Manuel de stylage pour les guides MASA

La plateforme en ligne MASA a pour objectif de mettre en place un outil permettant de mettre à disposition de la communauté scientifique des manuels de bonnes pratiques, traitant d’humanité numérique en archéologie (interopérabilité, thésaurus, archivage, *etc*.). La plateforme permet de déposer un guide dans un format de « traitement de texte » et de transformer le fichier dans un format « en ligne » et visible depuis un navigateur web. Plus précisément, la plateforme transforme automatiquement un fichier au format .odt (qui est le format de traitement de texte par défaut du logiciel OpenOffice) vers un format PDF, HTML ou TEI, mais c’est également un outil de stockage de données, en effet un manuel déposé peut ensuite être consulté ou téléchargé par n’importe quel utilisateur.

La procédure décrite dans ce guide indique comment passer d’un format .docx (le format par défaut des fichiers sous le logiciel Word) vers le format de sortie de la plateforme. L’étape qui est décrite en détail ci-dessous est la première à réaliser, avant la transformation, pour mettre en place cette transformation : le stylage du fichier. Cette étape s’appuie sur un certain nombre de règles et de consignes qui sont décrites dans le présent document. Il est essentiel de toutes les suivre, car si la chaîne de production est fonctionnelle, une seule erreur peut mettre en péril son bon fonctionnement. Une fois le fichier transformé, selon les exigences précisées dans le présent document, le fichier obtenu doit être inséré au sein de la plateforme, il sera ainsi stocké sur le server et il pourra ensuite être consulté et téléchargé au format voulu.(PDF, HTML ou TEI en fonction de la demande).

La plateforme MASA permet de créer de sauvegarder et partager des manuels d’une manière normalisée, il n’est pas possible de préserver des aspects esthétiques durant la transformation.

Stylage, encodage et production de fichiers structurés

Le concept de stylage

La plateforme MASA s’appuie sur le concept de stylage pour l’encodage automatique des fichiers. Par « stylage », on entend le fait d’appliquer des styles à des portions de textes au sein d’un logiciel de traitement de texte. Habituellement les styles sont utilisés pour mettre en forme plus rapidement des fichiers de texte de manière cohérente et homogène.

Lorsqu’un stylage est appliqué à un fichier et qu’il est transformé vers un format XML[[1]](#footnote-1), un balisage est appliqué à l’élément stylé, ce balisage est invisible mais c’est sur cela que nous nous appuyons pour réaliser la transformation. Le balisage de mise en forme du texte va ainsi être converti en un balisage sémantique. Ainsi en récupérant ce balisage issu du stylage et en le transformant il est possible d’obtenir un fichier encodé selon les normes d’un langage informatique comme le HTML[[2]](#footnote-2) ou la TEI[[3]](#footnote-3). Une fois le fichier transformé au format HTMl, il est également possible de le transformer au format PDF afin qu’il garde ses propriétés de mises en page en utilisant une transformation s’appuyant sur le langage PHP.

Au sein d’un fichier de traitement de texte classique, un certain nombre d’élément sont déjà balisés automatiquement, mais le stylage nous permet de mettre en place des précisions sur la nature des éléments présents au sein du texte. Par exemple, en appliquant un style de « Titre 1 » à une portion de texte, après la transformation, cette même portion de texte sera identifiée comme un titre de niveau 1 et aura pris la forme voulue pour tous les titres de niveau 1. Le balisage nous permet à la fois d’enrichir sémantiquement le texte en désignant des entités qui le compose mais également d’appliquer une charte typographique[[4]](#footnote-4).

La chaîne de production

Comme cela est présenté sur le site de la plateforme MASA, la marche à suivre pour opérer la transformation du manuel de son format natif (c’est-à-dire le format .docx) vers un format en ligne nécessite de suivre une procédure de plusieurs étapes (figure 1).



Figure 1 : les transformations du fichier

Première étape : le stylage du fichier

Dans un premier temps, il est nécessaire de procéder au stylage du fichier. Le stylage se fait grâce au logiciel de traitement de texte Word, il est nécessaire d’utiliser celui-ci et pas un autre, faute de ne pouvoir réaliser une transformation correcte du fichier par la suite. Cette étape de stylage est détaillée dans la seconde partie de ce document.

Deuxième étape : le passage au format .odt

Après cette étape de stylage, le fichier doit être ouvert dans le logiciel OpenOffice[[5]](#footnote-5) afin de le passer au format .odt. Pour se faire, une fois le fichier ouvert au sein d’OpenOffice, vous pouvez simplement enregistrer le fichier sous ce nouveau format (cliquer sur Fichier, puis Enregistrer sous puis dans la liste déroulante « Type » choisir « Texte OpenDocument (.odt) »). Cette étape est nécessaire afin de pouvoir passer ensuite de ce format .odt à un fichier html ou TEI (il n’est pas possible de passer du format .docx à l’un de ces deux formats).

Troisième étape : la transformation via la plateforme MASA

Une fois ce fichier obtenu il suffit d’utiliser la plateforme MASA pour réaliser la transformation vers le format voulu. Pour se faire il est nécessaire de sélectionner le fichier avec le « Parcourir » de lancer la transformation. Une fois cette étape réalisée, et si celle-ci n’a pas rencontrée d’erreur, le manuel aura été ajouté à la plateforme et en rafraichissant la page vous devriez voir son apparaître dans le cadre de la « Liste des manuels disponibles ».

Le guide peut alors être consulté en appuyant sur le bouton « Voir » à côté d’un manuel. Cette action donne accès une version HTML du manuel. Afin de récupérer le manuel dans sa version HTML, il suffit de faire un clique-droit sur la page du manuel au format HTML et de cliquer sur « enregistrer sous ».

Les boutons en haut de la page permettent d’obtenir le manuel au format TEI et PDF, au clique sur l’un d’entre eux, le téléchargement démarre automatiquement.

Concernant la forme visuelle du manuel à la sortie, celle-ci sera forcément différente de celle que vous aviez mis en place à l’origine sur Word avec le stylage fourni pour cette plateforme, le fond quant à lui restant inchangé. Les caractéristiques typographiques sont les mêmes pour tous les fichiers traités par la plateforme MASA et sont déterminés par le fichier CSS. Il n’est pour l’instant pas possible de personnaliser l’aspect visuel des documents en sortie.

PDF, TEI et HTML : le choix du format de sortie

Si le PDF est un format de fichier qui permet de visualiser directement le manuel via un lecteur PDF. Les deux autres formats de sortie qui sont proposés sur la plateforme sont deux formats pouvant être visualisé via un navigateur internet. Le HTML est le format utilisé par toutes les pages internet, il est donc à privilégier si l’on souhaite simplement afficher le manuel seul. Le format TEI est à utiliser pour insérer le manuel dans une page internet ou pour le modifier à nouveau dans un autre format.

Le stylage d’un manuel

Généralités

Un certain nombre d’éléments constitutifs de tout manuel ont été déterminés en amont et chacun de ces éléments correspond à un style dans le fichier « feuille\_de\_style\_MASA**»** présent sur la plateforme MASA, qu’il vous faut télécharger en cliquant sur le bouton « feuille de style ». Ce fichier apparaît comme un fichier Word vide car il ne contient que des styles.

Deux types de styles existent :

* Les styles de paragraphes qui concernent des paragraphes entiers, ils ne peuvent pas être utilisés sur une seule phrase, ils concernent les titres, les paragraphes, les légendes, *etc*.
* Les styles de caractères : une zone de texte en code informatiques, les liens hypertextes, des mots ou un groupe de mot en langue étrangère.

Afin que les différentes transformations puissent être réalisées, l’intégralité du fichier doit être stylée avec des styles valides, pour pouvoir voir quels styles ont été appliqués et si certains ne correspondent pas, vous devez passer votre fichier en mode « Brouillon ». Pour se faire rendez-vous dans l’onglet « Affichage » et cocher la case « Brouillon ». L’affichage de votre page va changer et vous aurez désormais une colonne à gauche de votre texte vous indiquant quels styles ont été appliqués sur les différentes portions de textes (figure 2).

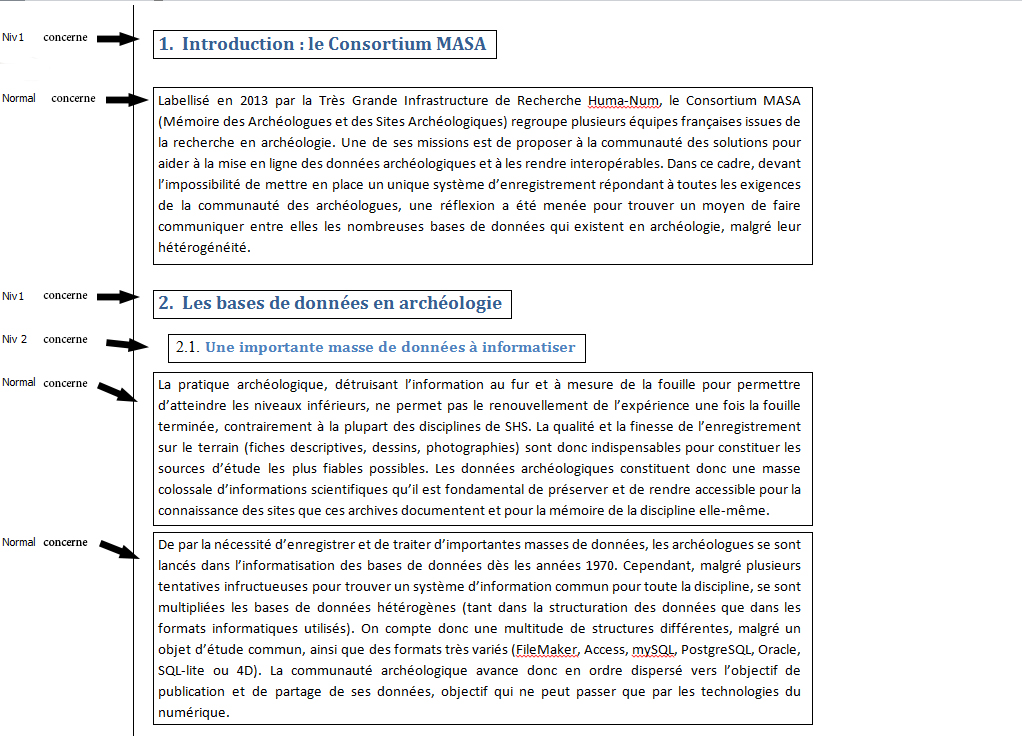


Figure 2 : Exemple du mode Brouillon

Si l’intégralité du fichier doit être impérativement stylé à la fin de cette étape, certains éléments ne sont pas à styler « à la main » car le stylage automatique de Word suffit, c’est le cas notamment pour les images, les tableaux et les listes, ces éléments sont développés plus bas dans ce document dans la troisième sous-partie. Les styles de caractères n’apparaissent pas dans la colonne de gauche en mode Brouillon, ils sont néanmoins pris en compte lors de l’export (figure 3).

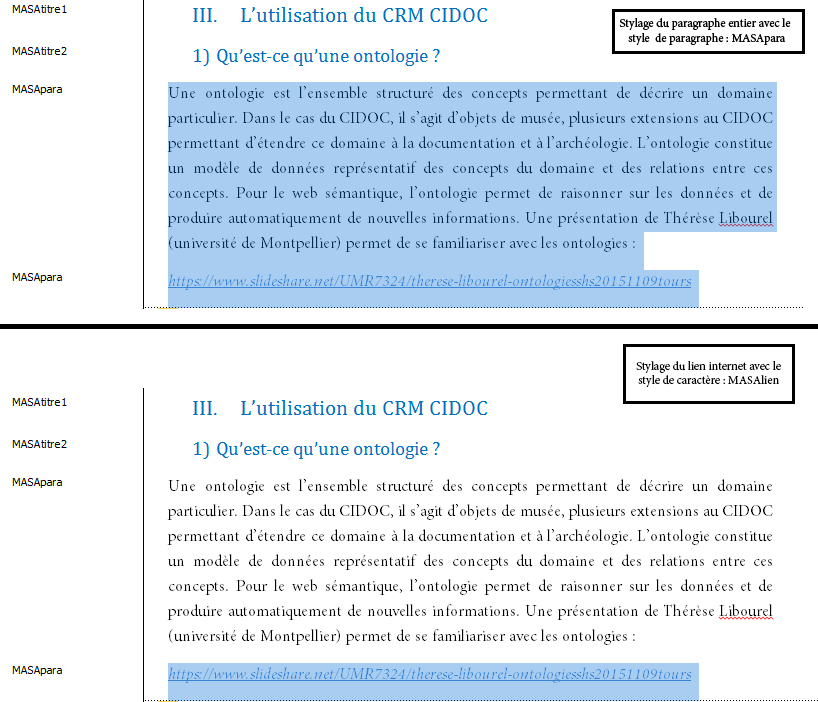


Figure 3 : Exemple d’utilisation d’un style de caractère

De plus il est possible de styler un paragraphe avec un style de paragraphe puis de styler une partie de ce paragraphe avec un style de caractère comme cela est présenter dans la figure 3 qui est présente dans la sous-partie dédiée aux exemples de stylage. En revanche, il n’est pas possible de styler une même portion de texte avec plusieurs styles de paragraphes, si cela est fait c’est seulement le dernier style qui sera pris en compte. Il est également important de retirer tous les espaces inutiles, à savoir les sauts de lignes, ces sauts de lignes sont plus visibles en mode Brouillon : un style est relié à un espace vide (figure 4)



Figure 4 : Exemple de saut de ligne à supprimer

La récupération des styles MASA

Comme évoqué précédemment, le stylage du fichier constituant le manuel se fait à partir du logiciel Word. La première étape est d’importer les styles qui vont être utilisés. Pour cela, il est nécessaire d’avoir préalablement téléchargé le fichier « feuille\_de\_style\_MASA**»**  qui est présent sur la plateforme et qui contient tous les styles. L’étape suivante consiste à importer ces styles au sein du fichier contenant le manuel afin de pouvoir les utiliser, pour se faire il vous faut afficher la fenêtre des styles en cliquant sur le bouton à droite de la liste de styles dans l’onglet « Accueil ».

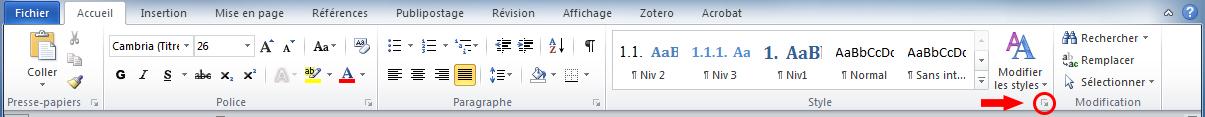


Figure 5: Emplacement du bouton pour la gestion des styles

Au sein de cette fenêtre nouvellement ouverte, cliquer sur le bouton gérer les styles : , puis sur « Importer/Exporter … », au sein de la nouvelle fenêtre, dans la partie de droite qui concerne le fichier au sein duquel il est possible récupérer des styles, cliquer sur « Fermer le dossier » puis « Ouvrir le dossier » et sélectionner le fichier « feuille\_de\_style\_MASA.docx ». L’ensemble des styles MASA devrait être présent au sein de l’encadré de droite. Vous n’avez plus qu’à les ajouter un à un à l’encadré de gauche (qui correspond au style du fichier du manuel) en cliquant sur le bouton  « <-- copier », ces styles MASA sont au nombre de 18. Vous avez maintenant l’ensemble des styles dont vous avez besoin pour le stylage du fichier. Si le manuel n’a pas encore été rédigé, il est possible de commencer à travailler directement à partir de ce fichier « feuille\_de\_style\_MASA.docx » et de l’enregistrer sous un autre nom.

Listes des styles

|  |  |
| --- | --- |
| **Nom du style** | **Eléments du texte à styler** |
| **Page de titre** | |
| MASAtitrePrin | Le titre principal du document (au sein de la page de titre) |
| MASAsousTitre | Le sous-titre du document (au sein de la page de titre) |
| MASAinsti | Institution ayant mis en place le document (laboratoire, consortium, équipe de recherche, *etc*.) |
| MASAauteur | Le nom du ou des auteurs |
| **Titres** | |
| MASAtitre1 | Titre de niveau 1 |
| MASAtitre2 | Titre de niveau 2 |
| MASAtitre3 | Titre de niveau 3 |
| MASAtitre4 | Titre de niveau 4 |
| **Paragraphe** | |
| MASApara | Un paragraphe de texte |
| MASAparaCtxt | Un paragraphe de texte traitant du contexte du projet exposé |
| MASAparaDef | Un paragraphe de texte s’apparentant à une définition d’un terme ou d’une notion. |
| MASAparaEtat | Un paragraphe de texte faisant un état de l’art d’une pratique. |
| MASAparaEx | Un paragraphe de texte listant ou développant un ou plusieurs exemples. |
| MASAparaSynt | Un paragraphe constituant une synthèse. |
| MASAparaTuto | Une paragraphe constituant un tutoriel ou le développement d’une marche à suivre. |
| **Suite de caractères spécifiques** | |
| MASAlien | Lien hypertexte |
| MASAtermEt | Mot en langue étrangère |
| MASAcodeInfo | Suite de caractère en langage informatique |

Le système de numérotation des titres des différents parties se fait de manière automatique via la transformation interne à la plateforme, ainsi il est important de supprimer le numéro de ce titre (qu’il s’agisse d’un « I. » « 1) » ou « A. ») lors du stylage pour qu’ils ne soient pas dédoublés. Les titres doivent être laissés comme ils le sont dans ce document.

Les cas particuliers (tableaux, images, listes …)

Le cas des tableaux

Il est tout à fait possible d’intégrer des tableaux au sein des manuels, leur traitement est automatique, il n’y a donc pas besoin de les styler avec des éléments particuliers. Il est toutefois possible de styler des éléments de textes au sein des cellules des tableaux (paragraphe, portion de code informatique, *etc.*).

Le cas des images

Les images n’ont pas besoin d’être stylées de manière particulière, néanmoins il est essentiel de styler la légende ou le titre avec le style qui leur est dédié. En mode Brouillon les images n’apparaissent plus, et son remplacées par un espace blanc stylé en mode « Normal », il convient donc le laisser tel quel.

|  |  |
| --- | --- |
| **Les images** | |
| MASAtitreImg | Le titre et/ou la légende d’une image |

Le cas des listes

Le traitement des listes est automatique, il est toutefois possible de styler des éléments présents au sein des listes comme des paragraphes, des portions de code informatique, *etc*.

Les notes de bas de page

Le traitement des notes de bas de page se fait également de manière automatique. Il convient seulement de s’assurer que la numérotation de celles-ci ne se réinitialise pas à chaque page et donc que la numérotation soit continue. En effet, la notion de page étant perdue lors d’un export vers un format internet, les notes seront toutes situées à la fin du texte à la suite de ce dernier. Un lien dynamique est ajouté de manière automatique afin de faciliter la lecture.

Les sommaires et table des matières

Les sommaires sont mis en place automatiquement lors de la transformation qui s’opère par le biais de la plateforme MASA, il est donc nécessaire de supprimer les sommaires et les tables des matières au sein du manuel.

Exemple de stylage

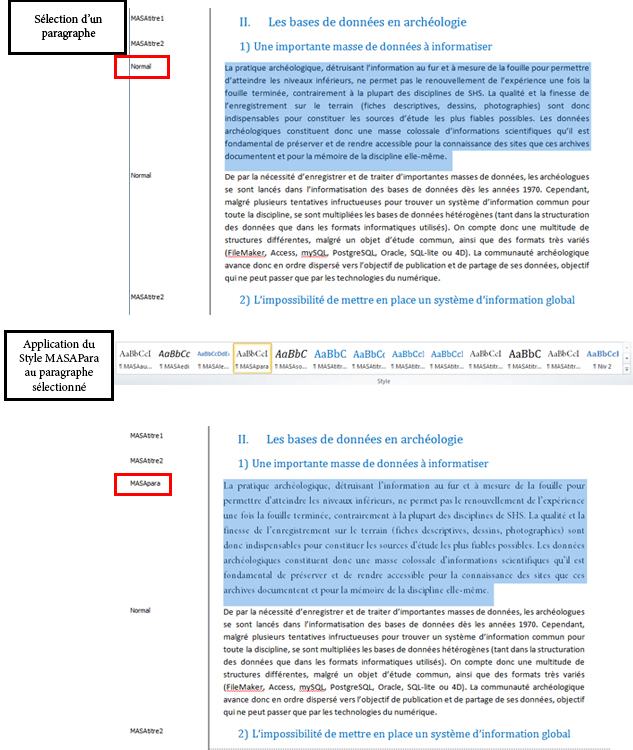


Figure 6: exemple du stylage d’un paragraphe

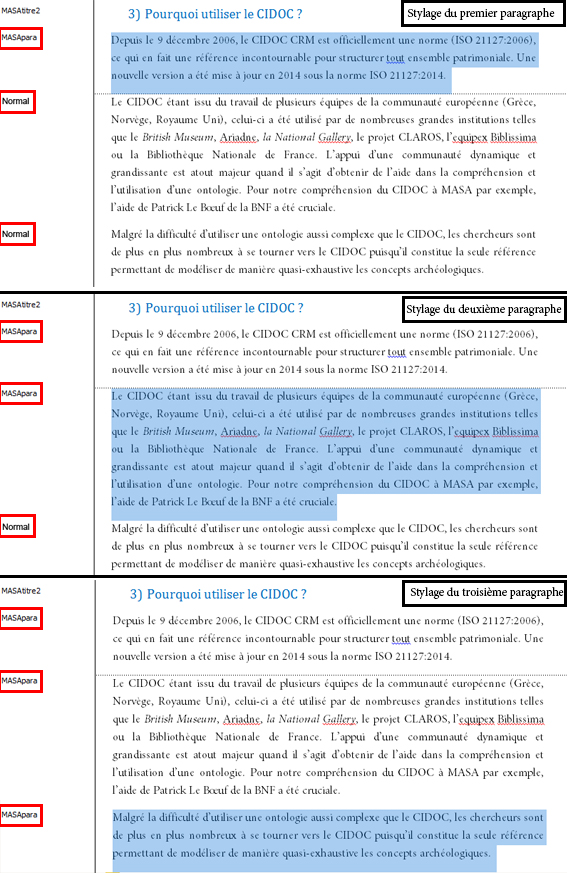


Figure 7 : exemple du stylage de plusieurs paragraphes

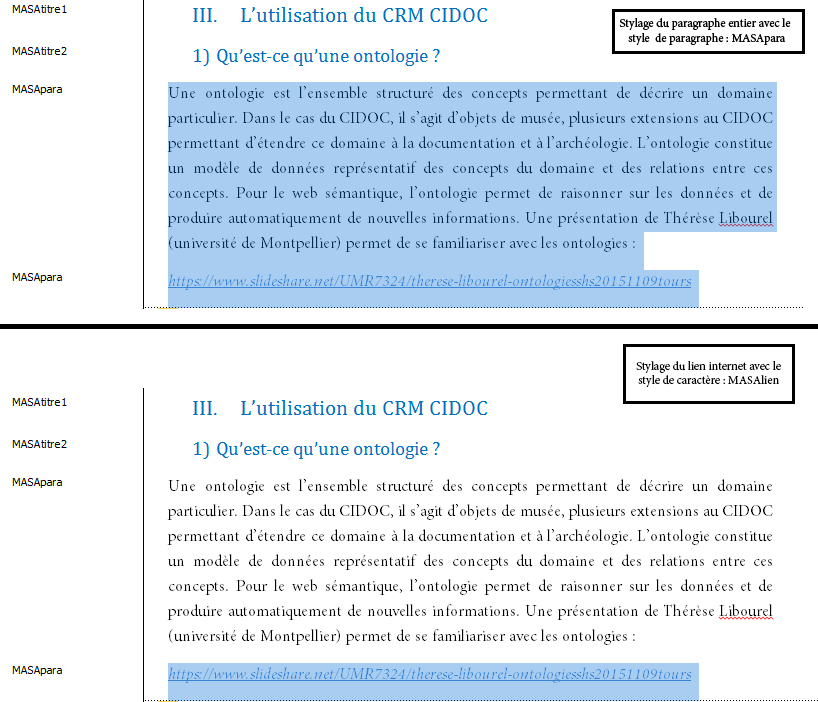


Figure 8 : exemple d’utilisation d’un style de caractère

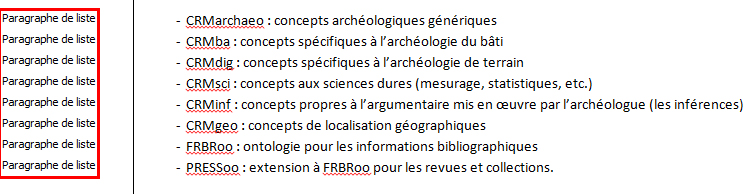


Figure 9 : Style automatique pour les listes

1. Le XML, eXtensible Markup Language, est un langage informatique de balisage générique, c’est-à-dire qu’il s’écrit en utilisant des balises qui permettent de structurer et de hiérarchiser un document. Pour en savoir plus sur le XML : <https://openclassrooms.com/courses/structurez-vos-donnees-avec-xml/qu-est-ce-que-le-xml> [↑](#footnote-ref-1)
2. Le HTML, HyperText Markup Language, est un langage qui permet de mettre en forme des contenus via un ensemble de balises. Pour visualiser une page en HTML, il est nécessaire d’utiliser un navigateur web. Pour en savoir plus sur le HTML : <http://glossaire.infowebmaster.fr/html/> [↑](#footnote-ref-2)
3. La TEI, Text Encoding Initiative, est un schéma XML qui permet de décrire la structuration d’un texte pour en savoir plus sur la TEI : <http://www.tei-c.org/index.xml> [↑](#footnote-ref-3)
4. Pour en savoir plus le concept de stylage et d’encodage des fichiers issus de traitement de texte, se reporter à l’article de Frédéric Glorieux sur le sujet : <http://resultats.hypotheses.org/267> [↑](#footnote-ref-4)
5. Le logiciel OpenOffice est téléchargeable à cette adresse : https://www.openoffice.org/fr/ [↑](#footnote-ref-5)