

Abstract

For data to be findable, integrable and reusable, it first needs to be normalized (so that data from different sources can be aligned) and, most importantly, it needs to be cleaned up, so it is free from original human and machine errors.

For both tasks, it is a standard practice to align data to well established standard ontologies and controlled vocabularies and to curate it, both manually and digitally. While there is no automated solution that can guarantee clean and well-aligned data, an efficient semi-automated solution can do the preliminary work, thus leaving curators with fewer, more complex cases.

Furthermore, resulting data dictionaries often need to be classified and tested for heterogeneity, so that they can be used in more structured, domain-specific, targeted, and harmonized fashion.

The annotation and mapping services can be used for streamlining data that comes from public resources and is often presented in a variety of formats and flavors. The goal of this project is to provide one-stop-shop for data collection, harmonization, alignment, and mapping.