

# Uma proposta de abordagem para gerenciamento de tarefas complexas em projetos de ciência cidadã

Marcela Mayumi Mauricio Yagui, Adriana S. Vivacqua

Programa de Pós-Graduação em Informática – UFRJ – Rio de Janeiro – RJ – Brasil

marcela.yagui@ppgi.ufrj.br, avivacqua@dcc.ufrj.br

**Abstract.** *Crowdsourcing and derivative approaches, such as citizen science, grew due to the advent of the web 2.0. However, understanding mechanisms, interactions, and how to manage tasks becomes difficult as the scalability of the crowd increases. This paper aims to show the design of a study to manage complex tasks in citizen science projects. Our proposal intends to contribute to an approach for sending, interpreting and annotating scientific data by non-specialists.*

**Resumo.** *Projetos crowdsourcing e de abordagens derivadas, como de ciência cidadã, cresceram devido ao advento da web 2.0. No entanto, entender os mecanismos, interações, e a forma de gerir tarefas, torna-se difícil à medida que a escalabilidade da multidão aumenta. Este trabalho tem como objetivo mostrar o planejamento de um estudo para gerir tarefas complexas em projetos de ciência cidadã. Nossa proposta pretende contribuir com uma abordagem para envio, interpretação e anotações de dados científicos por não especialistas.*

## 1. Introdução

O *crowdsourcing* é em uma abordagem na qual uma organização terceiriza tarefas para uma multidão virtual indefinida, em uma chamada direcionada. Este modo de trabalho ganhou força com o crescimento da web 2.0 e a presença on-line de pessoas em fóruns, redes sociais e *wikis* [Blohm et al. 2013; Brabham 2013].

À medida que um projeto *crowdsourcing* se torna maior, mais participativo e mais colaborativo, fica mais difícil entender: (i) mecanismos e interações entre a multidão; (ii) a multidão adequada para cada tarefa; (iii) metas, abordagens e níveis de granularidade das tarefas. Deste modo, pode-se considerar que há uma complexidade inerente a esse sistema, que é formado por meio da interação entre multidão, organização, tarefas e ambiente [Zou et al. 2014]. Em chamadas aplicadas, como projetos que envolvem pesquisas científicas, esse problema é intensificado devido ao fato de que apenas informações exatas e fidedignas devem ser consideradas nos resultados experimentais. Além disso, há a questão da propriedade intelectual das informações fornecidas pela *crowd* e a confiabilidade dos resultados obtidos.

A corrente *crowdsourcing* na qual são criados ambientes colaborativos on-line que fomentam a participação do público em etapas de pesquisas científicas é denominada *citizen science* (ciência cidadã, em português). A ciência cidadã possui um forte

potencial de engajamento da ciência com a sociedade, permitindo que estudantes ou interessados participem ativamente de um projeto científico real [Cohn 2008].

O objetivo deste trabalho é propor uma abordagem para gerir tarefas complexas em projetos de ciência cidadã. Na seção 2 apresentamos uma breve revisão da literatura e alguns estudos relacionados, contextualizando o estado da arte e as oportunidades de pesquisa. Na seção 3 mostramos o desenho de pesquisa elaborado para atingir o objetivo proposto. Na seção 4 apresentamos uma breve discussão e as considerações finais.

## 2. Revisão da Literatura

A *crowd science*, *citizen science*, *networked science*, ou *massively-collaborative science* é um tipo de *crowdsourcing* que fomenta a participação do público para apoiar as etapas de pesquisa científica, como por exemplo, com a contribuição de investigações, registros históricos e com coleta de amostras e evidências científicas [Franzoni and Sauermann 2014]. A ciência cidadã permite que interessados participem ativamente de um projeto científico real e apoiem pesquisadores a encontrar soluções científicas de relevância para a área de estudo [Cohn 2008].

Usualmente, a ciência cidadã possui diferentes níveis de participação da *crowd*: nível 1, onde as pessoas indicam áreas ou assuntos a explorar; nível 2, onde a *crowd* faz interpretações básicas de dados; nível 3, onde as pessoas ajudam na definição de problemas e na coleta de dados e amostras; e nível 4, onde ocorre a contribuição dos usuários na coleta, definição do problema e análise de dados [Haklay 2013].

Nesse sentido, alguns estudos vêm sendo realizados na área de ciência cidadã, principalmente aplicados em cenários das ciências naturais [Kullenberg and Kasperowski 2016]. Com relação à complexidade da tarefa, [Franzoni and Sauermann 2014] consideram que a interligação entre as tarefas implicam no grau de complexidade. [Staffelbach et al. 2015] discutem que a complexidade da tarefa se deve ao conhecimento técnico que deve ser aplicado em sua realização. [Amer-Yahia and Basu Roy 2015] relacionam complexidade ao grau de colaboração entre os participantes.

Para projetos *crowdsourcing*, alguns trabalhos visaram a entender a implicação da *crowd* na qualidade das tarefas. [Cui et al. 2017] criaram um método para identificar perfis e os relacionar às tarefas mais adequadas. [Cullina et al. 2016] propuseram um subprocesso dentro de projetos de *crowdsourcing* para selecionar a melhor *crowd* para projetos. [Hirth et al. 2013] analisaram a confiabilidade e os custos de abordagens como ‘decisão majoritária’ e ‘grupo de controle’ para validação de contribuições de uma *crowd*. [Beer et al. 2017] investigaram questões éticas e relacionadas à propriedade intelectual de dados gerados por sistemas *crowdsourcing*, sob a perspectiva das organizações criadoras das chamadas.

Deste modo, identificamos algumas questões envolvendo a ciência cidadã:

1- Como escalonar tarefas complexas envolvendo projetos de ciência cidadã? 2- Como escolher a *crowd* certa para cada tarefa? 3- Como validar as informações enviadas pela *crowd*? 4- Como garantir a propriedade intelectual das contribuições da *crowd* e dos resultados científicos obtidos?

Com o embasamento teórico inicial, propomos um desenho de pesquisa que será apresentado na seção a seguir.

### 3. Desenho de Pesquisa

Esta seção tem como objetivo mostrar os principais procedimentos da pesquisa proposta. A Figura 1 ilustra as etapas que serão detalhadas a seguir.

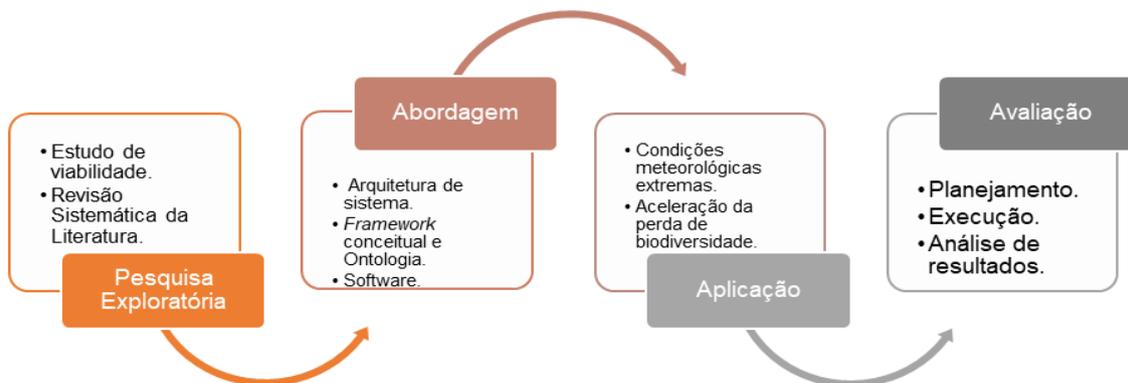


Figura 1 - Procedimento de pesquisa

#### 3.1. Pesquisa Exploratória

A etapa de pesquisa exploratória consistirá em realizar um estudo preliminar sobre o tema de pesquisa, para obtenção de uma familiaridade inicial com o problema, de modo a torná-lo mais explícito [Gil 1999]. Diante disso, esta etapa será realizada através do estudo de viabilidade de um projeto inicial e de uma revisão sistemática da literatura para o levantamento do estado da arte. As duas subetapas serão descritas a seguir.

##### 3.1.1. Estudo de viabilidade

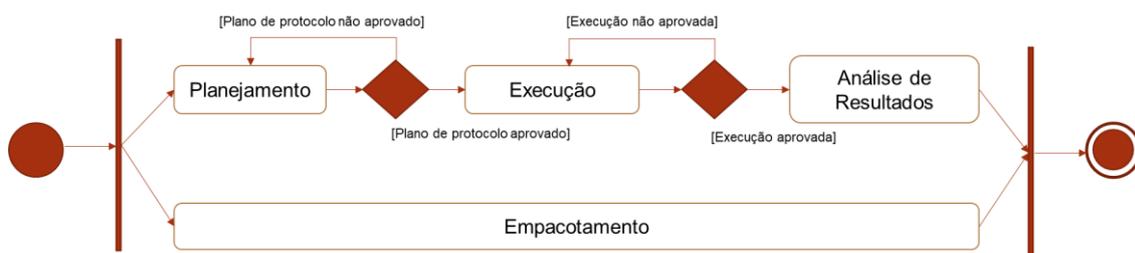
A primeira etapa desta pesquisa consiste no estudo de viabilidade do projeto. Para este estudo, serão consultados especialistas em áreas das ciências naturais (exemplo: clima, botânica, entre outros) a fim de verificar a relevância dos problemas sobre ciência cidadã inicialmente levantados. Além disso, serão consideradas a viabilidade de tempo, de recursos humanos e computacionais para a conclusão do estudo.

##### 3.1.2. Revisão Sistemática da Literatura

A segunda etapa da metodologia que será aplicada neste trabalho consiste na realização de pesquisas para encontrar bibliografias relacionadas com as áreas de foco deste estudo. Neste caso, a revisão sistemática da literatura (RSL) atuará como parte integrante do processo de pesquisa, pois para explorar uma área do conhecimento, é necessário que se conheça o estado da arte dessa área [Berwanger et al. 2007].

Para este trabalho, será utilizado o método de RSL proposto por Biolchini *et al.* (2005), onde o processo será composto por quatro etapas e suas respectivas avaliações: (i) planejamento; (ii) execução; (iii) análise de resultados; (iv) fase de empacotamento. A Figura 2 ilustra o processo definido por Biolchini *et al.* (2005).

A primeira fase desta etapa (seção 3.1) resultará na definição do tema de pesquisa. Após a definição do tema e verificação da originalidade, relevância, viabilidade e do estudo preliminar, ocorrerá a formulação do problema de pesquisa. A seção seguinte descreve a etapa de construção da abordagem.



**Figura 2 - Processo de RSL. Adaptado de Biolchini et al. (2005)**

### 3.2. Abordagem

Nesta etapa, será desenvolvida a abordagem. Após a definição do problema, será proposta uma abordagem composta por uma arquitetura de sistema, um *framework* conceitual para modelar o domínio escolhido por meio de processos, técnicas e ontologias para representar o conhecimento. Com isso, espera-se: estudar ontologias e sistemas relacionados; analisar padrões de metadados para compor bases de dados que suportem essas ontologias; utilizar ferramentas, como o Protégé, para construção e edição de ontologias e extração de informações de bases ontológicas de conhecimento.

Além disso, será implementado um sistema para apoiar o problema identificado. Será utilizada a tecnologia web por ser independente de plataforma e pela possibilidade de ter sua interface adaptada para utilização em *desktops* e *web apps*.

### 3.3. Cenários de aplicação

Esperamos aplicar a abordagem proposta em alguns cenários de emergências mundiais. Segundo o Relatório Riscos Globais 2019, elaborado no Fórum Econômico Mundial, os riscos ambientais estão entre os principais riscos por probabilidade e impacto para os próximos anos [World Economic Forum 2019]. No relatório, são citadas as condições meteorológicas extremas e a aceleração da perda de biodiversidade como algumas dessas emergências.

As condições meteorológicas extremas incluem comportamentos inesperados ou incomuns, acima da média, em séries de medições históricas relacionadas ao clima. Como exemplos, pode-se citar ondas de calor ou frio, ciclones, tempestades, entre outros [Neu et al. 2012; Tippett 2018]. Como consequências, as condições extremas podem causar desastres em diferentes graus, devido à frequência, intensidade e vulnerabilidade das populações afetadas [Tippett 2018].

Outro tema abordado foi a aceleração da perda de biodiversidade. No Brasil, por exemplo, as espécies nativas que ocorrem em biomas tais como Amazônia, Cerrado, Mata Atlântica, Caatinga e Pantanal apresentam grande variabilidade biológica, morfológica, anatômica e química, muitos dos quais ainda desconhecidos pela ciência. Contudo, a flora brasileira vem sendo progressivamente destruída [JBRJ 2019; Portal Brasil 2012; Souza Brito and Souza Brito 1993; Varanda 2009] e, deste modo, muitas espécies vegetais estão em fase de extinção, e outras já desapareceram antes de serem conhecidas pela comunidade científica [Portal Brasil 2012].

Os cenários descritos apresentam relevância para a comunidade científica (obtenção de dados fidedignos relacionados aos riscos ambientais) e para a sociedade. Deste modo,

espera-se conseguir uma participação do público em geral nas aplicações, com o envio, interpretação e anotações de tarefas relacionados aos temas.

### 3.4. Avaliações

Após a aplicação da abordagem em cenários de emergência, serão realizadas avaliações a fim de analisar a abordagem proposta. A avaliação será dividida em três etapas: (i) planejamento, onde serão definidos os participantes, o tema, escopo e tarefas do estudo; (ii) execução, onde serão coletados dados e registros solicitados por pesquisadores aos cientistas cidadãos. Para isso, serão observados dados referentes à implantação do software implementado em um servidor web; (iii) análise de resultados, onde os resultados serão avaliados de forma quantitativa e qualitativa. Para as análises quantitativas, serão coletados o tempo de realização de tarefa e quantidade de tarefas realizadas, para definir métricas, como a complexidade das tarefas. Como análises qualitativas, serão observadas a adaptação dos participantes, usabilidade do sistema e qualidade dos dados gerados pelo público.

## 4. Discussão e conclusões

Neste artigo apresentamos uma proposta de abordagem genérica para gestão de tarefas complexas em projetos de ciência cidadã. Com base na revisão da literatura, identificamos alguns problemas nessa área, como: (i) escalonamento de tarefas complexas em projetos de ciência cidadã; (ii) escolha da *crowd* certa para cada tarefa; (iii) validação das informações enviadas pela *crowd*; (iv) garantia da propriedade intelectual das contribuições da *crowd* e dos resultados obtidos. Por meio da execução deste projeto, esperamos contribuir para o campo da Ciência da Computação com a criação de uma abordagem para envio, interpretação e anotações de dados de cunho científico. Além disso, esperamos que a sociedade em geral compreenda a importância das pesquisas científicas para o entendimento de cenários de emergências ambientais.

## Referências

Amer-Yahia, S. and Basu Roy, S. (2015). From Complex Object Exploration to Complex Crowdsourcing. In *Proceedings of the 24th International Conference on World Wide Web.*, WWW '15 Companion. ACM.

Beer, J. De, McCarthy, I. P., Soliman, A. and Treen, E. (mar 2017). Click here to agree: Managing intellectual property when crowdsourcing solutions. *Business Horizons*, v. 60, n. 2, p. 207–217.

Berwanger, O., Suzumura, E. A., Buehler, A. M. and Oliveira, J. B. (dec 2007). How to critically assess systematic reviews and meta-analyses? *Revista Brasileira de Terapia Intensiva*, v. 19, n. 4, p. 475–480.

Biolchini, J., Mian, P. G., Natali, A. C. C. and Travassos, G. H. (may 2005). Systematic review in software engineering. , Technical Report ES. System Engineering and Computer Science Department COPPE/UFRJ.

Blohm, I., Leimeister, J. M. and Krcmar, H. (2013). Crowdsourcing: how to benefit from (too) many great ideas. *MIS Quarterly Executive*, v. 12, n. 4, p. 199–211.

- Brabham, D. C. (2013). *Crowdsourcing*. Cambridge, Massachusetts ; London, England: The MIT Press.
- Cohn, J. P. (2008). Citizen science: Can volunteers do real research? *BioScience*, v. 58, n. 3, p. 192–197.
- Cui, Q., Wang, S., Wang, J., et al. (2017). Multi-Objective Crowd Worker Selection in Crowdsourced Testing. In *SEKE*.
- Cullina, E., Conboy, K. and Morgan, L. (jan 2016). Choosing the Right Crowd: An Iterative Process for Crowd Specification in Crowdsourcing Initiatives. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. . IEEE.
- Franzoni, C. and Sauermann, H. (1 feb 2014). Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, v. 43, n. 1, p. 1–20.
- Gil, A. C. (1999). *Métodos e técnicas de pesquisa social*. Atlas.
- Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In: *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. p. 105–122.
- Hirth, M., Hoßfeld, T. and Tran-Gia, P. (2013). Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelling*, v. 57, n. 11–12, p. 2918–2932.
- JBRJ (2019). Centro Nacional de Conservação da Flora - CNCFlora. <http://cncflora.jbrj.gov.br/portal>, [accessed on Aug 12].
- Kullenberg, C. and Kasperowski, D. (14 jan 2016). What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE*, v. 11, n. 1, p. e0147152.
- Neu, U., Akperov, M. G., Bellenbaum, N., et al. (19 sep 2012). IMILAST: A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms. *Bulletin of the American Meteorological Society*, v. 94, n. 4, p. 529–547.
- Portal Brasil (2012). Flora é reconhecida como uma das mais importantes do mundo. <http://legado.brasil.gov.br/noticias/meio-ambiente/2012/04/flora-brasileira>, [accessed on Aug 12].
- Souza Brito, A. R. M. and Souza Brito, A. A. (may 1993). Forty years of Brazilian medicinal plant research. *Journal of Ethnopharmacology*, v. 39, n. 1, p. 53–67.
- Staffelbach, M., Sempolinski, P., Kijewski-Correa, T., et al. (2015). Lessons Learned from Crowdsourcing Complex Engineering Tasks. *PLOS ONE*, v. 10, n. 9, p. e0134978.
- Tippett, M. K. (10 dec 2018). Extreme weather and climate. *npj Climate and Atmospheric Science*, v. 1, n. 1, p. 1–2.
- Varanda, E. A. (2009). Atividade mutagênica de plantas medicinais. *Revista de ciências Farmaceuticas basica e aplicada*, v. 27, n. 1, p. 1–7.
- World Economic Forum (2019). The Global Risks Report 2019. *Global Risks 2019*. <https://wef.ch/2gN6SOd>, [accessed on Aug 12].
- Zou, G., Gil, A. and Tharayil, M. (dec 2014). An agent-based model for crowdsourcing systems. In *Proceedings of the Winter Simulation Conference 2014*. IEEE.