

Electronic Laboratory Notebook: The Academic Point of View

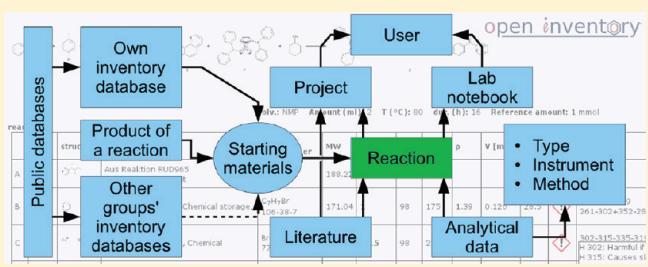
Felix Rudolphi^{*,†,§} and Lukas J. Goossen^{*‡}

[†]IT Department, MPI fuer Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Muelheim an der Ruhr, Germany

[‡]Department of Chemistry, Organic Chemistry, Technische Universität Kaiserslautern, Erwin-Schrödinger-Strasse, Geb. 54, D-67663 Kaiserslautern, Germany

Supporting Information

ABSTRACT: Based on a requirement analysis and alternative design considerations, a platform-independent electronic laboratory notebook (ELN) has been developed that specifically targets academic users. Its intuitive design and numerous productivity features motivate chemical researchers and students to record their data electronically. The data are stored in a highly structured form that offers substantial benefits over laboratory notebooks written on paper with regard to data retrieval, data mining, and exchange of results.



1. INTRODUCTION

In the past decade many companies in the chemical and pharmaceutical industry have introduced electronic means of experiment documentation to increase work productivity and facilitate knowledge exchange.^{1,2} The systems are either (1) text documents with embedded chemical structures, (2) enterprise solutions of specialized software companies,^{3–5} or (3) highly specialized custom-developed programs.⁶ Academic research groups could similarly benefit from electronic laboratory notebooks (ELNs). Earlier ELNs for academic users already allowed collaborative work,⁷ but they, as well as document management systems such as wikis,⁸ lack full integration of chemical structures and substructure search functionality. Until today, the use of ELNs in academia is scarce.⁹ What are the main reasons for this divide?

Text documents with embedded structure drawings do not offer substantial benefits over traditional handwritten laboratory notebooks, so this option is of limited value to group leaders, and time-consuming for the student or researcher. Custom-developed systems as used in companies are unavailable to the public, and the high specialization limits their utility for third parties.

Commercially available enterprise ELN solutions provide extensive functionality and would in principle be attractive also to academic research groups. However, such systems require an enterprise-level database infrastructure, and the overall costs for licenses, installation, and maintenance lie beyond the means of most nonprofit users. As the data storage and processing of closed-source commercial software are not transparent, researchers using such proprietary ELN solutions become dependent on the software supplier when customization, bug-fixes, updates, or data conversion tools are necessary.

What is it that distinguishes research groups at universities from large commercial companies? The differences go beyond size: The organization of academic research is more decentralized

because groups are more independent than in industry, hierarchies inside the groups are flat, and data-sharing between groups is the exception rather than the rule. The IT environments tend to be highly diverse and include various operating and data storage systems. Research groups frequently move to other institutions and, on such occasions, wish to migrate their entire data. These differences are not accounted for by the majority of commercial ELN solutions.

2. HISTORY AND REQUIREMENTS

When starting independent research at the Max-Planck-Institut für Kohlenforschung, Lukas Goossen programmed and implemented a custom ELN built on ISIS/Base,^{5a} especially designed for his research in the field of catalysis and organic method development.¹⁰ Although limited in its features, it already proved to be a valuable tool for the planning and documentation of parallel reactions. Using this basic ELN was very popular among the students, as they considered it less time-consuming than writing laboratory notebooks by hand.

This first-generation, file-based database was soon filled with thousands of reactions along with the corresponding chromatograms and spectra and became an increasingly valuable resource for all group members. However, with the release and distribution of Windows Vista (and its successors), costly updates of the desktop software and a conversion of the existing databases would have become inevitable.

At this stage, we decided to free ourselves from the ties to proprietary software by developing our own ELN software, with the intention of making it available to the public under an open

Special Issue: 2011 Noordwijkerhout Cheminformatics

Received: August 19, 2011

Published: November 14, 2011

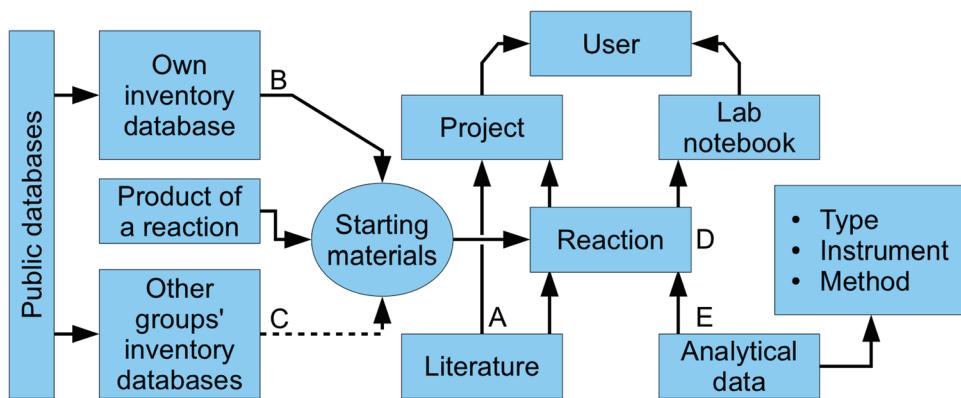


Figure 1. Relationships between database entities.

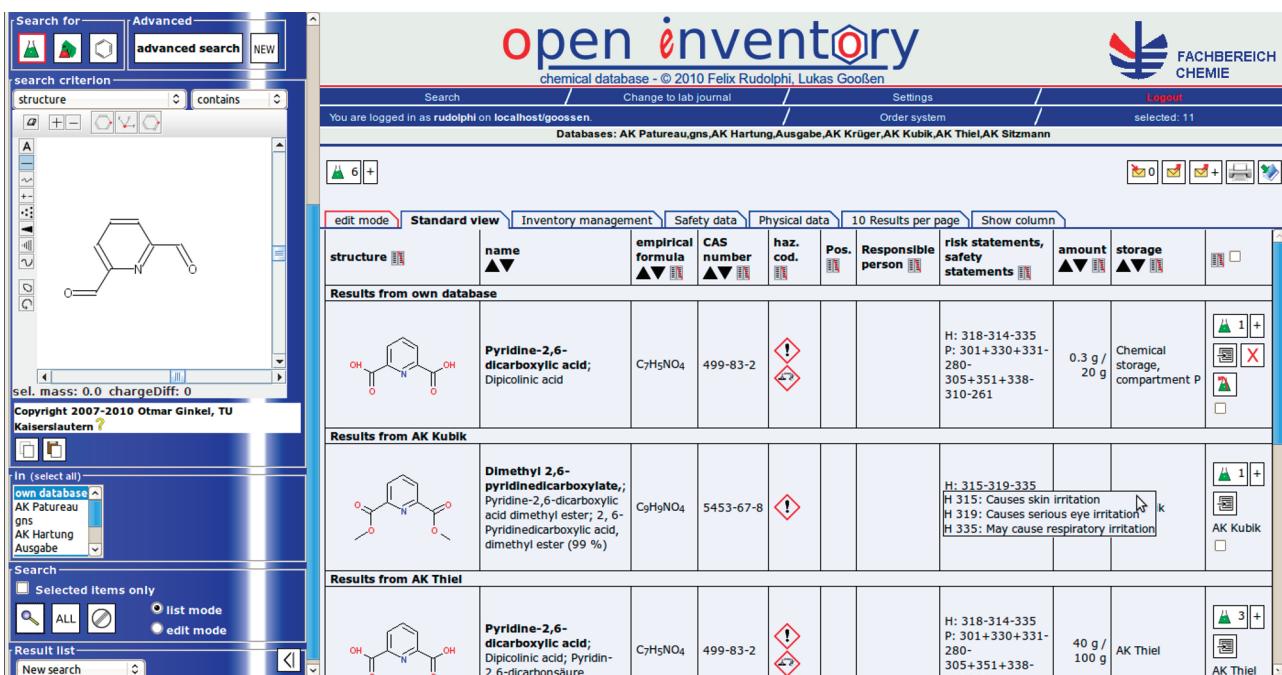


Figure 2. Substructure search in the chemical inventory.

source license. The lessons learned were that (1) an ELN is very useful for academic research, leading to a significant increase in work productivity and convenience, and (2) when seeking for a long-term solution, one must avoid dependencies on commercial software suppliers and on rapidly outdated technology platforms.

3. CONCEPT

When developing our own ELN, we wanted to substantially extend the functionality of the previous system and minimize the dependencies on underlying operating systems and third-party software. We considered two different options to realize our plans: The first was to create a desktop application using Java¹¹ as a platform-independent programming language, and the second was to build a Web-based system where the server generates output using platform-independent programming, e.g. Java Servlets,¹² JSP,¹³ or PHP.¹⁴

A Java desktop application has the advantage of offering direct access to vector graphics to display chemical structure formulas. Furthermore, the possibility to store data locally (which was not possible for Web browsers at that time) can be useful when working offline or with a slow network connection. In order to work offline, one would have to implement advanced synchronization logics to keep the data consistent. However, deploying and updating locally installed Java archives is time-consuming, and Java Swing user interfaces often lack stability and performance compared to native desktop programs or Javascript-driven Web interfaces.

A major advantage of a Web-based solution is that it renders software deployment obsolete and allows centralized updates. HTML in combination with style-sheets and Javascript allow the rapid design of user interfaces. Developers of Web applications can choose between multiple programming languages to generate HTML and Javascript code. Various software development frameworks can be used to create sophisticated applications within a short time. The key advantage of

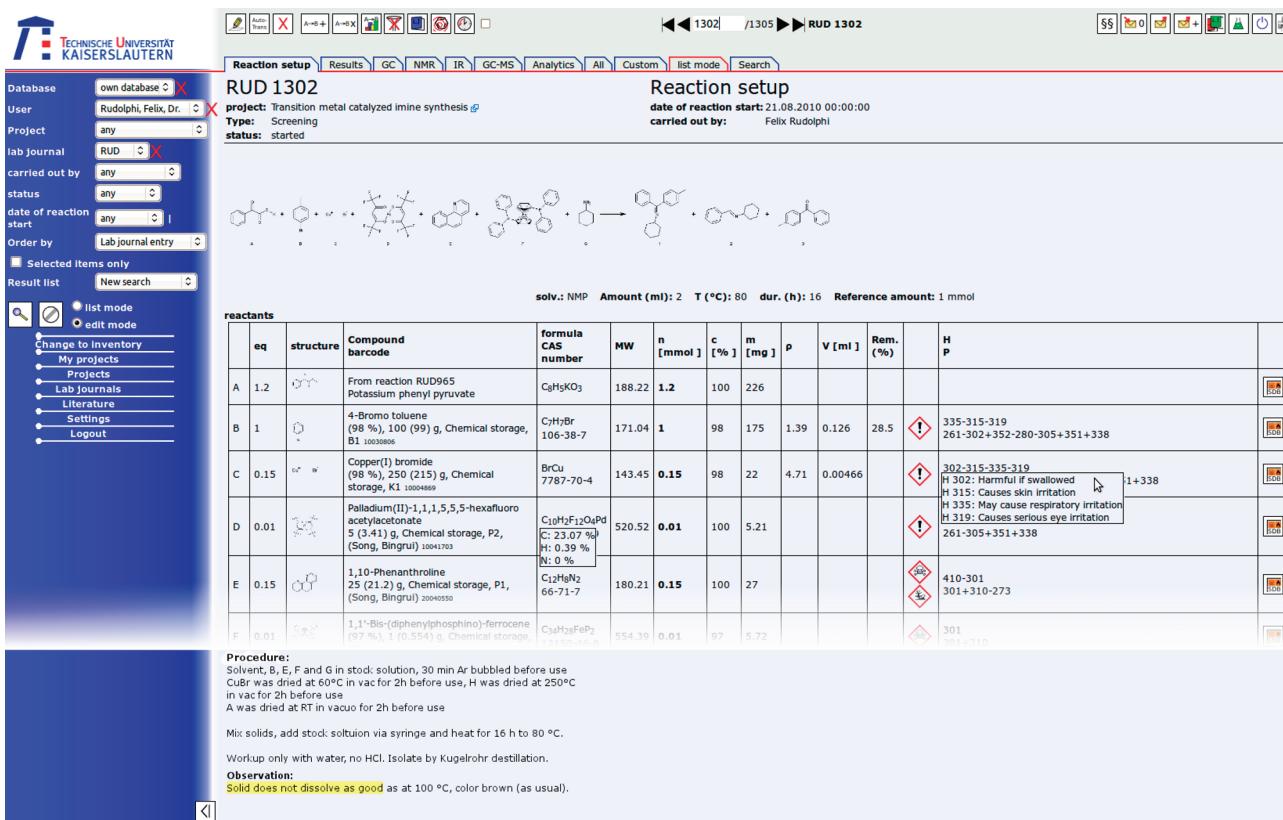


Figure 3. ELN sheet with equation, starting materials, description of setup, and observations.

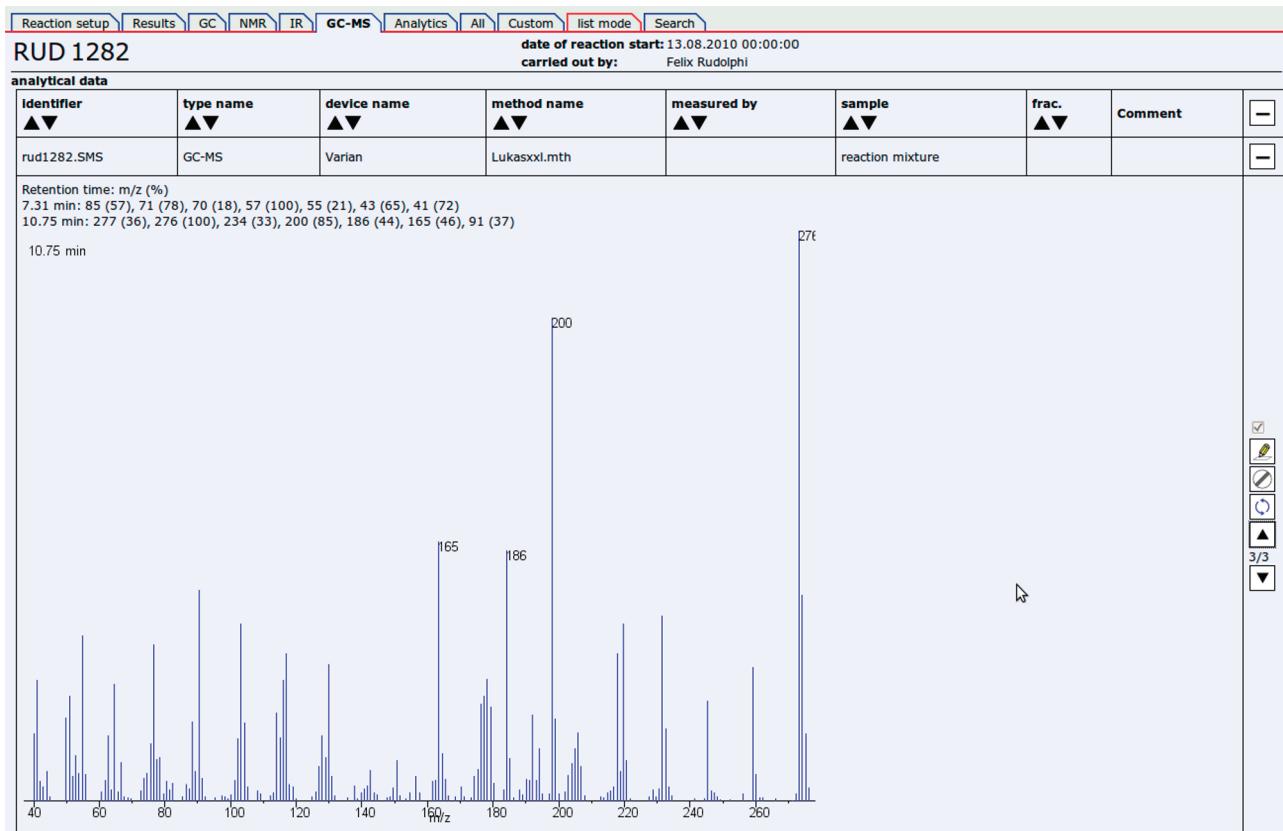


Figure 4. Preview image automatically generated from raw data.

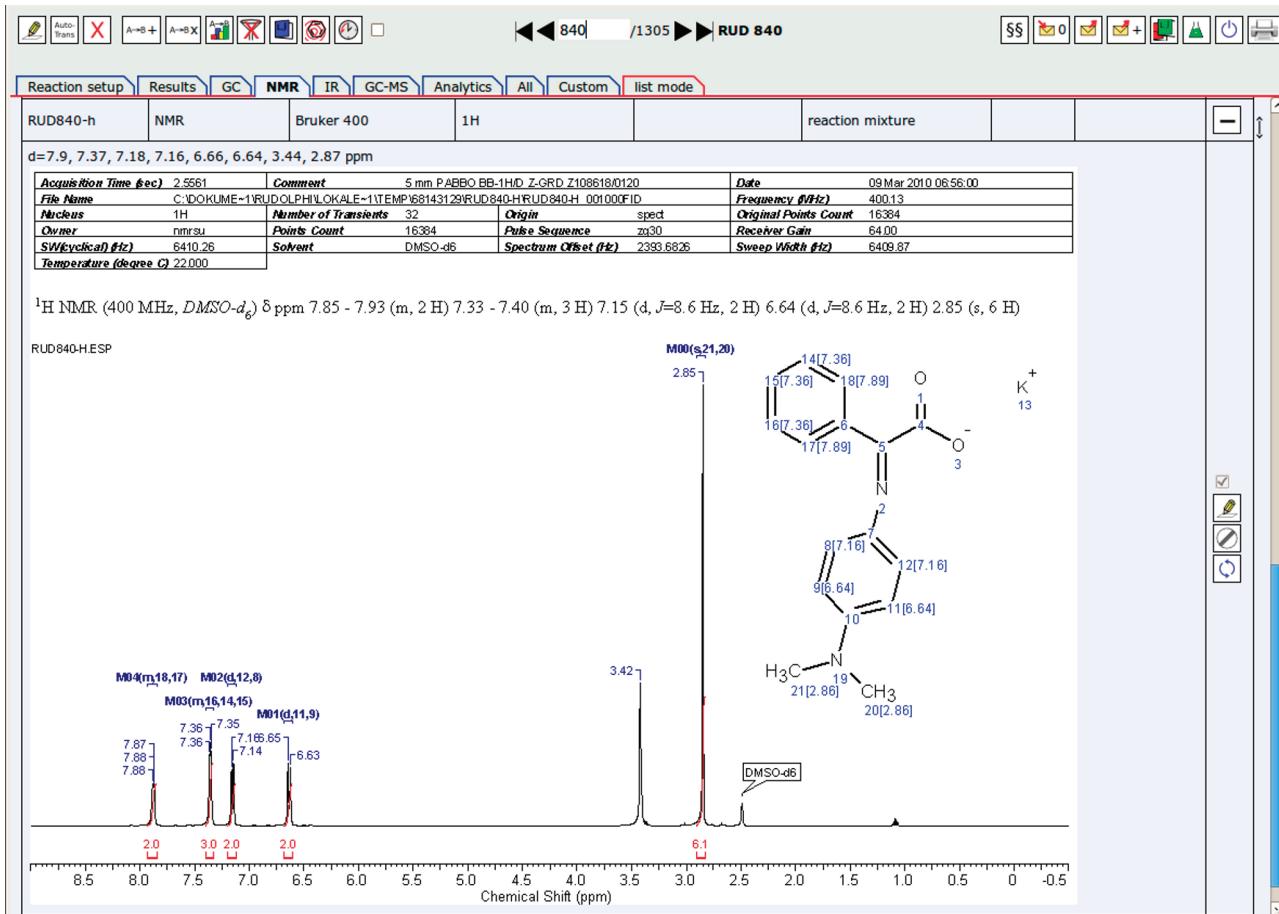


Figure 5. Image of a processed 1H-NMR spectrum.

using PHP as programming language is the availability of preconfigured LAMP/WAMP server software which can be installed without profound IT knowledge.¹⁵ This is ideal for academic research groups who often lack the support of an IT professional when setting up their database servers.

Realizing an ELN as Web application also poses challenges. The support for vector graphics in Internet browsers is still insufficient,¹⁶ so that structure formulas can only be entered using embedded Java applets^{17,18} or Flash¹⁹ and need to be rendered as pixel-graphics on the server. Moreover, the data transfer volume tends to be higher than with a pure Java solution, and offline work using Web applications is still in its infancy.²⁰

The key argument that convinced us to choose a Web-based solution built on PHP/MySQL²¹ was that this programming language is relatively simple and popular among students. A plethora of powerful libraries are freely available to simplify routine programming tasks. Moreover, at that time, Java was still a proprietary software platform, and we sought to minimize our dependency on such software. From today's perspective, considering the triumph of HTML technology and mobile Internet, this decision was a good one.

To deliver a desktop-like user experience and reduce data transfer volume, we planned to make extensive use of asynchronous communication and Javascript programming on the client computer.²²

4. REALIZATION

In January 2007, Felix Rudolphi started the programming of our PHP/MySQL-based ELN system, which we initially released in 2010 under the name *open enventory*.²³ Although the source code of the software is ‘open’, the data stored within *open enventory* is only accessible to authenticated users. To make the data open to a broader audience, we plan –in the future – to add an extension to the software that allows users to transfer selected experiments to a public repository.²⁴

open enventory integrates with and assists the entire chemical research process, thus minimizing media breaks in the information flow. It is licensed under the terms of AGPL v3.²⁵ This license allows anyone to use the software, to modify, and to extend it. Research groups that want to use this ELN can set up their own server and maintain full control over their data. The data storage and processing routines are transparent to all users. The interface is meanwhile available in English, German, Spanish, and French. Users can create localized versions of the software by translating dictionary text files to the desired language.

The software runs on all major Web browsers, e.g. Mozilla Firefox, Google Chrome, or Internet Explorer. A rapidly growing number of research groups at universities and research institutes as well several companies and security authorities have running installations of *open enventory*.

The Workflow. Most research projects start with literature studies. The ELN supports the researcher at this stage by providing a literature management system (Figure 1, A). At the

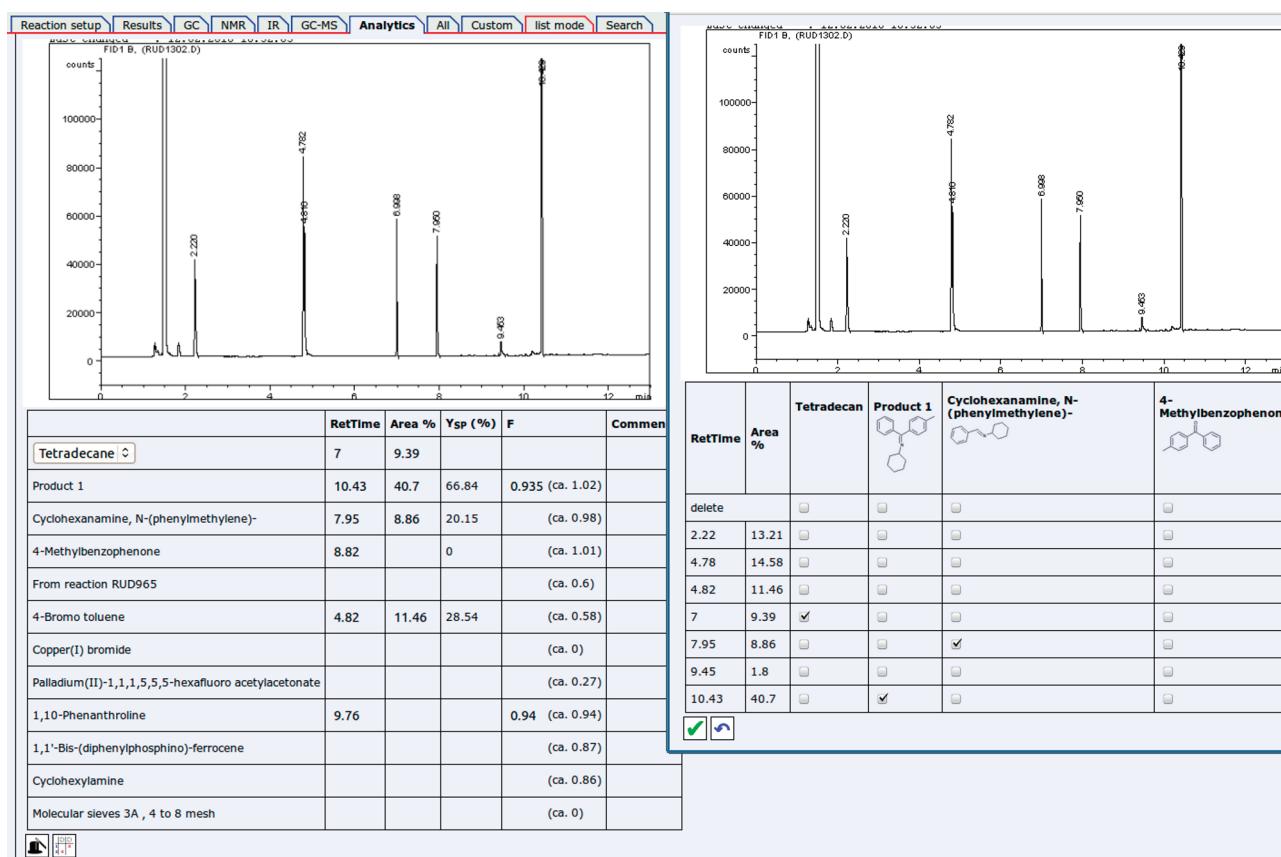


Figure 6. Graphical signal assignment and yield calculation from a chromatogram.

planning stage of chemical reactions, the ELN provides information on physical properties, in-house or commercial availability, and price and location of the chemicals required as starting materials (B/C). The ELN assists the user when planning experiments by automatically calculating molecular masses and filling in required masses and/or volumes for the drawn reaction equation (D).²⁶ When the researcher analyses the results of his experiments, the ELN provides tools for the automatic transfer, processing, and storage of analytical data and helps to design new experiments based on own results and that of research colleagues.

To reflect the independent but interconnected standing of workgroups in academia, we created tools that allow mutually linking otherwise independent databases. It can be exactly specified which type of information is shared within and between research groups. For example, collaborating groups can give each other read access to their chemical inventory, which allows them to share chemicals. Selected users from various research groups can also agree to share the data on all reactions belonging to a specific collaborative research project.

The Chemical Inventory. The chemical inventory is an integral part of the ELN. It assists in locating and ordering chemicals from various commercial suppliers and allows comparing prices. While shopping for chemicals, pluggable screenscraping modules allow extracting chemical structures, physical properties (e.g., molecular formula and weight, density, melting point) and safety data (e.g., risk and safety statements, hazard symbols as well as a link to the material safety data sheet) from the Internet pages of the suppliers and incorporating them into the local database. Once delivered, the new chemicals are

imported into the chemical stock with a few mouse-clicks and can be tagged with barcode labels that allow keeping track of them. When a chemical is later used within a chemical reaction, the fill level of the container is automatically updated. Researchers can search for chemicals e.g. using names, CAS numbers and chemical (sub)structures (see Figure 2). Many intuitive tools maintain chemical databases up-to-date with minimal effort, even when there is no dedicated personnel. This makes it easy to fulfill legal requirements of keeping track of stored chemicals.

Reaction Planning. When entering a new reaction, the ELN recognizes the starting materials from the reaction equation drawn by the user and checks which chemicals are in stock. The system automatically completes the entry, filling in known physical and safety data of the reactants as well as the storage locations and fill-levels of the corresponding containers (Figure 3). We built our own algorithm for the generation of canonical SMILES from structures drawn by the user. It internally disconnects ligands, normalizes mesomeric structures, and converts even ambiguous drawings of organometallic structures into definite SMILES.²⁷ These are used as keys for the identification of chemicals. If, in rare cases, multiple molecules in the database match the drawn reaction component, the user can select the appropriate molecule from a list of candidates.

Once the researcher has finished planning reactions, a printout can be generated to assist the reaction setup. It contains the exact amount of all the required reagents, the storage location of the corresponding containers, and all available safety data.

Recording and Analyzing Experimental Observations. A fundamental part of every lab notebook is the processing and

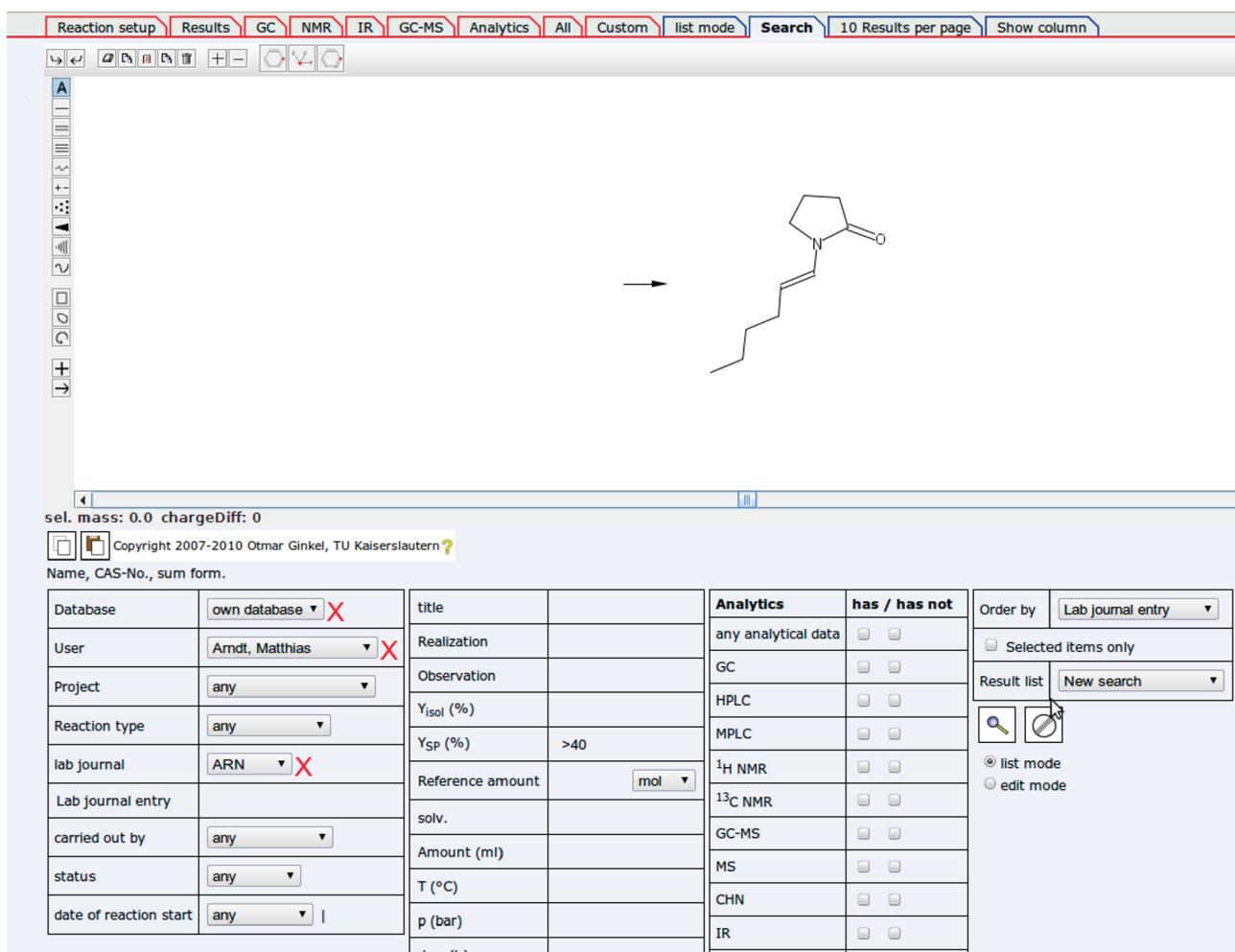


Figure 7. Precise locating of reactions within the ELNs.

storage of analytical data related to experiments. *Open inventory* is equipped with a wealth of tools that allow collecting data from analytical instruments connected via a local network or, alternatively, from local files at the client computer. When the users abide by a systematic naming scheme for their analytical samples and register the data path, file type, and processing software for all of their instruments within the ELN, this data transfer can be made fully automatic. By a simple mouse-click, all relevant new files found in the specified folders are automatically transferred to the appropriate data fields within the database. The software is able to generate and display preview images from files in many standard file formats, e.g. from pictures, PDF or MS-Office documents. It is also capable of extracting information and images from the proprietary file formats used by Bruker, Perkin-Elmer, Agilent, Varian, and other manufacturers of analytical instruments. This way, NMR, GC, GC-MS, and IR spectra can be imported directly from their original storage location and stored unchanged in the database. In addition, the raw data files are processed using a peak-finder algorithm to automatically build correctly formatted peak lists in a form suitable for inclusion in the Supporting Information of publications.

The original data files are zipped and stored in the database and, optionally, also on a file server. This prevents manipulation of the original raw data on the application level. A copy of the analytical data file is made available for processing, and graphical

information is extracted from the duplicate file during the transfer process. The processed data files and images generated from them are stored separately. The images and peak lists are embedded within the Web-page of the corresponding experiment. This allows the users to rapidly browse through their experiments (Figure 4).

The working copy of the analytical data can further be processed using software installed on the client computer. To do this, the user can download the compressed raw data from the database. The file extension of the compressed archive (SPZ) is linked to a locally installed, customizable 'macro'²⁸ which decompresses the archive to a temporary folder, identifies the contents, and starts the appropriate software if available. After the data have been processed by the user, the altered data are compressed and uploaded into the database, replacing the previous file and the old preview image, while a copy of the original raw data is immutably preserved. At a later stage, the processed data can be reopened to continue the processing, also on a different client.

Keeping the ELN separate from specialized data processing software makes the system highly modular and adaptive to the specific analytical instrumentation and software infrastructure at the research institute. This macro-based approach is a simple but highly effective way to establish a direct link between the data analysis software and the ELN, adding the full functionality of

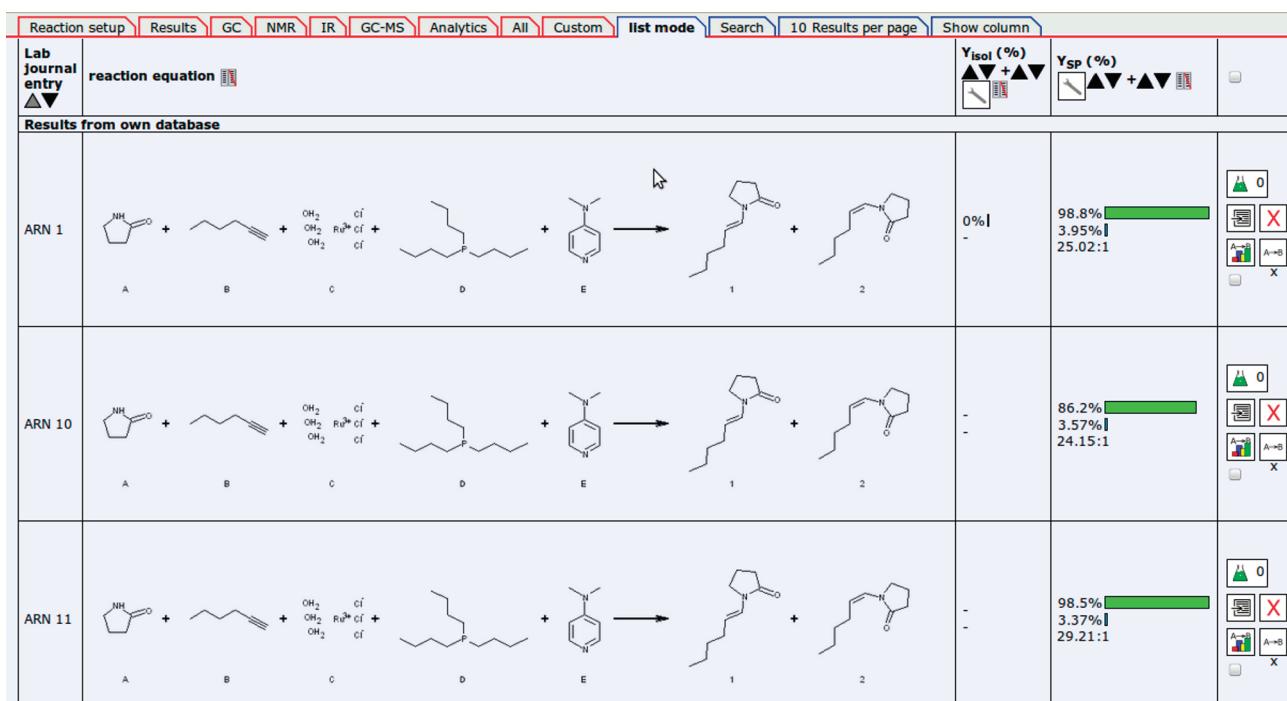


Figure 8. Search results.

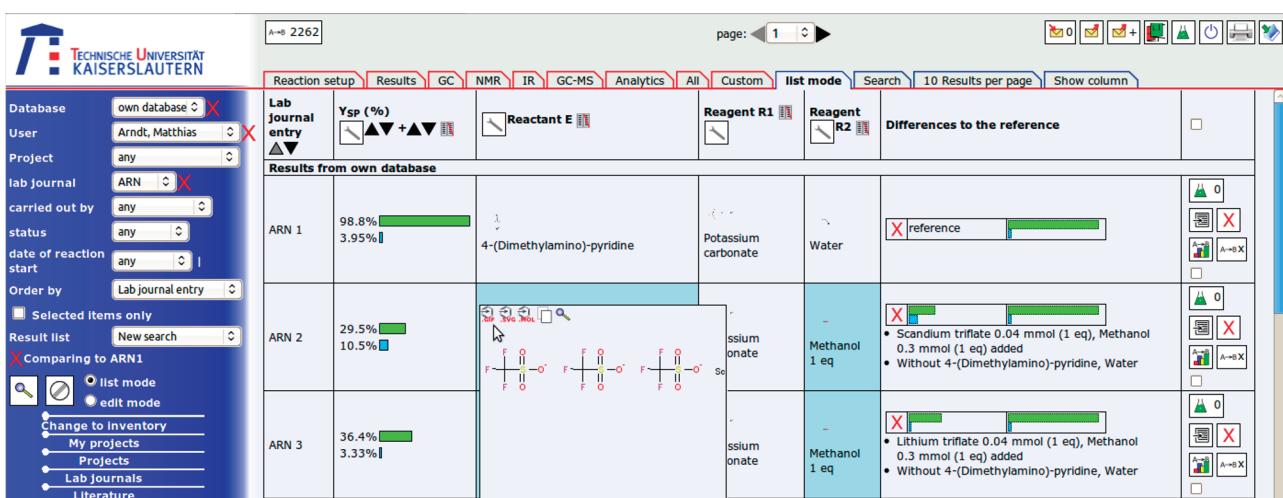


Figure 9. Automatic comparison of similar reactions.

existing data processing software at the research institute to the ELN (Figure 5).

The free access to the source code makes it possible for any user to add personal tools to the ELN. For example, we have created a tool that allows assigning signals in chromatograms to reaction components. The user simply opens the peak list and checks boxes indicating which signal belongs to the internal standard, which to a starting material, and which to a product. Based on this, the program calculates the chromatographic yield of the reaction. In this process, the tool gradually learns the retention times of products, and after a few manual assignments, it is able to automatically calculate chromatographic yields in screening series (Figure 6). Tools such as this dramatically reduce the time required for routine tasks and make the ELN

software very attractive for students. They are the key to convincing students and academic researchers to use an ELN. One has to keep in mind that the use of an ELN can be enforced only in industry. In contrast, students will only decide to use an ELN if it substantially simplifies every single step of their research process.

Data Mining. Keeping a record of all data generated for a single chemical experiment is the first step toward a better understanding of the underlying chemistry. The next step is to compare this data set with that of other experiments and to search for analogies or differences. In this context, an ELN that provides access to large amounts of data sets is an invaluable tool, vastly superior to handwritten journals, provided that it has powerful search and list functions. Such functions are key

Template: RUD 937, Reference amount: 1 mmol		User Rudolphi, Felix, Dr.	Target lab journal RUD	Overwrite last 0 entries	Type Screening	carried out by Dr. Felix Rudolphi											
Global factor	From reaction RUD 937 1.2 mmol 1.2 eq	4-Bromo toluene 1 mmol 1 eq	Copper(I) bromide 0.15 mmol 0.15 eq	Palladium(II)-1,1',1,5,5,5'-hexamethyl acetylacetone 0.005 mmol 0.005 eq	1,10-Phenanthroline 0.075 mmol 0.075 eq	Tri-o-tolyl phosphine 0.01 mmol 0.01 eq	structure	Compound barcode	Magnesium trifluoride 17 mg	[+]	Tetradecane 40 mg	n-dodecane 0.05 ml	[+]	copy procedure		Amount (ml)	T (°C)
1 1	1 1	1 1	1 1	1 1	1 1	1 1			Tri-o-tolyl phosphine 5 (-0.476) g, Fridge, Phosp... 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	2	125
2 1	1 1	1 1	1 1	1 1	1 1	2 2			Tri-o-tolyl phosphine 5 (-0.476) g, Fridge, Phosp... 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	2	125
3 1	1 1	1 1	1 1	1 1	1 1	1 >>			Triphenyl phosphine (2) C 50 (25500000000) g, Chemical storage, P3 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	2	125
4 1	1 1	1 1	1 1	1 1	1 1	1 >>			Tri-(2-furyl) phosphine (1) C 1 (0.983) g, Fridge, Phosphor-Box C 10033971	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	2	125
5 1	1 1	1 1	1 1	1 1	1 1	1 1			Tri-o-tolyl phosphine 5 (-0.476) g, Fridge, Phosp... 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP	2	125
6 1	1 1	1 1	1 1	1 1	1 1	1 1			Tri-o-tolyl phosphine 5 (-0.476) g, Fridge, Phosp... 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	4	125
7 1	1 1	1 1	1 1	1 1	1 1	1 1			Tri-o-tolyl phosphine 5 (-0.476) g, Fridge, Phosp... 10015773	1	1 1	1 1	1 1	<input checked="" type="checkbox"/>	NMP/Quinoline 3:1	2	145

Figure 10. Design of a reaction series, varying additives, and reaction parameters.

components of our ELN. They permit the rapid retrieval of entries in the ELN and list the results in a way that make it easy for the researcher to discover certain patterns. The database structure allows researchers to search not only within their own reactions but also within data sets for which their supervisor or colleagues have granted them access and within data generated by former group members.

Reaction substructures, text blocks, reaction parameters, the availability of spectra, and many other criteria may be used to search for data within the entire database or specified subsets (Figures 7 and 8). For substructure searches (which are not directly supported by the underlying MySQL database), we built an optimized PHP algorithm derived from the popular Match-MOL package²⁹ and implemented a caching layer for the search results on the application level. To compensate the reduced processing speed of the substructure search that results from the use of the interpreted language PHP, we incorporate large fingerprint sets for every molecule in the database.³⁰ Not only do they denote functional groups and rings (based on a fixed dictionary) within a molecule but also characteristic atom-bond sequences leading to hash values. When searching for substructures in the reaction database, a strict similarity search based on the fingerprint sets is performed first. This serves to substantially narrow down the range of compounds for which a substructure search must be performed. When searching in the reaction database, the substructure search algorithm avoids a substantial number of substructure searches due to the fact that many structures occur multiple times – which can be detected by checking their SMILES. Together, these strategies contribute to a significant performance improvement of the substructure search.

In the list view, the program highlights the differences between the experiments and visualizes the effects that these changes have on the reaction outcome (Figure 9). This makes it easy to optimize the reaction conditions and plan follow-up reactions.

Planning Follow-up Reaction Series. In many cases, reaction series designed to optimize reaction conditions contain various experiments that are only slightly modified in comparison to reactions that are already in the database. To simplify the

planning of such reaction series, we included a tool which uses an existing reaction as the blueprint and allows the reader to generate new entries simply by indicating the intended changes (Figure 10). This way, it takes only a few seconds to generate a whole series of systematic parallel experiments.

5. SUMMARY

In conclusion, the open-source software package *open enventory* offers all the advantages of professional ELN software solutions but is designed to specifically meet the demands of scientists in academic institutions. It supports the user at every stage of the research process by providing time-saving tools that simplify routine tasks. The software can be run fully independent of any proprietary software, using an Apache Web server with a MySQL database as platform. The download package contains the PHP scripts, a Java-based structure drawing-applet, and the macro to exchange data between the ELN and the software of analytical instruments.

The ELN has an integrated chemical ordering and inventory system and a literature database. For all experiments, the reaction equations, the experimental procedures, and all spectroscopic data are stored in a well-structured fashion. Search routines allow retrieving experiments based on text fragments, physical data, and chemical substructures. The use of this ELN rather than hand-written notebooks simplifies everyday laboratory work and provides a platform for sharing chemicals, analytical data, and experimental results, and every new entry adds to its value as a knowledge resource. More information on our project and a demo installation is available at <http://www.open-envntory.eu>.

■ ASSOCIATED CONTENT

S Supporting Information. Description of the fingerprints used to accelerate substructure searches. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: goossen@chemie.uni-kl.de.

Present Addresses

[§]Sciformation Consulting GmbH, Froebelstrasse 5, 61118 Bad Vilbel, Germany.

Notes

In memory of Dr. Klaus Angermund.

ACKNOWLEDGMENT

F.R. thanks the FCI and the Studienstiftung des Deutschen Volkes for fellowships. For creating the French version of the software, we thank Dr. Andreas Zumbuehl and France Favarger (University of Geneva) and Dr. Vincent Lemau de Talancé (University of Namur, Belgium). For creating the Spanish version of the software, we thank José Luis Núñez (ICIQ, Tarragona).

REFERENCES

- (1) This work was communicated on a poster presented at the fifth International Conference on Chemical Structures in Noordwijkerhout.
- (2) Achour, Z.; Laidboeur, T.; Gien, O.; Musolino, A.; Bon, X.; Grimaud, B. Sanofi-Synthelabo Chemical Development and the Development of an Electronic Laboratory Notebook. *Org. Process Res. Dev.* **2004**, *8*, 983–997.
- (3) *E-WorkBook Suite*, version 8.2; ID Business Solutions: Guildford, UK, 2010. ID Business Solutions. *E-WorkBook Suite*. <http://www.idbs.com/products-and-services/e-workbook-suite/e-workbook/> (accessed Jul 24, 2011).
- (4) *E-Notebook Enterprise*, version 13.0; PerkinElmer: Cambridge, MA, 2011. PerkinElmer. *E-Notebook Enterprise*. <http://www.cambridgeSoft.com/solutions/details/default.aspx?fid=165&pid=515> (accessed Jul 24, 2011).
- (5) (a) *ISIS/Base*, version 2.4; MDL Information Systems: San Leandro, CA, 2000. (b) *Symyx Notebook*, version 6.6; Accelrys Inc., San Diego, CA, 2011. Symyx Notebook. Accelrys Inc. <http://accelrys.com/micro/notebook> (accessed Jul 24, 2011).
- (6) For a recent overview over available commercial systems, see: Rubacha, M.; Rattan, A. K.; Housselet, S. C. A Review of Electronic Laboratory Notebooks Available in the Market Today. *J. Lab. Autom.* **2011**, *16*, 90–98.
- (7) For other ELNs designed for academic users, see: (a) Virtual Notebook Environment. <http://www.csi.uoregon.edu/vine> (accessed Oct 19, 2011). (b) DOE2000 Electronic Notebook Project. <http://www.csm.ornl.gov/enote> (accessed Oct 19, 2011). (c) Bradley, J.-C.; Samuel, B. SMIRP-A Systems Approach to Laboratory Automation. *J. Lab. Autom.* **2000**, *5*, 48–53.
- (8) Meanwhile, in some areas of research, document management systems or wikis have been adopted to serve as ELNs. For examples, see: (a) <http://onewebservices.wikispaces.com/GoogleAppsScripts> (accessed Oct 19, 2011). (b) <http://sourceforge.net/projects/eln> (accessed Nov 05, 2011).
- (9) An ongoing survey analyzes the extent of ELN usage in academia. See: http://dialamolecule.chem.soton.ac.uk/site/network/ELN_survey.html (accessed Oct 19, 2011).
- (10) (a) Goossen, L.; Döhring, A. Lewis Acids as Highly Efficient Catalysts for the Decarboxylative Esterification of Carboxylic Acids with Dialkyl Dicarbonates. *Adv. Synth. Catal.* **2003**, *345*, 943–947. (b) Goossen, L. J.; Winkel, L.; Döhring, A.; Ghosh, K.; Paetzold, J. Pd-Catalyzed Synthesis of Functionalized Arylketones from Boronic Acids and Carboxylic Acids Activated in situ with Dimethyl Dicarbonate. *Synlett* **2002**, *8*, 1237–1240. (c) Goossen, L. J.; Goossen, K.; Rodríguez, N.; Blanchot, M.; Linder, C.; Zimmermann, B. New catalytic transformations of carboxylic acids. *Pure Appl. Chem.* **2008**, *80*, 1725–1733. (d) Goossen, L. J.; Linder, C.; Rodríguez, N.; Lange, P. P. Biaryl and Aryl Ketone Synthesis via Pd-Catalyzed Decarboxylative Coupling of Carboxylate Salts with Aryl Triflates. *Chem.—Eur. J.* **2009**, *15*, 9336–9349. (e) Goossen, L. J.; Rudolphi, F.; Oppel, C.; Rodríguez, N. Synthesis of Ketones from α -Oxocarboxylates and Aryl Bromides by Cu/Pd-Catalyzed Decarboxylative Cross-Coupling. *Angew. Chem., Int. Ed.* **2008**, *47*, 3043–3045.
- (11) Java Desktop. <http://community.java.net/javadesktop> (accessed Jul 27, 2011).
- (12) Java Servlet Technology. <http://www.oracle.com/technetwork/java/javaee/servlet/index.html> (accessed Jul 27, 2011).
- (13) Java Server Pages. <http://www.oracle.com/technetwork/java/javaee/jsp/index.html> (accessed Jul 27, 2011).
- (14) PHP Hypertext Preprocessor, version 5.3.8; The PHP Group, 2011. PHP Hypertext Preprocessor. The PHP Group. <http://php.net/> (accessed Oct 20, 2011).
- (15) The package Xampp is particularly easy to install, but misconfiguration may pose security threats. Xampp, version 1.7.7; Apache Friends: Berlin, D, 2011. *Xampp. Apache Friends*. <http://www.apachefriends.org/en/xampp.html> (accessed Jul 27, 2011).
- (16) Scalable Vector Graphics. <http://www.w3.org/Graphics/SVG/> (accessed Jul 27, 2011).
- (17) *JME Molecular Editor*, version 2011.01; Peter Ertl, Novartis: Basel, CH, 2011. *JME Molecular Editor* Peter Ertl, Novartis. <http://www.molinspiration.com/jme/> (accessed Jul 27, 2011).
- (18) *MarvinSketch*, version 5.6.0.2; ChemAxon: Budapest, H, 2011. MarvinSketch. ChemAxon. <http://www.chemaxon.com/products/marvin/marvinsketch/> (accessed Jul 27, 2011).
- (19) Dallakian, P.; Haider, N. FlaME: Flash Molecular Editor - a 2D structure input tool for the web. *J. Cheminform.* **2011**, *3*, 6, DOI: 10.1186/1758-2946-3-6.
- (20) HTML 5, which is a W3C standard, supports local data storage, but the current implementation status is not yet sufficient. <http://www.w3.org/TR/2008/NOTE-offline-webapps-20080530/> (accessed Jul 27, 2011).
- (21) MySQL, version 5.5.16; Oracle: Redwood Shores, CA, 2011. MySQL. Oracle. <http://www.mysql.com> (accessed Jul 27, 2011).
- (22) ECMA international, ECMAScript language specification, version 5.1. <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-262.pdf>, 202–203 (accessed Jul 27, 2011).
- (23) (a) open inventory. <http://www.open-enventory.de> (accessed Jul 27, 2011). (b) The latest stable version of the software is available at <http://sourceforge.net/projects/enventory>. A text file within the downloadable archive describes the installation procedure.
- (24) For projects where researchers share their laboratory notebooks with the public, see: (a) <http://openwetware.org> (accessed Oct 19, 2011). (b) <http://usefulchem.wikispaces.com> (accessed Nov 05, 2011).
- (25) Afferro GNU Public License. <http://www.gnu.org/licenses/agpl.html> (accessed Jul 24, 2011).
- (26) (a) Rudolphi, F.; Goossen, L. J. Ein elektronisches Laborjournal als Open-Source-Software. *Nachr. Chem.* **2010**, *58*, 548–550. (b) Rudolphi, F.; Goossen, L. J.; Stesycki, M. Leistungsstark und durchdacht. *Chemie Report* **2010**, *5*, 16–17. (c) Rudolphi, F.; Goossen, L. J. Open-Source-Programm: Ein Elektronisches Laborjournal. *Chem. Unserer Zeit* **2011**, *45*, 56–57.
- (27) (a) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. (b) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.
- (28) The term ‘macro’ refers to a small script written in the language Autohotkey (for Windows), see: <http://www.autohotkey.com> (accessed Oct 19, 2011) or Tcl (for Linux) <http://www.tcl.tk> (accessed Oct 19, 2011).
- (29) Haider, N. Functionality Pattern Matching as an Efficient Complementary Structure/Reaction Search Tool: an Open-Source Approach. *Molecules* **2010**, *15*, 5079–5092.
- (30) A detailed specification of the fingerprints used can be found in the Supporting Information of this article.