

I principi FAIR nell'attività terminologica

Federica Vezzani
federica.vezzani@phd.unipd.it
Dipartimento di Studi Linguistici e Letterari (DiSLL)
Università di Padova - Italia

Nel quadro del programma *Horizon 2020*,¹ la Comunità Europea ha stabilito delle linee guida al fine di promuovere il libero accesso alle pubblicazioni scientifiche e ai dati prodotti da progetti di ricerca europei. L'*European Open Science Cloud* (EOSC)² è l'infrastruttura digitale che permetterà, a partire dal 2020, il raggiungimento di questi obiettivi tramite l'applicazione dei principi FAIR (*Findable, Accessible, Interoperable, Reusable*) [1] i quali descrivono un insieme di criteri per promuovere l'accessibilità, la trasparenza, la condivisione e il riutilizzo dei dati. In questo contesto, la corretta e adeguata rappresentazione di dati di tipo linguistico si configura come un compito necessario nel lavoro terminologico e, in particolare, nella realizzazione di risorse specializzate che raccolgono informazioni di tipo lessicale.

L'obiettivo della presente proposta è quello di rispondere alle esigenze descritte a livello europeo tramite la normalizzazione e standardizzazione della banca dati terminologica TriMED [2], nata per supportare la comunicazione in ambito medico e per facilitare la trasmissione di informazioni di tipo scientifico. TriMED raccoglie termini tecnici estratti da corpora medici e analizzati secondo un nuovo modello di scheda terminologica [3] che fornisce un elevato numero di dati linguistici i quali necessitano di essere presentati in modo accessibile, trasparente e riutilizzabile.

Per poter rendere TriMED una risorsa FAIR e, in particolare, per massimizzarne l'interoperabilità con altre risorse linguistiche, presentiamo la ristrutturazione delle schede terminologiche secondo le direttive dello standard internazionale ISO 30042:2008³ che definisce il formato TermBase eXchange (TBX). In questo modo, oltre alla realizzazione di schede terminologiche normalizzate, sarà possibile il collegamento diretto con 1) *Consumer Health Vocabulary*⁴ per supportare la comunicazione medico-paziente; 2) SNOMED CT⁵ per la codifica dei contenuti medici; 3) MeSH Terms⁶ per l'indicizzazione delle riviste biomediche di PubMed⁷; 4) IATE⁸ per supportare i traduttori nel processo di traduzione di testi scientifici. Per concludere, intendiamo mostrare lo stato di avanzamento del processo di standardizzazione discutendone i punti critici e presentandone i vantaggi dal punto di vista della ricerca terminologica.

Riferimenti bibliografici:

[1] Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016).

[2] Vezzani F., Di Nunzio G.M., and Henrot G. "Trimed: a multilingual terminological database." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*. 2018.

[3] Vezzani F., "Aménagement de la terminologie spontanée: un cas de collocation". *XIVe Journée Scientifique REALITER – Réseau Panlatin de Terminologie*. 2018.

¹ <https://ec.europa.eu/programmes/horizon2020/what-horizon-2020>

² <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

³ <https://www.iso.org/standard/45797.html>

⁴ <http://consumerhealthvocab.chpc.utah.edu/CHVwiki/>

⁵ <http://www.snomed.org>

⁶ <https://meshb.nlm.nih.gov/search>

⁷ <https://www.ncbi.nlm.nih.gov/pubmed/>

⁸ <https://iate.europa.eu/home>