

BUBI: An Interactive Water Utility Benchmarking Website

Jolma, A., Smits, K., Salminen, A.

Water Resources Laboratory, Helsinki University of Technology, Finland

The World Bank, Riga, Latvia (now at Queens University, Kingston, Ontario, Canada)

The World Bank, Washington, D.C. USA

Key words: benchmarking, World Wide Web, water utility, performance indicator

Abstract: A web-based benchmarking support system was developed to support management of water utilities. The tool helps experts to assess the performance of a utility and compare it with others. In addition the system is a tool for the utilities to publish their performance data. The web implementation allows co-operation and comparisons between utilities. A system of variables, comprising about 130 input variables and 120 output variables was developed and stored in the facilitator database. The facilitator hosts a website, which contains public, member only, and developers' sections. The website includes interactive pages to access the facilitator database for for example benchmarking report generation, discussion, data input, and database development. The facilitator system was developed on a Linux-based GNU system and it utilises also several other freely available software packages, including a relational SQL database, web server, graphing programs and libraries, all of which are glued together with Perl programs. A general finding of the study has been that a website is a good candidate as a central element of an environmental information system especially when co-operation support is needed.

1. INTRODUCTION

A water utility is either a water supply plant or a wastewater treatment plant; a key element in community's well being and in its environmental impact. Environmental and other performance indicators are being used to monitor how they succeed in this efficiently (Rogers 1999, Blankenship et al. 1998). Indicators and reports are targeted and prepared for top

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35503-0_29](https://doi.org/10.1007/978-0-387-35503-0_29)

R. Denzer et al. (eds.), *Environmental Software Systems*

© IFIP International Federation for Information Processing 2000

management of the utilities, general public, authorities, and financial support organisations. Benchmarking can be defined as the use of the performance indicators by the utility management in a search for best practices and improved quality of drinking water and effluent (Camp 1989, Blankenship et al. 1998). Benchmarking contains the idea of learning from others and from their experiences. The learning can also happen in a co-operative framework.

The water utilities are publishing information in the web, as are other industries. A recent survey of water quality data published on the web by water utilities in the U.S. is done by DeNileon (1999). The general impression he reports is the sometimes lacking accuracy, timeliness, and especially reliability of the presented data.

After the Baltic states (Estonia, Latvia, and Lithuania) became independent less than ten years ago, there has been much interest and work done in upgrading their water utilities. As a part of its Baltic Environmental project the World Bank initiated the development of a database, website, and tools for benchmarking the performance of the aided water utilities. A separate facilitator would manage this information system. After initial development, which introduced and proved the concept, another system, named BUBI after Baltic Utilities, Benchmarking Indicators has been developed and is planned to be operational soon.

This paper describes tools and techniques that were and are used and/or developed for the water utility benchmarking information system. The main attention will be given to the website and its supporting (facilitator) tools. Some tools were also developed to aid the utilities in preparing data for the facilitator database, uploading and downloading data and spreadsheets, and working on benchmarking locally.

2. BENCHMARKING

Benchmarking is a management tool, which is based on assessing the performance of a company or a plant using a set of indicators. In the water industry much work has been done to develop an indicator system for common use (e.g. Yepes and Dianderas 1996), however there is no universally agreed set. Camp (1989) divides benchmarking into five main phases: (i) planning, (ii) analysis (iii) integration, (iv) action, and (v) maturity. This study concentrated on the first phase, identification of the indicator set and data collection and developing a support system for activities in other, later phases.

Benchmarking is a process and a method for co-operation. Benchmarking includes an idea of striving towards set goals, "benchmarks", which can be certain abatement rates for various constituents, various financial ratios, etc.

Benchmarking can be used to create a common ground for comparisons and discussions for the reasons why the figures are different at different utilities.

3. INDICATOR SET

Much of the functionality of the BUBI website is due to the development of the system of variables, of which the indicator set is a subset. From the utilities point of view an important distinction between different variables is whether they are input variables or calculated ones. The variables are arranged into categories of which there are three levels. The main categories are basic data, customer satisfaction, financial, productivity, tariffs, water quality, and environment. Other attributes of variables are, e.g. the rate at which it is collected or calculated, the unit, and the publicity. The collection rate is monthly, quarterly or yearly. The publicity is private, shared, or public. Also the equations of the calculated variables are stored in the database. The system of variables is also a part of the data model of the facilitator database.

The facilitator database stores data from three different kinds of entities: (i) water utilities, (ii) cities, and (iii) countries. Usually there is only one water utility in one city but the separation was made to emphasise the financial separation of them. The water utility may have many plants, both for water supply and wastewater treatment, but this cannot be reflected on the database using the current datamodel.

At present there are about 130 input variables and 120 calculated variables.

4. THE WEBSITE

The website is divided into three major parts with different target groups and permissions.

The public part of the site, which will be open to anybody once the system is officially opened, contains general information about the project, about indicators which are used, and about the participating utilities. There are also a large number of tables, charts, and graphs available, which are made using the public data and indicators. It is assumed that there is an interest from the utilities side to publish data on their performance, and that there is an interest from the side of the customers of these utilities to compare how their local water utility measures with other utilities.

The member part of the website is accessible only by a valid username/password pair. This gives access to information, which is not public but shared among the members, and to the user's own data, if any, via interactive web forms. Super users can grant access to input and update data

of a given utility; city or country and they can also grant rights to use private data in benchmarking. The member part of the website contains also a number of general-purpose tools and information, e.g., discussion forum and documentation.

The developer section of the website contains tools to manage the database, i.e., edit or create variables, add new utilities and/or users or edit their permissions. It also contains a tool for defining the ready-made tables, charts, and graphs. Thus the developer interface provides a rather comprehensive control over the facilitator database and dynamic behaviour of the system.

The design of a website considered several things, including security, usability, functionality, and standards. The security is a many-faceted issue but two important things can be pointed out: the security of the connection and the security of the interactive pages. The standard http connection is not secure; anybody with an access to the Internet at the route of the connection may listen to it. This was not considered an important issue and standard http, instead of secure http (https), is used since the data, usernames and passwords that are transmitted are not of considered of much economic or other value to others. The security of the interactive pages is perhaps a more important issue since they may provide intruders with an access to the server. The standard method in Perl, taint checking was used, to filter out possibly dangerous input from the client to the scripts.

The website uses a rather common layout and navigational mode, which includes menus, navigation bars, direct links, and visual elements to emphasise certain parts on the screen. The home page of the member section is similar to this but uses different colours and has more direct links. The design of the site tries to be as horizontal as possible to avoid unnecessarily long paths to information. The functionality of the website is a crucial part of the functionality of the whole system. A design principle of supporting more than one way to do things was applied. This should make the system more robust and more adaptable to various users' preferences.

The architectural overview of the system is shown in figure 1.

5. BENCHMARKING REPORTS

Central elements of the website are the report generators. The report generators draws data from the facilitator database, arranges it for presentation, and generates a report based on the user's profile and user-defined parameters. The parameters are defined by a developer using a form and stored in the database in the case of ready-made reports or online by a set of forms by a member.

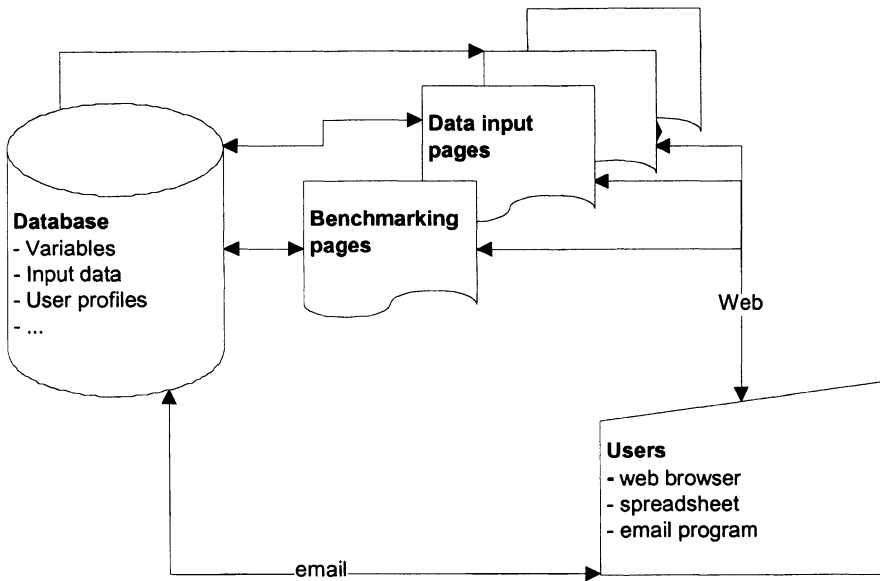


Figure 1. A schematic showing the three central elements of the benchmarking system: (i) the SQL database and programs, (ii) the HTML pages with varying degrees of dynamic content, and (iii) the user with various programs

The main form of the online report tool contains menus and other interface elements to select data/indicators, time period, utilities, and report format and style. The user can choose between a simple method, e.g., select an indicator category instead of picking individual indicators from a long list, and more advanced methods of setting parameters, i.e., a text input field to set the time period.

Currently the online report generator supports four formats for reports: (i) plain text, (ii) table, (iii) chart, and (iv) graph. The plain text format is suitable for transferring to other programs. The table format uses the standard HTML table. Charts are barcharts or area charts and suitable for comparing visually a small number of values or, in the case of area charts, additive values. Graphs are plots of timeseries. Typical publicly available reports include, e.g., total water produced per capita per day, tariff comparisons, and nutrient removal rates. Typical reports shared by participating utilities include, e.g., operating ratio, working ratio, the amount of unaccounted-for water, and total labour cost divided by the total operation and maintenance costs. Figure 2. shows an example of a productivity report on the public section.

The standard HTML does not allow very complicated forms or forms with dynamic interaction between interface elements. This has been overcome mainly by three means: (i) the interface has been kept as simple as possible, (ii) sequential forms, also known as "wizards" have been utilised, and (iii) some disfunctionality has been allowed, i.e., in some cases the form can be used to create a request which does not lead to a useful page. An example of a sequential form is the data input form which has the following sequence:

1. Ask for the object (generally an utility) of which data is to be edited/inputted (this can be skipped if the user is allowed to input data for only one object).
2. Ask for the specification of the data, there are panels for each data collection rates each with a number of checkbox options and drop-down menus to specify the data category.
3. The actual data input form with default values taken from the database.

Another report generator prepares a simple graphical comparison of utility's indicator values against the best and the worst.

6. THE SOFTWARE

The facilitator computer, which hosts the website and the database, is a Linux based system with mostly GNU or other free system software. The database is a free relational database, PostgreSQL, which is accessed using SQL. The web server is Apache and CGI software written in Perl 5 is used. Perl is used also for the other facilitator tool programs.

Software that was written for this system includes an email daemon, a program to evaluate values of calculated variables, a program to create HTML pages, e.g. the ready-made benchmarking reports, from templates and information in the database, and other smaller utilities. The email daemon accepts data submissions, certain database queries, and messages to be posted on the website. It duplicates the functionality of the website and is intended as a backup or an alternative means to use the system.

The program, which evaluates calculated variables, is probably the most complex single part of the system. The functionality is closely related to that of a spreadsheet model where the value of a cell depends on the values of other cells as defined by the equation in the cell. The equation is in this case are stored in the database and may include algebraic expressions of the code names of other indicators and standard and few custom functions.

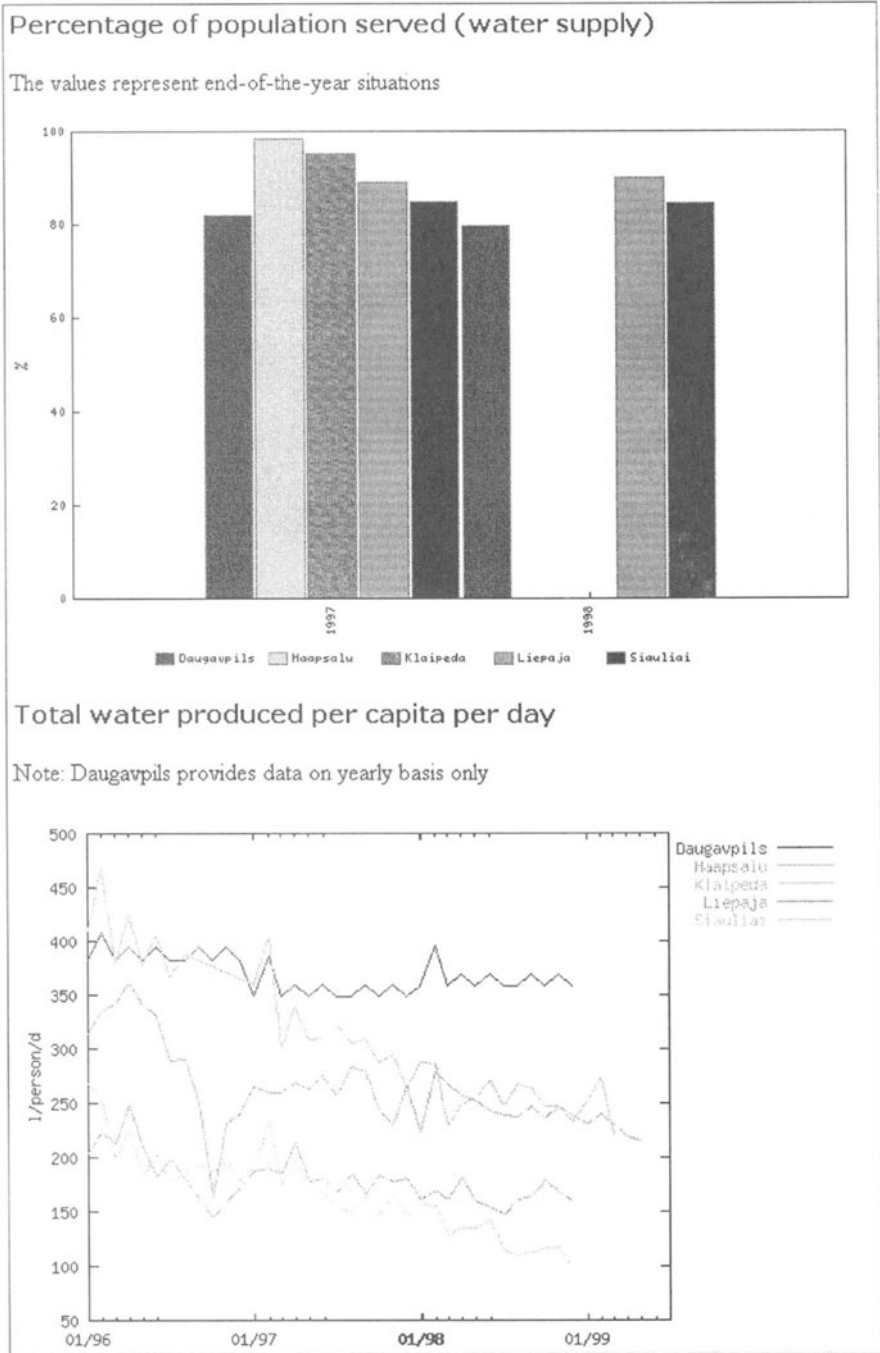


Figure 2. Graphs on the public section of the website

This system has two advantages over a corresponding spreadsheet model, namely that it allows the user to define more than one equation for an indicator and to directly input values for calculated variables, e.g. circumvent the normal calculation. The latter is useful if some input data is not available. The algorithm of the program is shown in Figure 3.

The program which generates web pages from report templates traverses a directory structure, searches for web page templates, checks if they are newer than the corresponding web pages, and emulates a web server in the sense that it generates the reports on the place of the include structures. The templates could be used as web pages, then only the webserver would generate the reports on the fly, but this method saves processing time since the contents of the pages do not change so often. The report templates are themselves generated from even plainer templates by adding the include structures which define the ready-made reports.

```

foreach object, i.e., countries, cities, and utilities {
  start from the first year there is data from and proceed month by month {
    gather all that is known, use also data from superobjects and appropriate
      data from previous time steps and longer time steps
    gather all equations that need to be solved
    solve custom functions
    substitute codes of known variables with their values
    evaluate until all that can be done is done {
      foreach equation that needs to be solved {
        evaluate (use Perl's eval)
        check if the value is new, same as the old, or different than the old
        add to the pool of known data if appropriate
        solve custom functions
        substitute codes of known variables with their values
      }
    }
  }
}
do database updates
do database inserts

```

Figure 3. Algorithm for calculating variables from other variables

Both the CGI and other software use two Perl 5 modules written for this system: one for basic tasks and the other for benchmark reports. The

modules are not object-oriented and simply encapsulate some common constants, variables, and subroutines. The basic subroutines are mostly written to encapsulate common database queries or parse input text.

The programs take advantage of a number of freely available Perl modules, which make them relatively small and more robust. The modules include CGI.pm, which greatly simplifies writing CGI programs, Postgres.pm, which provides a simple way to access data in PostgreSQL database, GIFgraph, which is a charting and graphing library, and several other modules. Besides a driver for GIFgraph, drivers for other charting and graphing libraries or programs were written, including Gnuplot, jsplot, and DISLIN.

7. DISCUSSION

The number of indicators in the system is very large. It is partly explained by the fact that there are both monthly and yearly versions of many indicators. There are also a large number of constituents, which are measured from the raw water, potable water, influent wastewater, and effluent wastewater. An additional reason for the complexity of the indicator set is that initial data is imported into it from other data sets, which have used and use different indicator sets. It seems that an important part of this project, at least in its beginning, will be discussions on what are the indicators and how are they obtained. Another matter complicating the development of the variable systems is the fact that the procedures and bookkeeping in many participating water utilities are not yet very well established. This is visible in the data in many ways: as missing values and as inconsistencies, for example incoming and outgoing funds or water are not in balance where they should be.

The design of a website, especially graphics and things related to usability, is time-consuming work, which requires skill. On the other hand the productivity, which can be achieved with the modern programming tools using the web infrastructure and exiting tools and programs, e.g. the standard web browser as an interface, is very high. The website and the facilitator tools are developed by freely available and freely usable software. Usually the source code is also freely available. The quality of this software is usually excellent and if problems arise, the developer and user community is responsive and the problem may be solved fast with their help or – the source code being available – the problem may be solved by changing it.

The modelling capabilities of the system are limited but extensible. More complex models that could be included into the system include forecasting models and models to calculate areal values. With these extensions the

system could be used for more comprehensive environmental management and monitoring.

8. CONCLUSION

A website can be a central element of a water utility benchmarking effort, especially if the effort includes co-operation.

9. REFERENCES

- Blankenship, L., Olstein, M. and Liner, B. 1998. Metric benchmarking. *Jour. AWWA*, **90**(6), 57–62.
- Camp, R.C. 1989. *Benchmarking: the search for industry best practices that lead to superior performance*. ASQC Quality Press, Quality Resources, 299 pp.
- DeNileon, G.P. 1999. Surfing the web for water quality data. *Jour. AWWA*, **91**(2), 46–54.
- Rogers, P. 1999. Editorial. *J. Water Resour. Plng. Mgmt.*, March/April, 74–75.
- Yepes, G. and Dianderas, A. 1996. Water and wastewater utilities, Indicators (2nd ed.), Water and Sanitation Division, The World Bank, May 1996.

Homepages of the software mentioned in the paper

DISLIN: <http://www.linmpi.mpg.de/dislin/>

Gnuplot: http://www.cs.dartmouth.edu/gnuplot_info.html

jsplot: <http://gd.tuwien.ac.at/visual/jsplot/>

Perl: <http://www.perl.org/>

Perl modules: <http://www.cpan.org/>

PostgreSQL: <http://www.se.postgresql.org/>