



consortium  
cahier

*Corpus d'auteurs pour les humanités : informatisation, édition, recherche*

# LES PUBLICATIONS NUMERIQUES DE CORPUS D'AUTEURS

- guide de travail, grille d'analyse et recommandations -

Ioana Galleron, Marie-Luce Demonet, Cécile Meynard,  
Fatiha Idmhand, Elena Pierazzo, Geoffrey Williams, Pierre-  
Yves Buard, Julia Roger<sup>1</sup>



## INTRODUCTION

Depuis plusieurs années, le nombre des publications numériques de ressources textuelles augmente de façon continue. De nombreux textes sont rendus disponibles via internet, soit de façon autonome, soit regroupés sous forme de collections, de bibliothèques et d'archives. Les porteurs de ces projets de numérisation, qu'ils soient novices ou expérimentés, se posent tous, à un moment ou à un autre, la question des meilleures pratiques en matière de réalisation, de publication et de préservation à long terme des éditions scientifiques numériques. Quels sont les normes et les standards - émergents ou établis - à respecter dans le domaine de la publication numérique? Quelles questions se posent à chaque étape du processus de réalisation, et où trouver des réponses pertinentes? Fondé sur l'expérience éditoriale de ses auteurs, sur l'observation de projets éditoriaux menés en Europe ou ailleurs, ainsi que sur une réflexion collective dans le cadre du groupe EVENT du Consortium CAHIER (HumaNum), ce guide se propose d'apporter une réponse à ces questions. Il souhaite venir en aide :

- aux chercheurs qui souhaitent se lancer dans des projets de publication électronique ;
- à ceux qui voudraient analyser et améliorer les publications électroniques qu'ils ont déjà réalisées ;
- aux évaluateurs de publications numériques.

Pour ce faire, il décrit trois types de publications numériques et cinq dimensions fondamentales (ou « caractéristiques mesurables ») que nous avons identifiées. Les frontières entre les trois types mentionnés ne sont pas étanches, et les résultats d'un même projet peuvent parfois relever de types différents, en fonction des besoins différenciés des concepteurs, de l'état d'avancement du projet et des ressources allouées à sa réalisation. Ils constituent toutefois, à nos yeux, des états assez stables et dotés de suffisamment de caractéristiques pour qu'ils puissent offrir des points de repère, contribuant ainsi à la structuration des travaux. De la même manière, en dépit d'une certaine superposition des problématiques, chacune des cinq dimensions pose des défis intellectuels et techniques spécifiques, qui méritent d'être discutés à part.

Outre la description des trois types de publication et des cinq dimensions mentionnées, nous proposons, à la fin de ce document, un tableau récapitulatif de tous les éléments décrits dans le guide : il est susceptible de servir de « check-list », ou de grille pour l'analyse des publications numériques.

## I. LES TROIS TYPES DE PUBLICATION

Nous utilisons ici le terme « types » mais il aurait été également possible de parler de « niveaux » entre lesquels une publication numérique peut évoluer, comme c'est le cas dans le guide des *Best Practices for TEI in Libraries* (2018)<sup>1</sup>. En effet, hormis quelques rares exceptions, les caractéristiques identifiées pour le premier type de publication (désigné ici comme « archive éditorialisée ») se retrouvent également dans le second type (appelé « édition de lecture ») et dans le troisième (nommé ici « édition enrichie »).

Il est à noter que, à la différence du guide des *Best Practices for TEI in Libraries*, les publications numériques en XML/TEI ne sont pas les seules prises en considération dans ce guide. En effet, d'autres options éditoriales et techniques, susceptibles d'être utilisées par les chercheurs, ont été prises en compte, comme par exemple la publication via un CMS. Toutefois, pour garantir l'accessibilité et l'évaluation des publications décrites ci-après, il est fortement recommandé d'utiliser les standards de description de données, partagés par les communautés académiques internationales. Un métalangage de description de documents tel que le XML<sup>2</sup> permet la structuration poussée des contenus manipulés, ainsi que leur exploitation dans des contextes variés. En outre, l'utilisation de vocabulaires de référence du type TEI<sup>3</sup> (pour l'encodage de documents textuels ou nativement numériques) ou EAD<sup>4</sup> (pour l'encodage des objets de recherche archivistiques) présente de nombreux atouts : dans le cadre d'un projet de publication qui a vocation à être largement diffusé, *a fortiori* dans un contexte d'Open Access, l'usage de vocabulaires standardisés est préférable par rapport à la création de vocabulaires maison *ad hoc*, potentiellement équivoques et abscons.

### I.1. L'ARCHIVE EDITORIALISEE

La notion d' « éditorialisation » d'une archive désigne ici un processus qui implique, d'une part, une scénarisation même minimale des ressources, avec un cheminement organisant le « vrac » des documents et, d'autre part, une activité qui engage une responsabilité scientifique. À la différence de ce qui peut être réalisé par une bibliothèque ou un centre d'archives pour organiser des contenus dans un environnement numérique, l'éditorialisation scientifique est orientée par une question de recherche qui se reflète dans le choix des documents, dans les médiations choisies, dans les métadonnées et dans les éventuels discours d'accompagnement.

Les « archives éditorialisées » mettent donc à la disposition de la communauté scientifique :

- soit une collection d'images (publication en fac-similé) ;
- soit une collection de textes « bruts » (textes générés par OCR ou obtenus par saisie manuelle) ;
- soit les deux, associés à l'aide de divers dispositifs.

<sup>1</sup> <http://www.tei-c.org/SIG/Libraries/teilibraries/>.

<sup>2</sup> Xtensible Markup Language. Le XML est une recommandation du W3C [<https://www.w3.org/XML/>].

<sup>3</sup> *Text Encoding Initiative* [<http://www.tei-c.org/>]. Voir aussi Lou Burnard, *What is the Text Encoding Initiative? How to use Intelligent Markup to Digital Resources*, Marseille, OpenEdition Press (Encyclopédie numérique), 2014 [<http://books.openedition.org/oep/426>].

<sup>4</sup> *Encoded Archival Description* [<https://www.loc.gov/ead/>].

L'objectif de ce type de publication est avant tout l'accès au texte, et le choix fréquent d'un format image en est une preuve. Aussi est-il particulièrement approprié quand il s'agit de permettre, le plus rapidement possible et avec un minimum de moyens, la consultation d'une ressource rare, en mauvais état de conservation, éloignée ou, enfin, difficilement accessible pour toute sorte de raisons. Par ailleurs, ce type de publication est souvent utilisé pour mettre en avant la cohérence d'une série de documents, souvent provenant de sources diverses, et dont la publication facilite la perception comme un ensemble.

Ce type de publication correspond approximativement aux niveaux 1 et 2 décrits dans le document *Best Practices for TEI in Libraries*. Ainsi, un texte obtenu grâce à l'océrisation et nettoyé des erreurs les plus grossières (comme, par exemple, des titres courants insérés dans le texte, des sauts de ligne intempestifs, etc.) relève de ce type de publication. Toutefois, il est essentiel que **ce premier type de publication** soit accompagné de métadonnées. Il s'agit là d'une étape qui précède une édition numérique plus élaborée.

## I.2. L'ÉDITION DE LECTURE (« *reading edition* »)

Que le texte ait été acquis par océrisation ou par transcription, il a été, pour **ce deuxième type de publication**, relu, corrigé et nettoyé des erreurs humaines ou de celles qui sont issues de la reconnaissance automatique de caractères. Ce type de publication va au-delà de la mise à disposition du texte : elle l'accompagne de toute une série d'éléments qui le mettent en perspective. Les choix d'édition, de transcription, de lecture et d'interprétation du ou des chercheurs qui ont supervisé le projet sont documentés de façon rigoureuse. Tirant parti des avantages du document numérique, les utilisateurs de ce type de publication peuvent suivre un parcours fondé sur l'interrogation avec des mots clés, ils peuvent analyser ou fouiller le texte à l'aide d'outils numériques, en extraire des informations qui répondent à leurs propres questions de recherche, etc. En outre, ce type d'édition est évolutif : même si le nombre et le type d'informations susceptibles d'être extraites restent limités, la réutilisation de l'édition et son enrichissement sont tout à fait possibles.

On peut comparer ce type de publication à ce qui est décrit comme le niveau 3 du guide des *Best Practices for TEI in Libraries*. Il peut s'agir d'une édition en XML/TEI qui se limite à l'encodage des grandes structurations du texte et fournit des informations de base, sans toutefois avoir recours à toute une série de balises optionnelles, permettant une manipulation poussée du texte.

## I.3. L'ÉDITION ENRICHIE

Le **troisième type d'édition** propose un texte profondément enrichi d'informations documentaires et contextuelles. En outre, celui-ci peut avoir été préparé de façon à permettre un affichage selon différents critères éditoriaux. L'édition enrichie permet souvent l'interrogation par facettes et l'exploitation étendue des données. Enfin, ce type d'édition suit nécessairement des pratiques harmonisées et se réfère à des standards soutenus par de larges communautés internationales, comme celle du consortium TEI.

Cette publication de troisième type se caractérise comme une édition à partir du moment où elle atteint une certaine stabilité : elle a été fixée par un chercheur dans un état précis (éventuellement pré-validé scientifiquement par des pairs) et déposée dans une archive ouverte. Dans tous les cas, elle est susceptible d'être citée de façon durable et, pour cela, elle a été pourvue d'un identifiant pérenne. Bien sûr, de nouveaux états du même texte, avec des enrichissements, peuvent être publiés ultérieurement, mais chacun fait l'objet d'un nouveau dépôt dans l'archive ouverte et dispose d'un nouvel identifiant.

## II. LES DIMENSIONS DES PUBLICATIONS NUMERIQUES

Après avoir présenté les trois types d'éditions numériques, nous allons, dans cette partie, présenter leurs cinq dimensions fondamentales. Rappelons d'abord que, comme la publication papier, la publication numérique est un objet composite. Si elle partage certains éléments avec l'édition papier, d'autres lui sont spécifiques. Elle consiste ainsi en :

- un texte ;
- des métadonnées et annotations (descriptives et analytiques) ;
- une présentation (documentation) du projet de numérisation et des enjeux de la publication ;
- une interface de consultation ;
- un plan de gestion des données.

La réalisation de ces cinq éléments implique des choix scientifiques et techniques qui seront discutés dans les pages qui suivent, en tenant compte des différents objectifs des types de publication décrits plus haut. Outre le guide des *Best practices* déjà mentionné, nous nous fondons également, pour la rédaction de cette partie, sur la lecture des *Guidelines for Editors of Scholarly Editions*<sup>5</sup> de la Modern Language Association, qui présente une section spécifique pour les éditions électroniques.

### II.1. LE TEXTE

Quel que soit le type de publication (papier ou numérique), quelques obligations issues de la tradition philologique sont à respecter :

- le texte publié doit être complet (ou, du moins, la sélection effectuée et les omissions doivent être signalées et justifiées) ;
- il doit être établi avec méthode et selon des règles qui sont explicitées et uniformément suivies ;
- dans la mesure du possible, il est également intéressant de permettre la confrontation de la version publiée avec celles qui l'ont précédée ou suivie.

#### II.1.1 Une numérisation attentive des aspects matériels

Au cours de la numérisation des sources primaires (codex, feuillets, pierres...), il convient de se rappeler que les images produites seront vraisemblablement la principale source des utilisateurs pour comprendre la matérialité du support d'écriture. Au cours de la numérisation des codex il est donc important d'inclure des captures de la reliure, des gardes et des annexes. Par ailleurs, une grande attention doit être apportée à la préparation du cahier de charges remis aux prestataires, afin qu'ils suivent de bonnes pratiques en termes de qualité (résolution des images, étalonnage des couleurs, inclusion d'une règle millimétrée, etc.)<sup>6</sup>

<sup>5</sup> Voir <https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions>

<sup>6</sup> Voir à minima les indications fournies par HumaNum, *Le guide des bonnes pratiques numériques*, p. 25-29 [https://www.huma-num.fr/sites/default/files/guide\\_des\\_bonnes\\_pratiques.pdf](https://www.huma-num.fr/sites/default/files/guide_des_bonnes_pratiques.pdf); et/ou les recommandations Minerva: <http://www.minervaeurope.org/interoperability/digitisationguidelines.htm>

### II.1.2 La publication de fac-similés

Dans le cas de la publication de textes en fac-similés – associés, ou non, à une transcription –, la lisibilité des images est essentielle, ce qui suppose à la fois une attention aux formats d'acquisition (qualité de l'image exprimée en dpi), et une juste évaluation des besoins de stockage et d'infrastructure matérielle pour la diffusion/communication de celles-ci.

De plus, la distribution des fac-similés sera accompagnée d'une description catalographique et codicologique des supports matériels, selon l'un des standards appropriés (de préférence EAD ou TEI<sup>7</sup>).

### II.1.3 La publication en mode texte

La publication d'un texte sous un autre format que le format image ou pdf implique des choix scientifiques qu'il convient d'explicitier. Le type de transcription est un bon exemple de ces choix : diplomatique, semi-diplomatique ou modernisée/normalisée? Afin de favoriser la lisibilité, et peut-être également pour des raisons pratiques et/ou juridiques, certains projets se contentent de proposer une transcription normalisée/modernisée, et, dans le meilleur des cas, de renvoyer le chercheur au mode image pour observer l'état primitif du texte. Or, un tel choix est synonyme de perte de données idiosyncrasiques et/ou d'histoire de la langue. Un chercheur intéressé par ces aspects ne pourra les exploiter avec des moyens automatiques ou semi-automatiques qu'en retournant à l'image de départ, et donc en dépensant de nouvelles ressources pour une nouvelle transcription. Il est donc recommandé de fournir au moins une transcription semi-diplomatique à laquelle on pourra ajouter, pour assurer une plus grande lisibilité du texte, une transcription normalisée/modernisée (v. *infra*, point II.4).

En ce qui concerne la qualité de la donnée produite (« le texte »), différents niveaux sont attendus en fonction du type de publication envisagé :

- pour une archive éditorialisée (publication de type 1), le texte « brut », légèrement nettoyé, peut suffire ;
- pour les éditions de lecture ou enrichies (publications de type 2 et 3), le texte aura été relu et corrigé, selon un processus clair et documenté. Un système sera mis en place pour la validation des transcriptions, et le processus adopté sera indiqué dans les métadonnées: noms des responsables, délais, thésaurus commun, normalisation des noms, dates, etc. Le lien entre document et métadonnées devra être clair, et l'accès à l'un et à l'autre, de l'un à l'autre, facile.
- en outre, l'édition enrichie peut apporter des informations sur les différents états du texte (variantes), la source des corrections effectuées, les difficultés de lecture, etc. Toutefois, ce type d'enrichissement n'est pas obligatoire, l'édition du troisième type pouvant avoir d'autres objectifs que philologiques.

## II.2. LES METADONNEES ET L'ANNOTATION

Les métadonnées sont un ensemble structuré d'informations permettant de décrire la ressource, de la classer, de l'organiser et de caractériser des données ou du contenu. Il existe au moins quatre types de métadonnées pour l'édition numérique :

- les métadonnées descriptives : elles permettent l'identification non ambiguë de la source (analogique et/ou numérique) ;

<sup>7</sup> Voir aussi les recommandations DeMarch [[http://www.bnf.fr/documents/ead\\_demarch.pdf](http://www.bnf.fr/documents/ead_demarch.pdf)].

- les métadonnées administratives : elles apportent des informations sur les caractéristiques des fichiers, les droits d'accès et d'usage, sur le processus de création des données ;
- les métadonnées structurelles : elles expliquent la composition ou l'organisation de la ressource : les pages, les chapitres, la table des matières, autres éléments constitutifs, etc. Ces métadonnées facilitent la caractérisation, la navigation, la présentation et la compréhension de la structure des sources ;
- les métadonnées techniques : elles précisent les caractéristiques techniques des données, les logiciels utilisés pour leur production et manipulation, leurs versions.

En outre, les éditions électroniques peuvent comporter des annotations permettant d'analyser et d'interpréter la source ; on parle à ce sujet de métadonnées d'enrichissement (ou de « balisage »), si ces annotations sont faites au fil du texte, au moyen d'un jeu réfléchi et prédéterminé d'étiquettes et d'attributs.

Quel que soit le degré de finesse dans le renseignement de ces différents aspects, les métadonnées doivent respecter les normes et standards internationaux.

### *II.2.1. Les métadonnées descriptives, administratives et techniques*

Selon l'institution qui traite l'objet et qui est responsable de la numérisation et/ou de la publication, les informations sur le texte numérisé peuvent être de différentes. Toutefois, il existe un certain nombre de champs qu'il convient de compléter dans tous les cas. À cet égard, les champs mentionnés par la norme Dublin Core non-étendu<sup>8</sup> peuvent représenter un minimum souhaitable, car ils apportent des informations sur le texte (titre, éditeur, date), la propriété intellectuelle (auteur, droits), l'instanciation, la gestion, la ressource (formats, dimensions), son contenu (type, mots-clés) et les modalités de préservation des documents. Dans une édition XML/ TEI, on retrouvera ces informations dans la structure minimale de l'en-tête (header TEI).

Lors de la saisie des métadonnées, une attention soutenue sera accordée à la normalisation de la présentation de celles-ci. Ceci implique, par exemple, le respect des recommandations internationales pour la saisie des dates (AAAA-MM-JJ), des noms de lieux (PAYS, Ville), des noms de personnes (NOM, Prénom), etc. Le cas échéant, il pourra être utile de recourir à des thésaurus pour faciliter l'exploitation postérieure des données.

Dans le cas d'une publication de type 3, on présentera obligatoirement un jeu de métadonnées plus étendu, en apportant toutes les informations nécessaires à la description du témoin de départ et à la caractérisation de la publication effectuée. La publication de type 3 utilise nécessairement un standard reconnu d'encodage des métadonnées, tel que METS, MIX, UNIMARC, XML-EAD, Dublin Core simple/étendu.

Par ailleurs, dans une édition XML/TEI, plusieurs éléments et sections du header permettent d'atteindre un très haut niveau de précision et de finesse dans la description de la source de départ et de l'édition produite. Sans entrer dans les détails, amplement présentés et commentés par les Guidelines de la TEI, il est à noter qu'outre la section obligatoire <fileDesc> le header permet de fournir des informations via les sections <encodingDesc>, <profileDesc> et <revisionDesc>. Celles-ci sont attendues dans une édition de niveau 3 même si les recommandations de la TEI les jugent optionnelles. Par ailleurs, une édition de ce niveau ne se contente pas de l'élément générique <p> à l'intérieur des sous-sections de <fileDesc> : titre, auteur, éditeurs scientifiques, différents membres de l'équipe ayant participé à l'encodage sont

<sup>8</sup> [http://www.bnf.fr/fr/professionnels/formats\\_catalogage/a.f\\_dublin\\_core.html](http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html)

précisés avec les éléments idoines, tandis que la source est décrite au moyen de l'élément <biblStruct> et son jeu très détaillé d'éléments fils<sup>9</sup>.

Enfin, les publications de type 3 feront référence à des fichiers d'autorité et à des vocabulaires contrôlés pour la description de différents éléments de la ressource.

### II.2.2. Les métadonnées structurelles et l'annotation sémantique

Dans une édition papier, les notes constituent traditionnellement l'espace privilégié dédié à l'apport des informations scientifiques. Tout en permettant d'insérer des notes, l'édition électronique dispose de systèmes plus élaborés pour enrichir le texte. La présence d'enrichissements sémantiques, à l'aide de balises par exemple, constitue un élément discriminant entre les différents types de publications numériques.

Offrant principalement des images et du texte peu travaillé (ou seulement une de ces alternatives), les archives éditorialisées n'impliquent pas vraiment d'encodage. Quand on y trouve quelques balises, celles-ci sont issues le plus souvent d'un système automatique de traitement de la source. À l'inverse, la richesse d'une publication de type 3 se mesure en fonction de la quantité d'information supplémentaire apportée par les chercheurs, souvent grâce à une coûteuse campagne d'annotation manuelle. Entre les deux, les publications de type 2 visent un niveau d'enrichissement plus modeste, se concentrant sur certains traits structurels. Elles se caractérisent également par peu de complexité documentaire : c'est vraiment l'accès au texte qui compte, dans un format *machine-readable*.

Avant de décrire les attendus d'annotation liés aux types 2 et 3, notons que l'enrichissement du texte peut être réalisé à l'aide de multiples technologies et outils : traitement ou éditeur de texte, éditeur xml ou html, etc. Tous les formats de fichier ainsi produits (.txt, .docx, .odt, .xml, .html, etc.) comportent des balises, même lorsqu'elles sont invisibles. Toutefois, et comme indiqué plus haut, l'utilisation d'un format ouvert, non-propriétaire (comme le XML) sera à privilégier, afin de permettre une ré-exploitation facile des données scientifiques.

Dans une publication de type 2, on s'attachera à coder :

- les grandes sections du texte (<div>, dans la TEI), et les titres (lorsqu'ils existent) ;
- si pertinent, les paragraphes (à l'aide des balises dédiées) ;
- pour le théâtre : les actes, scènes et tours de parole (si approprié aux caractéristiques du texte) ;
- pour la poésie : les strophes et les vers (si approprié) ;
- pour la correspondance : outre les <div> et les paragraphes, signalés plus haut, les destinataires, les expéditeurs, les lieux de rédaction et d'expédition, les dates (lorsque ces informations sont explicites dans le texte) ;
- pour les journaux, revues et gazettes : les sections et les articles, les rédacteurs, les dates de parution des textes (lorsque ces informations sont explicites dans le texte) ;
- pour les entrées des dictionnaires et des encyclopédies : le mot vedette est séparé du texte de l'entrée à proprement parler (se référer à TEI-Lex0<sup>10</sup> pour des recommandations plus détaillées).

<sup>9</sup> Le groupe de travail [Data\_Cahier] du consortium CAHIER propose un modèle de TEI Header utilisable par tous les projets d'édition.

<sup>10</sup> Bański, P., Bowers, J., Erjavec, T. (2017). « TEI-Lex0 guidelines for the encoding of dictionary information on written and spoken forms », *Electronic Lexicography in the 21st Century : Proceedings of ELex 2017 Conference*, Leiden [hal-01757108]



Dans une publication de type 3, l'encodage est plus élaboré, en lien avec l'objectif scientifique. De manière générale, les divisions sont typées selon une nomenclature unifiée, et les différents éléments de structure (scène et actes ; chapitres ; vers ou groupes de vers ; articles, etc.) reçoivent des identifiants uniques, avec une granularité aussi fine qu'approprié. Par rapport à une édition de type 2, les éditions enrichies se distinguent par un jeu plus élaboré d'éléments et d'attributs, comportant par exemple :

- dans le cas du théâtre, des attributs (comme @who dans la TEI) permettant d'identifier de façon non ambiguë le locuteur de chaque réplique, une liste des entre-parleurs avec des identifiants uniques (<castList> avec des <castItem>, dans la TEI), et le typage des didascalies ;
- dans le cas de la poésie, des éléments annotant la rime et le rythme ;
- dans le cas des dictionnaires, une annotation de l'information grammaticale, étymologique et d'usage (si possible), l'identification des sources des citations et leur équipement avec des liens hypertexte pointant vers des éditions extérieures, etc.

Outre les éléments précédents, et **en fonction de la question de recherche qui sous-tend l'édition**, on peut effectuer :

- un encodage complexe qui permet d'observer les différents états génétiques du texte (encodage génétique) ;
- un encodage qui permet d'accéder à différentes versions du texte grâce à différentes transcriptions (diplomatique, semi-diplomatique, modernisée) ;
- un encodage visant à montrer l'évolution d'un texte entre les différents témoins (encodage philologique) ;
- un encodage se concentrant sur des éléments qui permettent d'élucider la poétique ou l'esthétique du texte (éléments de structure, comme l'exposition ou l'exorde, les topoï, les péripéties, etc. : divisions non-explicites ; éléments de style comme les figures, les niveaux d'énonciation, etc. ; éléments participant à la construction de l'univers de fiction : personnages, descriptions, etc.)
- un encodage des entités nommées (noms de lieux, de personnes, d'institutions, dates et événements, etc.), à des fins d'analyse des réseaux ou de prosopographie. Des liens vers des fichiers d'autorité peuvent avoir été ajoutés à ces entités, permettant à la fois l'identification non-ambiguë et la mise en lien avec d'autres ressources du web ;
- etc.

Bien sûr, on répétera ici qu'on attend de ces projets qu'ils précisent leurs critères éditoriaux (en lien avec la question de recherche), leurs enjeux, les responsabilités et le partage du travail entre les différents collaborateurs.

### II.3. LA DESCRIPTION DU PROJET SCIENTIFIQUE

Un projet scientifique d'édition numérique est défini par la qualité de sa documentation, ce qui signifie que la description du projet est fondamentale pour les trois types d'édition. **En fait, une édition qui n'expose pas sa question de recherche et ne déclare pas ses critères de numérisation et de gestion des sources, n'est pas une entreprise scientifique.**

Une documentation de base du projet, et donc commune aux trois types d'édition, inclura *a minima* :

- la description des enjeux scientifiques du projet (motivation, apports, contexte scientifique, etc.), avec référence à des publications antérieures sur le sujet (état de l'art) ;
- la présentation de l'équipe et des responsabilités de chacun ;

- la composition du corpus et la localisation des sources numérisées ;
- la déclaration des caractéristiques techniques des logiciels ;
- les manipulations effectuées pour améliorer les données (amélioration des images, logiciels de gestion, etc.) ;
- si nécessaire, la documentation relative aux conditions d'usage des APIs offertes, des éventuels services de moissonnage, etc.
- si nécessaire également, l'explicitation des critères qui ont accompagné le choix des sources.

Pour les éditions de type 2 et 3, il faudra aussi ajouter des informations relatives à toutes les opérations effectuées, et notamment :

- les critères de transcription ;
- le traitement des erreurs présentes dans la source (qu'elles soient dues à l'auteur ou qu'elles aient été introduites par les scribes, les typographes, etc.) ;
- le traitement de la ponctuation et des graphies anciennes et/ou inusuelles ;
- les choix d'encodage.

Pour les éditions de type 3, il faudra aussi documenter tous les usages particuliers du balisage adopté, ainsi que toutes les adaptations ou customisations des standards utilisés (dans la TEI, cela signifie générer un ODD). On fournira aussi, par exemple, des informations sur la genèse des textes, sur l'état de conservation des sources, sur leur structure et on précisera la présence de différents types d'annotation et de couches de corrections. On fera également référence à une bibliographie spécifique.

Bon nombre de ces éléments figure parmi les métadonnées dont il a déjà été question précédemment. D'autres précisions exigent la rédaction d'un texte explicatif, qui doit être facilement accessible et clairement lié au texte édité à proprement parler. Si on utilise le format TEI pour l'encodage du texte, de telles descriptions peuvent être incluses dans le <teiHeader>, dans les sections spécialisées telles que <projectDesc> et <editorialDecl>. Une alternative qui a été récemment recommandée sur la liste de discussion de la communauté TEI est la création d'un fichier xml séparé pour l'introduction de l'édition scientifique, avec son propre header et <body> ; ce fichier est à relier à celui contenant l'édition scientifique à proprement parler par toute sorte de mécanismes de pointage (corresp ; copyOf ; etc.)

## II.4. L'INTERFACE DE CONSULTATION

Le propre de l'édition électronique est d'utiliser le Web pour communiquer un texte à ses lecteurs/ utilisateurs. Les questions qui se posent dès lors sont celles de l'accessibilité de l'édition, d'une part, et de sa réutilisabilité, d'autre part. L'interface de consultation doit être conçue en tenant compte de ces deux séries d'exigences.

### II.4.1 L'accessibilité

Le terme « accessibilité » n'est pas utilisé, dans ce qui suit, pour faire référence à l'adaptation des publications aux différents publics (non-voyants, mal voyants, dyslexiques, etc.), mais pour désigner la capacité offerte à un aussi grand nombre d'utilisateurs que possible de lire et explorer librement les textes.

La première exigence en termes d'accessibilité est de ne pas « enfermer » le texte dans une solution propriétaire, nécessitant des logiciels spécifiques pour y avoir accès. **C'est ici que**

**L'utilisation de standards ouverts prend tout son sens<sup>11</sup>.** L'emploi de formats de structuration, d'exploitation ou de protocoles d'échange suffisamment standard pour ne pas supposer un environnement matériel et ou logiciel particulier garantit la *disputatio*, caractéristique et moteur du travail scientifique. D'une manière générale, le respect des normes d'accessibilité proposées par le W3C est donc une bonne pratique.

Par ailleurs, pour les besoins de la recherche, il convient de donner accès au document en entier dans sa version numérique, **à son code source** et aux schémas d'encodage qui le valident. Pour satisfaire à cette attente, le recours aux plateformes de partage (type « git ») facilite la mise à disposition de ces codes et sources.

Une attente particulière de la communauté scientifique peut être de pouvoir lire le texte à la fois dans sa version d'origine (respectant la graphie, les habitudes scripturales, éventuellement les choix de mise en page de l'auteur initial), et dans une version modernisée. La proposition de plusieurs versions (diplomatique, semi-diplomatique, modernisée, type 1, 2, 3) est donc un élément supplémentaire d'accessibilité. Aussi, l'accessibilité signifie-t-elle la mise à disposition des versions au fur et à mesure de leurs enrichissements, des formats numériques dans lesquels le texte est proposé aux utilisateurs et, surtout, de la version brute du texte de façon à faciliter la genèse de nouvelles recherches.

Pour terminer, il est fortement recommandé de privilégier l'accès libre (non payant) aux sources, en attribuant à la publication une licence ouverte adaptée (de type Creative Commons). Si un embargo était nécessaire, une barrière d'authentification peut être proposée pour une durée limitée.

Enfin, afin d'augmenter l'accessibilité il est recommandé de multiplier les versions, à la fois du texte, et des formats numériques dans lesquels il est proposé aux utilisateurs.

#### II. 4. 2. La réutilisabilité

La conception d'une interface de consultation doit donner la priorité à la réutilisabilité de l'édition. L'esthétique ou le degré d'élaboration de l'interface sont importants mais ils ne sont pas nécessairement un indice de positionnement de l'édition électronique dans l'un ou l'autre des types décrits dans la première partie de ce document. Ainsi, il existe des éditions enrichies qui ne disposent pas d'interfaces de consultation très élaborées, et des archives éditorialisées qui peuvent avoir construit une interface attractive et dynamique, sans proposer de véritables enrichissement du contenu. S'il est toujours précieux de pouvoir visualiser les textes sous forme de vignettes, mais aussi en pleine page, de zoomer sur certaines parties de l'image, etc., la possibilité d'exploration libre de données produits selon un protocole validé scientifiquement constitue la première source de création de nouvelles connaissances.

Si l'on se tourne du côté des usages possibles du texte, on visera d'abord la garantie de sa lisibilité. Si une interface est proposée, le texte doit pouvoir être parcouru de façon cursive, grâce à un affichage en tant que page web (en HTML ou XML transformé), ou via un fichier .pdf, .epub, .odt, .rtf ou autre. Dans le cadre d'une archive éditorialisée, la publication des fichiers issus de l'étape d'acquisition du texte peut suffire à combler cette exigence : une série d'images, du texte « brut » mis en ligne satisfont à l'exigence de lisibilité. Avec les éditions de type 2 ou 3, des opérations supplémentaires doivent être effectuées, puisqu'on ne peut pas envisager de s'en tenir au simple dépôt, sur le web, d'un fichier probablement en format xml, plus ou moins richement encodé. L'objectif dans ces deux cas est à la fois de donner accès aux contenus et de permettre l'exploration des différentes métadonnées et annotations, pour présenter des informations structurées aux lecteurs, ou pour présenter le texte selon différents points de vue et perspectives.

<sup>11</sup> On notera en ce sens l'importance du recours à un jeu de caractères UTF-8 pour la saisie et l'affichage final.

Au-delà de la possibilité de lire le texte, les utilisateurs peuvent souhaiter l'interroger, en faire des extraits, l'enrichir à leur tour, etc. Nombreuses sont les éditions qui proposent des interfaces d'interrogation. Cependant, celles-ci ne peuvent jamais anticiper sur tous les questionnements possibles, sur toutes les questions de recherche susceptibles d'être poursuivies. La mise en ligne du fichier source apparaît, dans cette perspective, et comme indiqué plus haut, aussi importante que celle d'une version pour la lecture. Elle permet de reprendre le document numérique et de l'interroger avec ses propres outils, d'observer le système d'annotation choisi, d'en mesurer la pertinence et la rigueur, de mieux comprendre comment l'exploiter, éventuellement de l'affiner. Inversement, dans le cas des éditions très lourdement annotées une version simplifiée, fournissant le texte accompagné d'annotations structurales, favorise la lecture, ainsi que les opérations de transformation (comme l'ajout de ses propres codes, le passage vers l'édition papier, etc.). La mise en ligne du texte brut, en format .txt par exemple, est indispensable à certaines analyses textométriques, statistiques et à d'autres traitements automatiques.

On retiendra pour terminer que les interfaces suscitent des questions spécifiques au sujet de leur maintenance. Celle-ci demande des compétences différentes par rapport à la maintenance du fichier (ou des fichiers) résultant de la numérisation de la source. C'est peut-être sur cet aspect que le rôle de l'éditeur « matériel », par opposition à l'éditeur scientifique, est le plus important, car il est susceptible de mettre au point la meilleure vitrine pour le texte produit. On ne peut qu'appeler au développement des capacités institutionnelles d'édition numérique<sup>12</sup>, afin de permettre à un plus grand nombre d'éditeurs scientifiques d'accéder à des services de qualité pour la scénarisation de leurs éditions sur le web.

## II. 5 LA GESTION DES DONNEES

Compte-tenu de la grande instabilité du numérique par rapport au papier, le risque de perdre le travail effectué est un enjeu de taille. Une édition diffusée en ligne devrait rester consultable à long terme, et dans des conditions identiques, même si les techniques et les technologies évoluent. Dès lors, il est nécessaire de mettre en place un plan de conservation, qui prenne en compte tout particulièrement l'exigence de citabilité de l'édition, sans laquelle celle-ci ne peut pas jouer pleinement son rôle dans le milieu académique. Par ailleurs, la pérennité d'une édition numérique dépend d'un ensemble d'activités coordonnées et de l'anticipation des moyens pour les réaliser : moyens et compétences des éditeurs, aide des directions des systèmes d'information des universités, création d'équipes dédiées. Dans tous les cas, il revient au chercheur, responsable du projet scientifique, d'anticiper la gestion de ses données en préparant le Plan de Gestion des données de son projet (*Data Management Plan*). Le DMP est un document évolutif qui aide et explicite de quelles façons les données utilisées et générées par le projet seront utilisées. Initié au début du projet, ce plan doit être mis à jour de manière périodique pour s'assurer de son adéquation avec le déroulement de l'activité ou du projet.

En ce qui concerne le stockage et la conservation des données, l'utilisation d'entrepôts ouverts, publics et/ou institutionnels est recommandée<sup>13</sup>. Parallèlement, on peut recourir aux services d'un éditeur matériel, qui fournit parfois des solutions de pérennisation en plus de ses missions traditionnelles (mise en place de circuits d'expertise, préparation, mise en forme et diffusion des textes). Une troisième possibilité est le recours à des « centres d'humanités numériques » sur le modèle anglo-saxon, ou « cyber infrastructures », comme les appellent P.

<sup>12</sup> Sur le modèle de la chaîne éditoriale de Caen, Metopes (voir <http://www.metopes.fr> et [http://www.unicaen.fr/recherche/mrsh/document\\_numerique/projets/metopes](http://www.unicaen.fr/recherche/mrsh/document_numerique/projets/metopes) ) par exemple, ou des Presses de l'Université de Montréal (voir <https://apropos.erudit.org/fr/>)

<sup>13</sup> En France, HumaNum propose un service public de stockage des données, Nakala ; ailleurs dans le monde il existe des Gitlabs universitaires.

Mounier et M. Dacos<sup>14</sup>, qui réunissent les compétences techniques nécessaires pour gérer à long terme les données des projets<sup>15</sup>. Ceux-ci se dessinent moins comme des alternatives, que comme des institutions complémentaires à celles qui interviennent actuellement dans la gestion et la diffusion des résultats de la recherche.

La citabilité relève par conséquent des deux conditions suivantes :

- un format de nommage stable de la ressource sur le web. Les documents doivent être localisés par une identification univoque : dans ce cas, l'utilisation de systèmes d'identification tels que Handle ou DOI<sup>16</sup> est recommandée. Il est important de noter qu'un tel identifiant devrait être attribué de manière à permettre la citation raisonnée de toutes les parties de textes souhaitées par les éditeurs scientifiques. Ceci implique un travail en amont sur l'unité minimale citable : les éditions de type 3 offrent, comme on l'a vu plus haut, une granularité plus fine que celle de type 2, puisque l'articulation des identifiants pérennes avec des identifiants internes à la source (pour la citation de fragments) est possible.
- l'indication, dans l'édition même ou par l'intermédiaire de l'interface de consultation, d'une « modalité de citation conforme aux normes bibliographiques [et] aux usages et impératifs de la discipline <sup>17</sup> ». On précisera notamment les responsabilités scientifiques et éditoriales, ainsi que les titres (de l'œuvre, de la partie, du chapitre...)

Il est à retenir que le dépôt et l'allocation d'identifiants stables doivent concerner chaque version de l'édition ayant atteint un état satisfaisant et significatif d'achèvement. Tout comme dans l'édition papier, une seconde édition n'a pas vocation à remplacer la première, mais à apporter de nouvelles connaissances. La première édition doit rester consultable et citable, tout en permettant de mesurer les progrès de la connaissance scientifique. Par ailleurs, une adresse stable, de type PURL (Persistent Uniform Resource Locators), favorise la citabilité. Rappelons que « les url sémantiques ou explicites, les url comportant des identifiants de session incluent des éléments d'information non pérennes, et correspondent donc à de mauvaises pratiques<sup>18</sup> ».

Au terme de ce parcours, rappelons une fois de plus que les frontières ne sont pas étanches entre les différents types d'édition et les différentes dimensions décrites plus haut. Des décisions prises sur un des points mentionnés ont des conséquences sur les autres. Ainsi, la conception d'une campagne d'annotation a des répercussions sur la rédaction du DMP et influence naturellement le type d'interface de consultation que l'on proposera. De même, le choix de publier un texte via un CMS ou en tant que fichier XML conditionne non seulement l'ensemble de la chaîne de traitement, mais également le niveau que l'édition peut atteindre : il paraît difficile, avec la première option, de parvenir à une publication du troisième type. Des décisions mûrement réfléchies et adaptées au projet scientifique sont nécessaires sur chacun de ces points, afin de ne pas se retrouver contraint par des solutions d'une apparente facilité, qui ne tiennent cependant pas compte de l'ensemble des exigences entrant en jeu dans la confection d'une édition scientifique numérique.

<sup>14</sup> Marin Dacos et Pierre Mounier, *Humanités numériques*, rapport d'étude, 2014, p. 26 et 31. En ligne : [\[http://www.institutfrancais.com/fr/actualites/humanites-numeriques\]](http://www.institutfrancais.com/fr/actualites/humanites-numeriques)

<sup>15</sup> C'est le cas, en France, avec HumaNum par exemple.

<sup>16</sup> L'inconvénient de DOI étant que c'est un service payant, et privé.

<sup>17</sup> Collège de l'édition scientifique numérique, *Charte des bonnes pratiques de l'édition scientifique numérique*, BSN 7 [\[http://www.bibliothequescientificquenumerique.fr/charte-des-bonnes-pratiques/\]](http://www.bibliothequescientificquenumerique.fr/charte-des-bonnes-pratiques/).

<sup>18</sup> Idem.

### III. Grille d'évaluation des publications numériques de corpus d'auteurs

Cette grille résume les attendus liés à chaque type de publication, même si des zones de superposition et de flottement entre les différents types d'édition y apparaissent clairement. Elle précise, en outre, si l'élément ou la caractéristique concernée est indispensable ou bien si son absence ne peut pas être considérée, malgré tout, comme un élément diminuant la valeur de l'édition. Deux codes sont employés, E désignant une exigence essentielle, et O une demande optionnelle.

	Type 1	Type 2	Type 3
<b>Traitement des sources</b> <ul style="list-style-type: none"> <li>- Fac-similé (jpeg, pdf, tiff, etc. ; sauf restrictions liées aux droits)</li> <li>- Transcription (ou OCR, ou similaire), au format brut</li> <li>- Transcription selon des critères scientifiques établis par le projet (diplomatique, semi-diplomatique, normalisée)</li> </ul>	E E O	O E E	O O E
<b>Métadonnées</b> <ul style="list-style-type: none"> <li>• format standard (normalisé)</li> <li>• métadonnées descriptives (bibliographiques, y. c. les responsabilités dans la création de l'édition)</li> <li>• métadonnées administratives (techniques, de droits, etc.)</li> <li>• métadonnées structurelles</li> <li>• enrichissements (annotation et/ou balisage)</li> </ul>	E E E O O	E E E E O	E E E E E
<b>Description du projet scientifique</b> <ul style="list-style-type: none"> <li>• enjeux scientifiques (motivation, apports, etc.)</li> <li>• présentation de l'équipe et des responsabilités de chacun</li> <li>• éventuellement, critères ayant accompagné le choix des sources (témoins)</li> <li>• présentation du corpus</li> <li>• présentation des critères de transcription</li> <li>• précisions sur le traitement des erreurs présentes dans la source (scribes, typographes, lectures antérieures, etc.)</li> <li>• choix de traitement de la ponctuation et des graphies</li> <li>• choix d'encodage</li> <li>• informations sur la genèse du texte</li> </ul>	E E O O E O O O	E E E O E E O O	E E E O E E E O

<p><b>Accessibilité</b></p> <ul style="list-style-type: none"> <li>• version de lecture (interface minimale)</li> <li>• possibilité de lire le texte en différentes versions</li> <li>• possibilité de télécharger les sources                             <ul style="list-style-type: none"> <li>○ en format texte</li> <li>○ en format image (si libres de droits)</li> <li>○ en format xml</li> </ul> </li> <li>• possibilité de télécharger le texte en différents formats</li> </ul>	<p>E O O O O O</p>	<p>E O O O O O</p>	<p>E E O O E O</p>
<p><b>Plan de gestion des données</b></p> <ul style="list-style-type: none"> <li>• usage de standards</li> <li>• archivage pérenne</li> <li>• présence d'un identifiant permanent (handle, ark, purl)</li> <li>• programme de maintenance des données (périodicité, responsabilité)</li> </ul>	<p>E E O E</p>	<p>E E E E</p>	<p>E E E E</p>

#### IV. Exemples d'éditions de différents niveaux

##### Archives éditorialisées

L'archive des manuscrits de Stendhal [<http://pagella.bm-grenoble.fr/BMG.html?id=Bmg-0002150>]

Archives et chronologie des spectacles [<https://dezedo.org/>]

##### Éditions de lecture

Les Bibliothèques Virtuelles Humanistes [<http://www.bvh.univ-tours.fr/>]

*Artamène ou Le Grand Cyrus* [<http://www.artamene.org/>]

Manuscrits de Carlos Liscano [<https://liscano.nakalona.fr/>]

##### Éditions enrichies

Correspondance de Van Gogh [<http://vangoghletters.org/vg/letters.html>]

Édition en ligne des œuvres et de la correspondance de Descartes [<http://www.unicaen.fr/puc/sources/prodescartes/>]

Édition des dossiers de Flaubert pour Bouvard et Pécuchet [<http://www.dossiers-flaubert.fr/>]

Édition en ligne des *Essais* Montaigne [<https://montaigne.univ-tours.fr/edition-numerique-de-lexemplaire-de-bordeaux-des-essais-de-montaigne-1588/>]

Voir aussi : Marie-Luce Demonet, Alain Legros, Mathieu Duboc, Lauranne Bertrand, Alexei Lavrentiev (2016). Michel de Montaigne, *Essais*, 1588 (Exemplaire de Bordeaux), édition numérique génétique (XML-TEI/ PDF) [halshs-01337873]



## V. Bibliographie indicative

- Bachimont, Bruno (2007). « Nouvelles tendances applicatives. De l'indexation à l'éditorialisation », dans Patrick Gros (dir.), *L'indexation multimédia : description et recherche automatiques*, Paris, Hermès Sciences, 2007, p. 15-29.
- Benoit, J.-L., Bernet, Ch., Bonhomme, P., Romary, L., Viscogliosi, N. (2000). « Du document électronique à son usage : le rôle central de la normalisation », *Solaris*, décembre 1999 – janvier 2000 ; consultable aussi sur HAL [id : inria-00099209].
- Burnard, Lou, Katherine O'Brien O'Keefe, and John Unsworth (eds.) (2007). *Electronic Textual Editing*. New York: Modern Language Association of America.
- Driscoll, Matthew J. (2010). « The Words on the Page: Thoughts on Philology, Old and New », in Judith Quinn and Emily Lethbridge (eds.), *Creating the Medieval Saga: Versions, Variability and Editorial Interpretations of Old Norse Saga Literature*. Odense: Syddansk Universitetsforlag, p. 85–104.
- Driscoll, Matthew J., Pierazzo, Elena (2016). *Digital Scholarly Editing. Theories and Practices*, Open Book Publishers,
- Fraistat, Neil, Flanders, Julia (eds.) (2013). *The Cambridge Companion to Textual Scholarship*, Cambridge: Cambridge University Press.
- Gabler, Hans-Walter (2010). « Theorizing the Digital Scholarly Edition », *Literature Compass*, 7, p. 43–56.
- Gabler, Hans-Walter (2012). « Beyond Author-Centricity in Scholarly Editing », *Journal of Early Modern Studies*, 1, p. 15–35, [<http://www.fupress.net/index.php/bsfm-jems/article/view/10691>].
- Hirsch, Brett D., Jenstad, Janelle (2016). « Beyond the text : Digital Editions and Performance », *Shakespeare Bulletin* 34.1, The Johns Hopkins University Press, p.107–127.
- Kamp, Georg (2008). « Digital Publishing in the Faculties of Arts and Humanities : Prospects, Chances, and Risks », *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science*, 5/2, p. 77-80.
- Kline, Mary-Jo, Perdue, Susan H. (2008). *A Guide to Documentary Editing*, Charlottesville: University of Virginia Press, [<https://gde.upress.virginia.edu/>]
- Lebrave, Jean-Louis (2008). « De l'édition informatisée à l'édition électronique », in Aurèle Crasson (ed.), *L'Édition du manuscrit. De l'archive de création au scriptorium électronique*, Louvain-La-Neuve : Academia Bruylant, p. 169-188.
- McGann, Jerome (2001). *Radiant Textuality. Literature after the World Wide Web*, Palgrave.
- McGann, Jerome (2004). « Marking Texts of Many Dimensions », in Susan Schreibman, Ray Siemens, and John Unsworth (eds.), *A Companion to Digital Humanities*, Oxford: Blackwell, p. 198-217  
[<http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-3-4&toc.depth=1&toc.id=ss1-3-4&brand=default>]
- Mounier, Pierre (2018). *Les Humanités numériques : Une histoire critique*, Paris, Éditions de la Maison des sciences de l'homme.
- Nell Smith, Martha (2004). « Electronic Scholarly Editing », in Susan Schreibman, Ray Siemens, and John Unsworth (eds.), *A Companion to Digital Humanities*, Oxford : Blackwell, [[http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-3&toc.depth=1&toc.id=ss1-4-3&brand=9781405103213\\_brand](http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-3&toc.depth=1&toc.id=ss1-4-3&brand=9781405103213_brand)]
- Pierazzo, Elena (2011). « A Rationale of Digital Documentary editions », *Literary and Linguistic Computing*, 26, p. 463–77.
- Pierazzo, Elena (2014). « Digital Documentary Editions and the Others », *Scholarly Editing*, 35,

- [<http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html>].
- Pierrazzo, Elena (2014). *Digital Scholarly Editing : Theories, Models and Methods*, Farnham, Surrey: Ashgate, 2015 [disponible également sur HAL : hal-01182162]
- Price, Kenneth M. (2008). « Electronic Scholarly Editions », in Susan Schreibman and Ray Siemens (eds.), *A Companion to Digital Literary Studies*. Oxford: Blackwell, p. 434-450, [<http://www.digitalhumanities.org/companionDLS/>]
- Price, Kenneth M. (2009). « Edition, Project, Database, Archive, Thematic Research Collection : What's in a Name ? », *Digital Humanities Quarterly*, vol. 3, no. 3, [<http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html>]
- Price, Kenneth M. (2016). « Social Scholarly Editing », in Susan Schreibman, Ray Siemens and John Unsworth (eds.), *A New Companion to Digital Humanities*. Malden – Oxford – Chichester : Wiley Blackwell, p. 137-149.
- Renear, Allen H. (2004), « Text Encoding », in Susan Schreibman, Ray Siemens, and John Unsworth (eds.), *A Companion to Digital Humanities*, Oxford: Blackwell, p. 218-239.
- Robinson, Peter (2013). « Towards a Theory of Digital Editions », *Variants*, 10, p. 105-131.
- Rockwell, Geoffrey (2012). « Short Guide to Evaluation of Digital Work », *Journal of Digital Humanities*, vol. 1, no. 4 Fall [<http://journalofdigitalhumanities.org/1-4/short-guide-to-evaluation-of-digital-work-by-geoffrey-rockwell/>]
- Sahle, Patrick, et. al. (2014). *Kriterien für die Besprechung digitaler Editionen*, Version 1.1 [<https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/>]
- Schmidt, Desmond (2012). « The Role of Markup in the Digital Humanities », *Historical Social Research/Historische Sozialforschung*, 37/3, p. 125-146.
- Sinatra, Michaël, Vitali-Rosati, Marcello (2014). *Pratiques de l'édition numérique*, Montréal : Presses de l'université de Montréal. [<https://books.openedition.org/pum/306?lang=fr>]
- Vanhoutte Edward (2010). « Defining Electronic Editions: A Historical and Functional Perspective », in Willard McCarty (ed.), *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Cambridge: Open Book Publisher, p. 119-144.
- Van Zundert, Joris and Boot, Peter (2012). « The Digital Edition 2.0 and the Digital Library : Services, not Resources », *Bibliothek and Wissenschaft*, 44, p. 141-152.

---

<sup>i</sup> L'ordre de présentation des auteurs de ce guide cherche à rendre compte de l'intervention de chacun dans le texte. L'ensemble du guide a été travaillé de façon collective, et nombreuses sont les interventions ponctuelles des uns et des autres dans différentes parties du texte ; de ce fait, il s'agit véritablement d'une publication collective, et il ne faut surtout pas déduire un « poids » de chaque contribution à partir de l'ordre des signatures. Toutefois, outre la participation à la discussion collective, on retiendra que Ioana Galleron a pris en charge la coordination de l'ensemble et une partie de la section II.4, M.-L. Demonet et C. Meynard ont plus particulièrement rédigé la section II.1, Fatiha Idmhand la section II.2, Elena Pierazzo les sections II.2 et II.3, Geoffrey Williams la section II.4 et P.-Y. Buard et Julia Roger la section II.5.