POLITIQUE DE PUBLICATION DES DATASETS SUR LA PLATEFORME SODHA

Ce texte a pour but de présenter les critères d'évaluation pour la publication de datasets soumis par leurs déposants sur la plateforme SODHA.

En rassemblant ces critères dans un même texte, il s'agit ici de présenter la « politique éditoriale » SODHA.

---

1. Vue d'ensemble des critères de publication

Les datasets soumis pour publication (*Submit for Review*) doit remplir les conditions suivantes pour être effectivement publiés :

- ☑ Si l'accès à un ou plusieurs fichiers du dataset a été restreint (*Restricted*), le déposant a renseigné les conditions d'accès dans le champ du même nom, « Conditions d'accès » (*Terms of Access*). (Convention de dépôt, art. 10 ; Politique d'accès et de réutilisation, b.)
- ☑ L'étude au cours de laquelle les données qui constituent le dataset ont été rassemblées s'inscrit dans une ou plusieurs disciplines des sciences sociales et/ou des humanités numériques. (voir ci-dessous, 2.1)
- ☑ Les données ne contiennent pas d'informations sensibles (ex. des informations à caractère personnel qui permettent d'identifier des individus). (voir ci-dessous, 2.2)
- ☑ Les métadonnées ne contiennent pas d'informations ambigües. (voir ci-dessous, 2.3)
- ☑ Les fichiers du dataset peuvent être téléchargés et ouverts sans problèmes.
- ☑ Lorsqu'une nouvelle version d'un dataset déjà publié est créée, les changements doivent être documentés dans une instance supplémentaire du champ *Description*.
- ☑ Le dataset ne contient pas de fichiers exécutables ou binaires.

---

2. Critères de publication

Les critères de publication de l'archive SODHA sont d'ordre **fondamental** (portant sur le contenu des données) et **formel** (portant sur la forme, notamment les métadonnées). La question des **données à caractère sensibles** entre également en ligne de compte.

Fondamentalement, les datasets déposés sur la plateforme SODHA doivent être issus de recherches menées dans des domaines et des disciplines scientifiques qui s'inscrivent dans le cadre thématique de l'archive de données SODHA.

Formellement, les métadonnées qui constituent la documentation des datasets doivent suivre quelques règles de base.

En ce qui concerne la sécurité, les données ne peuvent pas contenir d'informations à caractère sensible (typiquement, des données à caractère personnel).

2.1 Niveau fondamental (contenu)

SODHA est une **archive de données** pour les **sciences sociales** et les **humanités numériques**. Son but est de recevoir et de mettre à disposition des jeux de données liés à une ou plusieurs des disciplines qui constituent ces domaines de la science.

En vue de s'inscrire dans un consensus concernant la définition de ce que sont les sciences sociales et les humanités numériques, SODHA se fie à des classifications de référence telles que la base de données de projets de recherche de BELSPO, FEDRA, la Norme flamande des disciplines de recherche mise au point par ECOOM-Hasselt ainsi que le rapport sur les humanités numériques publié par le *Massachusetts Institute of Technology* (MIT).

Par ailleurs, SODHA aborde les datasets que des déposants soumettent pour publication sur la plateforme avec une perspective « neutre » et « agnostique » quant au contenu des datasets : le rôle de notre archive n'est pas d'apprécier la qualité des données présentées par les déposants, le volume des jeux de données déposés ou encore la pertinence des questions de recherche choisies par les enquêteurs.

Si, à l'issue d'un examen au regard des classifications qui viennent d'être mentionnées, un dataset soumis pour publication sur la plateforme n'est pas considéré comme pertinent par rapport à ces critères, SODHA peut conseiller des voies de publication alternatives aux déposants (dans la mesure de nos connaissances).

2.2 Données sensibles

À l'heure actuelle, SODHA n'est malheureusement pas en mesure d'accepter des **données** qui contiennent des informations sensibles, typiquement (quoique pas seulement) des données à caractère personnel qui permettent d'identifier des individus. Cependant, des données à caractère personnel qui ne permettent pas, d'elles-mêmes ou par recoupement, d'identifier des individus peuvent être acceptées. L'infrastructure SODHA est en cours de développement de sorte à permettre à terme le dépôt et la diffusion de données à caractère sensible en toute sécurité sur notre plateforme.

Toutefois, nous encourageons vivement les déposants à créer sur la plateforme Dataverse de SODHA des notices de **métadonnées** relatives à des datasets contenant des informations sensibles. En d'autres termes, si des données de recherche ne peuvent être publiées car certaines d'entre elles peuvent poser des risques, les chercheurs concernés ne devraient cependant pas hésiter à **mentionner l'existence** de ces données car de telles informations peuvent être très utiles pour d'autres chercheurs engagés dans les mêmes thématiques de recherche. Consultez notre FAQ pour plus d'informations.

2.3 Niveau formel (métadonnées)

Au moment de déposer des données sur la plateforme SODHA, il est demandé aux déposants de remplir un certain nombre de champs de métadonnées obligatoires.

La cohérence formelle (en d'autres termes, le formatage) des métadonnées augmente les chances que des utilisateurs trouvent effectivement les données publiées sur la plateforme lorsqu'ils y effec-

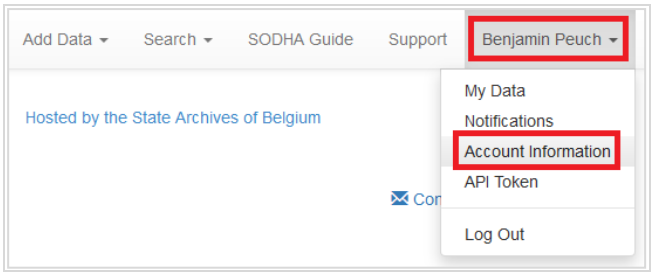
tuent des recherches. Pour utiliser un néologisme, cela augmente la « découvrabilité » des datasets. C'est pourquoi les règles suivantes sont d'application :

a. Les **noms** des individus mentionnés dans les champs Auteur, Personne de contact et Producteur doivent être écrits dans leur forme complète (notamment le prénom, qui ne peut pas être une simple initiale) et séparés par une virgule dans l'ordre qui suit : **[nom(s) de famille, prénom, éventuelle(s) initiale(s)]**

   Exemples
   Pirlot, Mélanie
   Van den Broeck, Jan-Albert
   Márquez Gutíerrez, Emelina Soraya
   de la Vega, Christian
   Von Humboldt Zimmerman, Dietrich
   du Val, Anne-Sophie Angélique
   Smith, Christian A.
   Pham, Van Duc
   Zhang, Ziyi

   On peut envisager des exceptions, par exemple dans le cas de jeux de datasets particulièrement anciens, pour lesquels la forme complète du nom d'un collaborateur ne peut être identifiée avec certitude. Dans de tels cas, veuillez s'il vous plait en informer les administrateurs de la plateforme via l'adresse sodha@arch.be.

   Le nom qui apparait dans le champ Déposant est automatiquement prélevé à partir des informations personnelles de l'utilisateur qui effectue le dépôt. C'est pourquoi nous demandons aux déposants de renseigner leur nom dans sa forme complète au moment de leur inscription ou en le modifiant par après si nécessaire (pour cela, sélectionnez le menu Informations personnelles) :



b. Afin d'inclure des **retours à la ligne** dans les champs de métadonnées (typiquement comme le champ Description), les déposants doivent avoir recours aux balises HTML **
**. L'encodage via la touche Entrée ne suffit malheureusement pas. La méthode pour y parvenir suit.

   Afin d'obtenir ceci :

> Ce dataset a été produit entre 2012 et 2014.
>
> Les données ont été collectées entre le 3 juillet 2012 et les 6 janvier 2013.

Des balises **
** doivent être ajoutées comme cela :

> Ce dataset a été produit entre 2012 et 2014.**

**
>
> Les données ont été collectées entre le 3 juillet 2012 et les 6 janvier 2013.

c. Les **paragraphes** doivent être séparés par une ligne vide, comme ci-dessus.

d. Les **titres** d'**ouvrages cités**, par exemple dans le champ Publication liée, peuvent être affichés en italiques à l'aide des balises ****. La méthode pour y parvenir est la suivante.

Afin d'obtenir ceci :

> D'hont, Ekaterina, *La Place du concept d'émotion en psychologie et dans les sciences sociales*, Brussels, Psykon Press, 2016.

Des balises **** doivent être ajoutées comme cela :

> D'hont, Ekaterina, ****La Place du concept d'émotion en psychologie et dans les sciences sociales****, Brussels, Psykon Press, 2016.

e. Lorsque **plus d'un élément** doit être encodé dans un **champ répétable** (comme le signale la présence du bouton +), le champ doit être alors répété autant de fois que nécessaire (plutôt que de surcharger une seule instance de champ avec plusieurs informations indistinctes) :

Bien distribuer l'information dans des instances de champs distinctes permet de garantir une bonne indexation des informations encodées par les déposants, ce qui augmentera la visibilité de vos données.

f. SODHA s'inscrit dans la mouvance des **données liées** (*linked data*) afin d'augmenter la visibilité des informations en ligne. C'est pourquoi, si les administrateurs SODHA découvrent des informations complémentaires relatives à vos données ou à d'autres publications que vous mentionnez, ils vous suggèreront de les ajouter à vos métadonnées.

2.4 Gestion des versions

Si un dataset a déjà été publié au moins une fois et que le déposant en crée une nouvelle version, les changements effectués doivent alors être documentés dans une nouvelle instance du champ *Description*. Pour ajouter une nouvelle itération de ce champ, cliquez sur le bouton « + » :

2.5 Fichiers exécutables ou binaires

Pour des raisons de sécurité, SODHA ne peut pas accepter les versements de datasets qui contiennent des fichiers exécutables.

2.6 Considérations générales

Lorsqu'un déposant soumet un dataset pour publication, un administrateur SODHA relira les métadonnées avant que l'ensemble ne soit publié.

Si certains passages des métadonnées contreviennent à ces règles, l'administrateur pourra effectuer par lui-même de **menues** corrections à cet égard. **En dehors de ces corrections, les administrateurs SODHA n'altèreront jamais les contenus soumis par les déposants de quelque façon que ce soit, et les déposants seront toujours prévenus de tout changement effectué dans leurs métadonnées par un administrateur SODHA.**

D'un autre côté, si certaines informations présentes dans les métadonnées s'avèrent ambigües (par exemple, à cause de fautes de frappe), les administrateurs SODHA prendront contact avec les déposants pour leur demander d'éclaircir ces éléments avant que le dataset ne soit publié.

3. Historique de version de ce document

Date	Numéro de version	Changements effectués
30-07-2021	1.0	Première version
05-08-2021	2.0	• Ajout du critère d'admission relatif à l'historique de version • Ajout du critère d'admission relatif aux fichiers exécutables
07-03-2022	3.0	Clarification du critère 2.2, portant sur les données sensibles, de sorte à distinguer les données à caractère personnel qui permettent d'identifier des individus des données à caractère personnel qui ne le permettent pas