

## **Toward an unified taxon ID system for biodiversity LOD and integration**

Guan-Shuo Mai<sup>1\*</sup>, Hsiang-Ying Li<sup>1</sup>, Chau-Chin Lin<sup>2</sup>, Kwang-Tsao Shao<sup>1</sup>

<sup>1</sup>Biodiversity Research Center, Academia Sinica

<sup>2</sup>Taiwan Forestry Research Institute

E-mail: trashmai@gmail.com

Scientific names are used as identifiers for taxa and are the keys for biodiversity data integration, but they could be less effective due to different representation forms. The misspelled names, synonyms, homonyms of taxa and different classification systems are confusing and further impede integration process. Here we propose a hash-based approach to generate taxon ID with parameters of a taxon's name and its hierarchical information, and so let everyone can generate taxon IDs locally and independently meanwhile they can be used globally as URI of LOD when combined with namespaces. They can be easily mapped and interlinked as a result of hashing function. With appropriately designed ontology describing taxa, scientific names, their canonical forms and relationship among them, we can sort out different problems emerged from biodiversity data integration by querying multiple biodiversity datasets altogether in different levels of semantic. Different perspectives of classification can be kept in the process instead of being discarded. We published LOD of Taiwan Catalogue of Life database and endangered species red lists from several Asian countries as a case to demonstrate how our concept works.