

# Guide de relecture d'un jeu de données avant publication (Qualité de l'entrepôt de données *DataSuds*)

Version 2.4, janvier 2022 – Luc Decker (administrateur de DataSuds <https://dataverse.ird.fr>),  
Hanka Hensens, Caroline Doucouré et Pascal Aventurier – Service IST/MCST, IRD, France



## Métadonnées

Champs	Réf.	Préconisations, recommandations - Conseils pratiques	Finalité et commentaires
tous	L1	De préférence, saisir toutes les informations en anglais. <a href="#">Dans ce cas, l'interface en anglais facilite la saisie : avant d'ouvrir les formulaires, changer de langue grâce au menu en haut à droite de l'écran.</a>	Visibilité et valorisation du jeu de données : recommandé, comme pour les articles scientifiques.
	L2	Ne pas mélanger différentes langues, sauf éventuellement pour le champ « Description » ( <a href="#">ajouter une séparation entre les langues avec le code &lt;hr&gt;</a> ) ainsi que pour le champ « Autre titre » (possibilité de renseigner une traduction : un titre dans une langue locale peut améliorer la visibilité du jeu de données.) et pour les mots-clés. <a href="#">La langue sélectionnée dans le menu s'applique uniquement à l'interface utilisateur.</a>	Clarté de la présentation du jeu de données. Dans l'entrepôt, les champs de métadonnées ne sont pas multilingues ; sauf exception, il n'est pas possible de saisir des informations traduites dans différentes langues.
	L3	Orthographe et grammaire : vérifier tous les textes saisis dans les métadonnées. <a href="#">Lorsque les saisies sont terminées, se rendre dans l'onglet « Métadonnées », copier/coller tout le texte dans un logiciel de traitement de texte (tel que Word), sélectionner la langue utilisée et lancer la vérification automatique de l'orthographe.</a>	Prouve que les informations ont été relues avant publication, témoigne du soin apporté dans la gestion des données. Réputation de l'entrepôt, des auteurs et de leur laboratoire.
Titre	T1	Caractérisation des données : type de données, contexte, période de collecte, localisation géographique - si applicable et pertinent. Si un autre jeu de données risque d'avoir le même titre, alors le titre devrait être plus spécifique. Autres possibilités : « <i>Replication data for...</i> » / « <i>Supplementary data for...</i> » [insérer le titre de l'article scientifique associé aux données]. <a href="#">Exemples : consulter les jeux de données publiés récemment dans DataSuds. Pour davantage de conseils : https://coop-ist.cirad.fr/rediger/article-scientifique/le-titre/1-le-titre-premier-niveau-de-selection-sur-le-web</a>	Selon la formulation et la précision du titre, un utilisateur de données potentiel ira - ou non - consulter plus en détails le jeu de données. Le titre doit être compréhensible par lui-même.
	T2	Longueur appropriée, approximativement entre 3 et 20 mots	Suivre les usages, comme pour un article scientifique.
	T3	Retirer les éléments qui n'ont en général pas leur place dans un titre : noms de fichiers, noms d'auteurs, citation complète d'un article, codes ou références internes au projet, parenthèses inutiles, caractères spéciaux. Le nom du projet peut être inclus mais ne constitue pas un titre à lui seul.	
Auteurs	A1	Personnes qui ont contribué à la production des données : rôle scientifique ou technique : conception, collecte, traitement, analyse. Le responsable du projet valide la liste des auteurs. <a href="#">Conseils : 1) Utiliser le bouton  pour ajouter des lignes au formulaire. 2) Attention, ce champ est prérempli avec le nom de la personne qui dépose (techniquement) les données : cette personne n'est pas nécessairement 1er auteur, parfois pas même auteur ... à éditer si nécessaire. 3) Une approche consiste à copier la liste des auteurs d'un article associé aux données - ou encore de modifier cette liste pour ajouter, mettre en avant des intervenants ayant joué un rôle important dans la collecte ou le traitement des données.</a>	Procéder comme pour un article scientifique dans le choix et l'ordre des auteurs.  Se référer aux <a href="#">Data Citation Principles</a>

	A2	Format : « Nom, Prénom », avec noms et prénoms en lettres minuscules. Exemple : Dupont, Jean	Suivre les usages, comme pour un article scientifique. Le format des noms d'auteurs est repris dans la citation du jeu de données qui est présentée sur sa page d'accueil.
<b>Affiliations des auteurs</b>	A3	Format : Structure 1 – Tutelles (si applicable) – Pays ; Structure 2 – Tutelles – Pays ; ... Exemple : "UMR DIADE – IRD, University of Montpellier, CIRAD, CNRS – France" <a href="#">Les informations préremplies dans le formulaire doivent en général être modifiées.</a>	Informations complètes sur les auteurs, utiles pour savoir où ils travaillent, pour les contacter, notamment en cas d'homonymie.
<b>Identifiants ORCID</b>	A4	Saisir les identifiants ORCID des auteurs lorsqu'ils sont connus. <a href="#">Ne pas retarder la publication s'il manque quelques identifiants, ils pourront être ajoutés ultérieurement.</a>	ORCID devient un standard.
<b>Description</b>	D1	Précision générale : est-elle suffisante pour le référencement adéquat des données par les moteurs de recherche ?	Principe FAIR « Facile à trouver » : les autres chercheurs potentiellement intéressés vont-ils trouver facilement ces données ? Quels mots clés sont-ils le plus susceptibles de saisir lorsqu'ils lancent une recherche ?
	D2	Contexte, périmètre, typologie des données - Résumer le projet scientifique associé ou/et <i>l'intérêt, l'objectif</i> de ces données - Résumer la liste des données déposées (« quoi ? ») ; comment, où et quand ont-elles été collectées/traitées ?	Répondre aux questions que les utilisateurs potentiels sont susceptibles de se poser avant d'aller plus loin et de commencer à utiliser les données. Participe au bon référencement du jeu de données. Principe FAIR « Réutilisable ».
	D3	Possibilité d'ajouter des liens cliquables vers des pages web telles que la description du projet, le site du bailleur, etc... (code <code>&lt;a href='https://...&gt; texte du lien&lt;/a&gt;</code> )	Facilite la démarche des utilisateurs et en général apprécié par les sites référencés.
	D4	Mise en page de la description : créer éventuellement des paragraphes ou/et insérer des sauts de ligne à l'aide du code <code>&lt;br&gt;</code> . <a href="#">Possibilité de mettre du texte en gras (code <code>&lt;b&gt;...&lt;/b&gt;</code>) ou en italique (code <code>&lt;i&gt;...&lt;/i&gt;</code>).</a>	Facilite la lecture de la description. Améliore la présentation du jeu de données
<b>Mots-clés</b>	D5	Saisir au moins 4-5 mots-clés. Précision : comme pour le champ « Description ». <a href="#">Utiliser un vocabulaire (thesaurus) de référence dans son domaine facilite la découverte par les scientifiques de ce domaine.</a>	Suivre les usages, comme pour un article scientifique.
<b>Publication connexe</b>	P1	Saisir les références complètes des publications liées au jeu de données. Préciser « <i>(submitted)</i> » le cas échéant. <a href="#">Utiliser le bouton  pour ajouter des lignes au formulaire.</a>	Les articles citent les données, avec leur identifiant DOI, et réciproquement. Augmente les citations.
	P2	Saisir l'identifiant pérenne (DOI) et/ou le lien (http...) des publications.	Facilite la navigation des utilisateurs.
<b>Renseignements sur la subvention</b>	M1	Ce champ optionnel permet de citer les bailleurs qui ont permis la réalisation du projet. Le contrat conclu avec un bailleur impose parfois que cela soit effectué systématiquement. Il est possible d'ajouter des liens vers des sites web.	Bonnes relations avec les bailleurs qui ont également besoin de visibilité et de montrer l'impact des financements accordés.

<b>Type de données, Période de collecte, Langue</b>	M2	Remplir ces différents champs de métadonnées (lorsqu'ils sont applicables) aide à bien décrire les données, bien que le formulaire de saisie autorise qu'ils soient laissés vides. <i>Note : le formulaire qui permet d'éditer les métadonnées comprend davantage de champs que le formulaire initial utilisé pour créer un nouveau jeu de données.</i>	Référencement des données et interopérabilité. Aide à la réutilisation.
<b>Métadonnées géospatiales</b>	M3	Ce champ de métadonnées est à saisir si les données ont été collectées dans un/des périmètre(s) géographique(s) déterminé(s) : pays, villes..., sauf si la localisation n'a aucune importance.	

## Fichiers déposés

Elément	#	Préconisations ou/et recommandations - Conseils pratiques	Finalité et commentaires
<b>Choix des fichiers de données à diffuser</b>	F1	<p>Il n'est pas toujours autorisé de diffuser (de partager) librement toutes les données d'un projet de recherche. Il convient de respecter des obligations légales et contractuelles, de vérifier également la propriété des données (exemples : partenaires impliqués ; réutilisation de données existantes : licences ou accords des fournisseurs). Les principaux points de vigilance sont décrits sur <a href="https://data.ird.fr/cadre-juridique/">https://data.ird.fr/cadre-juridique/</a></p> <p>Apporter une attention particulière en cas de données personnelles (exemple : résultats d'enquêtes, interviews) ou de données dites « sensibles », en particulier en santé et sciences sociales. Pour des raisons réglementaires, la publication <i>en libre accès</i> de données personnelles sensibles n'est pas autorisée, même si elles ont été pseudonymisées ou dé-identifiées : elles conservent leur caractère personnel. Sous certaines conditions, une publication en accès restreint est possible. Prendre contact (<a href="mailto:data@ird.fr">data@ird.fr</a>) pour conseil. Si les données ont été totalement <i>anonymisées</i>, des éléments concrets doivent prouver la qualité de ce traitement : appel à un expert reconnu, logiciel utilisé, indicateurs statistiques (k-anonymat, l-diversité, ...). Vérifier de plus que le formulaire de consentement ne s'oppose pas de manière explicite au partage des données ou à leur réutilisation.</p> <p>Dans presque tous les cas (sauf obligation de confidentialité portant sur l'existence d'un projet), il reste possible de publier seulement les métadonnées dans l'entrepôt, sans déposer de fichier de données.</p> <p>L'entrepôt permet de restreindre l'accès à certains fichiers, sous la forme d'un formulaire de demande d'accès qui devra être rempli par les utilisateurs : des autorisations d'accès leur seront accordées au cas par cas selon des conditions que le déposant aura définies.</p> <p>Il est recommandé de préserver également les données brutes dans l'entrepôt, en complément des données traitées, dérivées ou d'intérêt - d'autant plus si l'espace occupé par les données brutes est négligeable par rapport aux capacités de stockage actuelles.</p>	<p>Il ne s'agit pas d'ouvrir systématiquement et indistinctement l'accès à toutes les données : « aussi ouvert que possible, <i>aussi fermé que nécessaire</i> ».</p> <p>Des fichiers de données peuvent aussi être déposés ultérieurement, après vérifications complémentaires. Un jeu de données peut être facilement mis à jour, sous la forme de nouvelles versions (V2, V3, ...). Les utilisateurs peuvent consulter l'historique des versions et, par défaut, accéder aux versions antérieures.</p> <p>Conserver la possibilité de retraiter les données brutes pour 1) appliquer, tester, comparer de nouvelles méthodes (statistiques, intelligence artificielle, ...) qui pourraient être conçues dans le futur, 2) corriger ou améliorer un traitement de données, par exemple en cas d'erreur dans un logiciel qui serait découverte ultérieurement.</p>

<b>Volume de données</b>	<p>F2 La taille de <i>chaque fichier</i> déposé ne doit pas dépasser 975 Méga-octets. Cette limite est à considérer <i>après compression</i> éventuelle du fichier. La limite est arbitraire, fixée par la configuration de DataSuds ; il est possible de l'augmenter.</p> <p>Un jeu de données peut rassembler des dizaines ou même centaines de fichiers. Avant de déposer un grand volume de données (&gt; 5 Gb), prendre contact pour conseil auprès de l'équipe support de DataSuds : <a href="mailto:data@ird.fr">data@ird.fr</a></p> <p><i>Immédiatement après avoir été déposés, les fichiers ZIP sont automatiquement décompressés par l'entrepôt : leur contenu est extrait et présenté aux utilisateurs sous forme d'une liste de fichiers.</i></p> <p><i>Si l'on souhaite diffuser des données au format ZIP, créer et déposer un fichier ZIP temporaire qui contient lui-même le fichier ZIP qui doit être présenté tel quel aux utilisateurs.</i></p>	<p>DataSuds n'a pas été conçu pour les données dites « massives », en particulier pour les résultats bruts de séquençage génomique. Des entrepôts thématiques (par exemple <i>GenBank</i>) sont spécialisés dans la préservation et la diffusion de telles données.</p> <p>Il n'est pas toujours réalisable de conserver l'ensemble des données brutes d'un projet lorsque leur volume est très important. L'intérêt des données doit être mis en balance avec les coûts engendrés par leur stockage à très long terme. Il est parfois moins coûteux de préserver physiquement les échantillons. Même si l'utilisation de DataSuds est gratuite, la question est à considérer.</p>
<b>Nombre de fichiers</b>	<p>F3 Déposer au maximum quelques centaines de fichiers par jeu de données. Au-delà et si nécessaire, diffuser les fichiers de manière regroupée dans des archives au format ZIP, <i>suivant la méthode décrite ci-dessus.</i></p> <p><i>Pour déposer plusieurs fichiers en une seule passe, les regrouper également dans une archive ZIP.</i></p> <p>A l'opposé, il est possible de publier un jeu de données qui ne comprend aucun fichier, donc uniquement des métadonnées, et d'indiquer aux utilisateurs la procédure à suivre pour obtenir les données : personne à contacter, lien vers un site externe, attente de la fin d'un embargo, etc...</p>	
<b>Noms des fichiers</b>	<p>F4 Adopter des noms de fichiers spécifiques au projet et au contenu.</p> <p>Autant que possible, leur ajouter un préfixe propre au projet (un acronyme ou un identifiant) qui sera commun à tous les fichiers déposés, tel que « PROJETABC_stationN_dailyflow.csv ». Eviter les noms génériques tels que « data.csv », « datasuds_table1.csv », « tableau4.xls », « documentation.pdf » ...</p> <p><i>L'interface de DataSuds permet de renommer un fichier sans avoir à le redéposer : ouvrir le formulaire d'édition des métadonnées du fichier.</i></p>	<p>Evite que les utilisateurs ne mélangent ou confondent des fichiers aux noms identiques provenant de différentes sources. Ajouter un identifiant à la fin des noms de fichiers ne permet pas de les trier convenablement : un préfixe est préférable.</p> <p>Par défaut, Dataverse trie les fichiers par leur nom. Cela permet de prévoir l'ordre dans lequel les fichiers seront présentés aux utilisateurs et de les organiser.</p>
	<p>F5 Dans les noms de fichiers, utiliser uniquement des lettres, chiffres et caractères séparateurs (- ou _). Remplacer les autres caractères (spéciaux, accentués) et les espaces.</p>	Principe FAIR « Interopérable »
	<p>F6 Limiter les noms de fichiers à une longueur raisonnable : au plus 40 caractères sauf besoin particulier.</p>	

<b>Format des fichiers</b>	<p>F7 Autant que possible, déposer les fichiers dans un <b>format ouvert</b> tel que TSV (CSV) ou texte... se référer à <a href="https://fr.wikipedia.org/wiki/Format_ouvert">https://fr.wikipedia.org/wiki/Format_ouvert</a> et <a href="https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf">https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf</a></p> <p>De plus, afin de <i>rendre possible</i> leur archivage sur le long terme (un service actuellement en projet), les données devraient être déposées dans l'un des 64 formats actuellement acceptés par le CINES, consulter <a href="https://facile.cines.fr">https://facile.cines.fr</a>. Les données déposées dans d'autres formats seront exclues de cet archivage.</p> <p>Dans le cas de fichiers au format texte, utiliser de préférence l'encodage de caractères UTF-8 (<a href="#">un outil gratuit pour éditer et convertir les fichiers texte : Notepad++</a>). Dans le cas de documents PDF, les enregistrer avec l'option « format PDF/A ».</p> <p><i>Attention : avec la version française de Windows, les fichiers CSV enregistrés avec Excel utilisent le caractère point-virgule comme séparateur de colonnes. L'application Dataverse s'attend à ce que le caractère séparateur soit une virgule, selon le standard international : elle ne reconnaît pas correctement le contenu du fichier. Pour remédier à ce problème, enregistrer au format « Texte (séparateur : tabulation) » dont le caractère séparateur est mieux standardisé et renommer l'extension .TXT du fichier en .TSV. Autre solution : produire le fichier TSV à l'aide du logiciel gratuit LibreOffice qui permet de choisir le caractère séparateur aussi bien à l'ouverture qu'à la sauvegarde d'un fichier, grâce à l'option « Editer les paramètres du filtre ». Au cours de la conversion d'un fichier Excel, prêter attention à la présence de plusieurs pages, produire 1 fichier TSV par page.</i></p> <p>Si la conversion des données dans un format ouvert occasionne une perte d'information (exemple d'un tableau Excel : perte de la mise en page, couleurs et formatage), déposer également le fichier original en plus du fichier converti.</p>	Principe FAIR « Interopérable »
<b>Description des fichiers</b>	<p>F8 Remplir le champ « Description » attaché à chaque fichier, dans le formulaire d'édition des métadonnées des fichiers : résumer le contenu en quelques mots ou davantage.</p> <p>F9 Attribuer un ou plusieurs libellés (<i>tags</i>) à chaque fichier, comme le permet l'entrepôt : « Data », « Documentation », « Code ». Il est possible de créer des libellés personnalisés. <a href="#">Passer au préalable l'interface en anglais, pour ne pas utiliser la version française des libellés.</a></p>	<p>Aider les utilisateurs à identifier le contenu de chaque fichier</p> <p>Organisation, catégorisation des fichiers déposés</p>
<b>Noms de dossiers</b>	<p>F10 Définir un nom de dossier pour chaque fichier déposé. Pour le moins, il peut s'agir d'un nom de dossier commun à tous les fichiers, correspondant au nom du projet de recherche.</p> <p><a href="#">Le nom de dossier (« Chemin d'accès au fichier ») peut être saisi manuellement dans le formulaire d'édition des métadonnées des fichiers.</a></p> <p><a href="#">Il est possible de créer une arborescence avec plusieurs niveaux de dossiers (tel que dossier1/dossier2).</a></p> <p><a href="#">Si un fichier ZIP contenant des dossiers est déposé dans DataSuds, son contenu est extrait automatiquement et les noms de dossiers préremplis.</a></p>	<p>Organisation, catégorisation des fichiers déposés.</p> <p>Lorsqu'un utilisateur sélectionne plusieurs fichiers à télécharger -ou l'ensemble des fichiers-, l'entrepôt crée une archive ZIP qui inclut les différents dossiers qui auront été définis. Cela permettra aux utilisateurs de recréer automatiquement l'arborescence des dossiers et de préserver l'organisation des fichiers. Cela évitera aussi que les fichiers extraits par les utilisateurs ne se mélangent à d'autres par erreur.</p>

<b>Dictionnaire des données</b>	F11 Si les données sont déposées sous la forme de tables (TSV, CSV, Excel...), il est essentiel de préciser la signification de chaque variable, « champ » ou « colonne ». Cette information est en général documentée sous la forme d'un dictionnaire des données ( <i>data dictionary</i> ). Un dictionnaire des données est constitué d'un ou plusieurs tableaux de référence au format TSV/CSV/texte avec les colonnes suivantes : « nom de la variable », « contenu/signification de la variable », « unité de mesure », « signification des différents codes utilisés » (si applicables) ; format (optionnel). Il peut être commun à plusieurs tables ou fichiers ; il peut aussi être intégré dans un fichier de données. Si de nombreux codes sont utilisés, il est pratique de les documenter dans des tableaux séparés ( <i>code dictionary</i> ).	Principe FAIR « Réutilisable » : en l'absence de la signification précise des variables et des codes utilisés, la réutilisation des données s'avère difficile ou même impossible ; elle augmente le risque de mauvaise interprétation des données.
<b>Documentations annexes</b>	F12 En complément des fichiers de données, l'entrepôt accepte aussi toutes les documentations qui vont aider à comprendre les données, à préserver leur histoire, les conditions de leur collecte ou de leur production.  Voici des suggestions de documentations qui peuvent accompagner des données : fiche de présentation du projet de recherche (éventuellement une présentation Powerpoint) ; figures, schémas, cartes, photographies... ; formulaires vierges de collecte de données ou/et de recueil du consentement des participants ; guide de l'enquêteur ; guide de traitement des données ; procédure ou algorithme de traitement des données, code informatique ; notes techniques ; Plan de Gestion de Données. Une exception : les articles scientifiques (dans toutes leurs versions, dont les <i>preprint</i> ) ne doivent pas être déposés dans DataSuds. Certaines documentations sont considérées comme des « œuvres de l'esprit », telles que les textes rédigés : demander l'accord de leurs auteurs avant diffusion.	Principe FAIR « Réutilisable » : 1) Améliorer le potentiel de réutilisation des données et l'autonomie des utilisateurs, sur le long terme 2) Aider les utilisateurs à comprendre pourquoi la reproduction de l'expérience pourrait produire un autre résultat, par exemple si certains paramètres ou conditions expérimentales sont différents.
	F13 Insérer la citation du jeu de données dans une ou plusieurs documentations déposées avec les données.  Dès la préparation d'un jeu de données, sa future citation peut être connue à l'avance. Le DOI est pré-attribué et sera mis en service lorsque le jeu de données sera publié. Le titre du jeu de données et la liste des auteurs doivent être finalisés, puis copier la citation à partir de la page d'accueil du jeu de données et remplacer « version provisoire » par « V1 ».	Faciliter la citation des données, qui peuvent être ré-utilisées longtemps après avoir été téléchargées. Aider les utilisateurs à conserver trace de la source des données.

## Conditions d'utilisation

Elément	#	Préconisations ou/et recommandations - Conseils pratiques	Finalité et commentaires
<b>Attribution d'une licence</b>	L1	<p>Si les données sont ouvertes, le formulaire en ligne <a href="https://creativecommons.org/choose/">https://creativecommons.org/choose/</a> aide à choisir un modèle d'une licence qui devra être acceptée et respectée par les utilisateurs qui téléchargent les données. Tenir compte des exigences éventuelles des partenaires et des financeurs du projet, ainsi que de la nécessité de préserver la confidentialité des données dans certaines situations : respect des réglementations, déontologie et éthique.</p> <p>Il est possible d'attribuer une licence distincte en fonction des contenus, comme par exemple aux documentations qui accompagnent les données.</p> <p>Par défaut, l'entrepôt DataSuds attribue une licence «CC-BY» (parfois «CC0» = domaine public) aux jeux de données. <a href="#">Consulter l'onglet « Conditions d'utilisation » du jeu de données et vérifier que la licence correspond bien à ce qui est souhaité.</a> La licence CC-BY exige que le jeu de données soit cité lorsqu'il est utilisé ; elle est conforme aux préconisations du MESRI pour les projets sur financement public. Des exigences de citations peuvent être précisées dans un champ dédié.</p> <p>Si les modèles de licence standard (<i>Creative Commons</i> ou autres) ne conviennent pas, indiquer les conditions particulières à remplir pour accéder aux données. Il peut s'agir de la nécessité de signer un accord (<i>Data Use Agreement</i>) : contacter <a href="mailto:data@ird.fr">data@ird.fr</a> pour obtenir un modèle. Fournir des informations de contact (si possible pérennes : pas une personne précise) pour la soumission des demandes d'accès aux données.</p>	<p>Principe FAIR « Réutilisable » : répondre à la question « Dans quelles conditions est-t-il possible de réutiliser ces données ? »</p> <p>Rappel : « Aussi ouvert que possible, aussi fermé que nécessaire », d'abord respecter les exigences réglementaires, éthiques et contractuelles.</p> <p>En cas de changement de licence postérieure à la publication des données, la nouvelle licence ne s'appliquera pas rétroactivement aux utilisateurs qui ont déjà obtenu les données.</p>
<b>Formulation et cohérence de la licence</b>	L2	<p>Si une licence est attribuée, sa formulation (en anglais si les métadonnées ont été saisies dans cette langue), son logo et les liens éventuels doivent être complets, reproduits dans leur intégralité.</p> <p>Si des conditions particulières sont imposées pour l'usage des données, la licence doit être cohérente vis-à-vis de ces dispositions. En particulier, dans le cas de la licence CC-BY, les utilisateurs qui obtiennent un accès aux données ont le droit de les rediffuser librement par eux-mêmes. Cette licence ne convient donc pas aux fichiers dont l'accès est restreint.</p> <p>Si les données sont disponibles par ailleurs (tel que sur le site web du projet), la licence doit être cohérente et équivalente avec les indications données par le service qui héberge les données.</p>	Validité juridique de la licence

# Qualité de l'entrepôt de données DataSuds : Grille de révision d'un jeu de données

Version 2.4, janvier 2022 – Luc Decker (administrateur de DataSuds, service IST/MCST, IRD)



Collection (dataverse)

Relecteur

Jeu de données

Date

Critère	OK	partiellement OK	à améliorer, requis	à améliorer, conseillé	non applicable
L1. Métadonnées en anglais de préférence	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
L2. Langue unique (sauf description, autre titre et mots-clés)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
L3. Vérification de l'orthographe et de la grammaire	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T1. Titre : précision, spécificité	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T2. Titre : longueur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
T3. Titre : formulation et format	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A1. Auteurs : choix et nombre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A2. Auteurs : format des noms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A3. Auteurs : format des affiliations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A4. Auteurs : identifiants ORCID	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D1. Description : précision, pour bon référencement	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D2. Description : contexte, périmètre, typologie des données	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D3. Description : liens vers d'autres contenus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D4. Description : mise en page	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D5. Mots-clés : nombre, précision	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P1. Publications associées : citations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
P2. Publications associées : DOI et/ou liens	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
M1. Citation des bailleurs de fonds	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
M2. Saisie de Type de données, Période de collecte, Langue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
M3. Saisie des métadonnées géospatiales	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F1. Fichiers : permission de diffuser, droits & réglementations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F2. Volume de données déposées	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F3. Nombre de fichiers déposés	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F4. Noms de fichiers spécifiques au projet et à leur contenu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F5. Caractères utilisés dans les noms des fichiers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F6. Longueur des noms des fichiers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F7. Format des fichiers de données	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F8. Saisie des descriptions des fichiers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F9. Attribution de catégories (tags/libellés) aux fichiers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F10. Saisie de nom(s) de dossier(s)	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
F11. Dictionnaire des données (variables, codes, unités)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F12. Dépôt de documentations associées	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F13. Citation du jeu de données insérée dans la documentation	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>
L1. Choix d'une licence d'utilisation des données	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
L2. Formulation et cohérence de la licence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	OK	partiellement OK	à améliorer, requis	à améliorer, conseillé	non applicable