### **Concept Note: The Data4Good Center**

**Edward G. Happ, June, 2017 – version 6**

**1. Purpose**

The purpose of this document is to propose creating an Information Center at UMSI that is focused on technology-related and other data relevant to the nonprofit community and the citizens it addresses. There is an internal and an external perspective for this: *internally* in the sense that enabling more evidence-based decisions about the use of technology in nonprofit organizations can have positive, mission-moving impact on the work these institutions are engaged in; *externally* in the sense that a body of accessible information can serve the research communities, student projects, and ultimately the vulnerable citizens who seek information about their contexts.

**2. Background**

During the meetings at UMSI in June 2016, we discussed a number of projects I’ve undertaken at the International Federation of Red Cross and Red Crescent Societies (IFRC) and at NetHope that provide a base for a number of data analysis initiatives that UMSI could further develop. A Data-for-Good Center could serve a broader audience of those working on, supporting and interested in international humanitarian concerns.

**3. The Research Areas**

Five data products and related indices would provide the potential starting points:

1. *Crisis Technology Trends* – For NetHope, I have been tracking the technologies used in a variety of categories, in disaster relief response since the Indonesia Tsunami of 2004. These trends are important for identifying major shifts in technology and for providing an evidence base of successful applications of IT in crisis management.
2. *The IT Agility Index* – In times of rapid change or organizational shifts, a key IT strategy is to increase agility, or the ability to change more quickly as the organizational context changes. Starting in 2014, we developed an index to measure agility in IT. First we broke projects into smaller phases that delivered some value to users, and then we tracked the project phases delivered in less than 90 days divided by the total project phases completed. Tracking this agility index over time can be a good proxy for changes in IT agility. Applying this to the whole NGO sector would provide a new and important benchmark.
3. *IT Relevancy Index* – One way to track relevancy of IT in the NGO context is to measure the IT project portfolio against the IT Strategy Pyramid[[1]](#footnote-1). A baseline survey was conducted with NGO IT leaders in 2014. The goal was to determine the percent of projects and project budgets for each of the strategy pyramid’s four levels (the Pyramid Mix). The study confirmed the 80-20 rule of IT spending and time for the bottom of the IT pyramid, for a simple relevancy index. Tracking this index over time and among a broader group of NGOs was provide a strategic IT benchmark for NGOs.
4. *Local/Rural App Trends* – One of the items we tracked at IFRC were the number of IT applications and IT-related projects that originated in field and department programs that were outside the IT purview. Traditional called “shadow IT”, this became an important source of innovation and opportunity to take local successes to scale and avoid redundant time and cost on similar initiatives across a global organization. Tracking this more broadly would become an additional benchmark for NGO IT departments and could become an important source of IT trends.
5. *Big Data Trends* – Access to large external data sets and related analytics is only possibly to the largest organizations with the staff and budget to support it. Google’s work with flu trends is one example[[2]](#footnote-2). Providing analytics on technology and data trends would be useful for smaller NGOs. Ultimately providing data and analytics to vulnerable populations may become an important extension of this work, to the extent that the technology and technical knowledge grows among such communities.

**4. The Audience**

The needs and potential may be segmented into three audiences:

1. *The large, international org's* (like UN agencies, RC members, etc.) can afford to do their own data collection and analytics. They will however benefit from the larger data sets available if they pool their information. This is the goal of Humanity Hub that OCHA and the Dutch are building. Key obstacles are legacy methods, organizational hubris and report-oriented data, rather than comparable time series.
2. *The small NGOs* (less than $10M in revenue; developing country local NGOs), cannot afford the data analytics, for even internal data. This is the largely unserved market. As noted above, the external data (big data) and taking advantage of trends (and benchmarking) are beyond their reach. The concept of a Center of Data at UMSI could most benefit this segment. This could draw from OCHA, but does not interest them (I asked).
3. *The medium NGOs* fall somewhere in-between. They may not have the critical mass of data themselves, but would pay a modest subscription fees to access to good research data.
4. *Additional audiences* – these include Researchers, Granters, Students and ultimately vulnerable citizens themselves.

**5. The Proposal – the Data4Good Center**

Creating a center at a major university’s school of information for this work is a natural fit, both in terms of a focused interest area among faculty and students, and as a service to the nonprofit sector. Students in each of the levels would be an important partner for gathering data, analyzing it and presenting it to the audiences. A critical mass of data would also be a source for research in the area. Finally, such a center would help build the brand of a school that sees itself at the intersection of people, information and technology.

I propose to call this the *Data for Good Center,* or Data4Good for short. It would start small, with a technology focus. The first three data “products” above would be the initial phase. My sense is to start with the under-served NGOs and then move upscale.

Later phases could build on this base in the following ways:

1. Trend forecasts: from existing cross sector data and big data context
2. Unique benchmarking data like IT project portfolio mix and project agility index.
3. Standish-like NGO large project data
4. Consumer versus enterprise technology mix
5. Aggregate emergency response assessment data
6. Technology impact reports and benchmarks to grant organizations
7. IT assessment data; like IT Health-check, ICT capacity
8. Time series of open data sets and “PDF report-scraped” data for analysis

We would thus begin with data about technology, since it's a mutual core competency, and then look to market data (economic trends) as it impacts fundraising. Ultimately building the evidence base of data about programs (and trends like health trends) is where the beneficiary-citizen impact is most interesting. The next step would then make some of the relevant data directly available to communities/citizens.

Some potential phases for this initiative include:

1. Phase One – update and expand the three database-indices for NGO-IT leaders
2. Phase Two – expand to IT benchmarking data sets
3. Phase Three – integrate data sets from UN/OCHA HDC projects, as well as other NGO-relevant data sets from corporate partners
4. Phase Four – gather data from NGO PDF reports into time series for research
5. Phase Five- add the data science/analytics to help smaller NGOs better use larger data sets

**6. Next Steps**

The following steps should be considered:

1. Gauging the interest among faculty and students for such a center
2. Determining estimated costs and funding opportunities
3. Hiring/appointing the director for the center
4. Convening the right combination of (a) CIOs/IT managers, (b) large non-profits, (3) technology corporations, and others to contribute expertise and data.
5. Discussing issues of data quality, privacy and comparability.

**7. For further Information, please contact:**

Edward G. Happ  
Executive Fellow  
University of Michigan, School of Information  
Office: +1 734-764-6367  
Mobile: +1 203-979-5364   
Email: [ehapp@umich.edu](mailto:ehapp@umich.edu)

Website: [www.eghapp.com](http://www.eghapp.com)   
Blog: [http://eghapp.blogspot.com](http://eghapp.blogspot.com )  
Twitter: @ehapp  
SKYPE: eghapp  
  
Helping to Make Connections for Good  
Join my network on LinkedIn at [https://www.linkedin.com/in/edward-g-happ-23477b](https://www.linkedin.com/in/edward-g-happ-23477b%20)



**Appendix 2 – Data Sets Desired**

**Data Sets for the UM Data4Good Center**  
  
This is a working draft which will be amended as more information and feedback become available.  The Data4Good Center will be focused on the small-medium-sized NGOs who cannot afford the large-scale data science that would allow them to benchmark and model their humanitarian and conservation work at the intersection of technology, information and people.

1. I want to start with "small" data like the bench-marking indexes I've developed as a Global CIO (e.g., for NGO IT Relevancy and IT Agility indices, as well as basic NGO IT metrics).  Large tech and consulting companies can't provide that; we need to survey our own.
2. Next, I want to "scrape" the data tables from the NetHope organizations' (currently 53 NGOs) flagship reports going back 10 years (e.g. the IFRC's "World Disasters Report," Save the Children's, "State of the World's Mothers" report, etc.) This would give us some interesting time series for research, presuming we can get, or develop, the technology to do intelligent, context-sensitive scraping.  Tech companies may already have such technology.
   1. The first level of parsing the publicly available PDF reports is to separate out and tag the narrative, lists, images and data tables.
   2. The data need to be tagged by organization, publication title, source (URL), page, date, author(s), etc.
   3. The data tables need to be further parsed into data elements tagged by row and column identifiers as well as date (i.e. report date). Footnotes also need to be tagged and linked, as well as data definitions in the reports methodology and glossary.
   4. Data from the tables need to populate time series of data element values, program name, org name, row and column names, etc.
   5. The unstructured narrative data needs to be analyzed for themes derived from its context. Here, the more advanced AI engines will help.
   6. Analysis and visualization options from the data will come later. The first order of business for the project is assembling meaningful data.
3. Moving up the NGO-IT pyramid, are the data sets that may support NGO decision-making.  Things like long-term forecasts for climate change, economic, political and donor contexts in the geographic areas where international NGOs work.  Here, organizations and companies may already have some of these data sets, many of which are publicly available, but may be in PDF reports.
4. Somewhere in this progression, I'm interested in the data sets of voice and data telecommunications details and experiences by country and rural/urban regions within country.  These data sets would provide important input for the 2G Lab initiative I am starting.  The GSMA and Telegeography may be prime sources for this, and tech company initiative teams may have useful data sets here as well.
5. Ultimately, it's the data sets that may be useful for citizen-beneficiaries to directly use, like smart-routes with clinics, camps, WiFi hot-spots, etc. that migrant-refugees need.  I know NGOs with relief operations are interested in creating or expanding these.

As we continue to think more broadly, I expect the possibilities will grow.  The Data4Good Center is an initiative which creates the space for data research for NGOs to flourish.

1. I introduced the IT Strategy Pyramid in 2004 and described it in this Blog post (see item 4) <http://eghapp.blogspot.ch/2010/11/six-views-on-innovation.html> [↑](#footnote-ref-1)
2. Note that the Google Flu Trends project is no longer public, but general search trends are for the [public data](https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_#!ctype=l&strail=false&bcs=d&nselm=h&met_y=flu_index&scale_y=lin&ind_y=false&rdim=country&idim=country:US&ifdim=country&hl=en_US&dl=en_US&ind=false) resource. [↑](#footnote-ref-2)