



Integração de identificadores Wikidata para o aumento de FAIRness dos datasets.

Joaquim Santos^a,

^aCentre for Functional Ecology, Department of Life Sciences, University of Coimbra, Portugal

Resumo

A criação de datasets pressupõe a utilização de identificadores que de forma unívoca identifiquem as entidades envolvidas: pessoas, instituições, locais, eventos, etc. Os sistemas de gestão de dados utilizam identificadores próprios a nível local para o seu funcionamento. Esses identificadores são muitas vezes irrelevantes para utilizadores humanos e para outros sistemas, uma vez que apenas garantem a lógica interna do sistema em que operam.

A partilha de dados deve seguir os princípios FAIR (Findable, Accessible, Interoperable and Reusable). Mesmo em datasets que sigam estes princípios, os dados podem requerer um nível de interpretação que está sujeito a erros ou atribuição incorreta, por exemplo, confundir duas pessoas com o mesmo nome. A utilização de um sistema de identificadores universais que possam ser partilhados nos datasets permite um maior nível de FAIRness, uma vez que garante uma interoperabilidade mais eficiente entre sistemas.

A Wikidata é um sistema de dados estruturados que permite a interligação de entidades entre si (Linked Open Data). É uma base de conhecimento gratuita e aberta que pode ser lida e editada por humanos e por máquinas.

A utilização de Wikidata Q Identifiers (QID) em sistemas de gestão de colecções de História natural está a tornar-se uma prática comum, que visa permitir uma melhor integração de dados em agregadores (GBIF), mas também a criação de novas plataformas que possibilitam a interpretação e descoberta de novo conhecimento. São exemplo disso a plataforma Bionomia, e mais recentemente o WikiProject sobre expedições de investigação científica.

Esta prática pode ser aplicada em outro tipo de sistemas ou datasets com vista à sua partilha de forma totalmente interoperável.

Palavras-chave: Wikidata, Identificadores Globais, FAIR, Linked Open Data, Interoperabilidade

Designação do projeto/infraestrutura/iniciativa

Herbário da Universidade de Coimbra

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto UIDB/04004/2020 com o identificador DOI 10.54499/UIDB/04004/2020 (<https://doi.org/10.54499/UIDB/04004/2020>)

Público-alvo

gestores de repositórios e data centres, gestores de ciência, bibliotecários, curadores de dados, especialistas de informática, investigadores

Ligações web úteis

- Leachman, S., Meudt, H., von Mering, S., & Santos, J. (2024, Julho 9). Online workshop Upskilling in Wikidata for maximum impact IBC 2024. Zenodo. <https://doi.org/10.5281/zenodo.12741075>
- Bionomia: <https://bionomia.net/>
- WikiProject Research expeditions: https://www.wikidata.org/wiki/Wikidata:WikiProject_Research_expeditions