Linking external SQL databases and the Semantic Web: Pipelines for dynamic web publication with stable Uniform Resource Identifiers for database structural information and content schemes

Dagmar Triebel 1, Anton Link 1, Gregor Hagedorn 1, Andreas Plank 1, Markus Weiss 1, David Fichtmueller 1, Tanja Weibulat 1, Gerhard Rambold 4

Advanced mechanisms are required to dynamically link the world of stand-alone SQL databases with Wiki engines to generate Semantic Web representations with stable URIs. SQL database structural elements (ER models) as well as the contents of databases can be published in this way and thereby contribute to the Linked Open Data network. For realising this connection, data pipelines have been implemented, being initiated by export functions of the database clients of the Diversity Workbench framework (www.diversityworkbench.net).

A second pipeline is established for the publication of novel content schemes as required, e.g., for the research project 'MOD-CO: Towards an integrative and comprehensive standard for meta-omics data of collection objects' (http://www.mod-co.net, see poster number 1022). The terms, concepts, descriptors and descriptor trees are managed in DiversityDescriptions. This relational database has a generalized triple-structured design allowing the flexible organization of descriptor states, (hierarchical) dependencies, and interrelations between descriptors. A web publication of the database content is created dynamically, which allows for an appropriate citation of every individual MOD-CO schema element with a stable URI. Community-driven annotation might be organised separately in the same wiki.

One dataflow pipeline allows for the automated generation of Semantic MediaWiki (SMW) pages with stable Uniform Resource Identifiers (URIs) for publishing the database structural elements of entity-relation models of the Diversity Workbench (DWB). This allows for citing and persistently referencing each of the more than 2,000 DWB elements, as so-called concepts, and each of the 250 tables, as so-called concept collections. Furthermore, this will facilitate mapping efforts between DWB elements and, e.g., concepts of esta-blished TDWG data exchange standards.

The first two types of pipelines are building on work by Hagedorn, Endresen, O Tuama, Plank 2013, establishing the TDWG Terms Wiki (http://terms.tdwg.org) on the biowikifarm in the ViBRANT (Virtual Biodiversity Research and Access Network for Taxonomy) project. The pipelines use the same, unmodified SMW term templates for the SKOS (Simple Knowledge Organization System) compatible definition of concepts, classes and concept collections and thereby they are compatible with other schemes in the TDWG Terms Wiki. The work on DWB data pipelines is supported by the German Research Foundation (DFG) as part of the project GFBio (www.gfbio.org).

A third pipeline has been established for generating a single MediaWiki page from SQL database export for providing a citable publication for every version of a SQL database entity-relation model with all its information, tables and definitions.