

REVISTA BRASILEIRA DE
Entomologia
A Journal on Insect Diversity and Evolutionwww.rbentomologia.com

Review

Species descriptions and digital environments: alternatives for accessibility of morphological data

1 Thiago Sanches Ranzani da Silva

2 Universidade Federal do Paraná, Departamento de Zoologia Avenida Coronel Francisco, Laboratório de Sistemática e Biologia de Formigas, Curitiba, PR, Brazil

ARTICLE INFO

Article history:

Received 1 March 2017

Accepted 30 June 2017

Available online xxx

Associate Editor: Livia Pinheiro

Keywords:

Data accessibility

Interoperability

Annotations

Web-based taxonomy

ABSTRACT

Taxonomists' efforts throughout history provide significant amount of data that give support for establishing the specific identity of several groups of biological systems. In addition to identifying species, taxonomic research offers a wide range of biological information that can be used in other disciplines, e.g. evolution, ecology, integrated pest management. However, most of this information remains unappreciated due to certain aspects: (1) the advent of analytical tools have led to a shift in interest and investment in researches, focusing mainly in molecular studies; (2) the erroneous concept that the extensive data offered by taxonomic studies can be replaced by other datasets, separating it from its hypothesis-driven and investigative nature; (3) the final products found in taxonomic works are commonly restricted to a small group of researchers, due to its low accessibility and specific language. Considering this last aspect, web-based tools can be valuable to simplify the dissemination of the taxonomic product. Semantic annotation provide a condition in which species descriptions can be readily available and be far more extensive, enabling rapid exchange of countless data related to biological systems.

© 2017 Sociedade Brasileira de Entomologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Taxonomy is a fundamental science that provides support to all major fields in biology. Taxonomists produce much of the basic knowledge about species, providing crucial information to other disciplines that deal with any sorts of biological systems. However most taxonomic data remains unappreciated due to the fragmented condition and uneven accessibility of basic information on biodiversity. According to Miller et al. (2012), taxonomy allows species recognition through textual descriptions and images of specimens, and maps their distribution based on locality records, meaning that most of the vital knowledge used to preserve a particular species results from taxonomic works. With accelerated rates of extinction in many recent groups, summed to a loss of habitat in many regions of the world and the inexorable effects of climate change, taxonomy has never been more important (Isaac et al., 2004; Miller et al., 2012; Padial et al., 2010).

Biodiversity is one of the richest information provider fields for human knowledge (Peterson et al., 2010). However, only recent advances in cybertaxonomic infrastructure (i.e. the combination of digital and information technologies and informatics with

taxonomy) made possible the creation of tools that unite biodiversity information in a way that they are available to a wide variety of users, including ecologists, wildlife and land managers, likewise making the existing information promptly accessible to the global taxonomic community (Godfray, 2002; Miller et al., 2012). Although there are plenty of websites devoted to specific taxa and projects, which are useful to those interested in questions inside that domain, the true potential of online tools lies in mechanisms that combine data from primary sources in a way that users can easily filter and recombine, making their use flexible toward any purpose (Peterson et al., 2010; Thessen and Patterson, 2011).

Taxonomic works are, generally, a summary of a set of elements, including descriptions in textual form, scientific names and nomenclature acts, literature references, images, records of species' occurrences and, recently, DNA sequences (Miller et al., 2012). The role of a taxonomist is to establish connections between specimens and nomenclature, express interpretations and hypothesis through textual elements and document observations using images. In a traditional taxonomic publication, all of those elements are combined in a single document. However, if they were to be released simultaneously and maintained as data sets linked to the original publication in an external environment, there would be direct effects on the taxonomic community, as well on other researcher groups (Deans et al., 2012; Godfray, 2002; Miller et al., 2012; Padial et al., 2010). In order to link data from various fields of knowledge,

E-mail: tsranzanidasilva@gmail.com<http://dx.doi.org/10.1016/j.rbe.2017.06.005>0085-5626/© 2017 Sociedade Brasileira de Entomologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

disseminating phenotypic information in an open environment, the use of tools that enhance interoperability are vital. In this context, semantic annotations are the main choice to facilitate dissemination and retrieving of phenotypic data in a digital environment.

Semantic annotation is, basically, the process of attaching additional information to a concept, in a way that it is easily read and recovered by machines. It links multiple concepts across different domains, inside a digital platform. Hence, it enables the recovering of multiple knowledge, allowing the inclusion of particular concepts into broader perspectives. It can be used in ontology-based information retrieval queries for efficient data mining, spreading the current understanding of a specific concept to multiple domains. Therefore, its use may provide ways to extend the access to unappreciated data, such as those available in taxonomic works. Several studies applied semantic annotation to extend the application of phenotypic information to other domains of knowledge (e.g. [Ćwiek-Kupczyńska et al., 2016](#); [Hoehndorf et al., 2016](#); [Serrano-Solano et al., 2017](#)), suggesting that this process greatly enhances interoperability among systems that support this type of data.

Taking the aforementioned considerations into account, the intention of the present work is to discuss the importance of the taxonomic investigation, focusing on the means for retrieval of the taxonomic product, especially in the role played by semantic annotation in species descriptions. Discussions are based in a non-systematic literature review. Initially, I discuss the importance of the taxonomic work and the effectiveness of the taxonomist workflow to produce descriptions of biological systems. After that, I address the question of the accessibility of taxonomic data, focusing on the manner in which this type of information is presented, especially when data mining is concerned. In this sense, I discuss the construction of anatomical concepts and some conceptual disagreements concerning the application of homology assumptions in the elaboration of anatomic ontologies. Following this, I address the subject on how digital environments can be valuable to the dissemination of taxonomic products, emphasizing the role of semantic annotation in the extension of this kind of information.

Accessibility to the taxonomic information

Taxonomists provide much of the basic knowledge about biological systems, and are thus considered as information providers to the life sciences ([Wilson, 2002](#); [Gewin, 2002](#)). They provide essential information to innumerable studies, making it possible to grasp the real dimension of biodiversity, as well as the causes and consequences of their diversification. It is such an integrative evidence that is a requisite to other disciplines in biology, such as conservation ([Mooers, 2007](#); [Forest et al., 2007](#)), biomimetics ([Badarnah and Fernández, 2015](#); [Rajabi et al., 2015](#)), and evolution of body plans (evo-devo) ([Mallarino and Abzhanov, 2012](#)), being also directly related applicable to fields such as commerce, biosecurity, environmental issues, fishery, medicine, mining prospection, public health, wildlife management, etc. ([Kapoor, 1998](#)). In particular, ecomorphological approaches benefit from taxonomic/morphologic products, which provides useful information for investigating the contribution of size and specific identity as related to function (revision in [Sibbing and Nagelkerke, 2000](#)). The relation between ecology and morphology has been clearly demonstrated in various animal taxa ([Ricklefs and Travis, 1980](#); [Miles and Ricklefs, 1984](#); [Miles et al., 1987](#); [Douglas and Matthews, 1992](#); [Silva and Brandão, 2010](#)). Researchers assume that species are more similar functionally in cases when they are morphologically more similar ([Chalcraft and Reserits, 2003](#)).

Taxonomists examine specimens, classifying them in taxonomic concepts (i.e. hypothesis of species; [Berendsohn, 1995](#); [Berendsohn and Geoffroy, 2007](#) and [Kennedy et al., 2006](#)), organize their

associated information (e.g. where and when they were collected and by whom) and describe their physical appearance through text annotations ([Deans et al., 2012](#)). These annotations compiled by taxonomists are transcribed in “descriptions”, including diagnoses, which serve to formalize species that can be recovered by researchers in the future ([Deans et al., 2012](#); [Miller et al., 2012](#)).

Taxonomic descriptions are a valuable source of knowledge concerning phenotypic diversity in the living world. These evidences, however, are generally restricted to taxonomic works, written by and consumed almost exclusively by taxonomists. Therefore, species' descriptions are not easily obtained and other scientists seldom reuse the phenotypic data resultant from it ([Deans et al., 2012](#)). Even when phenotypic data is obtained during the elaboration of non-taxonomic studies, it is rarely included in the finished study and, thus, not made publicly available ([Zamir, 2013](#)) or it is presented in an extremely heterogeneous and unordered manner, making its discovery problematic and interpretation impractical ([Deans et al., 2015](#)).

Although electronic availability of taxonomic treatments are rapidly growing, reflected in the change of publication requisites ([ICZN, 2012](#)) that foster digitalization endeavors ([Balhoff et al., 2013](#)), phenotypic descriptions are still made in natural language normally applying specialized anatomical terminology. This kind of description poses a challenge when researchers try to data mine it (however, cf. [Cui \(2012\)](#) and [Thessen et al. \(2012\)](#) for mechanisms to convert natural language descriptions in semantic annotations), probably due to the excessive amount of homonymies and synonymies that exist between anatomical concepts ([Yoder et al., 2010](#)).

Descriptions are constructed based on analogies and are seldom reused. According to [Deans et al. \(2012\)](#), descriptions are repeated when a taxon is revised, a mostly redundant process leading to the accretion of various text annotations of phenotypes for a small number of taxa, in which most of it turns out to be unused information. Owing to this inefficiency, many researchers consider descriptions as a nuisance (cf. [Evenhuis, 2007](#)) and, in some cases, have argued in favor of the delimitation of a taxon exclusively with molecular data ([Hebert et al., 2003](#); [Cook et al., 2010](#)). [Deans et al. \(2012\)](#) mentions that the limited utility of descriptions may partially explain recurrent stagnation of funding and training of taxonomists (cf. [Agnarsson and Kuntner, 2007](#); [Carvalho et al., 2007](#)) and the subsequent lack of morale that pervades this fundamental science.

There are three central issues, raised by [Deans et al. \(2012\)](#), that may be responsible for encumbering the elaboration of phenotypic semantic annotations: (i) the need for a physical printed version to validate a formal description (although this is mostly changed; see [Cressey, 2011](#); [Rinaldo and Norton, 2009](#)); (ii) descriptions are normally composed in non-uniform natural language that are inefficient for data-mining; and (iii) descriptions do not commonly refer to logical definitions of concepts (i.e. homonymy, polysemy and synonymy are prevalent) ([Yoder et al., 2010](#)). Thus, the final product of taxonomic work remains largely unexploited and barely used by investigators in other disciplines.

However, according to [Patterson et al. \(2010\)](#), taxonomy possesses two singular features that makes its reuse possible in biodiversity informatics. The first is the universal use of the system of scientific names that enables treating names as metadata, which can be used to index information related to biodiversity. The second attribute refers to classification schemes. They transform lists of names into organizational structures (i.e. ontologies) responsible for grouping data, allowing the construction of generalized declarations and the inferences of taxa properties by users, permitting extensive or focused searches and facilitating browsing for information ([Patterson et al., 2010](#)). The value of names as metadata and classification as ontologies lead to the idea that names-based

Download English Version:

<https://daneshyari.com/en/article/8877221>

Download Persian Version:

<https://daneshyari.com/article/8877221>

[Daneshyari.com](https://daneshyari.com)