine-public-un-cas-de-copyfraud-reconnu-par-un-juge-francais/>.. Consulté le 13 avril 2014.; la confirmation a encore peu été commentée, mais elle est évoquée par le même auteur, L. Maurel, « Humain, robotique, animal : que devient l'auteur ? », - *S.I.Lex* -, 17 juillet 2017. URL : <https://scinfolex.com/2017/07/17/humain-robotique-animal-que-devient-lauteur/>.. Consulté le 29 août 2017.

47 A. Guerreau, « Pour un corpus de textes latins en ligne », *op. cit.*, parag. 20.

communauté savante comme le grand public de cette plus-value du travail philologique, de ce surcroît de savoir ?

Ce n'est pas le seul aspect pervers de ce mode de fonctionnement. En effet, devant les fumées peu engageantes produites par les annonces et injonctions contradictoires inscrites au frontispice des bases de données comme au seuil de l'enfer de Dante⁴⁸, la tentation est forte, pour le chercheur peu téméraire voulant seulement ajouter un texte à son corpus, de prendre des libertés dans l'exactitude des références bibliographiques de l'édition qu'il va utiliser – ou, au moins, de ne pas redistribuer les données sur lesquelles se fondent ses expériences. Quand on sait tout ce que l'établissement du texte et son sérieux (ou non) doit au lent travail et à l'application du philologue qui l'édite, il n'est pas difficile d'imaginer les conséquences de cet anonymat délétère, qui contrevient en outre aux fondements de l'éthique académique.

Ces deux problèmes devraient pouvoir suffire à eux seuls, s'il était besoin d'argument supplémentaires, à convaincre les éditeurs de texte d'opter pour la seule solution à même de faire véritablement valoir leur travail et de lui permettre d'atteindre son utilité pleine et entière : libérer d'eux-mêmes l'intégralité de leur travail éditorial, en le plaçant, à leur guise, sous une licence libre ou dans le domaine public⁴⁹.

Si l'on souhaite que la philologie prenne la place qui lui est due dans le paysage scientifique actuel et bénéficie des progrès scientifiques considérables contenus, en germe, dans le quatrième paradigme de la recherche scientifique, il importe certes de transformer la manière dont nous pratiquons l'édition de texte et dont nous offrons le produit de notre travail à la communauté savante ; mais il faut également nous transformer nous-mêmes, en adoptant une posture humble et ouverte devant nos sources, sans appropriation induite et sans dissimuler la possibilité de l'erreur – mieux, en fournissant à la communauté non seulement les résultats de nos recherches, mais aussi la manière dont ils ont été produits et peuvent être reproduits ; en leur fournissant, en somme, les moyens de réfuter nos énoncés, de faire mieux, de faire plus ou de faire autrement ; en acceptant que le mécanisme fondamental du progrès scientifique est la réfutation. L'erreur au centre : voilà qui devrait, en tant que philologues, nous réjouir !

ALTSCHUL (Nadia), « The Genealogy of Scribal Versions: A 'Fourth Way' in Medieval Editorial Theory », *Textual Cultures: Texts, Contexts, Interpretation*, vol. 1, n° 2 (2006), p. 114-136.

ANDREOLI-VERSACH (Patrick) et MUELLER-LANGER (Frank), « Open access to data: An ideal professed but not practised », *Research Policy*, vol. 43, n° 9 (novembre 2014), p. 1621-1633.

ANDREWS (Tara), « The third way: philology and critical edition for a digital age », *Variants: the Journal of the European Society for Textual Scholarship*, vol. 10 (2012). URL : <http://boris.unibe.ch/43071/>.

ATTARD (Isabelle), *Proposition de loi visant à consacrer le domaine public, à élargir son périmètre et à garantir son intégrité*, 21 novembre 2013. URL : <http://www.assemblee-nationale.fr/14/propositions/pion1573.asp>. Consulté le 5 septembre 2017.

BELL (Gordon), HEY (Tony) et SZALAY (Alex), « Beyond the data deluge », *Science*, vol. 323, n° 5919 (2009), p. 1297–1298.

48 Comme le note Alain Guereau, « on rencontre constamment des procédures qui visent à faire reconnaître diverses “interdictions” : les commerçants (...) bardent leurs CD et leurs sites de rodomontades et de gesticulations aussi mensongères que comminatoires et prétendent subordonner l'usage de leurs textes à la reconnaissance de leurs soi-disant “droits” (...). Tout cela contrevient à la loi, et n'a donc strictement aucune valeur juridique (nullité de plein droit : cliquez et signez tant que vous voudrez, cela ne vous engage à rien) » ; *Ibid.*, paragr. 23..

49 Je ne peux que reprendre ici l'appel lancé par P. Robinson, « Why digital humanists should get out of textual scholarship », dans , présenté à Social, Digital, Scholarly Editing, Saskatoon, [s.n.], 2013. URL : https://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship. : « Editors, hear this: if you want your digital edition to be used by others, to become the centre of other people's work, to be transfigured over and over into bright new forms of scholarship [...] you have to release it without restriction » (« Éditeurs de texte, entendez ceci : si vous voulez que vos éditions numériques soient utilisées par d'autres, soient centrales dans le travail d'autres personnes, soient métamorphosées sans cesse en de brillantes et nouvelles formes de recherche [...] il vous faut les offrir sans aucune restriction », je traduis).

- BERRY (David M.), « The computational turn: Thinking about the digital humanities », *Culture Machine*, vol. 12 (2011). URL : <http://www.culturemachine.net/index.php/cm/article/download/440/470>.
- BLACKWELL (Christopher), ROUGHAN (Christine) et SMITH (Neel), « Citation and Alignment: Scholarship Outside and Inside the Codex », *Manuscript Studies*, vol. 1, n° 1 (juin 2017). URL : http://repository.upenn.edu/mss_sims/vol1/iss1/2.
- BODARD (Gabriel) et GARCÉS (Juan), « Open source critical editions: a rationale », *Text editing, print and the digital world* (2009), p. 83–98.
- BORGMAN (Christine L.), « The conundrum of sharing research data », *Journal of the American Society for Information Science and Technology*, vol. 63, n° 6 (juin 2012), p. 1059-1078.
- BORGMAN (Christine L.), « The digital future is now: A call to action for the humanities », *Digital humanities quarterly*, vol. 3, n° 4 (2009), p. 233.
- BUSA (Roberto A.), « Foreword: Perspectives on the digital humanities », *A companion to digital humanities* (2004). URL : <http://www.digitalhumanities.org/companion/>.. Consulté le 24 août 2017.
- CAMPS (Jean-Baptiste), *La 'Chanson d'Otinel': édition complète du corpus manuscrit et prolégomènes à l'édition critique*, thèse de doctorat, dir. Dominique Boutet, Paris-Sorbonne, Paris, 3 décembre 2016. URL : <http://www.theses.fr/2016PA040173>.
- CAMPS (Jean-Baptiste), « Copie, authenticité, originalité dans la philologie et son histoire », *Questes: revue pluridisciplinaire d'études médiévales*, vol. 29 (janvier 2015), p. 35-67.
- CERQUIGLINI (Bernard), « Éloge de la variante », *Langages*, vol. 17, n° 69 (1983), p. 25-35.
- CLÉRICE (Thibault), « CapiTainS, l'édition numérique de texte et son exploitation », *Médiévales*, à paraître.
- CRANE (Gregory), BAMMAN (David), BABEU (Alison) et al., « EPhilology: when the books talk to their readers », *A companion to digital literary studies* (2007), p. 29–64.
- DUVAL (Frédéric), « Transcrire le français médiéval: de l' 'instruction' de Paul Meyer à la description linguistique contemporaine », *Bibliothèque de l'École des chartes*, vol. 170 (2012), p. 321–342.
- DUVAL (Frédéric), « Pour des éditions numériques critiques », *Médiévales*, à paraître.
- FROGER (Jacques), *La critique des textes et son automatisation*, Paris, 1968 (Initiation aux nouveautés de la science, 1 vol.).
- GRAY (Jim), « Jim Gray on eScience: A transformed scientific method », dans *The fourth paradigm: Data-intensive scientific discovery*, éd. T. HEY, S. TANSLEY, K. TOLLE, Washington, 2009, p. xvii-xxxi.
- GUERREAU (Alain), « Pour un corpus de textes latins en ligne », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* (avril 2011). URL : <http://cem.revues.org/11787>.. Consulté le 25 août 2017.
- HAYLES (N. Katherine), « How we think: Transforming power and digital technologies », dans *Understanding digital humanities*, 2012. URL : http://link.springer.com/chapter/10.1057/9780230371934_3.. Consulté le 24 août 2017, p. 42–66.
- IOANNIDIS (John PA), « Why most published research findings are false », *PLoS medicine*, vol. 2, n° 8 (2005), p. e124.
- ISABELLE (Attard), *Assemblée nationale - RÉPUBLIQUE NUMÉRIQUE(no 3399) - Amendement no 613*, [s.d.]. URL : <http://www.assemblee-nationale.fr/14/amendements/3399/AN/613.asp>.. Consulté le 16 novembre 2017.

- KUHN (Thomas Samuel), *The Structure of scientific revolutions*, 4e éd., Chicago, 2012. 501. ISBN : 978-0-226-45812-0.
- MAUREL (Lionel), « Humain, robotique, animal : que devient l'auteur ? », – *S.I.Lex* –, 17 juillet 2017. URL : <https://scinfolex.com/2017/07/17/humain-robotique-animal-que-devient-lauteur/>.. Consulté le 29 août 2017.
- MAUREL (Lionel), « Ces étranges « biens communs de la Nation » cachés dans la loi sur la Création... », – *S.I.Lex* –, 21 juillet 2015. URL : <https://scinfolex.com/2015/07/21/ces-etranges-biens-communs-de-la-nation-caches-dans-la-loi-sur-la-creation/>.. Consulté le 29 août 2017.
- MAUREL (Lionel), « Une victoire pour le domaine public : un cas de copyfraud reconnu par un juge français », – *S.I.Lex* –, [s.d.]. URL : <http://scinfolex.com/2014/04/13/une-victoire-pour-le-domaine-public-un-cas-de-copyfraud-reconnu-par-un-juge-francais/>.. Consulté le 13 avril 2014.
- MAZZONE (Jason), *Copyfraud*, Rochester, NY, 25 août 2005. URL : <https://papers.ssrn.com/abstract=787244>.. Consulté le 28 août 2017.
- MCCARTY (Willard), *Attending from and to the machine*, 2 février 2010.
- MCCARTY (Willard), « What is humanities computing? Toward a definition of the field », *Centre for Computing in the Humanities (King's College)* (1998). URL : <http://www.mccarty.org.uk/essays/McCarty,%20What%20is%20humanities%20computing.pdf>.. Consulté le 25 août 2017.
- NOSEK (Brian A.), ALTER (George), BANKS (George) et al., « Transparency and Openness Promotion (TOP) Guidelines » (2016). URL : <https://osf.io/preprints/vj54c/>.. Consulté le 1 septembre 2017.
- NOSEK (Brian A.) et ERRINGTON (Timothy M.), « Reproducibility in Cancer Biology: Making sense of replications », *eLife*, vol. 6 (janvier 2017), p. e23383.
- OPEN SCIENCE COLLABORATION, « Estimating the reproducibility of psychological science », *Science*, vol. 349, n° 6251 (2015), p. aac4716.
- OPENAIRE et CERN, « Zenodo », 2013. URL : <http://zenodo.org/>.
- PALUMBO (Giovanni) et RINOLDI (Paolo), « Prolégomènes à l'édition du corpus français de la 'Chanson d'Aspremont' », dans *Epic Connections / Rencontres épiques Proceedings of the Nineteenth International Conference of the Société Rencesvals, Oxford, 13–17 August 2012*, éd. M.J. AILES, P.E. BENNETT, A.E. COBBY, Édimbourg, [s.n.], 2015 (British Rencesvals Publications, 7, vol. 2), p. 549-576.
- PIERAZZO (Elena), *Digital scholarly editing: theories, models and methods*, Farnham, 2015. URL : [dx.doi.org/10.4324/9781315577227](https://doi.org/10.4324/9781315577227). ISBN : 978-1-4724-1211-9.
- PIERAZZO (Elena), « A rationale of digital documentary editions », *Literary and Linguistic Computing*, vol. 26, n° 4 (2011), p. 463-477.
- POPPER (Karl Raimund), *The logic of scientific discovery*, London, 2002. ISBN : 978-0-415-27843-0.
- QUENTIN (Henri), *Essais de critique textuelle (ecdotique)*, Paris, 1926.
- ROBINSON (Peter), « Why digital humanists should get out of textual scholarship », dans , présenté à Social, Digital, Scholarly Editing, Saskatoon, [s.n.], 2013. URL : https://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship.
- ROOS (Teemu), HEIKKILÄ (Tuomas) et MYLLYMÄKI (Petri), « Computer-Assisted Stemmatology Challenge », 2007 2006. URL : <http://www.cs.helsinki.fi/u/ttonteri/casc/>.. Consulté le 11 janvier 2012.

- SAHLE (Patrick), *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Norderstedt, 2013 (Schriften des Instituts für Dokumentologie und Editorik, 3 vol., n° Band 7-9). ISBN : 978-3-8482-6320-2.
- SAHLE (Patrick), *Catalog of: Digital Scholarly Editions*, Cologne, 2008. URL : <http://www.digitale-edition.de/>.
- SPERBERG-McQUEEN (C. M.), « How to teach your edition how to swim », *Literary and Linguistic Computing*, vol. 24, n° 1 (2009), p. 27-39.
- STÉRIN (Anne-Laure), « Le chercheur « fouille » dans des textes ou dans des données », Billet, [s.d.]. URL : <http://ethiquedroit.hypotheses.org/1528>.. Consulté le 29 août 2017.
- STUTZMANN (Dominique), *ICFHR2016: Competition on the Classification of Medieval Handwritings in Latin Script*, Paris, 2016. URL : <http://icfhr2016-clamm.irht.cnrs.fr/>.
- TEI CONSORTIUM, « TEI P5: Guidelines for Electronic Text Encoding and Interchange », 2015. URL : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>.. Consulté le 9 mai 2015.
- TGIR HUMA-NUM, « Nakala », 2014. URL : <http://www.nakala.fr/>.
- THOMASIN VON ZERKLAERE, *Welscher Gast digital*, éd. J. ŠIMEK, Heidelberg, [s.d.]. URL : <http://digi.ub.uni-heidelberg.de/wgd>.. Consulté le 19 juillet 2016.
- TURSKA (Magdalena), CUMMINGS (James) et RAHTZ (Sebastian), « Challenging the Myth of Presentation in Digital Editions », *Journal of the Text Encoding Initiative*, Issue 9 (janvier 2016). URL : <http://jtei.revues.org/1453>.. Consulté le 3 septembre 2017.
- Queste del saint Graal: Édition numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77)*, éd. C. MARCHELLO-NIZIA, version 0.8.1, Lyon, 2012. URL : <http://portal.textometrie.org/txm/>.