
De l'interopérabilité à la réutilisabilité des éditions électroniques

From Interoperability to Reusability of Digital Scholarly Editions

Ioana Galleron et Fatiha Idmhand



Édition électronique

URL : <http://journals.openedition.org/revuehn/350>

DOI : [10.4000/revuehn.350](https://doi.org/10.4000/revuehn.350)

ISSN : 2736-2337

Éditeur

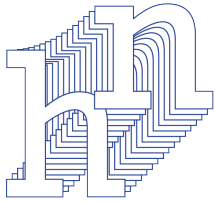
Humanistica

Référence électronique

Ioana Galleron et Fatiha Idmhand, « De l'interopérabilité à la réutilisabilité des éditions électroniques », *Humanités numériques* [En ligne], 1 | 2020, mis en ligne le 01 janvier 2020, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/revuehn/350> ; DOI : <https://doi.org/10.4000/revuehn.350>



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.



De l'interopérabilité à la réutilisabilité des éditions électroniques

From Interoperability to Reusability of Digital Scholarly Editions

Ioana Galleron et Fatiha Idmhand

Résumés

Si la qualité et la pertinence de l'édition scientifique numérique ont fait l'objet de plusieurs réflexions, sa « réutilisabilité » a été, jusqu'à présent, moins étudiée et pensée. Cet article analyse les raisons pour lesquelles la prise en compte de cet aspect est toutefois incontournable, en dépit de la tension dans laquelle il entre avec la poursuite d'objectifs propres à l'équipe éditoriale. Trois modes d'identification du point d'équilibre entre les besoins du ou des réalisateurs de l'édition et ses utilisateurs futurs sont évoqués : l'enquête préliminaire, le travail à partir d'une conceptualisation abstraite de l'activité scientifique et un mélange des deux. Enfin, l'article offre un exemple de réflexion sur la réutilisabilité à partir des « scholarly primitives » de John Unsworth.

While the quality and the relevance of the digital scholarly edition (DSE) have been repeatedly discussed and led to the formulation of various recommendations and grids of analysis, the possible “re-uses” of the published texts have less often been taken into account as a principle guiding the activity of the editorial team. This paper starts from the assumption that this aspect is essential in DSEs, in spite of the difficulties it raises. The scientific objectives of the editor are, indeed, sometimes in conflict with the expectations of the future users. The paper presents three modalities for finding a balance between the needs and wishes of the editorial team, and those of the wider scholarly public: a preliminary survey, an approach from a theoretical standpoint, or a mix of both. In the last part, the paper reflects on the re-uses of a text from the point of view of John Unsworth's “scholarly primitives”.

Entrées d'index

MOTS-CLÉS : humanités numériques, accessibilité, corpus d'auteur, édition électronique, édition savante numérique, évaluation, expérience utilisateur, modélisation de l'activité de recherche, réutilisation

KEYWORDS: digital humanities, accessibility, authorial corpus, digital editing, scholarly digital edition, evaluation, user experience, modelling of scholarly activity, reuse

¹ Avec la multiplication des ressources numériques, à laquelle on assiste surtout depuis les années 2000, une question essentielle revient de plus en plus sur le devant de la scène : comment s'orienter dans cet univers en pleine expansion où les projets les plus divers se côtoient et où les productions les plus disparates réclament, à mesure égale, le titre d'édition numérique ? Conscients de l'importance de la question, plusieurs chercheurs et sociétés savantes ont proposé au fil du temps des grilles pour accompagner, analyser ou évaluer ces productions. On peut ainsi citer les *Guidelines for Evaluating Work in Digital Humanities and Digital Media* de la Modern Language Association (MLA 2012), document accompagnant les *Guidelines for Editors of Scholarly Editions* de la même société savante, les plus récents *Criteria for Reviewing Scholarly Editions* de l'Institut für Dokumentologie und Editorik (IDE¹), ainsi que toute une série de contributions auxquelles il sera fait référence dans cet article. À l'aide de critères et de check-lists, ces contributions cherchent à aider à la fois les éditeurs potentiels et les utilisateurs de leurs produits, à situer sur une échelle de valeur les textes numériques trouvés sur la toile, et prendre conscience de leurs avantages comme de leurs limites. Tout récemment, les auteures de cet article ont eu elles-mêmes à se confronter à un tel exercice dans le cadre de la rédaction d'un guide de bonnes pratiques adressé à la fois aux membres de leur réseau, le consortium CAHIER de l'infrastructure des humanités numériques Huma-Num en France, et à la communauté scientifique en général².

² Il existe toutefois une dimension du travail éditorial sur laquelle les guides mentionnés ci-dessus ne s'attardent pas : celle de la « réutilisabilité ». Qu'il s'agisse des recommandations pour une « édition numérique ouverte », principalement techniques, ou de réflexions épistémologiques sur la spécificité de l'édition électronique par rapport à son équivalent analogique (l'édition papier), il y est rarement question de la prise en compte du futur utilisateur dans la conception et la réalisation même du travail éditorial. La grande majorité des éditeurs se contente de l'anticipation espérée et approximative d'une lecture que quelqu'un, quelque part, fera du texte. Pourtant, la question des usages est cruciale, ne serait-ce qu'en raison de la montée en puissance du critère de l'impact, académique et sociétal, dans l'évaluation des demandes de financement.

³ Cet état de fait est sans doute lié à la perception de la réutilisabilité comme un aspect moins important que la qualité, voire l'accessibilité de l'édition numérique, sur lesquelles les guides s'attardent volontiers. S'y ajoute la difficulté de préfigurer des réemplois, sans s'écarter de la question scientifique qui préside à la réalisation du projet. Comment penser

d'avance des questions de recherche inédites, des curiosités inouïes, susceptibles d'être provoquées par le texte ? Par certains côtés, la prétention est utopique, le propre de la recherche étant de se mouvoir constamment dans de nouvelles directions. Pourtant, il existe des fils d'Ariane sur lesquels on peut tirer pour orienter son travail, non seulement en fonction de soi, mais aussi selon les besoins des autres. Les lignes qui suivent essaieront d'apporter quelques réponses à cette question qui, sans se poser exclusivement dans le milieu des humanités numériques, résonne de façon particulière dans ce champ de recherche. Pour cela, nous analyserons brièvement, dans un premier temps, les grilles d'évaluation existantes, dans le but de comprendre les principes qui les sous-tendent. Dans un second moment, nous nous poserons la question des méthodes grâce auxquelles on pourrait penser et mettre en œuvre la réutilisabilité, puisqu'il est clairement difficile de la circonscrire et de l'anticiper. Enfin, nous passerons en revue, sans prétention d'exhaustivité, quelques conséquences pratiques de la mise en place de ce principe sur les décisions éditoriales.

L'édition numérique et ses évaluations

4 Nombreuses sont les analyses qui ont démontré et rappelé que l'édition électronique n'est pas (ou pas seulement) du « papier augmenté ». Le numérique permet de faire plus (plus de notes, plus de variantes, plus de textes d'accompagnement, etc.) mais surtout de faire autrement (voir, entre autres, Crane, Bamman, et Jones [2008] ; Sinatra et Vitali-Rosati [2014]).

5 À la lumière de ces nouvelles possibilités, quelles sont les principales obligations faites à l'éditeur scientifique numérique, telles qu'elles se reflètent dans la littérature existante sur le sujet ? Les *Guidelines* de la MLA annoncent que « *the scholarly edition's basic task is to present a reliable text* ». Sur cette base, ils déploient une série d'observations au sujet du respect du matériau source, de la théorie du texte sur laquelle se base l'édition, et du médium de publication. Pour Sperberg-McQueen (1996), les éditions électroniques doivent répondre à trois exigences : « *accessibility without needless technical barriers to use ; longevity ; and intellectual integrity* ». Dans ses recommandations aux évaluateurs, Sahle (2014) s'appuie sur le 7^e principe de Sperber-McQueen, qui affirme que « *Electronic editions without apparatus, without documentation of editorial principles, and without decent provision for suitable display are equally unacceptable for serious scholarly work* ». De même, pour Robinson (2013), « *a scholarly edition must, so far as it can, illuminate both aspects of the text, both text-as-work and text-as-document* ».

6 Quoique les formulations divergent et que les accents tombent différemment, ce que ces grilles et propositions ont finalement en commun, c'est de se concentrer, d'une part sur l'éditeur³, et, d'autre part, sur l'objet numérique⁴. L'évaluateur est ainsi invité à se focaliser sur les choix du premier et à apprécier le résultat du processus éditorial. La vie ultérieure et possible des éditions et leurs usages sont peu ou pas pris en compte ; quand ils sont mentionnés, les utilisateurs de l'édition électronique sont d'ailleurs souvent désignés de façon générique comme des « lecteurs » (« *readers* »). Cette étiquette, sans doute inclusive et comode, nous semble marquée par une conceptualisation propre au média

papier, qui fait de la lecture le mode privilégié de « consommation » du texte⁵. Or, cela ne correspond pas toujours à la réalité de l'emploi des éditions numériques scientifiques⁶ comme le montre, entre autres, la pratique du « *distant reading* », qui ne donne virtuellement pas de place, dans un premier temps en tout cas, à la lecture à proprement parler.

⁷ Certes, les guides cités plus haut signalent, à leur façon, la nécessité de tenir compte de l'utilisateur de l'édition. Dans la section 1.2.3. des *Guidelines* de la MLA, il est question de « *desired or potential audience* ». Sperberg-McQueen (1996) appelle dans son texte, qui a fait l'objet d'une communication à une conférence dès 1994, à rendre les éditions numériques accessibles au plus grand nombre, et recommande qu'elles « présente[nt] le même texte de plusieurs façons ». Dans les *Criteria* de l'IDE, on lit même que l'évaluateur d'une édition électronique doit se poser la question des interfaces « *which allow the reuse of the data of the SDE in other contexts* ». Dans les appels à projets, les questions scientifiques soulevées par l'édition doivent justifier leur contribution à l'état de l'art et leur apport à des recherches futures. Cependant, les conséquences pratiques de cette prise en compte sont rarement détaillées, à l'exception des observations sur l'interface de consultation à travers laquelle l'édition est donnée au public⁷. Les *Guidelines* recommandent des formats ouverts, invitent à se demander si « *the edited text can be easily republished, excerpted, or repurposed* », mais immédiatement, des questions sur la lisibilité du texte (« *is it suitable for photographic reproduction ? [...] does it provide PDF or other pretty-printing output ?* ») et sur l'archivage détournent la réflexion des autres usages rapidement énumérés plus haut. Le spectre des utilisations possibles envisagées est bien plus large dans les *Criteria*, mais il n'est présent qu'en filigrane et de façon disparate⁸.

⁸ Ce qui manque dans chacune de ces contributions – dont il ne s'agit pas, encore une fois, de minimiser l'importance ou la qualité –, c'est une réflexion suivie sur l'utilisateur et les réappropriations possibles des textes dans le paradigme numérique, ainsi que sur la façon dont les choix éditoriaux – ou, si l'on veut, les choix éditoriaux et de publication – stimulent ou bloquent ces réappropriations.

Splendeurs et misères de la réutilisation

9 Ancrées dans les pratiques séculaires des disciplines humanistes, les grilles d'analyse existantes continuent donc de privilégier le point de vue de l'éditeur et considèrent le texte numérique comme point d'aboutissement plutôt qu'en tant que possible point de départ. Dès lors, elles prennent peu en compte le fait que les éditions numériques sont appelées, plus que leurs pendants papier, à des évolutions ultérieures, et à subir des transformations dans lesquelles leurs premiers concepteurs peuvent ne plus jouer aucun rôle, comme le montrent Bachimont (2007) et Vitali-Rosati (2018) dans leurs travaux sur l'éditorialisation. En fait, le texte numérique peut être lu, mais aussi coupé, collé, modelé, intégré sous forme de lien hypertexte, annoté et ré-annoté, augmenté, transmis, etc. Jusqu'à un certain point, quelques-unes de ces opérations pouvaient déjà être réalisées à partir d'un texte sur un support matériel ; les généticiens des textes connaissent pléthore d'exemples d'écrivains ayant « extrait », par la découpe, des fragments d'autres textes. Toutefois, une telle pratique a toujours été limitée par la matérialité même du livre, qui conditionne le nombre et le type de réutilisations possibles : on peut prendre une loupe pour lire un passage écrit en petits caractères mais pas en modifier la taille, la forme ou la couleur ; on peut couper les mots d'un paragraphe et les coller sur une autre page au prix de la mutilation de l'objet d'origine ; on peut recopier à la main des extraits, mais l'opération est longue et fastidieuse. Par ailleurs, quels que soient le soin apporté dans la manipulation du livre ou les précautions pour le conserver, l'état de celui-ci se dégrade inexorablement avec le nombre de lecteurs. Tel n'est pas, *a priori*, le sort de la publication numérique. Comme elle est susceptible de supporter autant de lectures, successives ou simultanées, autant de manipulations que désirées, sans aucune perte de la qualité du support⁹, l'amplitude des réutilisations possibles constitue donc un des principaux avantages de l'édition scientifique numérique. Ses évolutions et transformations ne sont plus à déplorer, contrairement à ce qu'on faisait jadis à propos des adaptations, *excerpta* et autres réécritures « défigurant » le « texte d'origine » ou le « vrai texte » ; au contraire, avec certains *caveat* et garde-fous, liés notamment à la nécessité de respecter la paternité intellectuelle, ces réutilisations sont à encourager et à souhaiter.

10 Au cours des dernières années, le conditionnement des réutilisations par l'introduction, ou non, d'une dimension financière lors de la publication a fait l'objet de discussions assez intenses au sein de la communauté scientifique des humanités numériques. L'attention s'est surtout focalisée, à juste titre, sur les profits engrangés par certains éditeurs du fait de l'explosion de la demande d'accès à des ressources, sur fond de montée en puissance de la « *knowledge society* » et de généralisation du Web. Dans une bien moindre mesure, il a également été question de l'exclusion de certaines catégories de lecteurs (chercheurs des pays du tiers-monde, d'institutions moins bien subventionnées, etc.) induite par la barrière du paiement¹⁰. Cependant, l'ensemble de ces discussions et initiatives se concentre sur la démultiplication du nombre d'utilisateurs, et non sur la question des types et du nombre des usages. Certes, on peut considérer que l'utilisation des licences dites « ouvertes » implique une

réflexion sur ces sujets, ainsi qu'une catégorisation des réemplois possibles. Les « *creative commons* » proposent une typologie sommaire des usages (citation, modification, republication), parfaitement opérationnelle dans le contexte juridique où elle est apparue, mais qui nous semble assez insuffisante en lien avec la perspective adoptée ici.

11 Celle-ci se fonde sur le constat que, dans les faits, la réutilisation entre souvent en tension avec les principes mêmes qui sous-tendent la réalisation des éditions numériques. Les objectifs scientifiques des éditeurs sont divers et divergents :

- Le philologue se concentre sur l'établissement du texte, ou d'un texte, à partir des différents témoins, et en cherchant la meilleure façon de rendre compte de la multiplicité des lectures possibles d'un mot, d'une phrase ou d'un passage.
- Le généticien, qui ne vise pas directement l'établissement du texte, recourt au numérique pour donner à voir, dans sa chronologie, le processus de création de l'œuvre et l'ensemble de son dossier de genèse : ébauches, plans liminaires, œuvres consultées et états génétiques.
- L'historien de la littérature veut replacer l'œuvre dans le contexte, et se montre, à cette fin, attentif aux échos, reprises et allusions.
- Le poéticien, le rhétoricien, le stylisticien ou l'esthète s'attachent à l'analyse de la construction, aux récurrences, aux effets plus ou moins subtils des différents dispositifs de l'œuvre.
- Etc.

12 Dans l'univers du papier comme dans celui du numérique, chacune de ces perspectives entraîne des décisions lourdes de conséquences, portant sur la transcription, sur l'annotation, sur les discours d'accompagnement, sur l'organisation même du texte édité, qui laissent peu de place à d'autres approches. La plasticité du milieu numérique a été vue, très tôt, comme une promesse de ne pas avoir à choisir, comme si les différentes perspectives brièvement énumérées plus haut pouvaient être cumulées, « superposées » sur le même texte. Or, l'expérience nous montre de plus en plus que ce n'est pas si simple. En théorie, rien n'empêche le littéraire de se servir du texte préparé par le généticien, ou l'historien d'avoir recours aux productions du philologue ; les éditions XML-TEI, par exemple, sont réputées extensibles et susceptibles de recevoir autant de balises que désiré. Toutefois, dans la pratique, l'information utile aux uns est souvent occultée par les choix des autres, et réciproquement. Un texte de poésie dont les vers et strophes ont été encodés se trouve, du fait même de cette structuration, moins apte à l'analyse syntaxique. De même, la transcription normalisée, utile à l'historien, se fait au prix de l'écrasement de l'information sur les pratiques d'écriture et de segmentation, qui sont fondamentales dans d'autres travaux. Disons qu'en matière d'édition, numérique ou analogique, le bonheur des uns fait souvent le malheur des autres, et réciproquement.

13 Pierazzo (2011) suggère à l'éditeur scientifique une piste intéressante pour trouver le juste milieu entre ses propres objectifs et les besoins des autres : avoir davantage recours à l'enquête afin de prendre la mesure des besoins des utilisateurs. Cependant, elle ne détaille pas les modalités de cette enquête, une opération que l'on imagine peu triviale pour des chercheurs surtout formés à des méthodes d'analyse linguistique ou lit-

téraire. Par ailleurs, l'idée de l'enquête semble tournée vers les besoins des utilisateurs d'un site Web, dont il s'agit de prendre la mesure, et ne porte pas sur les choix de transcription, d'encodage ou d'accompagnement. Il est vrai que le principe de réalité, la limitation des ressources à laquelle toute entreprise d'édition est confrontée, empêche d'aller trop loin dans cette direction (voir aussi Robinson [2013]), et que l'essentiel est avant tout de fournir aux « récepteurs » des explications sur ce qui a été fait (« *decrypting the idiosyncratic habits of textual scholarship* » [Pierazzo 2011]). Mais ne peut-on pas aller un peu plus loin et, de la même manière que l'on recommande un format ouvert, facilement interopérable, afin d'éviter « l'enfermement » du texte dans un cadre technologique inaccessible à d'autres, déterminer une série d'opérations supplémentaires qui peuvent être réalisées sans coût rédhibitoire pour faciliter la réutilisation de l'édition ? Ne faut-il pas s'astreindre à faire, même si le projet éditorial ne semble pas en avoir besoin dans un premier temps, ce qui peut apporter un grand bénéfice à la communauté scientifique dans son ensemble, et même au-delà ? Comment identifier, néanmoins, le juste milieu entre ce que l'on veut faire, ce que l'on pourrait faire pour aider les autres, et ce qui est matériellement possible ?

14 Si l'on suit la piste de l'enquête suggérée par Pierazzo, il conviendrait de créer des cadres de discussion, des forums multidisciplinaires où l'idée d'une édition puisse être soumise et discutée. L'éditeur pourrait y présenter en amont ses choix et ses principes, et entendre de quelle façon ils rencontrent les attentes ou l'intérêt des autres. Il est question, sans doute, d'éditions scientifiques sur les listes de discussion des humanités numériques (comme la liste TEI-eng, par exemple), ainsi que dans différentes sociétés savantes (par exemple, Networked Infrastructure for Nineteenth-Century Electronic Scholarship [NINES]). Cependant, les messages postés sur ces listes portent le plus souvent sur les aspects pratiques et sur les difficultés techniques rencontrées dans la réalisation de l'édition. On n'y voit que rarement des demandes de discussion préliminaire, destinées à sonder les intérêts divergents, les lectures et utilisations différentes que l'on peut faire, à la lumière de telle ou telle discipline, d'un texte candidat à l'édition.

15 Dans une certaine mesure, le consortium français CAHIER offre un tel espace à ses membres, qui sont amenés, dès la demande de labellisation, à présenter leurs projets et les critères technologiques en amont ou au tout début de la réalisation effective. Le projet n'est accepté au sein du consortium que s'il s'engage à respecter les règles d'ouverture classiques du domaine des humanités numériques, telles que l'utilisation de logiciels open source, la réalisation d'éditions libres et ouvertes, etc. Cependant, pour toutes sortes de raisons, le consortium n'a pas encore engagé de discussion sur les possibilités concrètes de réutilisation des éditions, et le nombre de ceux qui se sont engagés dans cette voie reste limité. Leur retour est pourtant fondamental pour prendre la mesure de la valeur des éditions numériques, notamment lorsque celles-ci sont consommatrices des moyens du consortium pour le traitement des données.

16 Pour penser la réutilisabilité, un autre point de départ peut être l'anticipation abstraite des usages possibles. Une telle approche est facilitée par l'intérêt croissant, depuis les deux dernières décennies, pour la conceptualisation de la recherche dans les SHS. Celle-ci a débouché sur

plusieurs propositions de représentations schématiques, parmi lesquelles la plus célèbre, celle des sept « *scholarly primitives* » de Unsworth (2000). Dans son sillage ou en parallèle, à partir d'analyses plus ou moins étendues des pratiques, comme de leur transformation dans le contexte numérique, un certain nombre d'activités « types » des spécialistes des SHS a été conceptualisé. Ainsi, nous proposons, sur la base de telles représentations, de penser les réutilisations possibles d'une édition numérique, et d'interroger la façon dont les choix sous-tendant cette dernière peuvent faciliter ou handicaper lesdites activités.

17 Une telle approche a déjà été utilisée par le passé et s'est montrée fructueuse pour concevoir et évaluer les infrastructures de recherche destinées aux humanistes numériques (Schreibman *et al.* 2013), ou pour identifier des éléments manquants de la panoplie digitale du chercheur 2.0 (Gradmann *et al.* 2016). À notre connaissance, elle n'a pas été appliquée à l'analyse des éditions numériques et à la réflexion sur les meilleures pratiques dans ce domaine. Nous nous proposons de le faire dans la troisième partie de cet article, non pas avec la prétention de révolutionner les recommandations faites à l'éditeur ou les questions suggérées à l'évaluateur (on verra que notre texte retrouve certaines de ces recommandations ou questions, telles qu'elles ont pu être exposées dans les grilles et articles mentionnés plus haut), mais avec l'espoir de pouvoir les compléter et d'explicitier leur enjeu à la lumière de l'application systématique du principe de réutilisabilité.

Un modèle conceptuel de l'activité de recherche

18 Dans son article séminal, John Unsworth fait des opérations de découverte, d'annotation, de comparaison, de référence, d'échantillonnage, d'illustration (ou d'exemplification) et de représentation les bases de l'activité scientifique dans les sciences humaines, et, plus particulièrement, de l'analyse de texte. Sa proposition a été discutée à de multiples reprises (Benardou *et al.* 2010) et a suscité plusieurs questions quant aux définitions précises des opérations énumérées. De nombreuses réorganisations du modèle ont été proposées, par exemple dans Hennieke *et al.* (2017).

19 C'est à partir de tels modèles que nous proposons de penser la réutilisabilité en tant que principe de conception et d'analyse des éditions numériques, et plus largement des ressources créées par les sciences du texte¹¹. Pour faciliter la lisibilité, nous représentons l'activité de recherche comme un processus linéaire, tout en étant conscientes qu'elle comporte, en réalité, de nombreux retours en arrière (ou boucles de rétroaction), inévitables et même indispensables pour aller d'une question de recherche à la publication qui rend compte des réponses apportées. Une nouvelle question de recherche, issue ou liée à ces travaux, peut nécessiter la reprise des données, la génération de nouvelles données ou le croisement de ces données avec d'autres.



20 Il nous semble que l'édition numérique peut venir nourrir chacune des étapes de ce processus, avec plus ou moins de succès en fonction des choix effectués par ses concepteurs.

21 Ainsi, la publication même d'un texte, ou d'un ensemble de textes, peut constituer le point de départ de la formulation d'une question de recherche ; aussi l'édition traditionnelle ne s'est-elle pas trompée, en faisant de la mise en avant de l'intérêt du texte, dans la préface (ou autre lieu du paratexte), un passage quasi obligé. Le discours d'accompagnement suggère des pistes et essaie de susciter la curiosité du lecteur ; les éditions numériques ne devraient pas s'en priver, pas plus que leurs pendants « papier ». Or, dans les faits, on a plus de mal à intéresser les lecteurs à une introduction insérée dans le header d'une publication en XML-TEI ; on préfère hélas souvent fonctionner « en partie double », en publiant l'édition d'un côté, et en recourant à un canal traditionnel de diffusion (revue, colloque, actes, etc.) en parallèle afin de faire connaître le produit numérique mis sur le marché. Malgré l'émergence de nouveaux outils de moissonnage (comme WebOAI¹²), d'exposition des données (comme Isidore ou Europeana) ou de revues sur le Web (comme RIDE¹³), le numérique n'a pas encore véritablement réussi à créer ses propres espaces de publicité pour les éditions qu'il rend possibles. On ne s'attardera cependant pas beaucoup sur cet aspect, ni sur tous les différents types d'accompagnements de l'édition (notes, commentaires, annexes, etc.), qui peuvent mener à l'émergence d'une idée, avec des suites potentiellement fertiles. En revanche, la contribution de l'édition numérique à la seconde phase d'exploration, menant à la constitution d'un corpus, mérite qu'on traite de façon plus étendue des choix de l'éditeur, car c'est à ce stade qu'ils peuvent avoir leurs premiers effets sur les réemplois possibles.

22 En effet, l'utilisation de standards internationaux de description bibliographique, ainsi que la création de métadonnées riches, basées sur l'usage de vocabulaires contrôlés et normalisés, font partie des bonnes pratiques favorisant le réemploi. L'identification de l'édition et l'évalua-

tion de sa pertinence dépendent largement de ces opérations, de la façon dont elles « parlent » du texte et de quelques-unes de ses caractéristiques (sujet, conditions de production, taille, encodage des caractères, etc.). Plus la qualité des métadonnées est élevée, meilleurs sont le repérage de l'édition par des moteurs ou logiciels de recherche, ainsi que son inscription dans un catalogue de ressources. Il est vrai que l'identification de la ressource ne dépend pas uniquement de son « créateur », mais également des institutions chargées de constituer lesdits catalogues et listes de ressources, selon des modalités qui sont, elles-mêmes, en cours de mise en place ou de fiabilisation ; dès lors, il est probable qu'en dépit des efforts de collecte et de signalement effectués à différentes échelles (consortium TEI, plateforme Isidore, et autres), un chercheur intéressé par des éditions numériques dans une certaine langue et une certaine période puisse rater des ressources pertinentes pour sa recherche. Dans cette perspective, prendre en compte l'utilisateur suppose de vérifier, d'une part, que les métadonnées de l'édition respectent les critères de normalisation requis pour un meilleur moissonnage, et de s'assurer, d'autre part, que l'édition est bien moissonnée par un aussi grand nombre de moteurs de recherche que possible. En attendant la stabilisation des procédures de catalogage numérique, peut-être en même temps que l'émergence d'un écosystème éditorial prenant pleinement en charge la publication de sources numérisées selon des standards universitaires, il n'existe pas de moyen d'automatiser cette tâche.

23 Par ailleurs, l'agrégation d'une édition à un corpus va être plus ou moins facilitée par les choix techniques effectués, et en premier lieu par la possibilité matérielle offerte aux lecteurs de télécharger la ressource. En effet, nombreuses sont les éditions numériques qui « lient » l'utilisateur à une interface comme point unique d'accès et moyen exclusif pour parcourir le texte. Si, dans certains cas, des problèmes de droit se posent et empêchent la libre diffusion des fichiers, il est assez fréquent que cette exposition limitée et encadrée soit plutôt le résultat d'une difficulté à imaginer un utilisateur souhaitant s'affranchir de l'interface dans le but d'explorer les contenus avec ses propres moyens ou en fonction d'une question de recherche que l'éditeur, justement, n'avait pas anticipée. Un tel site Web, souvent construit avec d'importants moyens financiers, devient dans ces cas une barrière, et la « prise en compte de l'utilisateur » synonyme d'une volonté d'orienter son regard en l'empêchant de poser au texte ses propres questions.

24 De plus, il n'est probablement pas inutile de rappeler que la possibilité d'agréger une ressource à un corpus, lorsqu'elle est téléchargeable, dépend grandement de son format : image ou PDF, TXT, HTML ou XML ? Comme il a déjà été observé, l'interopérabilité et l'ouverture sont des aspects clés de l'édition numérique. Le texte préparé doit être à la fois manipulable informatiquement et libéré de la servitude à une technologie particulière. La proposition de différents formats, adaptés à des usages différents (PDF pour la lecture, TXT pour les explorations textométriques, XML pour enrichissement ultérieur avec de nouvelles balises) apparaît comme une bonne pratique. Certes, un fichier XML peut toujours être converti dans d'autres formats, mais le trouver préparé d'avance économise du temps aux chercheurs suivants et permet d'allouer des ressources à d'autres activités.

Si la question des choix permettant l'inclusion de la ressource dans les corpus constitués par d'autres a déjà fait l'objet de discussions et de recommandations par le passé, l'apport d'une édition numérique aux phases de préparation et d'extraction des données a beaucoup moins attiré l'attention, et s'avère sans doute plus critique. Pourtant, une fois le texte récupéré, dans un format manipulable informatiquement, les traitements et enrichissements dont il a pu faire l'objet peuvent être d'une importance capitale dans la stimulation, ou le blocage, de réutilisations futures. Sans doute chaque projet de recherche est-il unique, surtout dans les sciences du texte, où l'on pratique peu et envisage encore moins la reprise pure et simple de données générées ailleurs pour les interroger sur de nouvelles bases, quitte à les recombinaisonner avec d'autres. Aussi le passage du corpus aux données à interpréter donne-t-il lieu souvent, dans les humanités numériques, à des opérations de redécoupage et de ré-encodage, à des travaux de raffinement et de réorganisation, sur lesquelles il est difficile d'anticiper. Selon nous, il existe pourtant trois points sur lesquels une attention particulière peut s'exercer, parce que les décisions prises et les opérations effectuées peuvent faire gagner du temps ou peser considérablement sur la ré-employabilité des textes. Ces trois points sont : la transcription, le traitement des entités nommées et la mise en lien.

En effet, une transcription exclusive en version modernisée écrase définitivement la spécificité linguistique du texte d'origine, imposant un nouveau travail de saisie aux chercheurs intéressés par ces aspects. À l'inverse, une édition proposant une transcription diplomatique, ainsi qu'une normalisation ou modernisation, sert à la fois l'objectif de favoriser la lecture cursive et celui de pouvoir étudier les pratiques orthographiques ou morphosyntaxiques d'un locuteur donné, à une période donnée. Pour éviter cet écueil, nous suggérons par conséquent que l'édition numérique élémentaire¹⁴ intègre les deux modalités de transcription.

L'identification des entités nommées est une opération chronophage qui n'est peut-être pas nécessaire au projet d'origine ; pourtant, elle peut à son tour rendre des services considérables à la communauté scientifique dans son ensemble. On n'imagine pas d'édition papier sans notes de bas de page ou prosopographie, sans élucidation des noms de personnes, lieux, moments et titres dont il est question dans le texte (lorsque, bien entendu, cette identification est possible) : pourquoi en serait-il autrement dans l'édition numérique, du moins si elle prétend se situer à un certain niveau de sophistication¹⁵ ? D'autant plus que le Web offre de multiples bases de connaissances et que l'attribution d'identifiants pérennes permettant d'inscrire l'entité dans un réseau de « données liées » est dorénavant possible. L'identification des entités nommées permet, plus largement, de faire communiquer l'édition proposée avec d'autres ressources, et donc de l'inscrire dans un ensemble de discours, en réalisant l'idéal de l'hypertexte que décrivait déjà McGann en 2001.

Il existe certainement d'autres aspects du texte qui mériteraient d'être préparés préalablement à sa publication. Les linguistes peuvent souhaiter que les formes soient étiquetées et lemmatisées ; les philologues pouvoir aligner fac-similés et texte transcrit, ce qui peut amener à des choix parfois radicaux d'encodage de la structure¹⁶. Pour répondre à ces

attentes, les questions qui se posent, au-delà de celle du point d'équilibre scientifique et économique du projet, portent sur les techniques permettant de réaliser ces multiples opérations, tout en évitant les conflits de hiérarchie qui peuvent affecter les fichiers XML. Pour cela, une piste suggérée de plus en plus fréquemment est celle de l'encodage en stand-off, fondé sur une version « agnostique » du texte qui n'en reproduit que les constituants de base (pour faire simple : les mots) et qui utilise des fichiers à part pour donner à voir les différentes structurations¹⁷ et récurrences, au moyen de pointeurs vers les identifiants uniques des constituants de base. Sans entrer dans la discussion sur les avantages et les inconvénients d'une telle solution, nous retenons surtout le fait qu'elle suggère, et non sans une dose de paradoxe, que la réutilisabilité appelle à fournir deux ressources : l'une, aussi richement encodée que possible, et l'autre, plus « pauvre », c'est-à-dire une version à partir de laquelle d'autres travaux puissent être menés. Ainsi, la réutilisabilité ne passe pas seulement par l'accès libre et ouvert à l'édition numérique philologique, génétique, etc. réalisée *in fine*, mais également par la mise à la disposition d'un texte simple, à partir duquel de nouvelles opérations d'enrichissement pourront être menées.

29 En suivant les phases du processus de recherche schématisé plus haut, se pose ensuite la question de l'interface de consultation et de la façon dont elle permet de prendre en compte l'utilisateur. Comme observé dans les parties précédentes, l'importance de cet aspect a été généralement bien mesurée, à la fois par les auteurs d'éditions scientifiques et par les pairs proposant des recommandations et des guides pour un tel travail. Ce qu'on peut y ajouter toutefois, c'est la nécessité de mieux distinguer entre le rôle de l'interface dans la phase de consultation, d'exploration des ressources, menant à la constitution d'un corpus, et celui qu'elle peut jouer dans l'extraction des données et leur interprétation. Si le premier rôle est fondamental et impose la présence d'une interface comme élément constitutif d'une édition numérique, le second est optionnel et dépend dans une large mesure des ressources financières et humaines du projet. L'accès au fichier XML « derrière » l'interface peut, dans la plupart des cas, répondre de façon tout à fait satisfaisante à ce type de réemploi.

30 Enfin, la publication de la recherche, sous quelque forme que ce soit (présentation dans un colloque, communication sur un support papier, utilisation d'un format numérique) exige que l'édition soit citable. La stabilité de l'édition est donc un facteur essentiel pour sa réutilisabilité. Cet aspect n'est pas sans poser problème dans un contexte numérique caractérisé par l'aspiration à une amélioration continue du produit (culture du « *perpetual prototype* », selon Pierazzo). Cependant, tout comme l'impression de l'édition papier met un terme, sans doute provisoire, au progrès du savoir à propos d'un texte, l'édition numérique doit se fixer des paliers ou des états cibles, destinés à représenter un certain état d'avancement du travail, et à pouvoir servir de base pour les références ultérieures. Tout comme dans l'univers du papier, il n'est pas évident de décider à partir de quel moment les corrections et transformations apportées à la manifestation numérique de l'œuvre constituent une nouvelle version, ou bien une extension de l'ancienne ; mais, quelle que soit la décision sur ce point, il ne reste pas moins fondamental de stocker de façon pérenne des « arrêts sur l'image » périodiques, auxquels

la communauté scientifique puisse se référer et vers lesquels les liens continuent à pointer dans le long terme. Dès lors, l'édition numérique requiert elle aussi le respect des exigences d'archivage et de pérennité. Ainsi, la stabilité des textes édités passe également, comme la publication d'archives, par des identifiants uniques, principe recommandé par les communautés internationales. La granularité de cette identification est un élément qualitatif important décrit par les différents guides cités (*Guidelines*, les *Criteria*, mais aussi Bilder, Lin, et Neylon 2015) ; sans rechercher une utopique exhaustivité, sans doute l'édition numérique peut-elle se contenter, a minima, d'un handle ou d'une URI, moyen d'obtenir une URL durable et donc d'assurer, indépendamment de l'interface de consultation, la citabilité de l'édition.

Conclusions

31 La production de ressources numériques, surtout textuelles, attire de plus en plus d'acteurs, et a de beaux jours devant elle, même si les humanités numériques sont déjà appelées à se tourner vers de nouveaux horizons et de nouveaux défis avec la seconde vague (Schnapp, Lunenfeld, et Presner, s. d.). L'idée d'un corpus constitué significativement (voire majoritairement ou exclusivement) de « documents » trouvés sur le Net, autant que la fiabilité de ces derniers, ne sont aujourd'hui plus controversées. La bibliothèque virtuelle s'enrichit constamment de nouveaux ouvrages et permet aux chercheurs de s'affranchir de plus en plus des contraintes de temps et d'espace qui pèsent sur l'accès aux sources. Dans un tel contexte, la question de la réutilisabilité est appelée à se poser de façon de plus en plus aiguë aux communautés scientifiques, disciplinaires et transdisciplinaires.

32 L'ajustement entre ses propres objectifs de recherche, tels qu'exprimés dans l'édition effectuée, et les besoins des autres reste une question délicate, mais incontournable. L'avenir montrera si, pour la résoudre, les chercheurs s'engageront plutôt vers la mise en place d'enquêtes préliminaires, vers des *assessments* au moyen de modèles abstraits, vers une combinaison de ces deux modalités, ou encore s'ils inventeront d'autres voies de partage et d'échange.

Bibliographie

Bachimont, Bruno. 2007. « Nouvelles tendances applicatives. De l'indexation à l'éditorialisation ». Dans *L'Indexation multimédia : description et recherche automatiques*, édité par Patrick Gros. Paris : Lavoisier.

Benardou, Agiatis, Panos Constantopoulos, Costis Dallas et Dimitris Gavrilis. 2010. « A Conceptual Model for Scholarly Research Activity ». Communication à *iConference 2010*. University of Illinois at Urbana-Champaign. <https://www.ideals.illinois.edu/bitstream/handle/2142/14945/benardou.pdf>.

Bilder, Geoffrey, Jennifer Lin et Cameron Neylon. 2015. « Principles for Open Scholarly Infrastructures-v1 ». Figshare. <https://doi.org/10.6084/m9.figshare.1314859>.

Charles, Michel. 1995. *Introduction à l'étude des textes*. Paris : Seuil.

Crane, David, David Bamman et Alisson Jones. 2008. « ePhilology: when the books talk to their readers ». Dans *A Companion to Digital Literary Studies*, édité par Ray Siemens et Susan Schreibman. Oxford : Blackwell.

Galleron, Ioana, Marie-Luce Demonet, Cécile Meynard, Fatiha Idmhand, Elena Pierazzo, Geoffrey Williams, Julia Roger et Pierre-Yves Buard. 2018. « Les publications numériques de corpus d’auteurs – Guide de travail, grille d’analyse et recommandations ». Consortium CAHIER. <https://halshs.archives-ouvertes.fr/halshs-01932519>.

Gradmann, Stefan, Steffen Henniecke, Gerold Tschumpel, Kristin Dill, Klaus Thoden, Alois Pichler, Christian Morbindoni et Juliane Stiller. 2016. « Modelling the Scholarly Domain beyond Infrastructure ». Communication à *Jahrestagung DHd 2016*, Université de Leipzig.

Greatley-Hirsch, Brett, Peter J. Smith, Elisabetta Tarantino, Domenico Lovascio, Shirley Bell, Christian Griffiths, Kate Wilkinson, Sheilagh Ilona O’Brien et Louise Powell. 2017. « VII. Shakespeare ». *The Year’s Work in English Studies* 96 (1) : 367-465. <https://doi.org/10.1093/ywes/max008>.

Hawkins, Kevin, Michelle Dalmau, Elli Mylonas et Syd Bauman, éd. 2018. « Best Practices for TEI in Libraries ». <https://tei-c.org/extra/teiinlibraries/>.

Henniecke, Steffen, Stefan Gradmann, Kristin Dill, Gerold Tschumpel, Klaus Thoden, Christian Morbindoni et Alois Pichler. 2017. « Research Report on DH Scholarly Primitives ». ICT-PSP-297274. Commission européenne. https://dm2e.eu/files/D3.4_2.0_Research_Report_on_DH_Scholarly_Primitives_150402.pdf.

McGann, Jerome. 2001. *Radiant Textuality. Literature after the World Wide Web*. Basingstoke et New York, Palgrave Macmillan.

MLA, éd. 2012. « Guidelines for Evaluating Work in Digital Humanities and Digital Media ». Modern Language Association. <https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media>.

Pierazzo, Elena. 2011. « A Rationale of Digital Documentary Editions ». *Literary and Linguistic Computing* 26 (4) : 463-477. <https://doi.org/10.1093/lc/fqr033>.

Robinson, Peter. 2013. « Towards a Theory of Digital Editions ». *Variants* 10 : 105-131.

Sahle, Patrick. 2014. « Kriterien für die Besprechung digitaler Editionen ». Institut für Dokumentologie und Editorik. <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/>.

Schnapp, Jeffrey, Peter Lunenfeld et Todd Presner. s. d. « Digital Humanities Manifesto ». http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf.

Schreibman, Susan, Stefan Gradmann, Steffen Henniecke, Tobias Blanke et Sally Chambers. 2013. « Beyond Infrastructure – Modelling Scholarly Research and Collaboration ». Communication à *Digital Humanities 2013*, University of Nebraska-Lincoln, juillet. hal-00801439.

Sinatra, Michaël et Marcello Vitali-Rosati. 2014. *Pratiques de l’édition numérique*. Montréal : Presses de l’université de Montréal.

Sperberg-McQueen, C. Michael 1996. « Textual Criticism and the Text Encoding Initiative ». Dans *The Literary Text in the Digital Age. Editorial Theory and Literary Criticism*, édité par Richard J. Finneran. Ann Arbor : University of Michigan Press : 37-62.

Unsworth, John. 2000. « Scholarly Primitives : What Methods Do Humanities Researchers Have in Common, and How Might our Tools Reflect This ». Communication au *Symposium on Humanities Computing formal methods experimental practice*, Londres, 13 mai. <http://people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>.

Vitali-Rosati, Marcello. 2018. *On Editorialization : Structuring Space and Authority in the Digital Age*. Amsterdam : Institute of Network Cultures.

Notes

1 Nous faisons ici référence à la version 1.1 de ces *Criteria* conçus par Sahle (2014).

2 Le guide a été réalisé dans le cadre du groupe de travail « EVENT » (Évaluation et valorisation des éditions numériques de textes). Il a été publié à l’automne 2018, il est accessible sur la page internet du groupe de travail EVENT (<https://cahier.hypotheses.org/categorie/groupe-event>) et sur HAL-SHS (Galleron *et al.* 2018).

3 Voir en ce sens la fréquence du syntagme « the editor » dans les *Guidelines* de la MLA.

4 C'est plutôt le choix des *Criteria* de IDE, qui font moins référence à la figure de l'éditeur qu'à l'édition réalisée, à en juger par la fréquence du syntagme « *the Scholarly Digital Edition (SDE)* » ou « *the project* ».

5 Ce qui ne signifie pas que, très tôt, l'édition scientifique papier n'a pas cherché à prendre en compte des usages qui ne supposent pas la lecture, ou, plus exactement, qui se contentent d'une lecture ciblée, fonctionnelle, « pauvre » si l'on veut, à l'opposé de la lecture suivie, attentive et qui impose de parcourir l'ensemble du texte (ou sa plus grande partie) pour faire sens. De même, cette proposition ne méconnaît pas un des grands avantages de l'édition électronique par rapport au papier, qui est précisément de pouvoir fournir un texte enfin lisible, quelle que soit la quantité de notes et variantes qui lui est attachée (voir, par exemple, les considérations de McGann [2001, 57 et suivantes]).

6 Deux observations complémentaires sur ce point. La première est que l'on peut déplorer l'absence d'un observatoire sur la lecture de textes électroniques, qui puisse nous renseigner de façon fiable sur le public qui parcourt en ligne ou télécharge nos éditions. La seconde est que l'édition électronique portant le plus souvent sur des textes que les éditeurs papier voient comme comportant des caractéristiques les rendant peu rentables dans le cadre d'une édition traditionnelle (trop d'appareil, trop de notes, difficulté de s'orienter dans un maquis de variantes, etc. – voir Greatley-Hirsch *et al.* [2017]), ces textes sont « *scholarly* », non seulement parce que faits par des scientifiques, mais aussi parce qu'ils vont toucher en majorité un public scientifique (complémenté, dans certains cas, par un public étudiant). La limitation *de facto* de ce public n'est pas inintéressante dans la perspective de la réflexion sur les réutilisations possibles du texte dont il est question dans cet article.

7 Est significative en ce sens la suite du second principe de Sperberg-McQueen : « *Electronic scholarly editions [...] should not require a particular type of computer, or a particular piece of software: unnecessary technical barriers to their use should be avoided.* » Dans le même sens, est parlante la partie 2.7. de l'article de Pierazzo (2011), intitulée « *The needs of the others* », qui met finalement en avant « *website creators and, among them, digital editors* » et évoque une prise en compte des besoins des autres lors de la publication. Nous reviendrons sur cet article.

8 Voir, dans ces *Criteria*, les points 4.2. à 4.4. (« *Interface and Usability* », « *Browse* » et « *Search* »), mais aussi 4.9. (« *Technical interfaces* », point déjà évoqué) et jusqu'à 4.13.

9 Sans doute, le médium électronique a ses propres lois d'obsolescence, ses propres sources de dégradation (tout comme le papier et l'encre, en fonction de leur composition chimique). Nous sommes confrontés tous les jours à la périssabilité des ressources électroniques, ne serait-ce que sous la forme des « liens cassés » sur le Web. Mais cette périssabilité n'est pas indexée sur le nombre d'utilisateurs d'une ressource, sur le nombre d'accès à elle ; au contraire, la multiplication des usages joue plutôt en faveur de sa pérennisation : dans la sphère numérique, une utilisation fréquente est un argument de poids pour la mise en place de procédures de maintenance.

10 Des appels à une science ouverte, suivis d'initiatives politiques, se sont succédé, et le mouvement n'est visiblement qu'à ses débuts.

11 Par « sciences du texte », on entend ici à la fois la linguistique, l'histoire, l'analyse littéraire, la philologie, bref, l'ensemble des disciplines dans lesquelles le texte est un objet principal d'investigation, et non seulement un outil parmi d'autres.

12 WebOAI est un outil libre développé par le consortium CAHIER de Huma-Num qui implémente un entrepôt (*repository*) OAI et propose une interface pour consulter le catalogue des corpus numériques du consortium CAHIER. Il s'agit d'un agrégateur « intermédiaire » qui agrège les notices de la communauté CAHIER pour les déclarer ensuite aux moissonneurs généralistes comme Isidore, Gallica et Europeana. Il se base sur les *teiHeader* des fichiers numériques et définit les correspondances avec le schéma OAI CAHIER des principaux moissonneurs. WebOAI est une initiative communautaire pour faire converger les pratiques d'encodage TEI des métadonnées (<http://weboai.cahier.huma-num.fr/>).

13 RIDE, a Review Journal for Digital Editions and Resources : <https://ride.i-d-e.de>.

14 Pour plus d'informations sur la différence entre l'archive éditorialisée, l'édition élémentaire et l'édition enrichie, voir Galleron *et al.* (2018).

15 Voir la définition du niveau 4 (ou 5) de l'édition numérique tel qu'il est décrit dans Hawkins *et al.* (2018) ou la présentation du type 3 (édition enrichie) dans Galleron *et al.* (2018).

16 Robinson (2013) donne l'exemple d'un projet qui permet de « feuilleter » une archive page par page, mais pas selon les chapitres, les parties ou autres éléments « sémantiques » du texte.

17 Structurations en unités prédéterminées (paragraphe, chapitre, parties, etc.) ou unités non prédéterminées (incipit, péripétie, etc.). Voir Charles (1995).

Auteurs

Ioana Galleron

UMR 8094 Lattice, Montrouge, et université Sorbonne-Nouvelle, Paris, France

Ioana Galleron est professeure de littérature française et d'humanités numériques à l'université Sorbonne-Nouvelle. Elle s'intéresse en particulier à l'encodage en XML-TEI du théâtre du XVIII^e siècle, ainsi qu'à l'exploration des textes assistée par ordinateur (textométrie, stylométrie, XQuery).

ioana.galleron@gmail.com

Fatiha Idmhand

UMR 8132 ITEM, université de Poitiers, Poitiers, France

Fatiha Idmhand est professeure à l'université de Poitiers, chercheur à l'Institut des textes et manuscrits modernes et membre de l'équipe Archivos. Ses recherches portent sur la littérature hispanique contemporaine (Río de la Plata surtout) et sur les processus de création des œuvres. En lien avec ces recherches sur les manuscrits, leur préservation et leur analyse, elle développe des méthodes pour des approches informatisées des sciences du texte.

fatihaidmhand@yahoo.es

Droits d'auteur



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).