

Improving data discoverability in open repositories with folksonomy to taxonomy integration

Alena Begler

Saint Petersburg State University (Russia), Orcid 0000-0003-4375-1106

Today we have several standards designed to improve discoverability of the research data. These standards provide general meta-information, but they are not always useful to find relevant data for re-use. For example, it is not easy to find raw data from the specific task in any of the popular open repositories. A reason for this may be the lack of information navigation tools. Today for navigation in open repositories we have two main tools: 1) search field, sometimes with several standardized metadata fields; 2) tags, added by users (i.e. folksonomies). However, combined folksonomy-taxonomy systems work better than each of them separately (Kui & Tsui, 2011). Especially useful multi-faceted taxonomies (Cheung, Lee, & Wang, 2005). Therefore, repositories could be better suited for the information search with the integration of existing folksonomy of user tags into a controlled multi-faceted taxonomy. Facets of such a taxonomy may include both general and domain-specific facets. Examples of general facets: data type, data file extension, field of study, etc. Domain specific (for behavioral sciences): task characteristics (paradigm, specific task); sample characteristics (disorders, age, species); experiment type (behavioral, neurophysiological, computational). Each of these facets will consist of user tags (i.e. tag “flanker task” will be under “specific task” facet). Integration of existing folksonomy of user tags into such a taxonomy will give an opportunity to navigate through thousands of datasets and find relevant. Moreover, UI-integrated controlled taxonomy will improve the users work with tags, as maintainers of data will be aware of the meta-information about data, which is useful for their peers.