International Conference on Hydroinformatics

8-1-2014

Linked Water Data For Water Information Management

Edward Curry

Viktoriya Degeler

Eoghan Clifford

Daniel Coakley

Andrea Costa

See next page for additional authors

Authors
Edward Curry, Viktoriya Degeler, Eoghan Clifford, Daniel Coakley, Andrea Costa, Schalk-Jan Van Andel, Nick van de Giesen, and Christos Kouroupetroglou

LINKED WATER DATA FOR WATER INFORMATION MANAGEMENT

EDWARD CURRY (1), VIKTORIYA DEGELER (1), EOGHAN CLIFFORD (1) DANIEL COAKLEY (1), ANDREA COSTA (2), SCHALK-JAN VAN ANDEL (3), NICK VAN DE GIESEN (4), CHRISTOS KOUROUPETROGLOU (5), THOMAS MESSERVEY (2), JAN MINK (6), SANDER SMIT (7)

(1): National University of Ireland, Galway, Ireland.
(2): R2M Solution, Italy.
(3): UNESCO-IHE, Netherlands.
(4): TU Delft, Netherlands.
(5): Ultra 4, Greece.
(6): VTEC Engineering, Netherlands.
(7): BM Change, Netherlands.

The management of water consumption is hindered by low general awareness and absence of precise historical and contextual information. Effective and efficiency management of water resources requires a holistic approach considering all the stages of water usage. A decision support tool for water management services requires access to a number of different data domains and different data providers. The design of next-generation water information management systems poses significant technical challenges in terms of information management, integration of heterogeneous data, and real-time processing of dynamic data. Linked Data is a set of web technologies that enables integration of different data sources. This work investigates the usage of Linked Data technologies in the Water Management domain, describes the fundamental concepts of the approach, details an architecture, and discusses possible water management applications.

INTRODUCTION

According to the United Nations, approximately one-third of the world's population currently lives in water stressed regions [1]. Projections by the Organization for Economic Cooperation and Development (OECD) estimate that 47% of the world's population will likely be living in areas of high water stress by 2030 [2]. The growth in demand for electricity production, agricultural, and industrial uses are depleting the world's freshwater supply in both quantity and quality. A key factor is that water has not been adequately considered as a vital resource that needs to be managed. The result is that water infrastructure, business models, and behaviours at all levels of the water value chain reflect this fact. However, there is the significant opportunity to accelerate the development and implementation of ICT-based water awareness, management and conservation solutions.

In order to manage water holistically it is important to use water management decision support tools that present meaningful and contextual information about usage, price, and availability of water in an intuitive and interactive way to users. Different users will have different information requirements to manage their water, from home users managing their personal water usage, business users managing the water consumption of their commercial activities, to municipalities managing regional distribution and consumption at city level. In order to develop water information services for these diverse users it is necessary to leverage knowledge from across a number of different domains, including metering (household, neighbourhood, etc.), collection and catchment management, environmental, water quality, energy usage, utility information, end user feedback, occupancy patterns, meteorological data, etc. However, many barriers exist to their interoperability and there is little interaction between these islands of information. The design of next-generation water information management systems poses significant technical challenges in terms of information management, integration of heterogeneous data, and real-time processing of dynamic data.

Linked data technology leverages the existing open protocols and W3C standards of the Web architecture for sharing structured data on the web. Building on the successful application of linked data to solve similar problems within the Building Management [3] and the Energy Management [4] domains we propose its application in the Water domain. In this paper, we propose the use of linked data as an enabling technology for cloud-based water data services. The objective of linking water data is to create an integrated well-connected graph of relevant information for managing water. Representing water usage data within the linked data format makes it open allowing it to be easily combined with linked data from other relevant domain silos. This paper describes the fundamentals of this approach, details the main components of the initial envisioned architecture, and details the water management applications possible as a result of its implementation.

MOTIVATION

Sustainability requires information on the use, flows and destinies of energy, water, and materials including waste, along with monetary information on environment-related costs, earnings, and savings. [6] This type of information is critical if we are to understand the causal relationships between the various actions that can be taken, and their impact on sustainable performance. However, the problem is broad in scope, and the necessary information may not be available, or difficult to collect. Within the context of water management improving the sustainability of water consumption, especially through changing the way an household, organization, or city operates [5, 18], requires a number of practical steps that will include the need for a systematic approach for information-gathering and analysis.

Contextual Water Management

One of the key problems of modern water management is the lack of water information, management and decision support tools that present meaningful and personalized information about usage, price, and availability of water in an intuitive and interactive way to end users. This introduces limitations in the efforts to manage water as a resource, including:

- **User Awareness**: End users do not have access to water information (i.e. availability, consumption, pricing) at the moment water consumption decisions are being taken.
- **User Incentivisation**: Due to billing, pricing, awareness or metering aspects, end users may not have an incentive to make behavioural change.
- **Integrated Information Provision & Analysis**: Decision makers do not have access to information platforms to make organizational change. Personalized water information can

only be created by combining publically available water information with private water usage information only available to water service providers.

- **Benchmarking**: End-users do not know if their individual water consumption pattern is high or low compared to similar users.

Water Footprint's & Water Information Ecosystems

To successfully manage water data, an entity (a household, a company, a municipality, etc.) must consider all sources of water consumption, including indirect ones, augmented with water network distribution information. The demand for business transparency is driving multinational companies towards more holistic assessments of their water footprint and associated impact. Understanding all the freshwater sources and uses related to a business or product, decision-makers can identify environmentally conscious, programmatic changes to reduce their freshwater impact or footprint. Water footprint assessments are emerging concepts that require obtaining water data from many participants within an organisation's supply chain. Numerous data sources can be used for this purpose, including weather data, geo-location data, historical records, product usage data, user behaviour habits, etc. There is no single source to provide such data and a considerable number of different data sources must be integrated to collect the information necessary for an accuracy water footprint.

LINKED WATER DATA

Information integration projects typically focus on one-off point-to-point integration solutions between two or more systems in a customized but inflexible and ultimately non-reusable manner. The fundamental concept of Linked Data is that data is created with the mind-set of sharing and reuse. Emerging from research into the Semantic Web, Linked Data leverages the existing open protocols and W3C standards of the Web architecture for sharing structured data on the web. Linked Data proposes an approach for information interoperability based on the creation of a global information space. Linked Data has the following advantages:

- Separate systems that are designed independently can be later joined/linked at the edges.
- Interoperability is added incrementally when needed and where it is cost-effective.
- Data is expressed in a mixture of vocabularies.

Linked Data is facilitating the publishing of large amounts of structured data on the web. The resulting Web of Data can be considered as a web scale dataspace supported by Semantic Web technologies. The Linked Open Data Cloud represents a large number of interlinked datasets that are being actively used by industry, government and scientific communities.

Linked Data Principles

Linked data technology uses web standards in conjunction with four basic principles for exposing, sharing and connecting data. These standard principles are:

- **Naming**: Use URIs as names for things - the use of Uniform Resource Identifier (URI) (similar to URLs) to identify things such as a person, a place, a product, an organization, an event or even concepts such as risk exposure or net profit, simplifies reuse and the integration of data.
- **Access**: Use HTTP URIs so that people can look up those names - URIs are used to retrieve data about objects using standard web protocols. For an employee this could be their organization and job classification, for an event this may be its location time and attendance, for a product this may be its specification, availability, price, etc.

- **Format**: When someone looks up a URI, provide useful information using the standards - when someone looks up (dereferences) a URI to retrieve data, they are provided with information using a standardized format. Ideally in Semantic Web standards such as RDF.
- **Contextualisation**: Including links to other URIs so that people can discover more things - retrieved data may link to other data sources, thus creating a data network e.g., data about a product may link to all the components it is made of, which may link to their supplier.

Resource Description Framework

The Resource Description Framework (RDF) is the basic machine-readable representational format used to represent information. RDF is a general method for encoding graph-based data that does not follow a predictable structure. RDF is schema-less and self-describing, meaning that the labels of the graph describe the data itself. Data and facts are specified as statements and are expressed as atomic constructs of a subject, predicate and object, also known as a triple. The statement "Main Kitchen contains a Coffee Machine" is expressed in triple format as:

Subject - "Main Kitchen"
Predicate - "contains a"
Object - "Coffee Machine"

RDF is designed for use in web-scale decentralized graph data models. For this reason the statement parts need to be identified so that they can be readily and easily reused. RDF uses URIs for identification, expressing the previous statement in RDF then becomes:

http://lab.linkeddata.deri.ie/2010/deri-rooms#r315
http://vocab.deri.ie/rooms#contains
http://water.deri.ie#mr-coffee

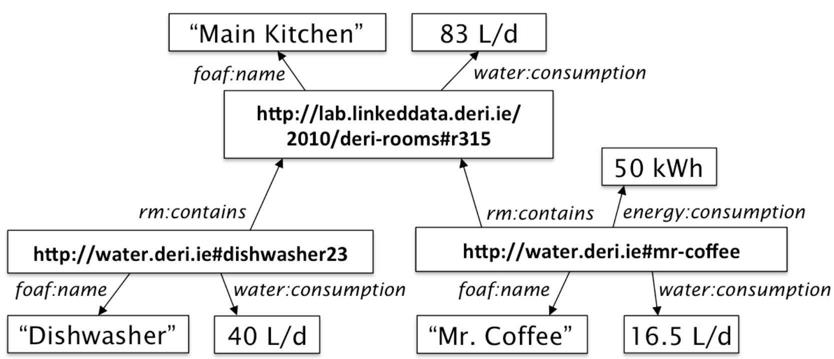

Figure 1. Example of Linked Water Data for an Office Kitchen

Universal Resource Identifiers that describe the data can be uniformly used across the system, even if they come from different sources. The graph structure of the linked data, as illustrated in Figure 1, easily supports optional parameters, and the evolution of parts of the data structure does not affect any other related data. The relations are described on a low-level, therefore allowing combining (linking) pieces of data together based on their relation types, and not on their provider or representation.

LINKED WATER DATA IN ACTION: WATERNOMICS

The goal of the WATERNOMICS project is to provide personalised and actionable information about water consumption and water availability to individual households, companies and cities in an intuitive and effective manner at a time-scale relevant for decision making. Access to this information will increase end-user awareness and improve the quality of the decisions from decision makers regarding water management and water government. WATERNOMICS will accomplish this by combining water usage related information from various sources and domains to offer water information services to end-users. The platform will enable sharing of water information services across communities of users by providing a convergence layer on top of existing water infrastructures with minimal disruption. The objective is to expose the data within existing systems, but only link the data when it needs to be shared. Representing water usage data within the linked data format makes it open, allowing it to be easily combined with linked data from other relevant domain silos.

Architecture

The main components of the initial envisioned architecture, as illustrated in Figure 2, are the sources of water usage metering on existing systems, the Linked water dataspace consisting of a linked data cloud & support services, and the resulting water management applications.

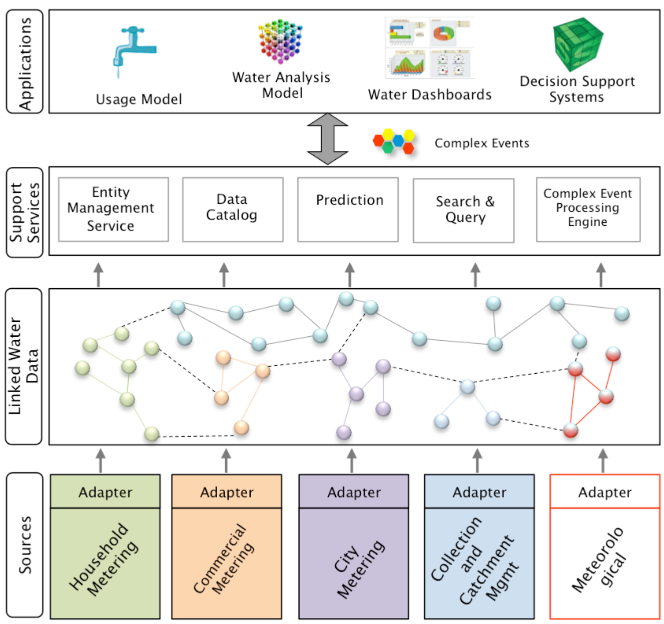

Figure 2. WATERNOMICS Platform

- **Water metering and data sources**: At the bottom of the architecture are the existing operational legacy information systems. Adapters will perform the "RDFization" process, which transforms multiple formats and legacy data and lifts it to the dataspace.
- **Linked Water Data**: The Linked Dataspace links at the information-level (data) not the infrastructure-level (system) by focusing more on the conceptual similarities (shared understanding) between information. The resulting Linked Water Cloud is rich with knowledge and semantics about water usage performance indicators and forms the basis for real-time water usage analytics and other applications with the help of support services

- **Support Services**: Dataspace support services are designed to simplify the consumption of the linked data cloud by encapsulating common services for reuse (e.g. cataloguing, search and query, prediction etc.).
- **Water Usage and Management Applications:** At the top of the architecture are the water usage and management applications that consume the resulting data and events from the linked water data.

Applications

The applications that may be built using linked water data are diverse; they include water awareness dashboards, decision support for the different targeted users (i.e. domestic users, organisations, cities), and water availability/forecasting, dynamic pricing, and water footprints.

Water Awareness Dashboards: Low user comprehension of water flows and usage is one of the biggest causes of water wastage. A lack of awareness on the amount of consumed water leads to the lack of incentives to monitor and affect the situation. Water awareness requires different information for household, company, and city level, and where different decisions are taken to manage water on these levels. Therefore water awareness dashboards need to be tailored for different needs of different water usage levels. The data collected by smart water meters is enriched with contextual linked data and processed in real-time, therefore allowing for deeper data analysis and faster reactions.

Water consumption forecasting: Hydro-meteorological forecasts predict natural demand and supply of water and can be used to prepare and adjust water supply. Forecasting systems can achieve different goals depending on the level of the system deployment. At the household level forecasts include analysis of occupants' behaviour and water consumption based on similar historical water usage. These forecasts can be incorporated to dashboards and be used as water saving goal drivers, forecasting models can leverage linked open data at the neighbourhood or city level. At the company level forecasts similar to those of the household level are also augmented by models or simulations of the water needs of subsystems within the organisation. Linked open data can be used to perform benchmarking between similar organisations to identify areas of potential water optimisation. Even more benefits can be gained on the city level.

Dynamic Pricing: Currently many municipalities use approximate calculations of water costs to charge residents for the yearly water consumption. The usage of historical water consumption data available to the municipality may be augmented by the open weather data repository and geo-location parameters repository to construct a more accurate model of consumption demands and supply. Depending on these external factors, and on the data obtained from the water distribution network repository, the price of water may vary in time and per location, based on how easy it is to obtain. This can enable water-pricing schemes that correspond closer to real costs of consumption and local conditions.

Water Footprint: Understanding the impacts of a product or service requires an analysis of all potential water consumption associated with a product, process or service for its entire life cycle. For example, a water footprint of a product would provide a quantitative cradle-to-grave analysis of the product/services global water costs (i.e., water used in raw materials extraction, through materials processing, manufacture, distribution, use, repair and maintenance, electricity generation, and disposal or recycling). Building a water footprint requires the gathering of water data from many participants within the supply chain. Linked Open Water Data can be a key enabler for the development of a global information ecosystem of water footprint inventory data on for products, services, and organisations.

RELATED WORK

As recent surveys show, a number of policies and standards for smart metering have been adopted in different countries, but most standards still contain a fragmented set of solutions [7] with little support for adding contextual data. Most policies and standards appear in the smart grid area, and are adopted by other areas [8]. Hydro-meteorological information is mainly described by drought indicators [9], notably Standardized Precipitation Index (SPI) [10] and Temperature Condition Index (TCI) [11]. Mostly these indices describe the present state of the system [12].

It has been shown that water consumption awareness and the strength of motivation greatly affect the potential for water saving. For example, in [13] the deployment of the experimental system that provided detailed water usage information in the shower showed the resulting decrease in water consumption. It also showed the division of users into two groups: those who continued to pursue conscious water behaviour even after the experiment was over, and those who returned to previous water habits after the removal of informational displays. An overview of pro-environmental behaviour models and key human-computer interaction (HCI) techniques to promote and motivate such behaviour are presented in [14]. In [15] a display to present gas, electricity, and water consumption in artistic way is proposed. In [16] a persuasive application to promote responsible attitude towards natural resources, food, and water during family interactions is described. The comparison between lightweight ambient and numeric displays is performed in [17]. Results showed that an abstract ambient display with color-coded visualization of water usage causes bigger water saving behaviour changes comparing to a numeric display. In [18] group-based feedback is used to reduce the consumption of paper within an office environment.

All of these techniques are complementary to linked water data. The approach we propose here aims to make it easier to implement such applications by reducing the cost of gathering the necessary data to drive the applications.

SUMMARY AND FUTURE WORK

Effective and efficient management of water resources requires a holistic approach considering all the stages of water usage. Development of a holistic Water Management Platform faces significant challenges due to heterogeneous data sources, complex data integration, evolving information requirements, and partial solutions' applicability. This paper presents a concept of Linked Water Data that allows creation of a holistic Water Management Platform. The objective of linking water data is to create an integrated well-connected graph of relevant information for managing water effectively. Representing water usage data within the linked data format makes it open, thus allowing it to be easily combined with linked data from other relevant data. The concept will be implemented as a part of the WATERNOMICS FP7 European project, and it will be tested in three water management pilot sites, on household, company, and municipality levels. The approach will also be considered within the design of Smart City Infrastructures for Water Management [19].

Acknowledgments

The research leading to these results has received funding under the European Commission's Seventh Framework Programme from ICT grant agreement WATERNOMICS no. 619660. It is supported in part by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

REFERENCES

[1] Food and Agricultural Organization of the United Nations (2007) Making every drop count [Online] Available: http://www.fao.org/newsroom/en/news/2007/1000494/index.html

[2] Organization for Economic Co-operation and Development " OECD Environmental Outlook to 2030" Environment & Sustainable Development, vol.1, pp. 1 – 523, 2008

[3] Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M., & O'Riain, S. (2013). Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics*, *27*(2), 206–219.

[4] Curry, E., Hasan, S., & O'Riáin, S. (2012). Enterprise Energy Management using a Linked Dataspace for Energy Intelligence. In *The Second IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT 2012)*. Pisa, Italy: IEEE.

[5] Curry, E., Guyon, B., Sheridan, C., & Donnellan, B. (2012). Sustainable IT: Challenges, Postures, and Outcomes. *IEEE Computer*, *45*(11), 79–81. doi:10.1109/MC.2012.385

[6] Curry, E., & Donnellan, B. (2012). Sustainable Information Systems and Green Metrics. In S. Murugesan & G. R. Gangadharan (Eds.), *Harnessing Green IT: Principles and Practices* (pp. 167–198). John Wiley & Sons, Inc.

[7] Liotta, A., Geelen, D., van Kempen, G., & van Hoogstraten, F. (2012). A survey on networks for smart-metering systems. In *International Journal of Pervasive Computing and Communications*, 8(1), 23-52.

[8] Yuan Xu, F., Zhou, L., Wu, Y. L., & Ma, Y. (2010). Standards, policies and case studies in smart metering. In *Power and Energy Society General Meeting*, 2010 IEEE (pp. 1-5). IEEE

[9] Barua, S., Ng, A. W. M., & Perera, B. J. C. (2010). Comparative evaluation of drought indexes: Case study on the Yarra River catchment in Australia. In *Journal of Water Resources Planning and Management*, 137(2), 215-226.

[10] Cancelliere, A., Di Mauro, G., Bonaccorso, B., & Rossi, G. (2007). Drought forecasting using the standardized precipitation index. In *Water resources management*, 21(5), 801-819.

[11] Kogan, F. N. (1995). Application of vegetation index and brightness temperature for drought detection. In *Advances in Space Research*, 15(11), 91-100.

[12] Boken, V. K. (2009). Improving a drought early warning model for an arid region using a soil-moisture index. In *Applied Geography*, 29(3), 402-408.

[13] Kappel, K., & Grechenig, T. (2009). Show-me: water consumption at a glance to promote water conservation in the shower. In *Proceedings of the 4th international conference on Persuasive Technology* (p. 26). ACM.

[14] Froehlich, J., Findlater, L., & Landay, J. (2010). The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1999-2008). ACM.

[15] Makonin, S., Pasquier, P., & Bartram, L. (2011). Elements of consumption: an abstract visualization of household consumption. In *Smart Graphics* (pp. 194-198). Springer Berlin.

[16] Lepe Salazar, F., Yamabe, T., Alexandrova, T., Liu, Y., & Nakajima, T. (2012). Family interaction for responsible natural resource consumption. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 2105-2110). ACM.

[17] Kuznetsov, S., & Paulos, E. (2010). UpStream: motivating water conservation with low-cost water flow sensing and persuasive displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1851-1860). ACM.

[18] Hasan, S., Medland, R. C., Foth, M., & Curry, E. (2013). Curbing resource consumption using team-based feedback: paper printing in a longitudinal case study. In *8th International Conference on Persuasive Technology (PERSUASIVE 2013)* (pp. 75–86). Sydney, NSW: Springer-Verlag. doi:http://dx.doi.org/10.1007/978-3-642-37157-8_11

[19] Ojo, A., Curry, E., & Janowski, T. (2014). Designing Next Generation Smart City Initiatives - Harnessing Findings And Lessons From A Study Of Ten Smart City Programs. In *22nd European Conference on Information Systems (ECIS 2014)*.