LSID versus HTTP URI: Two approaches and e-infrastructures for managing information about taxon names

Nina Laurenne[1,2], Jouni Tuominen[1,2], Arto Mertaniemi[1], Hannu Saarenmaa[3], Eero Hyvonen [1,2]

[1] Aalto University, School of Science, Department of Media Technology, Semantic Computing Research Group
[2] University of Helsinki, Department of Computer Science
[3] University of Eastern Finland, Digitarium
nina.laurenne@helsinki.fi, jouni.tuominen@aalto.fi, arto.mertaniemi@helsinki.fi, hannu.saarenmaa@uef.fi, eero.hyvonen@aalto.fi

The amount of biological information has increased during the last decades. The information is hidden in museum specimens, data bases of observations, and in literature. Integrating data from scattered sources is hard because different vocabularies are used. The biggest barrier for data integration is the changing nature of scientific names, which hinders the interoperability of information systems. A solution is to use machine-processable identifiers for identifying biological names.

Two approaches are presented for managing taxon names. The first one is a taxonomic database of the Finnish Museum of Natural History based on Life Scinece Identifiers (LSIDs). The scientific names of six butterfly checklists are cross-mapped and linked taxon names form a concept to which an LSID is given. The concept covers a currently valid name, synonyms, and their lexical variants, and references to original publications and the year of publication if available. A tool for mapping taxon names between checklists is provided.

The second approach is based on HTTP URIs and the taxonomic metaontology TaxMeOn is presented for depicting the information of the butterfly species lists. The meta-ontology is based on RDF/OWL and the key classes are: a scientific name, a taxonomic concept, a name status, a taxonomic rank, a reference to a publication, an author and a common name. The same relations are applied for mapping taxon names as using LSIDs. TaxMeOn provides functionalities for humans and machines for accessing the ontologies that are published in the ONKI Ontology Service. ONKI supports content indexing, concept disambiguation, searching, and query expansion.

Cross-linking taxon names between species lists helps a users to piece together the changes of scientific names and estimate the approximate amount of taxonomic treatments (none vs. many). Linking taxon names at the species level is straightforward, but at higher levels the problem is how to reconcile differing classifications of checklists.

The choice of an identifier used depends on needs, but despite the chosen identifier, the problem remains the same i.e. how to describe taxonomic information consistently without losing practicality. There is no significant difference whether LSIDs or HTTP URIs are used use to identify scientific names of checklists. However, the latter is more flexible as it allows interlinking the data with other Linked Data datasets increasing their interoperability.