Linked Data for Connecting Traditional Chinese Medicine and Western Medicine [6]

Jun Zhao[1], Anja Jentzsch[2], Matthias Samwald[3,4], and Kei-Hoi Cheung[5]
1. Department of Zoology, University of Oxford, Oxford, UK
2. Web-based Systems Group, Freie Universität Berlin, Berlin, Germany
3. Digital Enterprise Research Institute, National University of Ireland Galway, Galway,Ireland
4. Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria
5. Center for Medical Informatics, Yale University School of Medicine, New Haven, Connecticut, USA

Corresponding author: Jun Zhao jun.zhao@zoo.ox.ac.uk

Traditional Chinese Medicine (TCM), which is a type of alternative medicine, is receiving growing attention from patients and biomedical researchers in the western world. In spite of this growing attention, TCM has not been included as part of standard care in many western countries mainly due to a lack of scientific evidence for its efficacy and safety. In addition, many of the documentations about TCM are not available in English, creating a language barrier to patients, scientists, and physicians in the West. TCMGeneDIT (http://tcm.lifescience.ntu.edu.tw/) is a unique database that captures in English a rich body of knowledge about traditional Chinese herbs, including scientific evidence reported in scientific publications about the association of these herbs with diseases and genes. We re-formatted the TCMGeneDIT database in the RDF format (as Linked Open Data), making it programmatically accessible through a flexible query language (SPARQL) and a flexible Web service (SPARQL endpoint). This work represents collaboration between the BioRDF task force and the LODD (Linked Open Drug Data) task force of the Semantic Web for Health Care and Life Sciences Interest Group chartered by the World Wide Web Consortium (W3C).

We demonstrate how to use a novel data link creation tool, Silk (http://www4.wiwiss.fuberlin.de/bizer/silk/), to link the RDF-formatted TCMGeneDIT with other linked RDF datasets. These datasets cover clinical trials, drugs, and diseases, with Dbpedia (http://dbpedia.org/) acting as a central hub of linked data. We also present a tool, the aTag generator (http://hcls.deri.org/atag/generator/), which allows users to create RDF statements out of snippets of free text in an intuitive way. We report our experiences in creating data links in a large scale and our evaluation of the precision of the data linking. We demonstrate how this Web of knowledge can be used to allow patients and researchers to look for herbs/drugs for treating a given disease. The Linked Open Data approaches used in this work provide a standard and lightweight framework for publishing, sharing and linking data sets on the Web. The flexible SPARQL query language and protocol makes it possible to build a data integration application at a low cost. While the de-referenceable URIs of entities in the Linked Data ease navigation of this rich body of knowledge, we need to further investigate how best to improve the user interface for presenting this intricate network of knowledge to domain scientists.