The United Nations and Open Data

Donald Cowan,
Paulo Alencar
and Doug Mulholland
Computer Systems Group
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
Email: dcowan/palencar/dwm@csg.uwaterloo.ca

Fred McGarry
Centre for Community Mapping
Waterloo, Ontario Canada N2L 2R5
Email: mcgarry@comap.ca

Colin Mayfield
UNU-INWEH
Hamilton, Ontario Canada L8P 0A1
Email: colinmayfield@gmail.com

Technical Report CS-2017-07

***Abstract*—The United Nations through the United Nations University International Network on Water, Environment and Health (UNU-INWEH) has embarked on a number of projects related to the environment and supported by freely available data akin to open data. These projects all use a common knowledge platform called Knowledge Integration and Management United Nations University (KIM-UNU) developed by the University of Waterloo Computer Systems Group (UWCSG) and the Centre for Community Mapping COMAP). This paper describes the platform and its use in these specific projects on international water (IW:Science), sustainable land management (KM:Land) and safe water provisioning (HydroSanitas).**

***Keywords*—United Nations, environment, knowledge management, open data**

I. Introduction

The United Nations has a wide variety of research and development projects related to the environment across the globe. They also provide information systems to encourage best practices with respect to our natural environment. These studies and information systems provide an abundance of very valuable data and other information that is published in many different forms ranging from paper documents to various electronic formats. Of course like most studies these documents, whether paper or electronic, are distributed and stored in many different locations.

Ideally all the information about all such projects, past, present and future should be collected under one "electronic roof or warehouse," so that the information is easily accessible and can be used to direct future research and provide advice to various groups around the globe. Such an organization of the research results raises many issues including:

- How can this information be made available to the widest possible community of scientists and policy makers or even the general public?
- How can the different document formats be handled?
- How can one find all the studies related to a topic or geographic area?
- How can one search all the studies about a scientific or geographic area for specific knowledge?
- How can a working group form around a topic or area with the idea of sharing and expanding knowledge?

This paper provides an overview of a set of projects developed by the Computer Systems Group at the University of Waterloo (UWCSG) in partnership with the Centre for Community Mapping (COMAP) to address these problems. The projects were initiated and supported by the United Nations University International Network on Water, Environment and Health (UNU-INWEH).

II. Functions and Tools

The questions in Section I lead to a a number of technical solutions and tools that are described next.

A. Different Document Formats

Most documents produced over the last two decades have been created in an electronic format such as portable document format (pdf), Microsoft word, plain text or Hypertext Markup Language (html). Older paper documents can be scanned and converted through optical character recognition (OCR) to a machine-readable version. It would be convenient if all such documents could be rendered into a common format for later indexing and searching as described in Section II-B.

Most document types have a conversion command or tools to convert a document in a specific format into html. Therefore one decision was to use html as a common underlying format while retaining the document in its original form. Another approach can use IFilters [1]. IFilters are a Microsoft plugin, which can take documents in many different formats and parse each word so that the document can be indexed.

B. Indexing and Searching

With documents now in a common format they can be easily indexed and searched for single words and phrases using many different indexing and searching technologies. However, there are still a number of issues to be addressed.

Documents also have associated metadata such as that defined in the Dublin Core [2]. Searches could include indexed words and phrases, and metadata.

Each document is likely to be connected with one or more locations on the planet. Possible locations could relate

to items such as the subject matter (a study of the Mekong Delta), the author(s) of the document or places mentioned in the document. Hence each document could be geo-coded in multiple ways, so that the document is associated with one or more points or polygons, where a polygon would represent an area rather than a single location. A document could then be located by phrases and by indicating a location on the globe. Of course geo-coding is another form of metadata for a document and can be added to the search. In the systems described in this paper we have chosen to associate the document with a representative single point related to the subject matter.

Since research is ongoing, there will be new information added to the warehouse over time. Hence the search must not only find documents by phrase and location, but also over time. A search today may identify some documents, but the search tomorrow may produce new results. These new results should be marked so that the user knows new information has been added to the search results.

C. Accessing the Data

The research reports and other information from the United Nations are freely available much like open data [3]. Open data is based on the concept that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. However, republishing does imply citing the original source not only to give credit but to ensure that the data has not been modified or results misrepresented.

However, in contrast to true open data the United Nations does put some general conditions on use. The United Nations needs to have knowledge about who is accessing the data and for what use. This information will allow the UN to determine the effectiveness of its research funding and possibly where to concentrate its future efforts. Thus, everyone accessing the "open data" needs to make an official request for information and be appropriately authorized.

D. Forming a Working Group

Research groups can form around document sets or areas of interest to share or expand knowledge on environmental concerns. These groups can be viewed as a social network with a specific topic of collaboration in mind. With such groups the level of expertise and knowledge required is quite high.

There must be a process in place that provides each participant with a userid and password. In exchange for this entry ticket, each contributor must provide identification information and quantify his/her background so that the entire group is aware of all members of the group and has a way to assess all contributions that are published. This type of structure is called a mediated social network as it is managed on behalf of the group and provides a measure of review and control.

E. Content Management

What type of data structure supports the storage and retrieval of content? There will be documents of varying sizes ranging from a brief statement about a best-practice in producing clean water in an area of the world to a large multi-volume report on reclaiming deserts.

The document management system stores the documents and the metadata in a relational database, where each document has a unique identifier. The documents can be stored as binary large objects (BLOb) [4]. The index is stored separately with the document identifier and the offset of the indexed word recorded for each document.

III. Projects

The functionality and data structures described in Section II have been assembled to support several projects for the United Nations through UNU-INWEH. These projects are described in this section. The description for each project has been obtained from the UNU-INWEH site [5].

A. IW:Science

The theme of the IW:Science project is "Enhancing the Use of Science in International Waters Projects to Improve Project Results." This subject was explored and led to The IW:Science Synthesis Report "Science-Policy Bridges Over Troubled Waters Making Science Deliver Greater Impacts in Shared Water Systems," which was launched at the GEF IW Science Conference in September 2012 and is available at [6].

Through project partners and volunteer scientists the IW:Science Project's purpose is to enhance through knowledge integration and information sharing the use of science in the International Waters (IW) projects of the Global Environment Facility (GEF) [7]. The project will help strengthen priority setting, knowledge sharing, and results-based, adaptive management in current and future projects. The results are:

- The first ever synopsis of the science behind the IW portfolio of projects to date according to transboundary IW system types; river basins, lakes, groundwater, the coastal zone and land-based pollution, and large marine ecosystems and the open ocean.
- A systematic analysis of the IW portfolio and integrating knowledge from the wider scientific community with regards to critical emerging science issues, application of science for adaptive management and development and use of indicators to support IW projects.
- The creation of the IW:Science knowledge management system, a fully integrated database of IW documents and a suite of learning network and communication tools. This powerful tool made the synopsis and analysis process of about 5,500 documents possible while capturing new knowledge and review insights of the scientific working groups. The IW:Science KMS is powered by the KIM-UNU platform technologies described in Section III-D.

B. KM:Land

A major barrier to Sustainable Land Management (SLM) in drylands is the lack of knowledge management for mitigating land degradation. Knowledge Management includes information management and dissemination, as well as knowledge enhancement and capacity building for adaptive management.

The GEF-funded KM:Land project [8] aims to lay the foundations for a comprehensive system to track progress across the GEF Land Degradation Focal Area. This project is executed by UNU-INWEH in collaboration with the implementing agency UNDP using KIM-UNU described in Section III-D.

KM:Land is the first of a three-phase project, initiated in 2007, to support the enhancement of knowledge management principles in the GEF Land Degradation Focal Area.

KM:Land was designed to address the knowledge management gaps in the GEF LD Focal Area by providing the scientific-technical basis for selecting indicators to demonstrate the benefits, impacts and best practices of SLM projects in the GEF portfolio and through the creation of an online Learning Network.

The objectives of the KM:Land project are:

- To develop global and project-level indicators that demonstrate global environmental and livelihood-related impacts derived from actions on combating land degradation.
- To develop a framework for knowledge management and capacity building for SLM through the development of a KM:Land Learning Network .
- To define a process for establishing a monitoring and evaluation system that supports results-based management for SLM projects.

The Knowledge from Land (KM:Land) Learning Network aims to bring together Sustainable Land Management (SLM) experts and practitioners to share experiences and connect with colleagues worldwide on land degradation issues. At its core, it contains a well-established knowledge database which provides users with a common repository in which to see and access the wealth of scientific information that exists on SLM and land degradation issues. Users are encouraged to enhance this knowledge database continually through their own contributions. The database is searchable to facilitate the identification, extraction and synthesis of scientific knowledge.

Knowledge from Land also provides online interactive tools to facilitate discussions, and the sharing, creation and collaboration of work by individuals and groups. Users can participate in groups and discussion forums and can collaborate on wiki documents in the Communities of Practice.

The Knowledge from Land Learning Network is open to the public. However, a contributor to the database or a participant in the Communities of Practice must register as a user of the Learning Network.

C. HydroSanitas

An interactive knowledge portal of safe water provisioning for rural, remote and otherwise marginalized communities around the world. HydroSanitas [9] is powered by KIM-UNU. One of the most critical challenges facing the world today is ensuring an adequate supply of high quality drinking water despite rapid population growth and environmental change. This problem is especially acute in remote communities that do not have the financial, human and infrastructure resources to deal with the problems of providing good quality drinking water. Web-based information and communications technologies have the potential to foster new ways of understanding the problems of water supply while meeting these challenges with appropriate treatment solutions.

Information will be consolidated in a knowledge portal that will document proven approaches to protecting and treating drinking/waste water that are appropriate to the type of water resource, the physical environment and the social, economic and cultural contexts of the users.

The outcomes of the HydroSanitas Project are:

- A dynamic and interactive novel information portal where stakeholders can interact and share ideas.
- Opportunities to showcase technologies and methods for protecting and treating drinking water at demonstration sites.
- Practical knowledge to direct next-generation research and development and support local decision-making.

The innovations include:

- Building upon novel mobile and web-based communications to provide a practical and desperately needed example of how to leap-frog traditional information dissemination approaches and harness social networking for applied decision making, gap analysis and market assessment within the water sector.
- Developing tacit knowledge capture and multi-format information search capacity that can be applied to many different resource sectors.
- Developing a co-ordination mechanism for a fragmented resource sector.
- Providing social context for appropriate and sustainable drinking water solutions, especially in rural, remote and developing communities.
- Developing new IT methods to extract information from social media, explicitly recognizing the importance of social media not only as a communication tool, but as a data collection tool.
- Building upon existing water treatment technologies that have been demonstrated to be suitable for these communities and developing technologies to enhance water quality.

The key deliverables are:

- Decision support for appropriate, sustainable solutions.
- Enhanced knowledge portal that will include next generation dynamic and participatory synthesis and a knowledge sharing platform.
- New knowledge-mining techniques for appropriate water provisioning and protection, best practices and related resources.
- Enhanced participatory approaches and tools.
- Access to advanced methods and tools.

- Advanced strategies for analyzing the vulnerability and risk of water systems.
- This tool is designed to overcome the fragmentation that exists in the water sector in order to:
    1) transfer knowledge to communities for enhanced decisionmaking around providing their members with safe water;
    2) provide technology developers, researchers, consultants and practitioners with grass root knowledge of, and social and cultural context for, the needs, gaps and opportunities which exist in providing safe drinking water for rural, remote and otherwise marginalised communities around the world.

D. KIM-UNU

KIM-UNU [10], Knowledge Integration and Management United Nations University is the name of the technology platform developed for UNU-INWEH in partnership with the University of Waterloo Computer Systems Group and the Center for Community Mapping as a powerful and user-friendly Knowledge Management System (KMS). The platform has been developed in recognition that often there is a wealth of knowledge and information, representing a significant investment of development funds and project implementation which is not readily accessible, let alone capable of integration.

The content is easily located via a search interface. There is also a suite of communication and interaction tools allowing users to view, comment and access information as a group in either real-time or as a shared collaborative workspace. Integration of this learning network with the relational database of information provides the tools for the capture and storage of information input by users. This results in the continuous improvement of document classification, resulting in better refined searches for the next user, as well as the capture of newly created and/or tacit knowledge.

The platform has three integrated components:

- A relational database for document, map, photo, video, and spatial information upload combined with a powerful database management system to locate and extract the information;
- A suite of learning and communication tools to both facilitate live interaction between users and capture and save imparted tacit knowledge back into the system; and
- A user-friendly web portal that requires no programming or special IT skills to use, but more importantly, to run and manage.

Created as part of the GEF-funded IW:Science and KM:Land projects, the KIM-UNU platform saw the creation of the largest-ever compilation of GEF International Waters project documents. GEFs International Waters focal area has seen 20 years of work on 180 projects for international river, lake, groundwater, coastal and marine water systems spanning 172 countries and over US$ 6 billion of investment. The KIM-UNU platform allows working groups consisting of 75 scientists all over the world to mine the documents for science information across the entire portfolio at the click of a button, comment on documents and provide answers to science-specific questions that are fed back into the system.

The Knowledge Management System (KMS) based on KIM-UNU and developed for IW:Science, KM:Land and HydroSanitas has been noted as a useful and versatile KM and learning system for other multi-country or regional initiatives. It has already been re-applied by the GEF Scientific and Technical Advisory Panel to investigate GEF intervention in global hypoxia hot spots for large marine ecosystem areas globally [11]. In addition, various stages of trial systems and discussions for the replication of the KMS approach and software for other bodies of information are underway including the Lake Victoria Basin Commission (KMS implementation between Kenya, Uganda, Tanzania, Rwanda and Burundi).

The KIM-UNU platform is developed using Web Informatics Development Environment (WIDE) toolkit and technologies [12] created by UWCSG. The systems do not require programming or IT management skills to run and manage once they are developed. This makes for significantly reduced running costs and a platform to which any access-granted user can contribute.

IV. Conclusions

The United Nations through its many agencies has collected a large amount of “open data” over the decades. Of course this process has not stopped and in fact it is accelerating, as we attempt to understand more of the surrounding environment and how it can be sustained. This paper describes some specific applications to environmental problems around the world relating to international waters, sustainable land management and safe water provisioning. As well as presenting these three issues, the paper describes the underlying information technology that is used to manage the data and derive value from it in addressing serious global environmental problems.

References

[1] “IFilter,” http://en.wikipedia.org/wiki/IFilter, 2014.

[2] “Dublin Core MetaData Set,” http://en.wikipedia.org/wiki/Dublin_Core, 2104.

[3] D. Cowan, P. Alencar, and F. McGarry, “Perspectives on Open Data: Issues and Opportunities,” David R. Cheriton School of Computer Science, University of Waterloo, Tech. Rep. CS-2014-01, 2014.

[4] “Binary Large Object,” http://en.wikipedia.org/wiki/Binary_large_object, 2014.

[5] “United Nations University International Network on Water, Environment and Health (UNU-INWEH),” http://inweh.unu.edu/, 2014.

[6] “IW:Science,” http://inweh.unu.edu/iw-science/, 2014.

[7] “Global Environment Facility,” http://web.undp.org/gef/, 2104.

[8] “KM:Land,” http://inweh.unu.edu/km-land/, 2014.

[9] “HydroSanitas,” http://inweh.unu.edu/hydrosanitas/, 2014.

[10] “Knowledge and Integration Management Platform,” http://inweh.unu.edu/kim-unu/, 2014.

[11] “Hypoxia and Nutrient Reduction in the Coastal Zone - Advice for Prevention, Remediation and Research,” http://inweh.unu.edu/wp-content/uploads/2013/05/Hypoxia-and-Nutrient-Reduction-in-the-Coastal-Zone.pdf, 2011.

[12] D. Cowan, P. Alencar, F. McGarry, and C. Lucena, “A Web-based Framework for Collaborative Innovation,” David R. Cheriton School of Computer Science, University of Waterloo, Tech. Rep. CS-2012-02, 2012.