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

**Questionnaire détaillé**

*(Évaluation complète de la correspondance des données aux principes FAIR)*

Ce questionnaire peut-être utiliser lors de l’évaluation complète de la correspondance des données d’un corpus aux principes FAIR. À noter toutefois que ce questionnaire s’appuie sur l’hypothèse du versement des données dans un dépôt institutionnel de données quand la partie “métadonnées” est garantie avant tout par le dépôt. Dans cette optique, l’auteur des données garantit essentiellement les principes FAIR liés aux données. Cela explique le remplacement dans certaines questions du vocable “(méta)données” uniquement par le vocable “données”.

|  |  |
| --- | --- |
| **1. Données administratives** | |
| Nom du corpus |  |
| Adresse web du site où les données sont disponibles actuellement (sinon, adresse web du site du projet) |  |
| Responsable scientifique |  |
| Responsable technique (si différent) |  |

|  |  |
| --- | --- |
| **2. Données techniques** | |
| Quelle est la nature des données (textuelle, visuelle, sonore, etc.)? |  |
| Quels sont les formats des données utilisés (txt, csv, xml, jpg, mysql, etc.)? |  |
| Quels sont les standards des données utilisés (e.g.: XML / DTD, modèle personnel, etc.)? |  |
| Quelle est la taille des fichiers contenant les données (en mb)? |  |
| Quelles sont les unités de données (par exemple actes, personnes, images) et combien d’unités contiennent les données?  *(Par exemple 4000 actes, 232 personnes, 59 images)?* |  |
| Est-ce que des logiciels spécifiques sont nécessaires pour la lecture des données?  *(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.)* |  |
| Quelle est la plateforme d’affichage et de stockage des données (plateforme spécifique, CMS, site dédié, etc.)? |  |
| Par qui les données sont-elles hébergées?  *(Par exemple un site institutionnel, un site personnel hébergé sur un serveur commercial, etc. Si les données ne sont pas en ligne, elles peuvent être stockées sur un ordinateur personnel)* |  |

|  |  |
| --- | --- |
| **3. Correspondance aux principes FAIR** | |
| F1: Les données possèdent-elles un identifiant persistant et unique au monde (PID)?  *(Par exemple DOI, ARK, ISBN, etc.)* |  |
| F2: Les données sont-elles décrites avec des métadonnées détaillées? |  |
| F3:Le PID des données (voir F1) est-il mentionné dans les métadonnées? |  |
| F4: Les données et les métadonnées y associées sont-elles enregistrées ou indexées dans une ressource consultable (par exemple, un dépôt de données)? |  |
| A1.1: Les données sont-elles accessibles via un protocole de communication ouvert, gratuit et largement utilisé?  *(Il s’agit avant tout des protocoles informatiques libres et connus comme HTTP (=site web), FTP etc. Les autres “protocoles de communications” qui s’appuient sur des logiciels propriétaires et non universellement implantés (par exemple Skype, Microsoft Exchange Server, etc) ne répondent pas à ce critère).* |  |
| A1.2: Les données sont-elles accessibles via un protocole de communication qui permet, si nécessaire, de mettre en place une procédure d'authentification et d’autorisation?  *(FAIR data ne signifie pas OpenData. Le protocole via lequel les données et les métadonnées sont accessibles, doit avoir la possibilité, en cas de besoin, de restreindre l’accès aux données, par exemple uniquement aux utilisateurs possédant un login et un mot de passe.)* |  |
| A2: Les métadonnées seront-elles accessibles même lorsque les données elles-mêmes ne seront plus disponibles?  *(Normalement, ce critère est considéré comme rempli, si la condition F4 est satisfaite).* |  |
| I1: Les données utilisent-elles un langage connu et largement utilisé?  *(Par exemple XML, SQL, TXT, CSV, etc.)* |  |
| I2: Est-ce que les données sont explicites et compréhensibles dans l’état? Si non, est-ce qu’elles sont accompagnées d’une documentation détaillée?  *(Cela signifie que toutes les abréviations, tous les termes ambigus ou toutes les balises utilisées dans la description des données doivent être clairement expliqués ou avoir déjà une documentation existante).* |  |
| I3: Dans le cas où les présentes données peuvent être reliées à d’autres données (par exemple, les données textuelles peuvent être reliées aux images des manuscrits), est-ce que les métadonnées contiennent les références à ces données reliées (par exemple le PID des données reliées, voir F1)? Si oui, est-ce que le type de lien entre les données est explicitement mentionné? (Par exemple, “*ce texte est la transcription de ce manuscrit*” ou “*ce texte a été écrit par cette personne*”.) |  |
| R1.1: Est-ce que les données disposent d’une licence et est-ce que cette dernière est clairement mentionnée (par exemple dans les métadonnées)?  *(Les données peuvent avoir tous types de licence, y compris les licences “non libres de droits”).* |  |
| R1.2: Est-ce que les données contiennent (par exemple dans les métadonnées) un historique détaillé de la provenance de ces données?  *(Par exemple, les données peuvent être produites lors d’un autre projet et ensuite réutilisées, corrigées et modifiées par le présent projet).* |  |
| R1.3: Est-ce que les données utilisent des standards de présentation, d’organisation ou d’archivage communément utilisés dans le domaine scientifique auquel ces données appartiennent?  *(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. Si c’est le cas, ces conventions doivent être mentionnées, par exemple dans les métadonnées. Ne pas confondre avec I2. Le I2 vise à rendre les données compréhensibles / interopérables grâce à la documentation associée à ces données. Le R1.3 vise à faciliter la réutilisation des données grâce aux indications générales sur les standards et les conventions sur lesquels les données s’appuient. )* |  |