View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by IUPUIScholarWorks

Article

Overview, not Overwhelm: Framing Operational BI Tools using Organizational Capabilities

Saptarshi Purkayastha
Dept. of Computer and Information Science,
Norwegian University of Science and Technology, Trondheim, Norway

Jørn Braa
Dept. of Informatics,
University of Oslo, Oslo, Norway

Abstract

In contexts where fragmentation of information systems is a problem, data warehouse (DW) has brought disparate sources of information together. While bringing data together from multiple health programs and patient record systems, how does one make sense of huge amounts of integrated information? Recent research and industry uses the term, “Operational BI” for decision making tools used in operational activities. In this paper, we highlight the use of DHIS 2, a large-scale, open-source, Health Management Information System (HMIS) that acts as a DW. Firstly, we present the results of a survey done in 13 countries to assess how Operational BI Tools are used. We then show 3 generations of BI Tools in DHIS 2 that have evolved from action-research done over 18 years in more than 30 countries. Secondly, we develop the Overview-Overwhelm (O-O) analytical framework for large-scale systems that need to work with Big Data. The O-O framework combines lessons from DHIS 2 BI Tools design and implementation survey results.

Keywords

Operational BI; data warehouse; big data; health information systems; developing countries; DHIS 2; open-source; organizational capabilities

1. Introduction

It is evident from the literature that health information systems (HIS) in developing countries are fragmented [1] [2] [3]. This fragmentation causes problems of incomplete and inaccurate information [4] [5]. Yet, this fragmentation of systems is also commonly found in large enterprises, where different departments gather data in separate transactional systems that do not communicate with one another. Although our tools and embedded experiences are from the developing-country contexts, you’ll find reminisces of enterprises across the world and thus, the lessons learnt by us over the last 18 years should be valuable to many practitioners who design or implement Business Intelligence (BI) tools. At the outset, we will like to inform our reader about the practitioner focus of this paper. Thus, we have concise description of our research design and methodology. The paper focuses more on findings and historical evolution of our action-research. We also believe that it is not co-incidence that evolution of our BI tools has gone similar to what Chen et al. [6] describe as BI&A 1.0, 2.0 and 3.0. Our field-level requirements have directed us in the same direction, where we classify our generation of BI tools in a similar fashion.

Recent research points to the use of BI for managing the daily business operations [7]. This has been referred to as Operational BI as it is used to manage and optimize operational activities. Although most literature characterizes Operational BI as real-time and low-latency availability of data, there is also acknowledgment that “*Operational BI puts reporting and analytics application into the hands of users who can leverage information for their own operational activities*” [8]. Others have referred to this as “real-time”, but we prefer to use the term “right-time” and highlight operational availability to be the main difference, when referring to Operational BI in comparison to Real-time BI.

Corresponding author:
Saptarshi Purkayastha, Norwegian University of Science and Technology,
142 IT-bygget, Sem Sælands vei 7-9, Trondheim, Norway - 7419
Email: saptarsp@idi.ntnu.no

This is the author's manuscript of the article published in final edited form as:
Purkayastha, S. and Braa, J. (2013) Overview, not Overwhelm – Operational BI Tools for Big Data in Health Information Systems. Operational BI for Healthcare in Developing Countries.

Similarly, "use of information for local action" has been advocated by numerous researchers working with health information systems [9]. Through this paper's discussion section, we will see that it is not a co-incidence that public health researchers and operational science researchers have both pointed to the same need, but rarely shared the knowledge between the two domains. We look at a popular and arguably the largest [10] open-source Health Management Information System called DHIS 2 (its full-form: "District Health Information Software version 2" suggesting a health-only domain is irrelevant and is more of a backronym these days) and how its Operational BI tools are used in developing countries to manage country-wide or state-wide health systems.

The paper is organized as follows. In the following section we present the current conceptualization of Big Data and Operational BI. Here we also add the literature on Organizational Capabilities that is central to our survey of health data managers and implementers of DHIS 2. We briefly mention our research approach in Section 3. The findings of the survey are presented and discussed in Section 4 of the paper. In Section 5, we highlight the Operational BI Tools in DHIS2 and how these tools have evolved over time. Here we present a specific experience of integrating mobile data capture and the challenge posed by visibility of work due to data. In Section 6, we highlight the lessons learnt in the design of BI and Analytics (BI&A) tools for Big Data. In Section 7, we present the Overview-Overwhelm (O-O) analytical framework for understanding and managing Big Data and articulating the "bigness" of data through organizational capability perspective. We conclude the paper and highlight future work in the last section.

2. Conceptual Framing

2.1. Conceptualization of Big Data

Big Data is a term widely used in popular media. Words like "Petabyte Age", "Industrial revolution of data" are common, yet haven't we had huge datasets of petabytes running on super computers for bioinformatics, space research or other high-performance computing domains for at least the last 20 years? What makes the current times unique is that never before have the "masses" been involved in data creation exercise at this scale, nor has so much general computing power become available. Academic definition for the term is particularly hard to find, but industry reports have defined Big Data with wordings such as:

"Big Data is data that exceeds processing capacity of conventional database systems" (O'Reilly Media)
"Any amount of data that's too big to be handled by one computer" (Amazon)
"Big Data is data with attributes of high volume, high velocity, high variety and low veracity" (IBM)
"Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze" (McKinsey)

With the integration of multiple sources of data into a DW, we face the challenge of increased Volumes, increased Velocity and increased Variety of data. This has been commonly referred to as the "Three V" of Big Data [11]. Other researchers have also added the challenge of Veracity of data as the fourth Big Data challenge [12]. Weiss and Indurkhya [13] in computer science and Diebold [14] in econometrics were one of the first to suggest practical principles of data mining for Big Data. Still, it is difficult to apply these principles in the context of low-resource settings, where the constraints in themselves add to the "bigness" of incoming data. The nature of data in health systems is obvious to the "Four Vs", because with population increase, the medical observations, health indicators and health facilities are always on the increase. When any current incoming data point has to be analyzed with the vast snapshot of existing historical data, the exponential increase in complexity poses problems of processing, storage and analytics.

From these reports and related literature, we see that Big Data is not only large volumes of data, but also contains complex interconnections [15]. So the term in itself is somewhat poorly descriptive [16]. Should it be called large, quickly changing, complex datasets for people who dislike or find incorrect use of the term "*Big Data*"? We do not attempt to demystify Big Data here. We simply claim that national health DW contains data of similar attributes that needs to become meaningful information, through analytics so that it can improve health services and save lives.

2.2. Conceptualization of Operational BI

Over the last 50 years, BI has been expected to change the way in which information is used by organizations to make business decisions. These have been over the years re-invented as Decision Support Systems (DSS), Expert Systems and Executive Information Systems (EIS) [17]. Much effort has been tied to automation or support of human decision making. Still, systems that provide process-based information for decision making have only become top priority for

Chief Information Officers (CIOs) in the last 10 years [18]. Even the creation of job profile for CIO started in the early 90s with the implementation of ERP systems and ensuring the potentials of these systems are met [19]. This long hiatus can be attributed to the complexity of business processes and lack of availability of data from all parts of the organization that can enable making a holistic decision. BI provides insights to managers and information officers to make more informed decisions [20].

White [7] classifies BI into 3 main parts, namely strategic, tactical and operational. These are mainly classified based on Business Focus, Primary users, Time-frame and metrics of data. The basic premise in classifying the different forms of BI comes from the level at which information is used and when the information is used. In the health sector, information needs to be made available to people at all levels. This information is required for operational activities such as what diseases are more prevalent, where additional drugs and workforce is required, how to make them available given the available resources. As mentioned earlier, this has been referred to as "use of health information for local action" [21], where local practitioners can make use of information and adjust their work practices. With this conceptualization we see that Operational BI is used for the day-to-day activities of the users, who are information generators as well as information consumers at the same time.

Keny and Chemburkar [8] provide a slightly different conceptualization of Operational BI. They present the idea of granularity of information as the characteristic separating Operational BI from traditional BI. They suggest that while traditional BI relies on Key Performance Indicators (KPIs) to derive a holistic perspective on corporate performance, operational BI provides much more granularity to address the needs of operational functions. This characteristic of Operational BI is similar to the concept of "hierarchy of standards" [22], where it is advocated that each level of the health system should be able to manage their own set of indicators, with increasing granularity as we go to the lower levels. The higher levels only need the aggregate view of indicators from the lower levels.

Traditional BI has been complex to implement because it tries to capture all processes and business complexities. Thus, traditional BI implementations involve tricky data modelling and require highly experienced BI developers and data modellers. On the other hand, Kobielus [23] suggests that Operational BI tools can be used as "Do-It-Yourself" business intelligence, where users themselves can configure analytical instruments like indicators, metadata, data sources and create "mashups" (a view of data from different sources) using graphical tools. In IS research, it has been identified that mashups will become the basis for Web 3.0, where user-driven programming of the web [24] happens, as users themselves connect the different pages of information on the internet and create new information. This common conceptualization of Operational BI and Web 3.0 through similar approaches is interesting for future research and direction of Operational BI.

2.3. Organizational Capabilities

Most industry definition for Big Data deals with only technical artefacts such as database systems, but does not highlight the role of organizational capabilities in determining "bigness" of data. IS Researchers have discussed about Organizational Capabilities [25] perspective in Knowledge Management that we have found useful to understand how an organization looks at Big Data. This perspective is useful to define Big Data for an organization and manage Big Data to the best of available capabilities and how these capabilities can be put to best use.

A purely technical definition of Big Data talks about single computer or standard database systems that cannot meet computational demands of analytics. We argue that this term has to be more than technical alone and should encompass organization capabilities. Volume, Velocity, Variety and Veracity of data are relative terms. Just as the superlative network speeds of the 1990s are unacceptable today, we cannot measure computational capabilities available today by the standards of the fastest supercomputer in the world. Not everywhere is such computational power accessible. Hence, the term Big Data should be contextual and be defined by the socio-technical capabilities or Organizational Capabilities.

Our research model stems from the seminal study by Watson and Wixom [26] that studied implementation success of DW. Instead of implementation success, we postulate a model of comparing implementation factors (further classified as organizational capabilities) and the use of Operational BI tools – both of which help model the organizational effectiveness in terms of delivery of planned vision and effective use of data at different levels in the organization.

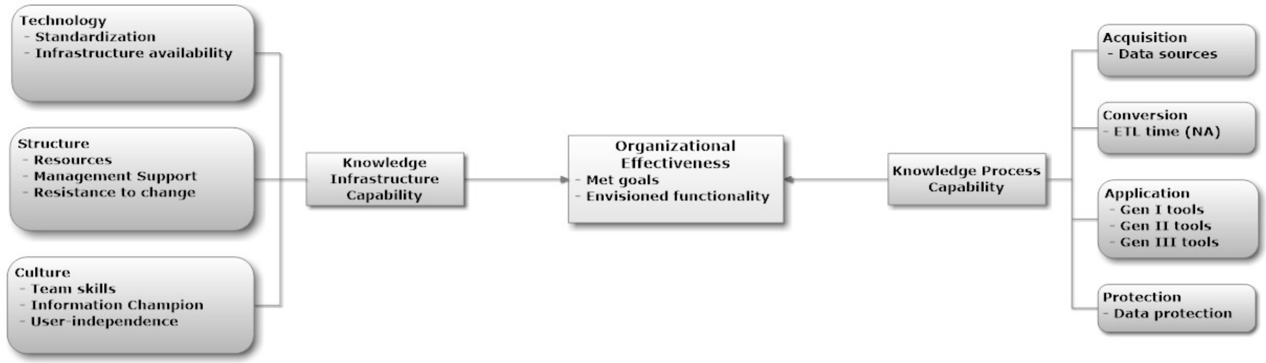

Figure 1. Our research model based on Knowledge management capabilities & organizational effectiveness – Gold, Malhotra & Segars (2001) and Implementation Factors on Warehousing Success – Wixom and Watson (2001)

In Figure 1, our research model is shown with Knowledge Infrastructure Capabilities being the formative factors on the left and Knowledge Process Capabilities being the reflexive factors on the right. Although, we classify these as factors at the start of the implementation, we realize that during the implementation process, these formative capabilities continue to evolve due to the reflexive factors. Other researchers have described this transition has been highlighted as Architectural Maturity [27] and we've seen similar changes during the adoption of DHIS2 systems, where technology, structure and culture evolve during the different stages of implementation. Under Technology capabilities, we list Standardization (metadata, user access, vocabulary) and Infrastructure availability. By infrastructure, we mean hardware (servers, networks, power supply) and software (OS, database systems) infrastructure that can allow installation, running and maintenance of the DW. Under Structure capabilities, we list Resources (amount of finance available, no. of people available), Management Support (memos, work orders) and Resistance to Change. Under Culture Capabilities, we list Team skills (competence), Information Champion (role) and User-independence (self-learning, training, decision making powers). These are also formative factors in defining Big Data for the organization. In Knowledge Process capabilities, we list Data sources (types, velocity), Application (BI Tools used – we describe this in detail in Section 5) and Protection (security, data integrity).

3. Research Approach

Our research has been conducted as part of a long term engagement in the field, doing actions and understanding the effects of our action. This empirical investigation can be broadly placed in Scandinavian Action Research tradition [28], which elsewhere has also been described as *Networks of Action* [29]. The *Networks of Action* approach is specifically designed for the resource limited conditions in the Global South. Our network, called HISP comprises of researchers, developers, implementers, representatives from ministry of health, all of whom share learnings between the nodes of the network. The 18 years time-span of the research project exceeds traditional 'projects' and is more akin to social movements [30]. One of the authors is the originator of the project in the mid-1990's and has participated in the design, development and implementation of HIS in many countries in Africa and Asia. The other author has been part of this network for the last 4 years, primarily engaged in Asia and is part of the core developer team for the DHIS2 software.

The survey respondents were implementers, who we define as "developers/consultants/customizers or local administrators; those who are involved in setting up the system". This includes international HIS researchers, public health consultants, in-country health bureaucrats. For the survey we defined Users as "end-users, people who enter data, look at reports or use reporting tools and perform day-to-day operations; those who are using the system". This survey did not include Users as respondents. We use PLS method to derive the associations between the different reflexive and formative organizational capabilities. The survey was modelled on [26], with a 5-point likert scale and similar technique of analysis.

4. Survey results and discussion

From our survey findings, we can see that some organizational capabilities have more direct co-relation with others, while others that we thought also had co-relation are not supported by the findings. We see that although most of the organizational capabilities are co-related to the use of Operational BI Tools, we see some factors not co-related, while others are more significant. We see from our study that structural and cultural organizational capabilities are more relevant in use of Operational BI tools than Technology factors.

Table 1: Hypothesis evaluations

Organizational Capability	Co-relation	Hypothesis Support
Technology	**Moderate co-relation**	
Higher Standardization efforts will be associated with higher use of BI Tools		Supported
Higher Infrastructure availability will be associated with higher use of BI Tools		Not supported
Structure	**High co-relation**	
Higher resources (finance, people, time) will be associated with higher use of BI Tools		Supported
Higher management support will be associated with higher use of BI Tools		Supported
Lower resistance to change will be associated with higher use of BI Tools		Supported
Culture	**High co-relation**	
Higher team skills will be associated with higher use of BI Tools		Supported
Information Champions at more levels will be associated with higher use of BI Tools		Supported
Higher user-independence will be associated with higher use of BI Tools		Supported
Acquisition	**No co-relation**	
Higher disparity of data sources will be associated with higher use of BI Tools		Not supported
Application (Use of BI Tools)	**Moderate co-relation**	
Higher use of BI Tools will be associated with meeting organizational impl. success		Supported
Higher use of BI Tools will be associated with meeting impl. success		Supported
Data Protection	**No co-relation**	
Higher data protection will be associated with meeting organizational impl. success		Not supported
Higher data protection will be associated with meeting impl. success		Not supported

Technology factors - Even though more infrastructural technology might be available, it does not necessarily mean that users would make use of Operational BI tools. This is an important learning because we often understand that since we have the technology and available data; it will not necessarily mean that these will be useful in meeting the organizational implementation success or envisioned functionality of the data warehouse. This seems to be another observation in the "Build It and They will Come" debate around Information Systems [31] [32] [33]. When there has been higher standardization effort, i.e. harmonization of datasets, formats we see that there is more likelihood that users will make use of the data. This was our hypothesis and is supported by the survey results.

Structural factors - We see resources within the organization as part of the structure. When more resources, better management support and lower resistance to change within the organization, we see better use of Operational BI tools and in turn better implementation and use of data.

Cultural factors - Team skills include implementer's inter-personal skills as well as the user skills in using the tools. Our hypothesis was that with higher skills there would be higher use of Operational BI tools. We also believed that with more Information Champions at all the levels of the organization, there would be better use of BI. Information Champions are people who are proactive in collection, reporting and use of data. Our results that support the view that more Information Champions mean better use of tools, is in some sense contrary to the findings from Wixom & Watson [26]. We find that this might be because there are other cultural and structural factors, similar to Beath [34] that are in support in the organizations that have been part of our survey.

Data Acquisition - To our surprise, data when coming from more disparate sources does not necessarily mean that users would want to co-relate them and view them. Thus, although the nature of Big Data might be co-relation between the data points, users will not want to make use of the co-relations just by default. We also thought that there may be a reverse co-relation (lower disparate data sources, meant use of BI tools) instead, but even that is not supported by our analysis. So, it may be concluded that even if data came from one department or from many departments (or health programs in our case), use of data is not affected by it.

Data Protection - We also found that data protection has no direct impact at least from our survey respondents on the use of BI tools or the organizational effectiveness in dealing with data. This may be so, because in the developing country context, there is little or no a law or structures around electronic data protection. Some countries in Africa and Asia are moving towards such legislations, but these are far from implementation.

From the survey of the countries, we see that Organizational Capabilities are vital in understanding use of Operational BI and in turn manage Big Data. Along with this understanding, in the next section we describe the evolution of the Operational BI tools in DHIS 2, that are inscribed [35] with the lessons that we have learnt in managing data for health systems.

5. Operational BI Tools in DHIS 2

The DHIS 2 is a tool for collection, validation, analysis, and presentation of aggregate statistical data, tailored (but not limited) to integrated health information management activities. It is a generic tool rather than a pre-configured database application, with an open meta-data model and a flexible user interface that allows the user to design the contents of a specific information system without the need for programming.

DHIS 2 has been implemented in more than 30 countries in Africa, Asia, Latin America and the South Pacific, and countries that have adopted DHIS 2 as their nation-wide HIS software include Kenya, Tanzania, Uganda, Rwanda, Ghana, Liberia, and Bangladesh. The software is developed through open-source collaboration and has been iteratively developed through application in “real world” through bottom-up, participatory software development. The software has its roots in Scandinavian Action Research tradition in IS development where user participation, evolutionary approaches, and prototyping are emphasized [36]. Its development started with the reforms in health sector in post-apartheid South Africa and has evolved into a large-scale web system that is now used for country-wide management of health information systems. Over this period, lot of BI tools have become part of the application, but its core agenda has been the use of information for local action through the use of flexible standards [22].

5.1. The first generation of Operational BI Tools

From Offline Pivot tables to Downloadable Pivots to Stuck with Big Data

The first generation of Operational BI tools in DHIS 2 included tools which allowed users to modify data attributes and observe their effects on indicators. These tools allowed data managers to introspect on the effect of changes to certain data points with respect to a set of existing longitudinal data points. In Microsoft Excel™ spreadsheet, such a tool is Pivot tables which allow creating unweighted cross tabulations. Pivot tables are commonly used for row-column interpolation and interchanging rows with columns and vice versa. Pivot tables have been used widely in understanding and analyzing information [37]. These have been important tools for integrated decision support systems based on data warehouses [38]. From early 1996 to early 2006, these tools have been used widely by users of DHIS v1. The pivot tables perfectly supported operational activities of health staff at facilities. As other researchers have reported, some effort was required to train users to work with pivot tables, but later it became one of the most used features in DHIS [39] [40] [41]. New incoming data points are fairly easy to add to existing pivot tables, but only till such data is limited in size. When we moved to DHIS 2, we decided to completely rewrite it as a web application. In such networked data collection and web application world, when we first tried to implement pivot tables, it proved to be extremely slow. Web browsers were not able to support pivoting even large, static dataset. High velocity data was impossible to accommodate through pivot tables. In 2011, Google provide a plugin for web-based pivot table interface through their Google Docs™ service. But the plugin has since been deprecated because large datasets are impossible to work with. To deal with the issue of large datasets, we decided to partition the data mart based on the facilities, user data and geographical locations. We also allowed users to select indicator dimensions, period, and location, so that pivoted outputs are manageable in size. In addition, we allowed users to connect to a web service and download data mart locally using a Java desktop *mydatamart* application. This desktop-based, downloadable pivot solution was better than the browser-based pivoting solution, but still could not handle 100s of Megabytes, nor high-velocity data as the offline data mart would become stale in a matter of minutes. So, although *mydatamart* is a recipe for flexible Operational BI and used in some implementations of DHIS2, yet it cannot deal with Big Data.

From Data mart Service to Nightly data marts to Caching proxies

The Data mart service is Extract-Transform-Load (ETL) process in DHIS 2. This backend process provides information for reports based on aggregations and disaggregations of transactional data. The reports are generated from the data mart, a common practice in DW tools. In the early days of moving to a web application, DHIS 2 would run the data mart service as soon as new reports were requested. This caused the same information to appear in the reports, even if data values had changed due to high velocity, while the latest data mart generation was not completed. This is a common problem with high velocity data. Over time, there were a few optimizations like having a separate reporting server, algorithm optimizations etc. Yet, we observed that data mart creation became more and more time consuming as data became high-velocity and high-variety. To be able to deal with this challenge of growing computational capacity, we decided to build a feature by which the data mart service would only be executed every night at a scheduled time. This meant that users would be able to store data, but the analytics could be run only on the previous day's data. This nightly data mart reduces the load on the infrastructure and makes the reports accessible to users, but reduces the operational nature of information. For e.g. in Ghana where 10 yr legacy data was imported, there were close to 33 million data points that needed rebuilding by the data mart service every night. This process is extremely technology intensive and even though using cloud hosted servers have alleviated some computational challenges, it still takes about 6 hours every night for completion. As more and more data is collected over the next few years, we are expecting even with "elastic computing", which provides on-demand computational resources, we will need more time (more than a night) to complete the data mart. To deal with this problem of correlating large amounts of data and still make information operationally available, one suggested technique which we have started to apply is using caching proxies. These are servers which will partition the requested data and remember it for a user/group of users. So users within a district or county will always request data from their proxies and these proxies will only deal with information that is generated, stored and processed for their geographies. These proxies refresh themselves based on their own caching rules that are written on an organizational unit's operational needs. This partitioning helps reduce the load and time for the data mart process. This feature is experimental and we are evolving new paradigms for partitioning based on operational needs.

5.2. The second generation of Operational BI Tools

Figure 2. Dashboards with Live Views

From cross-tabulations to Dashboards

From the early days of DHIS 2, graphs and charts have been useful visualization tools. The challenge with multi-dimensional data that has n-number of co-relations is finding a useful visualization for it. Health data is considered in essence to be longitudinal in nature and hence plotting values over time has been the standard approach to visualize it. Still, when co-relating data points, tables have often played a more significant role in such analysis. As the DHIS 2 community of users moved away from pivoting, it was realized that visualizing data through charts and placing them next to each other made it possible to do multi-dimensional visualizing of data. Thus, the DHIS 2 dashboards were born, where co-related data points would be placed next to each other, even overlapping if one wishes to and view data points longitudinally. This allowed users to create their own operational "mashups" and view co-related multi-dimensional data beside one-another. This has been extremely powerful for analytics even in low-level facilities, as it provided easy overview of status of operations. Each user has their own dashboard according to their operational needs. The dashboard is live view of maps, charts and tables. It uses data from the previously mentioned data mart service.

Data mashups on Maps

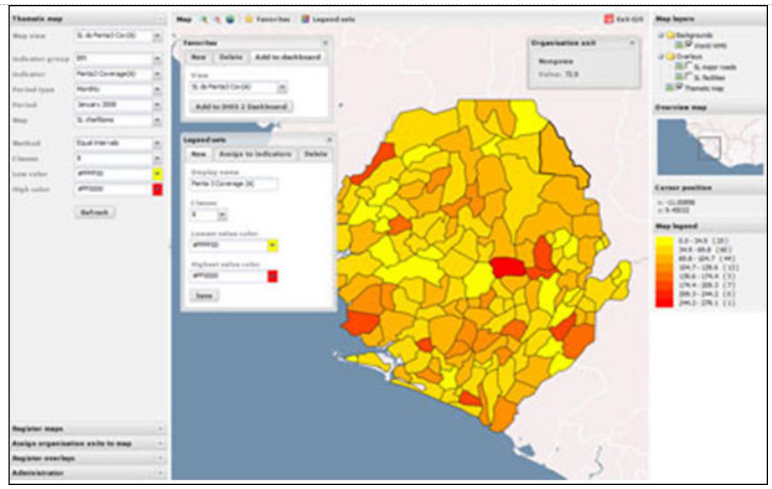

Figure 3. Data mashups on Maps

While the Dashboard provides useful overview, often the contextualization of permanent or semi-permanent data like population estimates, staffing, equipment, survey, geographical topography etc. can better visualized on a map. DHIS 2 learnt that as users wanted to co-relate live incoming data with permanent or semi-permanent data, geographical boundaries of facilities, districts, provinces etc. provided interesting ways in which users can visualize. Users create indicators from data points and use the values from different regions to plot legends onto a map. These are especially useful to compare geographically close locations for similar kinds of information. These maps allow contextualizing the incoming data with existing data and allow users to create their own "mashups" of data. Although this has been referred to a Geographic Information System (GIS) by other researchers [22], it is more along the lines of mash-ups ideology, where maps are used as a surface to represent data extracted from the system. The tool does not allow complex layering, manipulating geo-spatial data or CAD/CAM features that are expected of GIS systems. Instead the tool in DHIS2 allows thematic mapping, while letting the user define custom legend sets and save favourite map views. The tool also allows users to drill-down to the internal organization boundaries and look at the data represented by them. Source of data (organization unit) is automatically matched with the labels on the map/shape files. Users can also manually link data to geographic co-ordinates or polygons. Users are given the flexibility to represent any information from datasets onto any map. The indicators are then seen as different colour representations based on the range of data values.

5.3. The third generation of Operational BI Tools

Validation Rules

Validation rules are expressions to evaluate incoming data values with existing data values. These expressions are mathematical models that can be created by users based on existing data points. Validation rules are used to match data elements and find those that are different from the normal values. These include min/max ranges as well as comparison with the values of other data elements. Failing validation rules from a business logic perspective are outliers, but from our experience in the use of mathematical models, these outliers should not be considered as mistakes in the data, but instead should mean we need to question the reasons for the outlier. New insights can often be drawn from these outliers and hence DHIS 2 allows users to add unstructured comments that can be mined for further understanding the reasons for such outliers. The process of executing validation rules cannot be thought of as a simple data cleaning (in DW often part of ETL process) because it involves analysis of what data exists in the system and what is not being entered into the system. An important part of data that violates the validation rules is to look at the comments on such instances. DHIS 2 allows for adding comments, when data is beyond the min/max values or has problems with validation. For e.g. Validation rules are often created to verify outbreak alerts. Here, standard deviation is used to verify if the number of patients diagnosed with a disease is within the previous min-max range. Similarly, stock outs or buffer levels for inventory is managed through validation rules. So, we see that validation rules help in the local action because operational activities are better managed through the use of such rules. We also have a better understanding of data through these validation rules and comments entered on data that is failing validation rules.

Social network of Interpretations

While we have evolved through the different generations of Operational BI tools, we realize the strength of data analytics lies in the hands of users of these tools, rather than in the tools by themselves. Hence we have recently allowed users to share their interpretation of data, using all of the above mentioned tools. Users can annotate charts, graphs, tables and maps while sharing it with a group of users in DHIS 2. The sharing of these interpretations creates a sort of social network of health system analysts, where users comment and annotate each other's interpretations of data. We find that in many places newer insights have been created by leveraging the interpretations of different skilled analysts in this process. The interpretations are particularly useful when health staff make operational decisions based on interpretations of health staff located in another facility, which would otherwise be disconnected from each other's activities. These social networks allow making sense of the Big Data, but in fact also create more data for each other. Yet, this much more nuanced and extracted information is of greater value to the users, who feel excited and empowered to share their interpretations with peers and superiors. There are structural, cultural challenges that also arise in such social networks, but interpretation tools have indeed provided better extraction of information from large amounts of data. The interpretation tools in DHIS 2 are fairly new and only few countries are widely using it at the moment.

5.4. Case of Integrated mobile phone based data capture

HISP India is involved in the implementation of a large-scale mobile-based health information system in a northern state in India [42]. As part of the implementation, more than 5000 health workers have been given mobile phones that are used for reporting aggregate of health activities performed by field-level health workers. The health workers use a mobile application that captures data and sends this information as an SMS to the state health office. At about the same timeframe, clinicians working in outpatient ward of government health facilities were asked to report their activities on a daily basis through the use of similar SMS-based services. The clinicians started sending daily updates to the state office and a lot of data was being captured. On the other hand, since health workers were being given mobile phones, the state added 10 new data points on which health workers were also asked to report daily to the state. The primary aim for doing this from the state's perspective was to try and strengthen control over the health worker's activities; to know what the health worker was doing on a daily basis. Since, clinicians were already doing this, the state health department officials assumed that health workers would have no problems doing the same. However, as the implementation started, it was observed that health workers were resistant to daily reporting and voiced protests as part of the health worker union. On exploring the reasons for the resistance, it was found that the health workers only performed some of the duties on a daily basis on which they were asked to report. On the other hand, other activities that they performed on a daily basis were not being captured. Beyond that, they had fears that if they reported 'zeros', authorities would assume that they have not worked for the day, whereas in fact they provided other services. In protest to the daily reporting on new datasets, the health workers completely stopped using the mobile-based reporting system. Since the effectiveness of the full project was getting jeopardized, the state agreed that the health workers would rather report weekly on the same activities and the clinicians would continue to report their daily activities. This seemed to go well with all the parties involved and helped collect 'adequate' data about the workings of the health system.

6. Lessons learnt - How to design BI Tools for Big Data

Our experience is that organizational capabilities are most important when designing and implementing BI tools. Using Operational BI tools simplifies management of large datasets and users grow into data use with such tools. Since, our lessons come from experience in participating in a long drawn action-research project involving many different health systems/organizations, the evolution of the tools embed in them the lessons that we have learnt. We summarize some of the wisdom that has been embedded in the tools. This can be used by designers or implementers of other BI tools.

Lessons #1 – Organizational capabilities are difficult to change, evolve processes of data acquisition and analytics

Organizational capabilities have inertia that makes it hard to change. Since people have become accustomed to certain tools or have been trained to use a certain type of BI&A tool, it is extremely hard to get them to change. The Pivot tables was known to be a dead-end for large datasets, yet for users to be able to continue with their existing practices, we had to develop a transitional tool that allowed similar analysis. Similarly, we also see that certain users or organizational departments have capabilities that would allow them to generate data, while other users are not likely to be able to generate the same amounts of data. Organizational culture as seen from the case in India (see Section 5.4) plays a major role in response to acquisition of Big Data, not just processing or analyzing data. Particular care needs to be taken in cases which use ubiquitous technologies like mobile phones that technologically suited to generate Big Data

Lesson #2 – Data partitioning based on operational needs is the key to scalability

Partitioning data has been a common practice for managing data. But when it comes to Big Data, it becomes an essential task. Partitioning can be based on one or combination of the following constructs:

Time - as we did using nightly access to data
Space - as we did on geographical limiting
Activities - as we did using the hierarchy of standards notion

Lesson #3 – Outliers are not unclean data, but points of innovation

Outliers that have been created by a model are often considered to be unclean data or wrong data. In Big Data world, annotating such data and trying to find co-relations is very useful practice. It evolves the use of data, but also provides new insights into organizational practices. As we saw from sharing of interpretations and discussing finding in an enterprise social network creates more useful analysis within the organization about these outliers or different interpretations of data. Being able to extract further information from these discussions is expected to provide better insights into improving practices.

7. Towards Overview-Overwhelm Framework - Discussion

The conceptual view of transforming Big Data to meaningful and actionable information can be understood through the lens of the Overview-Overwhelm (O-O) analytical framework. Our starting point is the postulation that Organizational capabilities are constraining and enabling factors towards use of Big Data. We define the Overview space as optimal understanding of data, such that organizations can easily assimilate the information that is conveyed by such data. On the opposite end, we define the Overwhelm space, where there is wasteful use of resources to make sense of data or the information conveyed by the data is hard to grasp. In this regard, we refer to the work of Sandiford et. al. [43], which distinguishes approaches of IS in health systems, where one uses either an “action-led” approach or a “data-led” approach. When we bring Big Data into the picture, we see data-led approach resulting in Overwhelm space because often in countries, health reporting is done for the sake of reporting [44]. This creates large amounts of data that is rarely used [45]. Here we see upcoming technologies of mobile phone, ubiquitous sensors and wireless technologies contribute to generating Big Data. On the other hand, the Overview space can be created and maintained within health systems through an action-led approach. An action-led approach to Big Data is to process raw data into indicators; but go beyond creating indicators and further matching them with targets. Millennium Development Goals are such targets that countries implementing DHIS2 want to achieve. The Operational BI Tools in DHIS2 helps achieve this transition of data-led to action-led efforts or from Overwhelm to Overview information space. Organizations have large spaces in terms of information view in between the areas of overview and overwhelm. This space is often filled with data and no information or there is no scalable way in which data can be moved from Overwhelm areas to Overview areas. CIOs are individuals who need to take initiative to allow users to move data across this space. CIOs will find it hard to make this move for the whole organization. But with an army of users with operational needs for information, data can be moved as we have shown from our cases, from overwhelm to overview space. By the very nature of health data acquisition, we see that data has the inertia to move towards the Overwhelm space. We claim that Technology, Structure and Culture are formative factors towards changing this inertia and move information into the Overview space.

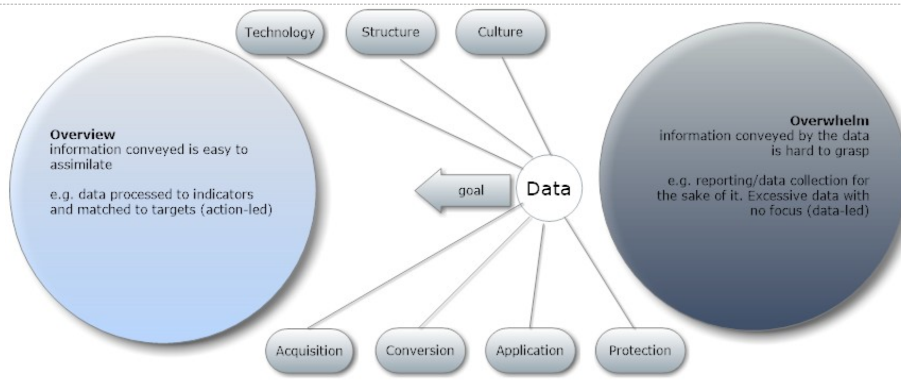

Figure 4. Overview-Overwhelm (O-O) Framework - A looking glass at organizational data management

The O-O framework can be understood as a looking glass to view Organizational capabilities. O-O helps define what Big Data is and how it can be managed. Each of the formative factors of Technology, Structure and Culture of the organization determine whether the acquired data is Big Data or not. Depending whether the data can be provided as an overview or whether it overwhelms the organization determines if it is Big enough or Not for the organization. Starting with the DHIS 2 first generation of BI tools, they moved from an Overview boundary to Overwhelm boundary and efforts were made to reverse that direction. But, the data became overwhelming for the technology, organizational structure and culture as time went on. That strand of BI tools did not adequately move back to Overview and hence is not a suitable toolset to manage Big Data. The second generation of BI tools through the use of visualizations provides overview of data. Still there are limitations to visualization and it does not completely cover the sharing of gained knowledge through the organization. The use of social networks through interpretation of data is much wider overview of the data.

8. Conclusion

We conclude the paper by realizing that Big Data should not be defined only through technical factors. It should also be described by organizational factors such as Structure and Culture. We highlight some of the lessons that have been learnt in design, development and implementation of 3 generations of Operational BI Tools in DHIS2. We specifically show that instead of Analytics in the broad sense, Operational BI tools will help transform organizations to be able to work with Big Data. In healthcare, our experience has been to move from data-led to action-led approach using Operational BI Tools. The O-O framework can help act as a looking glass through which we can define and manage Big Data in healthcare

Our further research points to how cloud computing can aid in sharing computing resources for processing Big Data. We have been working with countries to use IaaS, PaaS, and SaaS models for their data warehouses. It will be interesting to see what are the opportunities and challenges for countries to use Cloud computing for health data analytics.

Notes

1. DHIS 2 In Use – http://dhis2.org/inaction

Acknowledgements

We would like to thank all the funding organizations that have funded DHIS 2 development and implementation over the years. We would also like to thank ministries of health with whom we have worked towards implementing DHIS 2. We would like to acknowledge the efforts of the individuals from different partner organizations in these countries as well as colleagues and researchers who have participated in the design, development and implementation of DHIS 2.

Funding

The work is funded by the Norwegian Research Council – VERDIKT programme's research project called Global Health eInfrastructures (#193023) on which one author is principal investigator and the other author is a research fellow.

References

[1] Littlejohns P, Wyatt JC, Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. British Medical Journal. 2003 Apr 19;326(7394):860–3.

[2] Mossialos E, Allin S, Davaki K. Analysing the Greek health system: a tale of fragmentation and inertia. Health Economics. 2005;14(S1):S151–S168.

[3] Monteiro E. Integrating health information systems: a critical appraisal. Methods of Information in Medicine. 2003;42(4):428–32.

[4] Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Annals of Internal Medicine. 2006 May 16;144(10):742–52.

[5] Kimaro HC, Nhampossa JL. Analyzing the problem of unsustainable health information systems in less-developed economies: Case studies from Tanzania and Mozambique. Information Technology for Development. 2005;11(3):273–98.

[6] Chen H, Chiang RHL, Storey VC. Business intelligence and analytics: from big data to big impact. Management Information Systems Quarterly. 2012 Dec;36(4):1165–88.

[7] White C. The Next Generation of Business Intelligence: Operational BI. DM Review Magazine. 2005;

[8] Keny P, Chemburkar A. Trends in Operational BI. DM REVIEW. 2006;16(7):20.

[9] Wa O, Rh B. Surveillance. Information for action. Pediatric Clinics of North America. 1990 Jun;37(3):709–34.

[10] Webster PC. The rise of open-source electronic health records. The Lancet. 2011 May;377(9778):1641–2.

[11] Laney D. 3-D Data Management: Controlling Data Volume. Velocity and Variety", META Group Original Research Note. 2001;

[12] Zikopoulos PC, deRoos D, Parasuraman K, Deutsch T, Corrigan D, Giles J. Harness the power of Big Data: the IBM Big Data platform. New York; Singapore: McGraw-Hill; 2013.

[13] Weiss SM, Indurkhya N. Predictive Data Mining: A Practical Guide. 1st ed. Morgan Kaufmann; 1997. 228 p.

[14] Diebold FX. 'Big Data'Dynamic factor models for macroeconomic measurement and forecasting. Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society,"(edited by M Dewatripont, LP Hansen and S Turnovsky) . 2003. p. 115–22.

[15] Snijders C, Matzat U, Reips U-D. "Big Data": Big Gaps of Knowledge in the Field of Internet Science. International Journal of Internet Science. 2012;7(1):1–5.

[16] Manovich L. Trending: The Promises and the Challenges of Big Social Data. Debates in the Digital Humanities [Internet]. 2011 [cited 2013 Feb 14]; Available from: http://www.manovich.net/DOCS/Manovich_trending_paper.pdf

[17] O'Brien RC. Brief case: EIS and strategic control. Long Range Planning. 1991;24(5):125–7.

[18] Watson HJ, Wixom BH. The current state of business intelligence. Computer. 2007;40(9):96–9.

[19] Earl MJ. The chief information officer: Past, present, and future. Information Management: The Organizational Dimension. 1996;456–84.

[20] Lawton G. Making Business Intelligence More Useful. Computer. 2006;39(9):14–6.

[21] Stoops N, Williamson L, Braa J. Using health information for local action: facilitating organisational change in South Africa. In The digital challenge: information technology in the development context. Gateshead: Athenaeum Press Ltd. 2003;

[22] Braa J, Hanseth O, Heywood A, Mohammed W, Shaw V. Developing health information systems in developing countries: the flexible standards strategy. Management Information Systems Quarterly. 2007;31(2):381–402.

[23] Kobielus J. Mighty mashups: do-it-yourself business intelligence for the new economy. Forrester Research. 2009;

[24] Wong J, Hong JI. Making mashups with marmite: towards end-user programming for the web. Proceedings of the SIGCHI conference on Human factors in computing systems. 2007. p. 1435–44.

[25] Gold AH, Malhotra A, Segars AH. Knowledge management: An organizational capabilities perspective. Journal of management information systems. 2001;18(1):185–214.

[26] Wixom BH, Watson HJ. An empirical investigation of the factors affecting data warehousing success. Management Information Systems Quarterly. 2001 Mar;25(1):17–32.

[27] Chung R, Marchand D, Kettinger W. The CEMEX Way: The Right Balance Between Local Business Flexibility and Global Standardization. Case Study No. 2005;1501:3–1341.

[28] Engelstad PH, Gustavsen B. Swedish Network Development for Implementing National Work Reform Strategy. Human Relations. 1993 Feb 1;46(2):219–48.

[29] Braa J, Monteiro E, Sahay S. Networks of Action: Sustainable Health Information Systems across Developing Countries. Management Information Systems Quarterly. 2004 Sep;28(3):337–62.

[30] Elden M, Chisholm RF. Emerging Varieties of Action Research: Introduction to the Special Issue. Human Relations. 1993 Feb 1;46(2):121–42.

[31] Markus ML, Keil M. If we build it, they will come: Designing information systems that people want to use. Sloan Management Review. 1994;35:11–11.

[32] Hatakka M. Build it and They Will Come? – Inhibiting Factors for Reuse of Open Content in Developing Countries. The Electronic Journal of Information Systems in Developing Countries. 2009;37(0).

[33] Silvestre A-L, Sue VM, Allen JY. If You Build It, Will They Come? The Kaiser Permanente Model Of Online Health Care. Health Affairs. 2009 Mar 1;28(2):334–44.

[34] Beath CM. Supporting the Information Technology Champion. Management Information Systems Quarterly. 1991 Sep 1;15(3):355–72.

[35] Hanseth O, Monteiro E. Inscribing behaviour in information infrastructure standards. Accounting Management and Information Technologies. 1997;7:183–212.
[36] Greenbaum J, Kyng M, editors. Design at Work: Cooperative Design of Computer Systems. 1st ed. CRC Press; 1991. 312 p.
[37] Jelen B, Alexander M. Pivot Table Data Crunching. 1st ed. Que Publishing; 2005. 288 p.
[38] March ST, Hevner AR. Integrated decision support systems: A data warehousing perspective. Decision Support Systems. 2007;43(3):1031–43.
[39] Williamson L, Stoops N, Heywood A. Developing a district health information system in South Africa: a social process or technical solution? Studies in health technology and informatics. 2001;(1):773–7.
[40] Lungo JH. The reliability and usability of district health information software: case studies from Tanzania. Tanzania journal of health research. 2008;10(1):39–45.
[41] Sheikh YH, Bakar AD. Open Source Software Solution for Healthcare: The Case of Health Information System in Zanzibar. e-Infrastructure and e-Services for Developing Countries. 2012;146–55.
[42] Braa J, Sahay S. Integrated Health Information Architecture: Power to the Users : Design, Development, and Use. Matrix Publishers; 2012. 362 p.
[43] Sandiford P, Annett H, Cibulskis R. What can information systems do for primary health care? An international perspective. Social Science & Medicine. 1992 May;34(10):1077–87.
[44] Kanjo C. Pragmatism or Policy: Implications on Health Information Systems Success. The Electronic Journal of Information Systems in Developing Countries. 2011;48(0).
[45] AbouZahr C, Boerma T. Health information systems: the foundations of public health. Bulletin of World Health Organisation. 2005 Aug;83(8):578–83.