|  |  |  |
| --- | --- | --- |
| **Point de contrôle qualité** | **Problème observé et commentaire** | **Correction : comment ? importance ?** |
| 1. le jeu de données est difficilement accessible (format “image”, PDF, HTML...), le fichier est mal formé |  |  |
| 2. la licence est absente ou inhabituelle, le jeu de données n’est pas “open data” |  |  |
| 3. le fichier fait peu appel aux standards répandu et aux données pivots |  |  |
| 4. le fichier est mal documenté |  |  |
| 5. il existe des problèmes de syntaxe |  |  |

|  |  |  |
| --- | --- | --- |
| **Point de contrôle qualité** | **Problème observé et commentaire** | **Correction : comment ? importance ?** |
| 6. Valeurs aberrantes, suspectes, inexplicables, pas crédibles |  |  |
| 7. Il manque des données et cela n’est pas documenté (trous, données tronquées, valeurs vides, granularité / fréquence / maillage / fraîcheur) |  |  |
| 8. Trop de données : doublons, inutilement vieilles, précision / fréquence / maillage / fraîcheur |  |  |
| 9. Données posant problème avec la réglementation (données perso, relatives à la santé, la religion..., propriété littéraire et artistique, etc.) |  |  |
| 10. Les contenus posent problèmes : synonymies, non traduits (USA), cryptique (DAECPP), utilisation du 0 au lieu du “null”... |  |  |