ELSEVIER

Semantic integration of government data for water quality management

Zhiyuan Chen [a], Arrya Gangopadhyay [a], Stephen H. Holden [b,*],
George Karabatis [a], Michael P. McGuire [c]

[a] ITE 431, Information Systems, University of Maryland, Baltimore County (UMBC), 1000 Hilltop Circle,
Baltimore, MD 21250, USA
[b] SRA Touchstone Consulting Group, 1920 N Street, NW, Suite 600, Washington D.C., 20038, USA
[c] Center for Urban Environmental Research and Education, University of Maryland, Baltimore County (UMBC),
1000 Hilltop Circle, TRC 102, Baltimore, MD 21250, USA

Abstract

Normative models of e-government typically assert that horizontal (i.e., inter-agency) and vertical (i.e., inter-governmental) integration of data flows and business processes represent the most sophisticated form of e-government, delivering the greatest payoff for both governments and users. This paper concentrates on the integration of data supporting water quality management as an example of how such integration can enable higher levels of e-government. It describes a prototype system that allows users to integrate water monitoring data across many federal, state, and local government organizations and provides novel techniques for information discovery, thus improving information quality and availability for decision making. Specifically, this paper outlines techniques to integrate numerous water quality monitoring data sources, to resolve data disparities, and to retrieve data using semantic relationships among data sources taking advantage of customized user profiles. Preliminary user feedback indicates that these techniques enhance quantity and quality of information available for water quality management.
© 2007 Elsevier Inc. All rights reserved.

* Corresponding author. Fax: +1 202 338 6106.
  E-mail addresses: steve_holden@sra.com (S.H. Holden), georgek@umbc.edu (G. Karabatis).

1. Introduction

Normative models of e-government typically assert that horizontal (i.e., inter-agency) and vertical (i.e., inter-governmental) integration of data flows and business processes represent the most sophisticated form of e-government, delivering the greatest payoff for both governments and users. With this sophistication, though, comes great complexity of design and implementation that spans the domains of information systems, information policy and public administration. This paper concentrates on the integration of data supporting water quality management as an example of how it is possible to address this complexity in setting and implementing water quality policy.

Data integration problems arise from the implementation of the Federal Clean Water Act (CWA, under Sections 303(d) and 305(b)), which requires states, territories, and authorized tribes to report on the water quality status of jurisdictional waters every 2 years. These CWA mandates create unique data needs and problems, such as how to interpret information derived from multiple sources, of variable quality, using different formats, and collected according to different protocols and procedures. Existing tools for managing these data are not integrated nor do they provide any sort of data analysis capability to allow water resource managers to make informed decisions. The combination of both organizational and data complexity creates fundamental challenges to developing policies, based upon a robust information stream, which are responsive to a wide range of stakeholder interests. Such a problem setting represents the kind of challenge that sophisticated e-government systems are supposed to address.

We describe an interoperable system that builds upon the current digital government literature, allowing users to integrate water monitoring data across many organizations into a data warehouse for subsequent knowledge discovery. Our system makes the following contributions:

1.1. Strikes a balance between breadth and depth in data integration

Numerous organizations and individuals (e.g., volunteers) collect water quality data. Ideally we should integrate all possible data into a uniform format, but in practice this is difficult to accomplish because data sources do not agree on a universal format. We describe a hybrid approach that integrates the metadata (including information on how to access data sources, when and where the data are collected, which parameters are monitored, etc.) of all data sources but fully integrates data only from key sources, such as data from the Environmental Protection Agency (EPA) or U.S. Geological Survey (USGS). Thus, we enable water quality managers and policy makers to search the integrated metadata, to locate any source of interest, and to manually download data from that source; or access the fully integrated data as if the data were from a single source.

1.2. Enables policy makers to retrieve customized views of water quality data

Many aspects of water quality decision making (i.e., optimizing policy outcomes and minimizing public costs) depends on complete and consistent data. The decision making process, includes many diverse interests, including government agencies with different oversight and