

# Abordagem Colaborativa para a Melhoria da Qualidade de Dados em Bases de Dados Botânicas

Luís Alexandre Estevão Silva<sup>1,2</sup>, Ricardo Oliveira Barros<sup>1</sup>, Eduardo Dalcin<sup>2</sup>,  
Geraldo Zimbrão<sup>1</sup>, Jano Moreira de Souza<sup>1</sup>

<sup>1</sup>COPPE/UFRJ – Coordenação dos Programas de Pós-graduação de Engenharia

<sup>2</sup>Instituto de Pesquisas Jardim Botânico do Rio de Janeiro - Diretoria de Pesquisa Científica, Rio de Janeiro, RJ - Brasil

{estevao, jano, zimbrao, rbarros}@cos.ufrj.br, {estevao, edalcin@jbrj.gov.br}

**Abstract.** *The purpose of this study is to consider how research in collaborative technology can assist at development of management systems for botanical collections, at sharing the knowledge of researchers and promoting the spread of identifications and taxonomic revisions. For this approach a study was conducted in order to verify the quality of data and the main types of errors found in this type of biodiversity database application. A model for the development of an architecture for the system is proposed.*

**Resumo.** *O propósito deste trabalho é considerar como a pesquisa em tecnologia colaborativa pode auxiliar no desenvolvimento de sistemas de gerenciamento de coleções botânicas, compartilhando o conhecimento dos pesquisadores e, promovendo a divulgação das identificações e revisões taxonômicas. Para esta abordagem, foi realizado um estudo para verificar o nível de qualidade de dados e os principais tipos de erros encontrados neste tipo de aplicação de banco de dados na biodiversidade. Um modelo para o desenvolvimento de uma arquitetura para o sistema é proposto.*

## 1. Introdução

A biodiversidade tem recebido muita atenção em face da ampla divulgação e comprovação dos efeitos causados pelo aquecimento global na flora. Assim, diversas iniciativas ao redor do mundo foram postas em ação na busca da preservação das espécies. Porém, para que isto ocorra, primeiro o homem precisa conhecer melhor a natureza. Neste sentido, vários projetos de desenvolvimento de sistemas para a biodiversidade têm sido criados. Acontece, que os dados existentes sobre a biodiversidade em comparação com potencial ainda por ser descoberto é muito pequeno. Estima-se que o Brasil, que é o país de maior riqueza natural, tem apenas 10% desta riqueza conhecida. Assim, para permitir o desenvolvimento de pesquisas no campo da botânica, inúmeras instituições têm mantido grandes bancos de dados e obtido investimentos para o aprimoramento e divulgação destas coleções.

O Instituto de Pesquisas Jardim Botânico do Rio de Janeiro (IPJBRJ)<sup>1</sup> possui uma das maiores bases de dados da flora brasileira, com aproximadamente 450.000 registros *on-line*. Ocorre porém, que os dados apresentam diversos tipos de erros, sendo muitas destes com justificativas para sua ocorrência como por exemplo: registros foram incluídos no banco por digitadores sem formação em Biologia para avaliar os nomes dos espécimes e a dificuldade na identificação dos dados escritos nas etiquetas de identificação originais (algumas escritas em nanquim e latim), por exemplo.

Na Botânica, o taxonomista é quem identifica e valida os nomes das plantas coletadas, muitas ainda desconhecidas na ciência. Seu trabalho tem fundamental importância; o trabalho requer conhecimentos de várias outras áreas, como a Genética e a Ecologia. Afinal, é preciso conhecer cada espécie sob diversos aspectos para saber a que grupo de classificação científica pertence. Ele também verifica se uma espécie está em extinção ou não, se é rara, se é abundante e avalia o seu *status* ambiental. Dentre outras ações, a partir desta classificação e da geração das listas de espécies vulneráveis, são definidas as unidades ou corredores de conservação<sup>2</sup>. Porém, os taxonomistas são poucos e necessitam de anos de estudo para alcançar um nível adequado de experiência e conhecimento. Desta forma, este artigo tem por objetivos 1) realizar uma análise da qualidade dos dados existentes em institutos desta natureza, permitindo que as conclusões do trabalho possam ser utilizadas em outros institutos de pesquisas e jardins botânicos e, 2) permitir o desenvolvimento de um modelo que auxilie na construção de uma ferramenta para facilitar o trabalho dos taxonomistas e o compartilhamento de seus conhecimentos.

Um estudo da aplicação do trabalho proposto está sendo realizado para o Centro Nacional de Conservação da Flora (CNCFlora)<sup>3</sup> no IPJBRJ.

## **2. A análise dos erros**

A busca pela qualidade dos dados passa por uma análise dos tipos de erros que podem existir. Desta forma há a necessidade de um estudo das dimensões a serem avaliadas para a aplicação em uso.

### **2.1. As dimensões de qualidade de dados na taxonomia**

As dimensões podem ser compreendidas como as categorias de qualidade dos dados e por ser classificadas nas seguintes categorias: intrínseca, acessibilidade, contextual e representacional. Para estas categorias, podem ainda ser definidas outras 15 dimensões de qualidade [Strong, Lee e Wang 1997] (Tabela 1).

Além do estudo das categorias e das dimensões, existem tipos de erros que interessam diretamente para o gerenciamento da qualidade de dados. Em [Twidale, Marty 1999] [Beall 2005], estes erros são categorizados em erros de entrada de dados e erros emergentes. Um levantamento mostrou que os erros mais comuns de entrada de dados em aplicações com dados de coleções botânicas são:

---

<sup>1</sup> <http://www.jbrj.gov.br>

<sup>2</sup> [http://www.biodiversityhotspots.org/xp/hotspots/atlantic\\_forest/Pages/conservation.aspx](http://www.biodiversityhotspots.org/xp/hotspots/atlantic_forest/Pages/conservation.aspx)

<sup>3</sup> <http://cncflora.net>

a) *classificação incorreta* – são os erros ocasionados pela falta de conhecimentos taxonômicos; b) *erros de digitação* – estes erros são causados pela dificuldade de leitura de antigas etiquetas de identificação de algumas exsicatas (plantas desidratadas usadas em pesquisas); c) *campos parcialmente preenchidos* – alguns atributos foram parcialmente preenchidos, como é o caso dos dados de coleta da planta: ano, mês e dia; d) *coordenadas geográficas imprecisas* – muitas plantas foram coletadas sem GPS e as coordenadas acrescentadas posteriormente apenas com base no município onde foi realizada a coleta; e) *erros de migração de dados* – alguns longos campos textos podem ter sido importados parcialmente, ex.: campos de observação; e f) *falta de padronização nos nomes* – diversas formas de escrita para um mesmo valor, ex.: nomes dos coletores, plantas e das localidades.

**Tabela 1 – Categorias e dimensões de qualidade de dados**

Categoria	Dimensões
Intrínseca	Acurácia, objetividade, credibilidade e reputação
Acessibilidade	Acessibilidade e segurança de acesso
Contextual	Relevância, valor agregado, validade temporal, Completude e quantidade de dados
Representacional	Interpretabilidade, facilidade de entendimento, representação concisa e representação consistente

Quanto aos erros emergentes, também conhecidos como entropia, podemos citar os dados desatualizados, que são aqueles que apesar de terem sido incluídos corretamente, com o passar do tempo tornam-se inválidos, ex.: mudança de gênero em função de revisões ou mudança nos nomes de unidades geopolíticas. Outro importante fator que justifica a preocupação com a qualidade dos dados é a proliferação de erros, ocasionada quando o taxonomista não tem acesso aos nomes científicos atualizados por meio de outros sistemas de informações.

Outros tipos erros citados em aplicações botânicas [Dalcin 2005] são: g) *campos não-atômicos* – mais de um valor em apenas um campo, um exemplo seria o táxon (família, gênero, espécie e autor) em apenas um campo; h) *campos de entendimento duvidoso* – são valores que podem causar dúvidas quanto ao seu significado, um exemplo seria o valor ? ou *indet.* (indeterminado) em um campo reservado para a família de um espécime;

## 2.2. O uso de metadados em bases de dados botânicas

Para que a análise da qualidade dos dados possa ser realizada, um estudo dos metadados tem que ser realizado. Assim, foi realizado um estudo dos metadados usados em aplicações botânicas com o propósito de diminuir a dificuldade na interpretação dos valores armazenados nas bases de dados.

Metadados são comumente usados em sistemas que envolvem dados de diversas fontes, com o objetivo de uma busca por padronização das bases envolvidas [Moura,

Campos e Silva 2001]. No trabalho corrente, os atributos utilizados seguem o padrão de metadados Darwin Core<sup>4</sup>, padrão este utilizado em muitas aplicações botânicas.

Os tipos de erros apresentados na seção anterior foram levantados no contexto das bases de dados científicas onde a acurácia dos dados impacta diretamente nos resultados obtidos. Assim, podemos perceber a necessidade da associação dos metadados com as dimensões de qualidade [Pipino, Lee e Wang 2002]. Estas dimensões são apresentadas na Tabela 2 a seguir.

**Tabela 2 – Descrição das dimensões de qualidade**

Dimensões	Definições
Acessibilidade	O dado está disponível ou é fácil e rapidamente disponibilizado
Quantidade de dados apropriada	O volume dos dados é apropriado para a tarefa em execução
Credibilidade	Os dados são verdadeiros e tem credibilidade
Completude	Não faltam dados e são de suficiente tamanho e profundidade para as tarefas
Representação concisa	Os dados são representados de uma forma compacta
Representação consistente	Os dados são apresentados em um mesmo formato
Facilidade de manipulação	Os dados são de fácil manipulação e aplicados a diferentes tarefas
Livre de erros	Os dados são corretos e livres de erros
Interpretabilidade	Os dados estão em linguagens apropriadas e os símbolos e as definições são claras
Objetividade	Os dados não são tendenciosos, prejudiciais e parciais
Relevância	Os dados são aplicáveis e auxiliam na execução das tarefas
Reputação	Os dados têm uma alta reputação em termos de suas fontes e conteúdo.
Segurança	O acesso aos dados é restrito
Validade temporal	Os dados são suficientemente atualizados para a tarefa em questão
Compreensão	Os dados são facilmente compreendidos
Valor agregado	Os dados são benéficos e prevêm vantagens a partir de seu uso

Para implementar a associação entre os metadados e as dimensões de qualidade, foi realizado um levantamento e validação com os pesquisadores da área de modo a

<sup>4</sup> <http://splink.cria.org.br/digir/darwin2.xsd>

identificar os atributos principais para a análise qualitativa. Assim, foi verificado que alguns atributos têm uma contribuição diretamente proporcional a um ou várias destas dimensões, por exemplo, o atributo *detby* (determinador), que contém o nome do taxonomista que identificou a espécie, tem um peso maior de acordo com a reputação do pesquisador e seu tempo de experiência. Outros atributos, como por exemplo o campo *genus*, onde temos a identificação do nome do gênero, pode estar relacionado a várias dimensões. Neste campo, uma dimensão está associada também com o problema da entropia dos dados, tendo em vista que o nome do gênero pode ser alterado como resultado de um processo de revisão taxonômica, assim todas as determinações que foram baseadas naquele nome podem ser consideradas erradas, ocasionando sua proliferação.

Um estudo detalhado da associação das dimensões de qualidade de dados com bases de dados taxonômicas foi feito em [Dalcin 2005]. Como resultado deste trabalho foram selecionadas as seguintes dimensões: *credibilidade*, *completude*, *livre-de-erros*, *reputação* e *validade temporal*. As dimensões *credibilidade* e *reputação* são justificadas em face do grau de segurança assegurado na identificação dos nomes dos espécimes pelo taxonomista que realizou a identificação. A Tabela 3 mostra parte das relações entre as dimensões de qualidade utilizadas no presente trabalho com os principais atributos. Esta associação se faz necessária para a determinação da importância de cada atributo na base de dados. Nesta tabela os títulos das colunas representam, respectivamente: 1) credibilidade, 2) completude, 3) livre-de-erros, 4) validade temporal e, 5) reputação.

**Tabela 3 - Associação entre as dimensões de qualidade e os principais atributos**

Atributos	1	2	3	4	5
accession		x	x		
suffix		x			
family	x	x	x	x	
subfam	x	x	x	x	
genus	x	x	x	x	
cf	x	x	x	x	
sp1	x	x	x	x	
author1	x	x	x	x	
collector		x			x
number		x			
addcoll					
colldd	x	x	x		
collmm	x	x	x		
collyy	x	x	x		
detby					x

### 3. A colaboração na busca da qualidade dos dados

Inúmeras instituições disponibilizam em seus *sites* catálogos sobre os nomes das espécies, acrescentando-se ainda os livros, teses, dissertações, artigos e periódicos.

Porém, mesmo com a disponibilidade de informações oriundas destas fontes, o trabalho do taxonomista não pode ser substituído por um sistema de informação pois muitas são as dificuldades apresentadas em seu trabalho. Dentre estas dificuldades podemos citar que muitas das identificações somente são possíveis de ser realizadas examinando-se o material *vis a vis* e, em alguns casos, há a necessidade do processamento de análises laboratoriais das plantas, com potentes microscópios de varredura e exames de DNA. Assim, é comum o intercâmbio do material, o que sobrecarrega ainda mais o trabalho dos taxonomistas. Diante destes fatores, torna-se evidente a necessidade do trabalho colaborativo entre estes taxonomistas. Esta abordagem em por objetivo minimizar as dificuldades de trabalho deste profissional.

O Trabalho Cooperativo Apoiado por Computador (CSCW) refere-se à área de pesquisa que examina o projeto, a construção e o uso de ferramentas de colaboração. Neste ambiente, segundo [Tiwana 2000], “os processos de criação, compartilhamento e aplicação do conhecimento envolvem colaboração”. Podemos citar como alguns dos benefícios trazidos por estas ferramentas: a facilidade de comunicação, a definição de categorias de assuntos, a criação de ambientes de grupos de discussão. Estas vantagens explicam o crescente interesse das organizações em adotarem tais tecnologias, pois estas ferramentas tem auxiliado na disseminação de conhecimento tácito.

A motivação para o uso da tecnologia colaborativa no contexto deste trabalho visa permitir, além do compartilhamento do conhecimento entre os taxonomistas, a obtenção na melhora do grau de qualidade dos dados no trabalho dos taxonomistas. Assim, o desenvolvimento de um trabalho colaborativo para auxiliar estes profissionais torna-se de grande relevância. Na abordagem proposta neste trabalho a busca pela qualidade dos dados tem início no processo de identificação e correção dos nomes das plantas. Esta última observação faz alusão a semelhança que existe com o processo de realizado na entrada de dados de uma biblioteca tradicional, que possui menores índices de erros na catalogação, por ser realizado por um profissional de biblioteconomia. A entrada de dados botânicos nem sempre é realizada por taxonomistas.

O ambiente colaborativo, pelo controle realizado sobre os usuários, permite ainda a identificação daqueles que produzem a maior quantidade de erros e quais são os tipos de erros com maior incidência, facilitando qualquer tipo de auditoria ou análises de qualidade de dados. Na próxima seção será apresentado a proposta de solução baseada na tecnologia colaborativa.

#### **4. O modelo proposto para a abordagem**

Nesta seção é apresentado o modelo proposto para a melhora da qualidade de dados em um ambiente colaborativo. Para facilitar a compreensão de seu objetivo, podemos considerar que ele é a base para a integração da tecnologia colaborativa e da aplicação do controle da qualidade de dados. O modelo (Figura 1), desenvolvido a partir do trabalho de qualidade de dados realizado em [Barros 2009], controlará as ações dos taxonomistas de forma colaborativa e, aproveitará este ambiente para controlar os erros e proporcionar a padronização e difusão de atualizações taxonômicas nas bases de dados. Esse controle é possível por meio do uso das dimensões da qualidade de dados, da associação destas dimensões com os metadados e pela classificação dos usuários de

acordo com seus níveis de conhecimento. O uso de controles de qualidades em modelos conceituais permite ainda a definição de métricas [Stvilia 2007] [Peralta et al. 2004].

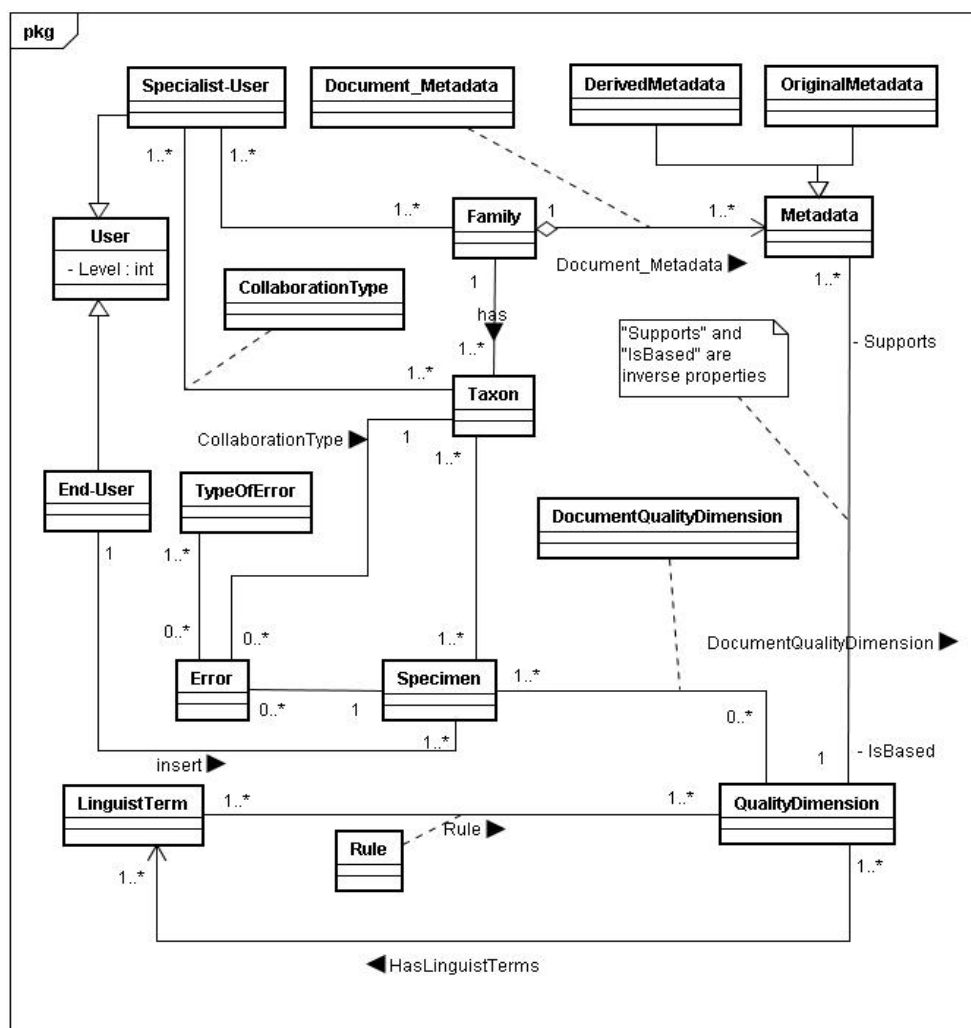


Figura 1. Modelo proposto

A seguir as principais classes do modelo proposto, em especial, as relacionadas com a colaboração nas atividades relacionadas à taxonomia são apresentadas:

**Classe Metadata** - Esta classe define o conjunto de metadados utilizados para a avaliação da qualidade da informação. Diferentes metadados podem ser associados com os dados, incluindo aqueles destinados a melhora da qualidade, de acordo com o conjunto de dimensões [Rothenberg 1996]. Os metadados podem ser classificados em originais e derivados:

- **Classe DerivedMetadata** – são metadados que não podem ser diretamente recuperados do documento, sendo necessária a aplicação de uma função computacional. Um exemplo, seria o atributo *PeríodoDeAtualização* que depende da *DataCadastro*;

- **Classe OriginalMetadata** – representa os metadados que podem ser diretamente recuperados do documento;

**Classe DocumentMetadata** – É uma classe de associação que descreve alguma informação do documento e do seu conteúdo;

**Classe DocumentQualityDimension** – É uma classe de associação que contém as dimensões de qualidade usadas para um determinado espécime;

**Classe QualityDimensions** – Esta classe define o conjunto das dimensões de qualidade que serão utilizadas na análise das bases de dados. Em termos de utilização, o modelo parte do princípio que é possível definir um relacionamento entre as dimensões de qualidade e os metadados. A Tabela 1 mostra o conjunto das dimensões de qualidade de dados capaz de suprir as necessidades de qualidade dos usuários para a aplicação;

**Classe LinguisticTerm** – A classe define o conjunto de adjetivos e advérbios relatados para as variáveis lingüísticas;

**Classe Rule** – O conjunto de classes que forma o sistema de regras do modelo tem por finalidade construir as regras mediante a inferência de conhecimento. Essa inferência ocorre sobre uma variável lingüística, em função de uma associação lógica fuzzy entre outras variáveis lingüísticas. Uma regra é representada por [ROSS 2004] e segue o formato: *REGRA: SE premissa (antecedente) ENTÃO conclusão (conseqüente)*. A seguir um exemplo da definição de uma regra:

REGRA: SE as variáveis de entrada < completeza = regular E reputação = boa E atualidade = boa >  
ENTÃO a variável de saída PC = boa.

Onde o PC – Prognóstico Composto de Qualidade de Informação pode ser obtido pela inferência de regras fuzzy.

**Classe Family** – Esta classe foi criada para relacionar os taxonomistas e as famílias de plantas nas quais são especialista;

**Classe Taxon** – Os táxons são os nomes científicos das plantas, uma composição formada pelo nome da família, do gênero, da espécie e do autor da espécie. Quando uma espécie é descoberta, o seu táxon é submetido a comunidade para análise e validação. Processo semelhante ocorre quando é realizada uma revisão;

**Classe Specimen** – Os espécimes são os indivíduos, os registros, das coletas das plantas. É comum que ao longo de sua carreira, o taxonomista efetue milhares de coletas para seus diversos estudos, como por exemplo, os levantamentos florísticos e análises de conservação. Em grande parte dos sistemas, a entrada de dados é efetuada sem críticas, de uma forma livre, sendo que muitos botânicos ainda fazem uso de planilhas para uma entrada mais rápida. Assim, toda a pesquisa realizada posteriormente pode ficar seriamente comprometida pela elevada quantidade de erros que podem ter sido cometidos durante o processo de inclusão;

**Classe TypeError** – Nesta classe os tipos de erros são listados de modo a facilitar a identificação do trabalho de limpeza a ser realizado;

**Classe Error** – O objetivo desta classe é o de manter registrado quais erros foram



cometidos e quando, associados com os usuários. Torna-se importante para o trabalho de busca da qualidade o acompanhamento temporal, pois este assinalará a evolução no nível de qualidade das entradas de dados de um usuário. Desta forma, o sistema poderá atualizar o *status* do usuário quanto a confiança de suas entradas;

**Classe User** – Os usuários do sistema são definidos nesta classe. Ela possui duas subclasses descritas a seguir:

•**Classe Specialist-User** – Este tipo de usuário é o taxonomista ou um pesquisador. Pode ter a especialidade em uma ou mais famílias de plantas e, por ser o detentor do conhecimento, a ele é permitido o acesso aos diversos táxons (nomes científicos das plantas);

•**Classe End-User** – O usuário final do sistema é controlado para que seja possível associar os espécimes com aquele que o inseriu na base. Assim, a qualquer momento será possível observar os autores de erros e também, divulgar para os usuários, as revisões e descobertas de nomes;

**Classe CollaborationType** – Todas as formas de colaboração ficam registradas nesta importante classe do sistema. Por meio desta classe, podemos registrar no sistema o trabalho realizada por um taxonomista. Alguns dos principais tipos de colaboração seriam: a identificação de uma nova espécie, a revisão ou classificação de um táxon, compartilhamento das referências bibliográficas e a análise da sinonímia.

#### 4. Implementação da arquitetura proposta

A proposta encontra-se ainda em fase de desenvolvimento. Entre seus módulos, apenas o que trata da qualidade de dados [Xexéo, Belchior, & da Rocha 1996] está terminado. Para sua implementação foi utilizado *Matlab Fuzzy Logic Toolbox* que é uma coleção de funções criadas para uso no ambiente de computação numérica MATLAB ([www.mathworks.com](http://www.mathworks.com)). Esta ferramenta permite a criação e edição de um sistema de inferências fuzzy para uso com o framework do MATLAB.

#### 5. Conclusão

A busca da qualidade de dados em sistemas de informação tem despertado uma atenção cada vez maior. É sabido que os resultados das análises realizadas em um banco de dados somente poderão ser efetivamente usados se estes dados forem confiáveis. Assim, foi proposto o desenvolvimento de uma arquitetura na qual um modelo conceitual pode proporcionar a melhoria da qualidade dos dados por promover e controlar as contribuições entre os usuários em um ambiente colaborativo; motivando-os a uma procura pelos erros e auxiliando na divulgação das atualizações e das revisões taxonômicas. Além, de permitir a indicação das fontes que geram a maior quantidade de dados com erros e os usuários responsáveis.

O trabalho apresentou como principais contribuições o estudo dos principais erros encontrados em bases de dados taxonômicas, o estudo das dimensões de qualidade de dados relacionadas com estas bases, a associação destas dimensões com o padrão de metadados Darwin Core e, apresentar a Botânica como um interessante campo de atuação de sistemas colaborativos e que necessita de tecnologias para a

limpeza de dados e o controle de erros. Outra contribuição foi a de permitir que outras instituições que utilizem estes tipos de dados em suas aplicações possam se utilizar do aprendizado alcançado, podendo este conhecimento ser aplicado pelos órgãos ambientais para a definição de ações de conservação da flora.

## Referências

- Barros, R. O. (2009) "Qualidade de Informação na Web: Um Prognóstico Fuzzy Baseado em Metadados, p. 231, tese de doutorado, Universidade Federal do Rio de Janeiro;
- Beall, J. (2005) "Metadata and Quality Problems in the Digital Library", *Journal of Digital Information*, <http://www.cndwebzine.hcp.ma/IMG/pdf/Beall.pdf>;
- Dalcin, E. C. (2005) "Data Quality Concepts and Techniques Applied to Taxonomic Databases" tese de doutorado, Universidade de Southampton, [http://www.dalcin.org/eduardo/downloads/edalcin\\_thesis\\_submission.pdf](http://www.dalcin.org/eduardo/downloads/edalcin_thesis_submission.pdf);
- Moura, A. M. C., Silva, L. A., Campos, M. L. M. (2001) "A Metadata Approach for Designing Configurable Interfaces in Digital Libraries", In: 12th International Workshop on Databases and Expert Systems Applications, Munique. DEXA 2001. Los Alamitos: IEEE Computer Society Press, 2001. v.1. p.942 – 947, [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?reload=true&arnumber=953176](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?reload=true&arnumber=953176), março;
- Peralta, V., Ruggia, R., Kedad, Z., & Mokrane, B. (2004) "A framework for data quality evaluation in a data integration system", *Proceedings of the Brazilian Symposium of Databases* (pp.134–147);
- Pipino, L. L., Lee, Y.W., & Wang, R.Y. (2002) "Data quality assessment", *Communications of the ACM*, 45, 211–218;
- Ross, T. J. (2004) "Fuzzy Logic with Engineering Applications", 2nd ed. ed. West Sussex, England;
- Rothenberg, J. (1996) "Metadata to support data quality and longevity", *Proceedings of the 1st IEEE Metadata Conference* (pp. 16–18);
- Strong, D. M., Y. W. Lee, R. Y. Wang (1997) "Data quality in context", *Communications of ACM* 40:103-110;
- Stvilia, B., Gasser, L., Twidale, M. B. and Smith, L. C. (2007) "A Framework for Information Quality Assessment", *Assessment*, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.8053&rep=rep1&type=pdf>, março;
- Twidale, M. B., Marty, P.F. (1999) "An investigation of data quality and collaboration", *Technical Report UIUCLIS - 1999*, University of Illinois at Urbana-Champaign, USA;
- Tiwana, A. (2000) "The Knowledge Management Toolkit, A Practical Techniques for Building a Knowledge Management System", Prentice Hall;
- Xexéo, G., Belchior, A., & da Rocha, A.R.C. (1996) "Aplicação da teoria fuzzy em requisitos de qualidade de software", *Memorias-I, XXII CLEI*, Santafé de Bogotá, Colombia, Junio (pp. 3–7).