MWSUG 2018 - Paper BL147

Data For Good as a Community Service Project at Work

David J Corliss, Peace-Work, Plymouth, MI

ABSTRACT

Many businesses large and small support volunteering within the community, often sponsoring volunteer work days where employees can attend a well-organized activity for a group doing good work in the community. Today, Data For Good volunteering an opportunity to supply a critical skill for charitable organizations making a difference. This presentation is designed to help business leaders set up a Data For Good project as community service and team building activity, with best practices for finding a good organization to support, recruiting participants, managing the volunteer work day, and sharing the story with the wider community. Descriptions of successful projects and how to do them include building a membership database, data dive events to develop recruiting models for community service groups, seasonal optimization of resources (e.g., food pantries), and others. Practical, proven processes are presented to make Data For Good your company's next community service project.

INTRODUCTION – DATA FOR GOOD

In many professions, pro bono work – literally, 'For Good" – is an ordinary, even expected part of every career. Physicians may take certain cases without charging for their services and volunteer at a free clinic. Accountants will volunteer for charitable organizations. Attorneys have a long history of pro bono activity, with the American Bar Association effectively requiring no less than 50 volunteer hours each year to remain a member in good standing. Other professional organizations such as the American Medical Association and American Institute of Architects encourage volunteer work while setting no hard and fast expectations, preferring instead to encourage participation and listing benefits of pro bono work in their standards of ethical practice while leaving the details to individuals.

In recent years, volunteer work in statistics and data science have become much more common. Often called "Data for Good" (D4G), this growing movement has been embraced by individuals, companies, and professional associations. A number of organizations have sprung up to design and manage statistical volunteering projects, recruiting volunteers and connecting them with projects and organizations that will benefit from analytic support. This addresses a prominent need in organizations pursuing good causes: they often have data but lack statistical skills and resources to perform analytics and cannot afford to hire highly skilled consultants. In these situations, statistical volunteers have the opportunity to make a real difference if the world doing the things we love the most.

One important avenue for Data For Good projects becoming much more widespread today is statistical volunteering through work. Companies often encourage employees to volunteer in the community where they work, with many offering time off for teams or individuals volunteering for local organizations. Common projects include cleaning up parks, volunteering at a school for a day, working in community food banks, and construction projects for Habitat For Humanity. While supported by the employers, which provide critical infrastructure and logistics, individual projects are usually proposed and led by individual employees. This makes Data For Good one kind of project many of us can consider for the next Team-Building service project where we work.

BENEFITS OF STATSICAL VOLUNTEERING

Statistical volunteering offers many advantages. In addition to using our technical skills to support causes and organizations we care about deeply, pro bono also work offers many benefits for career development. Staff early in their career can gain important practice and learn new analytic techniques. Volunteer consultants are able to broaden their experiences and gain practical experience in new techniques – something that can be very limited in professional situations, which often focus on the same handful of methods used over and over again for a commercial product. Statistical volunteering is often the best way to try out and enjoy a wide variety analytic tools and emerging technology. All of these are

good reasons to look for statistical volunteering on the resumes of job candidates and promote the practice among employees as one aspect of career in statistics and data science.

Statistical volunteering often pairs more experienced data scientists and analysts with those having less experience. This naturally leads to mentoring relationships than benefit both parties. All participants benefit from networking with people from different backgrounds, skill sets and work situations. More junior volunteers gain practical experience, the chance to work directly with the people needing the analysis – a quality often very restricted at work, where newer team members can rarely see anyone who directly uses their analysis. Statistical volunteers gain presentation skills, practical working experience and confidence in their work. As experienced is gained, taking a larger role on volunteer projects creates opportunities for leadership, furthering career development.

While Data For good offers substantial benefits for both businesses and for individuals' career development, the most significant impact of statistical volunteering will never appear on a balance sheet, product, or resume but in the lives impacted by the projects we perform. As statistical and data science professionals, the need for what we do every day on the job is great but volunteers are few. Data For Good applied to a community service project at work is growing rapidly because it is so much more than good business sense: it makes a real and tangible difference in our profession, our society and our world.

BEST PRACTICES

Company volunteer programs need a clear set of ground rules. Projects and the organizations receiving support should be approved through the HR department or, in the case of smaller companies, the Operations lead. Projects must not interfere with or delay regular work for company customers. Companies will often choose the cause they want to support and then seek employees to volunteer, generally preferring activities that support important needs in the community. As the company will want to let others know about the good they are doing in the community, communication with HR or operations, an employee newsletter and even the outside press might be a requirement. Employee volunteer projects often serve as team building activities, so it's a good chance several of your co-workers will be involved. Many companies will give a limited, specified amount of time off for volunteering through work, often one day a year or sometimes two. As the company is performing the work as a charitable activity, careful documentation is needed for tax purposes. This is usually done by a Team Captain who submits a request for an employee volunteer activity, helps recruit volunteers, records attendance in hours for each participating employee, and reports by to HR or Operations about how everything went.

One advantage of volunteering through work is the company often makes available analytic resources and software the volunteers use every day. The familiar analytic environment and software at work usually becomes the platform for pro bono projects as well, maximizing the efficiency and effectiveness of statistical volunteering.

Optimally, Data For Good volunteering through work should follow a clear, well-documented process carefully reviewed and approved by all appropriate teams and levels within the company. The critical point to keep in mind is that, as far as receiving support from the company is concerned, it makes no difference if the donation is Data For Good support: the exact same standards and process needs to apply to all charities and community partners receiving any kind of support, whether the substance of the donation is a grant, infrastructure support, or Data For Good volunteers on Company time. All are treated the same. Here are some points to consider:

- Companies should have in place a standard, documented process for vetting and approving organizations to receive charitable support. Most companies already have this – if not, one should be developed.
- The process should be managed centrally by an operations group such as HR. Different analytic teams can lead projects but need to work with HR to assure consistency and adherence to company rules and procedures.
- Organizations receiving Data For Good support should be carefully vetted using the standard process.

- Where possible, consider first those charitable organizations where a relationship already exists. Look at the community service organization, school, soup kitchen, animal shelter, or other organizations the company already supports with community volunteering opportunities and look at how data and analytics can do more help them in their important work.
- As with any company activity, the needs of the company come first. Data For Good volunteering must not interfere with company assignments or deadlines. Because of this, it is important to design volunteer projects with flexibility in terms of staffing and leadership, taking into account that some people will not be available at times due to pressing company commitments.
- Ideally, make the project a team building activity. Assemble teams of data and analytics co-workers to work together on the project.
- Where possible, link the project with analytics R&D. Consider how the volunteer project can be a way to try out new tools, software, and methods that can help the charitable organization while deepening the skill set of your teams.
- Just as with any other community service team building activity, carefully document the project and share it with others. If you have a media person or department, get them involved in the planning stages. Get the write up of your project in the company newsletter or email. Consider other media opportunities, looking into newspapers, TV or radio reporters looking for a story on people doing good in the community.
- Involve HR to see how past projects can help with recruiting talent in the future.
- Encourage employee feedback and use it to design new D4G community service projects inn the future.

VOLUNTEERING THROUGH WORK – EXPERIENCES AND OPPPORTUNTIES

While the same basic requirements and process can be applied, team-building Data For Good activities can take different forms. Here are some examples of successful community projects with observations on developing something similar for your company or organization.

THE CLOUDERA HACKATION

Companies often encourage employees to volunteer in the community where they work. Those engaged in data science and technology have sponsored projects where groups of employees apply the same skills they use on the job to support important causes on a pro bono basis. The number of companies supporting the Data for Good movement with community service projects continues to grow. In addition to large corporations with substantial HR departments to encourage and coordinate statistical volunteering, many small consulting companies and Analytic Services Providers are involved and often have the most flexibility in setting up Data for Good volunteering programs at work.

One of the most familiar kinds of projects is a Hackathon. This kind of activity meets many of the most important needs for a company:

- Just a few people are needed in advance to set it up
- The primary activity happens all on one day, scheduled for the event
- Allows for work schedules to come first: if a given person has deadlines to meet, that comes first and they can participate another time: the project won't fail if a few specific individuals aren't available on the day of the event
- Functions as a Team Building event as well as a community service project

In both Hackathons and Data Dives, a small team works in advance to identify a problem to address or organization to benefit, schedules the event and lets people know – usually by company email – and prepares a dataset. Teams of co-workers sign up to participate. The community service aspect and the

ability to do something people likes to do – data science – tends to result in greater [participation among employees.

In May 2016, Cloudera held a successful community service hackathon that compiled data on the spread of mosquito-borne viruses such as Zika. In addition to their own, Cloudera chose to invite participation teams from outside the company as well. They also worked with a university partner, the Texas Advanced Computing Center at the University of Texas, Austin, who supported the project with their large Hadoop cluster. More than 50 people participated on the day of the event, leveraging diverse skills sets, compiling data, applying data science to investigate mosquito-borne viruses, and increasing awareness of this problem. A full description if the project can be found on Cloudera's website here: https://vision.cloudera.com/data-citizens-gather-at-hackathon-in-austin-tx-to-fight-mosquito-transmitted-disease/. The Cloudera hackathon shows how a small project team can work can set up and promote a project that brings in strong participation by developing the infrastructure while leaving the data science to the participant team. Cloudera also leveraged an existing partnership - in this case, the University of Texas Advanced Computing Center – to bring in a needed resource. Participation is enhanced by holding an event on a single day, giving participants and managers alike the clearly defined, realistic, and time-bound commitment needed for successful project.

TERADATA PARTNERS WITH DATAKIND

One way companies can get started in Data For Good service projects is to partner with an existing organization with extensive experience. A great example of this happened with Teradata partnered with DataKind, a leading Data For Good organization, for a Data Dive in 2014. Teradata's D4G program, Teradata Cares, encourages and supports employees giving back to the community using their specialized skills. They worked with DataKind on a data dive at Teradata's major annual conference, now called Teradata Analytics Universe (formerly known as the Partners Conference). DataKind and the Teradata project team worked together to set up a data dive with data from four different community organizations. The data dive became one of the activities available for attendees at the two-day conference. You can read a DataKind blog about the event here: http://www.datakind.org/blog/teradata-partners-datadive/.

DataKind is one of the most prominent organizations in in the Data for Good volunteer movement. Founded in 2011 by data scientist Jake Porway, DataKind has grown rapidly to become a large, global organization making a great impact for social change, with thousands of volunteers working on dozens of projects. The organization combines the resources of a large number of volunteers with a smaller cadre of professional staff. With substantial financial support from foundations, grants and other donations supporting a multitude of projects, DataKind is a powerful engine for using data and analytic science to make a difference in communities and around the world. DataKind often collaborates with corporate data science efforts. Volunteers often work closely with a project team centered on a particular issue or organization. DataKind has a long history of collaboration with the SAS community and their founder, Jake Porway, was a keynote speaker at SAS Global Forum in 2015. Employees at the SAS Institute have donated to DonorsChoose, an organization support classroom projects by public school teachers, since 2006. DataKind has gotten involved, with volunteer data scientists mining DonorChoose data and providing the results to schools districts, government agencies and other organizations supporting public schools.

The Teradata – DataKind partnership shows how a company can benefit from working with an existing Data For Good organization. It also shows how a data dive makes a great first project (but doesn't have to be!) and can be offered as a Data For Good activity at a conference. One important characteristic of this event is that multiple community service organizations were supported, with teams or individuals able to choose which the want to help.

AN ON-GOING PROGRAM: MICROSOFT'S MYSKILLS4AFRICA

Microsoft's MySkills4Africa (https://myskills4afrika.microsoft.com) is an excellent example of a D4G activity through work that is on-going rather than event-based or a single project. Established in 2013, MySkills4Africa offers Microsoft employees from around the world the opportunity to develop practical and affordable technology in developing African nations. The Microsoft program partners with governments, NGOs, commercial enterprises and schools and universities. Characteristic of MySkills4Africa projects is

the opportunity to use advanced skills normally applied only at work and use them in service to others as part of the Data for Good movement. Many MySkills4Africa projects involve longer-term consulting projects working virtually with teams in Africa. Microsoft also sponsors 1-2 week hands-on projects in Africa. One project developed a team of Subject Matter Experts to partner with the Rwanda Ministry of Education. Months of preparation - all done as volunteers, in addition to work responsibilities – culminated in a two week trip to Rwanda to train professionals and small business owners in emerging technologies. At the same time, the employee team worked with Rwanda education officials to establish an on-going program to continue the good work.

INDIVIDUAL VOLUNTEERING

Not all or even most statistical volunteering goes through an organization: individuals working on their own make up a large portion of the Data for Good activities. Often, a personal connection to a group needing statistical help is an important factor. With the proper governance, these opportunities can be leveraged by a company to support community service projects.

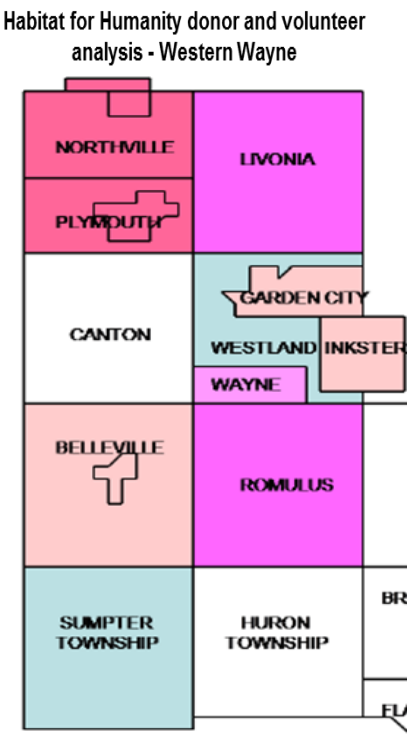
For individual or small group D4G community service projects, the same process applies but it is necessary for the individual proposing the project to perform most of the infrastructure tasks. Especially important is the need for careful vetting of the organization receiving support, making it very advantageous to work with a charitable organization where the company already has a relationship. Thorough documentation of the project is essential. Individual volunteer projects can be less formal, without a definitely timeline, and often can be put on a shelf for a time and re-started later when more time is available.

The nature of individual volunteering and the opportunities it offers can be seen in a project by the author for a local Habitat for Humanity chapter in 2005. After volunteering on a number of construction projects and getting to know the chapter's Board of Directors, the question was raised of how data science can be used to support the organization. While this was an individual project, it came about through work when the data scientist who proposed the project was a co-worker of a person of the organization's Board of Directors.

This community service project was performed by a volunteer for Peace-Work, s cooperative of statisticians, data scientists and other researchers applying analytics to issues in poverty, education and social justice. With a practice focusing on issue-driven advocacy, projects have included education performance metrics, and root cause analysis of homelessness, social media bullying, and other issues. Peace-Work operates by connecting volunteers with organizations and datasets, often from governments and other official sources, to address issues volunteers care about deeply.

While Peace-Work volunteers rely on the software tools of their own choosing, SAS is used in many projects. Peace-Work actively promotes the use of SAS University Edition as a powerful analytic tool with extensive statistical, machine learning and big data capabilities available to their researchers for free.

The Habitat For Humanity project performed database development and analysis chapter of donor and volunteers for a chapter in western Wayne County, Michigan, in the western suburbs of Detroit. Analysis of an anonymized list of donors and construction volunteers revealed most of the support for the chapter came from a small portion of their geographic area, with many cities largely unreached. A cluster analysis binned all of the communities in the chapter's area into one of three groups, based on level of charitable need in the community and their ability to meet it. Each of the three clusters, classifying all of the communities in the area served by Western Wayne County Habitat for Humanity, were matched with the zip code counts – no information that could identify an individual! Looking at the whole area, instead one small part where most of the board



members lived, produced a list of communities to target where the organization had very little presence. This Data for Good project had an unintended consequence – for the good! In addition to finding places to search for more donors and volunteers, it documented that the one place where the local Habitat group was building houses was very poor and weak infrastructure and especially poor schools. This produced a recommendation to build house in a different, neighboring community where the families would receive more support and the children attend better schools – doing more good for the families in new homes even if the land prices there were somewhat higher.

Another individual volunteering project paired a data scientist with a local community service organization providing a number of services, including a peer crisis counseling center. The project began when a statistician volunteering as a math tutor connected with the crisis counseling center on informal, on-going basis. The data scientist was given anonymized weekly totals for different types of contacts and the needs they addressed. This was used study seasonality at the call center. This analysis of variations by week enabled better understanding of future needs to help project required staffing levels.

These projects tell the story of how most individual D4G volunteering happens: a person is already helping an organization in a non-statistical way – working at the public library, walking dogs at the local animal shelter, volunteering at a school, house of worship or community center – any of a hundred things. Companies are advised to develop a list of organizations approved volunteering in advance, as much as possible, and help connect volunteers with community service organizations where the company has an existing relationship. With the proper vetting, process, and documentation, this same individual volunteering model can be applied at companies.

Individual statistical volunteering happens when people with analytic skills know how those skills can be used to help the organizations and causes they already support. It's such a rare skill, so needed by groups across our communities that companies who facilitate connecting volunteers can make a huge impact right where we live every day.

HOW TO GET INVOLVED

There are many excellent organizations in the Data for Good movement that can partner with companies seeking to develop a D4G community service program or provide good examples and thought starters, including

DataKind: http://www.datakind.org/

Statistics Without Borders: http://community.amstat.org/statisticswithoutborders/home

Peace-Work: davidjcorliss@peace-work.org

CONCLUSION

All of these opportunities and many more share a quality common found in Data For Good projects: an organization has data about its activities and the people it supports but doesn't have and can't afford the statistical expertise needed to make their important work in the community even more effective. At the same time, companies both small and large are looking for opportunities to build team spirit by working together on projects that benefit the community. D4G projects through work are a perfect opportunity to use the science we love to make a difference in our communities and around the world.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

David J Corliss, PhD
Peace-Work
davidjcorliss@peace-work.org
www.peace-work.org