

Aplicação da Ciência Cidadã em Banco de Dados de Herbários

Luís Alexandre Estevão Silva^{1,2}, Felipe Alves de Oliveira¹

¹Instituto de Pesquisas Jardim Botânico do Rio de Janeiro - Diretoria de Pesquisa Científica, Rio de Janeiro, RJ – Brasil

²Universidade Estácio de Sá – Diretoria de Pesquisa Aplicada, Rio de Janeiro, RJ - Brasil

{estevao, felipealves}@jbrj.gov.br

Abstract. *This article presents a proposal for citizen science application on databases of herbaria. These databases have a growing volume of data, but with the publication of high-quality images, it is possible that both the academic community as the general population, can participate in a process of acceleration of data entry and validation of same. To achieve this goal were studied the possibilities of contributions according to the level of knowledge of the participants in the botany. A methodology was analyzed to optimize the contribution process of citizens. The proposal was partially tested only with the academic community, with good results.*

Resumo. *O presente artigo apresenta uma proposta para aplicação da ciência cidadã em bancos de dados de herbários. Esses bancos tem um volume crescente de dados, porém com a publicação de imagens de alta qualidade, é possível que tanto a comunidade acadêmica quanto a população em geral, possam participar em um processo de aceleração da entrada de dados e validação dos mesmos. Para atingir esse objetivo foram estudadas as possibilidades de contribuições de acordo com o nível de conhecimento dos participantes na botânica. Uma metodologia foi analisada visando otimizar o processo de contribuição dos cidadãos. A proposta foi testada parcialmente, apenas com a comunidade acadêmica, com bons resultados.*

1. Introdução

Os diversos herbários distribuídos pelo Brasil e no mundo todo, têm feito um grande trabalho para tornar os dados e imagens de suas coleções acessíveis aos pesquisadores e estudantes [Peixoto e Morim 2002][Lanna et al. 2018]. Além do acesso aos dados, a publicação das imagens permite aos especialistas a visualização e identificação de amostras dentro das famílias estudadas por esses especialistas; aumentando assim o valor dos exemplares físicos e preservando esse frágil material do manuseio. Porém, para que os dados existentes nos diversos bancos de dados dos herbários sejam efetivamente utilizados, os curadores têm que dispensar um tempo razoável em atividades, como por exemplo: a digitação, a digitalização das exsiccatas, a limpeza e correção dos dados. Todas as atividades são importantes, porém a limpeza dos dados possui destaque pela consequência na geração de conhecimento científico.

Uma justificativa para o grande volume de erros nos dados é que as coleções possuem registros de coletas muito antigas, a partir de 1760 no caso do Jardim Botânico

do Rio de Janeiro [Forzza et al. 2008], com etiquetas de identificação escritas em nanquim e latim. Estudos anteriores sobre qualidade de dados [Silva et al. 2010] [Dalcin et al. 2012][Silva 2016], identificaram que no processo de entrada de dados também há uma grande possibilidade da entrada de novos erros, no momento da digitação. Esse crescente aumento do volume de dados sobre a biodiversidade tem provocado uma busca por formas mais eficientes de extração de conhecimento sobre a Flora. Nesse sentido, o desenvolvimento de sistemas de informações de coleções científicas tornou-se mais uma necessidade. O Instituto de Pesquisas Jardim Botânico do Rio de Janeiro desenvolveu o Jabot [Silva et al. 2017], que é hoje disponibilizado para uso em 40 herbários pelo Brasil, além do próprio Instituto. Porém, apesar de um sistema especialista de gerenciamento de herbários, o volume de trabalho para os curadores continua grande e, outras medidas podem ser buscadas para agilizar esse trabalho.

2. Projetos existentes de ciência cidadã na biodiversidade

Diversas iniciativas com a participação da sociedade têm surgido, uma delas é o e-Bird (<http://ebird.org>), além dos trabalhos [Bonter e Cooper 2012][Bonney et al. 2014] para observadores da natureza. Quanto a aplicação dos recursos da ciência cidadã na biodiversidade, alguns projetos também foram desenvolvidos [Wiggins et al. 2011] [Kosmala et al. 2016], permitindo nossa consideração de uso nessa proposta.



	<p>Dados taxonômicos – estudantes de botânica</p> <ul style="list-style-type: none"> • Transcrição taxonômica errada • Transcrição incompleta • Registro de uma nova determinação taxonômica
	<p>Dados gerais das coletas – público geral</p> <ul style="list-style-type: none"> • Nome do coletor errado • Identificação do nome do coletor principal • Completar a digitação de espécimes com dados informados mínimos • Data da coleta informada errada
	<p>Dados geográficos – estudantes de geografia</p> <ul style="list-style-type: none"> • Dados geográficos informados com erros • Dados geográficos não coincidem com a localidade • Possibilidade de indicação pelos dados da localidade

Figura 1. Possibilidades de uso da ciência cidadã nos dados de herbários

Esse levantamento e o estudo anterior sobre os tipos de erros nos dados, permitiu considerar 3 formas de uso da ciência cidadã nos bancos de dados de herbários: na taxonomia, nos dados gerais das coletas e dados geográficos (Figura 1). Há que se destacar que determinados erros podem exigir um conhecimento prévio na análise do erro; como por exemplo, erros taxonômicos sendo avaliados por estudantes de biologia e, erros de coordenadas geoespaciais, por estudantes de geografia. Como incentivo à participação dos estudantes, o sistema tem como controlar o total de horas trabalhada, convertendo em atividades complementares, necessárias à graduação desses alunos.

3. Metodologia proposta

Para permitir a aplicação da ciência cidadã uma proposta de metodologia foi formulada (Figura 2). A metodologia considerou que o material a ser digitado ou, pode ser definido de três formas: pela prioridade estabelecida pelo curador; pelo sistema, a partir das rotinas de qualidade de dados já ou, a partir das amostras que possuem apenas as imagens. Quando uma imagem é disponibilizada para validação, podem ser encontrados os seguintes tipos de erros: a) dados incompletos – o digitador não registrou o ano, por exemplo; b) dados com erros na digitação – o registro do dado foi feito com erro por uma dúvida na formação do nome científico, por exemplo. De forma a assegurar uma maior qualidade no trabalho realizado pelos participantes, o sistema apresenta as mesmas amostras para validação para três diferentes contribuidores.

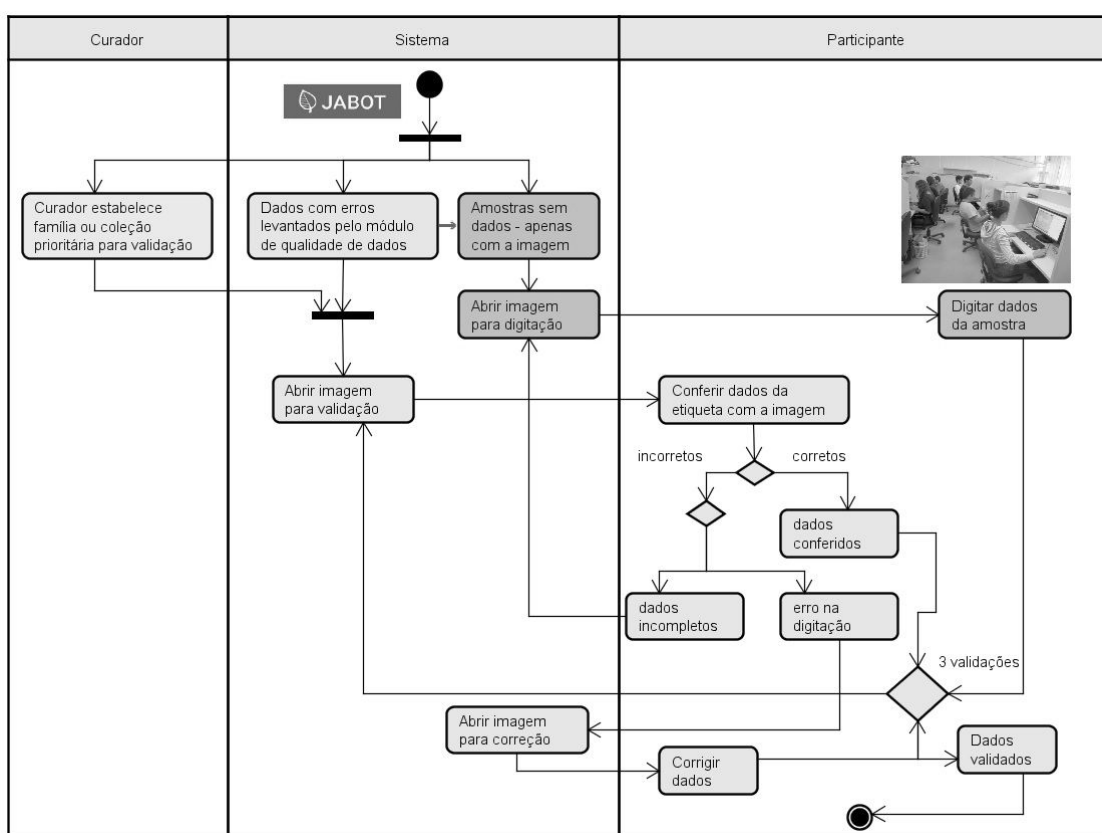


Figura 2. Metodologia proposta para a execução das atividades de entrada e validação de dados nos sistemas de informações de herbários.

4. Conclusão

Concluiu-se, que a proposta apresentada foi justificada pela constatação que os curadores, que normalmente tem uma equipe reduzida de funcionários nos herbários, têm um grande desafio de trabalho para avaliar e corrigir os erros de sua coleção. Com base nas características da aplicação: seu apelo histórico, grande valor científico para a botânica, volume de dados e presença de imagens em alta resolução disponíveis na web; e, com os recursos computacionais disponíveis na atualidade, a solução baseada na Ciência Cidadã pode atrair tanto a comunidade acadêmica, quanto os cidadãos da sociedade, auxiliando no trabalho de efetuar a entrada e conferência dos dados existentes. Alguns resultados promissores já foram obtidos por conta do uso de uma tela de correção de dados geográficos, permitindo comparar dados com a imagem.

Agradecimentos

Agradeço ao Programa Pesquisa Produtividade da Universidade Estácio de Sá pelo apoio.

Referências

- Bonney, R., Shirk, J.L., Phillips, T.B., et al. (2014). Citizen science: Next steps for citizen science. *Science*, v. 343, n. 6178, p. 1436-1437.
- Bonter, D.N., Cooper, C.B. (2012). Data validation in citizen science: A case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, v. 10, n. 6, p. 305-307.
- Dalcin, E.C., Silva, L. A. E., Zimbrão, G., et al. (2012). Data Quality Assessment at the Rio de Janeiro Botanical Garden Herbarium Database and Considerations for Data Quality Improvement. In *In: 8th International Conference on Ecological Informatics*.
- Forzza, R.C., Mynssen, C.M., Tamaio, N.; Barros, C.; Franco, L.; Pereira, M.C.A. *As coleções do herbário. 200 anos do Jardim Botânico do Rio de Janeiro*. Rio de Janeiro: Jardim Botânico do Rio de Janeiro, 2008. p. 45-55.
- Kosmala, M., Wiggins, A., Swanson, A.; Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, v. 14, n. 10, p. 551-560.
- Lanna, J., Silva, L.A., Morim, M., et al. (2018). Herbarium collection of the Rio de Janeiro Botanical Garden (RB), Brazil. *Biodiversity Data Journal*, v. 6, p. e22757.
- Peixoto, A. L. and Morim, M. P. (2002). Coleções Botânicas: documentação da biodiversidade brasileira". *Flora*, v. 55, n. 3, p. 21-24.
- Silva, L.A.E. (2016). A Data Mining Approach for Standardization of Collectors Names in Herbarium Database. *IEEE Latin America Transactions*, v. 14, n. 2, p. 805-810.
- Silva, L.A.E., Barros, R.O., Dalcin, E., Zimbrão, G.S.; Souza, J. (2010). Abordagem Colaborativa para a Melhoria da Qualidade de Dados em Bases de Dados Botânicas. In *II Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais, Belo Horizonte. XXX Congresso da Sociedade Brasileira de Computação - Computação Verde: Desafios Científicos e Tecnológicos, Belo Horizonte*. p. 535-544.
- Silva, L.A.E., Fraga, C.N., Almeida, T.M.H., et al. (2017). Jabot - Sistema de gerenciamento de Coleções Botânicas: A experiência de uma década de desenvolvimento e avanços. *Rodriguesia*, v. 68, n. 2, p. 391-410.
- Wiggins, A., Newman, G., Stevenson, R. D.; Crowston, K. (2011). Mechanisms for data quality and validation in citizen science. In *Proceedings - 7th IEEE International Conference on e-Science Workshops, eScienceW 2011*.