Adopt-a-Data-Scientist: A Blueprint
Prepared by Tim Hwang, Group C, FOCAS 2012, The Aspen Institute | August 7, 2012

The Problem With Data

While a great deal of time has been spent in trying to improve the *supply* of data to the public, considerably lesser effort has been expended in trying to design strategies to foster and grow *demand* for the data. This is a significant issue: the lack of a demand suggests that some of the main purposes at the core of open government -- better transparency, higher accountability and civic engagement -- might not be achieved merely by releasing the data and making that the end of the story. The lack of demand also suggests that it will be difficult in the long-term to build a large, supportive constituency around these issues. Both of these hinder, if not entirely handicap, the movement.

There are two parts to this. Part of the problem emerges from the requirements for working with data: even if a member of the public is interested in investigating a question, the costs of cleaning and processing data might be beyond the time and technical skills of a general member of the public. Another problem is data literacy: people might not know what questions to ask of the data that is available that might be interesting or relevant to them.

Moving Forwards

The solution to this problem cannot and should not be top-down. Given the broad range of constituencies in even relatively small geographic areas, it will be unlikely that a central authority will be able to correctly deliver data and insights which are relevant to every possible group (including those not regularly expressing their preferences online). Even if they could, such a central authority would be unlikely to be able to successfully meet the constantly shifting demands of the population for data. The nature of data is often *contextual* -- one is interested in it only as one needs it. To that end, "platforms" and "applications" which have been a focus on attention in recent years fail to capture the more frequent interest in data which is transient and one-off.

The correct approach is to start from strength: the demographic of people who already desire data-based insights but lack the time and technical skill resources to work with the information. These individuals combine two important characteristics: they are likely to be familiar with topics of local interest, and they already have a demand for data that could be converted to a willingness to pay for an answer to a question they have if the cost were to become low enough.

Our approach to those who are not so engaged with data should be linked with this first group. Rather than proselytize data writ large as something the general public should be interested in, the proper method is to teach by example. People become interested in data when it regularly and consistently provides them with insights which are immediately relevant and useful to *them* day-after-day, week-after-week. They do not

become engaged because someone tells them "data" is important. They do not become engaged because they happen to see a passingly interesting data visualization.

The Adopt-a-Data-Scientist Program

The Adopt-a-Data-Scientist (ADS) Program is design to foster grassroots interest in the possibilities of data, and to allow the creation of localized centers of demand for data.

Our proposal: promote the creation of local, self-organized groups of twenty members who will collectively "adopt" a local data scientist, data designer, or statistician. Each individual "adopter" commits to donating $100 per month to form a monthly $2,000 stipend which goes to pay the living expenses, research work, and training expenses of their data scientist. These scientists are to be recruited from the communities themselves, and are required to investigate questions of interest directed to them by the members collectively. These outputs might be reports, visualizations, or even just dumps of data. The projects might only require a day of research to complete, or a number of months. Whatever is needed. The scientists will serve at the leisure of the members from month-to-month, with longer contracts being signed if desired.

Anyone will be permitted to start an ADS Program in their area without having to require permission: they need only bring together the group of funding members. Organizationally, these groups will operate independently of one another, linked perhaps by some lightweight infrastructure like an e-mail list or a website from which data scientists might be matched to adopters and new findings might be attractively published.

The Data Residency

A beneficial outcome of such a project is human capital. The ADS Program effectively generates a grassroots network of data scientists who, through their consistent and dedicated service to their local communities, are deeply in touch with the demand side of the data equation. Collectively, one begins to generate a picture of the types of questions that might be relevant across communities, and the problems they face in acquiring the data they are most interested in. These are problems of general interest, and the people best suited to solve these problems are the data scientists themselves.

To that end, the ADS Program will include a three month residency program held once a year, eligible to all ADS adoptees who have served for a certain length of time with an adopter group. During this time, all participants will have room and board (as well as internet) covered, and are tasked with dealing with tackling a general problem of interest to all of the ADS groups. Unlike a traditional hackathon, the aim will be to build durable solutions that can be generally distributed beyond the ADS program.

In the initial phase, this residency will be funded by the support of a non-profit or consortium of non-profits. Once the number of local adopter groups reaches a certain scale, this could potentially be self-funded by the organization itself.