

Lab Service Wiki: a wiki-based data management solution for laboratories production services

Antoni Hermoso, Michela Bertero, Silvia Speroni, Miriam Alloza, Guglielmo Roma
Centre for Genomic Regulation, Barcelona, Catalonia, Spain

email: {toni.hermoso, michela.bertero, silvia.speroni, miriam.alloza,
guglielmo.roma}@crg.cat

Abstract. *Lab Service Wiki* is a Semantic MediaWiki implementation for the management of a production laboratory. Here we describe its implementation on a protein service lab. Users of the service enter information about a sample and the desired analysis to be performed by using a semantic-enabled form built on top of a wiki page. After submitting, a workflow is created, and the manager of the service can assign different experimental tasks to the lab operators. The final output is the generation of a report for the requester. Users and operators, according to their profile and granted permissions, can track the state of the requests and the associated experiments at any time. People interested in this implementation can access it at: <http://labservice.biocore.crg.cat>

Keywords: wiki, semantics, protein, workflow, laboratory

1. Introduction

Establishing a new lab-based service requires the implementation of dedicated data management systems to track and store experimental information in a proper way.

Nevertheless, many small and medium sized laboratories and research facilities still handle and track users' requests, experiment results, and analysis reports in a very rudimentary way. These 'outdated' practices consist of using only traditional paper-based notebooks for annotating Standard Operating Procedures (SOPs) and experiment results, no rule-based traditional emailing, phone calling, or even mailing in order to establish a communication with the requesters and assign concrete tasks to a lab member. Furthermore, the constant evolution of new laboratory technologies and the growing amount of data generated represent nowadays a daunting challenge in the implementation of a proper data management system.

Because of the more complex panorama we are facing nowadays, the enhancement of laboratory workflows has become a 'must' for a lab-based service [1], even with very qualified technicians. A proper defined workflow is highly required not only because it facilitates the already mentioned massive data handling, but also because it can help to better accommodate those quality assurance requirements that are currently demanded by upper authorities in most present-day facilities to ensure highest quality of the service.

Specialized literature and scientific software vendors has traditionally drawn a line between LIMS (Laboratory Information Management System) and ELN (Electronic Laboratory Notebook) systems [2]. Whereas the former ones are used for labeling and tracking samples along a workflow, managing lab inventory (such as reagents and mediums), and monitoring instruments, the latter ones are used for annotating raw, intermediary, and final experimental data, results, and reports associated to the samples, as well as ensuring the sharing of guidelines and relevant information among co-workers.

These two hypothetical systems are ideally meant to coexist in a service and they should be connected or even reside in a unique or shared informatics infrastructure, so that the different user profiles should not mind about the logic behind and simply perform their specific tasks in a natural and easygoing fashion.

With the advent of the Internet and the growth of affordable and easy-to-setup local network installations, laboratory management and annotation systems could be extended beyond the very experimental workplace [3]. Data could then be centralized, exchanged and processed in an in-house or outsourced server, and users in front of thin terminals, or even devices and equipment themselves, could act as clients against the central server.

Although there are many client-server applications in the market, a very convenient approach is using web-based solutions. This way, any modern browser may suffice, without any need to install additional software .

1.1. MediaWiki as a convenient approach

By the early 2000s, wikis started to emerge and being adopted as centerpiece tools in collaborating and group learning environments both in the Internet and in private networks. The most notable example is the non-profit online encyclopedia Wikipedia, built over the PHP-written MediaWiki software [4].

MediaWiki, as a web based wiki collaborative application, provides consistent concurrency handling and data integrity, ensuring that a user edit cannot overwrite a

coincidental other user's addition, and a familiar interface so no extensive training is needed for learning how to input data.

There have already been different approaches taking advantage of Mediawiki possibilities in biological laboratory data management. One example is ArrayWiki [5], a global public repository of microarray data and meta-analyses that host many relevant images and their original experiments. Another one is OpenWetWare [6], an online open-science community of, mostly, 'wet-labs' where diverse information such as protocols or courses is shared. It also features a wiki-based electronic notebook. By default, MediaWiki offers to these systems an open and well-known collaborative environment where trackability and authorship can be followed in a fine-grain basis.

Parallel to this, during the last few years there has been an increasing interest in applying semantic web principles, meaning and concepts rather than the style and content of common-day web, to MediaWiki installations. One notable approach is Dbpedia [7], an effort to structure Wikipedia information. Articles and relationships such as categories, and other tagged a posteriori, are exported as Resource Description Framework (RDF) files, and these can be used for building up complex searches using SPARQL query language.

Another project is Semantic MediaWiki [8], an extension to MediaWiki platform that can be quickly installed in and add semantic capabilities to plain wiki installations. As a complement of Semantic MediaWiki, a recommendable addon is Semantic Forms, an extension that allows to create forms that can conveniently edit wiki pages in a structured manner through web forms and link their fields to semantic properties.

One of the better known examples of Semantic Mediawiki applied to the biological area is SNPedia [9], a wiki-based database of Single Nucleotide Polymorphism (SNP). Semantic properties addition enable that potential users cannot only perform common full-text searches, but also field specific ones, such as chromosome locations or the technology —for instance, Microarray model— used to generate the data.

Taking advantage of these new web technologies, we started to develop *Lab Service Wiki* - a wiki-based laboratory management system, This web application is concretely meant to handle relevant experimental information related to protein cloning, expression, and purification steps, thus providing wet-lab researchers with a proper tool to meliorate the lab workflow and keep control of the overall laboratory activity.

A test implementation is available at: <http://labservice.biocore.crg.cat>

2. Lab Service Wiki

Our target facility, Protein Service, consists of 1 head and 3 technicians. It works as an internal service of potentially around 100 users in a research center. There is an average of 4 requests per month, which can have from one up to one hundred or more associated experiments.

Before any wiki implementation, researchers used to submit their requests through PDF-based forms sent by email to the service. Once received, the responsible of the service could plan a meeting with the requesters to further discuss the project and gather additional information. After its outcome, the request could be accepted, modified or denied and one or more experiments run based on the given request. As a final result, the researcher could receive the service product (the purified protein itself) along with a report describing the most relevant experimental information.

The drawbacks of this approach were multiple: first, all information related to requests, samples, and experiments were not likely to be annotated in a standard way; second, all changes to original experimental data could not be tracked accordingly; third, data files generated during the analysis ended up being spread among different physical and virtual media, and if they were not gathered all together, they could get lost after report generation. This panorama represented a serious hurdle to any effective action to be performed by an evaluation third-party.

2.1. Implementation

As explained above, because of its simplicity of use and extensibility, MediaWiki posed as a firm candidate for hosting a system that fulfills the given requirements. Despite setting up a plain wiki system with a set of templates, customized extensions and cron-programmed or resident web robots was a feasible possibility, using Semantic MediaWiki, and other related extensions, greatly simplified the design. Pages could be “tagged” and linked semantically in multiple ways, so there was no need to use any other external application to process them first (e.g., parsing wiki syntax with regular expressions) in order to associate them to specific content of other pages (translated in Semantic MediaWiki as property values).

First of all, to grant the right access to the users in *Lab Service Wiki*, we created the following different user profiles: 1) the Administrator, responsible of the creation of new templates, users management and their training; 2) the Researcher, customer who can submit requests to the service using pre-defined templates, view the status of his/her requests at any time, and retrieve the study reports when the experiments are complete; 3) the Lab Manager, responsible of the service who can create, edit, delete new experiments, associated to submitted requests, using predefined templates; and, finally, the 4) Lab Members, expert technicians who can add, edit experimental data, but cannot create or delete experiments.

Once researchers obtain an account, which is assigned by default to a generic group, they can therefore log in and generate a request using a template form. Even though the latter seems to be equivalent to the original PDF-based version, it takes advantage of the Semantic Forms extension and therefore provides searchable fields and other additional functionalities.

The request form is the starting data seed for the upcoming workflow and the different fields are coupled to predefined semantic properties. In order to avoid any misuse, different restrictions were introduced at the logical and input level. At the logical level, we defined different data types associated to the properties, such as string, number or boolean, and which values can be allowed. At the input level, we could define the default input type, for instance text, checkboxes and the possible values, which could also be filtered by using regular expressions. This last option is especially useful for refusing incorrect alphabet characters in biological sequences (nucleotide or amino-acid ones).

On one hand, the form cannot be submitted if users fill non-allowed input values in a restricted field. On the other hand, in case there existed a page with a not-allowed value, Semantic MediaWiki would depict a warning icon next to the conflicting value. Therefore, this could be studied and addressed by the wiki administrator. So, both logical and input restrictions should need to be kept compatible and in sync for ensuring data integrity and quality.

Using this form, researchers are required to input both sample and project information, and therefore can submit the new request. This action creates a new wiki page, that can be subsequently modified by the submitter at any time before the lab manager has accepted it.

Meanwhile, the lab manager receives a communication by email that a new request has been submitted. He/she can eventually modify some information (for instance, during a personal meeting or communication with the requester) and finally accept or reject the current request, selecting the value of the field 'status' (available options are Pending, Accepted, Discarded, Closed). It is important to say that only the lab manager can modify the field status and decide whether to accept or not the requests.

Thanks to the semantic annotation of the pages and different parser and user functions provided by several MediaWiki extensions, it is technically possible to avoid that the requesters can make any later modification after the status has been modified. The same mechanism is also used to prevent that other users apart from the original requester may access to any other request.

Once a request is accepted, the lab manager can generate and associate several experiments to it. Experiment wiki pages will reside in a different namespace restricted only to lab members by default. However, whenever desired, the lab

manager may choose to open the access of specific experiments to the requester so they can follow closely the development of the request.

The experiment page is also handled through web forms and, for convenience, split in different tab pages matching to the different stages and type of analysis (in our concrete case: cloning/subcloning, expression screening, scale-up purification and mutagenesis).

Provided request data is automatically passed from its original request to the experiment page. When suitable, thanks to Semantic Forms capabilities, some request information fields are also mapped to a corresponding field in experiment forms. This way, we can keep sample information intact and the lab members can modify mapped fields according to their expertise, overriding so user's initial suggestion.

Experiment stages will be conditioned by the request, so if the user did not want to perform any mutagenesis analysis (and it was not changed by a manager either), that tab will not be displayed in the experiment interface.

During the experiments, different types of data and media files can be produced. These can be also attached to the experiment pages. Semantic Forms provides an easier way to use interface for uploading files than MediaWiki's defaults. In case of a huge amount of data, such as large size files, or a file format that might not fit well inside wiki pages, linking URLs is always a suitable option.

2.2. Workflow, reporting and user permissions

The workflow of the experiment can be managed in more detail if necessary (*see Figure 1*), usually highly desirable in bigger laboratories with several workers, by selecting the lab members once they start to work in the experiment or when they become in charge of a certain stage. They could be notified by email when their username is invoked in a value field. The completeness of certain tasks can also be notified by the responsible, so the manager (or the same lab members group) can move to a next analysis, which can often depend in the completion of a previous one. After all tasks are finished, the manager can choose to create more experiment pages from the request if the outcome is not as expected, enable open access to the experiment results to the researcher, or even generate a report page from the data of the very request and the results of the different associated experiments.

Thanks to conditional clauses introduced in the different templates of the wiki pages, the different statuses should remain coherent and synced along the interconnected pages: Request → Experiments.

That is, once all experiments are finished and the status of the request is marked as closed, no other experiment associated to that request can be generated. The very

manager would not be able to modify this, for instance, by creating another experiment page and generating a new report once a former one was considered definitive, unless that task is requested to be performed by the wiki administrator.

If tight group-associated permissions are followed and wiki administrator only intervenes according to well-defined guidelines, there is no easy way of forging or tampering the workflow. Pages, ideally only through web forms, can be edited either by plain users, lab members or lab managers depending on the permissions granted to a group for a certain namespace. MediaWiki permissions also permit differentiating between editing and page creating permissions. For instance, as explained above, only lab managers would be able to open a new experiment page, but lab members would be able to edit them in collaborative fashion despite they cannot create them themselves.

Moreover, the semantic logic behind the different page types is never intended to be writable by the mentioned groups. That means neither templates, nor forms specification nor semantic properties. Updating them should be under the sole responsibility of the wiki administrator. Since certain edits could break the consistency and interlinking of the semantic data, and consequently also the user-specific permissions and the workflow, these kinds of changes are supposed to be performed on a stage server using a sample subset of the existing data.

The traceability of the workflow is ensured with the default MediaWiki 'recent changes' option and also by checking individual pages history. These two options can be restricted for different roles and at the user level with conditional clauses using semantic queries. It makes sense to disallow access to 'recent changes' access to plain users.

Another application of using inline-searching feature of Semantic MediaWiki is getting detailed table-like reports about the status and the current stage of the experiments for the lab manager and the workload of the facility to the potential clients. Researchers can also track their own pending and past requests from their own user page.

Different blocks of information can be viewable by the different roles. Common users might only see the number of requests on queue so they cannot get impatient if theirs are not processed as fast as they might have desired, but lab members could need more details, such as the number of experiments associated to each request, and their creation date, so they can make up their own priorities.

Of course, apart from all the semantic linking possibilities, lab operators can still use the system as a lab notebook, not only by adding comments in experiments themselves, but also creating pages that might summarize the experience gained from the different experiments in order to improve existing SOPs.

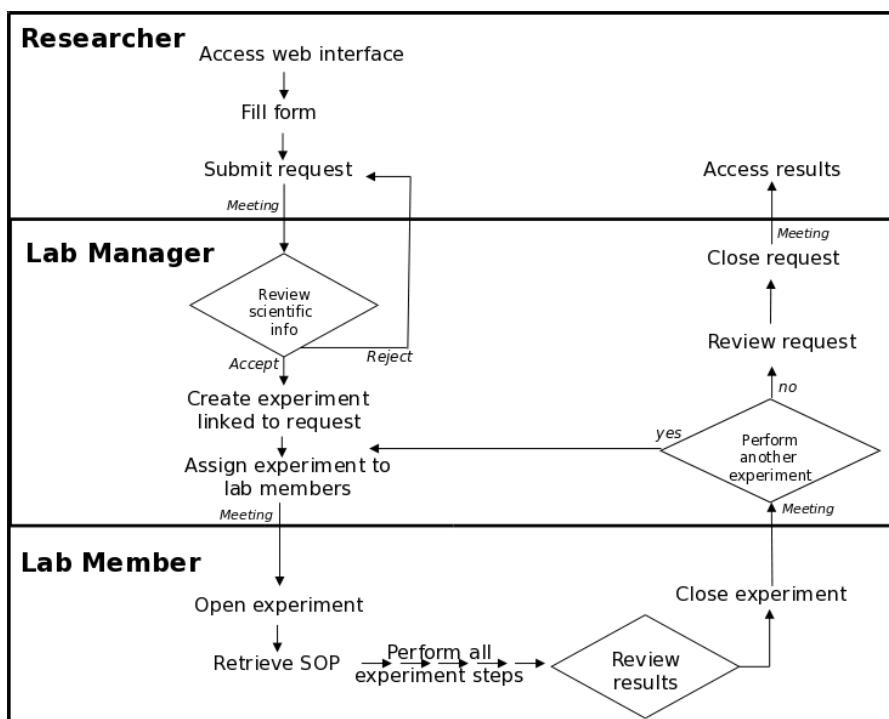


Fig 1. Simplified workflow of Lab Service Wiki.

3. Future advances

We could imagine about several features to improve the system. For instance, in the same facility, we could have access to an existing administration or catalog informatics system, which we could be interested in retrieving some data from. For this, we would need to use web robots against MediaWiki API, commonly written in Perl or Python scripting languages, which should mediate the connection to external databases and resources by querying them and updating accordingly the wiki. We could also trigger some applications to be run, for instance a sequence homology analysis. The result output could be linked externally within a wiki page, and also parsed in order to change a semantic field value.

Moreover, robots could also be used for automating some complex experiment workflows, so lab members do not need to generate hundred of pages from the same request if they expect to perform repetitive tasks. Since robots are to be put on move by triggers or in a periodical basis using system's cron, care must be taken to add the

necessary conditional logic requirements, in the wiki but also ideally in the very robot program, that may avoid any data breakage because of their failure. On the other, although they may have write rights, their editions should not be left unattended without validation by lab members.

As different experiments are performed, lab operators may notice that some data sets may repeat well enough to make them become a complex option value in a new simplified field that may encompass several previous ones. One solution for keeping backward compatibility with existing semantic definitions and, at the same time, trying to simplify the workflow (less fields to be filled), could be recurring to transclusion. In a few words, this means including the whole content of a page inside another one, so separate pages, as excerpts of information, can be kept apart for convenience and maintenance, and reused in the forms as many times as wanted.

We have centered the discussion upon a single research facility, but research institutions can also have many other hosted services not only in the same building, but spread in a campus, a city, a country or even all around the world. If different type of research analysis, using different equipment and in apart locations are to be performed upon the same sample, we might want sample information to be shared between the different experimental workflows. This is easier to be accomplished within the same MediaWiki installation, but it could also be worked out by using interwiki linking (as it is done between different languages versions of Wikipedia) and, more generally, thanks to well-designed web robots.

Semantic MediaWiki includes the feature to export existing relational information and semantic content as RDF files, which in turn could be analyzed by other software and used against other resources. And also, the other way around, external ontologies can potentially be imported into an existing Semantic Mediawiki installation. This may enhance the reporting we offer to the requestor by adding, for instance, functional genomics analyses by default thanks to Gene Ontology vocabulary [10]. Unexpected relationships may emerge if we datamine and process a bulk of experiments hosted in the system.

4. Conclusion

We described the implementation of *Lab Service Wiki* in our protein production service along with the proposed workflow to be used within the local research environment. Therefore, we consider that Semantic MediaWiki, as a concept empowered collaborative web system, is an excellent approach for designing a lab management and annotation system, which can be specifically adapted to the requirements of a modern day laboratory.

By using Semantic MediaWiki in contrast to a plain MediaWiki installation, we were able to link the content at a more detailed level that could be done by using only pages and categories. This way we assigned fine-grained permissions derived from semantic properties to active users and groups and, at the same time, both requesters and operators could benefit from specific searches and reports.

We also foresee many opportunities raised by the rational application of connecting different resources by web robots or by semantic content exchange.

Acknowledgments. We thank Oscar Gonzalez, Davido Castillo, and David Camargo for their technical support; as well as Luca Cozzuto, Francesco Mancuso, Ernesto Lowy, and our colleagues from other CRG core facilities for useful discussions.

References

1. Stephan, C., Kohl, M., Turewicz, M., Podwojski, K., Meyer, H.E., Eisenacher, M.: Using Laboratory Information Management Systems as central part of a proteomics data workflow. *Proteomics*, Jan 13 (2010)
2. Lims and ELN: 1 + 1 = 3. Scientific Computing. <http://www.scientificcomputing.com>
3. Ulma, W., Schlabach, D.M.: Technical Considerations in Remote LIMS Access via the World Wide Web. *J. Autom. Methods Manag. Chem.* 2005, 217--222 (2005)
4. MediaWiki. <http://www.mediawiki.org>
5. Stokes, T.H., Torrance, J.T., Li H., Wang, M.D.: ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. *BMC Bioinformatics* 9, Suppl 6:S18 (2008)
6. Waldrop, M.M.: Science 2.0. *Sci Am.* 298(5), 68--73 (2008)
7. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: 6. DBpedia – A Crystallization Point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Issue 7, 154--165 (2009)
8. Semantic MediaWiki. <http://www.semantic-mediawiki.org>
9. Cariaso, M., Lennon, G. SNPedia. <http://www.snpedia.com>
10. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genet.* 25, 25--29 (2000)