International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Communicating River Level Data and Information to Stakeholders with Different Interests: the Participative Development of an Interactive Online Service

Kit (C.J.A.) Macleod[A], Somayajulu G. Sripada[B], Antonio A. R. Ioris[C], Koen Arts[D] and René van der Wal[D, E]

[A]*Information and Computing Sciences Group, The James Hutton Institute, UK; kit.macleod@hutton.ac.uk*
[B]*Department of Computing Science, University of Aberdeen, UK; yaji.sripada@abdn.ac.uk*
[C]*School of Geosciences, University of Aberdeen, UK;a.ioris@abdn.ac.uk*
[D]*Dot rural Digital Economy hub, University of Aberdeen, UK*
[E]*Aberdeen Centre for Environmental Sustainability, School of Biological Sciences, University of Aberdeen, UK; r.vanderwal@abdn.ac.uk*

Abstract: Increasing the effectiveness of how river level data are communicated to a range of stakeholders (individuals, groups and communities) with an interest in river level information is likely to result in greater use of data collected by regulatory agencies. A range of interest groups, like those involved in recreational pursuits such as angling and canoeing, require certain but different information on changes in river levels state to allow effective scheduling of their activities. A range of options have been developed for communicating river level data to different audiences, but those fail to address group heterogeneity and information demands. To a large extent, those problems derive from a lack of understanding of information demanded by river water users, as well as the failure to comprehend how they perceive river level change. We are working with river users who span land managers and farmers, hydropower generators, recreational users e.g. those involved in canoeing and fishing and broader local communities as well as the public authority (SEPA) responsible for hydrological monitoring and provision of this information and cyberinfrastructure in Scotland. Currently, river level data is provided to members of the public through a web site without any formal engagement with river users having taken place. In our research project called wikiRivers, we are working with the suppliers of river level data as well as the users of this data to explore and improve from the user perspective how river level data and information is made available online. We are focusing on the application of natural language generation technology to create textual summaries of river level data tailored for specific interest groups. These tailored textual summaries will be presented among other modes of information presentation (e.g. maps and visualizations) with the aim to increase communication effectiveness. Natural language generation involves developing computational models that use non-linguistic input data to produce natural language as their output. Acquiring accurate correct system knowledge for natural language generation is a key step in developing such an effective computer software system. In this paper we set out the needs for this project based on discussions with the stakeholder who supplies the river level data and current cyberinfrastructure and present a detailed stakeholder identification, engagement and cyberinfrastructure development plan.

Keywords*:* hydrology; water; natural language generation; cyberinfrastructure; Scotland

1 INTRODUCTION

Pronounced fluctuation of water levels in response to rainfall and snowmelt is a natural property of many UK rivers and is in sharp public and professional focus due to its wide-ranging impact on both local economies and people's lives (Posthumus et al., 2009). There is a need for hydrological research to provide cross disciplinary integration in support of changing societal needs (Wagener et al., 2010). This requires research that spans the natural (e.g. hydrology), social (e.g. sociology) and computing sciences along with an analytical-deliberative approach framed by stakeholder requirements (Macleod et al., 2007). Increasing efforts to develop and subsequently use community-based geospatial cyberinfrastructure is improving people's access to data and information (De Longueville, 2010). Geospatial cyberinfrastructure "refers to the infrastructure that supports the collection, management and utilisation of geospatial data, information and knowledge for multiple science domains" (Yang et al., 2010). A recent review of the future needs for more effective geospatial cyberinfrastructure stressed the needs for: 1) understanding social heterogeneity in relation to problem/situation awareness; 2) improving how we can transform data collected to the information required; 3) greater use of semantic approaches for more effective knowledge exchange; 4) iterative development of geospatial cyberinfrastructure through engagement with stakeholders; 5) involvement of social and domain scientists to advance citizen science approaches; 6) use of cloud computing to geospatial cyberinfrastructure to provide open and transparent platforms; and 7) a geospatial cyberinfrastructure research agenda that includes a wider set of stakeholders including those from academia, governments and their agencies, NGOs and the wider public (Yang et al., 2010). There is a need to identify stakeholders through stakeholder analysis that includes key steps of: stakeholder identification; differentiating and categorising stakeholders; and investigating relationships between stakeholders (Reed et al., 2009). Where we follow Freeman's (1984) broad definition of stakeholders as "any group or individual who can affect or is affected by the achievement of the organization's objectives".

In this paper we are setting out how communication of complicated river level data sets will only be effective if the great variety of ways different users and stakeholders may use, understand and interpret river level data is taken into account. We are testing with stakeholders the development and use of graphics and text through natural language generation developed with stakeholders to provide more effective communication and use of river level data in Scotland. Here we describe how we have started and plan to continue to work with both users and suppliers of river level data to improve the existing online presentation.

1.1 River levels information: moving from users to stakeholders

Increasing the effectiveness of how river level data are communicated to a range of social groups is likely to result in greater use of data collected by regulatory agencies to support user needs and prevent water conflicts. Several different groups of water users, like those involved in recreational pursuits such as angling and canoeing, require information on changes in river water level to allow effective scheduling of their activities. Good hydrological information can also inform and enhance water regulation compliance. Likewise, accurate river water level information during spring and summer supports farmers' decision-making with regard to land treatments such as irrigation and pesticide application. Low-frequency events such as major floods can have devastating effects, leading to those people living in areas that are prone to flooding wanting to have access to reliable and real-time information on water levels notably when their rivers are in spate. Currently, in Scotland and elsewhere, river level data is provided to river users with little regard for the needs of these interest groups in terms of what information is presented and how this material may be understood. Existing water level monitoring takes places in rivers across the UK and these are collated by the authorities responsible for the monitoring and protection of waters (Scottish Environmental Protection Agency, SEPA[1], in Scotland; Environment Agency (EA)[2] in England and Wales). Both SEPA and the EA make a selection of this data available online[3], and although of considerable use, there are issues to do with access to and interpretation of this data. Given the diversity of interests in river water fluctuations, it is no surprise that the web pages maintained by SEPA for the dissemination of river level data receives the most visitors of all their online information services (up to 250,000 visits per month). This may be remarkable and indicative of a great need to improve the provision of this data, as the information provided is neither effectively summarised nor contextualised, thus failing to address the various interest groups which may all have somewhat different needs.

Various attempts have been made around the world to develop ways of communicating river level data to a wide range of audiences, but on the whole these fail to address group heterogeneity and particular information demands. To a large extent, those problems derive from a lack of understanding of information demanded by water users, as well as the failure to comprehend how river level change and data portrayal are perceived by people out with agencies tasked with the provision of such information. Doron *et al*. (2011) recently highlighted that though there are abundant water-related information sources, these are not accessible to users who do not posses specialist knowledge. They concluded that there is a need for improved access to environmental information sources based on user requirements. It is important to take into account that water users are creative and dynamic stakeholders, with multiple demands and ability to raise additional, unexpected data requirements. As a result, communication of hydrological data needs to be conceptualised not only as an element of the institutional responsibility of regulatory agencies, but also as part of a constant dialogue with river level data users. In practice, it means that online data portrayal should be constantly revised and improved according to stakeholder needs and feedback.

1.2 Natural language generation

Natural language generation is an area of research focused on developing computational techniques for automatically generating natural language (such as English) descriptions of non-linguistic information, that if done manually would take longer. An example is the automated production of weather forecast reports for winter road maintenance by the RoadSafe system (Reiter and Dale, 2000)[4] leading to messages such as "Road surface temperatures will fall below zero on all routes during the late evening until around midnight" based on numerical data from meteorological sensors. One specific advantage of natural language generation is that the same data may be used to generate different texts, i.e. text can be tailored to a specific audience.

In the wikiRivers project, we are working with those individuals, groups, organisations and communities who have an interest in using the online river level data provided by SEPA. We aim to improve on river user access, use and understanding of online river level data. We are particularly focusing on the

[1] http://www.sepa.org.uk/
[2] http://www.environment-agency.gov.uk/
[3] SEPA river level data - http://www.sepa.org.uk/water/river_levels/river_level_data.aspx and EA river and sea levels - http://www.environment-agency.gov.uk/homeandleisure/floods/riverlevels/default.aspx
[4] http://www.csd.abdn.ac.uk/~rturner/RoadSafe/

application of natural language generation technology to create textual summaries of river level data tailored for specific interest groups i.e. we can tailor the grammar based on their needs. In this context we consider tailored textual summaries of data embedded among other modes of information presentation (e.g. graphics including spatial information) with the aim to increase communication effectiveness. Acquiring domain knowledge, such as hydrological functioning of river systems and non-science stakeholder understanding of optimal river conditions for a wide range of river-related interests is fundamental to the development of an effective natural language generation based computer decision support systems.

Currently agencies such as SEPA use computer graphics to automatically generate graphs that visualize the underlying river level data. Since a natural language generation system works from the same river level data, the abstractions computed by the system (which are eventually expressed as words and phrases in the textual summaries) can be added as annotations to visualizations improving the user's ability to understand and information presented. Here, the text presents the summary and the visualizations present the details.

For example river level data is currently used by both canoeists and anglers for particular stretches of rivers in multiple locations in Scotland. These user groups access data on the current state and recent changes in river level to help them plan their activities. These groups (and within these groups) have different perceptual models of what particular river levels at a set location mean for river conditions for their chosen activity. These perceptual models differ between river users and for different geographical locations. In general anglers prefer river levels that are stable or falling in part, since the clarity of water is likely to be better. Whereas canoeists (depending on their ability and preferences) in general prefer river levels that are increasing. These general relationships are being used to convert the numeric river level data to more meaningful river level information using natural language generation.

2 METHODOLOGY

2.1 Stakeholder identification, engagement and cyberinfrastructure development plan

In line with the needs for greater involvement of stakeholders in terms of those supplying and using the information provided through cyberinfrastructure (Yang et al., 2010), we are developing, implementing and testing a broad and adaptive approach to stakeholder engagement (Figure 1).

2.1.1 Stage 1: Suppliers of river level data and existing cyberinfrastructure

Our starting point was the need to engage with the key stakeholder involved in the provision of river level data and existing cyberinfrastructure at the policy and operational levels in SEPA. Using telephone and face to face discussions with SEPA staff we aimed: 1) to learn about the policy and operational drivers and rationale for the historical development of the existing cyberinfrastructure; 2) to share our initial research ideas and allow these to develop in line with their aspirations and needs as suppliers of river level data; 3) to capture their rich views on the potential, as well as limiting factors for improving the communication of river level data to multiple interest groups; and 4) to learn from SEPA who are the main users of their river level data and how these may vary in time and space e.g. during a flooding event.

2.1.2 Stage 2: River level data user/stakeholder analysis

We are carrying out a stakeholder analysis that is systematic, yet critical and sensitive to stakeholder positions and needs based on active participation of the stakeholders themselves (Reed et al., 2009). We are currently developing a river user engagement strategy that will enable contact with individuals, interest groups and communities who are current users of river level data and have a desire to develop the cyberinfrastructure based on identified individual, group and community interests. Web analytics data that contains information at a monthly resolution on each of the 230 SEPA river level sites will be analysed along with aggregated data at a daily resolution to better understand what river level data is being used where and when. We plan to identify stakeholders (individuals, groups and communities) initially using snowball sampling interviews based on initial discussions with SEPA. The results of the snowball sampling will then be used to generate focus groups. These focus groups will be used to understand stakeholder knowledge of river level conditions in relation to their interests and current and future desires regarding river level information. We anticipate using social network analysis to understand the linkages between these stakeholders and their knowledge of river level processes and data (Wexler, 2001).

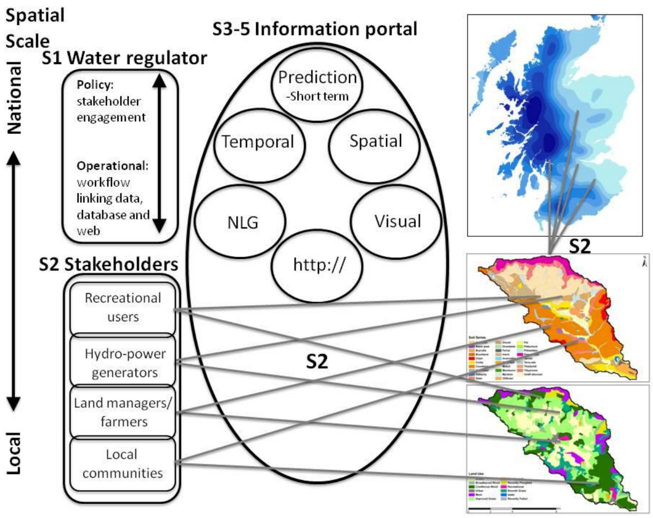

Figure 1. Stages in the wikiRivers stakeholder identification, engagement and cyberinfrastructure development. S1- interviews with collectors and suppliers of river level data. S2- river level data stakeholder analysis, including analysis of their interests in individual river networks in Scotland and what they require from the cyberinfrastructure. S3-5 Iterative development and testing of cyberinfrastructure and modelling of river level data with domain and stakeholder knowledge.

2.1.3 Stage 3: Developing cyberinfrastructure to enhance communication and use of river level information

Based on feedback from the snowball sampling interviews and focus groups we will construct project web pages and associated cyberinfrastructure that will support continued engagement about transforming river level data into information that meets the stakeholder's needs. Through discussions with stakeholders who span the supply and use of river level data in Scotland, we will select four to six case study areas where there are single or multiple stakeholder interests in river level information.

In order to further assess stakeholder interests regarding river level information, questionnaires will be applied across the case study areas supported by the project web pages. The results of these questionnaires will be complemented by semi-structured interviews, following ethical requirements and preserving anonymity. Interviews will be transcribed, coded and analysed using NVivo software. Results will be contrasted and compared within stakeholder interest groups, and across groups and case study locations.

2.1.4 Stage 4: Modelling river level data with domain and stakeholder knowledge
In the second year of the project, the data, information and knowledge collected through the focus groups, questionnaires and interviews will be organised and consolidated into conceptual models of river use and interest-specific requirements for river level information. These conceptual models will assist the development and enhancement of natural language generated texts portraying river level data. These, combined with graphical summaries, will be presented and discussed with the focus groups. We will use computational techniques for automatically generating multi-modal presentations of river water levels. We will work closely with expert and non-expert stakeholders to acquire the knowledge required for designing effective presentation schemes that combine textual descriptions with graphs and maps. Although analysis of spatio-temporal data in the context of natural language generation has been studied before, analysis of water level data over multiple river networks to a range of stakeholder interests is new. In addition, designing the presentation of information involving text, graphs and maps is a challenging task with the potential to impact on a wide range of other information presentation applications.

Based on stakeholder needs we are exploring the potential of established hydrological (Kundzewicz and Robson, 2004) and wider time series approaches for trend and change detection to provide short-term predictions of river state. We are currently testing the use and functionality of several approaches including the recently available WEKA time series API to provide this functionality http://weka.sourceforge.net/doc.packages/timeseriesForecasting/index.html?overview-summary.html

2.1.5. Stage 5: Testing of cyberinfrastructure with stakeholders
A key stage of the wikiRivers project will be the testing, further refinement and evaluation of the benefits brought by the cyberinfrastructure that we develop from the perspectives of both suppliers and users of river level data. This evaluation will in part include testing with the focus groups and a longitudinal survey that spans the project's duration.

3 RESULTS AND DISCUSSION

3.1 Stage 1: suppliers of river level data and existing cyberinfrastructure

The results of interviews to date highlight the following dominant issues from the stakeholder's perspective that have influenced the historical provision of river level data to the public:

3.1.1 Technical
River level sensor technology and the development of automated methods of data retrieval from monitoring stations have changed significantly over the past 20 years. Through our interviews with SEPA we have learned that they started to provide online river level data in 2000, were manually generated JPEG images were dumped into basic web pages. Then the SEPA web team took a more active interest and around 2003, images were then automatically generated for about 90 gauging stations that were chosen as being representative of different geographical areas. The most recent changes (in 2011), involved river level data from 230 sites being collected on a daily (and increasingly at 15 minute resolution using GPRS technology) basis. Commercial hydrological data acquisition (SODA) and storage (WISKI) software is used, along with daily queries that: 1) extract monitoring station meta-data that enables web pages to be automatically built; and 2) extract hydrological time series data that is then graphically presented. Only a small fraction of SEPA's data is made available e.g. they have river flow and additional rainfall data for the majority of these sites. They are considering moving

to a more reliable approach based on web services, but have little experience of implementing these.

3.1.2 Institutional

Through interviews we have discovered that the starting point for portrayal of water level data was a sense that the results from major financial investment in data collection needed to be visible to the public to ensure continued public funding. There is a strong desire at senior and operational levels in SEPA that greater access to the river level data is provided to water users, but there are internal issues of how and what is presented. Despite the wide use of the river level data web pages, operational staff struggle to secure time and resources to enable further development of these pages. There is not a proactive organisational attitude to data sharing.

Figure 2. An example of how online river level data is currently presented. A- river level data, B- current river state in relation to long term annual statistics.

3.1.3 River level data user

The data portrayal so far is characterised by the absence of any public engagement. So far, the only way users have been able to influence what has been portrayed is by directly contacting the SEPA web help desk. Depending on the query, these have been passed on to appropriate staff members to address. This approach has only provided random impressions of what river level users like and dislike about how the existing river level data is presented. River level data users have struggled to understand how the data is presented. Figure 2 shows an example of the current online presentation of river level data. Only the main panel (A) that graphically portrays two days of river level data was initially provided to the public. Feedback to the helpdesk highlighted that these stand alone river level data could not be interpreted by users of the information in a meaningful manner. This led to the recent addition of the sub panel (B) that plots the current river level state in the context of the long term flow duration statistics (median values). In our interview with operational SEPA staff they highlighted that users find this of limited use, since they do not understand the basis of these boundaries e.g. between high and normal, since the use of long term annual statistics does not take account of key seasonal thresholds in river level state that are of value to interest groups.

A first attempt at identification of users of SEPA's river level data has been made based on our discussions with SEPA. We have provisionally identified the main user interests as those involved in land management and farming, hydropower generation, a variety of recreational uses (primarily canoeing and fishing) and broader local communities. These will form the basis of our snowball sampling in stage 2 of our stakeholder identification, engagement and cyberinfrastructure development plan along with our analysis of the web analytics data.

4 CONCLUSIONS

In this paper we have set out the need for greater levels of engagement with users of river level data, our multistage stakeholder identification and engagement plan and initial results from multiple telephone and face to face meetings with stakeholders who are the principle suppliers of the existing river level data and associated cyberinfrastructure in Scotland. What we have learned is that there are technical, institutional and user issues and challenges to enhancing the communication to and use of river level data by various stakeholders. One major omission to date has been the lack of any formal engagement by the river data suppliers with the current users of the river level data, despite a long history of developing different means of data display, based primarily on technical advances. Fruitful application of cyberinfrastructure based on the use of natural language generation linked to visual summaries of data, is critically dependent on understanding the information stakeholders require, be them individuals, groups or communities and the barriers that might exist for these stakeholders to access this information, and how the information is understood and used.

ACKNOWLEDGMENTS

The wikiRivers project is part of dot.rural, an initiative funded through the Digital Economy RCUK cross-council programme. The James Hutton Institute is part funded by the Scottish Government.

REFERENCES

De Longueville, B., Community-based geoportals: The next generation? Concepts and methods for the geospatial Web 2.0. *Computers Environment and Urban Systems,* 34(4), 299-308, 2010.

Doron, U., Teh, T.-H., Haklay, M., and S. Bell, Public engagement with water conservation in London. *Water and Environment Journal,* 25(4), 555-562, 2011.

Freeman, R.E., Strategic Management: a Stakeholder Approach Basic Books, New York, 1984.

Kundzewicz, Z.W., and A.J. Robson, Change detection in hydrological records - a review of the methodology. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 49(1), 7-19, 2004.

Macleod, C.J.A., Scholefield, D., and P.M. Haygarth, Integration for sustainable catchment management. *Science of the Total Environment,* 373(2-3), 591-602, 2007.

Posthumus, H., Morris, J., Hess, T.M., Neville, D., Phillips, E., and A. Baylis, Impacts of the summer 2007 floods on agriculture in England. *Journal of Flood Risk Management,* 2(3), 182-189, 2009.

Reed, M.S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., Prell, C., Quinn, C.H., and L.C. Stringer, Who's in and why? A typology of stakeholder analysis methods for natural resource management. *Journal of Environmental Management,* 90(5), 1933-1949, 2009.

Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao, P.S.C., Basu, N.B., and J.S. Wilson, The future of hydrology: An evolving science for a changing world. *Water Resources Research,* 46, 2010.

Wexler, M.N., The who, what and why of knowledge mapping. *Journal of Knowledge Management*, 3, 249-263, 2001.

Yang, C., Raskin, R., Goodchild, M., and M. Gahegan, Geospatial Cyberinfrastructure: Past, present and future. *Computers Environment and Urban Systems*, 34(4), 264-277, 2010.