

Uma Proposta de Esquema de Dados de Relacionamento Entre Espécimes

Etienne Américo Cartolano Jr.

Escola Politécnica da Universidade de São Paulo, Depto. Computação e Sistema Digitais,
São Paulo, Brasil, 05508-900
cartolano@usp.br

Antonio Mauro Saraiva

Escola Politécnica da Universidade de São Paulo, Depto. Computação e Sistema Digitais,
São Paulo, Brasil, 05508-900
amsaraiv@usp.br

Pedro Luiz Pizzigatti Correa

Escola Politécnica da Universidade de São Paulo, Depto. Computação e Sistema Digitais,
São Paulo, Brasil, 05508-900
pedro.correia@usp.br

Tereza Cristina Giannini

Instituto de Biociências da Universidade de São Paulo, Depto. de Ecologia,
São Paulo, Brasil, 05508-900
giannini@usp.br

Renato De Giovanni

Centro de Referência em Informação Ambiental
Campinas, Brasil, 13084-791
renato@cria.org.br

Abstract

Pollination is considered one of the most important processes for biodiversity conservation. Studies show that successful pollination can improve the amount and the quality of plant fecundation and fruit production, stimulating its use in environmental programs and proposals of sustainable agriculture. However, the success of these actions is based on the knowledge on pollinators, their preservation and interaction with the environment. New initiatives have been created to facilitate and to stimulate the dissemination of this knowledge. The Inter-American Biodiversity Information Network – IABIN and the WebBee are initiatives that support pollinator's data and one of their main features is information about the relationship between pollinators and plants. This paper discusses the representation of plant-pollinator data and analyses the feasibility of doing this with the existing schemas for exchanging biodiversity data. A case study is presented that analyses a use case of a biodiversity portal where a user searches for relationship data.

Keywords: XML, biodiversity, schemas, relationship, species

Resumo

A polinização é considerada um dos processos mais importantes para a conservação da biodiversidade. Estudos mostram que a polinização bem conduzida pode melhorar a quantidade e a qualidade da fecundação das flores e da produção de frutos, incentivando o seu uso em programas ambientais e nas propostas de agricultura sustentável. No entanto, o sucesso dessas ações está diretamente ligado ao conhecimento sobre os agentes polinizadores, a sua preservação e

interação com o meio ambiente. Novas iniciativas têm sido criadas para facilitar e incentivar a disseminação desse conhecimento. A Inter-American Biodiversity Information Network - IABIN e a WebBee são iniciativas que contemplam informações sobre polinizadores e que têm entre seus principais serviços os dados de relacionamento entre polinizadores e plantas. Este trabalho discute a representação de dados sobre relacionamento polinizador-plantas e analisa a viabilidade do seu uso com os esquemas atuais para troca de dados de biodiversidade. Um estudo de caso é apresentado em que são analisados casos de uso de um portal de informações sobre biodiversidade onde usuários pesquisam dados de relacionamento.

Palavras-chave: XML, biodiversidade, esquemas, relacionamento, espécies

1. Introdução

Entre os vários organismos ameaçados destacam-se os polinizadores, que têm influência direta na manutenção dos fragmentos florestais e na agricultura [1]. A polinização tem sido considerada um dos mais importantes serviços de ecossistemas [2] e o uso e conservação dos polinizadores foram enfatizados pela Convention on Biological Diversity [<http://www.biodiv.org/>] que aprovou a São Paulo Declaration on Pollinators, documento que identifica as causas e conseqüências do declínio dos polinizadores [3]. No entanto, o sucesso dessas ações está diretamente ligado ao conhecimento sobre os agentes polinizadores, que inclui aspectos da biologia e da interação com o meio ambiente [4]. Novas iniciativas têm sido criadas para facilitar e incentivar a disseminação desse conhecimento, normalmente pulverizado entre centros de pesquisa e coleções particulares de pesquisadores. Estas iniciativas podem ter escopo regional ou global, mas em geral propõem-se a viabilizar e organizar de forma integrada a digitalização das informações sobre biodiversidade.

A WebBee [<http://www.webbee.org.br>] é uma rede brasileira que tem por objetivo ser um local de integração das informações sobre biologia e criação das diversas espécies de abelhas nativas e sobre seus habitats, englobando textos, imagens e vídeos voltados para diversos públicos [5]. Este sistema oferece vários serviços em forma de páginas Internet, entre eles o de relacionamento de espécies de abelhas e plantas (Figura 1). Atualmente o sistema trabalha de forma isolada, ou seja, não provê e nem recebe dados de outros sistemas de biodiversidade. Estudos em andamento propõem que a WebBee torne-se uma iniciativa regional brasileira para prover dados de espécies de abelhas para outros sistemas de biodiversidade, em especial para a Rede Interamericana de Informação sobre Biodiversidade (Inter-American Biodiversity Information Network) – IABIN.

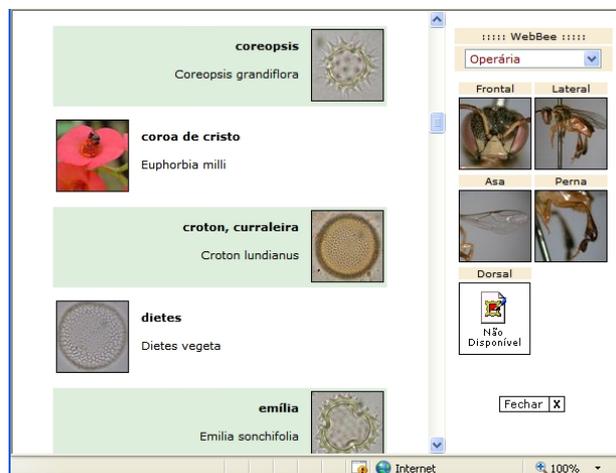


Figura 1 - Ficha de relacionamento entre espécie de abelha e plantas do sistema WebBee.

A IABIN [<http://www.iabin.net>] é um fórum criado para promover coordenação e colaboração técnica entre os países da América na coleta, compartilhamento e uso da informação sobre biodiversidade. Desta forma, visa-se contribuir para a tomada de decisão no manejo e na conservação dos recursos naturais, a educação, e o desenvolvimento sustentável da região. Entre os seus principais objetivos está a construção de uma infra-estrutura sob a plataforma Internet para a troca de dados sobre biodiversidade das Américas. A construção desta rede é feita por áreas temáticas, como Espécies e Espécimes, Áreas Protegidas, Ecossistemas e Espécies Invasoras. Cada área possui um grupo de trabalho responsável. Os grupos colaboram entre si e coordenam a transferência tecnológica para as bases regionais. Uma das redes é a Rede temática de polinizadores (Pollinators Thematic Network) - PTN.

O portal Internet desta rede temática de polinizadores pretende prover dados sobre espécimes de polinizadores, especialistas em polinizadores e polinização, literatura especializada e sobre relacionamento entre polinizadores e plantas. Parte desses dados será obtida com base em uma arquitetura distribuída baseada em provedores, o que requer a utilização de padrões.

Este trabalho aborda a questão da padronização de um esquema de dados para identificar um relacionamento entre espécies ou espécimes. O texto apresenta uma definição de relacionamento simples entre espécies ou espécimes e avalia a sua utilização sobre os esquemas de dados Darwin Core2, ABCD e Plinian Core.

Como estudo de caso, este esquema será usado para identificar o relacionamento entre polinizadores e plantas, onde a WebBee será o provedor dos dados. O consumidor dos dados será a ferramenta de portal do GBIF que será adaptada para uso na IABIN-PTN.

2. Padrões de Esquemas de Dados de Biodiversidade

Os sistemas de informação sobre biodiversidade baseiam-se tipicamente numa arquitetura distribuída na qual um portal serve de ponto de acesso unificado para informação fornecida por diversos provedores localizados em diferentes instituições - tipicamente museus, instituições de ensino e pesquisa que abrigam coleções biológicas [6]. Para permitir a troca de informação entre os sistemas heterogêneos foram definidos esquemas de dados que padronizam os itens de dados a serem trocados e sua forma de representação.

Esses padrões foram propostos e suportados por instituições e iniciativas como Biodiversity Information Standards – TDWG [<http://www.tdwg.org>] e o Global Biodiversity Information Facility, GBIF [<http://www.gbif.org>], e visaram, num primeiro momento, a digitalização e a disseminação dos dados sobre coleções biológicas. Tais padrões estão portanto relacionados com dados de espécimes. Os principais padrões adotados atualmente são o Darwin Core e o ABCD (Access to Biological Collection Data). Outro esquema de dados adotado pela IABIN, introduzido pela Rede Temática de Espécies e Espécimes da (SSTN), é o Plinian Core [<http://www.pliniancore.org>], sendo este relacionado com a troca de informações sobre espécies.

Esses esquemas podem ser analisados quanto ao suporte natural para troca de dados de relacionamento entre espécies ou espécimes, com ou sem adaptação de uso de elementos, e quanto à sua extensão.

- O **Darwin Core2 v1.4 (DwC)** é um conjunto simples de elementos identificados por marcadores, ou “tags”, que permitem estruturar dados de registros de espécimes ou de observações para serem compartilhados na Internet como um documento XML. O esquema DwC contém informações sobre a taxonomia do espécime, da localização geográfica da ocorrência e de elementos de referência, mas que devido a sua característica de generalidade, não suportam relacionamentos entre espécimes. No entanto, o DwC possibilita a extensão do seu esquema para usos específicos, não tratados naturalmente pelo esquema padrão. Neste processo o DwC é importado e acrescido de novos elementos, que para serem considerados extensões oficiais devem ser submetidos ao TDWG para validação. Existem algumas extensões aprovadas e em fase de validação: *geospacial*, *curatorial*, *palentology*, *monitoring*, *microbial*, *botanical*, entre outras.
- O **Access to Biological Collection Data v2.06 (ABCD)** é um esquema complexo para acesso e troca de dados sobre espécimes e observações. O ABCD é mais extenso que o Darwin Core e a sua estrutura é baseada em uma hierarquia que suporta a repetição de elementos e tipos complexos e organiza os metadados associados, atingindo aproximadamente 700 elementos. O ABCD prevê a associação binária entre dois espécimes dentro de seu esquema, desde que estejam em registros ABCD. No entanto, esta associação não permite discriminar o tipo de relacionamento.
- O **Plinian Core v.2.0 (PliC)** é um conjunto de conceitos que definem atributos básicos necessários para integrar e recuperar informações sobre espécies. Ele também é mais extenso que o DwC e carrega dados como: identificação do registro, taxonomia da espécie, ciclo de vida, comportamento, conservação e demografia, distribuição, uso, referências e até folclore. O PliC, entre outros esquemas de dados de espécies, é base para a criação de um novo modelo de dados de espécies organizado pelo grupo Species Profile Model (SPM) do TDWG. Os trabalhos estão em fase inicial, mas consideram desde já os principais elementos do Plinian Core v.2.0, entre eles o suporte ao relacionamento entre espécies, feito pelo PliC da seguinte maneira:

<Interactions>	→ Mutual or reciprocal actions or influences (<i>ComplexType</i>).
<TextInteractions/>	→ List of interactions (<i>NormalString</i>).
<Interaction>	→ Interactions in structured format (<i>ComplexType</i>).

<code><InteractionType/></code>	→ String with the type of interaction (<i>NormalString</i>).
<code><InteractionSpecies/></code>	→ Species involved in the interaction (<i>Type gisin: GUIDT</i>).
<code><InteractionComments/></code>	→ Notes about the interactions (<i>NormalString</i>).
<code><References/></code>	→ Bibliographic references (<i>Type tcs: ReferenceType</i>).
<code></Interactions></code>	
<code></Interactions></code>	

O esquema Plinian Core v.2.0 oferece um suporte natural e estruturado ao relacionamento entre espécies. É necessário verificar se ele atende aos requisitos dos principais serviços a serem oferecidos pela IABIN-PNT e suportados pela WebBee. Entretanto, os esquemas de dados de espécimes, DwC e ABCD, não oferecem um suporte adequado, e torna-se necessário um estudo sobre a adaptação destes esquemas, ou a criação de uma nova referência para a troca de dados de relacionamento entre espécimes, encontrados em coleções ou anotados em observações.

3. Definindo um Relacionamento entre Espécimes

Para determinar o conjunto de informações que caracterizam o relacionamento entre um espécime de abelha e os diversos espécimes de plantas, pesquisadores do Laboratório de Abelhas do Instituto de Biociências da Universidade de São Paulo (Beelab) analisaram as interações observadas em suas pesquisas. Foram descritos os seguintes atributos, geralmente observados:

- Nidificação – a planta é utilizada para estabelecimento de ninho.
- Polinização – a planta é polinizada durante a visita.
- Forrageamento – visita à planta para:
 - Pólen – coleta de pólen.
 - Néctar – coleta de néctar.
 - Óleo – coleta de óleo.
 - Resina – coleta de resina.
- Predação – predação de partes da planta.
- Dormitório – a planta é utilizada como dormitório por machos.
- Patrulha para acasalamento – a planta é patrulhada por machos na busca de fêmeas para acasalar.
- Visitante – quando não houver detalhe sobre o que ocorre durante a visita.

Esta análise feita pelo Beelab é em grande parte específica para um grupo de polinizadores, as abelhas, e claramente não pode ser generalizada para outros grupos. Existem interações típicas entre determinadas espécies, e tentar mapear essas especificidades em um único esquema de dados de uso geral tende a ser uma tarefa complexa e pouco eficiente. No mapeamento do relacionamento entre espécimes, um espécime deve apresentar uma ou várias interações com outros espécimes, provavelmente diferentes entre si. Com uma descrição específica teríamos um modelo diferente para cada tipo de interação, tornando o esquema complexo o bastante para ser pouco eficiente. A solução a ser adotada deve determinar um modelo mais simples de interação que contemple o relacionamento entre quaisquer espécies: abelhas, outros polinizadores, plantas, etc., independente da suas especificidades. Para determinar tal modelo, partimos para uma definição matemática de relação entre espécimes, que pode ter as seguintes configurações:

- Um espécime para um espécime – 1x1;
- Um espécime para muitos espécimes – 1xN;
- Muitos espécimes para um espécime – Nx1;
- Muitos espécimes para muitos espécimes – NxN;

Para simplificar a relação podemos adicionar um atributo de direção às relações, dando um sentido para interpretação dos dados. Com este artifício, as relações “1xN” e “Nx1” se equivalem em sua interpretação. Outra possível simplificação é a adoção de uma chave única, ou seja, apenas um espécime a ser relacionado. Neste caso, o relacionamento “NxN” pode ser substituído por “N” instâncias do relacionamento “1xN”. Sendo assim podemos mapear as relações entre espécimes com a seguinte expressão:

- Um espécime [relação] [sentido] um ou muitos espécime(s) – 1x1..N;

Onde [relação] indica o tipo de relacionamento entre os espécimes: predação, polinização, simbiose, forrageamento, competição e outras, que podem ou não ser específicas de uma interação e o atributo [sentido] dá um sentido à interpretação do relacionamento entre os espécimes.

Podemos então criar um esquema XML básico a partir desta expressão:

```

<specimensschema>
  <specimen>
    <taxonomy></taxonomy> → Identificação da espécie.
    <ocurrence></ocurrence> → Identificação única do espécime.
  </specimen>
  <interactions>
    <interaction>
      <relation></relation>
      <direction></direction>
      <references/>
      <specimen> → Espécime relacionado.
        <taxonomy> </taxonomy>
        <ocurrence> </ocurrence>
      </specimen>
    </interaction>
  </interactions>
</specimensschema>

```

4. Estudo de Caso – Portal IABIN

A ferramenta de portal desenvolvida pelo GBIF e disponível para uso público permite a busca por espécimes em coleções digitalizadas e publicadas por provedores cadastrados. A troca de dados feita na versão mais recente do portal utiliza padrões mantidos pelo TDWG, entre eles os esquemas Darwin Core e ABCD (Access to Biological Collection Data) e o protocolo TAPIR (TDWG Access Protocol for Information Retrieval), que suporta a troca de mensagens com os dois esquemas de dados. O Portal GBIF está sendo adotado pela Rede Temática de Polinizadores e pela Rede Temática de Espécies e Espécimes da IABIN, responsável pela adaptação para suporte ao esquema Plinian Core v.2.0, o que justifica sua escolha para esta análise. No caso da rede de polinizadores, um provável provedor de dados de relacionamento será o WebBee. A figura 2 ilustra a rede PTN com as alternativas de esquemas de dados de relacionamento.

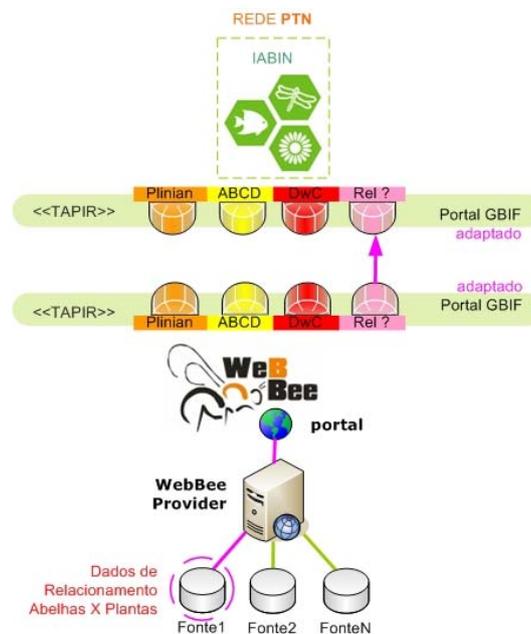


Figura 2 - Escopo do caso de uso de troca de dados de relacionamento.

Na arquitetura de portal adotada pela IABIN-PTN dois serviços distintos tratarão as informações de espécimes e espécies de polinizadores. Os dados de coleções e de observações já são tratados pelo portal GBIF, enquanto os dados de espécimes serão considerados apenas na ferramenta de busca de relacionamentos do portal em desenvolvimento.

Para analisar a usabilidade dos elementos de relacionamento do PliC e da expressão adotada “Um espécime [relação] [sentido] um ou muitos espécime(s) – 1x1..N” para os dados de relacionamento de espécimes, foram analisados casos de uso tomando como base o novo portal GBIF (<http://newportal.gbif.org>).

Podemos resumir a análise em dois casos:

- Caso de Uso 1: relacionamento a partir do detalhamento de uma ocorrência de espécime.
- Caso de Uso 2: relacionamento utilizando uma ferramenta de busca de relacionamentos.

Caso 1: No portal GBIF, um detalhamento de ocorrência de espécime, em coleção ou observação, ocorre quando um usuário seleciona um espécime a partir do nome científico ou popular, ou das ocorrências registradas em um determinado provedor ou nos provedores de um país. Neste caso de uso, selecionada a ocorrência, o portal exibiria uma ficha mais detalhada do espécime e no final, uma lista de relacionamentos. Cada item desta lista deveria ser uma ocorrência “Um espécime [relação] [sentido] um espécime [referência]” de relacionamento (Figura 3), ou seja, mostrando um relacionamento entre espécimes.

Occurrence Related Specimens

relation	direction	scientificName	institutionCode	collectionCode	catalogNumber	
[Competition]	[<-]	<i>ScientificName A</i>	Provider2	Coll21	1332	View
[Parasitism]	[<-]	<i>ScientificName L</i>	GBFI	Coll34	1233	View
[Parasitism]	[->]	<i>ScientificName T</i>	Reference1	Coll21	1332	View
[Predation]	[->]	<i>ScientificName E</i>	FAPESP	Coll34	1233	View
[Competition]	[<-]	<i>ScientificName P</i>	Reference2	Coll21	1332	View
[Predation]	[->]	<i>ScientificName B</i>	FAPESP	Coll34	1233	View

Figura 3 - Exemplo de relacionamentos nos detalhes de um espécime do Caso de Uso 1.

Neste caso de uso, o elemento “[relação]” em união com o elemento “[direção]” por vezes causa confusão na compreensão dos dados. Por exemplo, a relação “*Predation*” unida à direção “→” não identificava claramente em qual lado do relacionamento estava a presa. A solução encontrada foi utilizar nomes de relações já direcionadas: *ParasiteOf*, *ParasitedBy*, *InfectedBy*, *FlowerVisitedBy*, etc., em substituição ao elemento “[sentido]”. O elemento “[referência]” também foi descartado por ser redundante com o elemento de referência do esquema DwC original, pois esta referência é a fonte de todas as informações dos dois espécimes relacionados.

Podemos então adequar a forma proposta para:

```

<specimensschema>
  <specimen>
    <taxonomy></taxonomy> → Identificação da espécie.
    <ocurrence></ocurrence> → Identificação única do espécime.
  </specimen>
  <interactions>
    <interaction>
      <relation></relation>
      <specimen> → Espécime relacionado.
        <taxonomy> </taxonomy>
        <ocurrence> </ocurrence>
      </specimen>
    </interaction>
  </interactions>
</specimensschema>

```

No Caso de Uso 2, a ferramenta de busca de relacionamentos partiria do nome científico ou popular para pesquisar nas bases de dados de espécies, utilizando o esquema Plinian Core v.2.0. Filtros como relação e espécie relacionada poderão ser utilizados na pesquisa (Figura 4).

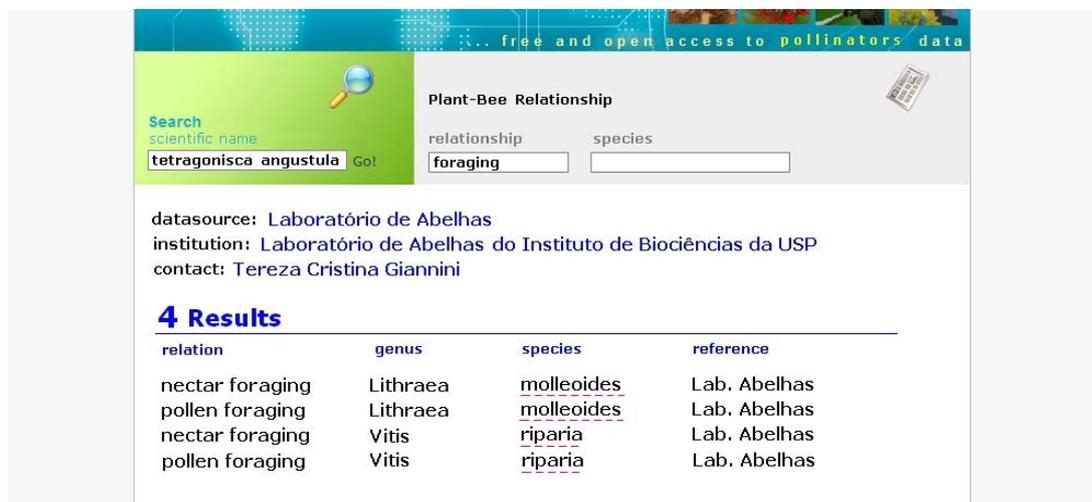


Figura 4 - Exemplo de busca de relacionamento de espécies do Caso de Uso 2.

5. Proposta de Extensão

Após as análises dos esquemas ABCD e Darwin Core e com base nos casos de uso, observou-se que estes esquemas não apresentam suporte natural aos dados de relacionamento entre espécimes e que uma adequação dos seus campos também não é trivial. Foi proposta uma extensão simples ao esquema Darwin Core, chamada “Interaction Extension”¹. Esta extensão se adapta à forma anteriormente proposta utilizando um elemento novo de relacionamento e elementos do próprio Darwin Core²:

<code><InteractionExtension></code>	
<code><RelationshipType/></code>	→ Novo elemento texto simples (VisitedFlowerOf, InfectedBy, etc.).
<code><Record-level Elements/></code>	→ Conjunto de elementos de registro do DwC.
<code><TaxonomicElements/></code>	→ Conjunto de elementos de taxonomia do DwC.
<code><IdentificationElements/></code>	→ Conjunto de elementos de identificação do DwC.
<code><BiologicalElements/></code>	→ Conjunto de elementos de biologia do DwC.
<code></InteractionExtension></code>	

Para contemplar as especificidades do relacionamento para polinizadores, foi criada uma especialização do “Interaction Extension”, chamada de “Pollination Extension”²:

<code><PollinationExtension></code>	
<code><PollinationEvidence/></code>	→ Indica se houve polinização.
<code><PollenRemoval/></code>	→ Indica se houve remoção de Pólen.
<code><NectarRemoval/></code>	→ Indica se houve remoção de Néctar.
<code><OilRemoval/></code>	→ Indica se houve remoção de Óleo.
<code><FlowerPredation/></code>	→ Indica se uma parte da planta foi removida.
<code></PollinationExtension></code>	

Dessa forma é mantida a generalidade do primeiro esquema, que pode ser utilizado para quaisquer tipos de interações, e atende às necessidades específicas dos relacionamentos de polinizadores identificadas pelos especialistas.

¹ <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/InteractionExtension>

² <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/PollinationExtension>

6. Conclusões

A troca de dados de relacionamento entre espécies e espécimes é uma característica de destaque da Rede Temática de Polinizadores da IABIN. Os esquemas existentes e usualmente adotados para troca de dados de espécimes (DwC e ABCD) e o recém introduzido esquema para troca de dados de espécies (Plinian Core) foram analisados com relação ao suporte que podem oferecer a relacionamentos.

Como resultado obteve-se que os padrões atuais de troca de dados de espécimes (ABCD e DwC) não suportam naturalmente o relacionamento de forma detalhada. No entanto, eles podem ser estendidos para que acomodem os dados de interação como módulos adicionais. Diferentemente, o padrão PliC suporta as informações do relacionamento de espécies de forma estruturada.

Para sustentar essa análise, um estudo com especialistas em abelhas foi feito e mostrou que os relacionamentos podem ser bastante complexos e dependentes do grupo estudado. Por esta razão, no caso de espécimes, optou-se por utilizar uma extensão simples e genérica do esquema DwC e uma outra extensão que detalha o relacionamento entre polinizadores.

Além dos aspectos técnicos é importante avaliar a viabilidade prática da aceitação do padrão na comunidade, que pode ser determinante no sucesso de qualquer iniciativa de padronização. Por isso, as extensões propostas foram publicadas no TDWG em uma ferramenta própria para a colaboração e a validação da comunidade científica (Wiki). O próximo passo dos estudos será determinar uma sugestão de extensão do Plinian Core específica para os dados de polinização.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela bolsa de mestrado concedida à Etienne A. Cartolano Jr. A FAPESP – Fundação de Amparo a Pesquisa do Estado de São Paulo, pelo apoio ao projeto BioAbelha. Aos pesquisadores do projeto IABIN –PTN pelas discussões sobre a questão da padronização.

Referências Bibliográficas

- [1] V. L. Imperatriz-Fonseca, D. De Jong, and A. M. Saraiva, Bees as pollinators in Brazil: assessing the status and suggesting best practices. Ribeirão Preto: Holos Editora, 2006.
- [2] C. Kremen, "Managing ecosystem services: what do we need to know about their ecology?," in Ecology Letters. vol. 8, 2005, pp. 468-479.
- [3] P. G. Kevan and V. L. Imperatriz-Fonseca, "Pollinating bees: the conservation link between agriculture and nature," Brasília: Ministério do Meio Ambiente, 2002, p. 313.
- [4] V. L. Imperatriz-Fonseca, F. A. L. Contrera, and A. d. M. P. Kleinert, "A meliponicultura e a Iniciativa Brasileira dos Polinizadores," in XV Congresso Brasileiro de Apicultura / I Congresso Brasileiro de Meliponicultura Natal, 2004.
- [5] E. A. Cartolano Júnior, A. M. Saraiva, V. L. Imperatriz-Fonseca, and R. S. Cunha, "WebBee - a Web - based Information Network on Bees," Revista de Engenharia de Computação e Sistemas Digitais, vol. 1, p. 10, 2003.
- [6] A. M. Saraiva, "Tecnologia da Informação na Agricultura de Precisão e Biodiversidade: Estudos e proposta de Utilização de Web Services para Desenvolvimento e Integração de Sistemas," in Departamento de Engenharia de Computação e Sistemas Digitais. vol. Livre Docência São Paulo: Escola Politécnica da Universidade de São Paulo, 2003, p. 209.