FutureTDM
Explore . Analyse . Improve

ORGANISATIONS

IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

Facts	
Project No:	665940
Program:	H2020 | CSA | GARRI
Duration:	09/2015 - 08/2017

The Plazi Approach

PLAZI
TAKING CARE OF FREEDOM

Plazi is a non-profit organization based in Bern, Switzerland. It aims to support and promote the development of persistent and openly accessible digital taxonomic literature, in other words, Plazi aims to improve the science of charting the world's biological diversity. [1]

Plazi maintains a digital taxonomic literature repository and participates in the development of new models for publishing taxonomic treatments in order to maximize interoperability with other relevant components (name servers, biodiversity resources, etc...). They advocate and educate the community on maintaining free and open access to scientific discourse and data, which they consider to be of vital importance.

Plazi: *We are data brokers. Our clients are anyone who wants to get the data. We take unstructured data and make it structured and accessible.*

We are interested in data in articles, not articles themselves. We take the data out of publications which we do not consider to be infringing copyright. The essential goal is that we make all literature accessible.

A huge amount of data sits on people's desks as copies of articles, which are not online and therefore not citable. TDM is used more in the field of biomedical science because they have well-structured data and tools to mine and ontologies.

Plazi: *The problem of TDM is that it does not follow the way science works. Our (biodiversity) literature is not made for it. It is almost impossible to get a machine to read it.*

TDM is used to abstract the data out of publications and make it available for research. Besides the many technical challenges of sharing data at global scale, Plazi has encountered various legal issues of data sharing.[2]

Plazi: *In the beginning, we had our entire system taken down... because we could not show rights clearance for each single work.*

CC0

This project has received funding from the European Union's Horizon 2020 (H2020) Research and Innovation Programme.

BIODIVERSITY DATA

250 years of scientific publications

500.000.000+ Printed Pages

1.900.000 Species 2-3 billion Specimens

17.000 new species per year

GOAL is to create a giant Phone book of the species of the world

TARGET is to link 1 Million taxonomic name usages

All the data used by Plazi is published data. Publications going back to 1753, the beginning of taxonomy. Anything published after that which is scientific and follows the code is part of the system. This is possible in Switzerland that allows a temporary copy for articles[3].

Plazi: *Our problem is not copyright; our problem is attribution. Scientists want to make sure they get attributed.*

The barriers Plazi faces are technical: not having a general legal regulation and working under the Swiss exception but not being able to rely on the same exception in the EU which limits the opportunity to extract data the same way as in Switzerland.

In addition, taxonomic literature also has a rich tradition of illustrating scientific articles with images and diagrams of species. Illustrations use standardized views of characteristics that have been discovered. This helps other taxonomists to compare, separate and diagnose species. Because of this standardisation, Plazi argues that these images should not fall under copyright and thus be made accessible through initiatives such as the Biodiversity Literature Repository in which over 105,000 images are openly accessible.[4]

Plazi: *Our conclusion is that most images in taxonomic literature do not qualify as copyrightable work in a legal sense. They are not novel forms of expression as would be required to be considered a copyrightable work. Rather, images are crafted to comply with conventions in the taxonomic domain. Thus, these images belong, like data, in the public domain.*[5]

TDM DRIVERS AND BARRIERS

- *Essential Biodiversity Variables should be shared as open data, making them available without charge or restrictions on reuse.*
- *Exceptions should be limited to sensitive data: (1) Data whose free accessibility could endanger certain aspects of biodiversity conservation; (2)Data that are qualified as confidential by the competent authority.*
- *Right holder(s) of research data (if any) should dedicate them to the public domain (by CC0-waiver, CC-BY-License or any similar instrument).*
- *Data, products and metadata should be made available with minimum time delay.*

[1] http://plazi.org/about/about-plazi/

[2] Agosti D, Egloff W (2009). "Taxonomic information exchange and copyright: the Plazi approach" (PDF). BMC Research Notes 2:53: 53.

[3] Willi Egloff & Donat Agosti Plazi, Bern (http://plazi.org) Globis-B Workshop Leipzig, 29.2./2.3.2016 Data Sharing Principles and Legal Interoperability.

[4] Willi Egloff & Donat Agosti Plazi, Bern (http://plazi.org) Globis-B Workshop Leipzig, 29.2./2.3.2016 Data Sharing Principles and Legal Interoperability.

[5] Egloff W, Agosti D, Kishor P, Patterson D, Miller J (2017) Copyright and the Use of Images as Biodiversity Data. Research Ideas and Outcomes 3: e12502.