Flexible End2End Workflow Automation of Hit-Discovery Research

Silke Holzmüller-Laue¹, Bernd Göde², and Kerstin Thurow¹

Abstract
The article considers a new approach of more complex laboratory automation at the workflow layer. The authors purport the automation of end2end workflows. The combination of all relevant subprocesses—whether automated or manually performed, independently, and in which organizational unit—results in end2end processes that include all result dependencies. The end2end approach focuses on not only the classical experiments in synthesis or screening, but also on auxiliary processes such as the production and storage of chemicals, cell culturing, and maintenance as well as preparatory activities and analyses of experiments. Furthermore, the connection of control flow and data flow in the same process model leads to reducing of effort of the data transfer between the involved systems, including the necessary data transformations. This end2end laboratory automation can be realized effectively with the modern methods of business process management (BPM). This approach is based on a new standardization of the process-modeling notation Business Process Model and Notation 2.0. In drug discovery, several scientific disciplines act together with manifold modern methods, technologies, and a wide range of automated instruments for the discovery and design of target-based drugs. The article discusses the novel BPM-based automation concept with an implemented example of a high-throughput screening of previously synthesized compound libraries.

Keywords
laboratory automation, end2end workflow, systems integration, human task support, Business Process Model and Notation (BPMN)

Introduction
Laboratory processes, like drug discovery, are dominated by high-throughput applications today. In drug discovery, compound libraries with hundreds, thousands, or millions of compounds are tested regarding their biological activity for defined targets. Less than 1% of the potential candidates for further investigations are the results of these tests. The large amounts of samples cannot be handled efficiently without automated workflows. But in reality we have a mixture of automated, semiautomated, as well as manual operations in a workflow from end to end. Respecting the usual terminology, workflow is considered as an operation-oriented business process. Compared to hierarchical factory automation, the structures of processes are distributed horizontally in many dependent activities and subprocesses at the workflow-automation level. These subprocesses go back vertically on the respective established laboratory automation system. On so-called automation islands, complex specific applications run fully automatically as a subprocess. Multiple devices such as liquid handlers, incubators, optical readers, and robotic arms to transfer plates and material are combined in a self-controlled laboratory automation system. Such a subworkflow is usually controlled by a process-control system (PCS) with a respective predefined method. Automation is established to execute the actual experiment. Preparatory and auxiliary subprocesses (storage, maintenance, and method development), analysis, interpretation, and knowledge extraction are time-consuming, often creative, working steps without sufficient integration. Especially in research labs, the automation of the fast-changing processes is not necessarily cost-efficient. Even in a laboratory equipped with modern automation technology, manual work is still required for a large number of routine laboratory tasks. Paper-based guidelines and standard operating procedures (SOPs) usually ensure such workflows with the required information about calibration of instruments, used reagents, and the
detailed workflows. The generated data amount is growing in automated material-generated and analytical processes. The data have to be transformed, managed, processed, interpreted, and evaluated. In each subprocess, different software systems are used for process control (classical PCS), data capture, data management, and data processing [laboratory information management systems (LIMS), content management systems (CMS), scientific data management systems (SDMS), electronic laboratory notebooks (ELNs), common spreadsheet programs, or data analysis systems]. Different software systems lead to a high number of various interfaces for data exchange, which are often handled manually. Thus, data exchange is often reduced to the absolutely essential information. Additional information that can also influence the process quality, such as storage conditions, is not available centrally and has to be requested from another department in the case of queries. Software vendors offer integrated laboratory platforms to improve efficiency and centralize the data for easy analysis and retrieval [1]. Laboratory execution systems (LES) as industry-specific manufacturing execution systems (MES) support guided workflows. But LES do not offer an open control and integration concept for any lab and information technology (IT) systems, or for message handling to involve manual operations. The modeling and control language of LES is mostly an in-house standard of vendors.

The goal of the novel model-based automation concept pursued by the authors is the flexible end2end automation to control and monitor overall workflow processes (Figure 1). The approach integrates all result-relevant subprocesses and activities independent of their different degree of automation. So the target workflow automation does not mean that the whole process is running completely by machines. But the human activities are embedded in a central time control and are supported with comprehensive just-in-time IT support for all IT systems that are used. This results in a guided workflow for the laboratory personnel. The workflow control distributes tasks or service calls to the participants (employees, systems, and services) according to the competencies or functions, sends notifications about the next tasks, and provides all execution-relevant information at the right time. As soon as the resources are available, the activities will be started. The synchronization of manual and automated tasks saves time and improves the throughput. Such end2end processes can be across systems and scientific disciplines or cross-departmental. The pharmaceutical market and in general the life sciences are one of the most regulated sectors in the industry. The regulatory requirements are bound to the comprehensive documentation of all steps, states, and results. Besides the control and coordination of the specific activities, an important task therefore is to involve data flow in the automation. The networking of

Figure 1. Business process management (BPM)-driven automation approach in the context of structured laboratory automation.
all relevant data as well as simplifying of data transport and data transformation by automation allow improving the decision support and maximizing the knowledge extraction from experimental data.

Methods

Business Process Management (BPM) as a New Approach for Integration and Automation

The scientific and business discipline BPM provides concepts, methods, and techniques to support the design, administration, configuration, enactment, and analysis of business processes. Traditional approaches of BPM are useful for controlling and improving business workflows. These techniques fall in the category that is called management BPM. Management BPM focuses on the high-level business processes that describe organization work processes to document, standardize, and optimize. Technical BPM is focused on automating business processes as well as managing the resources and interfaces. Technical BPM processes are more workflow-control and system oriented. They enable such low-level techniques as enterprise application integration (EAI), Web service orchestration, or other process-related techniques that are close to the technology. Nowadays, high-level process automation using BPM technologies is established in insurance and financial services, human resources, customer service, and e-commerce. Technical BPM and management BPM must align. The challenges in life science automation (LSA) are the complexity of processes and the heterogeneous automation environment with a lot of different IT systems. Here, end2end processes combine fully automated and semi-automated, partly hierarchical subprocesses on automation islands and systems as well as a lot of manual activities in a wide range of competencies.

Business Process Model and Notation (BPMN) 2.0 as a New Automation Language

With BPMN exists an established standardized graphical notation for process modeling, which pursues the objective to connect the different perspectives from specialty, business, and automation experts as well as IT departments to one another [2–4]. It is already acknowledged as the de facto standard for business process modeling. The acceptance as an ISO standard in September 2013 (ISO/IEC 19510:2013) will promote the worldwide unification of business process descriptions. A BPMN model in the current version, 2.0 (published in 2011), can be directly executed by a BPMN engine without translation into other workflow-control languages (e.g., Business Process Execution Language - BPEL). This fact was the trigger point for the qualification of BPMN as notation in LSA also with short life cycles of highly specific and interdisciplinary workflows. A business process is regarded as any sequence of activities from an initial state to one or more end states. An executable BPMN process model consists of the graphical process map with the activities (an XML-based BPMN diagram) and the process description with all information needed for the execution: process variables, roles, process participants, decisions, events, dialogues, forms, input and output variables of tasks including their mapping to process variables, and messages as well as interfaces to external systems or the business logic of tasks. The standard definition [2] and a manifold literature [3, 4] document the current version of BPMN. In this article, a few concepts and application properties are highlighted in the context of the application domain. The description of the executable process is separated by lanes and pools, which represent the process participants. Pools distinguish the involved organizational structures; lanes differentiate several departments of an organization or IT systems. The layout is process specific. A pool includes a completely closed process. The activity is the atomic building block for the processes. The process logic connects the activities with events and gateways for structuring the control flow. The process map can be simplified by using subprocesses (hierarchical modeling) and complex activities with “black box” behavior that is not modeled in detail. The last one meets the requirements of the typical situation in the structured laboratory automation, in which complex subprocesses are defined and performed with lab robots or specific IT systems (e.g., process control and data processing). Such “outsourced” subprocesses are represented in the process model by one activity, which implements the function call and the data exchange with the involved external IT system. BPM solutions in laboratory automation focus the process control and consider the data flow. The BPMN provides the modeling of the data flow with elements such as the data object, data storage, and the message flow. In the executable process model, these elements are transformed in process variables as well as input and output parameters of the activities. Data generated in the process can be made available for other participating systems, including required transformations.

Business process management systems (BPMS) are unique and complete software solutions that assist the modeling and execution of models, including task management, process administration, and process monitoring [5]. They represent generic automation systems for workflows. BPMS modules also can be embedded into software applications using application programming interfaces and Web services. Many vendors offer their BPMS as software as a service (SaaS). Other dedicated BP-modeling tools and process engines are available. A comprehensive list (as of July 2013, 76 products) is published by the Object Management Group (OMG) (www.bpmn.org). The OMG is an
international, open-membership, not-for-profit computer industry standards consortium that develops enterprise integration standards for modeling [UML, model-driven architecture, and BPMN] and visual design, execution, and maintenance of software and other processes (www.omg.org).

In the proposed automation concept, BPMS are the central building block in the form of the integration platform based on BPMN 2.0. Figure 1 shows the placement of the end2end workflow controllers (so-called workflow engines) in a typical collaborative automation environment of life science laboratories and the existing interface situation. For implementation of individual executable processes, you will need a BPM Suite and some software development for the integration of external systems. A BPMS will assist the standard-conform definition of interdisciplinary, interrelated business processes, including the data flow and interfaces to external services and third-party applications. The integration of information systems and technologies with one’s own integration facilities [e.g., an enterprise service bus (ESB), EAI, and manufacturing and laboratory execution systems (MES and LES, respectively)] reduces the development effort for interfaces. The process engine is an integral part of the BPMS. The engine orchestrates external services, calls dialogues for the user interactions, and automates the workflow.

**Process Repositories**

In the context of the special demands of research for flexibility, simple adaptability, and extendibility, the modularization and use of model repositories play an important role. Hierarchical modeling and software modularization are not new concepts, but they have here an additional benefit: They allow the scientists to be process composers. The purpose concept focuses on building a library of reusable components as building blocks for constructing process models. Completely implemented and configured subprocesses are managed hierarchies sorted in a process library. Domain experts can combine these building blocks or new individual end2end process models. Elementary automation skills are sufficient to do this successfully and adapt details, extend user forms, and modify the mapping between process variables and subprocess parameters. The laboratory personnel know such graphical modeling from instrument- or process-control systems (analytics, lab robots, etc.), which are standard in that industry. The experiences of the authors confirm that domain experts learn the shapes and symbols of BPMN very fast. After a short training, they can read models and discuss the workflows. The building of executable process models by the domain experts leads to a rapid application development and reduces the development costs significantly compared with the traditional development process, in which the domain experts define the requirements in a first step, and then the IT department implements the executable process. The growing responsibility of the domain experts for the process realization satisfies the demands and expectations of knowledge workers and results in high motivation of employees.

**Requirements for Integration in LSA**

Typical, recurring subprocesses of experimental workflows in the life sciences are, on one hand, manual activities such as the pre- and postprocessing of experiments; the planning of experiments; the preparation of chemicals and labware; the storage of chemicals, including the monitoring of environmental conditions; analysis and interpretation of experimental data; as well as final reporting. On the other hand, typical automated tasks or subprocesses include reaction processes (the syntheses of chemicals, biological tests, or elementary steps like heating, cooling, shaking, and stirring) and the analytical characterization of samples by measurement. These subprocesses can be multiple combined to process networks. Rules, decisions, events, conditions, and data objects can be added. Recursive subprocesses can be generated.

The activities and subprocesses mentioned above can be classified with regard to contents and independent of the application domain:

1. Controlling and monitoring of automation systems controlled by a PCS
2. Transports between automation islands (automated or human tasks)
3. Human task control and monitoring (work lists, notifications, and confirmations)
4. IT-assisted documentation (setting, adding, and querying of documentation and/or master data)
5. Data transfer (data, file, and directory moving), data transformation, data selection and filtering, and post-processing (e.g., normalization of measurements)
6. Data analytics, knowledge extraction, and reporting (also inclusive e-signature)
7. Archiving and data backup
8. Unpredictable ad hoc activities

Manifold existing laboratory IT and automation systems are to be integrated into the workflow control, because the specific systems include the knowhow of the vendors with several technologies (e.g., chromatography). To date, systems integration is a great challenge because of the diversity of interfaces and often the lack of suitable interfaces and standards. BPM-assisted process automation in homogeneous SOA environments can be regarded as solved by technology. But a synchronous access is suitable only for
elementary, automated functional coupling with short execution times of the called service. Many applications of LSA contain time-consuming subprocesses with execution times of several hours or days. In such cases, an asynchronous communication leads to a loose coupling of involved systems. Important strategies for integration are the use of message-oriented middleware (MOM) or an ESB as the basis for asynchronous communication between the process engine and the external systems. If there are no suitable interfaces on automation systems, the development of middleware is considered to realize an automated control flow.

LIMS as Documentation Systems for End2End Workflows

The accompanying documentation is of particular importance in life sciences applications. Plenty of information systems are established in the laboratory. The opportunities and integrated functionalities of such information systems [LIMS, ELN, SDMS, chromatography data systems (CDS), etc.] are merged more and more. Within LIMS, not only structured data can be captured, but also integrated ELN functionalities allow storing of unstructured data (e.g., formatted text, screenshots, photos, and files). In this way, the documentation of process execution, measurements, and results of experiments are combined. LIMSs are the suitable information system not only in service laboratories and quality control but also to support end2end workflow in research. A generic integrated documentation solution, LIMS, as storage (e.g., of data of the end2end workflow) simplifies the system’s architecture significantly. In the following example, an in-house-developed openLIMS is used. Besides typical functionalities for the management of samples, events, methods, and devices, openLIMS provides ELN functionalities for unstructured data, a Chemical Inventory System, as well as data-processing and visualization facilities [6, 7]. openLIMS contains several Web service functions for integration into a BPM application. In general, built-in service interfaces, the increasing integration level, and the growing workflow orientation of the modern LIMS of several vendors provide an easy way to apply BPM methods in the future.

Results

An Example with Selected Implementation Details

The search for medically relevant and biologically active substances is based on the availability of large compound libraries, which are synthesized in house or acquired by external providers. Up to hundreds of thousands of compounds per day are tested for their biological activity against several targets by using high-throughput technology. In primary screening, no or few replicates are used. The identified hits are candidates for further investigations such as the identification of dose–response curves, or cellular analysis in high content screening (HCS). Automation is essential for such high sample throughput. The end2end process model in Figure 2, which was written in BPMN 2.0 notation, shows all involved subprocesses in the first phase of drug discovery. The process path specified in the following with the syntheses of compounds and the biological primary screening with an enzyme-based assay is highlighted in this figure. The chosen bioscreening application is an example for other automated tests against further targets [8–10]. The results can be generalized. Because the chosen assay is enzyme based, the following description disregards the subprocesses of cell culturing and storage. Besides the main subprocesses (the synthesis of potential compounds—here with combinatorial chemistry—and their complex biological testing), required auxiliary processes define the overall workflow. Such additional subprocesses are the cell culturing for cell-based screening assays, the analytical characterization of synthesized compounds by measurement, the use of modeling in the planning of the synthesis,
the storage of chemicals and compound libraries, as well as the documentation of the process instances and the data archiving within LIMS. The process model in Figure 2 contains all these result-relevant activities and subprocesses, documents the interrelations, and correlates them. From the point of view of the integrated process control, several former delimited automation islands are connected by BPMN control instructions and the data flow. The integration of transport jobs and the manual preparation and postprocessing steps (manual or automated) between the automation islands improve the automation level by the central time and state control of the end2end workflow. The human workflow management of BPMS supports the manual auxiliary activities with work lists, roles and user management, systems and preselections, and so on.

The example shows the connection of several automation islands in three alternatives to integrate the process-and instrument-control systems into the central workflow control. The alternatives are distinguished regarding their automation level.

1. The PCS controlling the fully automated chemical synthesis does not support the external access. The integration can be realized only with user tasks. The means are notifications to the user for preparing devices, parameterization, or start and the request of execution confirmations with or without comments. The process-oriented workflow control makes available the required information (methods, parameter, etc.) completely and at the right time regarding the progress of process instances.

2. The instrument control system (CDS) used in the analytical department has a data interface. This enables the import of a sample list and archiving the results in data files.

3. The PCS controlling mobile robots as well as the PCS of the automation system for the biological screening offer functions as Web services. In addition, they communicate with external applications asynchronously via message queuing. One of these applications or communication partners can be the workflow control within the BPMS.

The laboratory information management system that is used, openLIMS, also offers a Web service interface. In this way, much functionality can be automated, such as the creation of new experiment documentation, adding further parameters or several data-processing functions.

The implemented system architecture of the realized example is shown in Figure 3.

**Synthesis of Compound Libraries.** The compound libraries are synthesized by a multicomponent reaction and combinatorial chemistry automatically [11]. Therefore, necessary experimental working steps are the automated synthesis of the compounds, the analysis of the product mixtures, the automated purification, the manufacturing of the dimethyl sulfoxide (DMSO) solutions, as well as validation. The analytical characterization of samples of
synthesized compounds by measurement in the validation phase gives information about the purity and other chemical properties. Both these data and the chemical structure are stored as master data of the chemicals within the LIMS. They are used together with screening measurements later in the data analysis for hit picking. The purity is the decision criterion about whether the compound is included in the library. Two automated subprocesses are involved in this part of the process (Figure 4) that are performed on a PCS-controlled automation system (synthesis) and with a CDS-controlled laboratory device system (chemical analytics).

Planning of Synthesis. The multicomponent reaction generates all combinations of \( n \) representatives of three educt classes. In the planning, the representatives are chosen based on criteria such as diversity, small molecular weight, and pharmaceutical potential (e.g., the so-called Lipinski rule of five). But also cost factors play a role. The combinations are built using the CombiChem module of ChemOffice (CambridgeSoft). The volume of combinations can be reduced further by model-based computational procedures (see below). The selection of promising candidates reduces the synthesis costs. The resulting list will be imported into the LIMS after the validated synthesis. The material supply is assisted by providing the inventory part of LIMS to check the stock. New materials are captured in a separated process of the material management by the appropriate personnel.

Computational Intelligence Supports Planning. By using the modern technologies of chemoinformatics during the planning, the bigger part of candidates can be discarded before the real synthesis. The reduced amount of candidates for synthesis and screening cut the costs significantly. The quantitative structure–activity relationships (QSARs) are often used to create a mathematical relationship between chemical structure and the biological reaction and to predict the prospective biological activity relating to a special target. celisca uses a fuzzy-based approach of computational intelligence implemented as a Web service [12, 13]. Based on these results, the planning will be corrected. The results of the screening experiments are used later to optimize and to train the fuzzy model via a post-subprocess. This feedback is included in the end2end process model (Figure 2).

Preparation of Experiments. Checking of devices, the assembly of the automation system, as well as the parameterization of the devices belong to the preparation steps, which are performed manually by the laboratory staff. The tasks are distributed according to competencies and skills and published in work lists by the BPMS. Selected tasks that are important for quality assurance, such as the provision of starting materials, are documented with a date–time stamp. Frozen material has to be defrosted early enough to ensure a correct dosage. The BPMS send all relevant information (parameters for device preparation or information about the provision of material) to the operators [via work lists, as a PDF document (e.g., SOP), or via e-mail] at the right time.

Fully Automated Synthesis on the PCS-Controlled Automation Island. For the synthesis of the compounds, a fully automated laboratory robot system (Figure 5) was available, which among other things can

- combine chemicals in liquid or solid form in the smallest amounts very precisely,
- handle different reclosable screw-cap vessels as well as crimp-cap vials,
- provide reaction conditions (tempering and mixing) up to 192-fold in parallel,
- carry out sample preparation tasks (dilution, filtration, and derivatization) for the analytical measurements, and
- analyze the reaction mixtures (optional) by gas chromatography–mass spectrometry (GC-MS) or high-performance liquid chromatography (HPLC).

The PCS that is used (in-house development [10]) for controlling the automated workflow on the automation island does not offer external access opportunities. Therefore, we go back to a manual initiation of this subprocess. In the BPMN model, this complex experimental process is represented by two user tasks: first, the notification of the operator to prepare the devices and to start the method, and a second human task with the request for the execution confirmation. Thereby, the workflow control burdens the laboratory personnel from routine documentation work: The workflow control knows the name of the operator of the current human task and can insert a name and time stamp into an information system automatically,
without interaction of the lab assistant and without the possibility for manipulations.

The automation system performs the dosage of chemicals, controls the reaction conditions, and creates the validation samples. The analysis of the reaction solutions and the fractions collected by the purification system as well as the validation of the library elements were carried out with a HPLC–time-of-flight–MS system in the analytical department. A monitoring during the fully automated synthesis is possible using a webcam. The link to the Web application for the monitoring is attached to the notification about starting the PCS method in the work list of the BPMS client.

Analytical Measurement of Reaction Products. The validation samples of reaction products have to be transported to the analytical department by mobile robots or laboratory staff. Here, the purity of the substances will be estimated. The used CDS has only a data interface, not a functional external access. The sample list created during the sampling step can be imported via the data interface. The corresponding operators receive the list as e-mail or with the task request in the work list of the BPMS client. Further parameterizations and the start of the measurement have to be performed manually. The results of the analytical measurement are chromatograms with information about the purity of the substances; 85% was defined as the lower limit in the previous investigations. The final analysis of the raw data using the CDS functions results in a report in an MS Excel document stored in the current report directory of the CDS. This file is added to the LIMS documentation for process recording and archiving. The analysis of this report to extract the positive validated samples is automated by an ESB. This ESB service monitors the report directory of the CDS and start the processing as soon as the current file is available. The ESB lists the validated samples and added the purity as a new parameter of the compounds to the planning data.

Storage of the Compounds and Monitoring of Conditions. The compounds are stored, usually frozen, if they are not processed immediately. The compounds dissolved in DMSO are transformed into 96-well microtiter plates in blocks of 80 compounds. For management of compound libraries, usually special inventory systems or enterprise resource-planning systems are used. The openLIMS includes this functionality. So there are Web service functions to query the inventory or document the withdrawal, which can be applied in the automated screening experiments.

Compound Registration. A list of validated compounds is created in the previous steps. The chemical structure information and further properties of these compounds are imported as chemical master data into the LIMS. The BPM process calls the appropriate Web service of the openLIMS. Parameters are directories, path information, and file names of the compound list and structure file. The master data and inventory are available for querying, experiment planning, or analysis hence.

Material Flow between Subprocesses by Using Mobile Robots. The steps synthesis, analytics, and biological screening are usually executed in separated laboratories or even buildings. The transport of the samples between the automation systems can be executed fully automated by mobile robots if the local and economic conditions allow this (Figure 6). These transport steps are controlled by the BPMS and also based on the end2end process model. In case of a Web service interface of the transport PCS, like in the realized example, the integration of the transport robots into overall process chains is uncomplicated. In our case study, the in-house-developed PCS manages, coordinates, and instructs several robots for local transportation tasks [14]. The Web service call to order the transport job includes all required information such as source and destination, kind of transport object, and the deadline. The first available robot closest to the source takes over the request. Besides the primary automation, the use of mobile robots avoids delays outside of normal working hours, providing that the compounds are processed
immediately on the following automation system. The use of mobile robots to connect complex automation systems results in an overall workflow control with high requirements for flexibility. The standardized BPM approach for the end2end laboratory automation guarantees this demand. Alternatively to the transport PCS, the control of mobile robots with a process-oriented solution can be realized in the form of a separate BPMN model. A separated process map implements the communication protocol. The application and transport process can communicate directly by messages.

Automated Biological Screening of Compound Libraries. The biological screening of compound libraries is realized as a fully automated application on a robot-assisted automation island controlled by the PCS SAMI® (Beckman Coulter, Brea, CA, USA). The system combines the robot ORCA (Optimized Robot for Chemical Analysis; Beckman Coulter) for transport of samples and labware, the liquid handler Biomek 2000, and several peripheral devices such as a shaker, incubator, barcode print and apply, and photon counter. An additional software module extends the PCS with external access via the Web service and asynchronous communication via message queuing (e.g., for status querying). Accomplished manual activities for documentation, preparation (provision of selected compounds and placement of labware on automation system), as well as analysis and interpretation of experimental results complete the automated core process. In an experimental run, up to 80 compounds are tested on four plates. For screening of the complete compound library, \( n \) iterations of process (shown in Figure 7) are required. Due to the characteristics of the system, a throughput of up to 1152 data points per day (three runs per day) is possible.

The user tasks are distributed to the roles of “biologist” and “laboratory personnel,” according to the required competencies. The biologist starts the subprocess with the planning of the current experiment within the LIMS. In this step, the compounds to be tested are defined. The workflow control disburdens the scientists from documentation tasks by automated execution of the template-based creation of the process documentation, the documentation of withdrawal in the material management, or macro-based raw data processing. Further support is the provision of dialogue components, for example of LIMS, for the selection of compounds or notifications via work lists and e-mail about the state or results of the current process instance. Time-critical manual activities are documented with a date–time stamp within LIMS as soon as the execution confirmation is received. Unique data for identification (ID) are required for a correct assignment of screening raw data to the experiment documentation. The LIMS centrally assign IDs used as barcodes of plates. The routing of information by the BPMS (as e-mail to the operators or directly to the services) allows direct digital reuse (e.g., for barcode labeling) and so avoids confusion. The integration of manual preparation steps into central time control also allows the monitoring time-critical process parts with alarm rules. For the execution of the method with the PCS SAMI®, diverse specific steps are necessary. These steps are packaged in a reusable BPMN subprocess: load method, schedule with current parameters, start the method, and polling of state in the message queue. This submodel receives the current parameters to call integrated services by the process engine on runtime. The subprocess is finished with an end state: done, error, or cancelled. This state decides the following process path. Here, as during the synthesis, a monitoring of the experiment on the automation system is possible per webcam, too. After a successful automated screening run, the automated data preprocessing of raw data follows. The
measurement results are imported from the process database into the LIMS automatically without the proposed workflow control. That is a function of the used LIMS. The event-based processing of raw data follows the normalization of data using openLIMS functionalities for data processing. Until now, high manual effort was required for this. The used data-processing macros are predefined within the LIMS. The parameters of the current instance are added by

---

**Figure 7.** Executable Process Model for the subprocess "HTS Screening with the GSK-3β Assay."

**Figure 8.** Challenges and effects of end2end process automation in life science automation.

---

**Table 1.** Common abbreviations and their meanings.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPM</td>
<td>Business process management</td>
</tr>
<tr>
<td>BPMN</td>
<td>Business Process Model and Notation</td>
</tr>
<tr>
<td>BPMS</td>
<td>Business process management suite (or system)</td>
</tr>
<tr>
<td>CDS</td>
<td>Chromatography data system</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide (organic solvent)</td>
</tr>
<tr>
<td>EAI</td>
<td>Enterprise application integration</td>
</tr>
<tr>
<td>ELN</td>
<td>Electronic laboratory notebook</td>
</tr>
<tr>
<td>ESB</td>
<td>Enterprise service bus</td>
</tr>
<tr>
<td>HCS</td>
<td>High content screening</td>
</tr>
<tr>
<td>HTS</td>
<td>High-throughput screening</td>
</tr>
<tr>
<td>LES</td>
<td>Laboratory execution system</td>
</tr>
<tr>
<td>LIMS</td>
<td>Laboratory information management system</td>
</tr>
<tr>
<td>LSA</td>
<td>Life science automation</td>
</tr>
<tr>
<td>MES</td>
<td>Manufacturing execution system</td>
</tr>
<tr>
<td>MOM</td>
<td>Message-oriented middleware</td>
</tr>
<tr>
<td>ORCA</td>
<td>Optimized robot for chemical analysis (Beckman Coulter)</td>
</tr>
<tr>
<td>PCS</td>
<td>Process-control system</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative structure–activity relationship</td>
</tr>
<tr>
<td>SAMI</td>
<td>PCS of lab robot system of Beckman Coulter</td>
</tr>
<tr>
<td>SDMS</td>
<td>Scientific data management system</td>
</tr>
<tr>
<td>SOA</td>
<td>Service-oriented architecture</td>
</tr>
<tr>
<td>SOP</td>
<td>Standard operation procedure</td>
</tr>
</tbody>
</table>
the workflow control. The automation of this working part disburdens the scientists from repetitive monotone tasks and avoids mistakes. The results allow the evaluation of the biological activity of the screened compounds. The scientist receives a notification in his or her work list to continue the creative data analysis. An automated generated e-mail includes information about the experimental results and links to the visualization of results. Both the statistics and the visualization simplify the decision making. The sending of notification with a work list or e-mail is one of the out-of-the-box functionalities of common BPMS.

Summary

The explained example points out how various subprocesses can be connected to the complex overall process with the objective to automate the end2end workflow in laboratories. The example shows furthermore how to get a consistent overall documentation in the context of the end2end workflow (e.g., for a better reproducibility). The cross-organizational collaborative that is characteristic of the described overall workflow is typical for life science applications. The end2end automation of such workflows opens up the potential for easier working conditions and better collaboration and communication between scientists, making drug-discovery research a truly multidisciplinary process.

The example highlights the following advantages of the proposed model-based automation approach:

1. Improving the control range in the experimental execution by the coordination of all involved activities and subprocesses. A higher automation level is achieved on the workflow level.
2. Efficiency improvement by the integration of manual activities into the central time control. The workflow control guides users step-by-step through a specific workflow, ensuring the corresponding SOP (controlled SOPs).
3. Improvement of decision making by real time-provision of data of end2end process instances.
4. Increased adaptability of process definitions by a model-based graphical approach.
5. The increasing of automated data transfer and transformations in the end2end view leads to efficient workflows without manual copying and searching data. Mistakes are avoided.
6. Increased quality and knowledge assurance: automation-assisted documentation assures reproducibility and auditability without additional efforts of the lab personnel (making more with less). Reliable data links are created that allow complex data analysis and maximizing of knowledge extracted from the experimental data. The automation approach for the quality assurance covers all dependencies of results within the end2end workflow optionally across organizational units.

Discussion

The advanced BPM opens a new effective way for high-level laboratory automation. The proposed process-oriented, model-based approach provides a particular flexible and adaptable solution at the focal point of the end2end workflow automation. That results generally in a higher degree of process integration. The process model combines fully automated, semi-automated, and manual activities. The overall workflow automation connects the often still-dominated islands of laboratory robots, workstation, and work benches. The open decision between fully automated, semi-automated, or manual process steps targets the cost efficiency and considers the current situation in the life science laboratories. Thereby, the model-based realization leads to a simplified and fast adaptability and modifiability of integrated workflow automation to changed requirements or laboratory systems.

Compared to specific custom-built solutions or not standardized MES or LES, the BPMN 2.0–based process modeling and control offer an open, standard-based approach for the adaptive workflow automation.

The process modeling with the graphical notation standard BPMN 2.0 meets the demands regarding the transparency of the processes for robust workflow control. Furthermore, it supports intraorganizational and interorganizational knowledge management. In life science laboratories, process transparency ranges across arbitrary detailed processes of the automation hierarchy and across the involved scientific disciplines. BPMN 2.0 assumes the role as potential automation languages for processes in the workflow layer. Although BPMN 2.0 is cross-industrial, the authors confirm its high acceptance in interdisciplinary research teams. A joint language contributes to understanding among domain experts, IT specialists, and also developers of laboratory automation.

BPMS-based realizations provide both an economic systems integration in the hybrid and heterogeneous systems environment as well as time- and state-controlled manual activities. Out-of-the-box features of the BPMS such as human task management and plenty of connectors to databases, e-mail-systems, and manifold other external systems contribute to it. That allows fast modifications to changed integration conditions like the replacement of devices or subsystems.

The increased support of manual activities in complex process chains is one of the important advantages of the introduction of BPM technology in laboratory automation. In common automation, concepts such as manual activities are usually not considered. The integration of manual steps into the central workflow control allows both the
coordination of effectively automated and human tasks and the improvement of quality assurance and quality control objectives. The stability and quality of the workflows are increased by the end2end process flow interaction. The control and the dedicated just-in-time IT support of human tasks benefit the process-oriented workflow automation independent of the automation level involved in laboratory systems. Because the workflow control calls relevant preselected functions, the often complex IT systems will be easier to handle and therefore more reliable. Work lists and automated checked time conditions support the coordination of simultaneous work of parallel workflows. Using mobile devices is a great convenience for improving productivity. The BPMS-based workflow automation takes up the advantages of mobile computing (e.g., for state and activity messaging). The provision of information and data capture will be location independent. Such smooth workflows worked in several labs and clean rooms. The proposed workflow automation leads to the significant relief of strain and stress of the lab personnel by many single effects in simultaneous executed workflows.

The paradigm shift noticed in quality assurance, from information silos in different departments to process-accomplished information flows, can be realized efficiently with the combination of workflow control and integrated information systems for the documentation and storage of process path information, including measurements and results. A complete, detailed, and nonredundant documentation of workflows is especially important for adherence to regulatory compliance, but also for reproducibility and knowhow retention. Besides the documentation, the linking of data and their context information benefits the data analysis and the evaluation. The purposed integration of generic LIMS and workflow control allows more documentation without more effort for the laboratory personnel. With the BPMN-based process models, we have transparent and well-documented processes that are easy to adapt. BPMN is a universal standardized process definition that is independent of vendors or specific products. Based on XML, the process models are transferable from one tool to another tool, so they are still readable in the future. The exact execution by the process engine leads to a unique quality.

Figure 8 summarizes the most important tasks and effects of the BPM-based end2end workflow automation. The meaning of one or another challenge and effect is different from that of the initial situation. The described application of pharmaceutical research focuses on a combination of all advantages that are highlighted in Figure 8.

BPM-driven laboratory automation benefits all application fields of life sciences, like medical, pharmaceutical, chemical, and biological labs. The flexibility of the model-based method is especially important for the research and development processes with relative short life cycles. Other than in manufacturing, in office services, or also in specialized service labs, processes in research and development are not always predictable and standardizable, but are high dynamic. Until now, this fact often prohibited the automation. With the proposed concept already in the development phase of new process models, the graphical process notation results in templates for automation, which reduce the time of development of the automated overall application. Besides the easy adaptability of the graphical process models, research processes benefit from the improved reproducibility of the end2end processes.

The importance and acceptance of BPMN 2.0 to analyze and control any business processes have risen respectfully in recent months. Other methods of business process modeling are increasingly being replaced by BPMN. It can be expected that advanced BPM will have more and more influence on laboratory automation as well as on quality assurance of life science processes in the future.

Acknowledgments
The German Research Association (DFG) is gratefully acknowledged for funding this project (HO 4566/1-1). We also wish to thank the German Federal Ministry of Education and Research (BMBF) and the Ministry of Education, Science and Culture of the State Mecklenburg-Vorpommern for the financial support of celisca.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The German Research Association (DFG) is gratefully acknowledged for funding this project (HO 4566/1-1). We also wish to thank the German Federal Ministry of Education and Research (BMBF) and the Ministry of Education, Science and Culture of the State Mecklenburg-Vorpommern for the financial support of celisca.

References


