Research data management in the field of Ecology: an overview

Cristiana Alves
CIBIO/InBIO, Portugal
cristianamaiaalves@gmail.com

João Aguiar Castro
INESC TEC, Portugal
joaoaguiarcastro@gmail.com

Cristina Ribeiro
INESC TEC, FEUP, Portugal
mcr@fe.up.pt

João Pradinho Honrado
CIBIO/InBIO, FCUP
Portugal
jhonrado@fc.up.pt

Angela Lomba
CIBIO/InBIO,
Portugal
angelalomba@fc.up.pt

Abstract

The diversity of research topics and resulting datasets in the field of Ecology (the scientific study of ecological systems and their biodiversity) has grown in parallel with developments in research data management. Based on a meta-analysis performed on 93 scientific references, this paper presents a comprehensive overview of the use of metadata tools in the Ecology domain through time. Overall, 40 metadata tools were found to be either referred or used by the research community from 1997 to 2018. In the same period, 50 different initiatives in ecology and biodiversity research were conceptualized and implemented to promote effective data sharing in the community. A relevant concern that stems from this analysis is the need to establish simple methods to promote data interoperability and reuse, so far limited by the production of metadata according to different standards. With this study, we also highlight challenges and perspectives in research data management in the domain of Ecology towards best practice guidelines.

Keywords: Biodiversity, Ecology, Research Data Management, Metadata tools, Literature review

1. Introduction

Ecology (the scientific study of ecological systems and of the biodiversity therein) is a challenging research community from the perspective of data management. Ecological and biodiversity data have been collected by researchers individually or as part of research teams, in the context of specific research questions and projects. Underlying data collection through time have been research topics such as the dynamics of specific habitats; the distribution and abundance of species; patterns and changes of environmental conditions; the processes that influence biological populations, communities, and ecosystems; and anthropogenic drivers of these processes (Berkley et al., 2009). Ecological and biodiversity data are collected by researchers using a wide variety of protocols tailored to address very diverse topics ranging from marine/terrestrial ecosystems to species distribution or genetics (Berkley, Jones, Bojilova, & Higgins, 2001). As a result, heterogeneous data are stored as independent datasets or databases that are dispersed throughout the research data facilities managed by ecological research communities. At the same time, to answer the multiple research questions, the need to share, describe and deposit data is a concern for many biodiversity researchers around the world.

Researchers are increasingly expected to take several measures regarding research data management (RDM), namely to comply with mandates that promote actions regarding data organization, sharing and publication. Benefits such as obtaining credit via citation or improving research workflows through collaboration may also encourage researchers to disseminate their data. Yet, availability of research data is not the same as existence of fit-for-reuse data (Tani, Candela, & Castelli, 2013). It depends, among other aspects, on specific metadata being provided to researchers so they can understand the data being accessed and evaluate their suitability. The inability to provide auxiliary information to

https://doi.org/10.23106/dcmi.952138986

contextualize research data is a practical impediment on data reuse (Thanos, 2017). In order to promote quality metadata, the European Commission (EC) is defining the principles to make data Findable, Accessible, Interoperable and Reusable, through the Guidelines on FAIR Data Management in Horizon 2020 (European Commission. Guidelines on FAIR Data Management in Horizon 2020., 2016).

The Research Data Alliance Metadata Standards Directory Working Group set out a metadata standards directory (Ball, Greenberg, Jeffery, & Koskela, 2016) for specific domains (Life Sciences, Engineering, Social and Behavioural Sciences) and for more general purposes. Nevertheless, the lack of resources in the long-tail of science (Heidorn, 2008) prompts researchers themselves to become active RDM stakeholders during the lifetime of projects, mostly to comply with funder or institutional policies and meet standards for good practice (Lyon, 2007). This means that research projects that do not have dedicated human resources to create standard-compliant metadata records place additional effort on researchers in the description of their data. Moreover, most standards are developed to describe data only at the end of the research workflow, with complex requirements that prevent researchers from adopting them consistently (Qin & Li, 2013). An evaluation of several metadata standards show that, although the flexibility to add new elements or modules to address community needs is a common objective in the development of scientific metadata standards, simplicity and sufficiency are not a top priority among them (Willis, Greenberg and White, 2012). Nevertheless, these features are likely to encourage researchers to describe their data, by making the process as easy as possible and focus on a minimal set of relevant metadata elements for the researchers to fill in.

Researchers are already metadata producers, yet in an ad-hoc sense and to fulfil specific, immediate needs (Mayernik, 2011). If provided with adequate tools, they are also more apt to describe context than information professionals. A promising path is to adopt metadata solutions that are tailor-made for researchers and their projects and can promote data reuse. Application Profiles, following the Singapore Framework logic of combining different standards components (Nilsson, 2008), are a practical implementation scenario to meet community-oriented metadata needs, offering the desirable flexibility but also enabling simplicity and sufficiency. The Minimum Information Framework, proposed in the geobiology community, for systematic documentation of sampling processes and particular contextual information about the site of data collection (Palmer et al., 2017), is a good example on how to design metadata tools driven by stakeholder needs and aiming at sufficiency (White, 2014).

The aim of this study is to present a comprehensive overview of the use of metadata tools in the Ecology domain through time. A meta-analysis focused on scientific literature on research data management in the field of Ecology was undertaken to support the identification and discussion of major initiatives, challenges and perspectives in research data management in this domain.

2. The Ecology domain

Ecological informatics is an interdisciplinary field that includes conceptual and methodological tools for the understanding, generation, processing and dissemination of various types of ecological data (Michener, Brunt, & Vanderbilt, 2002). Ecological informatics contributes to: (I) Experimental design phase; (II) Data plan; (III) Data acquisition and management; (IV) Quality assurance and control (QA/QC); (V) Metadata implementation; (VI) Data archival; (VII) Data access and dissemination; and (VIII) Data publication (Michener et al., 2002). For phases (I) and (II), designing the structure of datasets and implementing a logical structure within and among datasets can simplify data acquisition, entry, storage, retrieval and manipulation (Michener et al., 2002). In phase (III) the way in which data are acquired also affects data quality by influencing the amount of human error introduced into measurements. Phase (IV) refers to (QA/QC) strategies that are designed to avoid the introduction of errors, or data contamination into a dataset and the metadata in phase (V) is defined as "data about data" (NISO, 2004), so the datasets need to be described in their content, quality, structure, and accessibility (Michener et al., 2002). Different metadata standards have been developed to assure the description of datasets. Some are more generic and domain-neutral, like Dublin Core (Michener et al., 2002; Weibel, Kunze, Lagoze, & Wolf, 1998), while others are tailored to the biodiversity and ecological communities, such as Ecological Metadata Language (EML) (Michener, Brunt, Helly, Kirchner, & Stafford, 1997; Michener et al., 2002), and Darwin Core (Baker, Rycroft, & Smith, 2014). Others, like the EU INSPIRE Directive 2007/EC are specific metadata models, in this case for spatially explicit

https://doi.org/10.23106/dcmi.952138986

datasets (da Silva et al., 2014). Phase (VI) Data archival refers to assemblages of datasets packages that are stored, so users can locate, acquire, understand and use the data (Michener et al., 2002). Phase (VII) for data access and dissemination, and (VIII) for data publication ensure overall access to the datasets.

Ecological informatics is thus a framework that enables scientists to generate new knowledge through innovative tools, approaches and solutions that have been developed over the past decade, increasing scientists' efficiency and supporting faster and easier data discovery, integration and analysis; however, many challenges remain, especially in relation to incorporating Ecological informatics practices into mainstream research and education (Michener & Jones, 2012).

Ecological data cover a wide range of topics such as biodiversity surveys, measurements of environmental condition, inventories of species names and synonyms, species distributions, images and sounds, ecological interactions, behaviour, data set descriptions, and analyses and interpretations (Costello, Michener, Gahegan, Zhang, & Bourne, 2013). The variety of the ecological data makes it difficult to create simple, standardized methods to share resulting datasets, and consequently ecological data is currently described using several metadata models (D. Higgins, Berkley, & Jones, 2002) Further, usually data repositories have limited interoperability due to a lack of standards for data and communication protocols (Wieczorek et al., 2012). Inconsistent and ambiguous terminology in the description of biological datasets also creates obstacles in numerous aspects of data integration and use, including discovery, comparison, and quality assessment. It also makes data reuse by other scientists difficult (Baker et al., 2014; Wieczorek et al., 2012).

The need to start collaborative, multi-disciplinary research programs has been highlighted in order to overcome the challenge of efficiently and comprehensively collecting, documenting, communicating, and ultimately preserving primary research data (Jones et al., 2007). In fact, scientists, professional societies and research sponsors are recognizing the value of data as a product of the scientific enterprise and placing increased emphasis on data stewardship, data sharing, openness and supporting study repeatability (Michener & Jones, 2012). Various initiatives (from legal directives to informatics platforms) were developed to enable the sharing of ecological data:

1. Knowledge Network for Biocomplexity (KNB) (Berkley et al., 2009);
2. INSPIRE (Jones et al., 2007);
3. LTER (Michener, Porter, Servilla, & Vanderbilt, 2011);
4. Map of Life (Jetz, McPherson, & Guralnick, 2012);
5. GBIF (Costello et al., 2013).

Data repositories have also been growing rapidly and hold a tremendous promise for increasing the scope, coverage and societal relevance of ecological and biodiversity studies. Nevertheless, the data in these repositories still do not represent a reasonable portion of the massive ecological, environment and biodiversity data that are collected each year (Berkley et al., 2009).

3. Methods

For this review and for the meta-analysis performed, keywords or expressions based in the core area (i.e. Ecology, including Biodiversity), and then specific keywords from research data management (i.e. metadata and data management), were selected. The rationale behind the choice of keywords was to capture as many papers as possible in the Ecology domain and, more specifically, within data management. The Keywords selected were 'Metadata' OR 'Ontology-based approach' OR 'Data management' AND 'Ecolog*' OR 'Biodiversity', then redefined by the following scientific areas: Computer Science Information Systems, Computer Science Theory Methods, Computer Science Interdisciplinary Applications, Computer Science Hardware Architecture, Information Science, Library Science and Computer Science Software Engineering. This was done in order to capture papers within the research data management area ('Biodiversity' and 'Ecology' were not used because it was already in the keywords).

The time span of the search was 1900 to 2018. Searches were carried out between October 2017 and March 2018. ISI Web of Science (ISI WOS; http://webofknowledge.com/) was used, since it offers the widest coverage of published scientific literature (Buchadas et al., 2017; J. P. Higgins & Green, 2011).

https://doi.org/10.23106/dcmi.952138986

However, records gathered from Google Scholar that were absent from the ISI Web of Knowledge search were added to the final dataset. The inclusion criterion was to encompass works in the field of ecology and biodiversity with metadata methods (e.g. metadata models, data repositories, metadata language, data management). The selection was performed by individually examining first the title, keywords and abstract, and then the full text of the scientific manuscript. The exanimation of the papers and the decision on inclusion were made by an expert in the field of ecology and biodiversity.

4. Results and Discussion

The number of records retrieved from ISI Web of Science when using 'Metadata' as keyword was 15360. However, when including 'Ontology-based approach' as additional keyword, 15981 records were obtained. When including also keywords related to 'Ecolog*' OR 'Biodiversity' the number of records decreased to 368. After refining per scientific areas, the final number of records was 75 (in October 2017) and 126 (in March 2018) (Table 1).

Table 1 - Number of records retrieved in the literature search in the ISI Web of Science.

Keywords (General)	Keywords (Domain specific)	Results
Metadata		15360
"metadata" OR "Ontology-based approach"		15981
"metadata" OR "Ontology-based approach" OR "data management"	"Ecolog*"	288
"metadata" OR "Ontology-based approach" OR "data management"	"Ecolog*" OR "Biodiversity"	368
	redefined by scientific areas	75
"metadata" OR "Ontology-based approach" OR "data management"	"Ecolog*" OR "Biodiversity"	681
	redefined by scientific areas	126

The final subset of records included in this study was 93 (from the first and second literature search lists, 75 and 126 records, retrieved from ISI WOS) when applied the inclusion criterion (Fig.1).

When analysing the temporal evolution of the use of metadata tools in the scientific domain in the final subset of 93 records (Fig. 1), an increasing number of records in recent years is observed. One possible explanation is the gradual increase of the awareness towards the importance of metadata as a way to improve data management and data repository services.

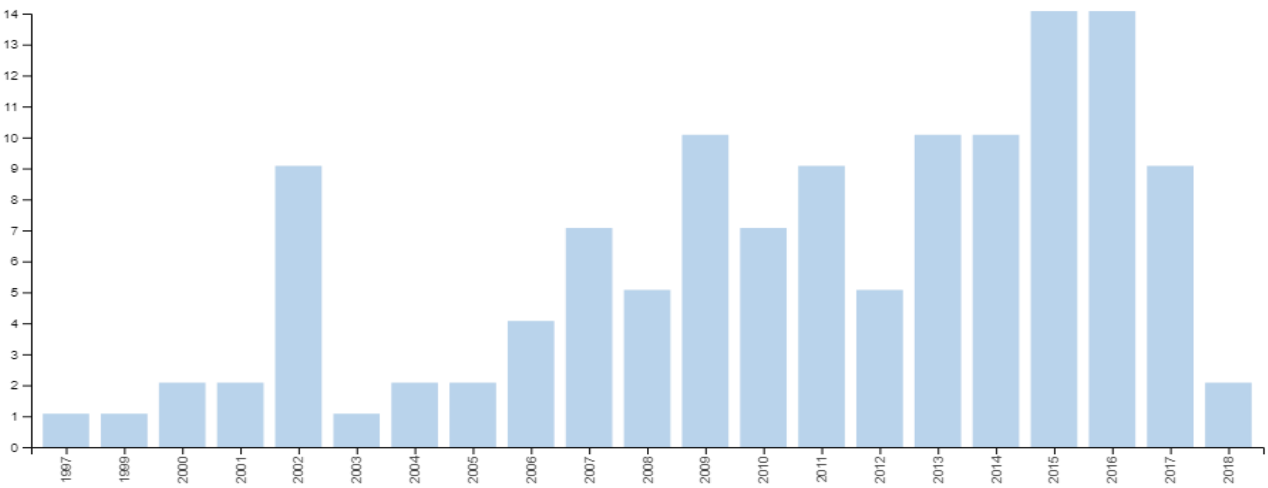

Figure 1 - Number of records retrieved from the literature search in ISI Web of Science per year (temporal overview).

https://doi.org/10.23106/dcmi.952138986

We found 40 different metadata tools either mentioned or used in the scientific manuscripts analysed from 1997 (the data of the first record) to 2018 (see Fig. 2). Each line of the graphic on Figure 2 corresponds to a metadata tool found in more than 3 records, while dots correspond to metadata tools that appear 3 times in the records analysed. The scale in the left refers to the number of records represented by the lines. The scale in right represent the number of records showed by the dots. Metadata tools that only appear once or twice were excluded from the graphical representation in order to simplify the visualization.

Most of these metadata tools are schemas and standards, such as: ABCD schema (Access to Biological Collection Data); FGDC Content Standard for Digital Geospatial Metadata; Crystallographic Information File (CIF); Darwin Core; Data Documentation Initiative; Directory Interchange Format (DIF); Dublin Core; Ecological Metadata Language; GML; Humboldt Core; ISO 19115 and its adoption by the INSPIRE metadata guidelines; ISO 19139; Macromolecular Crystallographic Information File (mmCIF); MIAME Notation in Markup Language (MINiML); Micro-Array Gene Expression Markup Language (MAGE-ML), ThermoML.
Other metadata tools range from metadata catalogues, ontologies, profiles and extensions to metadata editors and encoding standards, namely the: FGDC/CSDGM Biological Data Profile, Darwin Core (semantic web), Encoded Archival Description, Global Change Master Directory´s Interchange Format, iQL, Metacat, MIMOSA/ISO based XML schema, Morpho, NDG models, NEXML ThermoML, OBOE, SEEK, TERN Eco-Informatics data portal known as ÆKOS.
Our results showed that 14 different metadata tools were found to be used in the Ecology domain more than once in the records analysed.
The Ecological Metadata Language (EML) is a metadata standard widely applied in projects and platforms, since its year of implementation, 1997 (Aloisio, Milillo, & Williams, 1999; Michener et al., 1997). INSPIRE is based on the infrastructures for Spatial information established and operated by the European Union and was implemented in 2007. INSPIRE is the first "regional approach" and a legislative attempt to harmonize metadata standards for spatially explicit data (Filetti & Gnauck, 2011). The Darwin Core standard is used for sharing data about biodiversity and it first emerged in 1999 (Wieczorek et al., 2012).

Our review also revealed 50 different platforms/projects in ecology and biodiversity with the specific aim to encourage scientists to share, describe and publish their data. In Table 2, we list examples of such platforms/projects and the associated metadata standard. These examples, to date, are still available in the corresponding website and were mentioned in more than 1 record from the manuscripts analysed. Long-term Ecological Research (LTER) was initiated in 1980 for 6 sites, but this network has been increasing its reach globally since then (Michener et al., 2011). Likewise, the Knowledge Network for Biocomplexity (KNB) data repository has grown fast and now contains over 15,000 datasets (Berkley et al., 2009). The Taxonomic Databases Working Group (TDWG) was created to support the international collaboration of biodiversity informatics institutions and projects, to establish, adopt and promote standards and guidelines for the recording and exchange of data about organisms around the world (Veiga et al., 2017). The global initiative 'Map of Life' aims to gather, store and analyse data from species occurrences, fostering current knowledge on species distribution and contributing to reporting processes (Jetz et al., 2012). The Global Biodiversity Information Facility (GBIF) was created in 1999 and is currently the largest platform with more than seven hundred million occurrence records provided from more than 50 countries (Veiga et al., 2017). Other important initiatives in biodiversity and ecology are IPBES (Intergovernmental Platform on Biodiversity and Ecosystem Services) and GEO BON (Group on Earth Observation Biodiversity Observation Network) (Guralnick, Walls, & Jetz, 2017). Other datasets and data repositories retrieved in this review were: Forest Science Data Bank (FSDB), The Canopy Database Project, The Jalama Project, The Science Environment for Ecological Knowledge (SEEK), The BioCORE Project, The National Biological Information Infrastructure (NBII), Data ONE, The 'BEFdata' platform, BIOFRAG and IRBAS (The Intermittent River Biodiversity Analysis and Synthesis) (Cushing et al., 2007; Gil, Hutchison, Frame, & Palanisamy, 2010; Henshaw, Spycher, & Remillard, 2002; Leigh et al., 2017; Malaverri, Vilar, & Medeiros, 2009; Michener et al., 2007; Nadrowski et al., 2013; Pfeifer et al., 2014; Reichman, Jones, & Schildhauer, 2011).

https://doi.org/10.23106/dcmi.952138986

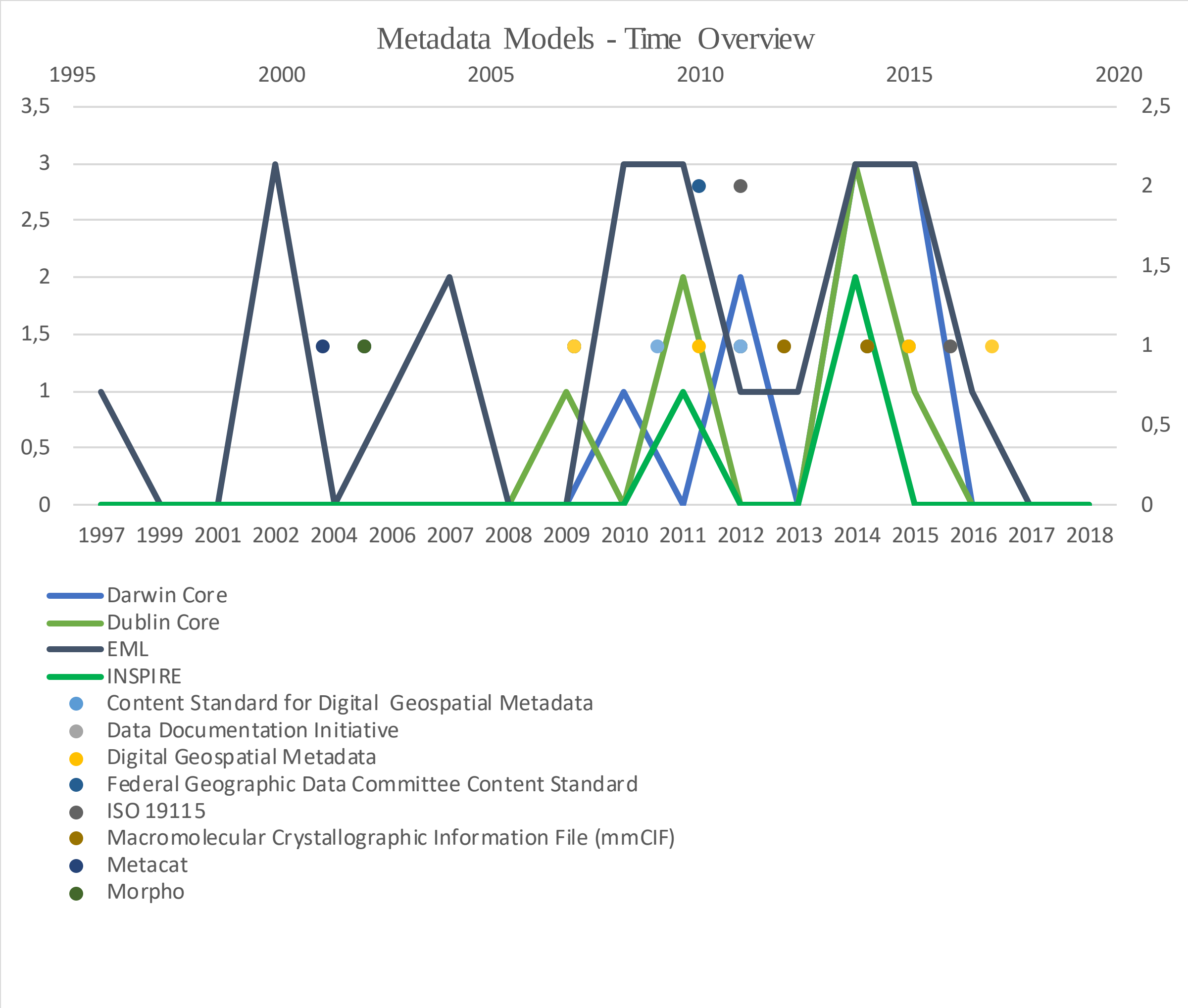

Figure 2 - Metadata tools retrieved from the records analysed through time (1997-2018).

Table 2 - Examples of platforms/initiatives developed and implemented to share, describe and publish data on the fields of ecology and biodiversity, and the associated metadata standards.

Platforms/Projects	Metadata Standard
Long-term Ecological Research (LTER)	EML
The Knowledge Network for Biocomplexity (KNB)	EML
Taxonomic Databases Working Group (TDWG)	Darwin Core
Map of Life Project	Darwin Core
Global Biodiversity Information Facility (GBIF)	Darwin Core

5. Conclusions and Future Perspectives

Since the 1990's the number of metadata tools referred and used by researchers in the field of Ecology has been increasing, alongside with the number of global and national/regional initiatives developed and implemented to share data according to common standards among researchers. With the development of metadata and initiatives to collect, store and share data among researchers, a wide range of metadata tools is currently available to researchers in the field. The 'big data' era further contributes to a pressing need to describe and publish data, so that it can be used within the same research area, as well as across research disciplines.

https://doi.org/10.23106/dcmi.952138986

With an increasing number of initiatives, platforms and repositories that can be used to deposit, publish and share their data with fellow scientists, researchers face new challenges. Such challenges relate e.g. to the lack of comprehensive metadata models that can be used to describe the various types of data used in the domain of Ecology. In many cases, researchers describe available datasets within the context of project consortia, when they are faced with the need to describe the data to be shared with fellow scientists. However, selecting and following a specific metadata model is not an easy task. A major challenge is to guarantee that previous metadata can be harmonized, so that previous work done by researchers is not lost.

Another relevant challenge is the complexity of the available metadata models. In fact, most metadata models available were not developed specifically to describe data in the domain of Ecology. A possible solution is proposed by Qin and Li (2013) consisting in a flexible ontology-based approach to break complex metadata standards into independent modules, so that metadata elements can be optimized for specific needs, while inconsistencies in naming conventions are also addressed. There is, thus, the pressing need to develop interdisciplinary research towards the development of suitable and easy to use metadata models and standards to foster data sharing and publication in the domain of Ecology.

Acknowledgements

This research is a result of the project TAIL – Research data management from creation to deposit and sharing - POCI-01-0145-FEDER-016736, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by National funds through FCT- Portuguese Science Foundation. (PTDC/EEI-ESS/1672/2014/2014).

References

Aloisio, G., Milillo, G., & Williams, R. D. (1999). An XML architecture for high-performance web-based analysis of remote-sensing archives. Future Generation Computer Systems, 16(1), 91-100. doi:10.1016/s0167-739x(99)00038-2

Baker, E., Rycroft, S., & Smith, V. S. (2014). Linking multiple biodiversity informatics platforms with Darwin Core Archives. Biodiversity Data Journal(2), e1039. doi:10.3897/BDJ.2.e1039

Ball, A., Greenberg, J., Jeffery, K., & Koskela, R. (2016). RDA Metadata Standards Directory Working Group.

Berkley, C., Bowers, S., Jones, M. B., Madin, J. S., Schildhauer, M., & Ieee. (2009). Improving Data Discovery in Metadata Repositories through Semantic Search.

Berkley, C., Jones, M., Bojilova, J., & Higgins, D. (2001). Metacat: A schema-independent XML database system.

Buchadas, A., Vaz, A. S., Honrado, J. P., Alagador, D., Bastos, R., Cabral, J. A., . . . Vicente, J. R. (2017). Dynamic models in research and management of biological invasions. Journal of Environmental Management, 196, 594-606. doi:https://doi.org/10.1016/j.jenvman.2017.03.060

Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z.-Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. Trends in Ecology & Evolution, 28(8), 454-461. doi:http://dx.doi.org/10.1016/j.tree.2013.05.002

Cushing, J. B., Nadkarni, N., Finch, M., Fiala, A., Murphy-Hill, E., Delcambre, L., & Maier, D. (2007). Component-based end-user database design for ecologists. Journal of Intelligent Information Systems, 29(1), 7-24. doi:10.1007/s10844-006-0028-6

da Silva, J. R., Castro, J. A., Ribeiro, C., Honrado, J., Lomba, A., & Goncalves, J. (2014). Beyond INSPIRE: An Ontology for Biodiversity Metadata Records. In R. Meersman, H. Panetto, A. Mishra, R. ValenciaGarcia, A. L. Soares, I. Ciuciu, F. Ferri, G. Weichhart, T. Moser, M. Bezzi, & H. Chan (Eds.), On the Move to Meaningful Internet Systems: Otm 2014 Workshops (Vol. 8842, pp. 597-607).

European Commission. Guidelines on FAIR Data Management in Horizon 2020. (2016). Retrieved from http://ec.europa.eu/research/%0Aparticipants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Filetti, M., & Gnauck, A. (2011). A Concept of a Virtual Research Environment for Long-Term Ecological Projects with Free and Open Source Software. In J. Hrebicek, G. Schimak, & R. Denzer (Eds.), Environmental Software Systems: Frameworks of Eenvironment (Vol. 359, pp. 235-244).

Gil, I. S., Hutchison, V., Frame, M., & Palanisamy, G. (2010). Metadata Activities in Biology. Journal of Library Metadata, 10(2-3), 99-118. doi:10.1080/19386389.2010.506389

Guralnick, R., Walls, R., & Jetz, W. (2017). Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography, n/a-n/a. doi:10.1111/ecog.02942

Heidorn, P. B. (2008). Shedding Light on the Dark Data in the Long Tail of Science. Library Trends, 57(2), 280-299. doi:10.1353/lib.0.0036

https://doi.org/10.23106/dcmi.952138986

Henshaw, D. L., Spycher, G., & Remillard, S. M. (2002). Transition from a legacy databank to an integrated ecological information system.

Higgins, D., Berkley, C., & Jones, M. B. (2002). Managing heterogeneous ecological data using Morpho.

Higgins, J. P., & Green, S. (2011). Cochrane handbook for systematic reviews of interventions (Vol. 4): John Wiley & Sons.

Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. Trends in Ecology & Evolution, 27(3), 151-159. doi:http://dx.doi.org/10.1016/j.tree.2011.09.007

Jones, C., Blanchette, C., Brooke, M., Harris, J., Jones, M., & Schildhauer, M. (2007). A metadata-driven framework for generating field data entry interfaces in ecology. Ecological Informatics, 2(3), 270-278. doi:http://dx.doi.org/10.1016/j.ecoinf.2007.06.005

Leigh, C., Laporte, B., Bonada, N., Fritz, K., Pella, H., Sauquet, E., . . . Datry, T. (2017). IRBAS: An online database to collate, analyze, and synthesize data on the biodiversity and ecology of intermittent rivers worldwide. Ecology and Evolution, 7(3), 815-823. doi:10.1002/ece3.2679

Lyon, L. (2007). Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report.

Malaverri, J. G., Vilar, B., & Medeiros, C. B. (2009). A TOOL BASED ON WEB SERVICES TO QUERY BIODIVERSITY INFORMATION.

Mayernik, M. (2011). Metadata realities for cyberinfrastructure: Data authors as metadata creators.

Michener, W. K., Beach, J. H., Jones, M. B., Ludascher, B., Pennington, D. D., Pereira, R. S., . . . Schildhauer, M. (2007). A knowledge environment for the biodiversity and ecological sciences. Journal of Intelligent Information Systems, 29(1), 111-126. doi:10.1007/s10844-006-0034-8

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., & Stafford, S. G. (1997). NONGEOSPATIAL METADATA FOR THE ECOLOGICAL SCIENCES. Ecological Applications, 7(1), 330-342. doi:10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

Michener, W. K., Brunt, J. W., & Vanderbilt, K. L. (2002). Ecological informatics: A long-term ecological research perspective.

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology & Evolution, 27(2), 85-93. doi:http://dx.doi.org/10.1016/j.tree.2011.11.016

Michener, W. K., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. Ecological Informatics, 6(1), 13-24. doi:http://dx.doi.org/10.1016/j.ecoinf.2010.11.005

Nadrowski, K., Ratcliffe, S., Bönisch, G., Bruelheide, H., Kattge, J., Liu, X., . . . Wirth, C. (2013). Harmonizing, annotating and sharing data in biodiversity–ecosystem functioning research. Methods in Ecology and Evolution, 4(2), 201-205. doi:10.1111/2041-210x.12009

Nilsson, M. (2008). The Singapore framework for Dublin Core application profiles. ttp://dublincore. org/documents/singapore-framework/.

Palmer, C. L., Thomer, A. K., Baker, K. S., Wickett, K. M., Hendrix, C. L., Rodman, A., . . . Fouke, B. W. (2017). Site-based data curation based on hot spring geobiology. PLOS ONE, 12(3), e0172090.

Pfeifer, M., Lefebvre, V., Gardner, T. A., Arroyo-Rodriguez, V., Baeten, L., Banks-Leite, C., . . . Ewers, R. M. (2014). BIOFRAG – a new database for analyzing BIOdiversity responses to forest FRAGmentation. Ecology and Evolution, 4(9), 1524-1537. doi:10.1002/ece3.1036

Qin, J., & Li, K. (2013). How portable are the metadata standards for scientific data? a proposal for a metadata infrastructure. Paper presented at the International Conference on Dublin Core and Metadata Applications.

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. Science, 331(6018), 703-705. doi:10.1126/science.1197962

Tani, A., Candela, L., & Castelli, D. (2013). Dealing with metadata quality: The legacy of digital library efforts. Information Processing & Management, 49(6), 1194-1205.

Thanos, C. (2017). Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies. Publications, 5(1), 2.

Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., Gendreau, C., Schigel, D., & Robertson, T. J. (2017). A conceptual framework for quality assessment and management of biodiversity data. PLOS ONE, 12(6), e0178731. doi:10.1371/journal.pone.0178731

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). Dublin core metadata for resource discovery (2070-1721).

White, H. C. (2014). Descriptive Metadata for Scientific Data Repositories: A Comparison of Information Scientist and Scientist Organizing Behaviors. Journal of Library Metadata, 14(1), 24-51. doi:10.1080/19386389.2014.891896

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., . . . Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLOS ONE, 7(1), e29715. doi:10.1371/journal.pone.0029715

https://doi.org/10.23106/dcmi.952138986