

Proposta de um sistema web para padronização, registro e integração de dados brutos em Ictiologia

Emerson Luiz F. Borges, Adriana Pereira Medeiros

Instituto de Ciência e Tecnologia - Universidade Federal Fluminense (UFF)
- Rio das Ostras – RJ – Brasil

emersonborges@macae.ufrj.br, adrianaedeiros@puro.uff.br

Abstract. *This paper describes the proposal of a web information system for formatting and structuring the sample events data and species occurrence in Ichthyology, according to the Darwin Core. The objective is promoting integration with SiBBr and other biodiversity information systems. In addition, the system will allow the insertion of water data from the sample environment and climatic data that will be used to assist researchers in understanding environmental factors that determine patterns of composition and abundance, and establishing cause and effect relationships.*

Resumo. *Este artigo descreve a proposta de um sistema de informação web para formatação e estruturação dos dados de eventos de coleta e ocorrência de espécies em Ictiologia no padrão Darwin Core. O objetivo é promover a integração com o SiBBr e outros sistemas de informação da biodiversidade. O sistema permitirá também a inserção de dados da água do ambiente de coleta e dados climáticos, que serão utilizados na compreensão de fatores ambientais que determinam os padrões de composição e abundância e estabelecimento de relações de causa e efeitos.*

1. Introdução

O Sítio Restingas e Lagoas Costeiras do Norte Fluminense (RLaC) compõe a rede de Pesquisas Ecológicas de Longa Duração (PELD). Seu objetivo é compreender a estrutura e o funcionamento dos ecossistemas de lagoas costeiras e restingas do Parque Nacional da Restinga de Jurubatiba e prever os efeitos de perturbações naturais e antrópicas sobre esses ecossistemas (Esteves e Petry 2008).

Uma característica comum das subáreas de pesquisa do PELD-RLaC é o registro inicial dos dados brutos (ainda não processados). Os pesquisadores da subárea de Ictiologia (estudo de peixes) registram dados brutos de espécimes como características, quantidade obtida em cada ponto de coleta, quantidade coletada por espécie, dados biométricos (medidas e peso), sexo, entre outros. Também são registradas as características do ambiente aquático como profundidade, salinidade, as quantidades de fósforo, nitrogênio, ph, carbono, oxigênio, clorofila e condutividade, e ainda, variáveis ambientais como temperatura, pressão atmosférica, umidade e quantidade de chuva.

A única forma de armazenamento desses dados é o registro em cadernos ou fichas de campo. Posteriormente, esses dados são inseridos em planilhas eletrônicas para gerar valores médios, desvios padrões e gráficos, com o objetivo de utilizá-los em monografias, dissertações, teses, relatórios e artigos científicos. O processo de criação dessas planilhas é feito de forma individual, sem nenhum tipo de padrão, dificultando o

compartilhamento dos dados brutos. De acordo com Rocha et al. (2004), esses dados estão sujeitos à degradação devido a problemas como: desvinculação de pesquisadores do projeto; morte de pesquisadores e problemas com hardware ou software que inviabilizam o acesso aos dados.

A disponibilização dos dados obtidos no âmbito de cada sítio de pesquisa no Sistema Brasileiro da Biodiversidade (SiBBR) é uma exigência do PELD (CNPq 2016). O SiBBR é uma plataforma on-line que pretende reunir dados e informações sobre a biodiversidade do Brasil, e foi desenvolvido para integrar e disseminar dados coletados e publicados por diversas instituições brasileiras, como universidades, institutos de pesquisa e agências governamentais (Gadelha et al. 2014). O principal problema enfrentado pelos pesquisadores de Ictiologia do PELD-RLaC é o mapeamento dos dados brutos coletados para um padrão que permita a integração com o SiBBR. Atualmente, esse mapeamento vem sendo efetuado manualmente, criando-se duas planilhas eletrônicas: uma nomeada *Event*, que armazena os dados referentes ao evento de coleta, e uma segunda planilha, nomeada *Occurrence*, onde são armazenados os dados referentes às ocorrências de organismos em cada evento discriminado na planilha *Event*.

Este artigo propõe um sistema de informação web que tem por objetivo permitir a inserção dos dados de coleta e de ocorrências de espécies em Ictiologia do sítio PELD-RLaC estruturados no padrão *Darwin Core*, para que posteriormente possam ser integrados ao SiBBR e outros sistemas. O sistema também permitirá a inserção de dados da água do ambiente de coleta e dados climáticos. Esses dados serão utilizados para auxiliar os pesquisadores na compreensão de fatores ambientais que determinam os padrões de composição e abundância e estabelecimento de relações de causa e efeitos.

2. O Padrão Darwin Core e a Ontologia BLO

O Darwin Core (TDWG 2009) é um glossário de termos de classes e propriedades criado para facilitar o compartilhamento de informações sobre a diversidade biológica. O núcleo do *Darwin Core* é baseado em categorias taxonômicas, suas ocorrências na natureza e suas interações com o ambiente. As classes do *Darwin Core* utilizadas neste trabalho são *Event*, *Occurrence*, *Location*, *Identification*, *Taxon*, *Organism* e *MeasurementOrFact*. A classe *Event* representa os eventos de coletas que ocorrem em determinado local em um determinado período de tempo. Cada instância da classe *Event* resultará em uma ou mais instâncias da classe *Occurrence*, que representa a ocorrência de espécies de cada evento. As espécies representadas em instâncias da classe *Occurrence* são identificadas com base numa classificação taxonômica representada pela classe *Taxon*. A classe *MeasurementOrFact* representa as instâncias de cada medida (comprimento, peso etc) de cada Organismo da classe *Occurrence*.

A ontologia BLO (*Batata Lake Ontology*) (Souza e Medeiros 2015) é uma ontologia de aplicação desenvolvida para descrever semanticamente dados de pesquisa e análise do Lago do Batata (Oriximiná/PA). Neste trabalho serão utilizadas as classes *Water*, *Rain*, *Air* e *Period*. Para cada evento de coleta são registrados os dados da água referentes ao ponto de coleta onde ocorreu o evento, esses dados são representados pela classe *Water*. Os dados climáticos, representados pelas classes *Rain* e *Air*, serão utilizados de acordo com um período de tempo, representado pela classe *Period*.

3. Arquitetura do sistema proposto

Na Figura 1, é apresentada a arquitetura do sistema *Freshwater Fish Ecology Data*,

composto por três camadas. Na primeira camada (1) o pesquisador poderá inserir, alterar, excluir e consultar dados dos eventos de coleta, ocorrências de espécies, medidas dos organismos e dados de água relacionados ao evento de coleta. Na camada intermediária (2), os dados de água (salinidade, ph, clorofila etc), assim como os dados climáticos (ar e chuva), serão fornecidos de acordo com as classes da ontologia BLO. O sistema permitirá que os dados dos eventos de coleta, ocorrência de espécies e medidas de organismos sejam fornecidos diretamente no padrão *Darwin Core*. A terceira camada (3) mostra que os dados fornecidos serão armazenados em um banco dados, juntamente com os metadados de cada subprojeto. Os dados de eventos de coleta e ocorrências de espécies serão exportados para o SiBBr através de comandos *SQL*. Todos os dados armazenados serão utilizados para gerar dados estatísticos apresentados em formato de gráficos e plotados em mapas de acordo com a necessidade de cada pesquisador. Após o seu desenvolvimento, o sistema será avaliado pelos pesquisadores do PELD-RLac executando casos de testes relacionados à inserção e consulta de dados, e à exportação desses dados para o SiBBr.

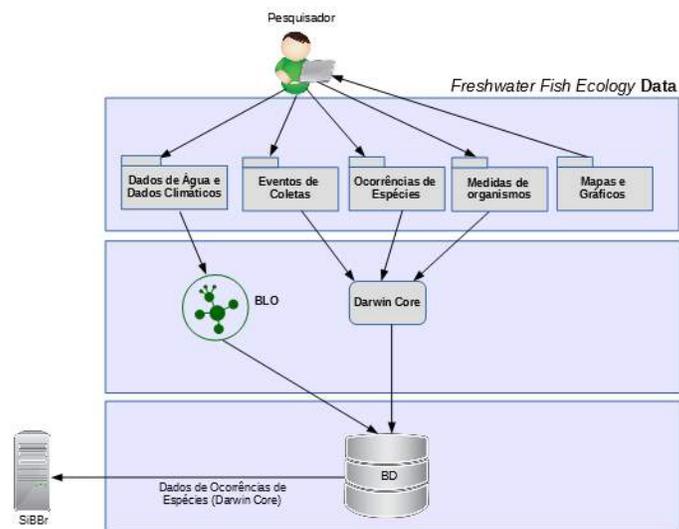


Figure 1. Arquitetura do Freshwater Fish Ecology Data

4. Trabalhos relacionados

O sistema de informação *speciesLink* (Canhos 2005) permite integrar dados dos acervos de coleções biológicas do Estado de São Paulo, fazendo o mapeamento desses dados para o padrão Darwin Core. O *Biodiversity Data Digitizer* (BDD) (Cartolano et al. 2010) é uma ferramenta projetada para facilitar a digitalização, manipulação e publicação de dados sobre a biodiversidade com base no padrão Darwin Core. O BDD possui módulos para interações entre espécimes, monitoramento, análises estatísticas, visualização de dados e sincronização de dados entre o seu banco de dados e planilhas eletrônicas. O *Brazilian Marine Biodiversity* (BaMBa) é uma base de dados integrada sobre a biodiversidade marinha brasileira utilizando os padrões EML e Darwin Core (Meirelles et al. 2015). BaMBa permite a inserção de dados marinhos, através do upload de planilhas a partir de um modelo fornecido, e a inserção de variáveis ambientais.

Esses trabalhos apresentam estratégias diferentes para padronização de dados da biodiversidade no padrão *Darwin Core*. BDD e BaMBa são os que mais se relacionam

com o trabalho proposto neste artigo. O primeiro por oferecer uma interface onde o pesquisador insere dados brutos no padrão *Darwin Core*. O segundo por apresentar uma estratégia para a integração com o SiBBr e outros sistemas. Porém, esses trabalhos não tratam o gerenciamento de dados de peixes de água doce, que é de extrema relevância para o estudo da dinâmica da ocupação no complexo de lagoas do Parque Nacional da Restinga de Jurubatiba realizado pelos pesquisadores do PELD-RLac.

5. Conclusão

O sistema facilitará a padronização dos dados, permitirá a integração com o SiBBr de forma automatizada e garantirá aos pesquisadores de Ictiologia do PELD-RELaC a preservação e a disponibilização desses dados. Assim, este trabalho poderá iniciar um processo de mudança de cultura entre os pesquisadores do Núcleo de Pesquisas Socioambiental de Macaé (NUPEM/UFRJ), onde estão alocados 90% dos pesquisadores do sítio PELD-RELaC. Essa mudança diz respeito às rotinas que envolvem o arquivamento e a disponibilização individual dos dados para uma cultura de integração, disponibilização e reutilização desses dados, independente do objetivo das coletas.

6. Referências

- Canhos, V. P. Sistema de Informação Distribuído para Coleções Biológicas: A Integração do Species Analyst e SinBiota.n. 2001/02175-5, p.2-51, 2005. <http://splink.cria.org.br/docs/outubro2005.pdf>.
- Cartolano, E. A.; Saraiva, A. M.; Veiga, A. K.; et al. Biodiversity Data Digitizer, 2010. Woods Hole, USA.
- CNPq (2016). “Política de Dados do Programa de Pesquisas Ecológicas de Longa Duração – PELD” (2016). http://www.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_ED/10157/4621110.
- Darwin Core Task Group, Biodiversity Information Standards (TDWG) (2009). “Darwin Core”. <http://rs.tdwg.org/dwc/>
- Esteves, F. D. A.; Petry, A. C. (2008). “Limnologia Em Pesquisas Ecológicas De Longa Duração: A Experiência nas Lagoas Costeiras do Norte Fluminense (PELD-SÍTIO 5)”. p. 1–11.
- Gadelha Jr., L. M. R., Guimaraes, P., Moura, A. M. C., Drucker, D., Dalcin, E., Gall, G., Tavares Jr, J., Palazzi, D. C., Poltosi, M., Porto, F., Moura, F., Leo, W. V. (2014). “SiBBr: Uma Infraestrutura para Coleta, Integração e Análise de Dados sobre a Biodiversidade Brasileira”. VIII Brazilian e-Science Workshop (BRESCI 2014). Proc. XXXIV Congress of the Brazilian Computer Society, 2014.
- Meirelles, Pedro Milet, et al. “BaMBA: towards the integrated management of Brazilian marine environmental data.” *Database 2015 – Journal of Biological Database and Curation* (2015): bav088.
- Rocha, C. F. D. Da; Assis Esteves, F. de; Scarano, F. R. (2004) “Pesquisas de longa duração na Restinga de Jurubatiba: ecologia, história natural e conservação”. RiMa.
- Souza, A. N. ; Medeiros, A. P. (2015). “BLO: Batata Lake (Oriximiná/PA) Application Ontology”. Seminário de Pesquisa em Ontologias do Brasil (ONTOBRAS), São Paulo. v. 1442.