

Gestão de dados da Biodiversidade: aplicação do padrão de metadados Darwin Core

Filipi Miranda Soares^a, Raíssa Yuri Hamanaka^b e Benildes Coura Moreira dos Santos Maculan^c

Resumo: No contexto da gestão de dados de pesquisa e da necessidade de padronização do armazenamento e tratamento dos mesmos para posterior reutilização, torna-se imprescindível o uso de metadados para a representação dos dados. Devido a importância do uso de metadados para a recuperação de documentos, o estudo objetivou analisar a aplicação do padrão de metadados Darwin Core (DwC) em um registro de ocorrência do Portal da Biodiversidade. Para avaliar a aplicação do DwC pelo Portal da Biodiversidade, foi feita uma correlação entre os metadados do DwC com os metadados do registro de ocorrência do repositório, buscando-se entender o significado de cada metadado e se eles correspondem de fato aos metadados do DwC, uma vez que a documentação do Portal da Biodiversidade apresenta o DwC como padrão de metadados adotado. Dos 34 campos do registro de ocorrências analisadas, a maioria tinha correspondência com os campos do padrão de metadados. Houve casos de duplicação de campos no repositório da biodiversidade, da criação de campos no repositório que não existiam no DwC e da existência de campo inadequado no repositório. A partir do exemplo do Portal da Biodiversidade é possível delinear a importância da curadoria digital, ao procurar agrupar conjuntos de dados semelhantes e estruturá-los de forma a possibilitar o reúso dos mesmos.

Palavras-chave: Gestão de dados de pesquisa. Metadados. Darwin Core.

Biodiversity data management: applying the Darwin Core metadata standard

Abstract: In the context of the management of research data and the need to standardize the storage and treatment of them for later reuse, it is essential to use metadata to represent the data. Due to the importance of using metadata for document retrieval, the study aimed to analyze the application of the Darwin Core (DwC) metadata standard in a Portal da Biodiversidade occurrence record. To evaluate the application of the DwC by the Portal da Biodiversidade, a correlation was made between the metadata of the DwC with the metadata of the record of occurrence of the repository, trying to understand the meaning of each metadata and if they correspond in fact to the metadata of the DwC, since the documentation of the Portal da Biodiversidade presents DwC as the metadata standard adopted. Of the 34 occurrence record fields analyzed, most corresponded to the metadata standard fields. There have been cases of duplication of fields in the biodiversity repository, the creation of fields in the repository that did not exist in the DwC and the existence of an inadequate field in the repository. Based on the example of the Biodiversity Portal, it is possible to delineate the importance of digital curatorship by seeking to group similar data sets and structure them in a way that allows them to be reused.

Keyword: Research data management. RDM. Metadata. Darwin Core.

-
- a Universidade Federal de Minas Gerais (UFMG). E-mail: filipivp2011@gmail.com. ORCID: <https://orcid.org/0000-0002-0674-7960>. Currículo: <http://lattes.cnpq.br/1290270601346793>
- b Universidade Federal de Minas Gerais (UFMG). E-mail: raissa0201@gmail.com. ORCID: <https://orcid.org/0000-0001-9516-5825>. Currículo: <http://lattes.cnpq.br/7188479615806578>
- c Universidade Federal de Minas Gerais (UFMG). E-mail: benildes@gmail.com. ORCID: <https://orcid.org/0000-0003-4303-9071>. Currículo: <http://lattes.cnpq.br/5336218259257800>

1 Introdução

No atual contexto do fenômeno Big Data, em que são gerados grandes volumes de dados não-estruturados em meio digital, torna-se de extrema importância a questão da organização do dado para sua posterior recuperação.

Inseridos nesse contexto estão os pesquisadores, produzindo um volume grande de dados científicos, o que impacta na definição das atuais fontes de informação

(para além do periódico científico) e no compartilhamento de dados, e, em consequência, na comunicação científica. Dessa forma, surgem desafios na gestão dos dados de pesquisa disponíveis em rede (SAYÃO; SALES, 2012).

Como solução para esses desafios têm origem o conceito de curadoria digital que “além de reduzir a duplicação de esforços na criação de dados de pesquisa [...] reforça o valor de longo prazo dos dados existentes quando os tornam disponíveis para a reutilização em novas pesquisas” (SAYÃO; SALES, 2012, p. 184).

Para garantir o reuso dos dados de pesquisa por pesquisadores, é necessário que os sistemas computacionais consigam acessá-los. Assim, tanto dados quanto algoritmos devem ser criados segundo os princípios de reusabilidade, encontrabilidade, acessibilidade e interoperabilidade (FAIR, 2016). Isso pode ser problematizado pelas seguintes questões: “se você disponibilizar seus dados para um cientista ou pesquisador que não esteve envolvido com seu projeto, seriam capazes de entendê-lo?”; “eles conseguiriam utilizar os dados efetiva e apropriadamente?”; “como criar, organizar, gerir, descrever, preservar e compartilhar dados, efetivamente?” (STRASSER et al., 2012, p. 1, tradução nossa).

A efetiva gestão de dados de pesquisa e sua posterior reutilização por demais pesquisadores, depende de um conjunto de práticas durante a coleta, o processamento e a análise dos mesmos. Essas práticas compõem o ciclo de vida dos dados (CVD), que é formado pelos seguintes elementos: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar e analisar (STRASSER et al., 2012). Para Sant’ana (2016), cada etapa ou elemento do CVD possui características em comum: privacidade, integração, qualidade, direito autoral, disseminação e preservação. Sayão e Sales (2012) dividem o CVD em ações: para todo o ciclo de vida (descrição e representação da informação, planejamento da preservação, participação e monitoramento e curadoria e preservação); sequenciais, que se repetem em cada etapa do ciclo (conceitualização, criação ou recebimento, avaliação e seleção, arquivamento, preservação, armazenamento, acesso, uso e reuso e transformação) e ocasionais (eliminação, reavaliação e migração).

O foco deste estudo foi o elemento “descrição” do CVD, definido como “dados que são descritos com precisão e minuciosamente usando os padrões de metadados apropriados” (STRASSER et al., 2012, p. 3, tradução nossa).

Conforme Costa (2017), há algumas iniciativas brasileiras em relação à gestão de dados de pesquisa, e dentre elas se destacam: o Programa FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) de Pesquisa em e-Science; o Portal da Biodiversidade; a Infraestrutura Nacional de Dados Espaciais no Brasil (INDE); as medidas de incentivo ao acesso aberto do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e o desenvolvimento do repositório de dados do Instituto de Energia Nuclear.

Para explicitar o elemento descrição (e seus fatores) do CVD, analisaram-se os metadados de um registro de ocorrência do repositório de dados do Portal da Biodiversidade em comparação com o padrão de metadados Darwin Core (DwC). DwC é um padrão internacional com base em taxon, composto por um conjunto de elementos (propriedades, atributos ou conceitos) desenvolvidos para padronizar o compartilhamento de informações sobre diversidade biológica.

2 Objetivo

Analisar a aplicação do padrão de metadados Darwin Core (DwC) em um registro de ocorrência do Portal da Biodiversidade⁵⁷.

3 Procedimentos metodológicos

O universo de pesquisa é todos os registros no Portal da Biodiversidade e como amostra foi selecionado o registro de ocorrência da espécie *Rhinella granulosa* (Spix, 1824), leigamente conhecida como sapo da caatinga. O procedimento adotado foi correlacionar os metadados do DwC com o significado atribuído a cada campo do registro e avaliar o seu uso. Assim, as categorias de análise selecionadas para análise foram os metadados adotados para descrição do registro de ocorrência, que foram correlacionadas com o conjunto de metadados do DwC.

4 Resultados

A análise determinou que o registro selecionado possui 34 metadados, apresentados em formato de folha de dados (CSV⁵⁸), baixado do Portal da Biodiversidade, que serão apresentados na ordem que aparecem no registro.

Os dois primeiros metadados do registro são <Nome da instituição> e <Sigla da instituição>. Eles representam um mesmo dado e possuem o mesmo equivalente no DwC, que é o metadado <institutionCode>. O primeiro campo de metadado está sem informações e o segundo apresenta a sigla ICMbio. Esses dois campos deveriam ser apenas um campo, pois significam a mesma coisa. Assim, seria necessário apenas definir o formato de entrada de dados: nome por extenso da instituição ou apenas a sigla.

Em seguida, aparece o metadado <Nome da base de dados>, que possui equivalente no DwC como <datasetName> e está preenchido corretamente. O campo seguinte foi denominado <Sigla da base de dados> e não possui equivalente no DwC. A melhor prática sugerida seria unir esses dois campos em apenas um e padronizar a entrada de dados: nome por extenso da base de dados ou apenas sigla.

Depois vem o campo <Responsável pelo registro> que possui o metadado equivalente no DwC <recordedBy>, com valor formatado de acordo com as recomendações do DwC para este campo.

Em sequência, são apresentados dois campos que equivalem a um mesmo campo no DwC, mas foram preenchidos com valores distintos: <Número do registro no portal> e <Número do registro na base de dados>, que equivalem semanticamente ao campo <catalogNumber> no DwC. O primeiro apresenta o valor numérico '1067665' e o segundo o valor "Nº Da Autorização/Licença Sisbio: 44832".

57 Disponível em: <<https://portaldabiodiversidade.icmbio.gov.br/portal>>. Acesso em: 30 set. 2018.

58 Comma-separated values é um formato de arquivo que apresenta dados tabelados.

Após, aparece o campo <Data do registro> que é ambíguo. Não é possível saber se o valor da data que preenche o campo é a data de registro na base de dados ou a data que o evento (ocorrência) aconteceu. No padrão DwC, o metadado para descrever a data de ocorrência é o <eventDate>, presente no registro como <Data do evento>. Quando baixado em formato CSV, o valor do campo <Data do registro> muda de '05/02/2017' para '06/02/2017', a mesma data presente no campo <Data do evento>. Não há um campo de data de registro no DwC. O campo com significado mais próximo seria <dcterms:modified>, que se refere à última data de modificação do registro documental da ocorrência na base de dados. Além disso, o DwC recomenda utilizar esquemas de codificação para preenchimento das datas, como a ISO 8601:2004(E), o que não foi aplicado no Portal da Biodiversidade. O valor do campo <Data do evento> apresenta outro problema: está definido como '06/02/2017 a 06/02/2017'. Se o evento aconteceu no dia seis de fevereiro e não foi inserido o intervalo específico de horas em que ele ocorreu, não é necessário inserir a data como intervalo de datas, uma vez que o evento aconteceu em apenas um dia.

O metadado seguinte, <Data de Carência>, não possui equivalente no DwC e não está claro ao que esta data se refere.

Depois, aparece o metadado <Nome científico>, que possui o equivalente <scientificName> no DwC. O valor desse campo no registro é 'Rhinella granulosa', que não está de acordo com as recomendações do DwC (DARWIN CORE TASK GROUP, 2015), que orienta indicar o nome científico completo, com autor e data. A melhor prática seria apresentar o nome 'Rhinella granulosa (Spix, 1824)'.
Em seguida vem <nome comum>, que possui o equivalente <vernacularName> no DwC, e está preenchido em conformidade com as melhores práticas recomendadas pelo padrão.

O campo seguinte, <Nome científico na base de dados>, não deveria existir. O nome científico da espécie é um identificador único, ou seja, não pode variar. Logo, é desnecessário que haja dois campos com o mesmo valor no registro de ocorrência.

O campo seguinte, <Nome científico na base de dados>, não deveria existir. O nome científico da espécie é um identificador único, ou seja, não pode variar. Logo, é desnecessário que haja dois campos com o mesmo valor no registro de ocorrência.

Logo em seguida aparecem os metadados <Nível taxonômico>, <Número de indivíduos>, <Reino>, <Filo>, <Classe>, <Ordem>, <Família> e <Gênero>, que têm valores que atendem às orientações de melhores práticas do DwC, e possuem como equivalentes, respectivamente: <taxonRank>, <individualCount>, <kingdom>, <phylum>, <class>, <order>, <family> e <genus>. Depois vem o metadado <Espécie>, que repete o mesmo valor de dois outros campos do registro de ocorrência, portanto, é redundante sua presença no registro. A sugestão seria substituir esse metadado por <specificEpithet>, que representa o epíteto⁵⁹ da espécie no nome científico.

A seguir aparecem os metadados <Estado de conservação>, <Categoria de Ameaça> e <Status de Sensibilidade>, que ainda não têm equivalência no DwC. Entretanto, uma extensão⁶⁰ está sendo desenvolvida para representar informações de estado de conservação. Após, aparecem metadados com informações de localização: <Localidade>, <País>, <Estado/Província>, <Município>, <Latitude> e <Longitude>, que possuem equivalentes no DwC, respectivamente:

59 Designa uma espécie ou subespécie diferente dentro de um mesmo gênero.

60 Disponível em: <<https://tools.gbif.org/dwca-validator/extension.do?id=http://purl.org/plic/terms/3.2.1/ThreatStatus>>. Acesso em: 25 set. 2018.

<locality>, <country>, <stateProvince>, <municipality>, <decimalLatitude> e <decimalLongitude>, e têm valores conforme as orientações do padrão.

Depois aparecem os campos <Outras informações da localidade> e <Jurisdição>, que não possuem equivalentes no DwC.

Por fim, o metadado <Destino do Material> tem o equivalente no DwC <MaterialSample>.

5 Considerações finais

Conforme o objetivo do estudo, foram analisados 34 campos de um registro de ocorrência do Portal da Biodiversidade em correlação com os campos do padrão de metadados DwC. Na maioria das análises, em cerca de 65% dos casos, houve correspondência do campo do registro com o campo do padrão de metadados. Entretanto, houve casos de duplicação de campos no repositório da biodiversidade, valores inadequados, criação de campos que ainda não existem no DwC e existência de campo inadequado no repositório.

O Portal da Biodiversidade reuni registros de ocorrência de nove bases de dados de biodiversidade brasileira, sendo uma iniciativa de gestão de dados que desde 2016 está integrado ao Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr), que está ligado ao Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC).

Essa iniciativa é um exemplo que permite perceber a importância da curadoria digital, que procura agrupar conjuntos de dados semelhantes e estruturá-los de maneira a possibilitar o seu reúso. Para tanto, é necessário manter interoperabilidade entre as diferentes bases, o que pode ser alcançado por meio do uso de padrões de metadados. Considera-se que na implementação do CVD, os metadados são essenciais, pois possibilitam a padronização da representação e/ou codificação dos dados para sua futura recuperação.

Acredita-se que a padronização dos dados coletados e armazenados em repositórios permite a democratização dos dados ao garantir o acesso efetivo aos mesmos.

Referências

COSTA, Máira Murrieta. **Diretrizes para uma política de gestão de dados científicos no Brasil**. 2017. 288 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, DF, 2017.

DARWIN CORE TASK GROUP. Biodiversity Information Standards. Darwin Core. 2015. Available from: <<http://rs.tdwg.org/dwc>>. Accessed on: 8 Sept. 2018.

FAIR principles for data stewardship. **Nature Genetics**, London, v. 48, n. 4, p. 343, Apr. 2016.

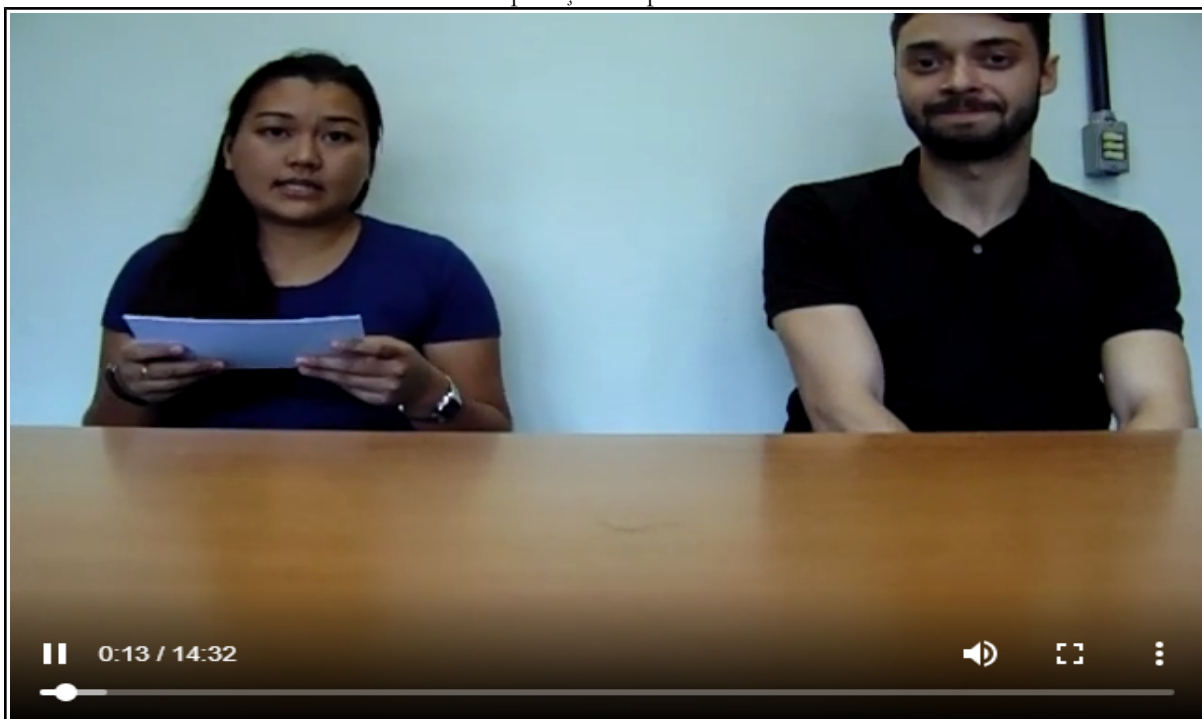
SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016.

SAYÃO, Luis Fernando; SALES, Luana Faria. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.: Est.**, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

STRASSER, Carly et al. Primer on data management: what you always wanted to know: but were afraid to ask. **DataONE**, Oakland, CA, p. 1-11, fev. 2012.

Vídeo da apresentação

Título: Gestão de dados da Biodiversidade: aplicação do padrão de metadados Darwin Core.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/video.php?id=12

Transcrição da apresentação

Olá nós vamos apresentar um artigo que submetemos ao 2º Workshop de Informação, Dados e Tecnologia (WIDAT), os autores do artigo foram eu Raíssa e Filipi e a orientadora Benildes nós dois somos mestrando do programa de gestão e organização do conhecimento da escola de Ciência da Informação da UFMG – a Benildes é orientadora do Filipi é professora também na escola do Centro de Informação tanto na graduação quanto na pós.

O título do nosso artigo é “Gestão de dados da Biodiversidade: aplicação do padrão de metadados Darwin Core”.

Nós vamos apresentar a estrutura do artigo, vou começar com referencial teórico o objetivo do artigo que ele vai explicar a metodologia e os resultados e depois eu vou apresentar as considerações finais.

Começamos contextualizando e falando um pouco do Big Data, esta grande quantidade de dados não estruturados.

É grande o volume de dados não estruturados e dessa influência esse grande volume de dados sempre existiu, mas agora que na ciência da informação está surgindo essa preocupação de lidar com esses dados e principalmente os dados científicos, os dados que são coletados pelos pesquisadores, e os desafios seriam no armazenamento desses dados pra posterior compartilhamento dos dados e para gerar de fato uma comunicação científica efetiva.

Então, pensando nesses dados científicos que devem ser armazenadas surgiu esse conceito de curadoria digital e essa curadoria digital gira em torno de evitar a criação de dados duplicados e aí também levou a questão do armazenamento de dados de forma padronizada para permitir o reúso desses dados então a questão do reúso de dados é a palavra-chave da curadoria.

Quando se fala em reúso de dados a gente pensa basicamente em duas características: como é que os dados serão reutilizadas pensando em disponibilizar esses dados em formatos abertos com uma série de critérios que estabeleça garantia da qualidade da disponibilização dos dados e também pensar na descrição de forma lógica e padronizada desses dados coletados pelos pesquisadores para que alguém que não tenha participado da coleta dos dados consiga entender e efetivamente utilizar esses dados.

Então, para permitir esse reúso dos dados científicos podemos pensar em gestão de dados, no nosso artigo essa parte do referencial teórico nós utilizamos basicamente três autores Strasser, Sayão e Sales e quando eles pensam no reúso de dados eles pensam no ciclo de vida dos dados, ou seja, é aquele conjunto de dados que se precisa ser submetido a uma série de etapas para que eles sejam armazenados e possivelmente compartilhados de forma efetiva, só que de acordo com o objetivo do nosso artigo nós focamos na etapa de descrição dos dados que eles definem como uma descrição detalhada dos dados através de padrões de metadados apropriados, então antes falar um pouco do objetivo do artigo eu queria citar a tese de Costa (2017) em que ela aborda as principais iniciativas de gestão de dados, aqui no Brasil uma delas é o Portal da Biodiversidade que é onde a pesquisa foi realizada, foi selecionado e aí chegamos no objetivo do artigo que é analisar os campos de um registro de ocorrência do portal da biodiversidade e comparar ele com campos do padrão da biologia, ou no caso, no Darwin Core.

Vou falar um pouquinho da metodologia que aplicamos e dos casos que a gente obteve. Primeiramente vou definir os conceitos o que é o portal da biodiversidade e o que é o padrão de metadados Darwin Core que a gente aplicou no portal da biodiversidade. Ele é um repositório de dados científicos sobre registros de ocorrência de espécies brasileiras que é mantido pelo incêndio de bio, e esse repositório aplica o padrão de metadados Darwin Core para descrever esses dados de registros de ocorrências e o que é o padrão de metadados.

Ele foi desenvolvido pelo Darwin Core que é um grupo da GBIF que é uma fundação internacional que publica informações sobre a biodiversidade e desenvolve uma série de ferramentas para o compartilhamento de informações sobre biodiversidade e então Darwin Core foi desenvolvido pelo compartilhamento de informação sobre espécimes e é importante diferenciar o conceito de espécie de espécime porque o conceito de espécie ele é mais amplo.

O que é uma espécie? É grupo de indivíduos uma população com as mesmas características filogenéticas que representa uma mesma espécie, então o conceito de espécie é construído com base nessas observações a essas características genéticas dos indivíduos que são categorizados como uma mesma espécie.

O espécime é uma amostra coletada dessa espécie em questão na atitude de análise então o biólogo vai a campo, ele captura um indivíduo, não lixo por exemplo, ele leva esse bicho para laboratório para analisar e fazer a inscrição taxonômica, a análise genética e todos os outros procedimentos possíveis dentro da biologia – os procedimentos de pesquisa.

Então esse bicho coletado é um espécime, mas esse espécime também pode ser a folha de uma planta, pode ser uma amostra de solo, enfim um espécime é um conceito de amostragem é uma mostra dentro da população. Um indivíduo é coletado dentro da população de uma espécie.

Bom, o Darwin Core ele trabalha com a representação da informação desses registros de espécimes que são os registros de ocorrência, então o biólogo foi a campo ele coletou um organismo ou ele fotografou um bicho, alguma coisa assim, isso é um registro de ocorrência. O que é esse registro de ocorrência? Tem basicamente a descrição taxonômica da espécie, qual foi o animal observado naquele local, quando ocorreu essa observação esse evento que como é chamado na biologia os dados desse local os dados do período de tempo que aconteceu a esse evento e outras relações típicas de dentro do contexto da biologia como habitat que esse bicho habita, enfim e todas os outros metadados conferidas dentro do Darwin Core levando em conta esses outros conceitos.

Então a metodologia que adotamos foi selecionar um registro dentro do portal da biodiversidade especificamente o registro da espécie que nela granulosa que é vulgarmente conhecido como o sapo da caatinga e a gente analisou campo a campo desse registro são 34 metadados, então analisamos o conteúdo de cada um desses metadados para ver se eles são equivalentes ao que é recomendado no glossário do Darwin Core e nas recomendações do preenchimento desses metadados.

Então o que a gente analisou dentre esses 34 campos no registro da biodiversidade? Existem alguns erros como por exemplo a questão de duplicatas de metadados. O primeiro exemplo que abordamos no artigo é o nome da base de dados e sigla da base de dados no glossário onde tem as orientações de preenchimento do quadro de metadados do Darwin Core, ele orienta que seja

adotado só um valor, esse é um campo único, então nome da base de dados que é dataset name que está no corpo do Darwin Core. Trata-se apenas de um metadado, então a sigla da base de dados ou nome da base de dados só podem entrar nesse campo, mas é preciso especificar qual o valor que vai ser adotado como padrão ou a sigla da base de dados, no caso do portal da biodiversidade esse campo aparece publicado tanto por extenso na base de dados quanto com a sigla e isso é considerado um erro no portal da biodiversidade.

Outro problema que foi detectado foi em relação ao preenchimento dos campos metadados porque o Darwin Core ele recomenda que seja seguida uma série de protocolos e vocabulários controlados e o preenchimento desses campos e formatos também, por exemplo, as datas elas devem seguir um padrão ISO e no caso do portal da biodiversidade eles não seguem esse padrão, então existem diversos formatos de entrada de data nos campos desses registros.

Um outro problema também envolvendo as datas é referente a data do evento, então um evento de registro de ocorrência da espécie ele tem início preciso específico, com dia da ocorrência a hora minuto segundo, se possível então a data de início e término do evento ela tem que ser mais específico possível e no caso do portal da biodiversidade esse dado não é apresentado dessa forma eles apresentam simplesmente o dia de início e o dia do término do evento, muitas vezes é o mesmo dia porque os eventos costumam ser curtos por efeito de observação ou de coleta de espécies geralmente duram bons minutos ou horas, mas geralmente não duram mais de um dia. Então isso é um uma outra inconsciência que observamos no padrão.

Por último, outro erro que foi detectado no registro do portal da biodiversidade foi que existem metadados que não estão presentes no núcleo do Darwin Core que é o núcleo central de metadados, no corpo de 176 metadados isso dificulta o compartilhamento de dados, porque existem dados nesse registro que não são compatíveis com os metadados desse conjunto do Darwin Core. Esses dados não podem ser compartilhados, então eles extrapolam o núcleo desse conjunto dos termos do Darwin Core.

E agora então eu vou passar a palavra para Raíssa que fará as considerações finais.

Nas considerações finais fizemos um resumo desses resultados que o Filipi apontou. Houve 65% de correção do registro de ocorrência com padrão de metadados Darwin Core.

Além disso nas considerações finais a gente realça a importância de iniciativas como o portal de biodiversidade, porque ele reuniu o registro de ocorrência da espécie de nove bases de dados brasileiras em uma só e além disso ele levou em consideração dois pontos importantes da curadoria digital, a questão de disponibilizar esse registro de ocorrência em um formato aberto e também a questão de dessas informações estarem devidamente descritas porque houve 65% de equivalência com padrão de metadados.

Então essa iniciativa que incentiva de fato o reuso daqueles dados, principalmente pelo fato de agrupar conjuntos de dados com temáticas semelhantes que estavam dispersos em meio digital em um único repositório nós enfatizamos muito essa questão, ou seja, a importância desse tipo de iniciativa em gestão de dados e principalmente a importância disso na democratização no acesso aos dados e para facilitar o uso efetivo dos dados e consequentemente facilitar a construção da ciência mesmo.

Agradeço ao professor Guilherme pelas contribuições em sala de aula e também pela oportunidade de participar no WIDAT.

Slides da apresentação

Título: Gestão de dados da Biodiversidade: aplicação do padrão de metadados Darwin Core.

GESTÃO DE DADOS DA BIODIVERSIDADE: aplicação do padrão de metadados Darwin Core

Filipi Miranda Soares, Raíssa Yuri
Hamanaka, Benildes C. Moreira dos
Santos Maculan

Disponível em: http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=12