Organic Process Research & Development **2004**, *8*, 970–982

Laboratory Information Management Systems for Life Science Applications

Kerstin Thurow,*,† Bernd Göde,† Uwe Dingerdissen,‡ and Norbert Stoll†

celisca, Center for Life Science Automation, F.-Barnewitz-Str. 4, 18119 Rostock, Germany, and Institute for Applied Chemistry, R.-Willstaater-Str. 12, 12489 Berlin, Germany

Abstract:

This paper discusses different aspects of laboratory information management systems in life science applications. Although different LIM systems are available on the market, the state of the art LIMS do not fulfill all requirements regarding flexibity and process connectivity. A fully web oriented LIM system is described in the article. The approach shown in the paper presents a system's model based on modern web oriented software technologies. It includes planning tools, a virtualized lab, and on-line process adaptation as well as an interface for computational data processing and data mining.

Introduction

Today the life sciences are understood as sciences related to life in all their branches and basic conditions which determine existence. Biotechnological applications, chemical process analyses and optimizations, and pharmaceutical and diagnostic applications only reflect a small part of the scope of the life sciences.

The research and business sector of the life sciences are linked to considerable resources and, with them, investments. Here it is not just a matter of expert knowledge but above all high-quality automated systems which, for example, permit experiments with a high throughput. The high laboratory investments require commercial specialization and utilization requirements as defined by 24/7 operation. Laboratory robotics, automated analytical measurement systems, or parallel reactor structures for chemical synthesis and processing determine, for example, the technical resources in the fields of drug development and pharmaceuticals. The productivity of high-throughput screening (HTS) internationally opens up considerable competitive advantages for the life science industry and makes many research projects realizable and affordable within reasonable time periods in the first place. HTS requires increased on-line process interfaces for which crossmanufacturer and inexpensive solutions are urgently being sought.

Increasing experimental throughput results in increasing data throughput and with it a confrontation of the specialized user with enormous, often still un-networked data quantities with a broad range of data types. Complex raw data structures must be compressed through a large number of algorithms which, unfortunately, are often dependent on the instrumentation.

IT solutions are in demand which, among other things, also automate the technical process data integration while paving the way for the cost-effective system integration constantly on the agenda in laboratory practice.[1,2] Under the term *laboratory information management system*, industry-related IT solutions for research, development, service, and production in the fields of chemistry, biology, environmental protection, or medicine are combined. The interdisciplinary object of the life sciences meanwhile characterizes a broad, large class of laboratory information management systems and offers a potential for new general system concepts in laboratory automation before the background of innovative Internet technologies.

Market for Laboratory Information Management Systems (LIMS)

The European market for laboratory software and systems, which comprises laboratory information systems (LIS) and laboratory information management systems (LIMS), has entered a phase of dynamic change. According to a new analysis of the business consulting firm Frost & Sullivan, the turnover volume of the entire market of Euro 172.3 million today is to grow to Euro 216.2 million in 2006.[3] A turnover growth from just under U.S. $125 million in 2001 to more than U.S. $305 million in 2007 is predicted for the U.S. market, which is equal to an annual increase in the turnover rate of 16.1%.[4]

The demand for laboratory information management systems (LIMS) that are used in the laboratory itself mainly comes from the pharmaceuticals sector. In the future the strongest demand is to be expected from the proteomic and genomic sectors. Due to the need of an effective management of growing data quantities, the requirement to replace antiquated operating systems and hardware platforms in existing LIMS installations and to promote networking between the institutions, the market volume is to expand further from a current Euro 73.9 million.

Important vendors of LIMS are Thermo Electron Informatics Inc., Labware Europe Ltd., Innaphase Corporation Inc., A J Blomesystem GmbH, MIPS n.v., Applied Biosytems, Triestram & Partner, and ACT Medisys (Misys) (see Table 1). However, in addition to the proven, familiar

* To whom correspondence should be addressed. Telephone: (+49 381) 54345-570/ Fax: (+49 381) 54345-571. E-mail: Kerstin.Thurow@celisca.de.
† celisca.
‡ Institute for Applied Chemistry.

(1) Branca, M. A. *Life Science Informatics: From Data to Drugs*; Cambridge Healthtech Institute: Newton Upper Falls, 2003.
(2) Munkvold, B. E. *Implementing Collaboration Technologies in Industry*; Springer-Verlag: London, 2003.
(3) *Strategic Analysis of the World Laboratory Information Management Systems (LIMS) Markets*; Frost & Sullivan: 2004.
(4) (a) Stolzenberg, T. *GIT Labor-Fachz.* **2003**, *47* (1), 67−68. (b) Cherrington, C. *GIT Europe* **2004**, *8* (2), 28.

10.1021/op040017s CCC: $27.50
Published on Web 09/17/2004

Table 1. **Vendors and products in the LIMS market**

company	product	remarks	interfaces
ISOFT	INTEGO, Lorenzo	highly integrated health information system for hospitals and laboratories	n.a.
AJ Blomesystem GmbH	blomesystem	individual LIMS development tool, 21 CFR 11 compliant, workflow definition, message system, ASP solution, high flexibility, custom specific adaptation	SAP R/3, Charisma, lab devices, Agilent CDS
Applied Biosystems	SQL*LIMS	Oracle tools for adaptation, 21 CFR 11 compliant, workflow definition, additional modules for clinical development, proteomics and genotyping, raw data archiving	SAP R/3, Baan, JD Edwards, BPICS, Movex, lab devices, standard CDS
Creon Lab Control AG	Q~DIS	dialogue generator, 21 CFR 11 compliant, workflow definition, Pocket-PCs, biometrical identification, additional modules for raw data archiving, project and document management, specification management, modular structure	SAP R/3, Baan, Prism, Navision, Charisma, CDS viar AIA, lab devices
Dialogue EDV Systementwicklung GmbH	diaLIMS	individual LIMS development tools or CSS for Web, ASP solutions, 21 CFR 11 compliant, additional modules for CRM, JAVA-applications	host connection via file, XML, database, lab devices
Innaphase Corporation Inc.	Pharma LIMS Suite	21 CFR 11 compliant, series of specific LIMS applications for pharmaceutical industry (Newton LIMS for batch release, quality control, stability and analytical development; Watson LIMS for bioanalytical work for preclinical and clinical applications; Galileo LIMS for in vitro investigations	
Labware Europe Ltd.	LabWare LIMS	21 CFR 11 compliant, biometrical identification, different languages, workflow definition, LabStation instrument interface server	SAP R/3, Charisma, lab devices, Agilent CDS
MIPS n.v. (Medical Information for Professional Systems)	GLIMS	for clinical labs and hospitals, object oriented, relational database management system, graphical user interface, SOPs	direct communication with analytical instruments and hospital applications
Thermo Electron Informatics Inc.	Thermo LabSystems, SampleManager	dialogue generator, 21 CFR 11 compliant, ASP solution, workflow definition, biometrical identification, additional modules for water management, raw data archive, instrument interface server adaptation through workflow	SAP R/3, AspanTech, OSisoft Atlas and other CDS, lab devices
	Nautilus	configuration and programming, 21 CFR 11 compliant, biometrical identification, additional modules for plate handling, robots, DNA array browser	SAP R/3, Atlas and Agilent CDS, lab devices
Triestam & Partner	Lisa	adaptation through Oracle tools, workflow definition, 21 CFR 11 compliant, communication possibility to process and product control systems	SAP R/3, Baan, PRISM, Navision, PI System, Softmatic, Agilent PE, Waters CDS, AIA, lab devices

systems of suppliers with many years of market presence, more specialized (and less extensive) and at the same time more economical solutions of new suppliers are also available. With the enrichment of the market with the increasing variety and diversity of the LIMS applications offered, the selection of the most optimal system for a specific application case becomes more difficult for the customer. The competitive structure in the LIMS segment is changing. In the LIMS segment smaller suppliers will probably leave the market, and fusions, acquisitions, and new entrants will probably provide for a restructuring of the competitive environment. The growth prospects increase in proportion to the degree that LIMS users make the transition to integrating laboratory systems with corporate software. However, the suppliers must note that this kind of change initially triggers the typical wait-and-see behavior among users that has also put the brakes on growth in the past.

Requirements for Modern Laboratory Information Management Systems

LIMS have been established as industry solutions for years. An LIMS supports a variety of action of project management, process planning, process implementation, and the evaluation and securing of results. LIMSs fulfill absolutely imperative documentation tasks. Quality standards according to ISO 9000 are often first made possible by an LIMS, as are implementations of the so-called good laboratory practice and other industry-specific quality standards. The functionality and scope of the LIMS applications have been continuously expanded over the past several years. LIMSs encompass both process-related procedures and administrative tasks of laboratory operation.[5]

Characteristic of modern LIMSs is the fact that the standard version of the software can be adapted to individual needs to match the user's requirements. The adjustment ability is an important factor in the success of the supplier and simultaneously expands the spectrum of application possibilities for a system. Today LIMSs can be integrated in the existing company processes.[6] The objective is no longer only the passing on of information but instead also its adaptation and compression. With the networking of information, the economic basis of the LIMSs is expanded, and due to the integration ability of an LIMS, a complete visualization of the company can be created.

General and Special Requirements for LIMSs in Life Sciences. Today the generally recognized state of science and research as the starting point for the challenges of relevant project topics in the LIMS sector includes:

1. cross-system and manufacturer LIMS integration standards for device integration in flexible laboratory automation;[7,8]

2. integrated, process-optimized data mining;[4]

3. modern solution concepts of web engineering for development and service of LIMS including new heterogeneous software technologies;[3,7,8]

4. solution concepts for IT-supported quality assurance, in particular for validation processes;[8,9]

5. IT security in LIMSs.[10]

The most important requirement for modern laboratory information management systems includes the option of direct integration of devices and the use of web-based systems for minimizing installation and maintenance costs.

A. Device Interfaces: In a laboratory, stand-alone devices are often used which separately process the individual process steps of a reaction, analysis, etc.. This resulted in nonhomogeneous system landscapes with more or less autonomous automation cells. These isolated solutions show serious weaknesses with regard to high-throughput analyses; complete laboratory automation is not possible.

A new objective is therefore the retrieving and archiving of data within a distributed system at a central point using a uniform platform. With an open, uniform solution, the time and costs involved can be reduced while simultaneously realizing the necessary functionalities in many cases.

The analysis data of the respective laboratory devices should be transferred completely automatically into an LIMS. However, no standardized method for connecting the devices to an LIMS has been specified yet. As in medium-sized labs, laboratory devices from various manufacturers are generally used, and the laboratory operators generally do not accept manufacturer-specific solutions.

With the LIMS applications available on the market, usually only manufacturer-specific interface and device connections based on RS-232 are supported.

B. Web-Based Systems: Many LIMSs already have Internet interfaces with which the customer of a service laboratory can access its LIMSs via the Internet. In this case, the Internet is used as a location for a terminal server, i.e., as an IP network. Existing systems are made available via the Internet. Additionally XML-based communication protocols are used.[11,12]

However, the real challenge is the use of actual web-related technologies.[13,14] First systems on the basis of a pure application service provider (ASP) are already in the pilot phase. ASP takes the function and data distribution for independently operating companies into account so that the LIMS can be provided on the Internet with all required components. All functions of a conventional client-server LIMS are available. Solutions of this kind open up all options to the lab for achieving a considerable competitive advantage through increased efficiency and better customer service.

Laboratory applications already available which are based on the use of web technology include those listed in Table 3.

(5) Breuer, M. *GIT Labor-Fachz.* **2001**, *45* (8), 824−826.
(6) Goossens, L.; Burton, P.; More, J.; Merrill, D. *GIT Lab. J. Europe* **2002**, *6* (1), 16−19.
(7) Wagner, J. *LABO* **2002**, *33* (13), 76−82.
(8) Ernestus, C. *IEE* **2003**, *48* (6), 32−34.
(9) Göde, B. *IP-LIMS/openLIMS − Projektdokumentation*, version 1.10; Universität Rostock: Rostock, Germany, 2003.
(10) Hamoudi, R. *Journal of the Association for Laboratory Automation* **2001**, *6* (5), 57−59.
(11) Lobin, H. Springer-Verlag: Berlin, Heidelberg, 2001.
(12) McIntosh, R. L.; Yau, A. *Journal Association of the Laboratory Association* **2003**, *8* (1), 38−45.
(13) Alonso, G.; Casati, F. et al. *Web Services − Concepts, Architectures and Applications*. Springer-Verlag: Berlin, Heidelberg, New York, 2003.
(14) Dumke, R.; Lother, M. et al. *Web Engineering*; Imprint der Pearson Education Deutschland: 2003.

Table 2. **Vendors and solutions for intrument integration**

company	product	device integration
Applied Biosystems	SQL*LIMS	RS232 communication, uni- and bidirectional instrument interface standard instrument interface for all instruments
IBS	CAQ=QSYS LIMS	realization of instrument interface for transformation of measurement data into unique transfer format
CSB-System AG	QLS/LIMS	RS-232
SHE Informationtechnologie AG	Quantos	interface for data exchange with analytical systems such as titration and GC systems (gas chromatographs), transfer of analytical data via RS 232 and USB interface
iCD.GmbH	LABS/Q	TCP/IP-server with specific DLLs for different devices, XML files with determined format, SQL tables
Labtronics	LIMSLink	transfer of text files to report files import of file into LIMS

Table 3. **LIMS using web technology**

product	company	remarks
IP-LIMS	Institute of Automation Engineering, Rostock University[a]	web-based LIMS
diaLIMS.net	dialogue EDV Systementwicklung GmbH	LIMS with all required components made available on the Internet and system operation ensured via application service providers (ASPs)
NetBrookingLIMS	Birkbeck College, University of London[b]	automation of work flows within genetic engineering labs
SEE-PLC-Animator	IGE+XAO GmbH	software for teleservice of PLC-based controllers
.NET Factory	AIT AG	creation of operating system for machines and systems without programming knowledge using .Net Remoting and XML web services as a communication interface[16]
LabRAT	LECIS	peer-to-peer architecture for laboratory automation using an XML-based communication protocol

[a] Available through celisca. [b] Research solution, not commercially available yet.

In the area of device interfacing, the importance of the Internet is increasing. This begins with the interfacing of simple digital inputs and outputs and ranges up to the integration of entire machines. The objective here is a so-called web-based management supported by an LIMS. The usually local separation between the analysis system and computer access can be overcome in this way by using this approach.

openLIMS: A Web-Based System for Life Science Applications

openLIMS: Product-Oriented Bundling of Modern IT Concepts. The many years of practical experience of LIMSs have shown that LIMSs can only be used as standard solutions in rare cases. Especially in research and in process and product development, laboratory procedures can, in particular with regard to the desired information networking, process the side to be controlled and the data analysis, often only to be very imprecisely defined a priori. This complicates static data modeling of an LIMS and requires approaches with a greater software flexibility and later, very broad and deep configuration options.

LIMSs are designed by formalizing actual laboratory processes. During this formalization of specific business processes, information interfaces to technical laboratory processes are included. This means that the machine or the machine system has to be abstracted as a database substructure. In addition to an exchange of user data, information of the access synchronization and status messages must also be taken into account. As a result, the visualization complexity of an LIMS increases as a general automation tool in all laboratories of the life sciences.

Especially for the needs of highly flexible life science applications in the fields of catalyst screening, drug develop-

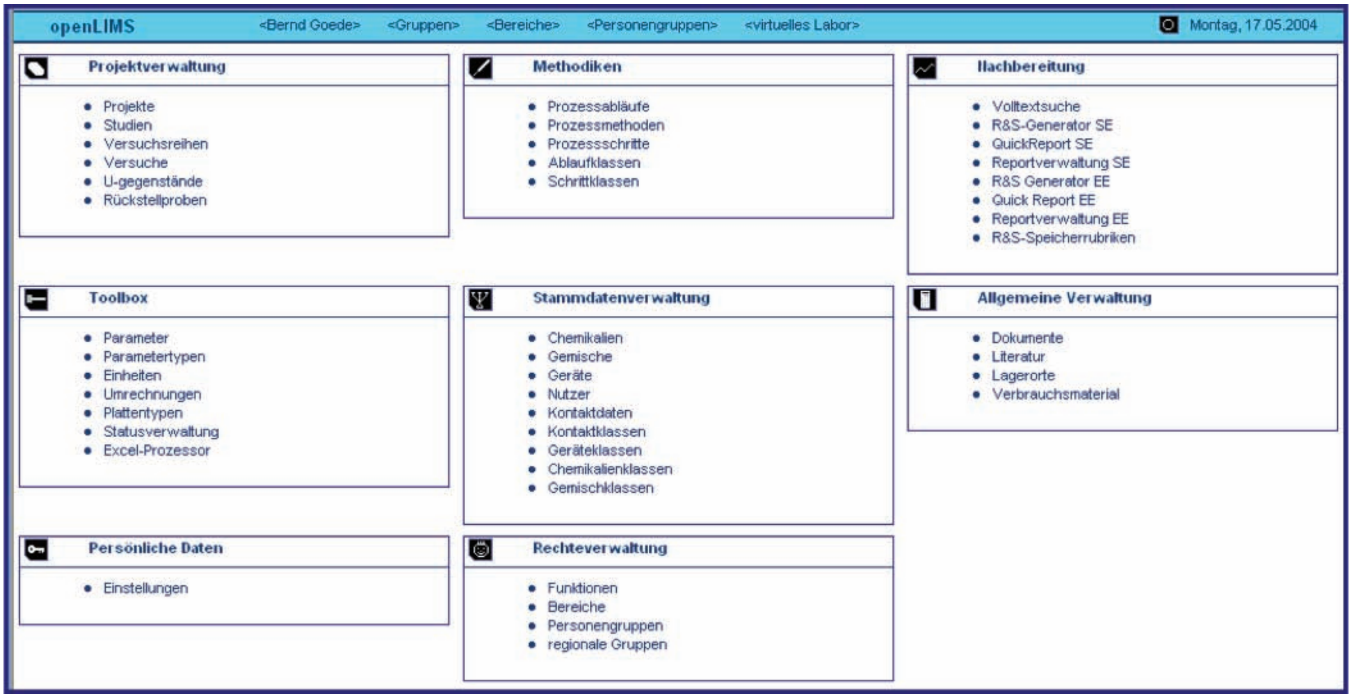

Figure 1. **Main Web Interface openLIMS.**

ment and drug screening, a laboratory information management system, openLIMS (Rostock University), was designed. Due to its open structure, it enables easy adaptation and reconfiguration in accordance with customer-specific requirements.

A major objective of the openLIMS concept is also to create user and technology-adaptive IT infrastructures with a database framework, basic functionalities, specification functions, and integration methods that far exceed the usual range of functions of LIMSs while at the same time opening up modern technologies for distributed systems.

In summary, at least 6 focuses of action result for the openLIMS approach:

1. open procedural hierarchy;
2. open process procedures and process parametrization;
3. open process data compression with higher, familiar technical languages/tools;
4. open data search and reports, as well as export interfaces;
5. integration-friendly software technologies;
6. universal, widespread IT standards and standard IP networks.

The following functional quantity is an example of an openLIMS configuration for chemical and biochemical applications (Figure 1).

1. Project management, structuring, and organization: structure and management of a project hierarchy on four levels; special project characteristics, e.g., test objects.

2. Process planning library: definition of process steps, methods, procedures and relvant classes.

3. Process execution: Dynamic documentation with online process coupling (project management/tests), automated result data calculation/automated result data import.

4. Process analysis: full-text search, search and report system for simple and linked database inquiries, search criteria option according to chemical structures and substructures, automated data export

5. Master data management: substances, devices, users, contacts and their classes, process parameters, IT and import algorithms, structured planning and implementation units, logical device units for process methods and their classes and information relationships, status definitions, authorization classes, personal user settings

The special role of the test object in the project hierarchy results from its relevance for chemical/biological applications. In the process, the test object can be part of the chemical database, can represent a test result, a measuring sample or can be based on any desired document as defined by the free parameter definition. Test objects can be integrated at any point in a process procedure. For other life science automation applications as well, which extend into the medical sector, similar crossproject and emphasized information relationships are also helpful.

Software Engineering for OpenLIMSs. Software engineering encompasses the entire manufacturing process for IT solutions from researching problem analysis to implementation to software maintenance, servicing, and training. Important innovations currently originate from Internet technologies. For modern software engineering, web technologies play an outstanding role. The great dynamics of modern software tools and the diversity, e.g., of server-client technology for web applications and their networking through web services explain the need for limiting a new performance area of web engineering under the roof of software engineering. A summary of relevant topical areas and methodologies is illustrated in ref 7.

The most important subareas of web engineering with regard to the limited application area of LIMSs/life science automation are structured as follows:

1. web protocols;

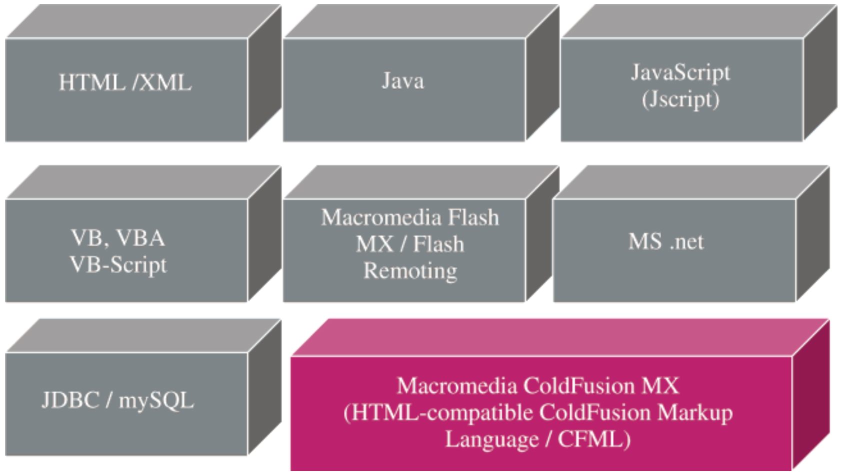

Figure 2. **Software technology for web-based LIMSs.**

2. web-server client architecture/web process allocation;
3. distributed systems and databases;
4. web document standards;
5. web security;
6. web interaction, web design and web software ergonomics;
7. web system validation/quality assurance of web systems.

The complexity and demands of the necessary development competence for the openLIMS system are reflected in combination with general elements of modern software engineering, such as process modeling of heterogeneously distributed operating processes.

Various heterogeneous software technologies are used in the openLIMS system as examples (Figure 2).

Parameter Concept in openLIMS. Most application fields can be derived from a core database model with a focus on an open process parameter principle.

Process parameters initially and directly define any desired process procedures in the lab which are applied to logical planning units (test tubes, MTPs, parallel reactors, robots). Process parameters which, for example, document the flow of materials or reaction conditions in the LIMS have units. Units of the same physical/geometric size are linked via conversion specifications. It must be possible to supplement both in an open system with selected end users. Parameters, units, and conversion functions can only be changed until they are used in the LIMS. Especially when using process parameters in the manipulation-protected test documentation, this restriction is understandable. Access to the parameter definition is tied to special administrator rights and validity checks.

In addition to the so-called free (simple) parameters, the term “table parameters” also exists in the openLIMS. A database object class or a database search is defined in agreement with the terminology of database management systems (DBMS). The importance of table parameters lies in operative information networking by the end user of the LIMS, even if the parameter definitions and the supplementing of the information model belongs in the hands of a selected group of users again. With the use of table parameters, a selection option (select fields, etc.) becomes active during parameter assignment.

The parameter principle takes both laboratory-related procedural processes and general descriptive characteristics of business processes into account which exhibit a time-dependent or time-independent combining of information objects. Enormous LIMS variabilities from chemical research to the telemedical sector can be achieved with relatively few parameter types. Parameter classification can be made as follows:

1. master parameters (unchangeable parameters);
2. simple process parameters;
3. structured process parameters;
4. unique status parameters (process status);
5. simple text parameters;
6. formatable text parameters;
7. parameters of chemical structures and structural relationships;
8. table parameters from the LIMS-DBMS or from networked DBMSs (third-party products);
9. graphic and image parameters;
10. file parameters of compatible file systems.

A few dialogue boxes for information acquisition/selection can be derived from these. Examples include:

1. entry of process parameters;
2. controlled acquisition/calculation of process parameters;
3. graphic editor for chemical structures and reaction equations;
4. text boxes;
5. WYSIWYG editor boxes;
6. indexing dialogue for database object classes in the LIMS;
7. graphic editor for entering hand drawings;
8. selection of files to be indexed.

Process Planning and Execution. Process planning and execution is at the focus of an LIMS. The latter is related to numerous information classes and in particular master data management. Business processes in the openLIMS are illustrated in a multilevel hierarchy. Here designations for hierarchical levels can be adapted to the user’s terminology.

In this example of the realization of LIMSs, a system structure with the following terms is used:

The respective master parameters can also be networked on an information level with any desired number of additional parameters (including the definition field: "Additional data"). Due to this flexibility, virtually all common business structures of R&D or contract labs can be reflected. Figure 3 shows a typical planning process for combinatorial synthesis. In the application development project sector, the LIMS is used both for planning experiments and for their evaluation. The assignment of processors and device parameters in order, for example, to also coordinate complex procedures is also very useful during planning.

Also, a management of resources, e.g., of chemicals or solutions, is possible. The freely plannable data quantities can be stored in databases and then assigned to the planned procedures. They are made available to the test planning as table parameters for this purpose. With regard to the chemicals database, this means that either the entire data stock of chemicals is offered for the parametrization of process steps or only selectively specified categories which can also be specified for specific users. With all subdatabases to be viewed as master data (chemicals, devices, disk types, locations, users, contacts, documents, files, etc.), it is possible to establish any desired number of information relationships between data records (documents with safety information, patents of a substance, literature on reaction, responsibilities of a device, supplier of a substance, location of a chemical, etc.) in accordance with the specific user requirements. Thus in the test planning library test procedures, subprocedures (methods) or elementary process steps can be defined and stored. As a result, the database-related openLIMS already offers highly effective access to information even without search mechanisms.

Virtual Laboratory. The virtual lab is part of the openLIMS, taking into account that an LIMS in the form of a web application itself already (globally) provides major business processes with corresponding virtual business environments. The functional limitation "virtual lab" in the openLIMS means in particular an integration interface for remote functions or standard communication services.

The following applications have been integrated in the virtual lab:

1. live visualizations for real-time process procedures (focus on procedural control of robot routines);

2. monitoring and control access to local or global process control solutions in single-user operation (mobile work, telework);

3. multimedia desktop conferences for telecooperation (project development, result evaluation, planning teamwork, etc.);

4. convenient service links to standard services (web, E-mail, IP remote access);

5. remote control of personal PC workplaces (single-user);

6. links to terminal servers (special data evaluation software, multiuser use possible);

7. selected http sensor/actuator technology for operative expert operation.

In the future, gateway additions through IP telephony and SMS services will be a possibility.

Figure 4 shows a typical example for a robot system in a virtual lab. Besides information regarding operator data, the option for remote control, data conferences, and life visualization are included.

All necessary information for the virtual lab will be maintained in the openLIMS in the master data management. The open parameter model leads to a highly flexible, comprehensive recording of information and information networking.

The service remote desktop, with which teleprocedures with an exact reference to the actual work at PC workplaces and their original working environment (software, interfaces, networking) can be realized, enjoys particularly frequent use. In general, all tasks for device presetting, for method control, for procedural monitoring, for measured value depiction, and for measured value conversion etc. can be realized with desktop remote control. An important condition for using the virtual lab is a separate access authorizations outside the LIMS authorization system.

openLIMS: Virtual Devices and online Process Communication. A conceptional focus of the openLIMS development are open system interfaces and on-line connectivity to the technical lab process, including a high flexibility with regard to planning and results parametrization.[17,18]

The technical lab process is represented by PC device links in the majority of cases for computer controlled processes. The interfacing functions or data include device controllers and the automatic data adoption of discrete information types (result images, chromatograms, spectrums, interface protocols, procedure protocols) etc.. Open structural standards such as XML play an increasingly important role for also lowering the interface expense between technical applications and for economically establishing on-line device environments for an LIMS.[11]

For on-line process integrations in the openLIMS, a principle of "virtual lab devices" was developed. It mainly contains the free visualization of data structures of logical communication and information interfaces. The concept of virtual devices/device systems allows measuring and robot systems to be adapted to specific planned projects at increasingly short intervals. This is possible, as the state of the art in robot system technology (modularity, system openness and configuration effectiveness) has improved considerably over the past several years.

Virtual memory modules as interfaces between the LIMSs and the lab process permit decoupling of the communication computing process to a large extent. This basic principle, known as a communication mailbox for many years now, is given new importance before the background of modern implementation possibilities where cost-efficient control of distributed nonhomogeneous process peripherals is concerned. There are several options for positioning the process

(15) Fllwood, L. *Curr Drug Discov* **2001**, 1 (5/6), 22−24.
(16) Föster, F. W. *GIT Labor-Fachz.* **2003**, 47 (8), 792−794.
(17) Knoff, F. *GIT Labor-Fachz.* **2004**, *48* (1), 46−47.
(18) Churchill, E. F. et al. *Collaborative Virtual Environments*; Springer-Verlag: London, 2001.

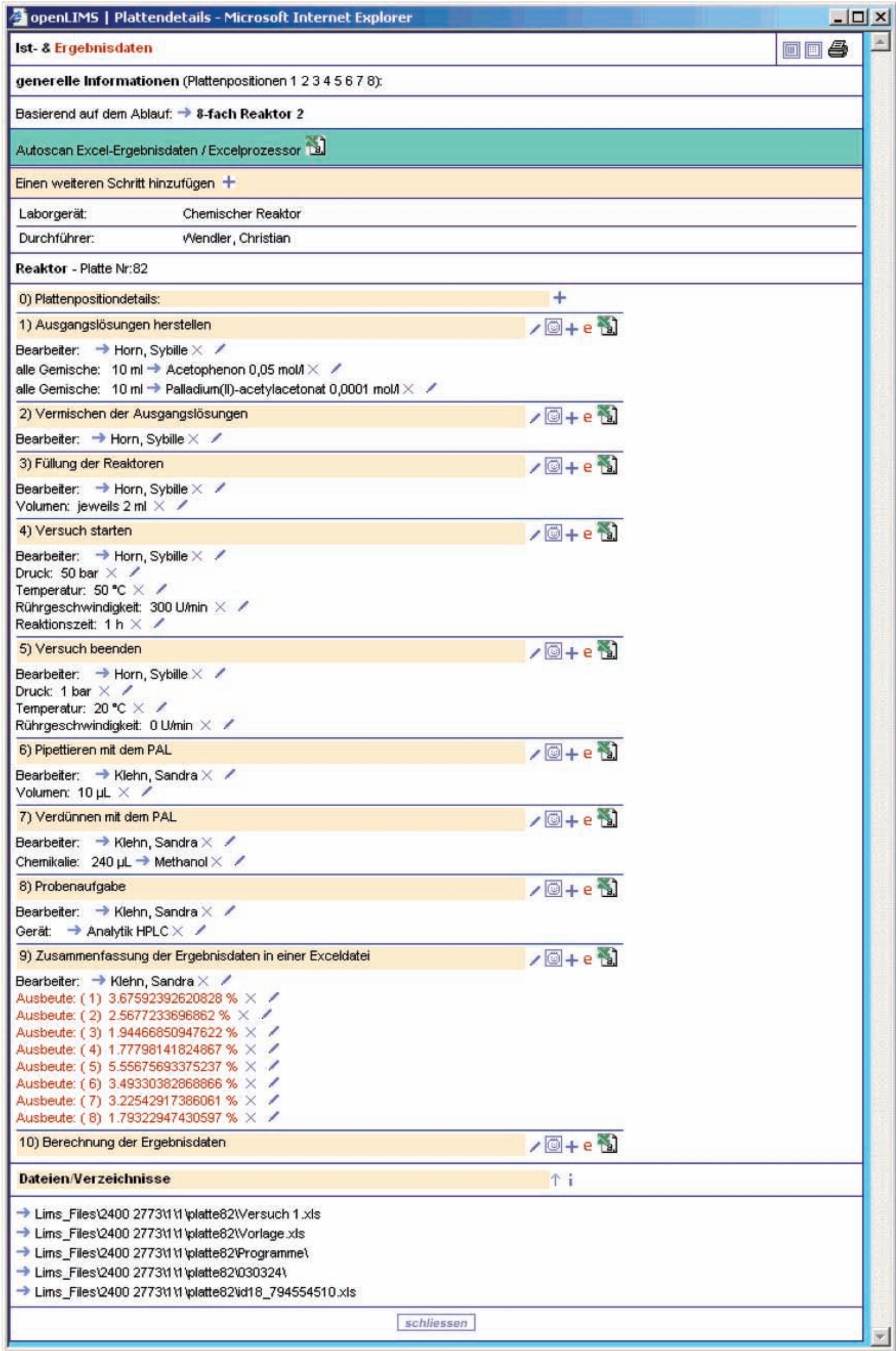

Figure 3. **Insight into a variation of process-accompanying test documentation.**

data mailbox in a networked system consisting of openLIMS applications and DBMS servers, raw data servers, and device/device-system peripherals. This is due to the fact that the free network management of DBMS interfaces (e.g., network ODBC/JDBC) or network file systems with a high reliability and security standard is ensured.

The integration concept of lab processes in the openLIMS system focuses on two methods (see Figure 5):

Figure 4. **Example of device master data in virtual lab.**

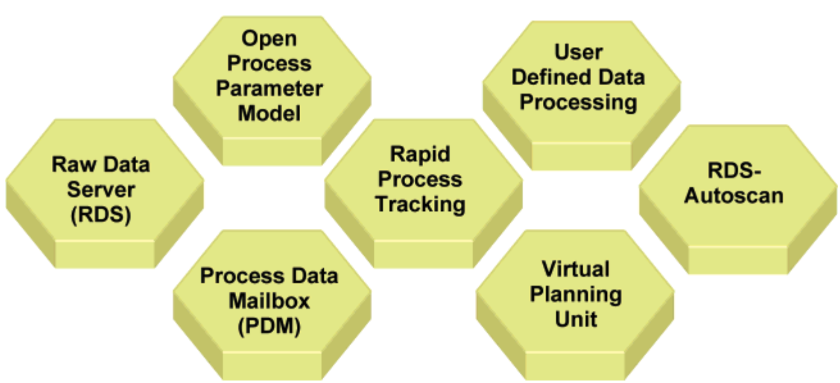

Figure 5. **Concept overview for on-line process interfacing.**

1. on-line process interfacing via raw data servers (file system) and server-controlled data processing or data import of structured and unstructured files;

2. database-supported interprocess communication between web application servers and distributed process databases which can be also be addressed via web services

(XML/SOAP) in addition to ODBC/JDBC.

Raw data are understood to be so-called primary information delivered from (mainly analytical) measuring systems (primary measuring results).

High quality-assurance demands in laboratory practice require the consistent data management, including with regard to raw data, to enable later tracing back to the measurement or recursive data evaluations (15). It can be assumed that many of the above devices supply not only raw data but also preprocessed information (integration characteristic values etc.). The raw data handling process control systems (PCS) are able to calculate secondary data according to a user's method from the raw data. This gives the term raw data server a certain fuzziness, provided it also bundles these data. The decisive criterion for limiting tasks of the RDS is the data indexing or data adoption in the DBMS openLIMS which is not carried out a priori. All information for the RDS is available in file formats. To protect the data integrity, the user of the LIMS is not able to change or delete these raw data.

The selection of a separate raw data server (RDS server) in the openLIMS operating concept was done in order to separate the central mass data server and DBMS server(s). This approach is based on a system optimization through computer and network load decoupling.

Online process interfacing via automated file processing by means of an openLIMS autoscan/Excel processor is a powerful tool for lab automation for the device integration of file-based mailbox communication. For demanding data calculating processes, this basic principle can be applied in a manner similar to those of other technical languages of statistical/visual test evaluation.

Numerous standard devices or robot-controlled lab systems were integrated on the basis of the presented concepts for on-line process interfacing so far. The spectrum of on-line process links in the openLIMS ranges from chemical analysis devices (GC, HPLC, MS) to chemical reaction technology (multiparallel synthesis reactors) to complex robot systems for biological applications.

The concept of database-coupled laboratory process links becomes especially important for HTS experiments using fast measuring/analysis processes (e.g., absorption readers). The live observation of highly parallel robot-supported processes demands effective communication between the process control component and the processing or display processes. In general, there is no clear-cut limit between (local) process control-system tasks and LIMS functionality according to the concept of virtual planning units and logical device units in the LIMS, as soon as process visualization tasks are integrated in the LIMS. The implementation concept of rapid process tracking (RPT) is provided in the openLIMS system for the generalized form of a device or method-specific live visualization of execution processes. It consists of:

1. the live process visualization of HTS process procedure structures on Microtiter plates;

2. an integrated data processing for the acquisition of decision-relevant information from raw data or from status messages of the lab system;

3. an adaptive, user-defined process visualization with detection and decision-supporting mechanisms;

4. the use of modern software technologies for ensuring the platform-independent openLIMS system platform even under the basic conditions of process visualizations with close to real-time speed

Open-Application Data Processing in openLIMS. Besides the generation of data and the storage and archiving of data, data processing is necessary in all applications. In agreement with the concept of automated raw data indexing for process procedures in the openLIMS and open-application data processing module, a so-called Excel processor (see Figure 6) was implemented as one example for a simple, very well-known and accepted data calculation system.

In addition to the LIMS web application (CF/Java **W**eb **A**pplication **S**erver), an automated data adoption and data processing option is created. At least one raw data server (RDS) is used for data connection on the file level to all process computers in the laboratory group and central file storage for further information processing and DBMS adoption in the openLIMS. A local broadband IP connection exists between an RDS and the LIMS server. The expense-minimized operation of raw data servers and the LIMS server on one computer is possible, too.

All Excel documents with definitions for processing instructions (algorithms) and data input and result fields or field ranges transferred to the LIMS define an *Excel processor* (EP). It simultaneously forms a *library* for automated Excel data processing so that new algorithms of the Excel processor can be created on the basis of existing processing modules. Each processing module of the Excel processor (*EP module*) has a *file import definition*, which refers to a source document, e.g., the lab robot technology or chemical analysis (e.g., Excel files with peak data). These source documents are linked to the desired EP module in the process procedure steps, e.g., of test planning or executions, and are therefore indexed in the openLIMS or adopted in the DBMS.

The use of the EP module in the LIMS leads to an automated data compression of laboratory raw data according to any desired algorithms with simultaneous information adoption in the LIMS. The use of the EP module in test procedure steps (evaluation) is also stored in the openLIMS as an indexed Excel file, i.e., with the relevant I/O values.

In HTS experiments such as the biological testing of potential drug candidates on robotic systems, the quantity of raw data generated exceeds the possibilities of Excel. For these purposes technical solutions based on data mining systems able to handle mass data can be connected to the LIMS. Thus, the integration of third party software components for data processing is a possible option. This refers especially to QC/QA and validation processes. The possibility of integrating such software packages depends on their ability to inplement software interfaces modules.

openLIMS Support for Data Mining

A familiar problem of all HTS applications is the enormous amount of data and information. Sometimes the display of those data, i.e., by software modules such as

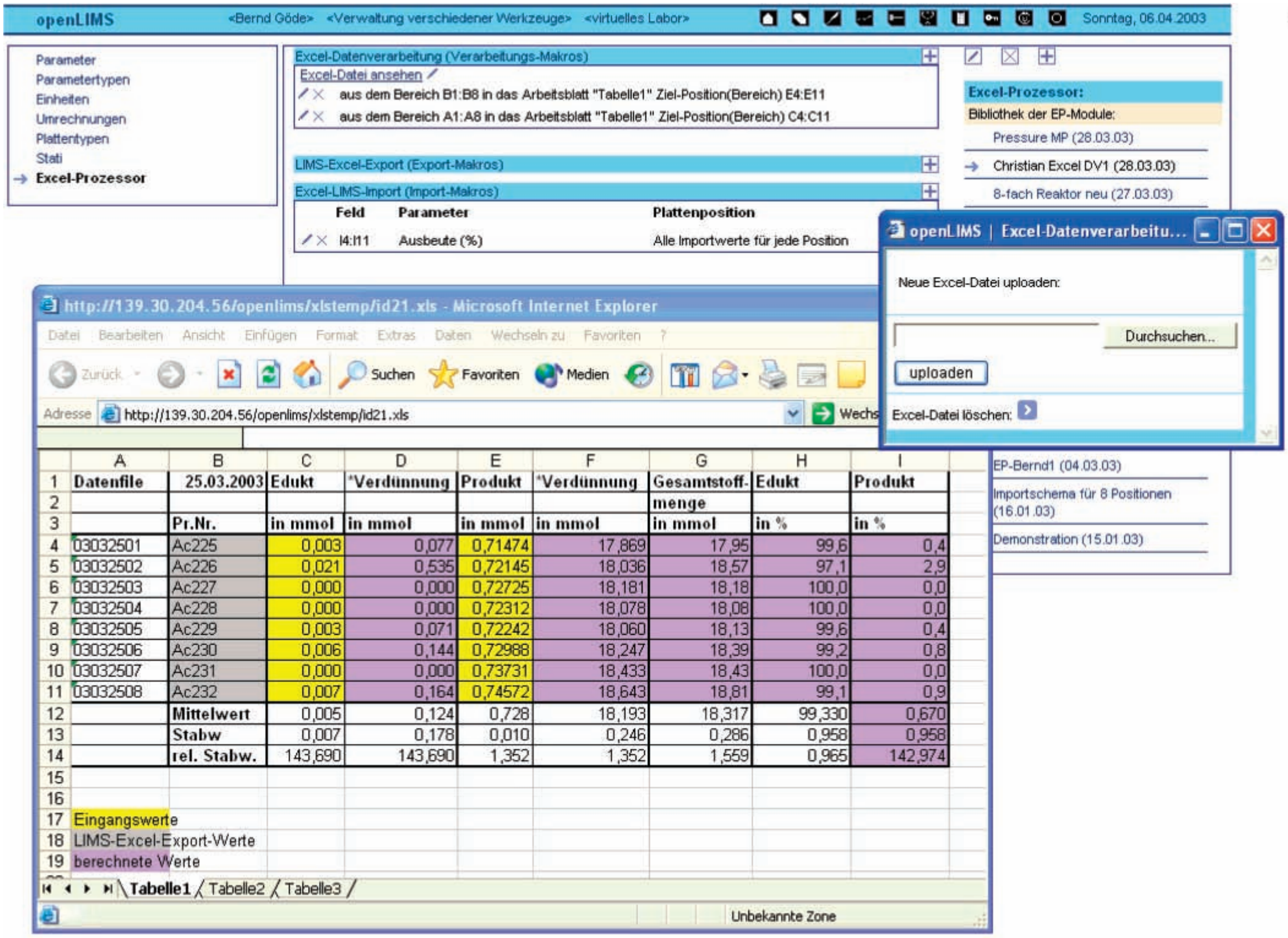

Figure 6. **Excel processor in openLIMS environment.**

SPOTFIRE with a direct access to database arrays is important. The system described in this article has been adapted to this case. The general concept of the system is to provide the user with an open system to integrate additional algorithms. The increasing throughput of experiments with a large number of parallel test executions is followed in the usually exponential context of mass data records. In addition to the planning data, these data records usually contain the actual process variables and in particular the analytical results, usually in the form of structured and unstructured raw data and already partially compressed information. For the further evaluation of HTS experiments, interfaces from the LIMS to the statistical-analytical data evaluation are essential. In the process, part of the selection of relevant test data records and their linking to reference data can be effectively supported by the LIMS. However, this requires that the data export be preceded by open database inquiries with regard to the complete (networked) test documentation.

Today the demanding overall process of iterative information acquisition in complex, extensive database structures is summarized in the term "data mining". Data mining comprises numerous commercial tools, including ones for data selection, for statistical and analytical data analysis, and for graphic visualization of information and information relationships. Data mining is not only aimed at the management and processing of scientific information. On the contrary, it is used for the problem and industry-oriented detection and display of information patterns and information relationships in complex data. To a limited degree, result trends and results can be predicted/estimated, e.g., in order to also initiate new evaluation steps or new experiments.

An LIMS supplies a major part of the starting data material for data mining processes. In the higher processing stages, it is a matter of database-related networking of a wide range of sources, in addition to LIMSs, e.g., databases on chemicals, on compounds, on reactions, or on experiment data from third-party suppliers in a cooperative group.

Simple evaluation processes can already be carried out in the openLIMS by using the Excel processor (EP) module. However, the module clearly distinguishes itself here from the requirements of modern data mining. The main support of data mining lies in the automated calculation of derived process variables, which simplify the interpretation of laboratory procedure documentation in the form of relevant process data records. A data mining tool based on a MATLAB implementation for a Fuzzy based algorithm has been implemented as a tool for processing complex HTS data arrays.

Open Report&Search Generator. The open parameter concept of the openLIMS and the often specialized research

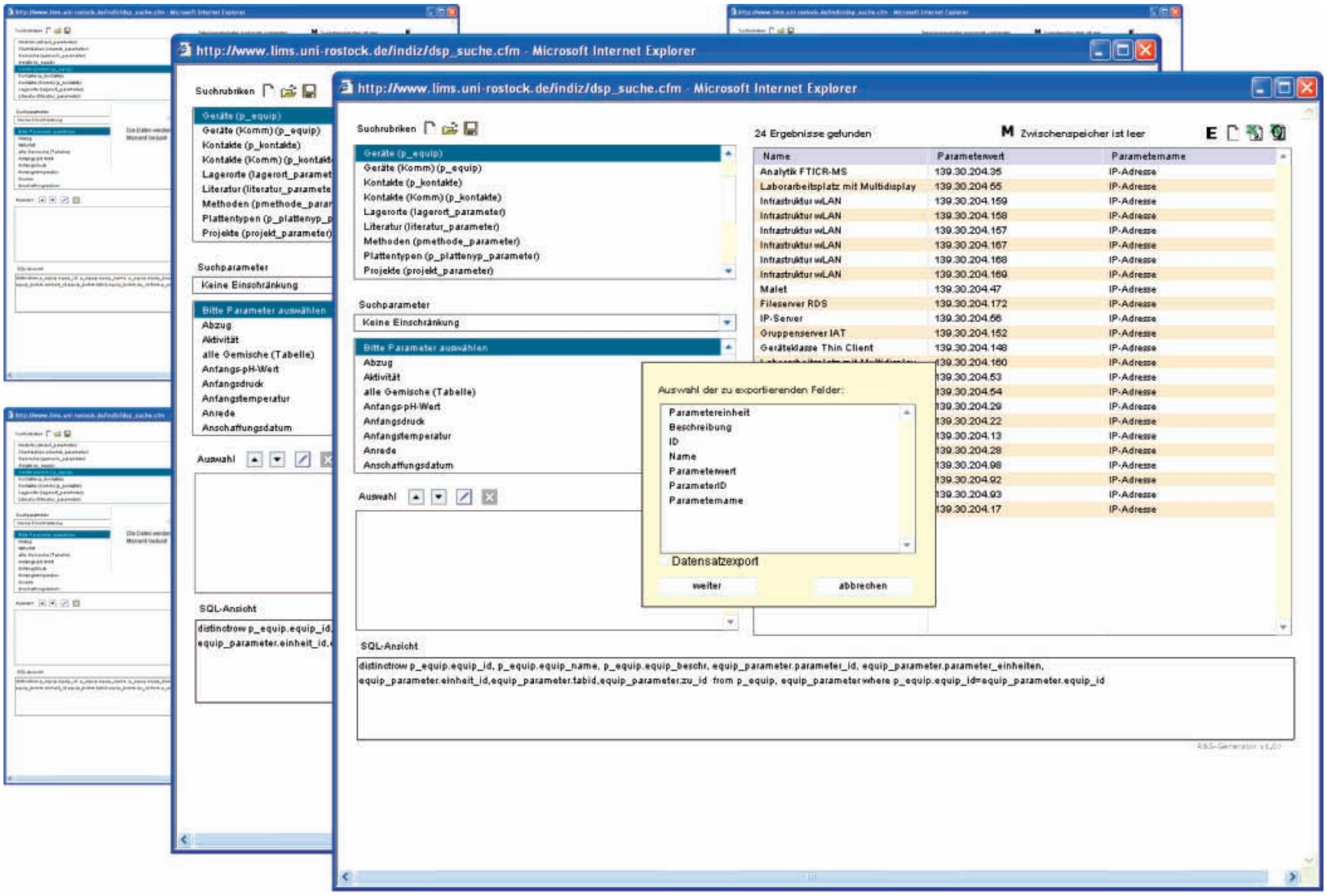

Figure 7. **R&S Desktop.**

requirements which grow with the application also require new solution methods for the implementation of database searches and database reports. Conceptionally, the openLIMS development provides a universal tool for the definition of database inquiries (search, DBMS search) and for the generation of lists (DBMS reports). This tool takes the respective current provision and use of free and table parameters into account. In addition, it also supplies procedures in the form of database inquiries and display schemes, which are then available as the growing menu option "Quick search".

Another important interaction results from the adoption of database inquiries in table parameters, which can be available for information networking with process visualizations.

With regard to terms, the open search and report concept in the openLIMS is defined under the designation Report&Search generator (R&S generator for short).

The R&S generator mainly fulfills three tasks in the openLIMS system:

1. the predefinition of database searches and reports for the user-defined function expansion of the openLIMS software solution;

2. the operative, largely universal information search in the openLIMS;

3. the preparation and carrying out of information exports to other IT applications.

The R&S generator is realized in two versions. The simplified functions in the SE version (Small edition) are based on a limitation of unlinked database searches, i.e., information is sought for an information class (in the technical jargon referred to as an entity or database object class). In the process, both master parameters and assigned free parameters for table parameters can be specified as search criteria. The "and/or" combination of several of these criteria is possible. For many end users, sufficient report results can already be achieved with the R&S generator SE. Figure 7 shows a typical screen for the R&S generator.

The realization of a report and search generator with expanded functional characteristics is the result of increased requirements for data selection of networked search criteria. In particular data selection steps in the iterative process of data mining demand an unpredictable search structure in networked mass data. Experiments with certain actual process parameters, in dependence on certain search chemicals and under the basic condition of the use of selected disk types (e.g., logical reactor devices), are sought. For this class of networked searches, a convenient draft environment with the designation "search desktop" was created. On it the hierarchical search specifications can be developed and saved as a macro. The search desktop in the openLIMS permits the development of all networked researches on database object classes and their relationships to each other which are common for databases.

Secure, Legally Binding Business Processes with LIMSs. The openLIMS system concept supports the use of the digital signature and encryption. To meet the requirements for secure, legally binding business processes in anonymous IP networks, as well as for quality assurance, electronic document storage and electronic signature processes are also becoming increasingly important in laboratory information processing.[19] Here it is important in the LIMS system concept to integrate established security products as seamlessly as possible in the LIMS business process visualizations.

In the openLIMS system the central process-documentation interface is available for the application of identity management software. In the context of B2B communication, for example, all information from planning to the execution to postprocessing of the results is combined in a test documentation and visualized within a short time. Here time information is also provided in the database model, which is relevant for proof of the process procedure. This information carried out as a time stamp can assume a special documentation character which will lead to the use of certified (tamper-proof) time information.

At the heart of secure B2B communication lies the electronic signature on a certain process status of the lab activity, e.g., of a test documentation at the point in time *T* and in addition on the protected exchange of test results. The generalization of the term "test" can in turn be illustrated with the analogies life science experiments, sample analysis for environmental analysis, contract measurement of substance samples or material exploration of a cell treatment, medical diagnosis, etc..

The signature solution of the company Secrypt was used as the basis for an example of the integration of security technologies for telecooperation in the LIMS.[20] The identity management software sold under the name digiSeal provides a broad range of additional functions which were utilized for the openLIMS system:

1. asymmetrical data encryption for the openLIMS test documentation;

2. electronic signature of the test documentation with SmartCard-based certificates;

3. use of certified time stamps;

4. generation of verifiable, machine-readable print documents of the test documentation;

5. indexing of any desired signed documents in the test documentation and in the document management of the openLIMS.

With the electronic, digital signature, several tasks of secure legal and business traffic are carried out in the application segment:

1. proof of the *identity* and the *binding character* of the signatory of a test report (test documentation in the LIMS); this proof of identity is also called authenticity;

2. proof of the document *integrity*, i.e., a test documentation signed once is protected against tampering with the contents;

3. guaranteeing of the *confidentiality* during the communication of test documentation, i.e., protection of the information against viewing by unauthorized third parties.

As a result, there are generally recognized and often more effective electronic analogies to the signed paper document or to the "envelope principle" with the digital signature.

Test documentation can be digitally signed at any desired (recorded) point in time and assigned again to the project hierarchy in the LIMS as an indexed file.

Besides the electronic signature, additional features to, e.g., maintain the integrity of analytical or HTS data have to be implemented into the openLIMS. Audit options are required in order to be compliant to regulations such as DIN or 21 CFR Part 11. The integration of these features is part of the further development for the openLIMS which originally has been designed for combinatorial catalysis and combinatorial synthesis.

Summary

The article describes the background, general requirements, and realization of an LIM system for life science application. The complete open structure of the systems allows an easy adaptation to customer needs. Due to its complete web-based realization, maintenance costs can be lowered. The integration of devices such as analytical system, reaction technology, or robots is one of the main advantages of the system.

Acknowledgment

The authors wish to thank the Ministry of Science of the State Mecklenburg-Vorpommern (Germany) and Degussa AG − Project House Catalysis for supporting the LIMS research. We also thank Ms. Kristina Rimane for market and vendor analysis.

Received for review June 1, 2004.

OP040017S

(19) *Guidance for Industry − 21 CFR Part 11; Electronic Records; Electronic Signatures Validation*; U.S. Department of Health and Human Services, Washington, D. C., 2001.

(20) *secrypt GmbH*; White Paper Identity-Management-Software digiSeal: Berlin, 2003.