Water Institute of SA

192
DATA HANDLING AT THE DASPOORT LABORATORY

CM Esterhuyse

Waste Water Treatment
City of Tshwane Metropolitan Municipality
PO Box 1022
Pretoria
SOUTH AFRICA
Tel: +27 12 358 0702
Fax: +27 12 358 0696
kerneelse@tshwane.gov.za

ABSTRACT

Today's laboratory leader must have insight into the operation of a laboratory as a profitable business that competes in the global sphere which involves more than just analyzing samples. This article discusses the evolution of data handling at the Daspoort Laboratory from using spreadsheets to a secure data storage solution that is centrally managed while being readily available to legitimate users, potentially anywhere in the world, via the Internet.

INTRODUCTION

The Daspoort Laboratory of the City of Tshwane Metropolitan Municipality (CoT) generates huge volumes of analytical data from a number of monitoring points, e.g. various Waste Water Treatment Works (WWTW), rivers, boreholes, industrial effluents, and public municipal swimming pools. The Daspoort facility consists of different laboratories analyzing industrial effluent, wet chemistry laboratory, sludge analyses laboratory, two instrument laboratories, and a microbiology laboratory housed in one building. Prior to the improvement project, all of the laboratories' data were being stored in a number of separate spreadsheets.

In addition to laboratory testing, the Waste Water Treatment Section manages online auto-analyzers at all their WWTW sample points. These analyzers each perform the same four analyses on four different monitoring points every hour. The online data were captured and stored separately from the laboratory's data but were rapidly becoming too difficult to manage due to the increasing volume and number of users.

USING SPREADSHEETS

In a typical Daspoort Laboratory scenario, a spreadsheet was initially created by a single user for their own use. The need to share information led others to start using it and perhaps modify it so that it grew bigger and more complex and over time often evolving into a business-critical application. Spreadsheets are well suited to certain tasks, but have limitations when used as an enterprise tool. The following conditions are often an indication that an alternative solution should be considered: large volumes of data; multiple simultaneous users; access control required; audit trail required; limited editing of data and macros; and inconsistent layout by the different users.

MULTI-USER DATABASE

As soon as a spreadsheet application meets several of these conditions, a multi-user database application is often a better solution. The advantage of a database solution is that it addresses all of the conditions but still allows data to be easily extracted into a spreadsheet for analysis, graphing, etc. – without requiring any expert knowledge. The advantages of a multi-user database application are outlined in Table 1.
Organizations have traditionally been reluctant to abandon their spreadsheets in favor of a database because of the high cost involved. However, this trend is changing due to the availability of free, high quality database packages (e.g. PostgreSQL, MySQL).

Table 1: Advantages of a multi-user database application:

Parameter of interest	Multi-user database	Spreadsheet data storage
Security and control	Various levels of security can be implemented – database level, application level, network level, etc.	One password to protect a spreadsheet from changes. No levels of security can be implemented. If you lose the password, there is nothing you can do.
Concurrent access	Designed to support concurrent use for both reading and updating.	With password protection, spreadsheets cannot be read and updated both concurrently by various users.

Web access	Employees, customers, anyone can have access.	Access can be made available to a file server.
Scalability	Designed to support a few or up to even thousands of users, depending on hardware configuration.	Files can be read by multiple users, but updated only by one user at a time.
Support for relationships between entities	Avoids duplication of data.	Data can be duplicated.
Audit trail	Application can implement audit trail to any level of detail.	Audit trail limited to last user that saved the spreadsheet.
Data validation	Both standard and custom data validation possible.	Both standard and custom data validation possible.
Reporting	Many options available due to established standards (SQL, ODBC, etc).	Limited reporting options available.
More robust	Spreadsheet data, programs (formulas and macros) and formatting are held in one place – potential for errors are limited / minimized.	Spreadsheet data and formulas are held in various places (each cell), and potential for errors are high.
Searching and retrieval and trend analysis	Relational databases are optimized for efficient processing of large data volumes (indexing, caching, etc).	Spreadsheets have limited rows and columns for trending of large data volumes.

LABORATORY INFORMATION MANAGEMENT SYSTEMS (LIMS)

LIMS is a computer software application that operates on a multi-user database (3) consisting of tables of sample results logged manually or automatically into the system. A LIMS allows the electronic collation, calculation and dissemination of data, often received directly from instruments (4). It can include word-processing, spreadsheet and data-processing capabilities and can perform a variety of functions, including sample registration and tracking, analyses assignment and allocation, worksheet generation, processing captured data, quality control, financial control, and report generation. LIMS can be restricted to the laboratory itself, or it may form part of a company wide computer system (Intranet), or even part of the world-wide-web (Internet). The LIMS reporting can be either electronic files or hard copies. A well designed system enables high levels of Quality Assurance (QA) to be achieved from point of sample entry to the production of a final report. Validation requirements include management of access to the various functions, and audit trails to catalogue alterations and file management.

SELECTING A LIMS

Depending on the laboratory's mission, financial resources, and availability of "computer literate" people, there are three recognized routes in implementing a LIMS (3)—(i) in-house development, (ii) customized software, or (iii) purchase available commercial software. In-house development is the most complex route requiring expert personnel with programming and laboratory skills. The second route utilizes an external consultant to customize software to conform to laboratory operational requirements; this option can be risky and expensive if the consultant is not carefully selected. Commercially available software is recommended as the safest and best alternative (3) but one should still shop around or retain a reputable consultant since there is a tendency to buy extra functions that may not be needed in the laboratory.

LIMS purchasers have a well-developed methodology for evaluating which system best meets their needs (6). The Daspoort Laboratory's requirements were carefully documented and weighted against the functionality provided by commercially available solutions. The functional specification included the following criteria: must support at least 12 users simultaneously; support manual entry of laboratory analysis results; capable of direct interface to laboratory instruments for electronic capture of data; automatic import of data from online auto-analyzers and perform hourly data reduction to determine a number of significant quality control parameters; data should be centrally stored, accessed and automatically backed up daily; billing information must be generated; reports for customers must be easy to generate; access to system functions must be defined by user security category; must generate an audit trail to track changes made to analysis data; use open source components that run on standard Personal Computer (PC) hardware; web-based and accessible from any workstation with a web browser; a variety of reports can be easily generated and emailed or saved, e.g. billing report for Administration department, results report for customers, permit report for monitoring authority, productivity report for lab manager; reports generated in PDF (Portable Data Format) format; data easily accessed with external tools like MS-Excel for customized analysis and reporting; built-in capability for remote support and data backup; and modular design ensures that new functionality can be added in a structured manner.

The Daspoort Laboratory studied the experiences of similar industries (7; 2, 1) to aid in selecting the appropriate option. Chapman (2) followed an in-house development option, Avis (1) bought a commercially available software database

management system, and Van Ginkel (7) followed the custom written strategy. It was found that The National Department of Water Affairs and Forestry's (DWAF) Resource Quality Studies (RQS) had a laboratory complex similar to the Daspoort Laboratory, but on a much larger scale. After evaluating the options and constraints, Daspoort Laboratory Management decided to follow the same strategy as the RQS Laboratory by employing a consultant to custom write the LIMS.

SELECTING A VENDOR

Leitham (6) discussed a list of considerations for employing a consultant or vendor, which might include the following: the long-term viability of the vendor must be considered to meet the clients future needs; vendors must be evaluated based on their ability to advance and keep pace with more rapidly changing needs; vendors will be differentiated by their ability to reliably adapt and modify their products in an era of innovation and rapid change. The vendor should be able to adopt new technology, e.g. .NET (**Microsoft**, Redmond, WA) or Java® platforms (**Sun Microsystems**, Santa Clara, CA); the vendor should be able to explain the business benefits of the targeted new technology and evaluate these new technologies; and there should be a clear business benefit to the customer to use whatever technology is selected.

In the end, the Daspoort Laboratory employed the same consultant used by RQS Laboratory for the development of their LIMS. This consultant has been in business since 1991, is used by several similar industries, and has a good reputation within the industry to advance and keep pace with changing needs. The consultant is an active user of open source software and Java ®, and was therefore able to keep the cost to develop, install and maintain the LIMS for the Daspoort Laboratory low. Delivery was in a very short time frame of three months.

APPLICATION SOFTWARE USED

Open source software played an important part in this project. Open source software (7) is software freely available on the Internet. Anyone may use it, but, more importantly, users are encouraged to improve upon it. By sharing improvements and ideas and pooling resources with thousands of others around the world via the Internet, the open software community is able to create powerful, stable, reliable software at very low cost.

The basic idea behind open source is very simple: When programmers can read, redistribute, and modify the source code for a piece of software, the software evolves. The best-known open source product is the Linux operating system (OS) which can be used as a server, desktop, or embedded system application. Customers continue to expand the role of Linux servers into an ever-increasing array of work environments in both the commercial and technical segments of the market. Some of the other well-known open source software packages include *Apache web server, Sendmail, Perl,* and a variety of *database* systems including MySQL, PostgreSQL, MaxDB, Firebird, Hypersonic SQL, Cloudscape and Ingres.

The Shuttleworth Foundation, established in October 2000 by entrepreneur Mark Shuttleworth, is one of the organizations promoting the use of open source software in South Africa. This software is an enabler for developing countries that gives access to high quality software components that can be enhanced and combined to build highly cost-effective solutions.

The Daspoort Laboratory Management took a strategic decision to use open source components in the proposed solution whenever possible, based on the positive experience by the consultant on other projects and the requirement to keep costs low. They also decided that this would be a web-based application because of these associated advantages: centralized deployment and updating of the application – everything is on the server. This also makes remote support easier; no extra software required on the workstation – only a web browser; most users are familiar with web browser software; and well suited to remote access by customers and remote offices.

The dominant technologies for building scalable and robust web-based applications are Microsoft .NET and Java Enterprise Edition (J2EE). The Microsoft solution was ruled out because it requires a number of commercial Microsoft products (Windows Server, SQL Server) and significantly larger server hardware requirements which would add significant cost to the solution. In addition, backup and disaster recovery options and remote access and control for support purposes would require additional commercial software with the consequent increase in capital expenditure and ongoing annual licensing costs. In order to provide the disaster recovery facilities that are a part of the current solution, all these costs would be doubled. Although the core Java technology is not open source as such, it is freely available for use. In addition, a wide variety of high quality open source Java applications, tools and software libraries are available.

Based on the core criteria of open source and Java, the following key components were selected for the Daspoort Laboratory Information Database (LabInfo) system: Linux OS (server); PostgreSQL Relational Database Management System (RDBMS); Apache Tomcat J2EE server; Struts Web Application Framework; IText Java library for generating PDF documents; and JfreeChart Java library for generating charts

DELVELOPMENT AND IMPLEMENTATION

In the development phase, the laboratory manager tested the LabInfo system on a daily basis and forwarded any bugs that were found to the consultant via an Internet based reporting system. The changes, monitored by both the laboratory manager and the consultant, were kept on the Internet reporting system and could be easily tracked. Changes were made directly on the LabInfo system by the consultant through daily updates via the internet. Refinements or

improvements include simplified manual data entry, improved online-analyzer data layout, and statistical quality control of data within defined limits. The purposes of these limits are to alert the data-capturer when the data entered were out of the specified or calculated range. The upper and lower limits are defined as three times the above or below the average value for a sampling point and specific analyses. These statistical limits are re-calculated each day using the new data captured during the day. However, they can be set manually or automatically daily calculated values. The development and testing phase ran simultaneously, and was completed within three months with most of the development finalized, and the bugs repaired.

Training sessions were conducted with all laboratory personnel and other users on sample registration, data capture, reporting, and data export to MS-Excel.

The LabInfo system has now been running at Daspoort Laboratory for almost five years, and has delivered numerous improvements in the operations of the laboratory, including: simultaneous browser-based access to the system by multiple users. Users log-in and are given access only to the functionality that they need; improved control, approval, and auditing of analysis results; essential reports are now easier and quicker to create and distribute. They are also more accurate, since there is no re-entering of data. Reports are generated in the widely used PDF format. Often-used reports include: billing report; customer results report; laboratory worksheets; effluent quality and effluent compliance report (includes a number of graphs only for online auto analyzers); analyst productivity report ; and WWTW's performance report; selected users access the database via MS-Excel for customized analysis and reporting. This process is much more efficient than before because data is no longer scattered over multiple spreadsheets. Data integrity is ensured as the Excel spreadsheets have read-only capabilities.

The open source Daspoort LID system was delivered for under one hundred and twenty thousand Rand which compares very favorably with alternative commercial solutions that were considered that were in excess of one million Rand.

BACK-UP AND CONTINUOUS SUPPORT

Good disaster planning is essential for any business-critical system such as the Daspoort LabInfo system. In addition to a daily back-up of the database to a writable CD (with a future option of a DVD writer when the volume of stored data justifies it), changes are replicated via the Internet to a standby server housed at the consultant's premises, but owned by the client. In the event of a disaster, the standby server can be deployed to site at short notice.

The consultant runs an Internet web-based job tracking system that enables the Daspoort Laboratory to log any problems or requests to change or optimize the LabInfo. All work is recorded, thus keeping the laboratory manager informed of progress. The Daspoort Laboratory made good use of this system during the development and testing of the LabInfo system, and continues to do so.

A support contract was signed with the consultant, which include database management and provision for future developments. The Daspoort Laboratory advertised in the media on a yearly basis for a consultant to manage and continuously develop the existing laboratory information data base system, using the present key components and existing database.

FUTURE DEVELOPMENTS

As stated previously, the Daspoort LabInfo system was designed as a modular system to facilitate future enhancements. Current developments include the management of equipment and chemicals to establish traceability and linkage to analysis results. Next is instrument integration, i.e. the ability to generate work lists for PC-based instruments and then import the results electronically, eliminating the need for reentry of data with the associated human error.

CONCLUSION

The Daspoort LabInfo system is fulfilling its mandate of effective information management. Data is now securely stored and managed in one place, while being readily available to legitimate users – potentially anywhere in the world via the Internet and a web browser.

The cost to implement the system was notably contained through the prudent use of industry-standard open source software.

A modular design has made it possible to add or replace functionality in a structured manner, which will ensure the longevity of the system.

ACKNOWLEDGEMENTS

The author would like to express his sincere gratitude to the City of Tshwane Metropolitan Municipality for allowing him to report this successful research, planning, development and, implementation of the Daspoort LabInfo system.

REFERENCES

1. AVIS, R. 2003 Personal Communication from Mr R. Avis, Laboratory Manager, Cydna Laboratories, Johannesburg, 18 August 2003.

2. CHAPMAN, A. 2003 Personal Communication from Me A. Chapman, Laboratory Manager, Erlab, Kemptonpark, 18 August 2003.

3. DE BEER, W.H.J. 2005 *Quality and Productivity Management IV: Study Guide.* Tshwane University of Technology

4. EURACHEM, 2002. Guide to Quality in Analytical Chemistry, *An Aid to Accreditation*, [Online]. Available at: <http://www.eurachem.ul.pt/guides/CITAC%20EURACHEM%20GUIDE.pdf.>
Accessed: 17/02/2006

5. LEITHAM, D. 2006 The Future of Information Technology: Selecting a Forward-Looking LIMS Vendor. *American Laboratory*. [Online]. March 2006. Available at: <http://www.isc-newsletter.com/newsletter/mar06/lin-2306/art1.pdf>
Accessed: 24/03/2006

6. OPENSOURCE INITIATIVE, 2006. [Online]. Available at: <http://www.opensource.org>
Accessed: 24/03/2006

7. VAN GINKEL, C. 2003 Personal Communication from Me C. van Ginkel, Specialist Scientist, Resource Quality Studies, Pretoria, 18 August 2003.