***[Nom\_du\_corpus]***

**Modèle des métadonnées**

*(Utiliser lors du versement dans le dépôt de données)*

**Rappel**: Il est utile de rappeler que les principes FAIR ne sont que des recommandations qui visent à rendre les données scientifiques plus accessibles et plus compréhensibles et leurs sauvegarde et réutilisation plus pérennes. La *FAIRititude* [FAIRness] complète des données n’est cependant qu’un objectif théorique difficilement atteignable. Par conséquent, chaque champ des métadonnées doit être rempli le plus complètement possible, mais dans des limites raisonnables.

|  |  |  |
| --- | --- | --- |
| **Champ** | **Qualité+** | **Métadonnées** |
| Identifiant (PID):  *(L’identifiant est attribué automatiquement par le dépôt lors de la mise en ligne).* | **obligatoire** | *attribué automatiquement par le dépôt lors de la mise en ligne* |
| Date de la mise en ligne (des données): | **obligatoire** | *attribué après la mise en ligne* |
| Intitulé (du jeu de données du corpus): | **obligatoire** |  |
| Auteur(s) du corpus (affiliation):  *(Par “l’auteur du corpus” on entend la ou les personne(s) qui ont créé le jeu de données et non la ou les personne(s) qui ont écrit les textes des documents contenus dans le corpus. Par exemple, si Jean-Paul Machin, chercheur au CNRS, a créé le corpus qui contient les textes de Dante Alighieri, on retient, comme auteur du corpus, J.-P.Machin et non Dante Alighieri.)* | **obligatoire** |  |
| Contributeur(s) (affiliation, rôle):  *(Il s’agit de toutes les personnes qui ont contribué, d'une façon ou d'une autre, à l’élaboration du corpus. Voir* infra *la liste des différents rôles possibles des contributeurs telle qu’elle est proposée par le dépôt Zenodo)****\**** | **optionnel** |  |
| Description (des données):  *(Présenter ici le corpus, ces données, le programme de recherche qui a permis la construction de ce corpus, etc.)* | **obligatoire** |  |
| Format de fichier (FAIR: I1):  *(Voir:* <https://en.wikipedia.org/wiki/List_of_file_formats>*)* | **optionnel** |  |
| Logiciels spécifiques nécessaires pour la lecture des données: (nom, version, système d’exploitation) (FAIR: I1):  *(Par exemple des logiciels de lexicométrie pour les données lexicostatistiques ou des logiciels de l’analyse de réseaux pour les données prosopographiques, etc.*  *Les logiciels de traitement de texte (format doc, txt, csv, etc.), de bases de données (format sql), de visionnage des images (jpg, tiff, etc.), de lecture des fichiers XML (sauf des cas particuliers comme, par exemple, les logiciels lexicométriques TXM ou Philologic4 qui s’appuient sur un formatage XML particulier), etc. ne sont pas des logiciels spécifiques et ils ne doivent pas être mentionnés.)* | **optionnel** |  |
| Si format ou logiciel particulier, expliquer clairement pourquoi ils ont été choisis (FAIR: I1): | **optionnel** |  |
| Si les données ne sont pas explicites en l’état, renvoyer vers les explications/documentation/vocabulaire pour expliciter les données (FAIR: I2): | **optionnel** |  |
| Provenance et historique des données (FAIR R.1.2):  *(Il s’agit d’indiquer l’historique et la provenance d'un jeu de données en tant qu’objet digital. Par exemple, les données peuvent être produites lors d’un autre projet et ensuite réutilisées, corrigées et modifiées par le présent projet.*  *Le contenu de ce champ peut, par exemple, être le suivant: “Les données qui concernent les VIIIe et IXe siècles ont été extraites d’un jeu de donnée sous format sql produit lors du projet “X” et corrigées ensuite par Monsieur Untel. Les données qui concernent le XIIe siècle ont été extraites d’un jeu de donnée sous format xml produit lors du projet “Y” sans aucune correction supplémentaire. Toutes les autres données ont été saisies manuellement par Monsieur Unautre et Madame Uneautre sous format des documents JSON dans la base MongoDB”.*  *Pour plus de détail sur la notion de la provenance des données, voir par exemple:* <https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/28>*)* | **optionnel** |  |
| Standards, conventions des données utilisées (FAIR: R1.3):  *(Bien qu’un “standard” commun ne semble pas exister en sciences humaines et sociales, certains projets ou programmes peuvent s’appuyer sur des conventions similaires. Par exemple, l’organisation des données dans le même type de tableau, l’utilisation des mêmes principes de balisage, etc.).* | **optionnel** |  |
| Version (des données):  *(Il est conseillé d’utiliser la nomencalutre du “versioning semantic”. Pour plus de détails sur le “versioning semantic” voir* [semver.org](http://semver.org) *. Pour la première version d’un jeu de données on met habituellement “1.0.0”)* | **optionnel** |  |
| Langue principale des données:  *(Si, par exemple, les textes sont en latin, mais ils sont accompagnés par les analyses en français, on indique ici la langue principale des données qui est donc le latin).* | **optionnel** |  |
| Mots-clés: | **optionnel** |  |
| Droit d'accès et licence des données (FAIR: R1.1):  *(Il existe différents types du droit d’accès et de licences. Voir infra la liste des droits d’accès proposée par le dépôt Zenodo. Le nombre de différentes licences est bien trop élevé pour les mentionner toutes).****\*\**** | **obligatoire** |  |
| Financement(s):  *(Il s'agit de tous les financements qui ont permis l’élaboration de ce corpus. Si possible, le financement doit être accompagné des références précises qui permettent son identification. Par exemple: ANR-14-CE31-022 " ÉCRIT ET GOUVERNANCE…”).* | **optionnel** |  |
| Données liées (FAIR: I3):  *(Mentionner, le cas échéant, le PID, l’intitulé des données liées et le type de lien. Par exemple “ces textes (PID, intitulé du jeu de données avec les textes)* est la transcription *de ces photos de manuscrits (PID, intitulé du jeu de données avec les images) ”* | **optionnel** |  |

**+** Champ obligatoire dans le dépôt de données Zenodo.

**\* Liste des rôles des contributeurs (proposé par le dépôt Zenodo):**

* Contact person
* Data collector
* Data curator
* Data manager
* Distributor
* Editor
* Hosting institution
* Other
* Producer
* Project leader
* Project manager
* Project member
* Registration authority
* Related person
* Research group
* Researcher
* Rights holder
* Sponsor
* Supervisor
* Work package leader

**\*\* Les différents types des droits d’accès (proposé par le dépôt Zenodo):**

* Open Access
* Embargoed Access
* Restricted Access
* Closed Access