Research Ideas and Outcomes 2: e10617
doi: 10.3897/rio.2.e10617

Research Presentation

Online direct import of specimen records into manuscripts and automatic creation of data papers from biological databases

Viktor Senderov‡, Teodor Georgiev§, Lyubomir Penev‡

‡ Pensoft Publishers & Bulgarian Academy of Sciences, Sofia, Bulgaria
§ Pensoft Publishers, Sofia, Bulgaria

Corresponding author: Viktor Senderov (datascience@pensoft.net)

Reviewable v1

Received: 23 Sep 2016 | Published: 23 Sep 2016

Citation: Senderov V, Georgiev T, Penev L (2016) Online direct import of specimen records into manuscripts and automatic creation of data papers from biological databases . Research Ideas and Outcomes 2: e10617. doi: 10.3897/rio.2.e10617

Abstract

Background

This is a Research Presentation paper, one of the novel article formats developed for the Research Ideas and Outcomes (RIO) journal and aimed at representing brief research outcomes. In this paper we publish and discuss our webinar presentation for the Integrated Digitized Biocollections (iDigBio) audience on two novel publishing workflows for biodiversity data: (1) automatic import of specimen records into manuscripts, and (2) automatic generation of data paper manuscripts from Ecological Metadata Language (EML) metadata.

New information

Information on occurrences of species and information on the specimens that are evidence for these occurrences (specimen records) is stored in different biodiversity databases. These databases expose the information via public REST API's. We focused on the Global Biodiversity Information Facility (GBIF), Barcode of Life Data Systems (BOLD), iDigBio,

and PlutoF, and utilized their API's to import occurrence or specimen records directly into a manuscript edited in the ARPHA Writing Tool (AWT).

Furthermore, major ecological and biological databases around the world provide information about their datasets in the form of EML. A workflow was developed for creating data paper manuscripts in AWT from EML files. Such files could be downloaded, for example, from GBIF, DataONE, or the Long-Term Ecological Research Network (LTER Network).

Keywords

biodiversity informatics, bioinformatics, semantic publishing, API, REST, iDigBio, Global Biodiversity Information Facility, GBIF, PlutoF, BOLD Systems, ecological informatics, Ecological Metadata Language, EML, Darwin Core, LTER Network, DataONE, DwC-SW, semantic web

Introduction

On 16 June 2016, V. Senderov and L. Penev held a webinar presenting two novel workflows developed at Pensoft Publishers, used in the Biodiversity Data Journal (BDJ), and soon to be used also in other Pensoft journals of relevance: (1) automatic import of occurrence or specimen records into manuscripts and (2) automatic generation of data paper manuscripts from Ecological Metadata Language (EML) metadata. The aim of the webinar was to familiarize the biodiversity community with these workflows and motivate the workflows from a scientific standpoint. The title of the webinar was "Online direct import of specimen records from iDigBio infrastructure into taxonomic manuscripts."

Integrated Digitized Biocollections (iDigBio) is the leading US-based aggregator of biocollections data. They hold regular webinars and workshops aimed at improving biodiversity informatics knowledge, which are attended by collection managers, scientists, and IT personnel. Thus, doing a presentation for iDigBio was an excellent way of making the research and tools-development efforts of Pensoft widely known and getting feedback from the community.

Our efforts, which are part of the larger PhD project of V. Senderov to build an Open Biodiversity Knowledge Management System (OBKMS) (Senderov and Penev 2016), focused on two areas: optimizing the workflow of specimen data and optimizing the workflow of dataset metadata. These efforts resulted in the functionality that it is now possible, via a record identifier, to directly import specimen record information from the Global Biodiversity Information Facility (GBIF), Barcode of Life Data Systems (BOLD), iDigBio, or PlutoF into manuscripts in the ARPHA Writing Tool (AWT). No manual copying or retyping is required. Moreover, we created a second, data paper-based workflow.

The concept of data papers as an important means for data mobilization was introduced to biodiversity science by Chavan and Penev (2011). The data paper is a scholarly journal publication whose primary purpose is to describe a dataset or a group of datasets, rather than report a research investigation. Data papers serve to increase visibility, provide peer review, permanent scientific record, and credit and citation capabilities (via DOI) for biodiversity data. Thus, data papers support the effort for data to become a first class research product, and are a step forward in the direction of open science (Chavan and Penev 2011, Chavan et al. 2013).

Using this workflow, it is now possible to generate a data paper manuscript in AWT from a file formatted in recent EML versions.

Presentation

A video recording of the presentation is available. More information can be found in the webinar information page. The slides of the presentation are attached as supplementary files and are deposited in Slideshare.

During the presentation we conducted a poll about the occupation of the attendees, the results of which are summarized in Fig. 1. Of the participants who voted, about a half were scientists, mostly biologists, while the remainder were distributed across IT specialists and librarians, with 20% "Other." The other categories might have been administrators, decision-makers, non-biology scientists, collections personnel, educators, etc.

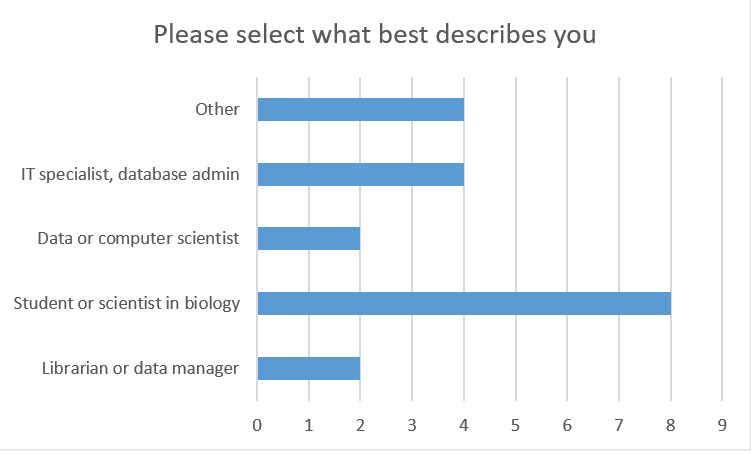

Figure 1.

Poll results about composition of audience during live participation.

At the end of the presentation, very interesting questions were raised and discussed. For details, see the "Results and discussion" section of this paper.

Larry Page, Project Director at iDigBio, wrote: “This workflow has the potential to be a huge step forward in documenting use of collections data and enabling iDigBio and other aggregators to report that information back to the institutions providing the data."

Neil Cobb, a research professor at the Department of Biological Sciences at the Northern Arizona University, suggested that the methods, workflows and tools addressed during the presentation could provide a basis for a virtual student course in biodiversity informatics.

Methods

Both discussed workflows rely on three key standards: RESTful API's for the web (Kurtz 2013), Darwin Core (Wieczorek et al. 2012), and EML (Fegraus et al. 2005).

RESTful is a software architecture style for the Web, derived from the dissertation of Fielding (2000). It is out of the scope of this paper to fully explain what a RESTful API is, but a short summary follows (after Kurtz 2013):

1. URI's have to be provided for different resources.
2. HTTP verbs have to be used for different actions.
3. HATEOAS (Hypermedia as the Engine of Application State) must be implemented. This is a way of saying that the client only needs to have a basic knowledge of hypermedia, in order to use the service.

On the other hand, Darwin Core (DwC) is a standard developed by the Biodiversity Information Standards (TDWG), also known as the Taxonomic Databases Working Group, to facilitate storage and exchange of biodiversity and biodiversity-related information. ARPHA and BDJ use the DwC terms to store taxonomic material citation data.

Finally, EML is an XML-based open-source metadata format developed by the community and the National Center for Ecological Analysis and Synthesis (NCEAS) and the Long Term Ecological Research Network (LTER, Fegraus et al. 2005).

Development of workflow 1: Automated specimen record import

There is some confusion about the terms *occurrence record, specimen record*, and *material citation.* A DwC Occurrence is defined as "an existence of an Organism at a particular place at a particular time." The term specimen record is a term that we use for cataloged specimens in a collection that are evidence for the occurrence. In DwC, the notion of a specimen is covered by MaterialSample, LivingSpecimen, PreservedSpecimen, and FossilSpecimen. The description of MaterialSample reads: "a physical result of a sampling (or sub-sampling) event. In biological collections, the material sample is typically collected, and either preserved or destructively processed." While there is a semantic difference between an occurrence record (DwC Occurrence) and a specimen record (DwC MaterialSample, LivingSpecimen, PreservedSpecimen, or FossilSpecimen), from the view

point of pure syntax, they can be considered equivalent since both types of objects* are described by the same fields in our system grouped in the following major groups:

- Record-level information
- Event
- Identification
- Location
- Taxon
- Occurrence
- Geological context

Taxonomic practice dictates that authors cite the materials their analysis is based on in the treatment section of the taxonomic paper (Catapano 2010). Therefore, in our system, as it is a document-authoring system, we take the view that these objects are material citations, i.e. bibliographic records that refer to a particular observation in the wild or a specimen in a museum. As a supplementary file (Suppl. material 1), we've attached a spreadsheet listing all DwC terms describing a specimen or occurrence record that can be imported into AWT as a material citation. From here on, we will refer to the objects being imported as specimen records, and to the objects that are part of the manuscript as material citations.

At the time when development of the workflow started, AWT already allowed imp ort of specimen records as material citations via manual interface and via spreadsheet (Suppl. material 1). So, the starting point for the development of the workflow was to locate API's for downloading biodiversity and biodiversity-related data from major biodiversity data providers and to transform the data that was provided by these API's into DwC-compatible data, which was then to be imported into the manuscript. We chose to work with GBIF, BOLD Systems, iDigBio and the PlutoF platform.

In Suppl. material 2 we give all the necessary information about the API's and how to map their results to DwC: endpoints, documentation, and the mapping of terms. GBIF and iDigBio name their fields in accordance with DwC, whereas for PlutoF and BOLD Systems there is a direct mapping given in the spreadsheet.

In order to abstract and reuse source code we have created a general Occurrence class, which contains the code that is shared between all occurrences, and children classes GbifOccurrence, BoldOccurrence, IDigBioOccurrence, and PlutoFOccurrence, which contain the provider-specific code. The source code is written in PHP.

* Note: we are using the term *objects* here in the computer science sense of the word to denote generalized data structures.

Development of workflow 2: Automated data paper generation

Data papers are scholarly articles describing a dataset or a data package (Chavan and Penev 2011). Similarly, metadata are data about a dataset (Michener 2006). Ecological metadata can be expressed in an XML-language called EML (Fegraus et al. 2005). It formalizes the set of concepts needed for describing ecological data (Fegraus et al. 2005).

It is broad enough and allows dataset authors from neighboring disciplines (e.g. taxonomy) to annotate their datasets with it. We asked ourselves the question: would it be possible to convert such metadata into a data paper manuscript? As the data paper manuscript in AWT is also stored as XML (for format details see the Pensoft API), all that was needed was an XSLT transformation mapping the fields of EML to the fields in the data paper XML. We have created two such transformations, for EML version 2.1.1 and for EML version 2.1.0, which we've attached as Suppl. material 3. The workflow has been tested with EML metadata downloaded from GBIF, DataONE and the LTER Network, however, it can be used for EML metadata from any other data provider.

Data resources

The presentation this paper describes is available from Slideshare: www.slideshare.net/ViktorSenderov/online-direct-import-of-specimen-records-from-idigbio-infrastructure-into-taxonomic-manuscripts.

Results and discussion

Workflow 1: Automated specimen record import into manuscripts developed in the ARPHA Writing Tool

Implementation: It is now possible to directly import a specimen record as a material citation in an ARPHA Taxonomic Paper from GBIF, BOLD, iDigBio, and PlutoF (Slide 5, as well as Fig. 2). The workflow from the user's perspective has been thoroughly described in a blog post; concise stepwise instructions are available via ARPHA's Tips and tricks guidelines. In a nutshell, the process works as follows:

1. At one of the supported data portals (BOLD, GBIF, iDigBio, PlutoF), the author locates the specimen record he/she wants to import into the Materials section of a Taxon treatment (available in the Taxonomic Paper manuscript template).
2. Depending on the portal, the user finds either the occurrence identfier of the specimen, or a database record identifier of the specimen record, and copies that into the respective upload field of the ARPHA system (Fig. 3).
3. After the user clicks on "Add," a progress bar is displayed, while the specimens are being uploaded as material citations.
4. The new material citations are rendered in both human- and machine-readable DwC format in the Materials section of the respective Taxon treatment and can be further edited in AWT, or downloaded from there as a CSV file.

Discussion: The persistent unique identifiers (PID's) are a long-discussed problem in biodiversity informatics (Guralnick et al. 2014). Questions of fundamental importance are: given a specimen in a museum, is it (and how often is it) cited in a paper? What about the quality of the database record belonging to this specimen? In order for us to be able to

answer these questions to our satisfaction, specimens need to have their own unique identifiers, which are imported as part of the material citation and allow the researcher to scan through the body of published literature to find which specimens have been cited (and how often). In practice, however, this is not always the case and we have to rely on the DwC triplet, (institutionCode, collectionCode, catalogNumber), to positively identify specimens (Guralnick et al. 2014). In the next paragraphs, we discuss how the information provided by the repositories can be used to track material citations.

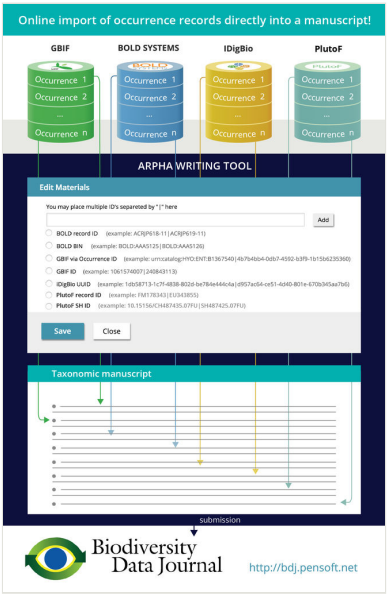

Figure 2.

This fictionalized workflow presents the flow of information content of biodiversity specimens or biodiversity occurrences from the data portals GBIF, BOLD Systems, iDigBio, and PlutoF, through user-interface elements in AWT to textualized content in a Taxonomic Paper manuscript template intended for publication in the Biodiversity Data Journal.

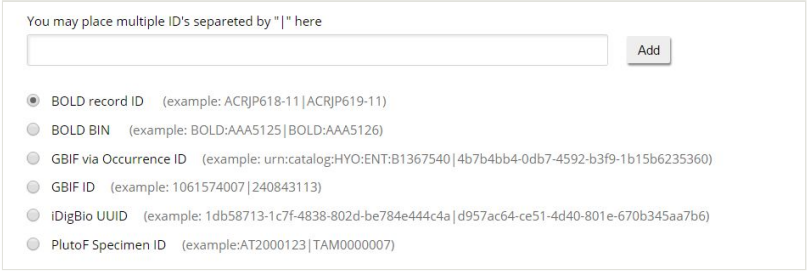

Figure 3.

User interface of the ARPHA Writing Tool controlling the import of specimen records from external databases.

GBIF: Import from GBIF is possible both via a DwC occurrenceID, which is the unique identifier for the specimen/ occurrence, or via a GBIF ID, which is the record ID in GBIF's database. Thanks to its full compliance with DwC, it should be possible to track specimens imported from GBIF.

BOLD Systems: In the BOLD database, specimen records are assigned an identifier, which can look like `ACRJP619-11`. This identifier is the database identifier and is used for the import; it is not the identifier issued to the specimen stored in a given collection. However, some collection identifiers are returned by the API call and are stored in the material citation, for example, DwC catalogNumber and DwC institutionCode (see mappings in Suppl. material 2). In this case, we have what is called a DwC doublet (Guralnick et al. 2014), which provides us with the minimum amount of information to attempt an identification.

A feature of BOLD Systems is that records are grouped into BIN's representing Operational Taxonomic Units (OTU's) based on a hierarchical/ graph-based clustering algorithm (Ratnasingham and Hebert 2013). It is possible to import all specimen records from a BIN in a single step, specifying the BIN ID. This streamlines the description of new species from OTU's as it allows the taxonomist to save time and import all materials from the BIN.

iDigBio: iDigBio provides its specimen records in a DwC-compatible format. Similar to GBIF, both a DwC occurrenceID, as well as DwC triplet information is returned by the system and stored in our XML making tracking of specimen citations easy.

PlutoF: Import from PlutoF is attained through the usage of a specimen ID (DwC catalogNumber), which is disambiguated to a PlutoF record ID by our system. If a specimen ID matches more than one record in the PlutoF system, multiple records are imported and the user has to delete the superfluous material citations. PlutoF does store a full DwC triplet while no DwC occurrenceID is available for the time being.

Ultimately, this workflow can serve as a curation filter for increasing the quality of specimen data via the scientific peer review process. By importing a specimen record via our workflow, the author of the paper vouches for the quality of the particular specimen record that he or she presumably has already checked against the physical specimen. Then a specimen that has been cited in an article can be marked with a star as a peer-reviewed specimen by the collection manager. Also, the completeness and correctness of the specimen record itself can be improved by comparing the material citation with the database record and synchronizing differing fields.

There is only one component currently missing from for this curation workflow: a query page that accepts a DwC occurrenceID or a DwC doublet/ triplet and returns all the information stored in the Pensoft database regarding material citations of this specimen. We envisage this functionality to be part of the OBKMS system.

Workflow 2: Automated data paper manuscript generation from EML metadata in the ARPHA Writing Tool

Implementation: We have created a workflow that allows authors to automatically create data paper manuscripts from the metadata stored in EML. The completeness of the manuscript created in such a way depends on the quality of the metadata; however, after generating such a manuscript, the authors can update, edit, and revise it as any other scientific manuscript in the AWT. The workflow has been thoroughly described in a blog post; concise stepwise instructions are available via ARPHA's Tips and tricks guidelines. In a nutshell, the process works as follows:

1. The users of ARPHA need to save a dataset's metadata as an EML file (versions 2.1.1 and 2.1.0, support for other versions is being continually updated) from the website of the respective data provider (see Fig. 4 as an example using the GBIF's Integrated Publishing Toolkit (IPT)). Some leading data portals that provide such EML files are GBIF (EML download possible both from IPT and from the portal), DataONE, and the LTER Network.
2. Click on the "Start a manuscript" button in AWT and then select "Biodiversity Data Journal" and the "Data paper (Biosciences)" template (Fig. 5).
3. Upload this file via the "Import a manuscript" function on the AWT interface (Fig. 6).
4. Continue with updating and editing and finally submit your manuscript inside AWT.

Figure 4.

Download of an EML from the GBIF Integarted Publishuing Toolkit (IPT)

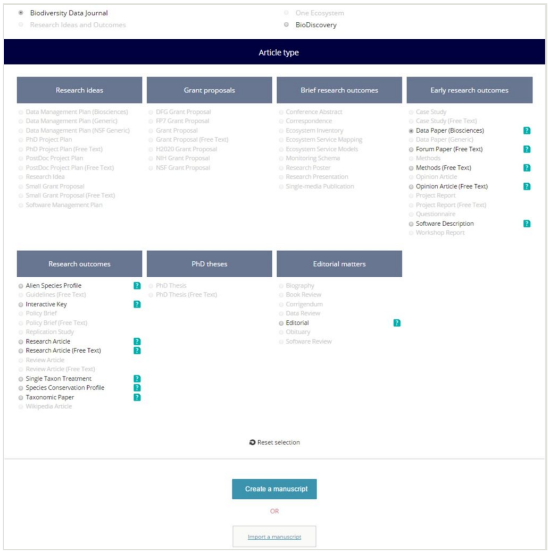

Figure 5.

Selection of the journal and "Data Paper (Biosciences)" template in the ARPHA Writing Tool.

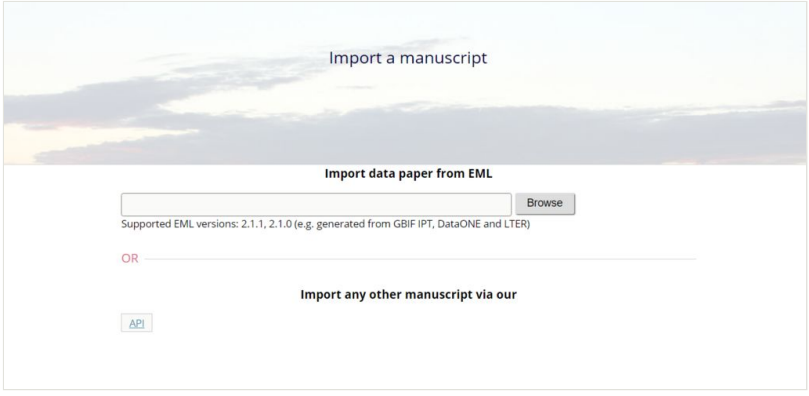

Figure 6.

The user interface field for uploading EML files into ARPHA.

Discussion: In 2010, GBIF and Pensoft began investigating mainstream biodiversity data publishing in the form of "data papers." As a result this partnership pioneered a workflow between GBIF's IPT and Pensoft's journals, viz.: ZooKeys, MycoKeys, Phytokeys, Nature Conservation, and others. The rationale behind the project was to motivate authors to create proper metadata for their datasets to enable themselves and their peers to properly make use of the data. Our workflow gives authors the opportunity to convert their extended metadata descriptions into data paper manuscripts with very little extra effort. The workflow generates data paper manuscripts from the metadata descriptions in IPT automatically at the "click of a button." Manuscripts are created in Rich Text Format (RTF) format, edited

and updated by the authors, and then submitted to a journal to undergo peer review and publication. The publication itself bears the citation details of the described dataset with its own DOI or other unique identifier. Ideally, after the data paper is published and a DOI is issued for it, it should be included in the data description at the repository where the data is stored. Within less than four years, a total of more than 100 data papers have been published in Pensoft's journals (examples: Dekoninck et al. 2013, Desmet and Brouilet 2013, Gutt et al. 2013, Pierrat et al. 2012, Shao et al. 2012, Tng et al. 2016). The workflow and associated author guidelines are described in Penev et al. (2011).

The present paper describes the next technological step in the generation of data papers: direct import of an EML file via an API into a manuscript being written in AWT. A great advantage of the present workflow is that data paper manuscripts can be edited and peer-reviewed collaboratively in the authoring tool even before submission to the journal. These novel features provided by AWT and BDJ may potentially become a huge step forward in experts' engagement and mobilization to publish biodiversity data in a way that facilitates recording, credit, preservation, and re-use. Another benefit of this usage of EML data might be that in the future, more people wil provide more robust EML data files.

Feedback: The two workflows presented generated a lively discussion at the end of the presentation, which we summarize below:

1. Are specimen records imported from GBIF and then slightly changed during the editorial process then deduplicated at GBIF? Answer: Unfortunately, no. At GBIF, deduplication only occurs for identical records.
2. Are we leaving the identifiers from GBIF or iDigBio in the records? Answer: Yes. We have made the best effort to import specimen record identifiers. This has been discussed in the previous sections.
3. How will the tool reduce the input time for constructing a manuscript? Answer: AWT reduces the time for creating a manuscript in two significant ways. First of all, the workflows avoid retyping of specimen records or metadata. Secondly, another time-saving feature is the elimination of copying errors. Creating of data paper manuscripts from EML saves, as a minimum, the effort of copy-pasting metadata and their arrangement in a manuscript.
4. What are the major hurdles or challenges left in having this become a mainstream tool? How mature is the tool? Answer: We believe that the main hurdles in this becoming a main-stream tool are visibility and awareness of the tool by the community. As the stability of the software is already at a very good stage.
5. Is it possible to track the usage of museum specimens for data aggregators? Answer: Yes, see question 2 and discussion in the present section.
6. How do you go to the article page where collection managers can search for data published from their collections on the Pensoft website? Answer: We are working on the streamlining of this functionality. It will be part of the OBKMS. Currently, we markup collection codes against the Global Registry of Biodiversity Repositories (GRBio) vocabularies, and the reader can view the records from a particular collection by clicking on the collection code.

Acknowledgements

The authors are thankful to the whole Pensoft team, especially the software development unit, as well as the PlutoF, GBIF and iDigBio staff for the valuable support during the implementation of the project. Special thanks are due to Deborah Paul, Digitization and Workforce Training Specialist from iDigBio, for giving us the opportunity to present the workflow at the webinar as part of the iDigBio 2015 Data Management Working Group series. We also thank also our pre-submission reviewers for the valuable comments.

Funding program

The basic infrastructure for importing specimen records was partially supported by the FP7 funded project EU BON - Building the European Biodiversity Observation Network, grant agreement ENV30845. V. Senderov's PhD is financed through the EU Marie-Sklodovska-Curie Program Grant Agreement Nr. 642241.

Hosting institution

Pensoft Publishers, Bulgarian Academy of Sciences

Author contributions

The workflows were developed by:

- V. Senderov - main author and implementor of the XSLT and RESTfull API workflows.
- T. Georgiev - project manager, co-architect.
- L. Penev - scientific advisor, vision.

References

- Baskauf S, Webb C (2014) Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF. http://www.semantic-web-journal.net/content/darwin-sw-darwin-core-based-terms-expressing-biodiversity-data-rdf-1. Accession date: 2016 7 14.
- Catapano T (2010) Journal Article Tag Suite Conference (JATS-Con) Proceedings 2010. Bethesda (MD): National Center for Biotechnology Information (US). 6 pp.
- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12: S2. DOI: 10.1186/1471-2105-12-s15-s2
- Chavan V, Penev L, Hobern D (2013) Cultural Change in Data Publishing Is Essential. BioScience 63 (6): 419-420. DOI: 10.1525/bio.2013.63.6.3

- Dekoninck W, Brosens D, Vankerkhoven F, Ignace D, Wegnez P, Noé N, Heughebaert A, Bortels J (2013) FORMIDABEL: The Belgian Ants Database. ZooKeys 306: 59-70. DOI: 10.3897/zookeys.306.4898
- Desmet P, Brouilet L (2013) Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland. PhytoKeys 25: 55-67. DOI: 10.3897/phytokeys.25.3100
- Fegraus E, Andelman S, Jones M, Schildhauer M (2005) Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. Bulletin of the Ecological Society of America 86 (3): 158-168. DOI: 10.1890/0012-9623(2005)86[158:mtvoed]2.0.co;2
- Fielding RT (2000) Architectural styles and the design of network-based software architectures. PhD Dissertation. Dept. of Information and Computer Science, University of California, Irvine., 180 pp.
- Guralnick R, Conlin T, Deck J, Stucky B, Cellinese N (2014) The Trouble with Triplets in Biodiversity Informatics: A Data-Driven Case against Current Identifier Practices. PLoS ONE 9 (12): e114069. DOI: 10.1371/journal.pone.0114069
- Gutt J, Barnes D, Lockhart S, de Putte Av (2013) Antarctic macrobenthic communities: A compilation of circumpolar information. Nature Conservation 4: 1-13. DOI: 10.3897/natureconservation.4.4499
- Kurtz J (2013) What is RESTful? ASP.NET MVC 4 and the Web API. URL: http://dx.doi.org/10.1007/978-1-4302-4978-8_2 DOI: 10.1007/978-1-4302-4978-8_2
- Michener W (2006) Meta-information concepts for ecological data management. Ecological Informatics 1 (1): 3-7. DOI: 10.1016/j.ecoinf.2005.08.004
- Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D (2011) Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Zenodo 1: 1. DOI: 10.5281/ZENODO.56660
- Pierrat B, Saucède T, Festeau A, David B (2012) Antarctic, Sub-Antarctic and cold temperate echinoid database. ZooKeys 204: 47-52. DOI: 10.3897/zookeys.204.3134
- Ratnasingham S, Hebert PN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8 (7): e66213. DOI: 10.1371/journal.pone.0066213
- Senderov V, Penev L (2016) The Open Biodiversity Knowledge Management System in Scholarly Publishing. Research Ideas and Outcomes 2: e7757. DOI: 10.3897/rio.2.e7757
- Shao K, Lin J, Wu C, Yeh H, Cheng T (2012) A dataset from bottom trawl survey around Taiwan. ZooKeys 198: 103-109. DOI: 10.3897/zookeys.198.3032
- Tng D, Apgaua D, Campbell M, Cox C, Crayn D, Ishida F, Liddell M, Seager M, Laurance S (2016) Vegetation and floristics of a lowland tropical rainforest in northeast Australia. Biodiversity Data Journal 4: e7599. DOI: 10.3897/bdj.4.e7599
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1): e29715. DOI: 10.1371/journal.pone.0029715

Supplementary materials

Suppl. material 1: Species occurrence Darwin Core template 1, version 1

Authors: Pensoft Publishers
Data type: spreadsheet
Brief description: A template for an occurrence or specimen record to be imported as a material citation.
Filename: Species_occurrence-1_v1_DwC_Template.xls - Download file (157.00 kb)

Suppl. material 2: Data Portals APIs and Mappings

Authors: Viktor Senderov, Teodor Georgiev
Data type: spreadsheet
Brief description: This spreadsheet contains the information about the specimen API's of GBIF, BOLD Systems, iDigBio, and PlutoF. It lists the endpoints and the documentation URLs in the sheet named "APIs". In the sheet named "Mappings" it lists how to map the non-DwC compliant APIs (BOLD and PlutoF) to DwC-terms.
Filename: oo_95899.xlsx - Download file (32.66 kb)

Suppl. material 3: XSLT transformations for data paper generation

Authors: Viktor Senderov
Data type: Zipped XSLT's
Brief description: This archive contains XSLT transformations from EML v. 2.1.1 and v. 2.1.0 to Pensoft data paper format.
Filename: oo_95900.zip - Download file (7.86 kb)