

Open Data on the Web

Problem

Beyond the obvious issues of openly licensing of data, many of the significant barriers to the constructive use/reuse of data are technical in nature. Both in understanding, of obscure focused vocabularies used to describe resources & concepts, and access methods & formats.

Proposition

As a consumer of data you often need to mix data from more than one source to add value to your processes and enriching the experience of those using your output. Identifying enrichment resources delivered by services can often simplify the required effort for the interlinking of data without significant infrastructure and data maintenance procedures. Few would embark on building a mapping interface, follow the same model by consuming live data from such as: dbpedia, VIAF (for creative people & organisations), BBC (animals), WorldCat.org (bibliographic description).

Such intermixing of data can be said to be lubricated by the use of generic well understood vocabularies. A prime example of this being Schema.org. Often underestimated as just a way to markup web pages for Rich-Snippet display in Google results, Schema.org is becoming a general purpose high-level vocabulary capable of providing structured description across many sectors. Backing by the major search engines brings stability, reach and a large community that understand it. It brings double benefit developers in easily consuming data from disparate sources whilst providing and obvious choice in the way to publish and make visible their own efforts.

Person

Richard, as Chair of the Schema Bib Extend W3C Community Group is closely involved with initiatives to extend Schema.org to improve its capability in bibliographic data description, and as Technology Evangelist at OCLC works with the teams engaged with sharing the millions of bibliographic resources from their membership.