

# Solve for Good: A Data Science for Social Good Marketplace

Rayid Ghani, Lisa Green, Alberto Bengoa  
Center for Data Science and Public Policy  
University of Chicago  
rayid, lisa, alberto@solveforgood.org

Mohak Shah  
LG Electronics  
mohak@mohakshah.com

## ABSTRACT

Solve for Good is a platform for social good organizations to pose their problems that need data intensive help, and for volunteers to help solve those problems. Once the projects are submitted by the organization, they go through a scoping process (done by scoping volunteers and guided by our Data Science Scoping Process). Once a project scope is finalized, it becomes available for data science volunteers to start working on. The finished work is reviewed by a QA team consisting of volunteers and staff of the organization that submitted the project.

Solve for Good comes out of our experience working with government agencies, non-profits, universities, volunteers, professionals, students, and the private sector over the last several years. We repeatedly get contacted by governments, non-profits, and other social good organizations asking for help with data projects. We also have smart, passionate individuals who contact us offering their help, often in a volunteer capacity, on weekends, evenings, or for a few days or weeks. Solve for Good is our attempt at linking these two. We are just starting out, and looking for help in doing this better, and getting feedback from you. Join us at <http://www.solveforgood.org> as a volunteer to help solve problems, as an organization to submit problems, as partners to help us expand the platform and provide resources to run it, and as corporations or foundations to loan volunteers and donate resources used in solving these problems.

## 1. THE PROBLEM

Social Good organizations from the public and private sector have been maturing in their data collection abilities over the past few years. They are collecting more and better data, and are beginning to invest time and effort into analyzing it to make better decisions. At the same time, data scientists are clamoring to help them in nontraditional ways - by attending hacknights and weekend hackathons, offering to volunteer their skills, and providing in-kind help. Unfortunately, most of these efforts do not result in demonstrable added value and social impact without a lot of hand-holding and support from external organizations. The main reasons for this lack of significant and sustained impact are the following:

1. Non-profits and government organizations have no systematic and scalable way of accessing and using a large pool of volunteers who can help them in a reliable, consistent manner. Even when they find themselves with such access, they don't have a way of translating their needs into a well-defined technical or data science problem. Problem owners may un-

derstand their needs but don't speak the technical language, while the volunteers may be data science experts but don't understand the business context or have a deep understanding of social good problems.

2. Potential volunteers do not have a way of connecting with organizations who need their time and skills. They may offer help and expertise, but often get turned down because the organization does not understand how data and technology could help, because they lack the resources to do the necessary prep work, or because the problem identified by the volunteer is not real or urgent. Further, Individual volunteers rarely have a complete skill set sufficient to solve a problem on their own, in addition to a lack of operational context.
3. In the few cases where all these problems are overcome and something does come out of such a collaboration, the resulting system is rarely maintained, extended, or even put into production in the first place. This is typically because there is seldom an effective hand-off from the initial team of volunteers to the people in the partner organization with the skills to use and maintain it (if they even exist).

In discussions with hundreds of government agencies and non-profits, and in experiences running data science for social good programs, we have found that there are many volunteers who want to offer part-time help to solve technology and data-related problems. Yet, organizations, are unable to use these resources effectively. There is no consolidated list of people to go to and look for specific skills. They do not have the resources to manage the overhead in coordinating volunteer efforts. They lack the resources to translate the business problems into technical problems, or to break them into self-contained modules that could be handed over to volunteers. The result is frustration on both sides.

## 2. THE OPPORTUNITY

If we could solve this problem effectively and in a scalable manner, it would:

- Create long-term connections between social good organizations and volunteers.
- Create connections among the volunteers that can be sustained across other efforts, thereby building a productive community that is passionate about using technology and data for social good.
- Solve important problems requiring skills that these organizations struggle to recruit.

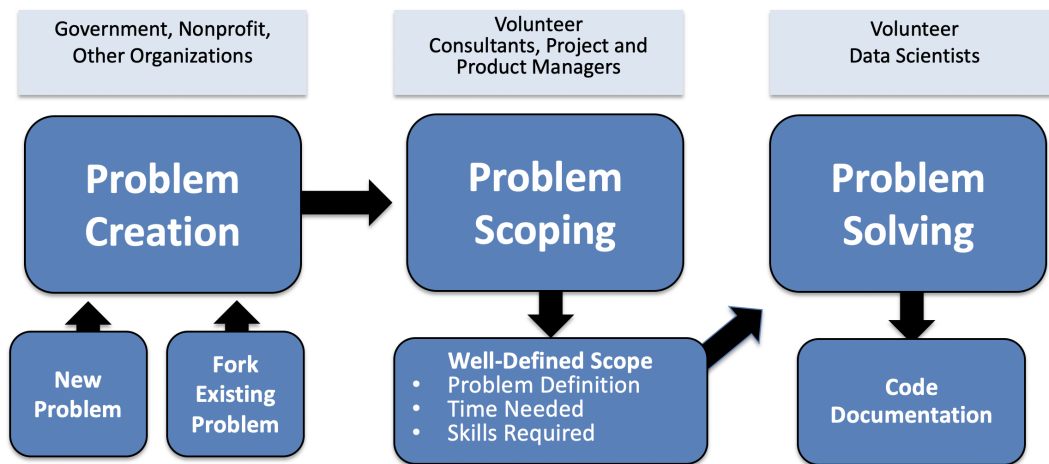


Figure 1: Solve for Good Workflow

- Give high social impact organizations a better internal understanding of how to formulate, scope, and solve these types of problems.

### 3. SOLVE FOR GOOD: A DATA SCIENCE FOR SOCIAL GOOD MARKETPLACE

Solve for Good is an online platform for social good organizations to post projects they need help with, where volunteers with expertise in data science project scoping can ask questions and turn those needs into well-defined problems, and where data scientists can team up to solve those problems collaboratively. The community will serve as a marketplace, a discussion board, and a collaborative workspace. It has the following steps in the workflow:

#### 3.1 Problem Creation

Solve for Good enables organizations to use data-driven methods to better achieve their missions. Governments and non-profits can have a low-overhead way to post problems and to describe their context and potential impact. They can post their project to connect with skilled volunteers who can:

- Understand their goals and help scope projects.
- Work with them and other volunteers on those projects.
- Review the solutions and give feedback.
- Help them understand and operationalize the results.

We are adding additional features such as allowing other organizations to vote on the importance and urgency of a given problem, or can fork it into a related problem that they are facing.

#### 3.2 Problem Scoping

Problem experts can engage in discussions to scope out the problem, detail the requirements, and come up with the metrics used to evaluate the solution. Problem scoping and definition often require a lot of experience in the domain of the problem, in scoping projects across domains, and an understanding of what makes a good data science project. These are not skills that are typically taught in academia and require practical experience. We have spent the past

several years developing project scoping methodologies<sup>1</sup> that are available open-source. We also plan to hold workshops and webinars to train volunteers in this process. An example of such a training workshop is a "Scopeathon" run at the University of Chicago every year.<sup>2</sup>

#### 3.3 Problem Solving

Figure 2 illustrates the project page that shows which stage the project is in. Once the project has been scoped, data science experts with the appropriate skills can collaborate and work individually or organize teams to tackle the tasks. We use Github to maintain the associated code repository, wiki, and mailing list. This minimizes the need to develop the infrastructure from our end, offers free publicity, and encourages open-source development. The finished work is reviewed by a QA team consisting of volunteers from our platform and staff of the organization that submitted the project.

In addition, we plan to integrate with existing lightweight collaboration platforms such as slack and trello. We will also plan to partner with cloud platforms (AWS, Azure, Google Cloud) to provide shared technical and data infrastructure to the teams working on these projects.

### 4. PARTNERSHIPS

We are partnering with the Center for Data Science and Public Policy at University of Chicago<sup>3</sup>, Data Science for Social Good Fellowship<sup>4</sup>, Data Science for Social Good Foundation<sup>5</sup>, Carnegie Mellon University, and plan to extend the partnerships to organizations such as Datakind, DrivenData, and Code for America as well as other civic tech, meetup, and industry groups to get problems as well as volunteers to sign up and actively participate in this marketplace. We are also in discussions with specific government and nonprofit organizations to join the effort.

### 5. HOW TO GET INVOLVED

<sup>1</sup>Data Science and Public Policy Project Scoping Guide - <http://www.datasciencepublicpolicy.org/home/resources/data-science-project-scoping-guide/>

<sup>2</sup><http://southsidecivic.net/scopeathon>

<sup>3</sup><http://www.datasciencepublicpolicy.org>

<sup>4</sup><http://www.dssgfellowship.org>

<sup>5</sup><http://www.datascienceforsocialgood.org>

# Data Integration - Matching records

Metropolis Department of Public Health

11 followers.

Match records of people coming from disparate data sources from various health services that the city provides



♥ Health

Project progress bar: New (active), Scoping, Scoping QA, Staffing, In progress, Final QA, Completed. Status: This project is waiting to be scoped.

Information | Scope | Discussion

Figure 2: Project Page

We are just starting out, and looking for help in doing this better, and getting feedback from you. Join us at <http://www.solveforgood.org> as a volunteer to help solve social good problems, as an organization to get help, as partners to help us expand the platform and provide resources to run it, and as corporations or foundations to loan volunteers and donate resources used in solving these problems. Solve for Good needs a global community of volunteers that are passionate about using technology and data for social good. If you have project scoping, management, or data science skills and a passion for social impact, join us to:

- Apply your skills for work that matters

- Gain data science experience and build your portfolio of work
- Showcase your skills and increase your visibility

## 6. ACKNOWLEDGEMENTS

We are grateful for the support of SIGKDD Social Impact program, the team at the Center for Data Science and Public Policy and the fellows at the Data Science for Social Good Fellowship for helpful discussions and beta testing, and for the KDD attendees in 2018 for beta testing the website.