Big Data Analytics for developing countries – Using the Cloud for Operational BI in Health

Jørn Braa
Department of Informatics,
University of Oslo,
Oslo, Norway

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

ABSTRACT

The multi-layered view of digital divide suggests there is inequality of access to ICT, inequality of capability to exploit ICT and inequality of outcomes after exploiting ICT. This is evidently clear in the health systems of developing countries. In this paper, we look at cloud computing being able to provide computing as a utility service that might bridge this digital divide for Health Information Systems in developing countries. We highlight the role of Operational Business Intelligence (BI) tools to be able to make better decisions in health service provisioning. Through the case of DHIS2 software and its Analytics-as-a-Service (AaaS) model, we look at how tools can exploit Cloud computing capabilities to perform analytics on Big Data that is resulting from integration of health data from multiple sources. Beyond looking at purely warehousing techniques, we suggest understanding Big Data from Organizational Capabilities and expanding organizational capabilities by offloading computing as a utility to vendors through cloud computing.

KEYWORDS

Cloud Computing; Analytics-as-a-Service; AaaS; Big Data; Digital divide; Organizational capability; Operational BI; Health Information Systems; DHIS 2

INTRODUCTION

Availability of resources for processing data has evolved over the years due to changes in computing technology such as from mainframes to client-server to cloud computing (Rajan & Jairath, 2011). In the past, with every major shift in technology infrastructure, larger datasets are processed in much shorter timeframes. Yet, within organizations in developing countries, observing smooth, linear and longitudinal technology shifts are rare when compared to organizations in developed countries. Many researchers refer to this as digital divide (Cullen, 2001; Hoffman & Novak, 1998; Barzilai-Nahon, 2003). Digital divide in such cases should not be understood from the simplistic notion of limited access to technology, but rather as a multi-layered concept (Wei et al, 2011) - of inequality of access to ICT (first-level), of inequality of the capability to exploit ICT (second-level) and of inequality of outcomes (e.g., learning and productivity) after exploiting ICT (third-level). Since the spread of internet and easy access to internet-based software services, organizational users in developing countries have started to expect that their work applications will provide the same characteristics, as services offered using cloud computing - like On-demand self service, multi-device access, resource pooling, rapid elasticity and measured service (Mell & Grance, 2011). In developing country organizational settings, we can see that internet access has becoming widely available. Even though not through high-speed wired connections, but using mobile networks (ITU, 2013). Our research is specifically in the Health Information Systems (HIS) domain, working with the ministries of health in developing countries where health systems are our organizations of study. Here, we ask our first research question – How can cloud computing contribute to bridge the multiple levels of digital divide within health systems?

Buyya et al. (2009) suggested that computing will one day be the 5th utility after water, electricity, gas and telephony due to cloud computing. Some of the currently available service models for providing such computing utility include Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS) and Analytics-as-a-Service (AaaS). In the future we expect more types of service models being created for cloud computing. Overall, we suggest that such computing utility will be useful to health systems, only if they allow health care providers to make better decisions during their daily health care provisioning. Recent research points to the use of Business Intelligence (BI) for managing the daily business operations (White, 2005). This has been referred to as Operational BI, as it is used to manage and optimize operational activities (Marjanovic, 2007). 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" (Keny & Chemburkar, 2006). Others have referred to this as "real-time" (Azvine et al., 2006; Watson et al, 2006), 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. Similarly, "use of information for local action" has been advocated by numerous researchers working with health information systems (Kimaro, 2006; Mosse & Byrne, 2005; Stoops et al., 2003). Here we ask our second research question – What service models of cloud computing can provide information for local action, so that better decisions can be made?

Through the examples in this paper, we suggest some strategies to exploit cloud computing such that they may be used to aid decision making in the health system. We look at the popular and arguably the largest (Webster, 2011) open-source Health Management Information System (HMIS) called DHIS2 and how its Operational BI tools are used in developing countries to manage country-wide or state-wide health systems. In the paper, we highlight some of the new developments in DHIS2 that exploit Cloud Computing utility for Operational BI, moving from traditional warehousing approaches like data marts to more elastic and on-demand real-time analytics tools in DHIS2. In the discussion section, we suggest how these tools help manage Big Data from Organizational Capabilities perspective (Gold et al., 2001). We use the Overview-Overwhelm framework (Purkayastha & Braa, 2013) as a mechanism to understand the organizational view of Big Data and how "bigness" of data needs to be defined in terms of an organization's capability.

The next section of the paper explains Cloud Computing definitions, directions, concepts and service models. We also look at theoretical concepts of organizational perspective as well as Operational BI that have framed our research design. The research method is explained in the next section, followed by the case of Operational BI Tools in DHIS2. We discuss the synergies of cloud computing for Big Data analytics for health systems, especially from a developing country perspective in the next section. In the last section, we conclude with future directions for Analytics in the cloud.

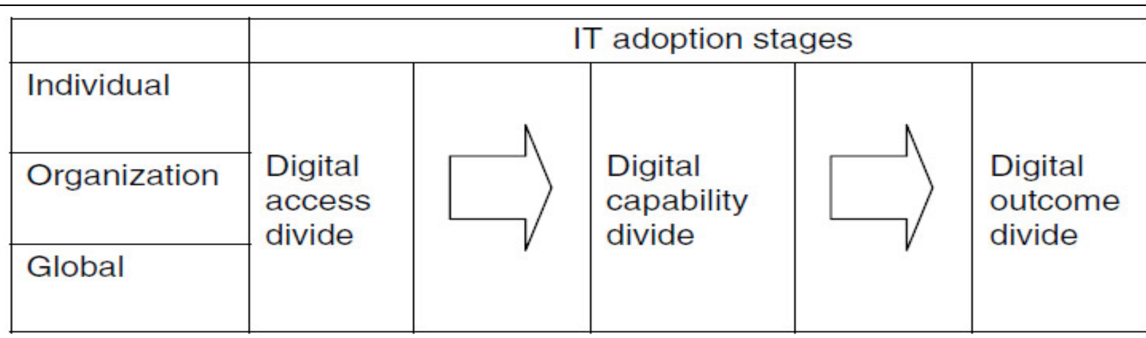
CONCEPTUAL BASIS

The NIST (2011) defines cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This definition of cloud computing is quite useful to understand that the central theme is access to configurable computing resources. Although this definition is probably the most widely accepted, many researchers (Mowbray, 2009; Qian et al., 2009; Weiss, 2007) have highlighted that cloud computing is a vaguely defined, but much described phenomenon. Since the publication of core papers by Google in 2003 to commercialization by Amazon in 2006 and now to learning from the consumer internet and building an “Industrial Internet” of devices, appliances, aircrafts etc. (Bruner, 2013), there has been a steep rise in the number of research articles describing the phenomenon of cloud computing. A search of “Cloud computing” in the Scopus database of published articles in Title, Abstract and Keywords yields 12714 results, all published since 2008!! Yet, in this research spree, little has been said about the role of cloud computing in enhancing organizational capabilities and even less about outcomes resulting from such capabilities.

DIGITAL DIVIDE AS A FACTOR OF ORGANIZATIONAL CAPABILITIES

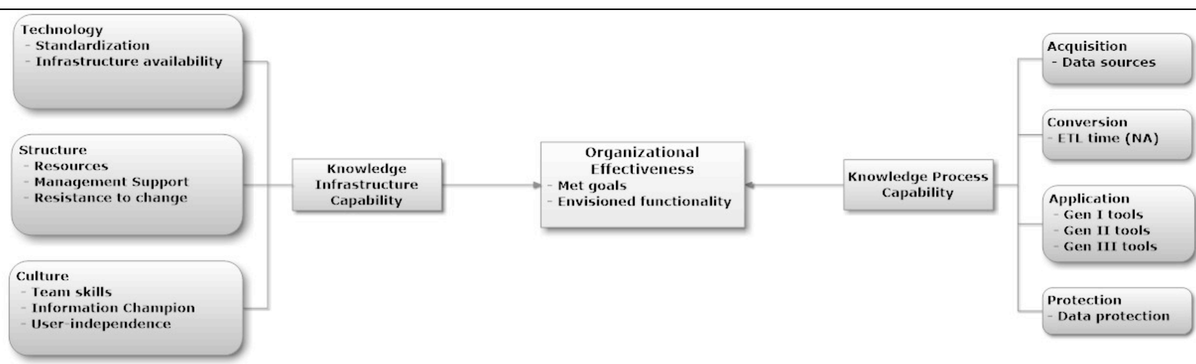
The inequality of capabilities and outcomes is central to the conceptualization of digital divide (Wei et al, 2011) that we believe cloud computing has impacted in the developing countries and in low-resource contexts even in developed countries.

Figure 1: Three-Level Digital Divide Framework (Wei et al., 2011)



We see that the digital divide ranges from individual to organization to global. We look at this model in our research context through the organization level. The first level of digital divide in our cases refers to the gap between health systems with regards to both their opportunities to access ICT and their use of the Internet for a wide variety of activities. The second level refers to inequality in the ability to use an accessible technology. In the above conceptualization of digital divide, the model relies on concepts from Social Cognitive Theory, particularly self-efficacy – "*the belief in one's capability to organize and execute the courses of action required to manage prospective situations*" (Bandura, 1997 p.2). While Bandura (1986) articulates cognitive factors in terms of the individual's capabilities, we draw parallels with Organizational Capabilities (Gold et al., 2001) as the organization's capability to organize and execute courses of action. In health systems with regards to HIS, such capabilities mean ability to establish and ensure smooth functioning of the large infrastructure of medical services and health information capture and the ability to process this information into organizational knowledge (ibid.). To conceptualize capabilities of organizations for knowledge management, we use the Organizational Capabilities framework in terms of enlisted factors as shown in Fig 2. under Technology, Structure, Culture, Acquisition, Conversion, Application and Protection. With the specifics to DHIS2, we label the Operational BI tools under Application as Gen I, Gen II and Gen III tools (Purkayastha & Braa, 2013). The resultant inequality in Organizational Effectiveness or outcomes can be considered as the third-level or Digital Outcome Divide. This can be understood as a gap between meeting goals of healthcare and envisioned functionality after the implementation of HIS.

Figure 2: Organizational Capabilities Framework in Knowledge Management (Gold et al., 2001)

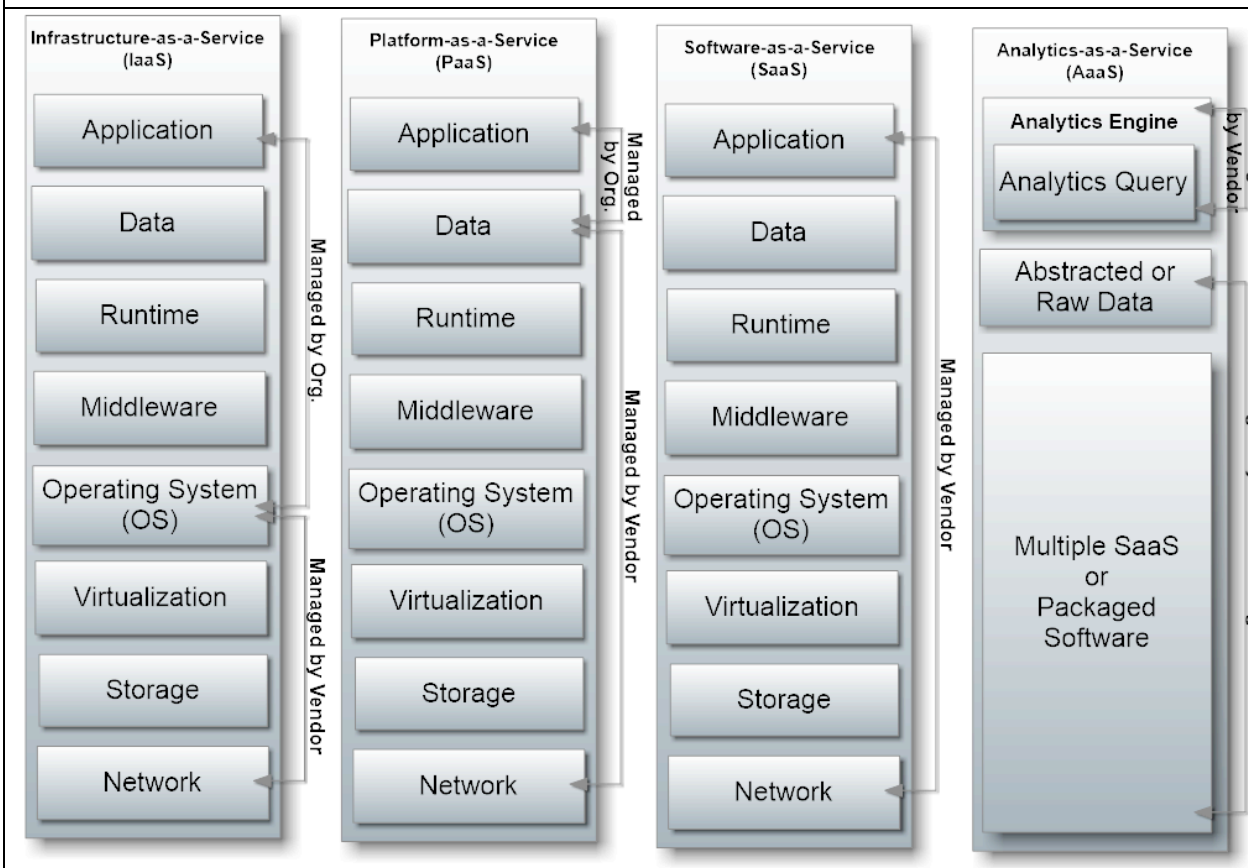


When we talk about access to ICT such as hardware and software within the health context, it is expected to be met through procurement of these in a cost-effective manner. Typically in developing countries, the debate is whether information systems should consume the available limited resources or whether these resources should be put to improving medical services. While this debate is relevant to the context, this paper and our research focus is on the design and use of health information systems. Thus, the access divide, capability divide and outcome divide are understood by us in terms of information processing, instead of health outcomes.

While bringing data together from different health programs and patient record systems, we challenge general techniques for information management. This has been referred to as "Big Data" challenge. With the integration of multiple sources of data into a data warehouse, 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 (Laney, 2001). Weiss and Indurkhya (1998) in computer science and Diebold (2000) 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 also obvious to the "Three V", 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. What if we only compare the incoming data point to a somewhat smaller snapshot of data? A snapshot of data that is relevant to the locale? - Locale being geography and/or health programs and/or facilities and/or health providers etc.

The multiplicity of service models for cloud computing started when consumer internet companies began offering internet-based alternatives to desktop applications, hosted over a number of distributed virtualized servers. So, an organization "*does not buy software license for an application such as enterprise resource planning (ERP), instead a business signs up to use the application hosted by the company that develops and sells the software, giving the buyer more flexibility to switch vendors and perhaps fewer headaches in maintaining the software"* (Dubey & Wagle, 2007 p.1). Such software delivery model from Application Service Providers (ASPs) (Günther et al., 2001) eventually became known as Software-As-A-Service (SaaS). When the software that's made available as a service is a database system, the model is often labeled as Database-as-a-Service (DaaS). From an historical research perspective, it is important to recognize that early visions of turning software into a service (Turner et al., 2003) suggested use of multiple low-level services and providing a larger service using SOA (Gold et al, 2004; Papazoglou, 2003). But when considered today, SaaS is generally a single ASP product with little integration with other open services (Sun et al., 2007; Wei & Blake, 2010) and sometimes deliberately as walled-gardens to cause lock-ins (Cerbo et al., 2012). Thus, in the SaaS model, as shown in Figure 3, all parts of the service stack such as the hardware, storage, networking, OS, runtime, data and application is managed by an external vendor. On the service stack, when the application is managed internally by the organization and everything else is managed by the external vendor, the service model is commonly known as Platform-as-a-Service (PaaS). In the PaaS model, the vendor provides a platform on which applications are written to access underlying services. Social networking platforms, app hosting platforms and custom-CRM platforms on the cloud are examples of PaaS.

Figure 3: Relevant Cloud Computing Service Models



At the lowest spectrum of the service models, when an external vendor manages the hardware (generally based on virtualization) and the organization internally manages the middleware, runtime, data and application, it is known as Infrastructure-as-a-Service (IaaS). The IaaS model relies on the principle that hardware investments become obsolete and organizations do not have to own such capital costs upfront. Rather, organizations rent server clouds that can host runtime, middleware, data and applications for them and the infrastructure maintenance is left to the vendor. IaaS providers offer customizable OS, Runtime, CPU, Disk etc from the stack and offer on-demand changes to the hardware. This is referred to as "elastic site" (Marshall et al. 2010), so that when the application needs more resources, the underlying infrastructure grows as much as the application needs. Such elastic computing allows running analytic services at reduced cost since larger infrastructure is rented only when analytics need to be performed (Buyya et al, 2010 p.106). At other times, basic computing resources from IaaS provider are used and thus lower cost of renting it. When such analytics is provided as a service by an external vendor, this is referred to as Analytics-as-a-Service (AaaS). So the Analytics Engine consisting of algorithms and query processor is setup by the vendor, while rest of the transactional data and required software, platform and infrastructure is at the organization. The AaaS vendor generally hosts the Analytics Engine on IaaS and performs the analytics quickly, possibly co-relating with external data in public domain and gives results back to the organization.

RESEARCH METHOD

The research has been conducted as part of an ongoing, long term action-research project. Our empirical investigations can be broadly placed in the Scandinavian Action Research tradition, which focuses on long term engagement in the field, doing actions and understanding the effects of the actions. Our approach is inspired by networked action research (Elden and Chisholm 1993; Engelstad and Gustavsen 1993), and has elsewhere been termed *Networks of Action* (Braa et al. 2004). The Networks of Action approach is specifically designed for the resource limited conditions in the Global South. The 18 years time-span of the research project exceeds traditional 'projects' and is more akin to social movements (Elden and Chisholm 1993). This network is called HISP (Health Information Systems Programme) due to the origins in the research project and comprises of researchers, developers, implementers, representatives from ministries who share knowledge and learning between the different nodes of the network. The authors of this paper are part of the central co-ordination node of the network.

Data for this research has been generated through active engagement with a global network of developers, implementers and users, who work in health systems of a number of developing countries. Emails, developer meetings, design discussion notes have been secondary sources of data. Other secondary sources of data include earlier published work by the authors and colleagues as well as consulting assignment for countries where DHIS2 has been widely deployed as the national system for health information management. The analysis of the data has been done with an interpretative perspective. Different cycles of interpretation were shared between the authors and multiple members of the network have participated in discussions about the use of cloud services. 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 second author has been part of the research project for the last 4 years, primarily engaged in Asia and is part of the core developer team for the DHIS2 software.

OPERATIONAL BI TOOLS IN DHIS2

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 an action-research initiative over the last 15 years. The software has its roots in Scandinavian Action Research tradition in IS development where user participation, evolutionary approaches, and prototyping are emphasized (Greenbaum & King, 1991). 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 (Braa et al., 2007).

The Operational BI Tools in DHIS2 have followed 3 large generational changes, similar to what has been identified as BI&A 1.0, 2.0 and 3.0 (Chen et al., 2012). These changes from Pivot tables to Data marts to data mash-ups for interpretations within inter-organizational social networks have been done to meet Big Data needs (Purkayastha & Braa, 2013). These highlight the need for information at every level and information that has been analyzed and understood at each level. In the next sections, we describe how DHIS2 deployments have evolved because the Operational BI tools to manage Big Data require larger computational power.

MOVING FROM PACKAGED OFFLINE DEPLOYMENT TO CLOUD

The first versions of DHIS were developed as a Microsoft Access™ application that was meant to be used at each district. Districts would then share indicators and the national level would co-relate the data between districts. This decentralized system was widely acknowledged as an empowering tool for health workers (Williamson et al, 2001). When work on the next major version called DHIS2 started nearly 10yrs after the first release, it was developed as a web application that could be accessed over a browser and could be deployed on a variety of proprietary or open-source OS, Runtime and Databases. Thus, the focus of DHIS2 was for web deployments, but not losing sight that information had to be available for local action such that facilities and districts could analyze and access their own data, as well as probably get comparisons with other health facilities. Since internet access was still limited in countries like Sierra Leone, there were offline installations of DHIS2 at each district office (Kossi et al., 2013). Thus, the DHIS2 web application was installed at District health offices like packaged software and accessed over a web browser. At nearly the same time, DHIS2 was also being implemented in Kenya, where internet penetration over mobile networks was increasing. In Kenya, the plan was initially to do many standalone implementations in districts, but due to availability of mobile internet through USB dongles, it was decided that the Kenya deployment would be done on an internet-based web server and all facilities across Kenya would use mobile internet to access the DHIS2 system through a computer (ibid.). This has been described as Participatory Design in the Cloud and empowered rural communities to participate in the design of the datasets in the HMIS system in Kenya (ibid.). This was a major change from how DHIS2 had been deployed earlier.

Although deployment in Kenya may be described as a cloud deployment, it is really a hosted deployment using IaaS provider like Linode or AWS. These IaaS vendors allowed hosting DHIS2 outside Kenya's health department's servers. This puts the management of infrastructure with an external vendor instead of expecting Kenya's health department to build expertise in ICT management capabilities. Interestingly, since the Kenya deployment, most DHIS2 deployments are now hosted on IaaS externally and in the next sections of the paper we describe how DHIS2 service model has moved to exploit cloud capabilities even more by moving to PaaS, SaaS and eventually AaaS.

DEPLOYING DHIS2 AS A PAAS

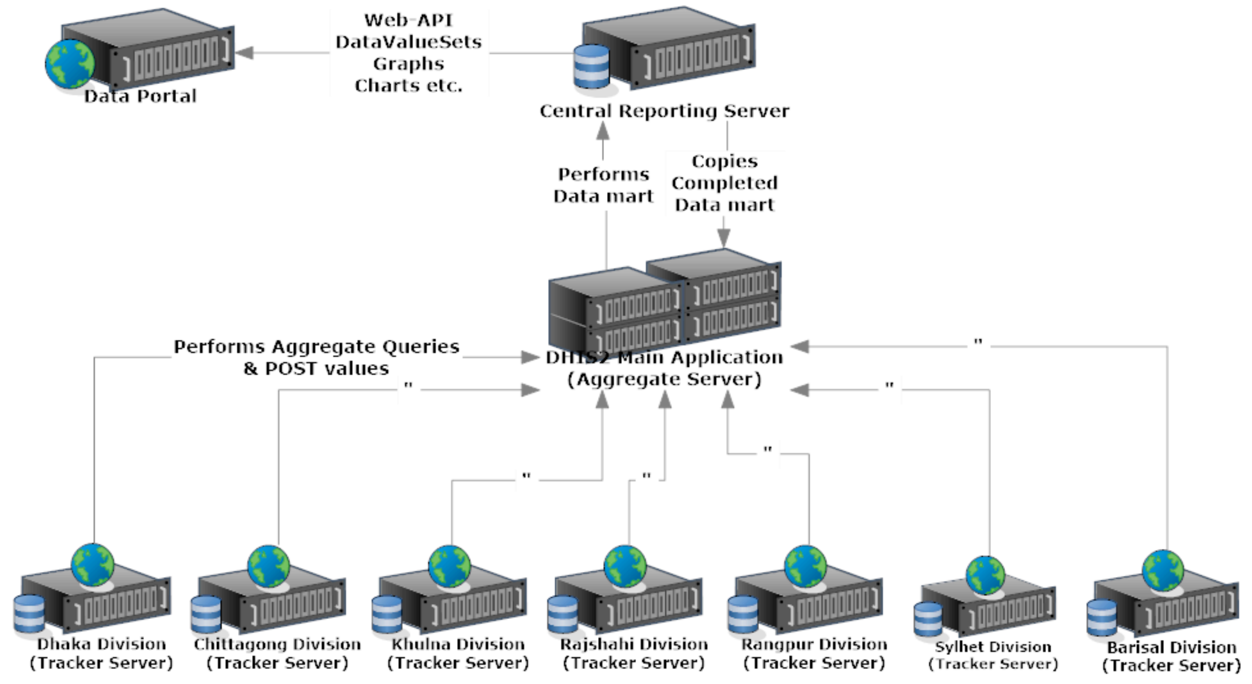
DHIS2 has often been described as a platform by many researchers (Manoj et al., 2013; Lowe et al., 2012). DHIS2 in these instances is a platform for integrating health programs and sharing or comparing indicators from different sources of data. As DHIS2 deployments moved to a hosted environment, the ease of access to data over the internet at any-time, gave incentives to health programs to use DHIS2 as their health data repository (Sæbø et al, 2011). DHIS2 deployment in Kenya took more than a year for starting to get data from all facilities, but it continues to evolve by integrating other health programs. Interestingly, cloud-hosted DHIS2 deployments today not only integrate aggregate data from health programs, there are also attempts to integrate patient-level health information coming from multiple Electronic Medical Records (EMR) systems. In these cases DHIS2 deployment from the national ministry acts as a platform provider for different health programs in the country. The DHIS2 platform is available for customization by the different administrations, where they include their own datasets that are representative of their work practices and still are able to co-relate to data elements from other programs. In Malawi, the DHIS2 system has been deployed as a much more complex PaaS model. Lowe et al., (2012) describe the use of Ministry of Health's DHIS2 PaaS to integrate climate information with rural telemedicine. They use the PaaS platform for statistical and dynamical disease prediction models to be rapidly updated with real-time climate and epidemiological information. This permits health authorities to target timely interventions ahead of an imminent increase in malaria incidence (ibid). Yet, we have only been able to scratch the surface of the PaaS model in the DHIS2 deployments. With Open Data and further activities of Open Government (Janssen, 2011), it seems that DHIS2 PaaS deployment could scale to much larger levels in the future.

These will enable governments to become PaaS providers to citizens, so that better transparency and information is available to citizens to built different applications on top of these. Recent versions of DHIS2 have also included an App framework that allows installation of new user-interfaces and customizations to the DHIS2 REST interfaces. This not only simplifies customizations for the users, but extends DHIS2 platform capabilities to a completely new level.

CONSIDERATIONS FOR CLOUD DEPLOYMENTS

A medium sized country in South-East Asia has deployed DHIS2 as packaged software within the Ministry of Health's datacenter since the year 2011 to manage routine HMIS. Unlike Kenya, they did not want to use IaaS provider mainly because of privacy and data ownership issues. Their DHIS2 deployment is mainly dependent on 2nd generation of Operational BI Tools (Purkayastha & Braa, 2013) like charts, comparative reports between facilities etc. These Operational BI tools require creation of data marts from the ETL process of the DHIS2 data warehouse. At the current stage, the country database for 2 years contains 36 million records until sub-district level consisting of ~9000 organization units. Below this level, there are ~16000 health facilities from which more granular data entry will soon start. The plan is then to capture patient-records from these facilities by converting paper registers from programs into electronic program tracking with DHIS2. Given the current capacity of the server, this scaling to lower level of health facilities and higher granularity of data is extremely computing resource intensive. At present, the data mart service on their DHIS2 deployment on a single, expensive server does not get finished within an expected period of nightly maintenance, so that during the next working day, the resultant analytics can be used. It requires more than 16hrs to complete the process of building the data mart. After the data mart is completed, the resulting analysis is somewhat stale and loses meaning for the facilities that want to track patients on a real-time basis. Thus, to be able to manage such Big Data, the architecture of the HMIS needed to change. Figure 4 shows a federated architecture that allows distributing load and data between multiple servers. This new federated architecture requires buying new servers and setting up of required infrastructure is expensive. The maintenance effort to manage data across multiple servers is also an overhead in the performance, whenever data needs to be co-related between organizations that are hosted in different servers.

Figure 4. – An expensive federated architecture in a SE-Asia country to manage Big Data



The real use of the federated system is only nightly when the analytics service is running, so that patient records from the last day's activities can be used for services that need to be provided on the next day. The limited use of analytics in their current 16hr long data mart process has resulted in lack of perceived usefulness of system by data managers, health administration and district health officers. Other than being expensive, with Big Data the Total Cost of Ownership (TCO) of the system has considerably increased due to maintenance, setup, bandwidth etc. This is expected to increase exponentially as more patient records are captured and more programs are included in tracking of the patient. The 3Vs of Big Data has now resulted in high-demand for computing, only during a short period of that time, that if rented out through an AaaS provider can decrease the TCO.

EXPANDING DHIS2 TO AAAS SERVICE MODEL

While the DHIS2 PaaS service model has been adopted in a number of countries, providing a platform for health programs, organizations, international donors etc, there have been till date limited number of SaaS providers for DHIS2. But in the last year or so, AaaS service model is starting to gain momentum, particularly with integration with patient-level data and need for dynamic reporting. When DHIS2 is hosted by a ministry of health, it publishes a list of indicators that allow meaningful use and minimum datasets that are useful for monitoring and evaluation of services provided by health providers and facilities. The transactional data from these services are located in paper registers or EMR System in facilities. A facility can then capture these paper records as transactions in the DHIS2 tracker system or a separate EMR. They can also just calculate aggregates on paper and report on these as aggregate values into DHIS2. The DHIS2 can then generate the required indicators for the facilities.

While the above mentioned workflow is common, there is a growing need for existing EMR systems with patient-level records to be able to perform analytics on data in their systems and compare with public health data that is available, as well as comparative studies between EMR systems for disease surveillance, outbreak alerts etc. Such analytics are computationally intensive tasks as we have seen earlier. Rather than expecting powerful computers available at facilities, the suggested model of AaaS is where an AaaS vendor performs analytics and provides results back to the EMR system, along with the earlier mentioned workflow of reporting to the ministry of health or another monitoring organization. AaaS service provided by DHIS2 relies on the hosted IaaS and covers the computational load on behalf of the EMR systems. This offloading of resources at the central location, allows better use of the underlying IaaS over time. E.g. while the analytics service from Philippines is running, there is still data entry happening in the Kenya system and Kenya does not require the analytics to be run, until the time Philippines begins its data entry. If such data can be well abstracted such that data between the organizations are separated with strict access controls, there are enough cost savings to justify the bandwidth required to send data to an analytics provider. Thus, the offloading computing utility is better shared with actually sharing the patient-records from the EMR systems. The AaaS service provider only needs to cover the analytics algorithm, abstracted data and the time-to-load from the incoming data from the EMR systems.

Since the Analytics engine runs on IaaS service provider, it can expand its required computing resources on-demand, multi-device access, resource pooling, rapid elasticity and measured service. The EMR systems individually rent only a miniscule amount of the AaaS computing resource and hence do the same type (or even advanced) analytics at a much more reduced pricing.

The DHIS2 analytics API can also be combined with the latest technology innovations from HTML5 standard such as web sockets for low-latency IO and quick response, when compared with earlier AJAX or similar web technologies. The DHIS2 analytics API also produces images, pdf, excel sheets etc. based on the parameters sent to the web service. These resources can be cached on the EMR client-side and reduce the amount of bandwidth required on each web service call. The resources also allow simplified implementation of the client system. The client-side does not have to embed e.g. charting libraries or libraries that convert to different data formats. These data formats can be directly retrieved from the DHIS2 Analytics API over standard HTTP responses.

ANALYSIS & DISCUSSION

To think of our earlier concepts on digital divide, we see each service model as being able to increase organizational capabilities and thus building bridges for digital divide. The DHIS2 deployments at Kenya in comparison to Sierra Leone highlights the significance of deployment over IaaS, as providing access to many different types of users. Shared access to information across different hierarchical boundaries enables improved access to services. This has been made possible by moving for a packaged software model to an offline deployment model to use of a cloud service provider. We can think of this as a bridge to reduce the access-level digital divide. So, while we actually centralized the infrastructure through the cloud deployment in the IaaS model, the access to data actually became much more decentralized. This decentralization helps bridge access-level digital divide as it provides the same services to a larger group of facilities and health providers by allowing them to connect over the internet at any location.

On the other hand, then different organizations, facilities, health programs and users share the data and underlying infrastructure, we see that they can share their interpretations that have been generated from Operational BI tools. This we refer to as Gen III Operational BI tools that are closely linked to building inter-organizational social network over which users share their data interpretations. Deployment over PaaS service model allowed users in Kenya to share their interpretations that are generated from Operational BI Tools. This has helped improved shared understanding of information that is generated in DHIS2. We consider this to be a large leap of capabilities that are now available to organizations due to PaaS deployment. Without this shared access to data and its interpretations, such knowledge capabilities would be very hard to achieve within the organization. While we believe, that such knowledge generation capabilities are as much about access first and then about use, when we see certain users share interpretations, it encourages other users to comment and share their own interpretations of the data. Thus, PaaS helps bridge capability-level digital divide.

As EMR systems are primarily transaction systems that capture all the services provided by facilities or providers to patients, the amount of data generated by singular, disconnected systems become much more meaningful when they can share their data with other EMRs. In terms of continuity of care, it provides patients with much more flexibility to get services at different locations. Yet, when different EMR systems want to communicate with each other about outcomes from patient transactions, there are challenges of semantic and syntactic interoperability (Sheth, 1999). These challenges can be simplified if outcomes can be understood through a common set of minimum datasets leading to meaningful use of data. Such services are easily applicable through AaaS service models, because they are dependent on providing analysis without all the systems having to talk to each other. The analysis from AaaS allows EMR systems to realize the similarities or dissimilarities in the outcomes of the services provided in different facilities. The limited capabilities of the EMR systems and the organizations that use these systems can be increased in terms of analytic capabilities through the use of AaaS services. Thus, AaaS service model helps in bridging the outcome-level digital divide.

While the move from packaged software deployment model to IaaS deployment model helped bridge the access-level digital divide, there is simply more to it than meets the eye. Countries like Sierra Leone and Ghana during deployment had limited technology capabilities within their organizations. The IaaS deployment model allowed focusing organizational resources towards data management, health management and service provisioning rather than learning new technologies for infrastructure management. On the other hand, use of PaaS allows bridging capability divide between the organizations. The shared indicator repository for example, allows organizations to share best practices in health information management at a country-level or multi-country organization (e.g. WAHO). Over the extended DHIS2 platform, the countries can deploy their own application (e.g. Liberia, Togo, Gambia all deploying applications on the WAHO DHIS2) and these applications are only parts that the country needs to manage. Rest of the service stack is all managed by WAHO. Some indicators and datasets are specific to the country level, like the hierarchy of standards (Braa et al. 2007). This data from the service stack is managed by the countries, while the WAHO-level indicators and datasets including the analytics is managed by WAHO. Thus, WAHO here acts as a vendor for the country managing the PaaS infrastructure and providing services to the member states.

DHIS2 deployments off late have not yet adopted the SaaS service model. Basically this is due to a limitation of a vendor that can provide services to multiple countries. Still, there is emerging possibility that universities in East Africa can provide such SaaS services to countries or organizations in the East Africa region. There is a strong business case for a DHIS2 SaaS provider. Global funding agencies might also play an important role in providing seed funding to a vendor who can provide DHIS2 SaaS services here. Though, the conflicts between the organizations need to be resolved first for such a SaaS model to take-off. Countries that are in conflict would not want to have their computing resources shared with others on the same SaaS provider. This could become a strong deterrent and it will be up to the vendor to resolve such conflicts on an urgent basis. Yet, we have not studied independent variables of power in our research and is thus out of scope for this paper.

Although this creates a type of dependence on an external vendor, the organization focuses on building its own capabilities of structure and is able to better utilize its inherent, sentient capabilities. While our surveys do not clearly derive a co-relation between data security and organizational effectiveness (Purkayastha & Braa, 2013), this might be a result of the openness to use cloud computing. If the organizations involved in DHIS2 deployments were in countries that had stricter guidelines to data ownership and place where data needs to be hosted, the results would probably be different. But given that we have been able to actively deploy DHIS2 in countries using cloud computing, we are convinced that the co-relation between data security and perceived use of information system is less strong than what Wixom & Watson (2001) suggest based on studies in developed countries. In DHIS2 deployments, our findings highlight that cultural and structural factors play a larger role with a stronger co-relation between these factors and organizational effectiveness of Operational BI Tools (Purkayastha & Braa, 2013)

In the earlier mentioned SE-Asia country, where data ownership issues meant that they had to invest in a lot of infrastructure, has resulted in a higher TCO for the DHIS2 implementation. The overall direction for the use of DHIS2 on cloud computing is thus positive. There are still concerns of data ownership and many countries or ministries in these countries have fear of snooping, changes to data by vendor organizations have been highlighted in our discussions. We also see that some ministries of health would be willing to offload abstracted patient-level information to AaaS. This finding would be interesting to a number of IaaS service providers and would like to move their focus to marketing their solutions to AaaS developers.

BIG DATA ANALYTICS ON NEXT-GEN HEALTHCARE INDUSTRY

While we see that there is a general shift in the HIS industry from a single EMR to Health Information Exchanges (HIE), there is lack of clarity on how this shift will happen in the developing world. There is the technical challenge of moving from patient health records in a single EMR system to shared health record coming from multiple transactional systems such as EMRs, Lab Information Systems, Human Resource Systems, Drug Logistics Systems etc. But beyond this technical challenge of integration, there is also a larger organizational integration challenge that will involve lot of workarounds (Ellingsen et al, 2013). For example in the U.S, HIE implementations have faced persistent challenges over the past 20 years (Vest & Gamm, 2010) and we expect similar challenges, but also additional challenges of resource constraint and policy formulation will be faced by implementations in developing countries.

The main driving force behind implementing an HIE is mainly of public health significance. Shared health records may allow continuum of care and better patient-care based on preserved medical history might be a side-effect, the larger goal of HIEs is to be able to monitor health programs, government spending and detect disease outbreaks. This primary goal of HIEs has been rightly acknowledged by American Reinvestment & Recovery Act (ARRA), 2009 and similar policies around the world. These policy frameworks and laws put “meaningful use” through the sharing of required set of indicators from health-care providers to the forefront and monetize the use of EMRs through these indicators. This suggests that in the future of healthcare for medical providers to get paid, analytics will play a major role in medical practices (Frisse et al, 2011). As shown earlier, it is difficult for small EMR systems to build in such analytics, purely from an economic ROI perspective, and thus AaaS services will have an important role to play in “meaningful use” of medical information. In terms of HIE analytics, the core technical challenge has been Big Data Analytics and being able to handle large scale transactions from non-standard medical terminology in patient records. Both of these problems may be addressed through large-scale computing resources that can be made available through the cloud. For example in recent years, we see about 73% of trading in the US stock markets uses Algorithmic Trading (AT). This has been possible through co-relating data from various sources and allowing algorithms to self-learn and decide on investment strategies (Chaboud et al., 2013). To achieve the same in health care is somewhat optimistic because of the Variety challenge of Big Data, it is not impossible. Given the exponential rise in computing capabilities through the use of virtualization in data centers that provide cloud computing services, it is quite possible that machine learning algorithm for health care can be effectively designed in the near future. With shared medical terminologies between providers, the Variety challenge can be solved and AaaS cloud models can then easily deal with the Volume and Velocity problems. The Veracity challenge of Big Data is a little more complicated in health care than other industries. Especially in developing countries, reliable diagnosis is a challenge and often medical practitioners re-order tests (Alvarez, 2005). Even with advanced electronic system, this challenge of quality and Veracity of data is difficult to solve.

The strategy that we have followed with DHIS2 is that EMR systems should be able to work together through common medical terminologies at the patient-level and exchange metadata with a DHIS2 data warehouse to send public health information. DHIS2 can then act as an analytics engine and provide feedback about data quality to the EMR systems. Standardizing shared metadata is often the biggest challenge in HIE implementations. Because of the flexibility of defining metadata in DHIS2, different kinds of transactional systems can be integrated through it and made part of an HIE initiative. DHIS2 thus participates in the development of the Open Health Information Exchange (OpenHIE – http://ohie.org) initiative that has been started to create open-source standards, metadata exchange and applications that can be brought together to create an HIE implementation in developing countries.

CONCLUSION

We see that cloud computing service models will continue to evolve over time, with more and more combination of the current models as well as new types of computing utilities being provided. These could become widely accepted, not just based on cost-savings to organizations, but will based on how these models can improve organizational capabilities. How vendors that provide such utility can become assets to organizations to offload tasks that are outside an organization's core competencies. These bring focus to the organization's activities and helps consolidate their existing capabilities. In our experiences in the health system, providing health services is the core competency of the organizations and offloading computing-related tasks to external vendor who provide reliable and efficient services will be the direction for the future.

To answer our original research questions, we definitely see that cloud computing helps bridge digital divide. IaaS (and probably SaaS) helps bridge access-level digital divide. PaaS model helps bridge capability digital divide. AaaS model helps improve outcomes and bridge outcome-level digital divide. In terms of characteristics of cloud services that allow information for local action, the on-demand and scalable services are the core to provide better decentralized access to data and analysis. Especially in terms of cost-savings, AaaS that's running on IaaS is a great leap forward. Capabilities to manage Big Data Analytics that were impossible to meet with limited hardware are now possible with the use of AaaS providers.

In terms of future direction of this research, we'd like to study the privacy and security concerns related to AaaS providers in much more details. In the light of recent events related to government snooping of private information, these discussions are relevant for the future of cloud computing. Privacy is highly relevant to the health context and hence will be important for future studies.

REFERENCES:

Alvarez, R. (2005). Health care has to move into the hi-tech age. *Bulletin of the World Health Organization*, *83*(5), 323–323. doi:10.1590/S0042-96862005000500003

Azvine, B., Cui, Z., Nauck, D. D., & Majeed, B. (2006). Real Time Business Intelligence for the Adaptive Enterprise. In *The 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, The 3rd IEEE International Conference on E-Commerce Technology, 2006* (pp. 29–29). doi:10.1109/CEC-EEE.2006.73

Bandura, A. (1986). Fearful expectations and avoidant actions as coeffects of perceived self-inefficacy. *American Psychologist*, *41*(12), 1389–1391. doi:10.1037/0003-066X.41.12.1389

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Barzilai-Nahon, K. (2006). Gaps and Bits: Conceptualizing Measurements for Digital Divide/s. *The Information Society*, *22*(5), 269–278. doi:10.1080/01972240600903953

Braa, J., Hanseth, O., Heywood, A., Mohammed, W., & Shaw, V. (2007). Developing health information systems in developing countries: the flexible standards strategy. *MIS Quarterly*, *31*(2), 381–402.

Braa, J., Monteiro, E., & Sahay, S. (2004). Networks of Action: Sustainable Health Information Systems across Developing Countries. *MIS Quarterly*, *28*(3), 337–362.

Bruner, J. (2013). *Industrial Internet*. O'Reilly Media, Inc.

Buyya, R., Broberg, J., & Goscinski, A. M. (2010). *Cloud Computing: Principles and Paradigms*. John Wiley & Sons.

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, *25*(6), 599–616. doi:10.1016/j.future.2008.12.001

Cerbo, F. D., Bezzi, M., Kaluvuri, S. P., Sabetta, A., Trabelsi, S., & Lotz, V. (2012). Towards a Trustworthy Service Marketplace for the Future Internet. *The Future Internet* (pp. 105–116). Springer Berlin Heidelberg.

Chaboud, A., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2013). *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market* (SSRN Scholarly Paper No. ID 1501135). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1501135

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS Q.*, *36*(4), 1165–1188.

Chisholm, R. F., & Elden, M. (1993). Features of Emerging Action Research. *Human Relations*, *46*(2), 275–298. doi:10.1177/001872679304600207

Cullen, R. (2001). Addressing the digital divide. *Online Information Review*, *25*(5), 311–320. doi:10.1108/14684520110410517

Diebold, F.X. (2000), Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting," Discussion Read to the Eighth World Congress of the Econometric Society, Seattle, August 2000. Retrieved from http://www.ssc.upenn.edu/~fdiebold/papers/paper40/temp-wc.PDF

Dubey, A., & Wagle, D. (2007). Delivering software as a service. *The McKinsey Quarterly*, *6*, 1–12.

Elden, M., & Chisholm, R. F. (1993). Emerging Varieties of Action Research: Introduction to the Special Issue. *Human Relations*, *46*(2), 121–142. doi:10.1177/001872679304600201

Ellingsen, G., Monteiro, E., & Røed, K. (2013). Integration as interdependent workaround. *International Journal of Medical Informatics*, *82*(5), e161–e169. doi:10.1016/j.ijmedinf.2012.09.004

Engelstad, P. H., & Gustavsen, B. (1993). Swedish Network Development for Implementing National Work Reform Strategy. *Human Relations*, *46*(2), 219–248. doi:10.1177/001872679304600205

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of management information systems*, *18*(1), 185–214.

Gold, N., Mohan, A., Knight, C., & Munro, M. (2004). Understanding service-oriented software. *IEEE Software*, *21*(2), 71–77. doi:10.1109/MS.2004.1270766

Greenbaum, J., & Kyng, M. (Eds.). (1991). *Design at Work: Cooperative Design of Computer Systems* (1st ed.). CRC Press.

Günther, O., Tamm, D.-W.-I. G., Hansen, D.-V. L., & Meseg, D.-K. T. (2001). Application Service Providers: Angebot, Nachfrage und langfristige Perspektiven. *Wirtschaftsinformatik*, *43*(6), 555–567. doi:10.1007/BF03250818

Hoffman, D. L., & Novak, T. P. (1998). Bridging the Racial Divide on the Internet. *Science*, *280*(5362), 390–391. doi:10.1126/science.280.5362.390

ITU, The World in 2013: ICT Facts and Figures. (2013). *ITU*. Retrieved September 5, 2013, from http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, *28*(4), 446–456. doi:10.1016/j.giq.2011.01.004

Keny, P., & Chemburkar, A. (2006). Trends in Operational BI. *DM REVIEW*, *16*(7), 20.

Kimaro, H. C. (2006). Strategies for Developing Human Resource Capacity to Support Sustainability of ICT Based Health Information Systems: A Case Study from Tanzania. *The Electronic Journal of Information Systems in Developing Countries*, *26*(0). Retrieved from https://www.ejisdc.org/ojs2/index.php/ejisdc/article/view/272

Kossi, E. K., Sæbø, J. I., Braa, J., Jalloh, M. M., & Manya, A. (2012). Developing decentralised health information systems in developing countries –cases from Sierra Leone and Kenya. *The Journal of Community Informatics*, *9*(2). Retrieved from http://ci-journal.net/index.php/ciej/article/view/861

Laney, D. (2001). 3-D Data Management: Controlling Data Volume. *Velocity and Variety", META Group Original Research Note*.

Lowe, R., Chadza, T., Chirombo, J., Fonda, C., Muyepa, A., Nkoloma, M., … Zennaro, M. (2012). A platform to integrate climate information and rural telemedicine in Malawi. In *EGU General Assembly Conference Abstracts* (Vol. 14, p. 9503). Presented at the EGU General Assembly Conference Abstracts. Retrieved from http://adsabs.harvard.edu/abs/2012EGUGA..14.9503L

Manoj, S., Wijekoon, A., Dharmawardhana, M., Wijesuriya, D., Rodrigo, S., Hewapathirana, R., … Dissanayake, V. H. W. (2013). Implementation of District Health Information Software 2 (DHIS2) in Sri Lanka. *Sri Lanka Journal of Bio-Medical Informatics*, *3*(4). doi:10.4038/sljbmi.v3i4.5431

Marjanovic, O. (2007). The Next Stage of Operational Business Intelligence: Creating New Challenges for Business Process Management. In *40th Annual Hawaii International Conference on System Sciences, 2007. HICSS 2007* (p. 215c–215c). doi:10.1109/HICSS.2007.551

Marshall, P., Keahey, K., & Freeman, T. (2010). Elastic Site: Using Clouds to Elastically Extend Site Resources. In *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* (pp. 43–52). Washington, DC, USA: IEEE Computer Society. doi:10.1109/CCGRID.2010.80

Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technolog. *National Institute of Standards and Technology*.

Mosse, L., & Byrne, E. (2005). The role of identity in health information systems development: a case analysis from Mozambique. *Information Technology for Development*, *11*(3), 227–243.

Mowbray, M. (2009). The Fog over the Grimpen Mire: Cloud Computing and the Law. *SCRIPT-ed*, *6*(1), 132–146. doi:10.2966/scrip.060109.132

Papazoglou, M. P. (2003). Service-oriented computing: concepts, characteristics and directions. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering, 2003. WISE 2003* (pp. 3–12). doi:10.1109/WISE.2003.1254461

Purkayastha, S., & Braa, J. (2013). Overview, not overwhelm – Operational BI tools for Big Data in health information systems. *Operational BI for healthcare in developing countries*. Retrieved from http://folk.ntnu.no/saptarsp/Operational_BI_healthcare_LMIC.pdf

Qian, L., Luo, Z., Du, Y., & Guo, L. (2009). Cloud Computing: An Overview. In M. G. Jaatun, G. Zhao, & C. Rong (Eds.), *Cloud Computing* (pp. 626–631). Springer Berlin Heidelberg.

Rajan, S., & Jairath, A. (2011). Cloud Computing: The Fifth Generation of Computing. In *2011 International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 665–667). doi:10.1109/CSNT.2011.143

Sæbø, J. I., Kossi, E. K., Titlestad, O. H., Tohouri, R. R., & Braa, J. (2011). Comparing strategies to integrate health information systems following a data warehouse approach in four countries. *Information Technology for Development*, *17*(1), 42–60.

Sheth, A. P. (1999). Changing Focus on Interoperability in Information Systems:From System, Syntax, Structure to Semantics. In M. Goodchild, M. Egenhofer, R. Fegeas, & C. Kottman (Eds.), *Interoperating Geographic Information Systems* (pp. 5–29). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-5189-8_2

Stoops, N., Williamson, L., Krishna, S., & Madon, S. (2003). Using health information for local action: facilitating organisational change in South Africa. *The digital challenge: information technology in the development context. Gateshead: Athenaeum Press Ltd.*

Sun, W., Zhang, K., Chen, S.-K., Zhang, X., & Liang, H. (2007). Software as a Service: An Integration Perspective. In B. J. Krämer, K.-J. Lin, & P. Narasimhan (Eds.), *Service-Oriented Computing – ICSOC 2007* (pp. 558–569). Springer Berlin Heidelberg.

Turner, M., Budgen, D., & Brereton, P. (2003). Turning software into a service. *Computer*, *36*(10), 38–44. doi:10.1109/MC.2003.1236470

Vest, J. R., & Gamm, L. D. (2010). Health information exchange: persistent challenges and new strategies. *Journal of the American Medical Informatics Association*, *17*(3), 288–294. doi:10.1136/jamia.2010.003673

Watson, H. J., Wixom, B. H., Hoffer, J. A., Anderson-Lehman, R., & Reynolds, A. M. (2006). Real-Time Business Intelligence: Best Practices at Continental Airlines. *Information Systems Management*, *23*(1), 7–18. doi:10.1201/1078.10580530/45769.23.1.20061201/91768.2

Webster, P. C. (2011). The rise of open-source electronic health records. *The Lancet*, *377*(9778), 1641–1642. doi:10.1016/S0140-6736(11)60659-4

Wei, K.-K., Teo, H.-H., Chan, H. C., & Tan, B. C. Y. (2011). Conceptualizing and Testing a Social Cognitive Model of the Digital Divide. *Information Systems Research*, *22*(1), 170–187. doi:10.1287/isre.1090.0273

Wei, Y., & Blake, M. B. (2010). Service-Oriented Computing and Cloud Computing: Challenges and Opportunities. *IEEE Internet Computing*, *14*(6), 72–75. doi:10.1109/MIC.2010.147

Weiss, A. (2007). Computing in the clouds. *netWorker*, *11*(4), 16–25. doi:10.1145/1327512.1327513

Weiss, S. M., & Indurkhya, N. (1997). *Predictive Data Mining: A Practical Guide* (1st ed.). Morgan Kaufmann.

White, C. (2005). The Next Generation of Business Intelligence: Operational BI. *DM Review Magazine*.

Williamson, L., Stoops, N., & Heywood, A. (2001). Developing a district health information system in South Africa: a social process or technical solution? *Studies in health technology and informatics*, (1), 773–777.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Q.*, *25*(1), 17–32. doi:10.2307/3250957