A proposal for a RDA Interest Group

Biodiversity Data Integration IG

Abstract

Biodiversity is infinitely complex and must be generalised, approximated and abstracted in order to be represented and understood. However, this should not contradict the epistemological richness of biodiversity science as a discipline or deprive biodiversity scientists the flexibility to express their knowledge regardless if their precise terminology has been standardised. This interest group would like to increase the effectiveness of biodiversity e-Infrastructures by promoting the adoption of common tools and services establishing data interoperability within the biodiversity domain, enabling the convergence on shared terminology and routines for assembling and integrating biodiversity data.

In this context taxon names provide key ontologies on connecting biodiversity data. The absence of an appropriate global name-reference system hinders an efficient and dynamic cross-referencing of taxon names, the functional re-use of biodiversity information and a single access to 'all names in use'. It also hinders the further development of a communal (virtual) research environment, supporting science as a community effort. As an example, the Global Names Architecture (GNA) would like to create more stability, compatibility and community in names resolution, by including an objective (nomenclatural) layer, stabilising the cross-referencing of taxon names and by including advanced indexing services for name discovery, recognition and re-use, which will also optimise the use and uptake of all associated biodiversity information. GNA will optimise the discovery and integration of biodiversity data by developing and improving authority files and names backbones (like CoL and PESI), checklist mapping routines and repositories (like the GBIF Checklist Bank), harmonise web portal APIs, build workflows to secure the proper resolution and data cleaning for e-Science application, and support the on-going virtualisation of the research domain. Gradually the names architecture will advance towards a global names references system, including necessary components like the Global Names Usage Bank (GNUB) and Global Names Index (GNI).

Other recognised essential foundations for data interoperability includes systems for developing, managing and sharing controlled vocabularies and ontology’s, like the GBIF Knowledge Organization System (KOS), and infrastructure models allowing an integrated access to vernacular names, also serving the non-scientific community. The promotion and adoption of common standards tools and best practices amongst biodiversity researchers as well supports the further internationalisation of existing infrastructures. Therefore an initial focus of this IG includes the drafting of an agenda contributing to the defragmentation of species (data) management, reducing redundancy on species data tools and resources.

Members of this Interest Group are supposed to collaborate with relevant other RDA Working and Interest Groups on data citation, meta-data, persistent identifiers, big data analysis, data publishing, and so on.
Announced short term subgroups and potential Working Groups includes:
- Global Names Architecture
- Defragmentation of species data-management
- Vernacular names infrastructure

The general objective is to make this Interest Group a sustainable component of major European biodiversity informatics initiatives, like LifeWatch and EU BON. Preparatory activities are supported by EC FP7 projects, like PESI, VIBRANT, pro-iBiosphere and BioVel and global organisations like TDWG and GBIF.

Contacts and chairs
Nicola Nicolson <n.nicolson@rbgkew.org.uk> – Global Names Europe leader
Rich Pyle <deepreef@bishopmuseum.org> – Global Names US leader
Paul Kirk <P.Kirk@kew.org> – Global Names Architecture chair
Heimo Rainer <heimo.rainer@univie.ac.at> – Vernaculars names infrastructure leader
Nicolas Bailly <n.bailly@cgiar.org> – Defragmentation of species d-m leader
Yde de Jong <mail@yjong.net> – PESI coordinator