

RDA-CODATA INTEREST GROUP ON LEGAL INTEROPERABILITY OF RESEARCH DATA

Case study: Plazi Treatment Repository

Willi Egloff & Donat Agosti, Plazi, Switzerland

Background

A large part of our knowledge on the world's species is recorded in the corpus of biodiversity literature with well over hundred million pages, including millions of treatments describing the species and higher taxa. Additionally it is represented in natural history collections by an estimated 2-3 billion labeled specimens. This body of knowledge exists in great part in paper-print form and is not directly accessible through the Internet. At the same time, a complementary body of knowledge based on DNA analyses is rapidly developing. In order to allow access to and re-use of all the taxonomic information contained in this literature to all persons who have a requirement for biodiversity data, it must be digitized, linked together and organized in an open biodiversity knowledge management system (OBKMS). Progress towards this goal is hampered by numerous factual, technical, economic, sociological, but also legal factors.

Discussion

Plazi was founded as an association with the primary goal of transforming both the printed and the digital taxonomic literature into semantically enabled enhanced documents. Activity up to now includes

- the creation of test bodies of literature
- XML schemas modelling its logic content (TaxonX and TaxPub, a taxonomy domain specific National Library of Medicine Journal Archiving Tag Suit version)
- the development of a mark-up editor and web services (GoldenGATE) allowing tagging and the enhancement of documents with links to external resources, such as specimens, images, or reference vocabularies
- a repository for publications and issuance of bibliographic identifiers
- a dedicated server to serve the marked-up content
- semantic tools to mine information

- development of stable identifiers for treatments allowing citation of treatments similar to bibliographic references.

Plazi puts the taxonomic information (species names, taxonomic treatments, a.s.o.) and the metadata of the publications in the public domain, while respecting existing copyright restrictions for the access to the original documents. The actual workflow is based in Switzerland and is organized according to Swiss copyright law. Although this legislation provides for a rather wide range of exceptions and limitations to copyright and can be qualified as fairly user-friendly, the repository is far from being openly and freely accessible to all persons who have requirements for biodiversity data. The goal of an open biodiversity knowledge management system cannot be reached on the mere basis of the actual Swiss copyright law.

Case study

Plazi has worked up to now with several test bodies of taxonomic literature covering ants and fish, and more recently plants and animals in general. The main task was to develop procedures and strategies that would allow the computerised extraction of taxonomic information, especially names and taxonomic treatments, as well as bibliographic references from printed publications and digital sources. The extraction process requires repeated reproductions of the original texts and is governed by art. 19 Swiss Copyright Code, which provides for a far reaching exception for the use of copyright protected works for the purpose of internal documentation within a private or public institution. In order to fulfil the criterion of “internal documentation”, access to the original documents must be restricted to Plazi members. The result of the extraction process consists of data and standardised information which do not qualify as works in the meaning of copyright. This body of information is stored in the Plazi repository which is openly and freely accessible [1].

The borderline between the copyright protected document and non protected content of this document is not clear yet. It will be part of the case study to define supplementary criteria for this distinction.

Plazi's next step will be the extension of the treatment repository up to 1 million treatments. Furthermore, we discuss the inclusion of visual and audio data. This enlargement again can only be realized if it is compatible with existing copyright legislation. For that reason, Plazi undertakes an extensive assessment of European copyright legislation, in order to detect further legal barriers to the building up of a open repository for taxonomic information and to the development of an open biodiversity knowledge management system as well as possible solutions to remove these barriers. This work is executed in straight relation to two ongoing EU funded projects (Pro-iBiosphere and EU-BON) where Plazi is responsible for developing a common data sharing policy [2].

The study will analyse the actual Plazi workflow and compare different national copyright legislations. A first study referring to 8 European countries is already completed and will be published in one of the next editions of “Zookeys” [3, 4]. An evaluation of the virtual application of these copyright legislations to the Plazi approach as well as protection on the international level will be added. Special attention will be given to regulations which favour

collective solutions as the so called “extended collective licenses” in the Scandinavian countries or the national implementations of the “copyright exceptions for research purposes” provided for in the EU-Directive 2001/29.

Close attention will also be paid to taxonomic names and their classifications, since names are indispensable to link all the legacy biodiversity knowledge with other relevant ongoing efforts. We will further discuss opportunities for cross-disciplinary application of our findings, as copyright barriers have similar effects to other scientific domains.

As a result, the study should allow to implement legal interoperability approaches into the structure of the Plazi repository. The solutions developed within this project may serve as source of inspiration to similar solutions in other scientific domains.

References

- [1] Agosti, D, Egloff W 2009. Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2009, 2:53 (<http://www.biomedcentral.com/1756-0500/2/53/abstract>)
- [2] Patterson DJ, Egloff W, Agosti D, Eades D, Franz N, Hagedorn G, Rees J, Remsen DP 2014. Scientific names of organisms: attribution, rights, and licensing. BMC Research Notes 2014, 7:79 doi:10.1186/1756-0500-7-79 (<http://www.biomedcentral.com/1756-0500/7/79>)
- [3] Egloff, W, Patterson, DJ, Agosti, D. Taxonomic literature and knowledge in European legislature. Zookeys (in preparation)
- [4] Pro-iBiosphere. 2013. D2.4.1 Draft policy on Open Access for data and information [PDF](#)

Contacts

Willi Egloff, Plazi: willi.egloff@advocomplex.ch

Donat Agosti, Plazi: agosti@plazi.org