JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

ChemTrove: Enabling a Generic ELN To Support Chemistry through the Use of Transferable Plug-ins and Online Data Sources

Aileen E. Day,*[,†] Simon J. Coles,[‡] Colin L. Bird,[‡] Jeremy G. Frey,[‡] Richard J. Whitby,[‡] Valery E. Tkachenko,[§] and Antony J. Williams[§]

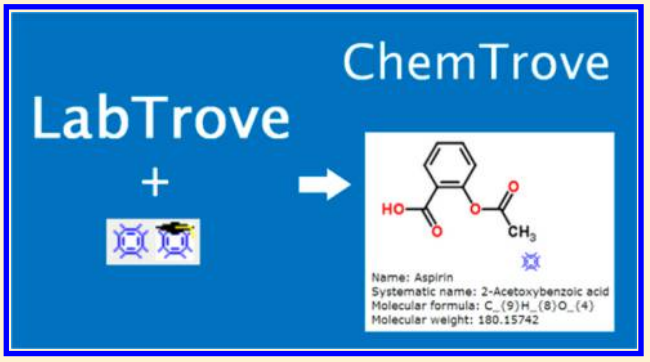
[†]Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge CB4 0WF, United Kingdom

[‡]Chemistry, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

[§]Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, North Carolina 27587, United States

Supporting Information

ABSTRACT: In designing an Electronic Lab Notebook (ELN), there is a balance to be struck between keeping it as general and multidisciplinary as possible for simplicity of use and maintenance and introducing more domain-specific functionality to increase its appeal to target research areas. Here, we describe the results of a collaboration between the Royal Society of Chemistry (RSC) and the University of Southampton, guided by the aims of the Dial-a-Molecule Grand Challenge, intended to achieve the best of both worlds and augment a discipline-agnostic ELN, LabTrove, with chemistry-specific functionality and using data provided by the ChemSpider platform. This has been done using plug-in technology to ensure maximum transferability with minimal effort of the chemistry functionality to other ELNs and equally other subject-specific functionality to LabTrove. The resulting product, ChemTrove, has undergone a usability trial by selected academics, and the resulting feedback will guide the future development of the underlying ELN technology.



INTRODUCTION

Electronic Lab Notebooks (ELNs) are ubiquitous in commercial settings, particularly in the pharmaceutical sector, but little used in academia. In general, enterprise ELNs do not readily lend themselves to the academic environment as they have been built to suit commercial purposes.[1] The drivers for academic use, most notably the motivation to publish work, are somewhat different to those of the commercial setting, which is primarily concerned with the protection of IP. Academia is concerned with progressing understanding by knowledge sharing, but most currently available commercial ELNs are not well suited to this process, even though ELNs have the potential to provide a significant reduction in the barriers to dissemination and enhance the provenance of chemical information.

Several attempts have been made over the past decade to create electronic systems to store data from academic activities. Much of this work stemmed from the digital repositories community, and examples exist where this approach has been applied to the research chemistry area.[2−6] However, repositories have historically been designed as systems to archive and curate and do not provide a method for data capture, particularly at the point of creation, and neither do they work well for recording contextual information. There is a requirement for an academic ELN to encompass all of these functions.

The Dial-a-Molecule[7] network funded by the Engineering and Physical Sciences Research Council (EPSRC) is aimed at the Grand Challenge: "In 20−40 years, scientists will be able to deliver any desired molecule within a timeframe useful to the end-user, using safe, economically viable and sustainable processes". The roadmap for "Dial-a-Molecule" identifies a key step to be the ability to predict the outcome of unknown reactions; knowledge generated from known reaction outcomes is critical. Despite the large reaction databases that have been compiled over the last century, for example, Beilstein,[8] Reaxys,[9] SPRESI,[10] and CASREACT,[11] our ability to predict reaction outcomes is still poor. One reason is that only a small subset of reactions carried out are ever reported in the literature and then with minimal experimental detail. Furthermore, it is rare for failed or suboptimal reactions to be reported, thereby discarding crucial information that could facilitate prediction. Electronic capture of full reaction data and outcomes in ELNs and their integration with searchable data repositories is thus a crucial enabling technology for "Dial-a-Molecule". The ePro ELN[12] by Novalys is an example of an ELN that allows captured reactions (both successful and failed) to be published to their platform and searched for using their search engine, but

Received: October 1, 2014
Published: February 13, 2015

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

these are all commercial products, rather than open source. The move to electronic capture of data relating to how reactions are carried out and what the outcomes are is a crucial step, and putting the technology and culture in place to allow public access to the data is the next step. Barriers to the uptake of ELNs by academia are the cost, concern about having data "locked-in" to a commercial solution, and the perceived overhead in the use compared with paper.

A significant driver for the adoption of ELNs is that academia is increasingly being held accountable for the management of the research data it generates. Mandates from funders[13] require researchers and their institutions to provide infrastructure to support research data management and to hold information for some considerable time after the project funding finishes. Additionally, there are now requirements from funders[14] that research outputs are made openly available. Furthermore, the recent Finch Report[15] recommended policy development to support Open Access by both the "author pays" and the "institutional repository" routes. This report also states that both Open Access publishers and academia should seek to innovate in the area of research data publication. The recent Royal Society report "Science as an Open Enterprise"[16] highlights the benefits of opening up scientific information.

Because commercial ELNs tend to be process-oriented, their adoption within academia has been very limited, so the community now needs assurance that there are types of ELNs that can work for them. There are some generic digital notebooks that have achieved widespread adoption such as Evernote[17,18] and OneNote;[19−21] these demonstrate a requirement for a flexible and shareable notebook. These features, coupled with a rigorous approach to data curation and tailoring with tools to support the research chemist, have the potential to form the basis of an ELN for academia. A review on the development of note taking in the digital era by Bird et al.[22] provides considerable background on this topic.

LabTrove[23] has been developed by the University of Southampton since 2005 with the intention of providing both structured and unstructured discipline-agnostic ways of recording data in order to serve the academic community. As an open source electronic (laboratory) notebook for researchers to plan experiments and save and share their results, its aim is to engage a developer community that can build specific add-on tools that can be incorporated via the open source model. "Out-of-the-box" it contains generic tools for researchers from any discipline to upload, display, and share data of a range of file types (either within LabTrove itself or via links). LabTrove offers an open solution toward the aims of the Dial-a-Molecule network but lacks the chemistry-specific and time-saving features that would provide greater motivation for its adoption, such as structure drawing and automatic stoichiometry calculation. It also needs to allow compound and reaction data to be published from the ELN directly to repositories. LabTrove therefore provides an ideal platform on which to develop supplementary features to support the work of the academic research chemist.

A Dial-a-Molecule meeting in December 2011 set up a collaboration between LabTrove and the Royal Society of Chemistry (RSC). The evolving interfaces to ChemSpider[24−27] provided an opportunity to integrate such a chemistry platform with LabTrove at the Web browser level.

ChemSpider was initially developed as a low-budget proof-of-concept project, by a small team, simply to contribute a free resource to the chemistry community. Following a two-year period of expanding the database content to over 20 million chemicals, adding functionality to the system to facilitate database curation and crowd-sourced depositions of data and developing a series of related projects, ChemSpider was acquired by the Royal Society of Chemistry (RSC) in 2009.[28] The original strategic vision of providing a structure-centric community for chemistry was expanded to create the world's foremost free access chemistry database and to make subsets of the data available as open data. ChemSpider has undergone continuous development to increase the number of compounds, the amount of information for each, and to improve the presentation, accessibility and quality of this data. Tools and interfaces have also been developed to allow other platforms and software to search this database, deposit to it, and integrate with it. A rich programming interface composed of multiple Web services enables integration with internal projects such as the RSC publishing platform, which hosts hundreds of thousands of scientific articles, the Marinlit marine natural product database,[29] and the Merck Index.[30] The exposed interfaces have also allowed analytical instrument vendors to integrate their software to permit them to identify structures.[31]

ChemSpider content is primarily aggregated and assembled by the contribution of data from various online resources and database hosts, the sources of the data being listed in the data sources Web page.[32] ChemSpider also offers a crowdsourcing platform whereby members of the chemistry community may contribute both new data or participate in the curation and validation of existing data.

We decided that the integration of ChemSpider with ELNs would be a path by which *direct* deposition of data could be performed. The intent was to allow users of ELNs to deposit selected records, or data contained within an ELN record (accompanied by metadata supplied in the elnItemManifest format[33]), into ChemSpider at the click of a button. In this way, scientists could share their data with the community and maintain provenance to their original ELN record. A proof-of-concept was delivered via integration with the IDBS ELN system in 2011.[34] However, the approach used was not easily extensible; although it was reliant on generic ChemSpider Web services that could in theory be used by any system or platform, most of the programming that it entailed was reliant on that ELN's specific API. As such, the code could be used only by that particular ELN, and would have needed to be largely rewritten to use it in another ELN, or indeed even that particular ELN if its architecture changed substantially, which indeed it did. Clearly, a more transferable method to offer ChemSpider functionality to ELNs was required, the basis of which would be to use the increasingly common trait of ELNs to be Web-based systems, accessed via an Internet browser, rather than desktop applications.

In this paper, we describe how the generic LabTrove ELN has been enhanced with chemistry-specific features by integrating plug-ins and compound data provided by a third party via the Internet. Although a number of ELNs are currently available to the chemical community,[35] with a range of tools for chemists, the approach described here is novel in two ways. First, in contrast to other more discipline-specific ELNs, which tend to be developed and provided in a single package by a single provider, this has been done through the use of reusable Web parts. This use of plug-in and widget architecture is not common in other ELNs, but is certainly appropriate in a collaborative project. As such, it is the most extensible way to separate the functionality from different

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

providers, which might also be targeted at different user groups, so that in the future other subject-specific plug-ins could be added to the ELN. It is also the most extensible way to allow future integration of the chemistry functionality and data with other ELNs. The second novel approach taken here is to connect the ELN to an external Web-based data source providing automated retrieval of compound information from an extensive online database. This approach differs from most other ELNs that rely on built-in libraries, which are more fixed in nature in that they tend to grow only incrementally, for example, via updates. Using an online database as a source is more dynamic, allowing the body of information on which a system relies to grow at the same rate as the body of chemical information on the Web.

The result of this collaboration is a first version of ChemTrove (LabTrove enhanced by ChemSpider lookup plug-ins and templates). This paper describes the development of ChemTrove and the main outcomes of user trials.

APPROACHES

A primary expectation of an ELN for chemistry is to include a chemical depiction of the molecule of interest in a notebook entry. In the generic version of LabTrove, this would be performed by an import operation: drawing the molecule in a separate chemical editor program and uploading the resulting file, either in a chemical file format (e.g., ChemDraw's .cdx[36] or MDL molfile[37]) or converting it into an image file and attaching it to the entry. The advantage of saving the file in its native chemical format is that no information in it would be lost, and the file could potentially be edited in the future. However, a user would not be able to view the chemical structure directly from within the ELN due to being dependent on having the appropriate software program installed to download and then view files in that format. Therefore, the first and simplest tool to be added to the LabTrove entry edit page was the "insert from ChemSpider" tool. This tool allows a user to search ChemSpider for a compound and click on it to retrieve an image (in PNG format) of the compound and other basic compound information from ChemSpider and include it in the ELN entry, thus avoiding the need to draw molecules already in ChemSpider. However, ELN entries often relate to novel molecules that might not yet be in ChemSpider, or researchers might prefer to depict a molecule already in ChemSpider in a different way, for example, its orientation. As a result, the second tool to be added to LabTrove was the "draw compound" tool, which allows a user to draw a molecule when editing an entry and for the resulting structure to be rendered on the page. This also gave the opportunity to provide a tool that enables structure visualization to be more interactive and is akin to the use of the ChemDoodle sketcher[38] within the LabArchives ELN.[39]

Another tool that many other chemistry ELNs provide is a tool to aid in the creation of stoichiometry tables, which define the compounds that take part in a reaction and their amounts. An "edit stoichiometry table" tool would allow known compounds in the reaction to be defined via ChemSpider compound lookups so that the user would not need to draw the compound or type in its name, molecular weight, or formula.

As well as providing tools that retrieve data from ChemSpider, an additional aim is to provide tools to publish compound and reaction data from ELNs directly to the Royal Society of Chemistry's databases such as ChemSpider SyntheticPages.[40] Extensive back-end development of the Royal Society of Chemistry's databases is required before this will be possible, as reported recently in a general overview of the developing architecture.[41] However, as a step toward the aim of publishing understandable compound and reaction data from LabTrove to ChemSpider, the LabTrove "template" functionality has been used,[42] which allows templates to be created and applied to new posts to define text and placeholder textboxes within them. Using this functionality, LabTrove templates have been written to structure entries that contain compound and reaction data such that they will be more understandable than a free format block of text and references to files when deposited to the repositories.

While historically ELNs tended to be desktop applications, the majority now have Web front ends that are accessed via an Internet browser. This has the advantage that no software needs to be installed or kept up to date locally to use the ELN, which in theory can be accessed on mobile devices (assuming caveats regarding operating system and support for programming languages, e.g., iOS and Java compatibility) as well as computers. Mobile access is of particular importance for ELNs because there is not always a networked computer in the laboratory where the experiment is being conducted. LabTrove is one such Web-based system, and details of its architecture are described elsewhere.[42]

ChemSpider is also a Web-based system, and its architecture has gradually been adopting jQuery widgets[43] to render components of its Web pages so that those components can be displayed in multiple locations and can also be reused without code duplication. As such, the logical approach when adding chemistry functionality to LabTrove via ChemSpider was to do so by concentrating development on jQuery widgets that could be called from within the LabTrove pages as required, rather than hard-coding directly into the LabTrove code. Moreover, the widgets could easily be integrated with other ELNs or any Web site (e.g., blogs) if required. Although the widgets could be stored and invoked locally within the LabTrove code, for easier maintenance and updates, it was decided to call widgets hosted on a ChemSpider server. To show one of these widgets in an HTML page or dialogue popup box, all that is required is to include references to the widget files on the ChemSpider server in the page header, at least one placeholder to be populated by the widget in the page body, and JavaScript functions to initialize the widget. By hosting the widget files on the ChemSpider Parts server, it was possible to fix critical bugs by simply updating the files on the server, so that ChemTrove administrators would not need to update the code. For substantial changes in functionality, the widgets hosted on ChemSpider Parts are versioned, so to upgrade ChemTrove to use an incremented new version of the widgets, all that is needed is to change the URLs of the referenced widget files.

All three of the new tools involve adding content to LabTrove pages when editing them. Edits to LabTrove pages are managed by a popular open source editor control (TinyMCE[44]), which is used to edit many other Web sites and blogs, for example, WordPress.[45] As such, the most elegant way of incorporating the widgets with LabTrove was to simply write a new "ChemSpider tools" plug-in for TinyMCE (with its own button in the editor toolbar), which would in turn call the ChemSpider widgets. A generic plug-in for this editor could also potentially be added to any blog or Web site that uses this editor. With this more general user base in mind, the added chemistry functionality has been split into two TinyMCE plug-

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

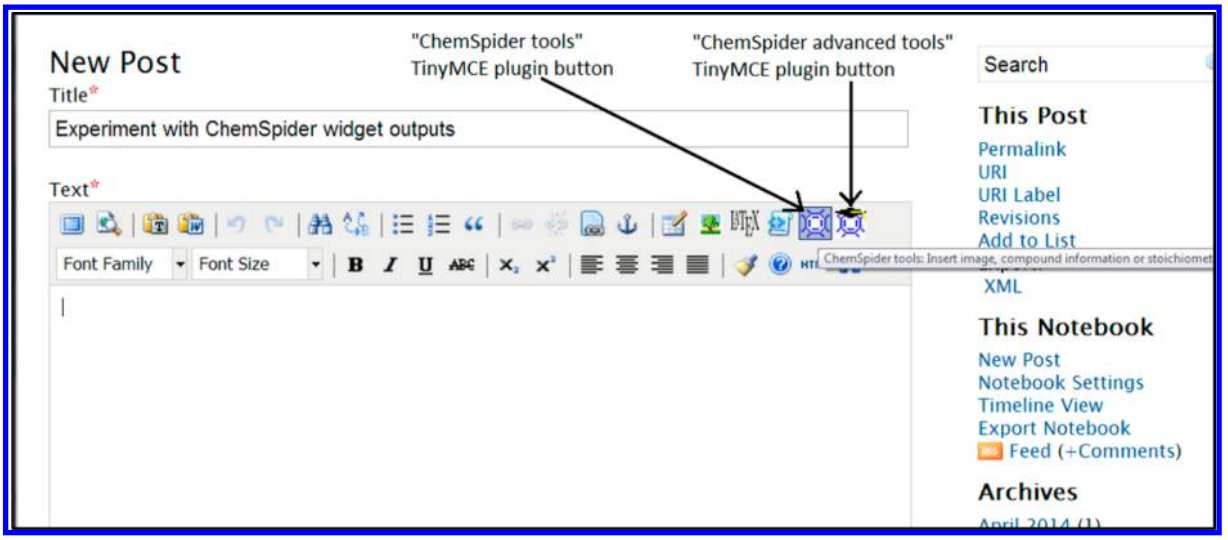


Figure 1. ChemTrove edit page with "ChemSpider tools" and "ChemSpider advanced tools" TinyMCE plug-in buttons.

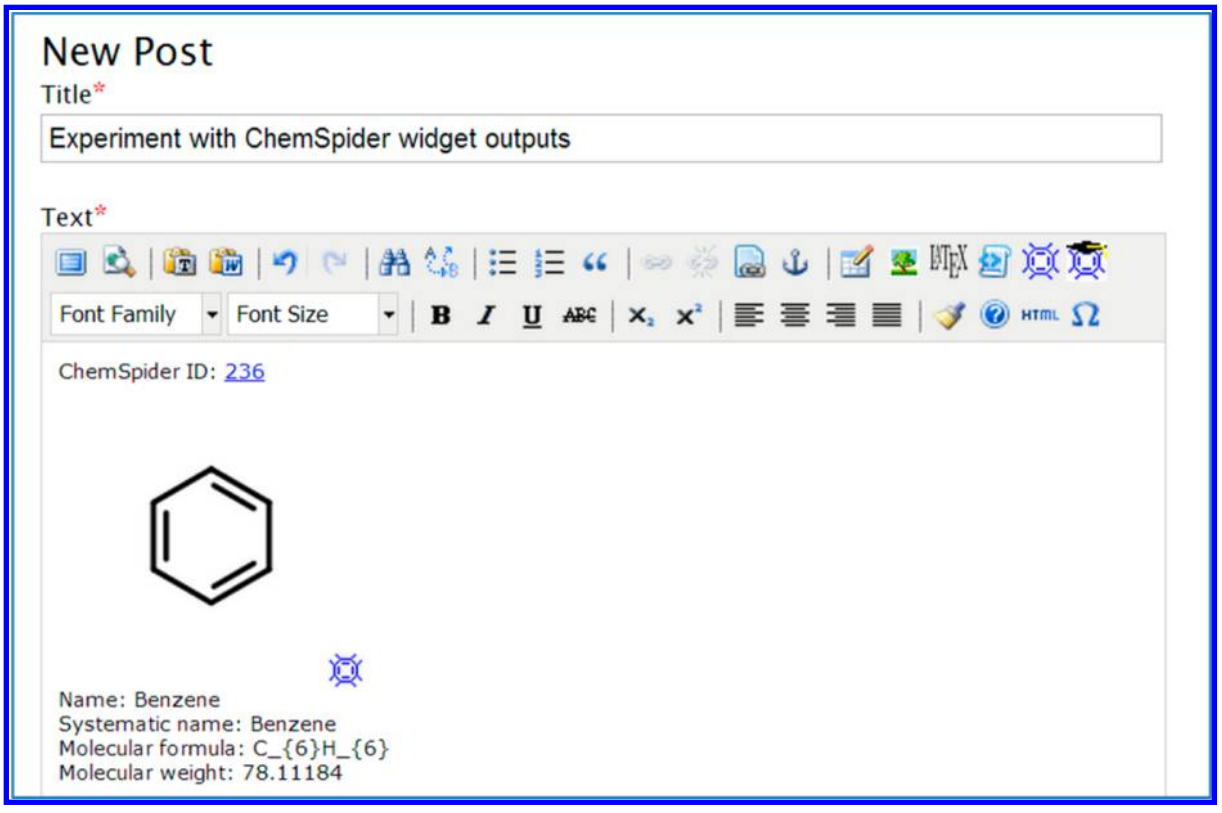


Figure 2. ChemTrove page containing output of "Insert from ChemSpider" tool.

ins: a standard "ChemSpider tools" plug-in and a "ChemSpider advanced tools" plug-in. The "ChemSpider tools" plug-in invokes the "insert from ChemSpider" or "edit stoichiometry table" tools, which both output pure HTML into the page body and do not require any additions to the page header to be rendered correctly. The "ChemSpider advanced tools" plug-in invokes the "draw compound" tool, which inserts HTML elements that call a ChemSpider widget. Practically, this would mean that to integrate the "ChemSpider advanced tools" plug-in with a Web site or blog that uses the TinyMCE editor would need an extra step (to add the required widget references to the header of any page that shows the resulting content), rather than the standard "ChemSpider tools" plug-in, which can just be installed like any standard TinyMCE plug-in. Figure 1 shows the LabTrove edit screen with editor buttons that invoke the "ChemSpider tools" and "ChemSpider advanced tools" TinyMCE plug-ins.

To transform LabTrove into ChemTrove, all that is required is for an administrator to change two settings in the main LabTrove configuration file. The programming has effectively been layered so that the main development has been on the ChemSpider Parts widgets, and these have been wrapped into two TinyMCE plug-ins, which have in turn been wrapped into a LabTrove "chemSpiderAdvanced" plug-in (requiring minimal changes to the LabTrove code itself). This layered approach would make it straightforward to integrate the ChemSpider

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

widgets with other ELNs or Web sites (most easily those that use the TinyMCE editor). Further development of LabTrove using the plug-in architecture that was exploited by this project could also allow users to access other databases (chemistry or otherwise) or other discipline-specific features. LabTrove is open source, and its code, including that of the LabTrove and TinyMCE plug-ins that invoke the ChemSpider widgets, is available to view, contribute to, download, and locally install via SourceForge.[46] The ChemSpider widget code is proprietary to the Royal Society of Chemistry at present, although they are free to use, and the technical details of some of the methods used by the widgets relevant to this project are summarized in the Supporting Information. The Royal Society of Chemistry has embraced the use of open source code,[47] although for the time being, the ChemTrove chemistry components have not yet been released as open source as this project was to investigate proof-of-concept capabilities.

CHEMTROVE TOOLS

The "Insert from ChemSpider" ChemTrove tool is straightforward in its functionality for allowing a user to specify the compound information that they would like to add to a LabTrove entry (compound image, name, systematic name, molecular formula, molecular weight, and/or a link to the corresponding ChemSpider page). A user can then enter a search term (e.g., compound name, SMILES,[48] InChI,[49,50] trade name, etc.) to search ChemSpider and retrieve to the edit page formatted HTML containing the requested information. The compound image is dynamically retrieved from ChemSpider and is the most up to date depiction of that molecule (specified by ChemSpider ID) rendered as a PNG image. A set of screen captures that demonstrate the use of this tool are shown in Figure S1 of the Supporting Information, and the resulting output in a ChemTrove page is shown in Figure 2.

The "Draw compound" ChemTrove tool allows a user to draw their molecule in a pop-up window using the JSDraw structure editor[51] and generates the MDL molfile representation of the resulting structure. It then renders the molecule in an interactive cartoon style (with the ability to rotate and zoom in on the molecule, etc.) using the lightweight version of the JSMol three-dimensional compound viewer.[52] The lightweight version of JSMol was integrated as the structure viewer, rather than the full version of JSMol, because it was more straightforward to write a jQuery wrapper to invoke it, which would be more usable when viewing LabTrove on mobile devices. Screen captures demonstrating its use are shown in Figure S2 of the Supporting Information, and its output is shown in Figure 3.

The "edit stoichiometry table" ChemTrove tool guides a user through a series of dialogue boxes to make a stoichiometry table containing all of the chemicals used and produced in a reaction as separate rows, performing calculations and unit conversions to make this task easier. The technical details of this widget are given in the Supporting Information.

A set of screen captures that demonstrate the use of this "edit stoichiometry table" tool are shown in Figures S3–S5 of the Supporting Information, and the resulting output on a ChemTrove page is shown in Figure 4. The resulting stoichiometry table is displayed somewhat differently to those constructed by other ELNs, which tend to have many more columns—one for each value for that chemical. It was necessary to amalgamate all of these values into a maximum of four

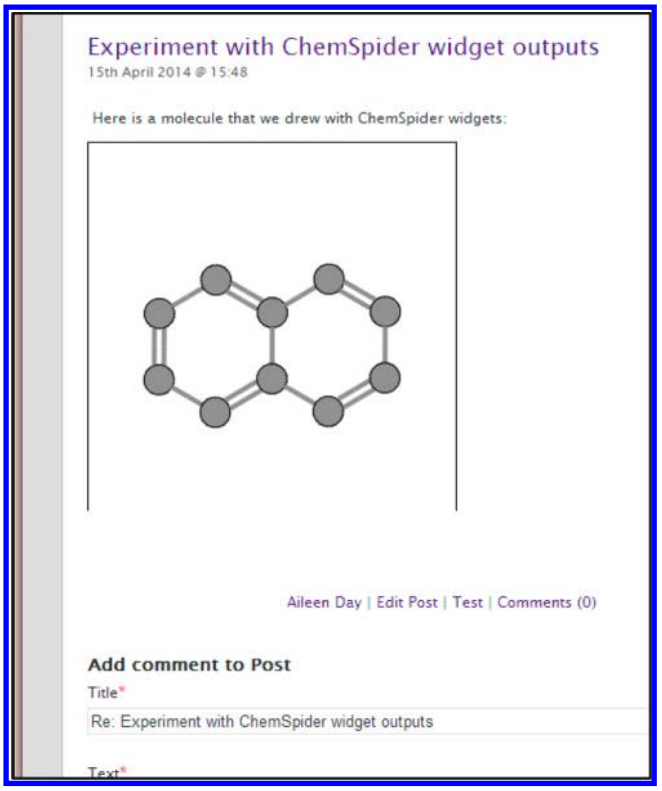


Figure 3. ChemTrove page containing output of "Draw compound" tool.

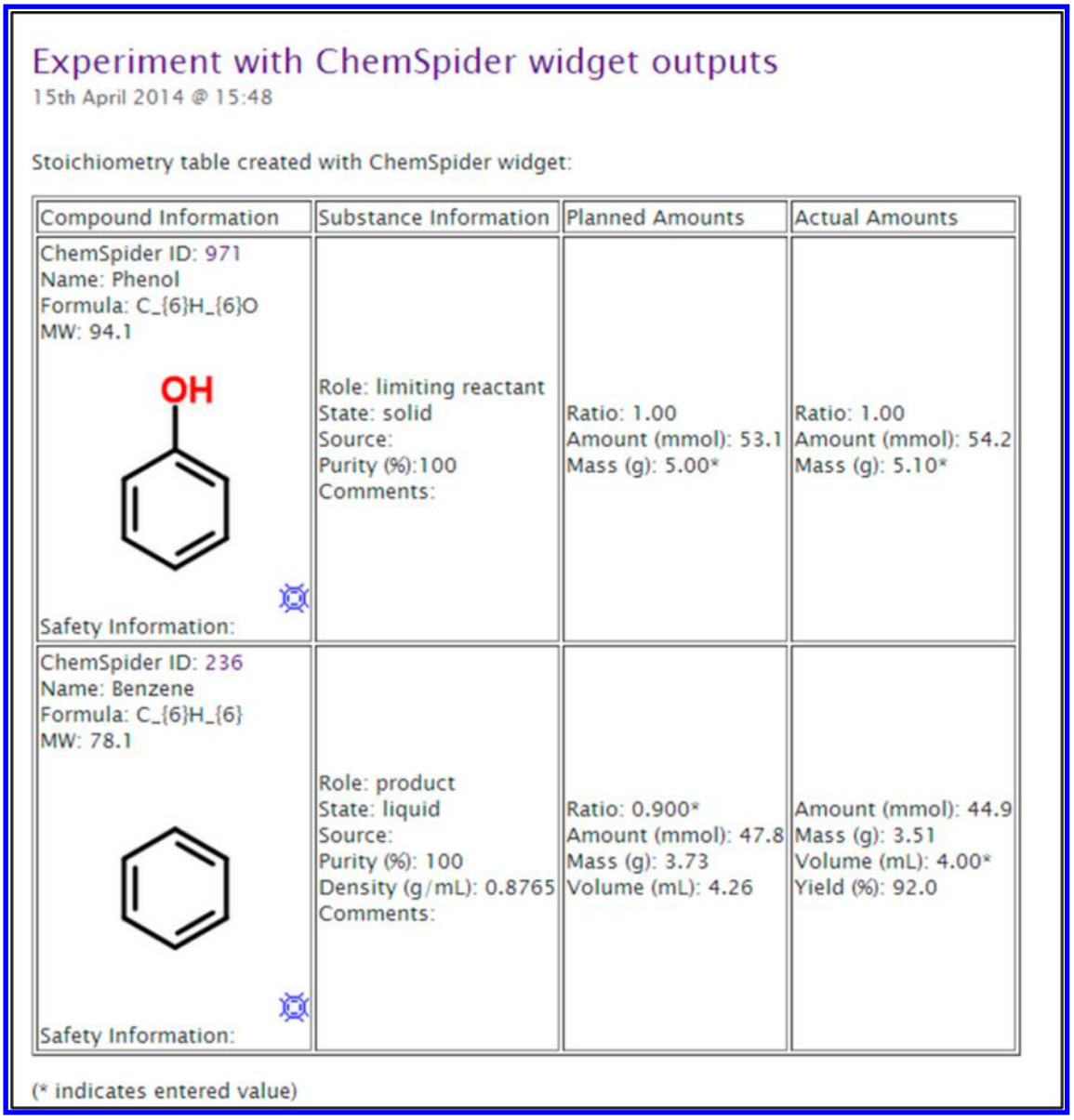
columns (with multiple values being shown in each) due to the constrained width of page layouts in LabTrove.

CHEMTROVE TEMPLATES

The first LabTrove template, "ChemSpider compound and associated data", was written to structure a ChemTrove entry so that the compound information in it would be suitable to be published to ChemSpider. This is of particular use to users with compound characterization data who wish to capture this information in a structured format. The template is shown in Figures S6–S8 of the Supporting Information. The fields in it correspond to all of the fields that can be populated when depositing a compound to ChemSpider using the deposition system on its Web site (via manual entry and SDF file deposition). In addition to this, it is possible to upload spectra files for the compound into the template (because uploading spectra from ELNs into ChemSpider will be a valuable way to share them). The template functionality in LabTrove is straightforward to implement using the online documentation,[53] but is limited to creating text boxes that contain free format text. For this template to be of more use, it should be possible to specify fields to be restricted drop-down lists or file upload boxes. As an interim step, a help page was written (to which a link was added to the template) providing guidance as to how to populate each field (e.g., if certain values were permitted for them or to flag fields for which files should be uploaded).

Similarly, the "ChemSpider SyntheticPages style reaction" template was written to structure a ChemTrove entry so that the compound information in it would be suitable to be published to ChemSpider SyntheticPages using the same fields that can be entered when submitting a page via that Web site. The template is shown in Figures S9–S11 of the Supporting

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501–509



Experiment with ChemSpider widget outputs

15th April 2014 @ 15:48

Stoichiometry table created with ChemSpider widget:

Compound Information	Substance Information	Planned Amounts	Actual Amounts
ChemSpider ID: 971 Name: Phenol Formula: C_{6}H_{6}O MW: 94.1 OH Safety Information:	Role: limiting reactant State: solid Source: Purity (%):100 Comments:	Ratio: 1.00 Amount (mmol): 53.1 Mass (g): 5.00*	Ratio: 1.00 Amount (mmol): 54.2 Mass (g): 5.10*
ChemSpider ID: 236 Name: Benzene Formula: C_{6}H_{6} MW: 78.1 Safety Information:	Role: product State: liquid Source: Purity (%): 100 Density (g/mL): 0.8765 Comments:	Ratio: 0.900* Amount (mmol): 47.8 Mass (g): 3.73 Volume (mL): 4.26	Amount (mmol): 44.9 Mass (g): 3.51 Volume (mL): 4.00* Yield (%): 92.0

(* indicates entered value)

Figure 4. ChemTrove page containing output of "edit stoichiometry table" tool.

Information. However, a reaction needs to be recorded in a much higher level of detail for publication to ChemSpider SyntheticPages than would be recorded in a day-to-day ELN, so a more simplified "Reaction template" was also written for more general use. The template is shown in Figure S12 of the Supporting Information.

The use of these templates was tested during a student intern project[54] to digitize theses data and publish it. This resulted in 1035 new LabTrove entries (358 that use the "ChemSpider compound and associated data" template, 198 using the "ChemSpider SyntheticPages style reaction" template, and 479 using the "Reaction template") hosted in a public notebook at the University of Southampton.[55] In parallel, 208 new compounds (and over 600 spectra) that reflect the content of these formatted LabTrove pages were submitted to ChemSpider.

CHEMTROVE TRIAL AND FEEDBACK

In order to test the suitability of ChemTrove as a day-to-day ELN for academics and to receive feedback to develop it and the associated ChemSpider plug-ins, a restricted cloud-based trial was arranged. The trial was conducted over a three month period starting early in 2014 and tested by selected academics around the world. A further longer trial ChemTrove instance was hosted by the RSC and ChemSpider. In all, 57 users participated in both of these trials. These were from 14 research groups and three universities in the United Kingdom (Imperial College London, University of Southampton, and University of Warwick) and six in the United States (Cornell University, Indiana University, University of Florida, Tulane University, University of Oregon, and Georgia Tech University). They were from disciplines including synthetic organic chemistry, analytical chemistry, supramolecular chemistry, biochemistry, nanomaterial synthesis, cheminformatics, library science, health and safety, and information technology, and at a variety of stages of their career (including masters students and Ph.D. students as well as group leaders). ChemTrove was demonstrated to each participant in person or via Web conferencing facilities. Further initial training materials were then made available and included online (LabTrove doc-

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

umentation[56] and demonstration YouTube videos[57]). Between them, these trial participants created 78 LabTrove entries (experiments). Feedback from the trial was primarily from these initial demonstration sessions and via questions and comments shortly afterward. The participants were then provided with nine exercises (stored in the form of ChemTrove pages in a training notebook), which led them through performing basic ChemTrove tasks and using the ChemSpider tools. They were then encouraged to generate their own example pages and to use ChemTrove in place of (or in parallel with) their usual laboratory notebook.

Feedback was encouraged via a designated ChemTrove notebook, either in free format with their own suggestions and observations or using a questionnaire template that was circulated at the end of the trial. The questionnaire was split into four main sections: "About you", "Your overall aims", "What did you do", and "Suggested new features/issues for prioritization". The "About you" section asked background questions about the ChemTrove user such as their academic position and details of the platform from which they accessed ChemTrove. "Your overall aims" asked questions about the current lab notebook used, its limitations, what they would like to do in ChemTrove, its advantages and disadvantages compared to their current lab notebook, and whether they would like to continue using it after the ChemTrove trial ended. The "What did you do" section asked which ChemTrove tools and templates they used and for links to example pages created using them. The "Suggested new features/issues for prioritization" section comprised a list of a large number of issues with ChemTrove and feature requests for it, which were to be marked for prioritization as either High, Medium, or Low priority. This list of features included all suggestions from the initial demos and training sessions. It was grouped into subsections of "General LabTrove usability", "LabTrove templates", "Stoichiometry table widget", "Structure drawer widget", and "Other Chemistry functionality". By asking these questions, we aimed to get an idea of what tools and features were most important to users of different kinds, how they used them, and to be able to identify the most important new features and issues to be addressed in the future development of LabTrove and the ChemSpider widgets in order for ChemTrove to become a more usable ELN.

The response rate to the questionnaire was 7%, and the results are shown in Table S1 of the Supporting Information. Although the response rate to the formal questionnaire was low, there was a lot of less structured feedback given in the form of users' first impressions, feedback, and questionnaires during and immediately after their initial demonstrations. Furthermore, 12 masters students left feedback in a different format, as shown in Table S2 of the Supporting Information. All of this feedback has been taken into consideration in making plans for the future development of the underlying technologies.

The ChemTrove tools supplied by the ChemSpider widgets were generally regarded as important, but in need of some future work to increase their functionality, such as the following:

- The "Insert from ChemSpider" tool's most important issue to address was to allow searching based on drawing a structure as well as the existing text search (and if this structure was not found in ChemSpider to simply allow it to feed into the "Draw compound" ChemTrove tool to be rendered directly in the ChemTrove page).
- The "structure drawing" widget's issues that need addressed are as follows:
  - Make sure it works on all browsers and platforms.
  - Allow reactions (not just molecules) to be drawn and rendered.
  - Allow ChemDraw files to be uploaded simply and rendered directly in an entry.
  - Render the drawn molecule with an alternative, more "chemical" structure viewer (rather than the JSmol cartoon-style).
- The "edit stoichiometry table" widget needs to give more support to compounds not in ChemSpider (so that they can be drawn, and the molecular formula and weight can be calculated and added) and should be able to retrieve compound densities from ChemSpider.

Future Development. The functionality requested from user feedback should be relatively straightforward to address in future versions of the underlying ChemSpider widgets and TinyMCE plug-ins. This development needs to be performed in conjunction with wider changes and updates to the widget architecture. The feedback and observations that we received with regard to the Chemistry plug-ins were very positive, with few issues, although there were comments directed at the additional functionality that would be desirable in the underlying LabTrove system.

The approach taken for this project was to develop the chemistry functionality via widgets and install it into LabTrove using plug-ins, to allow maximum extensibility to other systems.

As such, the ChemSpider lookup functionality described here could, with minimal effort, be transferred to other Web-based ELNs. No additional programming would be required to integrate other ELNs that use the TinyMCE editor, but the ELN would need to be modified to add ChemSpider references to ELN pages that show drawn molecules. ELNs that do not use this editor would need some development to add an editor button that invoked the widgets and added the results to the experiment, but much of the LabTrove plug-in code could be reused. ELNs that already feature chemistry-specific functionality could benefit from additional ChemSpider lookups (to save time drawing molecules, or additional information). There would, however, be more advantage integrating the ChemSpider widgets with more generic ELNs that do not currently have the ability to create stoichiometry tables or draw molecules.

Similarly, other chemistry functionality and lookups could be added to ChemTrove by simply adding to the functionality of the plug-in created, for example, to add PubChem[58] lookups similar to those possible in Avogadro[59] via the JSMol interface. Likewise, more biological functions could be wrapped into a BioChemistry plug-in that included Protein Data Bank[60] lookups, which could be invoked via the JSMol interface (again similar to Avogadro), or to provide similar functionality to the ChemCalc[61] interface by using the Web services (together with other third party components) underlying it.

CONCLUSIONS

We have described an initial project to integrate the plug-ins, widgets, compound, and reaction data provided by ChemSpider to augment an open source ELN to support chemists. This project was approached so that the additional functionality and

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

underlying chemical data could be developed in a continuous and dynamic way, rather than the more static approach of most ELNs that rely on less frequent software and database updates being distributed by the ELN vendors and implemented by the ELN administrators. We have demonstrated that it is possible to add compound data from ChemSpider, draw and render compounds, and create stoichiometry tables, and as the result of feedback, we now have a better idea of what is required by users.

■ ASSOCIATED CONTENT

Supporting Information

Screen captures of the ChemTrove tools and templates, full technical details, and full questionnaire results. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: daya@rsc.org.

Funding

The authors thank EPSRC for funding the CombeChem eScience pilot (GR/R67729/01), which formed the basis for much of the work described here, and also for funding the Dial-a-Molecule Grand Challenge Network (EP/K004840/1 and EP/H034447/1).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.E.D. thanks the University of Southampton for supporting a visiting position.

■ REFERENCES

(1) King, A. Notebooks Go Digital Chemistry World [Online], 2013, *5*, 64−67 http://www.rsc.org/chemistryworld/2013/05/electronic-lab-notebook-review (accessed September 29, 2014).

(2) Downing, J.; Murray-Rust, P.; Tonge, A. P.; Morgan, P.; Rzepa, H. S.; Cotterill, F.; Day, N.; Harvey, M. J. SPECTRa: The Deposition and Validation of Primary Chemistry Research Data in Digital Repositories. *J. Chem. Inf. Model.* **2008**, *48*, 1571−1581.

(3) eBank UK, http://www.ukoln.ac.uk/projects/ebank-uk/dissemination/ (accessed December 16, 2014).

(4) Repository for the Laboratory (R4L), http://r4l.eprints.org/ (accessed December 16, 2014).

(5) CLARION, https://clarionproject.wordpress.com/ (accessed December 16, 2014).

(6) Harvey, M. J.; Mason, N. J.; Rzepa, H. S. Digital Data Repositories in Chemistry and Their Integration with Journals and Electronic Notebooks. *J. Chem. Inf. Model.* **2014**, *54*, 2627−2635.

(7) Dial a Molecule Grand Challenge, http://www.dial-a-molecule.org (accessed September 29, 2014).

(8) Heller, S. *The Beilstein Online Database: Implementation, Content, and Retrieval*; ACS Symposium Series 436; American Chemical Society: Washington, DC, 1990.

(9) Lawson, A. J.; Swienty-Busch, J.; Geoui, T.; Evans, D. The Making of Reaxys—Towards Unobstructed Access to Relevant Chemistry Information. In *The Future of the History of Chemical Information*; ACS Symposium Series 1164; American Chemical Society: Washington, DC, 2014; pp 127−148.

(10) Roth, D. L. SPRESIweb 2.1, a Selective Chemical Synthesis and Reaction Database. *J. Chem. Inf. Model.* **2005**, *45*, 1470−1473.

(11) CASREACT (accessed through the SciFinder Interface), http://www.cas.org/content/reactions (accessed December 14, 2014).

(12) ePro ELN, Novalys, http://www.enovalys.com/epro/ (accessed December 14, 2014).

(13) For a selection of funders policies and an overview see: RCUK Common Principles on Data Policy, http://www.rcuk.ac.uk/research/datapolicy/ (accessed December 16, 2014); Overview of Funders' Data Policies, http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies (accessed December 16, 2014); Dissemination and Sharing of Research Results, http://www.nsf.gov/bfa/dias/policy/dmp.jsp (accessed December 16, 2014).

(14) EPSRC Policy on Access to Research Outputs, http://www.epsrc.ac.uk/about/access/roaccess/ (accessed September 29, 2014).

(15) The Finch Report, http://www.researchinfonet.org/publish/finch/ (accessed September 29, 2014).

(16) Science as an Open Enterprise: Open Data for Open Science, http://royalsociety.org/policy/projects/science-public-enterprise/ (accessed September 29, 2014).

(17) Evernote, http://evernote.com (accessed December 16, 2014).

(18) "How Academics Use Evernote To Make Life Easier" Blog Post, http://www.academicpkm.org/2013/07/30/how-exactly-do-people-use-evernote-in-academia/ (accessed December 16, 2014).

(19) OneNote, http://www.onenote.com/ (accessed December 16, 2014).

(20) "OneNote as a Research Log Book" Blog Post, http://mathpen.wordpress.com/2013/01/15/onenote-as-a-research-log-book/ (accessed December 16, 2014).

(21) "Electronic Lab Notebooks" Blog Post, http://postdocexperience.scienceblog.com/2012/11/19/electronic-lab-notebooks/ (accessed December 16, 2014).

(22) Bird, C. L.; Frey, J. G. Chemical Information Matters: An e-Research Perspective on Information and Data Sharing in the Chemical Sciences. *Chem. Soc. Rev.* **2013**, *42*, 6754−6776.

(23) LabTrove, http://www.labtrove.org/ (accessed September 29, 2014).

(24) ChemSpider, http://www.chemspider.com (accessed September 29, 2014).

(25) Pence, H. E.; Williams, A. J. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, 1123−1124.

(26) Williams, A. J. ChemSpider: Integrating Structure-Based Resources Distributed across the Internet. In *Enhancing Learning with Online Resources, Social Networking, and Digital Libraries*; ACS Symposium Series 1060; Belford, R. E., Moore, J. W., Pence, H. E., Eds.; American Chemical Society: Washington, DC, 2010; pp 23−39.

(27) Williams, A. J. Public Compound Databases − How ChemSpider Changed the Rules Making Molecules on the Web Free. In *Collaborative Computational Technologies for the Life Sciences*; Ekins, S., Hupcey, M. A. Z., Williams, A. J., Eds.; Wiley: New York, 2011; pp 363−386.

(28) RSC Acquires ChemSpider Press Release, http://www.rsc.org/AboutUs/News/PressReleases/2009/ChemSpider.asp (accessed December 16, 2014).

(29) MarinLit Website, http://pubs.rsc.org/marinlit/ (accessed December 16, 2014).

(30) Merck Index Website, https://www.rsc.org/merck-index (accessed December 16, 2014).

(31) Little, J. L.; Williams, A. J.; Pshenichnov, A.; Tkachenko, V. Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 179−185.

(32) ChemSpider Data Sources, http://www.chemspider.com/DataSources.aspx (accessed December 16, 2014).

(33) Coles, S. J.; Frey, J. G.; Bird, C. L.; Whitby, R. J.; Day, A. E. First Steps Towards Semantic Descriptions of Electronic Laboratory Notebook Records. *J. Cheminf.* **2013**, *5*, 52−62.

(34) "ChemSpider Plugin for IDBS ELN" Blog Post, http://www.chemspider.com/blog/chemspider-plugin-for-idbs-eln.html (accessed September 29, 2014).

(35) LIMSwiki.org, http://www.limswiki.org/ (accessed September 29, 2014).

(36) ChemDraw CDX file format, http://www.cambridgesoft.com/services/documentation/sdk/chemdraw/cdx/ (accessed September 29, 2014).

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509

(37) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244−255.
(38) ChemDoodle, http://www.chemdoodle.com (accessed December 16, 2014).
(39) ChemDoodle in LabArchives ELN, http://web.chemdoodle.com/labarchives (accessed December 16, 2014).
(40) ChemSpider SyntheticPages, http://cssp.chemspider.com (accessed September 29, 2014).
(41) Williams, A. J.; Tkachenko, V. E. The Royal Society of Chemistry and the Delivery of Chemistry Data Repositories for the Community. *J. Comput.-Aided Mol. Des.* **2014**, *28* (10), 1023−1030.
(42) Milsted, A. J.; Hale, J. R.; Frey, J. G.; Neylon, C. LabTrove: A Lightweight, Web Based, Laboratory "Blog" as a Route towards a Marked Up Record of Work in a Bioscience Research Laboratory. *PLOS One.* **2013**, *8*, No. e67460.
(43) jQuery, http://jquery.com/ (accessed September 29, 2014).
(44) TinyMCE, http://www.tinymce.com/ (accessed September 29, 2014).
(45) WordPress, https://wordpress.com/ (accessed September 29, 2014).
(46) LabTrove Code Repository Hosted by SourceForge, http://sourceforge.net/p/labtrove (accessed September 29, 2014).
(47) Day, A. E., Williams, A. J., Batchelor, C. R., Kidd, R. J., Tkachenko, V. E. Utilizing Open Source Software to Facilitate Communication of Chemistry at RSC, Published in Open source software. In *Life Science Research: Practical Solutions to Common Challenges in the Pharmaceutical Industry and Beyond*; Harland, L., Forster, M., Eds.; Woodhead Publishing Series in Biomedicine; Woodhead Publishing: Cambridge, 2012; pp 63−87.
(48) 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.
(49) Stein, S. E.; Heller, S. R.; Tchekhovski, D. An Open Standard for Chemical Structure Representation — The IUPAC Chemical Identifier. *Nimes International Chemical Information Conference Proceedings* **2003**, 131−143.
(50) Heller, S. R.; McNaught, A.; Stein, S. E.; Tchekhovski, D.; Pletnev, I. InChI - the Worldwide Chemical Structure Identifier Standard. *J. Cheminf.* **2013**, *5*, No. 7.
(51) JSDraw Structure Editor, http://www.scilligence.com (accessed September 29, 2014).
(52) Lightweight Version of JSMol, http://wiki.jmol.org/index.php/Jmol_JavaScript_Object#Lightweight_JSmol (accessed September 29, 2014).
(53) Creating a LabTrove Template, http://www.labtrove.org/documentation/Creating_a_template (accessed September 29, 2014).
(54) "Southampton University Internships to Transfer Thesis Data into LabTrove and ChemSpider" Blog Post, http://www.chemspider.com/blog/southampton-university-internships-to-transfer-thesis-data-into-labtrove-and-chemspider.html (accessed September 29, 2014).
(55) University of Southampton's Department of Chemistry Instance of Labtrove, http://blogs.chem.soton.ac.uk/whitby_group (accessed September 29, 2014).
(56) LabTrove Documentation, http://docs.labtrove.org/2.3/lt/Main_Page (accessed September 29, 2014).
(57) Getting Started with ChemTrove YouTube Playlist, http://youtu.be/kEMMZP1Qm0Q?list=PLY2XyqyPZY3_3bi0Mmv2vpikmzim1o-iF (accessed September 29, 2014).
(58) PubChem, https://pubchem.ncbi.nlm.nih.gov/ (accessed December 16, 2014).
(59) Avogadro Cross-Platform Molecular Editor, http://avogadro.cc/wiki/Main_Page (acessed December 16, 2014).
(60) Protein Data Bank, http://www.rcsb.org/pdb/home/home.do (accessed December 16, 2014).
(61) Patiny, L.; Borel, A. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J. Chem. Inf. Model.* **2013**, *53*, 1223−1228.

DOI: 10.1021/ci5005948
J. Chem. Inf. Model. 2015, 55, 501−509