D5: Data Driven Documents Driving Development

Fahad Pervaiz
fahadp@cs.washington.edu

Trevor Perrier
tperrier@cs.washington.edu

ABSTRACT

The collection of health system data is recognized as an important tool for international development. With the collection of large amounts of data there exists a need to have better tools for exploring and analyzed this data. In this class project we designed and built a plugin for the District Health and Information System v2 (DHIS2). Our main goal was to create a simple interaction using familial visualization elements such as bar-charts and maps. We wanted to make it easy for someone to explore data reports at multiple levels of the organizational hierarchy for the country. The end goal of this project is a DHIS2 application that can be run in the cloud for any instance of DHIS2.

INTRODUCTION

In the last decade there has been a renewed focus on metrics and evaluation in the international development community [5]. That fact that you must first know the current state of a countries health system before attempting to improve it is not that revolutionary; however, collecting and aggregating large amounts of information in low resource environments is no trivial task. One system that attempts to solve this data collection problem is the District Health and Information System v21 [1].

DHIS2[1] is tool to collect, manage, visualize and explore your data. It lets you manage individual as well as aggregate data with flexible data models, which has been in use in the field since 2008 when the stand alone desktop version was updated to a cloud based architecture. DHIS2 is being used in 46 different countries with most countries in Africa. It has a highly adaptable and extendable web api and underling data model so that each country can customize it to their particular needs.

Using DHIS2 health administrators in the capital can track detailed information about the country's health system such as infant mortality, diarrhoea, and antenatal clinic visits. However, this creates an overwhelming amount of data that must be parsed and analyzed by health officials. DHIS2 has a built in visualization tool [Fig 1] which is very complicated and requires multiple steps to change time periods or view data from different facilities.

For our class project we worked on creating a simplified data explore plugin for the DHIS2[2].The goal this project was to create a new interaction model for viewing reported. We wanted to remove the multiple levels and often unused options provided by the default DHIS2 visualizer - such as which axis the data should appear on, what type of chart to create, or which time periods to include. The intended result is a set of vizulizations that make it easier for a district health administer to explore the data collected by their DHIS2 system.

BACKGROUND

On of us, Fahad Pervaiz, has been working extensively with data quality issues in DHIS2 for the last year as well as assisting in the deployment of a new submodule to track vaccines on a per refrigerator basis. In November of last year Fahad attended the DHIS2 academy in Lagos Nigeria. DHIS2 academies bring together local managers from neighboring countries to show a glimpse of their current use of DHIS2 and review the design principles, setup and maintain of the software. At the DHIS2 academy Fahad was able to interview and work with many people who use DHIS2 on a daily basis. We used these interviews as starting motivation for this project.

Through other research projects both of us are doing we have connections with Seattle based NGO PATH. PATH, in partnership with the Gates Foundation and UNICEF, is working on DHIS2 deployments in Loas, Etheopia, and Tanzania. Talking to district and provincial managers

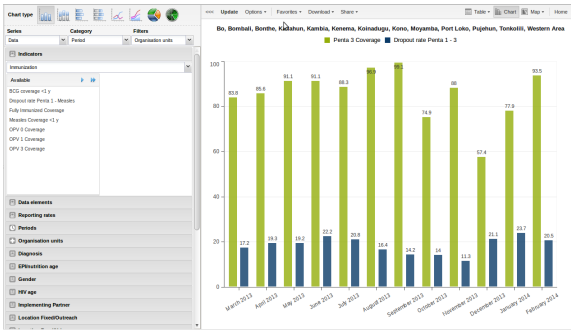

Fig 1. The current DHIS2 visualization tool is designed to be used by experts and is hard to explore data with.

[1]http://www.dhis2.org

in Tanzania and Laos, we found out that DHIS2 is most commonly used to build bar chart graphs and maps. These are the most useful visualizations because most district managers are not very comfortable using other graphs like area, line, pie and scatter charts. They also stressed how difficult it is to limit data in a graph to specific geographic areas or explore lower level areas. For example, a common task is to determine which health facilities are are reporting abnormal data. Currently the way this is done is by creating multiple graphs using the available menu system. The purpose of this visualization was to let user investigate and explore their datasets easily, and we wanted to do this in a way that would be both familial and intuitive to possible for

Currently, Health Ministries conduct training of their managers and staff so they can use this complicated graphing tool. This certainly is not a scalable solution as users keep changing the departments or position within a department so a lot of energy is spent to train the users.

The director of BID (Better Immunization Distribution) in Tanzania said that visualizations in DHIS2 are used for primarily for the following tasks:

- Evaluate the relative performance of administrative areas.
- Determine changes in performance over time.
- Investigate highs or lows to find poor performing health units.

Based on this and other feedback we decided to focus our visualization on the following two areas:

- Slide through the monthly data reports using just bar charts to see the changing trends
- Geographical drill down on bar charts and maps to investigate the datasets

STRUCTURE OF DHIS2

In order to further understand the motivations behind our chosen design it is important describe the data structure of DHIS at a high level. There are several main data types that we will be dealing with. First, Organization Units form a hierarchical tree going from the country level down through provinces, districts and then individual health facilities. Second, what DHIS2 refers to as a Data Set is analogous to a report or collection of Data Elements, or fields, to be filled out by specific organization units at regular intervals. Data Elements can be combined into calculated fields referred to as Indicators. It is these Indicators that health officials are usually the most interested in.

At this point is is also important to note that there are numerous data quality issues with most DHIS2 instances. For this project we used the publicly accessible demo instance which is a weekly dump of data for Sierra Leon. In this data set about 10% of health facilities do not submit reports for the given time period. Besides data being absent sometimes the data is inaccurate. For example the demo data set we use for our visualization is Penta Coverage. This is the percentage of children under six months of age who have received a penta vaccine. This percentage is calculated from the raw numbers and therefore reflects inaccuracies in the data. For this reason the data often has percentages greater than one hundred.

Design

The major goal of this project was to create a new interaction paradigm for exploring DHIS data. In the current visualization tool To explore the data organizational units over time, the desired month or year must be selected from the “Periods” panel [Fig 2]. Next the data must be selected from the “Indicators” panel. Lastly, you must browse through a tree of all

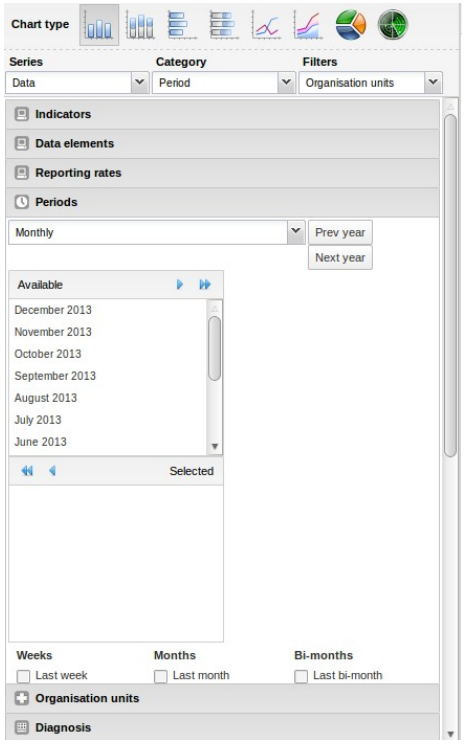

Fig 2. The current DHIS2 options panel with many options and menus.

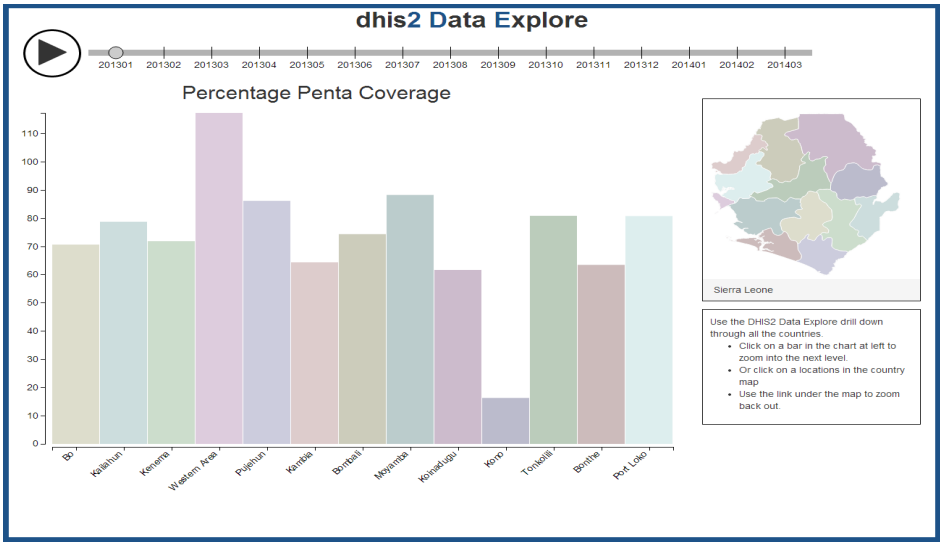

Fig 3. A view of Penta Coverage in our Data Explore for DHIS2. Each bar in the chart is a link into the data for that province or district. The slider at the top can be used to view data for different reporting units.

organizational units and select the ones you wish to plot. This requires 3 to 4 clicks as well as scrolling to the correct option on each panel. Additionally, there is no smooth way to transition the chart to next time period without going back through the menu system. As district managers in Nigeria reported to us this process is further complicated by the numerous options in each panel are overwhelming to quite a few novice computer users.

To simplify this, in the design process we focused on ways of showing data at different levels of the countries organizational tree and across multiple reporting periods. We made the decision to limit the number of options available and make user experience designed for better exploration of the data. In our initial story boarding we decided to only show bar graphs since those would be what our users were most familiar with. We also decided that showing anything other then organizational units on the x-axis would not be very intuitive. The issue is that even by fixing these options the DHIS data set is still multidimensional. For each data element there are multiple levels in the organization hierarchy as well as multiple reporting periods. The next phase of our design was focused on these issues.

In DHIS2, to perform a drill down operation you must reselect the admin area from organization unit panel that

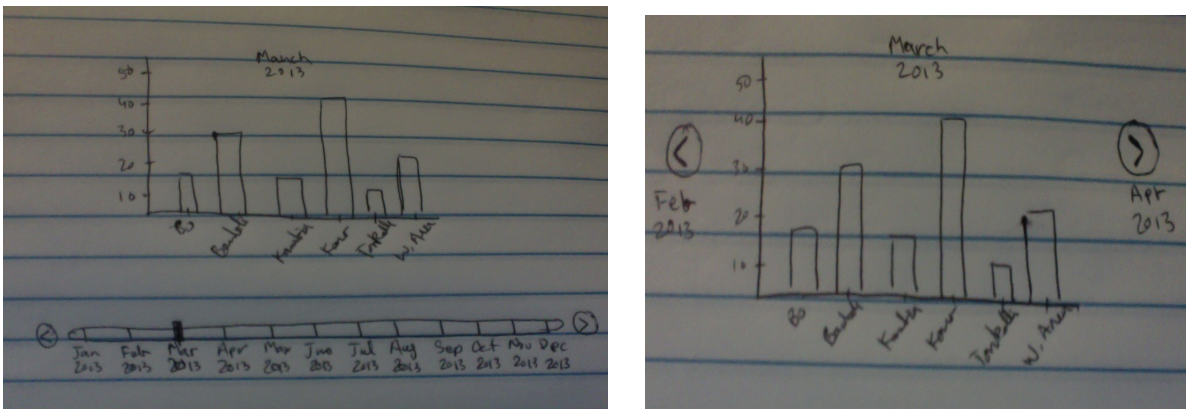

Fig 4. Images from our initial storyboarding. Very early on we settled on using bar charts and had a timeline for sliding through different reporting periods on the bottom.

is one of many option panels as show in Figure 2. This is a very cumbersome and complicated for quite a bit of users. For our design we decided to make the bars themselves links to the data in the next organization level. We wanted to do this in a way that allowed the user to see how the data was changing. Following the work in [2] on animated transitions to unstack bar charts we decided to split a bar into its constituent parts and animate the creation of the next level down.

This worked well for drilling down the organizational hierarchy, but to go back up there was no natural place to click. Instead we created breadcrumbs of where in the hierarchy the user currently is and allowed them to jump back up one level. We also added a map so that the location of health facilities could easily be identified. In our storyboarding we had thought about also including a breadcrumbs map that showed where in the country the current province or district was. However, we decided that users should be familiar enough with the countries political units to not need this.

Our storyboards also examined ways to go through all reports for various time periods. We decided to put a slider at the bottom of the bar chart with all available time periods and allow the user to scroll through the reports quickly. The slider metaphor also enables non-linear browsing of time periods which was very difficult in the DHIS2 visualizer. During our poster session and demo three changes were suggested to how we implemented the slider. First, the slider was directly under the bar chart which made it look like x-axis labels. To fix this we moved the slider avove both the bar chart and country map. Second, in our original implementation the y-axis would scale with the data for each time periods report. This was very distracting and as was pointed out actually made it difficult to compare reports across time. Lastly, the labels for each bar were attached to the bar. This meant that as the time period changed and the heights of the bar charts changed labels would jump up and down dramatically. We moved the labels to the bottom where the slider used to be. This made the labels stay in place and the transitions much less distracting.

FUTURE WORK

There are several next steps for this class project. First, because we were focusing on the design and implementation of the basic data explore we used only a few data sets. We need to add a selection menu that will allow you to choose different indicators or data elements to chart. One potential issue this brings up is that some data elements have negative values and the current bar chart implementation does not handle them very well.

The second major item for future work is to hook out data explore into the DHIS2 api so that live data can be loaded into the visualization. This can be done as part of the DHIS2 application store so that any running instance of DHIS2 can use our visualization.

Because the DHIS2 data set is so rich there is room for more design on more concise ways of showing this data. For example it would be interesting to explore the space of small multiples across data elements and reporting periods to really give uses a comprehensive over of the data at a quick glance. We decided to avoid this visualization paradigm because we felt that the bar charts would be more accessible to our intended audience and that small multiples would be an overwhelming amount of data. However, they would make a nice addition to a health metrics dashboard.

REFERENCES

1. Braa, Jorn, and M. Humberto. "Building collaborative networks in Africa on health information systems and open source software development–Experiences from the HISP/BEANISH network." *IST Africa* 3 (2007).

2. Heer, Jeffrey, and George G. Robertson. "Animated transitions in statistical data graphics." *Visualization and Computer Graphics, IEEE Transactions on* 13.6 (2007): 1240-1247.

3. Heer, Jeffrey, and Ben Shneiderman. "Interactive dynamics for visual analysis." *Queue* 10.2 (2012): 30.

4. Nyella, Edwin. "Challenges in health information systems integration: Zanzibar Experience." *Information and Communication Technologies and Development (ICTD), 2009 International Conference on*. IEEE, 2009.

5. Stansfield, Sally, et al. "The Case for a National Health Information System Architecture; a Missing Link to Guiding National Development and Implementation." *Making the eHealth Connection, Bellagio* (2008).