Scientific Data Infrastructure for Sustainability Science Mobile Applications

Katherine G. Herbert, Emily Hill, Jerry Alan Fails, Joseph O. Ajala, Richard T. Boniface, Paul W. Cushman
Department of Computer Science
Montclair State University
Montclair, New Jersey, USA
{herbertk, hillem, failsj , ajalaj1, bonifacer1, cushmanp1}@mail.montclair.edu

***Abstract*— With the recent ecological and volatile climate issues, numerous concerns have arisen which have led to the rapid growth of sustainability sciences. In sustainability studies and related areas such as crisis data management, multiple communities need to interact and contribute data, and then have this data modeled for them in an effective manner. The National Science Foundation Advisory Committee for Environment Research and Education states that, "Because of the complex relationships among people, ecosystems, and the biosphere, human health and well-being are closely linked to the integrity of local, regional and global ecosystems." In our work, we look towards developing a mobile application platform that allows data integration for multiple information sources that allows the user flexibility to learn about and actively participate in understanding and helping their environment. In this paper, we address our current work with the scientific aspects of this data.**

Keywords – Scientific Big Data, Sustainability Science, Data Integration

I. INTRODUCTION

Recently, we have seen multiple "Storms of the Century", where large weather systems have come into a region causing havoc on the local community and environment. Various weather models are showing that there may be a fundamental shift in the weather patterns for these areas, potentially leading to more severe weather and storms that can impact these regions similarly in the future. This has highlighted additional environmental issues that have put attention on the concept of sustainability. As a part of sustainability, the field looks towards how decision maker actions can be impacted in making choices that will help environmental and ecological concerns.

As part of the decision making process, data collection, integration and dissemination are key to making informed sustainability related decisions. Data are important to the society as they serve as primary mechanism for communicating scientific results to users and community at large. Many areas of human lives including health, weather, safety, social, environmental, and security are elucidated by these data.

Our project looks to help scientists create, generate and administer data so as to manage the increasing expectations of the society at different times on different critical issues. This platform gives scientists a flexible platform that enables them to create a database that meets their needs, which will facilitate an easier and faster decision making process. It also helps to make the data available in a format that other stake holders and decision makers. This can include citizen-scientists, executives and other policy makers, who may not have scientific background.

This project is part of a larger project to develop a Sustainability Studies Mobile Toolkit (SSMT), which we are currently developing to address the data collection, analysis and decision making issues in the field for environmental, ecological and biological scientists. This overarching project looks to creating a collaborative mobile environment for scientists, citizen-scientists and other decision-makers to collaborate on sustainability related issues. For this project, the toolkit enables the ecological/biological scientist to record and interact with his or her data via a tangible interface device (iPad or Android). It visualizes data to help scientists and other decision makers design and display analytics for scientist-decision makers to better understand the implications of their work in the field. It also helps facilitate the scientist to release his or her work to the citizen-scientist via the iPad or Android application so that the decision maker with an average citizen's understanding of science can observe, participate and become involved in sustainability concerns. In the interest of space, we have not included a full survey of the state of the art for this field. However, please visit bit.ly/GeotaggerRW to learn more about our project and related work.

II. SCIENTIFIC DATA CHALLENGES

For the scientific decision maker, many factors become involved when using a mobile system to help with data services. When working with our ecological collaborators, there were many challenges that needed to be overcome before our collaborators could adopt the technology. Essentially, the technology needed to be flexible enough to have a zero-time learning curve, allow for collaborative exploration of the experiment (since the experiment could be a location in a South American bay or a controlled location in a university laboratory) and easy, reliable data collection that enables the scientist to control dissemination of the data for publishing and protection of data for research purposes.

The scientific data itself also presents challenges. It can take different forms depending on the focus of the scientists and the interests of the end users. Sometimes scientists work on similar projects but focus on different results. The implication is that scientists need an application that provides them with flexible data models to accommodate this need. When looking at this data only, the scientists can interact

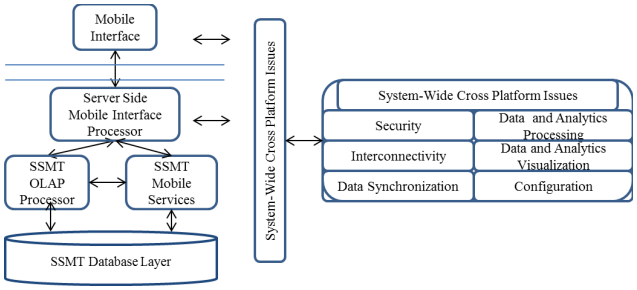

Figure 1. The Sustainability Science Mobile Toolkit structure

with data from their research. When coupled with aninteractive citizen-science application, the data set size explodes increasing the challenges that exist in providing analytics on this data. Therefore, there is a need to provide a platform that will be flexible enough to enable scientists to capture dynamic scientific data and present it in a way that is meaningful to their colleagues and other end users.

III. Scientific Data Framework

Figure 1 illustrates the architecture for our system, the Sustainability Science Mobile Toolkit (SSMT). In this system, we have a number of data integration layers and data analytic layers to service the various populations using the tool. Important for the scientist user is the flexibility to specify the structure of their data. We initially used a NoSQL approach to solving their data needs. We used this platform to create a common storage platform that meets the need of scientists that are working on related but different projects and ensure that their focus on individual project interests is protected. The platform is designed to provide a flexible and dynamic platform to help scientists collaborate on projects and promote sharing of data and information. The system allows for sharing and merging various scientific data generated from different or similar research. Our scientists' current practices of using spreadsheets or word processing applications for this task can be cumbersome, time consuming, and the final output may not be presentable. For resilience purposes, the tool needs to cache on the client side. It was challenging to accomplish this with our original NoSQL implementation. We have moved to using a relational database platform so that client-side caching can mirror the server-side data. This caching is critical when the server or network services are unavailable. Uploading and synchronizing with the server resume once network services are restored.

IV. Conclusion and Future Work

In order to learn about and develop the field of sustainability science, we must come to understand how all factors influence one another. Sustainability science melds science, health, human behavior, and decision-making, with each of these areas having their own history, requirements and concerns. Many views of sustainability tend to look at the macroscopic problems, taking a higher-level view that does not enable adequate consideration of interdependencies between these cross-disciplines. To better understand complex systems and make more informed decisions about community interactions that directly affect the success of habitat restoration and sustainability, scientists need to collect data capturing these complex interactions. This data capture needs to be modeled quickly so scientists can make decisions about their experimental models in the time frame necessary, and the data must be accessible and interpretable to other decision makers so that their findings can permeate other sustainability-impacting activities. This can result in extremely large amounts of data collected that have many different needs, including real-time data analytics, data synchronization, and security.

A significant challenge that scientists face in understanding and modeling these systems is to distill meaningful patterns from large amounts of field and experimental data [4]. Thus, there is a critical need for tools to convert this data into information that can support building sustainability ecosystems [2]. A scientist's data collection must be natural, intuitive and aid rather than hamper their work. In addition, observations made by scientists in the field need to be translated back to the scientist in a form that he or she can use to make decisions immediately. These data need to be available and in a format so that other decision makers, who may not have a scientific background, can access and interpret the data. This data is diverse, often representing years of study, consisting multimedia, heterogeneous data sets that can manifest many Big Data issues. Ideally, such a *decision dashboard* can be integrated into a mobile platform so that there is seamless integration of data collection and decision-making. Currently, we are working towards this dashboard [1], with prototypes of these systems for both the scientist as well as decision makers.

Acknowledgment

We would like to acknowledge the PSE&G Institute for Sustainability Science for funding this research and Drs. Jennifer Bragger, Paul A.X. Bologna, Jennifer Krummins and Kirsten Monsen for their feedback and participation.

References

[1] Fails, J.A., et.al. "GeoTagger: A Collaborative and Participatory Environmental Inquiry System", the 17th Annual ACM Computer Supported Cooperative Work and Social Computing Conference, Baltimore, MD, February 15th-19th, 2014.

[2] Levin, S.A., and Clark, W.C., Eds. 2010. Toward a Science of Sustainability: Report from Toward a Science of Sustainability Conference, Airlie Center, Warrenton, Virginia, Dec 2009. CID Working Paper No. 196, Center for Int. Dev.: Harvard University.

[3] National Science Advisory Board, 2003. "ComplexEnvironmental Systems Synthesis for Earth, Life, and Society in the 21st Century: A 10-Year Outlook for the National Science Foundation", http://www.nsf.gov/geo/ere/ereweb/ac-ere/acere_synthesis_rpt_full.pdf.

[4] Walters, C., 1997. Challenges in adaptive management of riparian and coastal ecosystems. Cons. Ecol 1(2):1. http://www.consecol.org/vol1/iss2/art1/.