Semi-Automatic Service Provision Based on Interaction of Data Warehouses for Evaluation of Water Resources

Dalė DZEMYDIENĖ¹*, Saulius MASKELIŪNAS¹, Giedrė DZEMYDAITĖ², Arūnas MILIAUSKAS¹

¹Institute of Mathematics and Informatics, Vilnius University
Akademijos 4, Vilnius, Lithuania
²Faculty of Economics, Department of Economic Policy, Vilnius University
Saulėtekio 9, Vilnius, Lithuania
e-mail: daledz@mruni.eu, saulius.maskeliunas@mii.vu.lt, giedre.dzemydaite@ef.vu.lt, orunas@gmail.com

Received: February 2016; accepted: November 2016

Abstract. Our research is devoted to development of information infrastructure for e-service semi-automatic provision by using distributed data warehouses (DWs) of water protection domain. Development of software for semi-automatic service provision is based on artificial planner and structure of goals adapted for specialized needs of end users. Such e-service preparation mechanism can work under the unified coherent framework for solving the environment protection problems by evaluation of the processes of water consumption and contamination. The possibilities of integration of distributed DWs of water management sector into web portal meeting the requirements of conceptual interoperability are presented. Design approach is based on development of decision support system (DSS) that is designed as multilayered system with multi–componenetal, interoperable structure of databases (DBs), which are under responsibility of different public administration institutions such as European Environment Information and Observation Network (EIONET) and national environment protection agencies. The infrastructure of EIONET is used for supporting and improving data and information flows. The Water resource management information system (WRMIS) became the kernel component of DSS. WRMIS prototype facilitates data flows between the institutions and gives access to data for relevant institutions and the public providing e-services using proposed DSS. The research investigations are made according to the requirements of European Union Water Framework Directive, Sustainable Development Strategy and ReportNet as the EIONET infrastructure for supporting and improving data and information flows. Additional means are integrated in the structures of DSS as knowledge representation techniques based on conceptual schemas, data flow diagrams, and decision-making rules. The on-line management techniques are based on assurance of interoperability by using Open Web Platform W3C standards for web service development, such as XML, SOAP, HTTP, etc.

Key words: e-services, decision support system (DSS), semi-automatic e-service composition, data warehouses (DWs), water consuming and contamination processes.

*Corresponding author.
Introduction

Our universe of discourse concerns the representation of full scale of interconnections between data warehouses (DWs) for possible semi-automated e-service composition in the recognition of abnormal situations in environment protection sector related with the water resource management and contamination processes (Dzemydiené et al., 2008a, 2015). We would like to propose an approach for the multi-dimensional situation evaluation that allows us to support interoperability of distributed DWs design as well as for analysis possibilities by developing of structural constructions of meta-data models for controlling of DWs. Such techniques help us in designing of structures for e-service provision based on semi-automatic e-service composition. Such technological platform requires special organizational–technological applications, too. Goals for semi-automatic e-service provision are related with such activities for evaluation of data flow processes (Portchelvi et al., 2012) and special components for recognition of abnormal situations of water consuming and pollution. On the other hand, most of the systems are found to be both heterogeneous and scattered, causing difficulties to use data and information in an integrated manner.

As related works in the area of decision support systems (DSS) development can be mentioned works proposed by Mysiak et al. (2003) in the area of environment protection and under the MULINO project for water resources management, and the methodology of designing of environment DSS is proposed in the work of Poch et al. (2004). Some authors as J. den Boer, E. den Boer, J. Jager have proposed the decision support tools for waste management and treatment (den Boer et al., 2007). Development of Lithuanian districts and regions in accordance to sustainable development requirements and based on risk analysis is provided in Dzemydaitė and Galinienė (2013), Galinienė and Dzemydaitė (2012) research studies. The approach of evaluation of quality of surface water by using GIS and a heavy metal pollution index (HPI) model in a Coal Mining Area of India presented in Tiwari et al. (2015). How the network of pipe replacements influence the water supply network quality is presented in Van Dijk and Hendrix (2016).

The amounts of environment protection data collected and stored in DWs have dramatically increased in recent years. In addition, there is a lack of strong traditions for coordination and sharing of data and knowledge. The aim of our research is to present the coherent framework for the interoperable DWs. We suggest that data require adjusted multi-dimension structures of data warehouses. One layer of this framework is devoted to the orchestration of means which help to integrate the software for e-service provision. The artificial planner is implemented for adaptable e-service provision based on integrated semi-automatic tools.

The examples of e-service composition are presented using workflow structures. For the on-line e-services provision we are designing the infrastructure of interoperable data warehouses and present possibilities to realize the data flows between different institutions of public administration. Shown results illustrate the multi-aspect data analysis possibilities of water consumption, contamination and treatment in districts of Lithuania.
1. Components of Interoperable Data Warehouses for Evaluation of Water Consumption, Contamination and Treatment Processes

Continually we have evaluated the situation of usage of water resources, wastewater effluaxes, waste utilization, contamination levels, protection efficiency, etc. in the different regions and districts.

Comprehension of information is associated with modern information processing, storage, and transmission technologies, new kind of hardware. Related information systems development approach is presented in Gudas and Lopata (2016), Dzemydiénė and Dzindzalieta (2012). Additional means are integrated in the structure of framework such as knowledge representation techniques, conceptual schemas, data flow diagrams, the on-line management techniques for evaluation of processes and special components for provision of e-services and recognition of abnormal situations of pollution.

The main framework of the approach represents the layers which are needful for securing requirements of interoperability (see Fig. 1). The layers of interoperable functions are represented in left part of the framework and detailed parts show the integrated components which are needful for technological solutions, implementation of protocols, and composition of interoperable parts of the system for provision of e-services for water management sector.

The evaluation of the region has complex infrastructure and decision making processes follow requirements of adaptability for providing e-services in web environment (Dzemydiénė et al., 2015; Swanson et al., 2004). The on-line e-services allow to access the information in distributed DWs and to recognize the situation of water quality in rivers and lakes, marine, ground water, and point sources of emissions by enterprises.
2. Assurance of Interoperability Requirements Between Integrated Parts of DWs

The structure of main parts of the Water Resource Management Information System (WR-MIS) was presented in our previously described works (Dzemydienė and Maskeliūnas, 2011; Dzemydienė et al., 2008a, 2008b, 2015). The purpose of this work is aimed at the problems of provision of e-services by integrating the structures of distributed information systems, while estimating proposals in what ways to simulate situations, intellectualize environment pollution estimation support by evaluating the emissions from contamination objects.

Most of the existing environmental IS have evolved during a long period of time. High data storing, analysis and reporting requirements should be met (Saarenmaa et al., 2002; Bonino da Silva Santos et al., 2009). For assurance of interoperability we are using the XML, SOAP (simple object access protocol), and HTTP protocols. The collaboration processes of stakeholders at water management and wastewater treatment sector of Lithuania are presented in Fig. 2.

---

**Fig. 2.** The collaboration processes of stakeholders at water management and wastewater treatment sector of Lithuania.²

---

²Denotation: EC – European Commission; EEA – European Environmental Agency; EPA – Environmental Protection Agency; EPD – Environmental Protection Department; MoE – Ministry of Environment of the Republic of Lithuania; REPD – Regional Environmental Protection Department; ReportNet – infrastructure of EIONET for supporting and improving data and information flows.
WRMIS is working as an umbrella integrating all needed components of the domain. As the example of interoperability between distributed DWs we would like to present realization of information system (IS) working as the collection of data from Marine and transitional water sector (ISMA). ISMA was developed for Marine Research Centre (MRC) in Klaipėda under the DANCEE project “Implementation of the EU Water Framework Directive, Lithuania” (DANCEE, 2004).

The ISMA is working as one part of sub-systems which we integrate in common structure of WRMIS. The main peculiarities of ISMA are following:

- ISMA is functioning in local area network. All data are integrated into one universal ISMA database, stored in one central database (on one computer in the local network). On other computers there is installed small client software.
- DB data is integrated into one MDB file. The separation of the user interface and the business logic from the database prevents changes in the developed application in case the database and its structures will be moved to a more powerful database tool.
- ISMA presents filtered data visualized in the form of tables and charts, exports data to (and imports from) Microsoft Excel, generates EuroWaterNet data and reports.
- The development tool: Delphi 7.0; the application communicates with database using ADO and SQL interface. Data was imported to ISMA from delivered sources, and the developed data entry forms are used.

The main strategic aim was to develop flexible and universal ISMA application which:

- could be adjusted to continuously changing requirements in MRC and Environmental Protection Agency (EPA), without a need to re-program the developed ISMA application, just modifying setup and database peculiarities by ISMA administrator (on the other hand, remaining understandable, not too complicated for administrator and ISMA users, “intuitively” manageable);
- would fit not only data management needs of Marine Research Centre, but the data management needs, requirements of any other Water Framework Directive-compliant monitoring (i.e. Surface water monitoring in rivers and lakes by EPA and Regional Environmental Protection Departments, for Biological data, etc.).

For these purposes we deal with the decision support integrating the components of different DWs. The parts of the data warehouse and web applications are developed by standardized means of web service development tools. The special parts are developed by using the meta-models of the unified networking of data warehouses for interoperable, conceptually adequate different DB (see Fig 1).

3. Infrastructure for e-Service Provision by Using Integrated DWs

A dynamic e-service orchestration model is developed on the basis of all interconnected components of the framework presented in Fig. 1, and additionally include the mecha-
nisms for assessing and controlling data flows in distributed processes and determining responsibilities. E-service oriented architecture integrates coordination engine of e-service provision and all components needed for the realization of e-service (see Fig. 3).

For illustration of such e-service provision engines we would like to present the example of the implementation of data grouping and meta-data construction on the base of Marine Research Centre (MRC) DW. Such data will be joined in one DW with simple structure; however, such large amount of data should be divided logically by Data Groups (DGs) according to existing (initial) data representation and institutional structure (according to the needs of data operators and MRC departments). Such distribution will be represented by data entry and analysis forms.

The expanded tree structure of realization of grouping and data inputs of such ISMA sub-system is presented in Fig. 4.

There are different classification levels:

By Observation types:
- Hydrological observations (H);
- Meteorological observations (M);
- Hydrobiological observations (B);
- Hydrochemical observations (C);
- Ecotoxicological observations (T).

By Station types:
- Marine Stations MS;
- Lagoon Stations LS;
- Posts (all) PS;
- Marine Posts MP;
- Lagoon Posts LP.
By *Sample* types:

- Water samples smW;
- Sediment Samples smS;
- Biota Samples smB.

In some cases posts could be sub-grouped:

- Recreational (R);
- Costal (C);
- Open (O).

"Analyse" menu item allows to filter ISMA database data. Available expandable/collapsible filters (see: Fig. 5):

- by station names, water types, station attributes (e.g. Marine, Posts, etc.);
- indicated date interval, season, time;
- depth of samples, parameter value interval, species;
- writing custom SQL query.

Filtered data can be presented in a table form ["Query" tab], as the list of calculated statistical parameters ["Statistics" tab], and in a graphical form ["Chart" tab].

The "Query" tab is presented at the beginning; the "Apply filters" button shows filtered data in table-view here.

The "Statistics" tab allows to calculate statistical parameters of filtered data, with the "Refresh" button to present (re)calculated statistics for the filtered data.
The “Chart” tab gives chart-view of analysed data; here: “Refresh” button applies to the selected filters: the chart is drawn, and below this chart – the filtered values of parameters and resume information of prepared chart are presented; “3D” checkbox allows to prepare three-dimensional charts.

Hydrological observations data with its complex structure could be classified by Observation Object types: Bacterial plankton (bioBP); Chlorophyll (bioCH); Phytoplankton (bioPP); Zoo benthos (bioZB); Zoo plankton (bioZP); Macrophytos (bioMP).

Using such definitions, specific Data Groups could be described shortly, for example:

- H-LS: hydrological observations data collected from lagoon stations;
- T-MS(R): ecotoxicological observations data collected from marine recreational stations;
- T-MS-smB: ecotoxicological observations data collected from marine stations biota samples.

4. The Illustration of Results of e-Service Provision Using Multi-Aspect Analysis of Observable DWs data

The values of the observed parameters may change dynamically, depending on time and the events occurring. Solution of different problems is interfered with one to another.
Example of one kind of e-service provision: the end user can provide the goal such as: *KeepWasteWaterContaminationAtSafeLevel* that can be further decomposed to lower level goals like: *TrackContaminationLevels*, *ReportIncreasedContaminationLevels*. These are also a subject for further decomposition (see Fig. 6).

The tasks are accomplished supported by program modules which are chosen from prepared list of service patterns. A task can be an *AtomicTask* or a *ComplexTask*. Tasks are directly mapped to a Hierarchical task network (HTN) planning domain. An *AtomicTask* corresponds to an operator in HTN (*HTNOperator*), a *ComplexTask* is mapped to *CompoundHTNTask*, which uses *HTNMethods* for a task decomposition.

The overall solution architecture of the e-service semi-automatic composition system is presented in Fig. 7.
Having a knowledge base conformant with this model, the system can decompose the goal to tasks and map them to a HTN planning domain. In an e-service composition scenario all tasks should be members of ComputationalServiceTask, which is performed by a ComputationalService (an e-service). This way the planner result is transformed to e-service execution calls. Our approach requires a domain specification based on GSO (Bonino da Silva Santos et al., 2009) and a SHOP2 planning (Nau et al., 2003) before a composition start.

One of the main three parts of infrastructure is the integrated middle-unit software, the basis of which is creation of cooperation environment for message interchange as well as using the standard means of network services. In process of public e-service provision this software should be grounded by knowledge control components that ensure decision-making and management activities which allow us to collect and send services by user query requests and data as well as to make related transactions (Sirin et al., 2007). The software packages are interconnected, and some integrated means of applied programs are used. The schemas of workflow representing the back-up processes are developed and implemented (see Fig. 8).
The system provides information via e-services. List of e-services provided by the system may not be complete yet; therefore a composition of several services may be needed. Here we promote ability to create composition from contextual information and user requirements, without a software developer involvement. That will minimize time needed to create the required composite service.

Along with the integrating functions of the applied programs the mentioned programming tools of the service bus are applied. These functions are realized in the Collaboration interface structures of coordinating engine. The Service bus performs and makes several additional tasks related with decisions: shy transform XML documents from one format into another, use common data from different information and operation systems, redistribute the run of processes according to the loading of systems and in case of failure they may pass the message interchange function to another server, etc.

Real-time subsystem is embedded in the target system as a concurrent computing system related with the monitoring of data. The monitoring subsystem connected with expert subsystem must detect the faults of process performance. The time for obtaining a solution is often strictly limited. These conditions impose strict deadlines on the obtaining a decision and maintaining the functioning correctness. The task structure relationship with information elements, the course of decision-making processes and presentation of alternative variants of decisions are represented in the sub-models.

Conclusions

The results presented in this paper describe the developed framework of information infrastructure for e-service provision in water protection domain area allowing the integration of different data warehouses and data bases which are distributed by different responsibilities of institutions. Development of software for semi-automatic service generation is based on artificial planner and structure of goals adapted for the needs of end users. Such e-service generation mechanism can work under the unified coherent framework for solving interrelated problems of the application domain. The European Environment Information and Observation Network have responsibilities for ReportNet – the infrastructure for supporting and improving data and information flows. All countries provided data about specific environment data changes. The Geological Survey of Lithuania; Lithuanian Hydro-Meteorological Service; Ministry of Environment of the Republic of Lithuania; Marine Research Centre are such institutions responsible for providing and maintaining such data in Lithuania. Water resource management information system combines multi-dimensional infrastructure of information. The consideration of real time process control and enterprise functioning is organized by the integration of sustainable development requirements for retrospective analysis of contamination processes. The components of decision support information infrastructure are described for aims of assistance in contamination evaluation of wastewaters.

The developed levels of representation of dynamical aspects of observable processes (monitoring subsystems, the WRIMS and others which help in realization of communication between DWs) help in multiple objective decision making. The presented approach
shows possibilities to construct semi-automatic e-service composition on the precisely developed infrastructure of the interconnected DWs. For realization of such system the interoperability requirements have to be supported for all levels of data structures.

Acknowledgements. Authors thank to European Environment Information and Observation Network (EIONET), Lithuanian Environment Protection Agency for information support and collaboration possibilities in the Water Resource Management Information System (WRMIS) development.

References


---

**D. Dzemydiene** is a leading professor, working at the Mykolas Romeris University (Lithuania) and at the Institute of Informatics and Mathematics of Vilnius University. She holds a diploma of applied informatics and engineering in 1980 at Kaunas University of Technologies, PhD in mathematics–informatics in 1995 at Vilnius University, habilitation doctor procedure in 2004, and professor diploma in 2009. She is the supervisor of doctoral studies in the field of informatics engineering. She published over 100 scientific articles and 5 books. She is the organizer of international conferences of Baltic databases and information systems and member of editorial boards of some international journals. Her research interests include development of knowledge management systems, design methods of business management systems, intelligent service support in mobile applications.

G. Dzemydaitė is a doctor of economics, working as a lecturer at the Department of Economics Policy of Vilnius University (Lithuania). She obtained a diploma of Master of economics from Vilnius University, and obtained her PhD of social sciences, economic in 2016. Her research interests include: innovations and economic development methods of regional development, indicators of sustainability.

A. Miliauskas is a doctoral student of technological sciences in the field of informatics engineering at the Software Engineering Department of the Institute of Mathematics and Informatics at Vilnius University (Lithuania). His research interests: software development, artificial intelligence methods, e-service composition tools.

Iš dalies automatinis elektroninių paslaugų teikimas sąveikų duomenų saugyklių pagrindu vandens ištekliams vertinti

Dalė DZEMYDIENĖ, Saulius MASKELIŪNAS, Giedrė DZEMYDAITĖ, Arūnas MILIAUSKAS