

Assessment of FAIRness of open data sources in life sciences

Filip Pattyn¹^[0000-0003-0858-6651], Bérénice Wulbrecht¹^[0000-0002-9444-1709], Kenny Knecht¹^[0000-0002-1049-3684] and Hans Constandt¹^[0000-0003-0858-6651]

¹ ONTOFORCE NV, Ottergemsesteenweg-Zuid 808, 9000 Gent, Belgium
filip.pattyn@ontoforce.com

Abstract. The life sciences domain heavily relies on the availability and usability of open data. Some of these sources are very extensive and are considered as key or golden sources for one type of information in the field. Despite ongoing data structuring, data standardization and data linking efforts, many of these data sources aren't by default easily usable for integrated analyses in automated workflows. The FAIR (Findable Accessible Interoperable and Reusable) data principles can be used as a guideline to overcome this. There's currently no repository storing information about how FAIR a data source currently is. An assessment of the FAIRness of the most used life sciences related open data sources could delineate what the status is of the usability of open data in the field. We analyzed the FAIRness of the largest and most used data sources available in the linked data search and navigation platform DISCOVER and we calculated what the effort is to convert these sources to make them FAIR enough to integrate them in the semantic web based DISCOVER platform. The results show that it requires still a lot of work to upgrade these popular data sources to a higher level of FAIRness.

Keywords: FAIR data, Linked Data, Data Quality.