Applying Information Analytics to Enhance Rural Water Management Efficiency and Sustainability

Geoff Riggs
IBM GBS – Montréal, Canada
griggs@ca.ibm.com

Cecilia Laverty
IBM GBS – San Diego, United States
cecilia.laverty@us.ibm.com

Nithya Rajamani
IBM Research – Hyderabad, India
nitrajam@in.ibm.com

Hew Crooks
Safe Water Network – New York, United States
hcrooks@safewaternetwork.org

Ravindra Sewak
Safe Water Network India Director – New Delhi, India
rsewak@safewaternetwork.org

KEYWORDS
Analytics, Optimization, Performance Monitoring, Efficiency, Sustainability

ABSTRACT:

Water scarcity is a global issue that impacts the full spectrum of society from developed cities to rural villages. While water provisioning projects are numerous, there is an opportunity to augment management capabilities used to monitor and assess the actual performance and sustainability of decentralized water provisioning networks. IBM and Safe Water Network have developed a business analytics tool that seeks to optimize operations, enhance decision-making, and facilitate the expansion of remotely monitored Safe Water Stations to additional villages in rural India.

CONTEXT

By 2050, India will be among the world's most water scarce nations. Water scarcity is a global issue that impacts the full the spectrum of society from developed cities to rural villages. Ensuring safe and accessible water for remote communities in developing countries is particularly important, as they represent 80% of the world's inhabitants who lack access to improved water supplies. India has appreciated this need since its independence and subsequently included policy and funding to address the rural drinking water crisis since its first Five Year Plan in 1951. Guided by these quinquennial action plans, the State and Federal governments have collectively invested about Rs. 1,50,973 crore (27.96 billion USD) on rural drinking water over the last 60 years. Moreover, it is expected that the Indian Government will allocate Rs. 1,66,000 crore (30.74 billion USD) in the upcoming Five Year Plan for 2013-2018.

Per the last census, India alone has over 640,000 rural villages housing about 69% of the country's population. The Federal and State efforts to supply rural communities with safe drinking water is significant especially given the area of focus, but piped access is currently at about 35%. Consequently, decentralized water infrastructure which utilizes naturally occurring water resources is important to achieving sustainable water security for India's villages.

There are numerous responses to the challenge of developing sustainable decentralized water solutions in India. Many solutions have an infrastructure component, which requires installation and maintenance costs, and sometimes micro-financing funding arrangements. Often the management of the low technology water apparatus (*e.g.,* hand pumps, rainwater harvesting tanks), is eventually transferred to the community. As such a pervasive challenge, especially in rural areas, is the difficulty in obtaining accurate status data to support monitoring and evaluating the performance of the water operation. A lack of good data is part of the impetus behind the rapid deployment of cellular devices throughout India and many developing countries. Affordable and abundant information sources that can send and receive data and reports are affecting a sea change in the ability for even small organizations to better track the actual water provisioning performance of remote or rural towns and villages.

The prototype project described in this paper proposes that information analytic techniques which have previously applied to larger, complex operational environments, can enhance the viability and sustainability of small water aid organizations that work with comparatively modest amounts of data. While the evaluation phase is still in progress at time of writing, we hope to demonstrate the following benefits: First and foremost, performance metrics that can objectively demonstrate project goals are being met are essential for rural water development projects to remain attractive and accountable to investors. Second, analytics that display real-time (or more current) operations information supports more responsive and even proactive maintenance needs, which ultimately supports cost savings. Third, gleaning better intelligence from data analysis can help water providers better understand their consumers by shedding new perspectives on sales behaviors, drivers and impediments. Fourth, richer, easy-to-interpret performance reports support more timely and comprehensive business decisions. Fifth, more robust data analytics and reporting helps to create more meaningful baseline performance objectives, which in turn supports planned growth activities. Lastly, to avoid being hampered by the surging volumes of data becoming more prevalent even in remote villages, aid organizations like Safe Water Network will need more efficient tools and processes. Information analytics (sometimes referred to as Business Intelligence) tools and techniques can reduce the often significant effort associated with manual data manipulation, allowing staff to focus on higher value activities.

US-based NGO (Non-Government Organization) Safe Water Network approached IBM to jointly design, develop and pilot a right-sized business analytics tool that seeks to optimize operations, enhance decision-making, and facilitate the expansion of remotely monitored Safe Water Stations to additional villages in rural India. The prototype aims to enable more efficient and sustainable management and monitoring of decentralized water kiosks that currently utilize a modest amount of control technology (*i.e.,* remote monitoring) and data (collected through web browsers, spreadsheets) water resources through the application of information analytics tools and techniques. This data-centric project supports a broader Smarter Villages vision, wherein all the critical functions of rural or remote communities can benefit from next-generation information management techniques typically associated with larger enterprises and municipalities.

SAFE WATER NETWORK OPERATIONS

Safe Water Network has chosen to focus these initiatives in areas that include a rural population without access to piped water and living under $4 USD a day. A majority of India's population live in rural areas and approximately two-thirds (53.7%) of those people fall in the lowest 40% of economic prosperity. Thus more than half the rural population falls within the economic target levels set by Safe Water Network.

Safe Water Network employs two primary methods to improve water supply: rainwater harvesting, and water purification kiosks. Safe Water Network has currently facilitated the implementation of 20 water kiosks in villages around India operated by local communities or entrepreneurs. The kiosks use ground water and Reverse Osmosis (RO) technology to extract and treat local groundwater and a water charge is recovered from the individual villagers for the purified water. Safe Water Network has developed a remote monitoring platform that collects both operational and sales data at 16 of the kiosks based in India, which collectively service approximately 50,000 villagers. The remote monitoring infrastructure typically consists of networked remote monitoring system, radio frequency identification (RFID) cards, manual and automated data capture means, and access to a regional cellular network. The data is typically consolidated via web forms and spreadsheets, and then uploaded to central servers for management analysis and decision-support. Sometimes manual transcription (handwritten) at the village level is required to get financial data into a transferable digital format.

The treatment systems require monitoring of the source water, plant operations, and reject water. Automated tracking of many attributes is accomplished through on-site sensors. Some of these measurements consist of water quality attributes (*e.g.,* Total Dissolve Solids (TDS)) and others are specially related to plant operations (*e.g.,* pressure, water levels, recovery percentage). PH and residual chlorine are presently read and manually fed on the local console.

RFID cards are issued to each household in the village, and are loaded with payment credits with which water can be purchased. The same cards are used to track sales data. The sales data is then transferred using through 2G and 3G cellular communication systems. Data concerning distribution point performance (*e.g.,* sales and throughput per site, tracking sales per user, and buyer behavior) can be tracked and evaluated on a hierarchical basis (*e.g.,* country wise, state wise, and district wise).

Remote monitoring of both the treatment system and Point-of-Sales data is currently accomplished using a web-based application. Automated alerts, alarms and reports are also generated and transmitted to appropriate staff. The remote monitoring application provides a visual representation of the water treatment plant operations including monitoring water flow and various water quality measurements throughout the system.

This monitoring capability is vital for distributed management and central awareness of individual plant operations. The ability for remote monitoring and automatic operation of many of the water treatment plant operation is critical for the expansion of this program.

SMALL ORGANIZATION DATA CHALLENGES

The ability to monitor and evaluate operational performance begins with collecting timely, reliable, and accurate data. Data collection for rural water development projects often relies on field level operators, community members, or similar to manually collect data. Recently, with the ever-widening use of affordable Information and Communications Technology (ICT) like mobile phones, richer information is becoming more prevalent and accessible. However the means by which organizations are able to capture, process and convert data into business intelligence is often inefficient, not standardized, and only supports basic reporting needs.

Data management is not a core competency of most decentralized water provisioning organizations, yet most must process an ever-increasing amount of data, and from it attempt to extract business intelligence. As such, data can be considered both a risk and an asset; if not used wisely it represents extraneous effort, and potentially ill-informed decisions. If managed properly, data can be converted into actionable insights. As information technology targeting larger enterprises matures and decreases in cost and complexity, the capabilities ‘trickle down’, and can be applied to smaller, more numerous organizations that employ the foundational technology infrastructure necessary to support advanced analytics and optimization techniques. Like many small organizations that are limited in their ability to devote significant resources to data management, Safe Water Network has found that basing operational and strategic decisions on a series of heavily customized spreadsheets is inefficient. Managers spend significant amounts of time conducting ad-hoc data analysis, which impedes optimal resource allocation and decision making. Additionally, as the number of tele-metered kiosk locations grows the challenge of translating data into organizational intelligence increases. Consequently, traditional data processing techniques become unsustainable and unfit for purpose because they are unwieldy, and cannot facilitate the desired continued expansion of the organization. Thus, inefficient and unsustainable information management methods affect the organizations' ability to serve more of the thousands of water-stressed communities in rural India.

Strategically, information sharing and collaboration is increasingly being used by consortiums of organizations that are tackling multi-faceted, complex problems, such as water scarcity. Complex systems dynamics considers interactions and possible trade-offs between overlapping complex environments such as natural and human environments, by facilitating the analysis of a high number of attributes, factors, and actual and possible events. Systems dynamics is a component of modern analytics and optimization tools and techniques, in that previously complex datasets and information can be organized and presented much more easily, with the aim of sharing information across domains to support more holistic solutions. Sharing reliable and well-structured information facilitates trust and innovation, both of which are important advantages when addressing complex water environments that comprise unpredictable environmental (climate change, water levels), social (new technologies, old traditions) and economic (funding, sustainment) challenges. Small organizations like Safe Water Network actively engage other NGOs, funding organizations and industry partners. In the modern, interconnected, data-centric world, it is essential for organizations of all sizes to have well modeled, current, reliable and easily accessible information. The ability to share and leverage intellectual capital is a critical enabler of more efficient, innovative operations.

Sharing information is also a vital success factor to help villages adopt new water provisioning services. Cultural norms and behaviors can often be at odds with new methods of resource utilization, and technology-supported water distribution may meet with resistance from community elders. Encouraging the use of new tools and techniques may first involve educating villagers on the negative impact of current practices. Once a new source of safe water is made available, a plan must be in place to ensure it is adopted by a reliable, committed local entrepreneur or community, with the help of local key opinion leaders. At each stage there are important data points that can inform management of potential implementation challenges, and once the kiosk is operational, timely and comprehensive performance tracking becomes essential. Even the modest amounts of data that Safe Water Network and similar NGOs process can be daunting for often over-worked supervisors and managers. When data is properly structured in a relational database, and user-friendly 'views' of that data are made available through an intuitive interface, the real-time challenges that face new kiosks can be flagged more quickly, and the local community can be engaged more effectively.

TECHNOLOGY DESCRIPTION

IBM's business analytics prototype provides enterprise-class capabilities in a solution architecture that is tailored for Safe Water Network and entities of similar size and complexity. The solution is designed to enhance Safe Water Network's existing information processes and response times so that managers and village operators can act more efficiently based on insights gained from better information.
The organization's operational and financial performance metrics are extracted from the various source systems, validated, and loaded into a centralized data mart. Running the raw data through this ETL (Extract, Transform, and Load) process ensures that quality information is stored and that Safe Water Network has confidence in knowing that it is taking action on accurate information. By conforming to standard business dimensions of Accounts, Customers, Time and Safe Water Station, Safe Water Network can now gain a complete picture of operational and financial performance.

Once the information has been captured in the centralized data mart, it is modeled to reflect how Safe Water Network views its business. Organizing the model into business processes of Sales and Plant Production allows Safe Water Network to intuitively interact with it. The IBM analytical tool allows Safe Water Network managers and supervisors to interactively drill down into current and historical details to shed light on anomalies, or create their own custom analysis for strategic planning and reporting. Instead of basing critical capital and operational planning on manually generated spreadsheets, management will be able to quickly interpret and process richer information with enterprise and site-level views, as well as reports that can be easily customized to be useful for different audiences, from senior managers down to village-level operators.

The first accomplishment of the solution allowed Safe Water Network to gain a very accurate understanding of a Safe Water Station's operational performance. Safe Water Network captures a tremendous amount of operational metrics such as temperature, pressure, pH, voltage, etc but their previous tools did not permit easy conversion of raw data into actionable business intelligence. The IBM prototype allows Safe Water Network to analyze a plant's operational performance for only when it is on-line and in a graphical format that is easy to understand. Using analytics allows Safe Water Network to improve operational efficiencies, avoid equipment failures and maximize production up-time. The image below is one of several report views that the Safe Water Network management team uses to gain insight of kiosk operations.

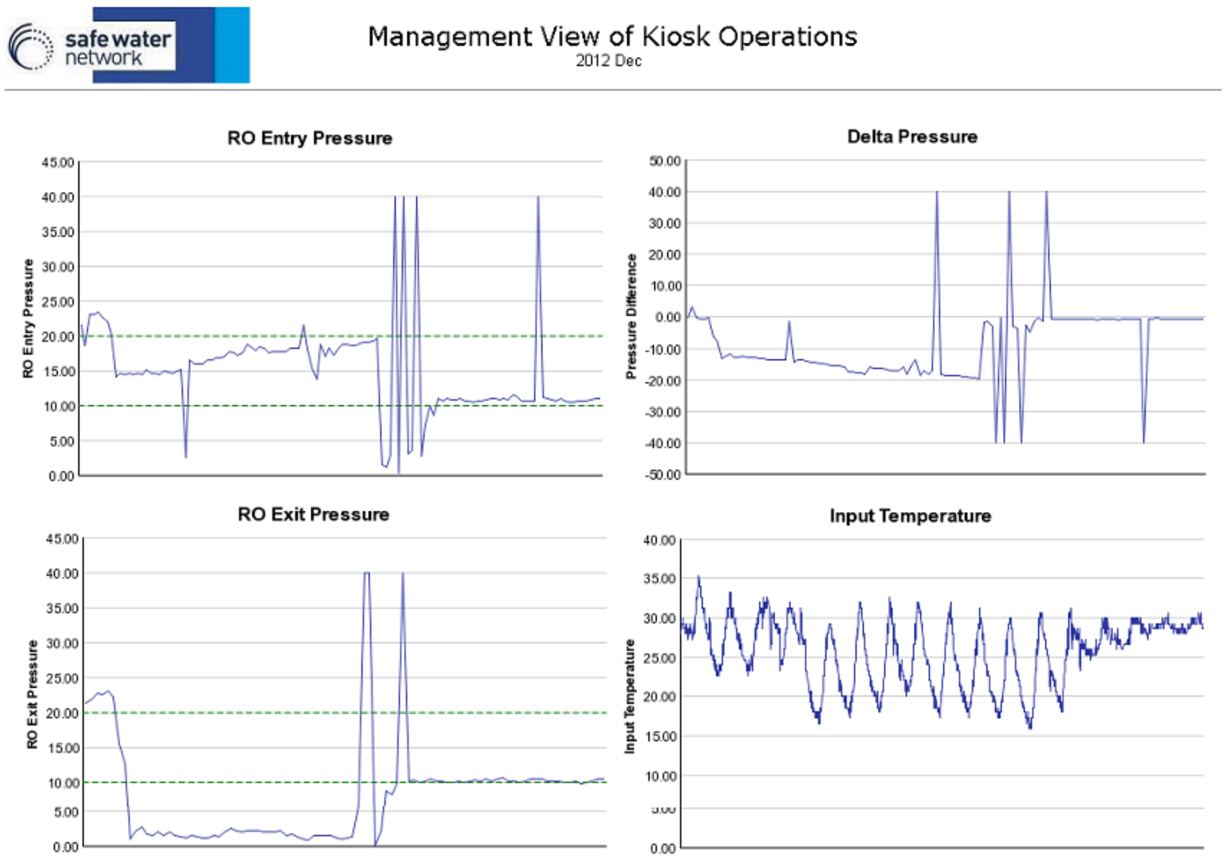

Figure 1. Operational Analytical Dashboard, displaying different operational characteristics for a particular site

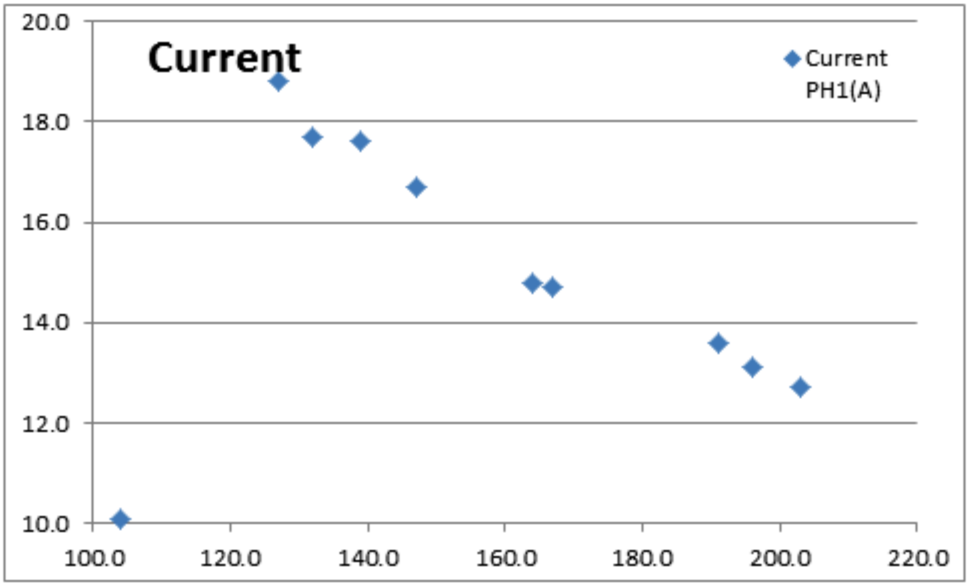

Figure 2. Impact of voltage on motor current, from village Asmadpur Charoli, 24-hour period

The report above highlights the current produced by the pump motor at different current levels. This is an important real-time operational data point because supply voltage will lead to burning of motor as it is operating at almost 50% higher than its rated currents. Monitoring and prevention is more efficient than costly repairs and a clear view of operational status supports more timely intervention.

SUSTAINABILITY OF SOLUTIONS

The challenges of clean water accessibility, availability, quality and operational efficiency in rural India can be addressed by improved tracking of the performance of the water network, providing a real-time (or near real-time) picture of issues, problems and throughput. Supporting this type of management is a new generation of information analytics tools and techniques often referred to Business Analytics and Optimization, capabilities which fall under the broad heading of ICT, and are central to the prototype project described in this paper.

As with many other ICT initiatives, the first beneficiaries of this type of analytics approach have been larger utilities in developed countries, such as DC Water in the United States, and large agricultural users in water-constrained environments. Indeed the technology community has developed a myriad of products to address urban needs for improved water management; however, there has been comparatively less attention paid to rural needs, especially in developing countries. IBM and Safe Water Network are shifting the paradigm by targeting rural and relatively low-technology villages and communities with an efficient and sustainable business analytics solution.

Sustainability is a primary consideration when designing water management solutions with lasting impact, and encompasses economic, social and environmental criteria, as well as the design and construction of the solution itself. Economic (*i.e.,* cost) structures must be sound for the water provider to continue to operate; the pivotal nature of water in functional human systems, along with strong cultural attributes of water use are lynchpins in determining social sustainability of a project; and natural ecosystems' interaction with the water provision system are a *sine qua non* condition for the durability of these water systems.

For any organization to build a truly sustainable water operation, the sustainability of the water infrastructure must extend to all its peripheral functions: governance, finance, human resources, marketing, and of course, technology; and so, as with physical water infrastructure, sustainability has been a fundamental consideration in the design of the ICT solution for Safe Water Network.

Collectively the project team has defined sustainability as a set of conditions that allow for lasting positive impacts of ICT for Safe Water Network and its objectives. WaterAid outlined the obstacles to sustainability in the water sector:

> "There are many inter-related reasons why the achievement of sustainability poses such a challenge to the WASH (Water, Sanitation, and Hygiene) sector. There are three particularly important reasons which stand out. The first is the limited capacity (in the sense of knowledge, skills and material resources) of communities, local government institutions and other service providers to manage systems. The second is the inadequacy of financial revenues to cover the full operation, maintenance and capital maintenance costs of infrastructure. The third relates to the historical approach to service delivery of different actors in the WASH sector. This has been carried out in a fragmented way, with competing agendas and a general disregard or lack of understanding of government frameworks. WaterAid has sought to combat this issue in the past. Although coordination between different sector players has seen improvements in recent years, hindrances have been entrenched for a long time and their legacy continues to frustrate progress."

An ICT solution must therefore address these challenges.

The first challenge of limited local capacity can be overcome by providing a solution that is *accessible*; in addition, the inherent design of the solution, primarily meant to address *security*, places a high emphasis on cloud computing, where complex resources are located and managed centrally by an ICT provider rather than by Safe Water Network or its local plant staff. This reduces local technology requirements such as power and security for physical equipment and therefore one less thing for on-site operations to have to manage.

Accessibility: To ensure accessibility, the tool must ensure that the content is accessible from different geographic areas and through different devices on different platforms. The ICT solution at Safe Water Network provides access to data around the world, from any internet connected device, such as a personal computer or a digital tablet, and as such will make better information more readily available to remote managers and local site supervisors and operators.

Security: The solution's security is maintained by backing up all software and data in a hardened data center that resides in the United States while adhering to all applicable regulations. As such, there is no physical technology footprint required in the villages, ensuring that local capabilities (or lack thereof) are not barriers to ICT-based implementations. The fundamental principle of cloud computing, which is used in this case, is to centralize most functions that can be efficiently centralized, while pushing out to the "edge of the network", in this case the village, only minimal infrastructure needs (an internet connected device.) Centralized functions are managed in a shared infrastructure, leveraging the scale of the cloud provider to provide benefits to smaller entities, such as Safe Water Network, its partners and the villages it serves.

The revenue hurdle can only be overcome by providing a low-cost, *affordable and scalable* infrastructure.

Affordability: Cost affordability is a fundamental design consideration for non-governmental organizations like Safe Water Network, and therefore the analytical prototype used is purpose-built for small and medium sized organizations. While NGOs are often categorized as public sector organizations by ICT providers, their size and scale is more akin to small and medium-sized enterprises that share more of their characteristics.

Scalability: Scalability works in two ways, first by ensuring support for future growth; in this case, the platform used can be scaled up to one that could be used by a Fortune 100 company without having to rebuild any of the analytical content. The complement to this scalability is the notion of variable cost; the scaling approach means that at the entry point, when needs are relatively lower, costs are lower, and increased capacity is only acquired when needed. Safe Water Network does not need to invest proactively to anticipate and account for future growth, because the ICT capacity will simply follow the system's requirements as they evolve. This also lowers the need for Safe Water Network to accurately predict its needs and the time-scale of its growth, allowing it further flexibility to grow as varying conditions (political, economic and social) allow.

Finally, WaterAid suggests that a fragmented set of competing entities may undermine the possibility of cogent, sustained improvements in water provision. Finally, WaterAid suggests that a fragmented set of competing entities may undermine the possibility of cogent, sustained improvements in water provision. Safe Water Network pursues discussions with multiple state governments and other national structures about using remote monitoring technology to provide improved visibility and performance to government-sponsored and funded projects. Well-structured and timely data that has been converted into information that is easier to share allows SWN to demonstrate higher accessibility of performance information, and facilitates sponsor confidence.

CONCLUSION

Information sharing is one of the most fundamental and powerful mechanisms to achieve better performance from any collaborative enterprise. As more effective means of information sharing has spread to the farthest corners of the planet, so too has the ability to better monitor, measure and manage a full spectrum of operations. Water is so critically important to sustaining life and prosperity that it is obvious that every means possible to ensure its quality and equitable distribution must be employed. In large urban environments and complex facilities, advanced information management techniques are becoming the norm. In areas of extreme water stress, the need to employ these same capabilities is obvious and urgent. Strained human and natural resources must constantly compete for equally scarce funding, yet many projects are not tracked appropriately, and much funding ultimately is wasted on unsustainable initiatives.

Unlike water, data is more and more an abundant resource, but in and of itself it presents limited value. Turning data into insight reveals hidden value, and is a strong enabler for behavioral change. More and more, villages will generate and share data. The same information management techniques employed by highly evolved organizations can and should be applied to water provisioning operations that use and rely on good data. Good information provides strategic confidence, and facilitates stronger partnerships and relationships.

As analytical capabilities and skills are institutionalized, the Safe Water Network staff will be better equipped to process the surging volumes of data generated in their environment, and hence have more time to analyze and extract the embedded intelligence. With more time and new insights, the same staff should be better able to identify anomalies and trends that support more effective operational and strategic decisions. Efficient, accurate monitoring of site-level operations will help Safe Water Network to allocate support where most needed, and more quickly. A system-wide view of performance trends will support long-range planning and identification of where the greatest needs might arise, and the factors present in previous successful kiosk sites that should be replicated in future sites.

Ultimately, a stronger information management structure will facilitate easier comparison and cross-referencing of water information with data about other critical village services, such as education, health and sanitation, and more broadly to communications and logistics services. All of the fundamental support structures of a village do or will use data to strengthen their efficiency and sustainability. By providing real-world context and, Safe Water Network can support this shift to a Smarter Villages paradigm with the evolution of this analytics prototype for water kiosk management.

AKCNOWLEDGEMENTS

The authors would like to acknowledge the time and input of Hew Crooks, Senior Vice President, Operations (United States) Safe Water Network organization. The authors also benefited from input from Jean-Francois Barsoum, IBM Senior Consultant (Canada), Carey Hidaka, IBM Senior Consultant (United States), Ned Bader, IBM Project Software Engineer Manager (United States), and a strong IBM technical team consisting of David Zyla (Canada), Sadanand Shastri (United States), Mezjan Dallas (India), Amit Merchant (India) and Rob Chillingworth (Canada). The IBM sponsors and support team for this project are Cameron Brooks, Dale Davis and Lara Kinas.

REFERENCES

Chen, K., Hellerstein, J.M., and Parikh, T.S., (2011): Data in the First Mile, 5th Biennial Conference on Innovative Data Systems Research (CIDR '11), Asilomar, California 9-12 January.

Chandramouli, C., (2011b): Census of India 2011 Provisional Population Totals Paper 2 of 2011 India Series 1, Office of Registrar General & Census Commissioner, India, viewed 23 February 2013, On line at; <http://www.censusindia.gov.in/2011-prov-results/paper2/data_files/india/paper2_1.pdf>.

DCWASA Platform for Spatio Temporal Analytics, The Computerworld Honors Program, On line at: http://cwhonors.org/viewCaseStudy2010.asp?NominationID=96&Username=dcws

Hillery, Jon (2013): Viewed on February 23, 2013, On line at: <http://www.globalgiving.org/projects/safe-water-to-children-in-india/updates/

Indian Water Review (2013): Govt plans Rs 166,000 crore spend on rural water supply, news story released 25 February, viewed on 26 February 2013, On line at: <http://www.indiawaterreview.in/Story/News/govt-plans-rs-166-000-crore-spend-on-rural-water-supply/1072/1.

JMP (2012): Progress on Drinking Water and Sanitation 2012 Update, viewed on 24 February 2013, On line at: www.wssinfo.org/fileadmin/user_upload/resources/JMP-report-2012-en.pdf.

Maheshwari, P., Kapoor, S., Kummamuru, K., and Pingali, G., (2011): ICT Enabled Smart Rural Aggregation Platform, 2011 Annual Service Research and Innovation Institute (SRII) Global Conference, IEEE Computer Society, pp. 597-600.

Market Feasibility Study, Safe Water Network and John Hopkins University, On line at: <http://www.safewaternetwork.org/pdf/Market%20Feasibility%20Study%20Note%20Apr%2021%202010.pdf>

Ministry of Drinking Water and Sanitation (2012): Annual Report 2011-12, viewed on 27 December 2012, On line at: http://www.mdws.gov.in/documentreports/term/64.

Terri and Nasscom (2011): Sustainable Tomorrow: Harnessing ICT Potential, viewed 26 February 2013, https://docs.google.com/gview?url=http://www.nasscom.org/sites/default/files/researchreports/TERI-NASSCOM_Green_ICT_Executive%2520_Summary_Dec2011.pdf&chrome=true

Thompson, P., Hope, R. A., and Foster, T., (2012): Is silence golden? Of mobiles, monitoring, and rural water supply, Waterlines, 31:4, pp. 280-292.

Water for People (2013a): *Akvo FLOW*, viewed 23 February 2013, On line at: http://www.waterforpeople.org/flow-mapping/.

WaterAid (2011) Sustainability framework, On line at: www.wateraid.org/publications

Water and Related Statistics. Viewed 20 February 2013, On line at: http://www.cwc.nic.in/ISO_DATA_Bank/ISO_Home_Page.htm

Weber, J.S., and Toyama K., (2010): Remembering the Past for Meaningful AI-D, 2010 Association for the Advancement of Artificial Intelligence (AAAI), Spring Symposium Series, Palo Alto, California, pp. 97-102, viewed 15 February 2013, http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewFile/1125/1360.