

DataDive: an Open data initiative

Nikki Roda and Claire Barco

In October 2011, during the official Fall Study Break at the University of Michigan, first-year School of Information student Nikki Roda spent her time volunteering in her hometown of New York City. At the first ever DataDive, Nikki worked together with other volunteers to analyze and interpret large data sets for local non-profits that lacked the resources to do it themselves. The event was hosted by an organization called DataKind (formerly Data Without Borders). As a result of their participation in the event, the social organizations were able to visualize their impact and reach, enabling them to more easily connect with their donors.

Roda, a passionate data scientist with experience supporting nonprofit organizations and their struggle to use data effectively, envisioned a similar gathering that would assemble representatives from nonprofit organizations, U-M statistics and data sciences departments, and members of the community. Roda, along with her collaborator and fellow SI student Claire Barco, hoped that this DataDive would provide a proof of concept of how groups and individuals can come together to collectively address the data analysis and visualization needs for area nonprofits and local organizations. Unlike previous DataDives, this event was focused on engaging the Ann Arbor and Greater-Detroit area: the purpose would be to promote data literacy to both local social organizations and student volunteers.

The A2DataDive brings together local nonprofits and volunteer community members for a weekend of service to help organize, manipulate, calculate, and visualize data. The model connects nonprofits in need to data science students, practitioners, and enthusiasts from the University of Michigan and in the region in an effort to 1) provide a creative channel to give back to the local community and 2) to build a strong, interdisciplinary community of practitioners who can share knowledge, skills, resources and data to support our broader community needs.

A DataDive provides the collaborative space over a 48-hour period for real work to happen. It starts with the selected nonprofits presenting

the problem they are trying to solve with supporting data to the community of data scientists volunteering their time and skills to the event. A twelve hour day is devoted for the data scientist participants clean, collect, sort, merge, plot, and analyze the data -- providing data analytic resources that the non-profits need to address their problems. Half of the next day is saved for finishing, organizing, and presenting the work that was completed to all the participants. As a connection is established between the non-profits and the data scientists, there are often continued opportunities for these volunteers to support the non-profits in the future.

The first ever A2 DataDive, held in February 2012, featured the datasets of Focus: HOPE and the African Health Open Educational Resources Network. Focus: HOPE works in northwestern Detroit to bridge the racial divide and develop a healthy and happy community for all Detroit residents. Their main focus is the HOPE Village Initiative, taking to heart the phrase "It takes a village to raise a child". They provide a number of different services, including training classes and food assistance for needy families and individuals throughout the Detroit area.

The African Health OER Network (AHON) is a social organization that aims to provide Open Educational Resources for medical and nursing students in resource-constrained schools throughout Africa. Their goal is to supplement the students' education and improve quality of care in their nations in the long-term. From Ann Arbor, they create a network through which individuals and organizations can make and distribute videos, readings, and other educational materials, free-of-charge.

The A2DataDive attracted a wide range of skills and perspectives, and included representation from many of the academic programs that now feature data analysis in their curriculum. More than 70 people contributed their skills and expertise and provided a weekend of service, including alumni, members of the Ann Arbor Chapter of the American Statistical Association, Domino Corporate Headquarters, Detroit, Data Driven Detroit, ProQuest, and even two Code for America Detroit Fellows¹. The event offered learning opportunities for attendees to become more familiar with open source tools and techniques for analyzing data through a series of workshops hosted by a range of experts from across campus.

1 <http://codeforamerica.org/2012-partners/detroit>

There is a growing need for data-based skills across institutions and organizations and this event illustrated the power of peer-teaching and teamwork. This event also showcased the ability to share data and skills across institution/community boundaries and the importance of open practices in today's networked world. These were all values that were shared by one of A2 DataDive's principle sponsors, Open.Michigan², an initiative led by the Office of Enabling Technologies³ at the University of Michigan Medical School. The event aligns well with Open.Michigan's mission to encourage researchers, learners, and instructors to maximize the impact and reach of their scholarly work through open sharing.

This model illustrates how open teaching, learning, research practices can positively impact the activities of a DataDive in gathering, using, and supporting data across disciplines or organizations. As such, all content developed by the participants, including presentations, code, lists of resources, documentation and visualizations, are all licensed under a Creative Commons license and hosted on the Open.Michigan collection⁴. This allows for data scientists, nonprofits and learners across the world to access and use the resources developed during this participatory, local event.

According to event co-organizer Emily Puckett Rodgers of Open.Michigan, "Everyone was very impressed by how a mix of community members, along with U-M students, staff, faculty, alumni, and retirees, were able to take on this challenge and produce results for these organizations with data needs."

"Several participants said 'I learned so much from the other people in my team,'" Rogers added.

After the successful proof of concept, the A2DataDive model was officially adopted by the School of Information's SOIAR (Student Organization for Information Analysis and Retrieval) student group and is co-sponsored by the School of Information and the Open.Michigan initiative.

Based on feedback from the participants, the volunteer organizers of the A2DataDive have expanded the model to include 'data jams' that take place before the datadive. These are designed as light-weight

2 <http://open.umich.edu>

3 <http://sitemaker.umich.edu/enablingtech/home>

4 <http://open.umich.edu/education/si/resources/datadive/winter2012>

learning opportunities to introduce participants and nonprofits to the elements of data assessment, analysis. The A2DataDive organizers partnered with Justin Joque, a data librarian at the University of Michigan's Clark Library⁵, to coordinate and host the first data jams, held in October and November 2012. While more formally structured as workshops, these events still enable group work and result in useful resources for participating nonprofits. The resources developed for these workshops are also captured as open educational resources and hosted on the Open.Michigan collection.

In addition to the data jams, organizers have also partnered with the developing School of Data. Based on the Peer to Peer University (P2PU) community and platform, the School of Data seeks to address the growing need for professionals, students, and citizens to increase their capacity to understand, harness, and manage data. This concept and the DataDive model are very closely aligned in goals and implementation.

In the summer 2012, the School of Data organizers reached out to the Open.Michigan initiative and the University of Michigan to partner in developing content to support this concept. The P2PU platform⁶ and the School of Data emphasize experiential learning and offer a chance for the University of Michigan's School of Information to engage in setting frameworks for a new kind of learning. Data divers from the School of Information can use the P2PU platform to capture but organize the resources and output that get created in data jams and the A2DataDive. By using this structure and considering how we share our process of learning, the A2DataDive organizers can create 'challenges' and 'courses' to guide others through our learning process.

These new partnerships illustrate the adaptive, peer-based nature of this model. When individuals, organizations, and groups come together to solve real-world problems and go about it using open models of participation, data sharing, and publish their results in accessible, adaptable ways, they can be responsive to community interest, need and new opportunities. In this respect, the A2DataDive model represents an emerging standard for activism and learning that is based on an open infrastructure.

5 <http://www.lib.umich.edu/clark-library>

6 <https://p2pu.org/en>

Portions of this article were adapted from a white paper prepared for the Dean of the University of Michigan School of Information, the article “A2DataDive brings community together to analyze data for nonprofits,”⁷ and a blog post “A2DD’s New Partnership with School of Data”⁸. All resources are licensed under a Creative Commons Attribution 3.0⁹ license.

Nikki Roda and Claire Barco

Email: a2datadive-team@umich.edu

7 <http://umhsheadlines.org/05/a2datadive-brings-community-together-to-analyze-data-for-nonprofits>

8 <http://a2datadive.org/new-school-of-data>

9 <http://creativecommons.org/licenses/by/3.0>