**IT Think Paper #24: Open IFRC Data for Stakeholders**

**Edward G. Happ, Global CIO**

**October 25, 2015**

There is a growing desire among the RC National Societies and others to access and analyze the data IFRC collects annually. This data is largely trapped in a plethora of printed reports and PDF files that are not conducive to analysis. The IFRC publishes over 200 topical reports and nearly 2,000 grant reports annually. The data tables with these is conservatively in excess of half a million data points. Add to this the data in our internal data warehouse and file repositories and the data points exceed 2 million. And these are refreshed at least annually; some change monthly.

Some summary data has begun to be shared through applications like FDRS and platforms like FedNet and the ifrc.org website. However, these are not data sets one can readily pull into analytical applications. The temptation is to collect more data and place it in repositories that have more a presentation rather than an analytical goal. Indeed, most knowledge-bases that organizations create are based on the library model of categorizing and arranging at the document level rather than the data level.

This problem exists on the data level for the enterprise applications that IFRC has evolved, such as Finance, Pledge Tracking, and HR Information systems. Sharing data among enterprise systems became a key productivity gain by establishing the Data Warehouse, BizTalk middleware and Business Objects retrieval, analysis and report system. These are built on commercial off-the-shelf systems.

Recently we realized that as the number of consumer and local applications grew, the need for standard data sets, called master data, was needed to ensure common definitions of data across applications and clear ownership over who was responsible for managing data access and ensuring data quality. This project included a number of department workshops and was completed in 2015.

The successes that we have managed for our data warehouse tools and processes for master data management can be applied to a broader scope of data and made available for a wider audience, including those outside the IFRC and even the RC Movement. There is no need to build a special database, and tools with a high degree of costly customization. The investment should be made in data collection processes and access control with existing tools

Our goal should be to maximize the exposure of existing data to the end of greater transparency and analytical opportunity. The approach was implied by Professor Mukesh Kapila, former IFRC USG, namely that we take an R&D approach to our data “create an [R&D] network around this.” As in the academic community, we allow the researchers determine how the data is used and key themes identified. The key is turning reports into accessible data sets.

As Maarten van der Veen, Project manager for the Coalition of Humanitarian Innovation (HIC) at the Netherlands RC recently wrote:

“The benefit of having access to open data is first of all that we don’t have to define all business problems beforehand, but that we can solve problems when they arise, because the data is available. This kind of flexibility is crucial during disaster response and for planning programs.”

We have asked Finance (Natig) to provide some initial data sets that Maarten requested from the data warehouse. We are calling for a Proof of Concept project for open data sets to follow.

More recently, the Emergency Operations Center (EOC) project has become an opportunity for us to rethink our approach to data. Indeed, we noted in our EOC report that “The #1 impact of the EOC may not be that we gain a decision-support tool for emergency response, it is that we reorient our thinking and practice of data toward data sets rather than data reports.” Let us think about our data audience as primarily the researcher rather than the reader. We just may find that print media (on or offline) is the derivative expression.

**NOTES FROM FUTURE PERSPECTIVES ROUNDTABLE – 13-Oct-2015**

(Mukesh Kapila, Roger Bracke, Ian Steed, Elise Baudaot, Ed Happ et al.)

Knowledge cannot be put into a box.

Need to build capacity to handle knowledge and local accountability; Route is through research.

Last two years of reports is ~2,000 pages of info; we don’t realize the wealth of data we have.

He created a list of 20 questions to pose to this data, such as “What’s the right CB strategy for NSs?”

Need to create an [R&D] network around this

EMAIL TRAIL

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**From:** Edward HAPP
**Sent:** Wednesday, October 14, 2015 3:16 PM
**To:** MvanderVeen@redcross.nl; Roger BRACKE <roger.bracke@ifrc.org>
**Cc:** Natig VELIEV <natig.veliev@ifrc.org>; Joseph OLIVEROS <joseph.oliveros@ifrc.org>
**Subject:** Re: Data on IFRC operations and national society capacity

Maarten,

Allow me to connect you with Roger Bracke, who hosted a Futures Perspectives Roundtable yesterday with Mukesh Kapila. Part of the discussion was about the research benefits of aggregating data sets.

I believe there's an intersection with your interests below. I would broaden the discussion further including the value of open data that may be used across the RC Movement, including our volunteers as an interested "crowd" source.

…

Ed

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

On Oct 7, 2015, at 2:55 AM, Veen van der, Maarten <MvanderVeen@redcross.nl> wrote:

Hi Edward,

The benefit of having access to open data is first of all that we don’t have to define all business problems beforehand, but that we can solve problems when they arise, because the data is available. This kind of flexibility is crucial during disaster response and for planning programs.

However, there are a few business problems that we can already define:

- Who is doing what, where and when (4W). This is a main activity that information managers do during disaster response. First of all we do this to support the Shelter Cluster (IFRC). The idea of having data about the history of appeals is to quickly be able to assess the parties that are already in a country for a running operation.

- The 4W challenge also applies to any regular programming that national societies do for their international programs. Currently it is not easy to find what national societies are working in what countries, and what has been done/tried. Complementarity is therefore not that easy, as we don’t know what countries are being supported by other NS, or where disaster response or regular programs are ongoing.

- Accountability: having a map of ongoing operations is a great way of showing the RCRC impact. Pooling data from different sources and combining this in an impact map/dashboard will help the RCRC show its relevance

- Identifying high risk areas. Being able to go back in history to show where the RCRC have worked, and what it did (i.e. shelter support, wash, etc), is a great way of analyzing where in the world the risks are highest. That data we can use to make priority maps, and to also prioritize other activities that we are doing, such as Missing Maps (putting the most vulnerable on the map).

The difference between having reports (pdfs) and data is evident. We can rework the data, analyze it for a specific purpose, map it, dashboard it, etc.

Hope this clarifies it. Currently we are in an explorative phase of what is possible and useful. So as soon as we do anything useful with the data we can feed it back to your team and have a chat about it.

With kind regards

Maarten van der Veen

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

From: Edward HAPP [mailto:edward.happ@ifrc.org]

Sent: dinsdag 6 oktober 2015 09:01

To: Veen van der, Maarten <MvanderVeen@redcross.nl>

Cc: Natig VELIEV <natig.veliev@ifrc.org>; Joseph OLIVEROS <joseph.oliveros@ifrc.org>

Subject: Re: Data on IFRC operations and national society capacity

Maarten,

Could you provide more background on the business problem you are trying to solve? What questions are you looking for the data to answer? Who is the audience. Before solving the ease and speed of access it would be good to have these on the table. Thx,

Ed

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

On Oct 5, 2015, at 9:25 PM, Veen van der, Maarten <MvanderVeen@redcross.nl> wrote:

Hi Edward,

No problem that we could not meet this afternoon. I came unannounced.

I mentioned earlier that we able to contribute to find ways to better dashboard the RCRC data, to provide new insights, and to help it improve our understanding of the RCRC impact in the work that we do. We do this every day for different RCRC responses and shelter cluster deployments. I have come to understand there are two datasets being maintained by IFRC:

- a dataset of information on the capacity of NS that they fill annually through an online survey.

- A dataset of all the appeals (IFRC operations)

-

All the data that we would like to map out is available in PDFs on the IFRC website, but we don’t have access to the structured data. It would be a waste of time to data mine all those PDFs if more structured data is readily available.

I spoke to Natig today who showed me that it is a 15 minute job to get the basic data from the data warehouse exported and sent to an email address. I am not sure how structured the NS data is, and if and how easily it can be provided.

I would like to double check with you if it is ok if we try the following:

- IFRC provides a basic dataset to us with the structured data.

- Simon Johnson (British Red Cross) and myself will look into different ways of analyzing, mapping and dashboarding that data

- We share these results with your team, so that you can see what is possible

- We work from there to see what is desirable, possible and the best way to disseminate some of the results

Looking forward to your response,

With kind regards

Maarten van der Veen