

MOOC Wikidata : la qualité des données



WIKIMEDIA
FRANCE

[MOOC] Wikidata

- Une formation en ligne ouverte à tous et toutes
- Gratuite
- D'une durée de 4 à 6 semaines
- Qui apprend à devenir autonome sur Wikidata



WIKIMEDIA
FRANCE

La qualité des données quand...

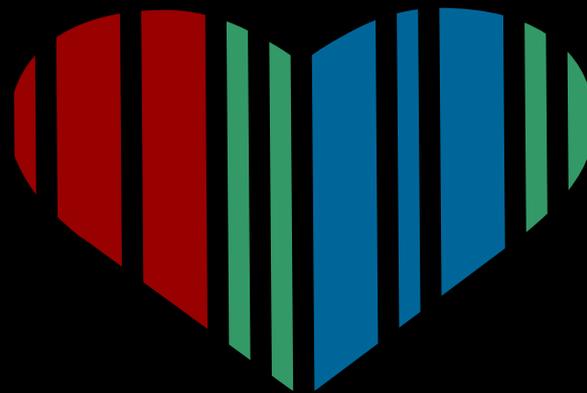
n'importe qui
peut enregistrer
n'importe quoi ?



**La plupart des
internautes
contribuent
pour améliorer
les données**



WIKIMEDIA
FRANCE



Il est possible de détecter de vandalisme grâce à des outils, créés et améliorés par la communauté

- Sarabadani, A., Halfaker, A., & Taraborelli, D. (2017, April). Building automated vandalism detection tools for Wikidata. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1647-1654).
- Heindorf, S., Scholten, Y., Engels, G., & Potthast, M. (2019, May). Debiasing vandalism detection models at wikidata. In The World Wide Web Conference (pp. 670-680).



WIKIMEDIA
FRANCE

ORES (Objective Revision Evaluation Service)

Filtres Mettre en valeur les résultats

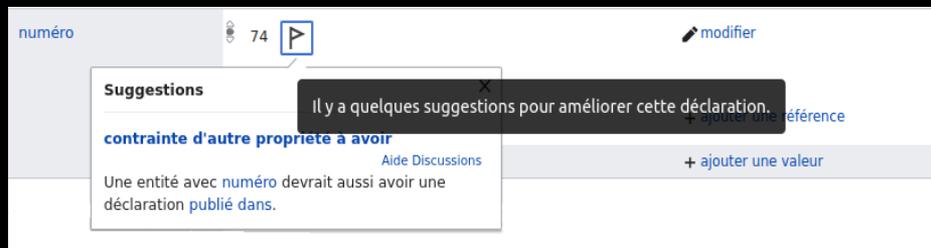
Prédictions de qualité de la contribution [Comment ça marche ?](#)

- Très vraisemblablement bonnes**
Filtre précis pour trouver presque toutes les modifications sans problèmes.
- Ont probablement des problèmes**
Avec une haute précision, trouve la plupart des modifications problématiques.
- Fortement susceptibles d'avoir des problèmes**
Très précis pour trouver les modifications incorrectes ou dommageables les plus évidentes.

Prédictions des intentions de l'utilisateur [Comment ça marche ?](#)

- Très vraisemblablement de bonne foi**
Très précis pour trouver presque toutes les modifications de bonne foi.
- Sans doute de mauvaise foi**
Avec une précision moyenne, trouve le plus d'éditions problématiques volontaires.

La qualité des éléments de Wikidata est variable (comme sur Wikipédia)



numéro 74  [modifier](#)

Suggestions

contrainte d'autre propriété à avoir [Aide Discussions](#)

Une entité avec **numéro** devrait aussi avoir une déclaration [publié dans](#).

Il y a quelques suggestions pour améliorer cette déclaration. [référence](#)

[+ ajouter une valeur](#)

Les triplets RDF (déclarations) peuvent contenir des références

nature de l'élément



forêt

✓ publier supprimer ✕ annuler ?

+ ajouter un qualificatif

▼ 1 référence

supprimer

affirmé dans

Quelles aires protégées pour l'Afrique de l'Ouest ?

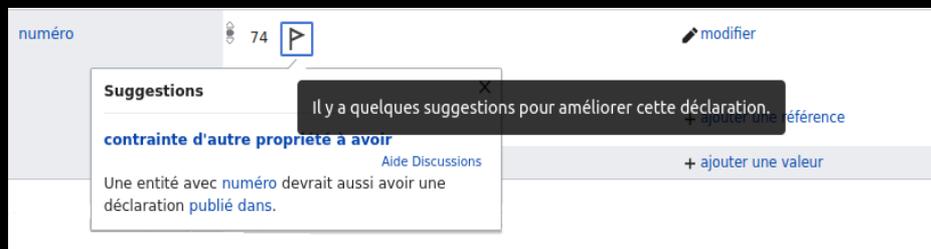
page(s)

345

+ ajouter

+ ajouter une référence

Wikidata fait des suggestions pour améliorer les données



numéro 74   [modifier](#)

Suggestions

contrainte d'autre propriété à avoir [Aide](#) [Discussions](#)

Une entité avec **numéro** devrait aussi avoir une déclaration [publié dans](#).

Il y a quelques suggestions pour améliorer cette déclaration. [référence](#)

[+ ajouter une valeur](#)

En résumé...

- Tous les enregistrements dans la base sont traçables (= pas d' "anonymat")
- Qualité variable des éléments
- Toujours possible de signaler des insertions non-fiables
- Existence d'un outil prédictif (ORES)
- Possibilité de demander à la communauté



