Extensible open source content management systems and frameworks: a solution for many needs of a bioinformatics group

Sean D. Mooney and Peter H. Baenziger

Submitted: 31st July 2007; Received (in revised form): 25th October 2007

Abstract
A common challenge for bioinformaticians, in either academic or industry laboratory environments, is providing informatic solutions via the Internet or through a web browser. Recently, the open source community began developing tools for building and maintaining web applications for many disciplines. These content management systems (CMS) provide many of the basic needs of an informatics group, whether in a small company, a group within a larger organisation or an academic laboratory. These tools aid in managing software development, website development, document development, course development, datasets, collaborations and customers. Since many of these tools are extensible, they can be developed to support other research-specific activities, such as handling large biomedical datasets or deploying bioanalytic tools. In this review of open source website management tools, the basic features of content management systems are discussed along with commonly used open source software. Additionally, some examples of their use in biomedical research are given.

Keywords: content management system; CMS; laboratory information management system; LIMS; website design; data management; open source software

INTRODUCTION
In general, content management systems (CMS) are software tools to manage webpages and websites. When our community thinks of CMS tools, they often think of wikis such as MediaWiki (http://www.mediawiki.org/). Indeed, use of community-driven database annotation has attracted much interest in science [1–4]. However, there are many open source tools for exchanging content, developing community-driven knowledge and maintaining websites. These systems provide a vast array of useful tools and plugins that can solve almost any administrative task. Other reviews of the many options have been previously written [5, 6]; here we provide an updated review with a focus on how a content management system may be used to facilitate the work of a biomedical researcher.

The ‘content’ in CMS is defined rather loosely; most systems, however, consider each of the following as examples of content: traditional webpages, wikis, discussion forums, social networking, event management, document management and groups (Table 1). One especially useful aspect of popular open source CMS tools is their extensible and modular design, allowing for development of new content capabilities. This characteristic enables groups with web programming expertise to develop custom applications for their group,
without having to maintain the underlying web architecture.

One of the benefits of CMS tools is that they facilitate website and content maintenance, allowing the development group to place more effort into actual development for their scientific needs, instead of the virtual administrative needs (authentication, authorisation, etc.). Though content management systems provide some level of maintenance-free development by simplifying administrative and system tasks, they are not without caveats and the wide range of choices can make choosing a CMS difficult. Complicating this further, security of these systems is not always up to a level required by biomedical research groups so informatics groups must understand the security of a system before choosing.

These tools are increasingly being used in biomedical research environments as both applications to manage scientific research data and intranet systems for managing laboratory group needs such as meetings and documents. Bioinformatic service providers may find that CMS can provide an attractive alternative to either home growing their own web infrastructure or using expensive laboratory information management systems (LIMS).

### Popular open source CMS programming languages and software

The lowest level solution discussed here is a content management framework (CMF). These are usually an extended programming language and an application programming interface (API) for development of websites. CMFs are generally database driven and can provide efficient development of any CMS feature. One commonly used CMF is Zope (http://zope.org/), an open source example founded on the Python programming language and based on the Zope Public License. Zope provides a powerful API that has been used as the basis for full CMS applications, including Plone (described below). Ruby on Rails (http://www.rubyonrails.org/) is another tool that, like Zope, is an open source (based on the MIT license and the Ruby license) CMF for development of web portals and content management applications. Another framework, Struts2 (http://struts.apache.org/), is fuelled by the recent merging of two prior frameworks and is now managed by the Apache Software Foundation. This java-based framework brings together the best of its predecessors: WebWork and Struts. CakePHP (http://www.cakephp.org/) is another choice that is useful for PHP developers. There are many web frameworks available; a comparison is available on Wikipedia (http://en.wikipedia.org/wiki/Comparison_of_web_application_frameworks).

A higher level approach appropriate for non-developers and those willing to extend existing comprehensive systems is a CMS application. These are generally fully functional and have easy-to-use web-based installation procedures. With an existing Apache web server, a MySQL database and super user access, the authors were able to setup several of these systems, each taking less than one hour.

The simplest CMS applications are single, complete and web-based, such as a wiki or blog. These tools are exceptionally useful in scientific and academic environments, but may lack interoperability with other content management system applications. One important group of these applications is Document Management Systems (DMS), which provide web-based document management. KnowledgeTree (KT, http://www.knowledgetree.com/) is a document management tool that has versioning features similar to the code management tool CVS (http://www.nongnu.org/cvs/) and an easy-to-use web interface. Licensed under the

### Table I: Types of CMS

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td>Management of documents, often includes version control</td>
</tr>
<tr>
<td>Groupware</td>
<td>Tools for managing collaborative projects in groups or teams. Can include mailing lists, calendars, contact lists and group-based content access control</td>
</tr>
<tr>
<td>Blogs</td>
<td>Published text that commonly becomes the root of a discussion thread</td>
</tr>
<tr>
<td>Forums</td>
<td>Discussion thread management</td>
</tr>
<tr>
<td>e-Commerce</td>
<td>Management of online sales and customers</td>
</tr>
<tr>
<td>e-Learning</td>
<td>Management of curriculum materials and courses, often called learning management systems (LMS)</td>
</tr>
<tr>
<td>Image galleries</td>
<td>Management of image galleries</td>
</tr>
<tr>
<td>Wikis</td>
<td>Collaboratively developed webpages and related content</td>
</tr>
<tr>
<td>Portals</td>
<td>Traditional CMS tools that manage basic web portals, and often support some of the features above</td>
</tr>
</tbody>
</table>

This is a list of the most commonly used CMS applications or stand-alone tools.
KnowledgeTree license, it is open source but has several requirements regarding web layout including the mandatory use and location of logo and copyright statements. One of the most useful features of KT is its ability to apply workflows. These workflows enable the creation of states of documents and transitions that determine how documents move through states. States can control permissions and actions can be defined that apply externally developed plugins. Workflows could be used easily for document review, publication status or even, with the help of an external application, data analysis. Another very popular open source tool is MediaWiki (http://www.mediawiki.org/wiki/MediaWiki), which is the PHP-based software behind the well-known Wikipedia (http://www.wikipedia.org/). MediaWiki provides easy-to-manage, community-driven webpage content. MediaWiki sites do not necessarily have to be publicly available; with plugins, group level access control is available. These tools are popular and deserve such popularity.

The next advancement of content management systems has been focused on extensibility to provide the functionality of the tools above with an infrastructure that provides a framework for interoperability. For the user, this means that wiki pages, blog posts, and group calendars can be tightly integrated. Several commonly used tools are described here; these are the finalists for the Packt 2006 Open Source CMS Award (http://www.packtpub.com/award). Generally, each of these tools have semi-automated install scripts, making installation very easy. To install one of the following tools, the zipped source file is uncompressed into a directory on the live webpage tree, permissions are set to make those pages writeable by the web browser and after directing a browser to the directory you follow the instructions. Webpage permissions may need to be reset and installation materials may need to be removed.

Drupal (http://www.drupal.org/) is one of the most popular open source CMS tools. Often used for community-driven websites, it runs such popular media websites as ‘The Onion’ (http://www.theonion.com/) and ‘Ain’t It Cool News’ (http://aintitcool.com/). Like many other projects, Drupal is built on Apache, MySQL and PHP and is extensible using modules. Hundreds of unique and useful modules are available, making it extremely flexible. A module tree is available on the Drupal website (http://drupal.org/project/Modules). Xoops (http://www.xoops.org/), like Drupal, is built upon PHP and MySQL and licensed under the GPL. Its name is an acronym representing ‘eXtensible Object Oriented Portal System’. Xoops is well supported, but there are fewer modules available than for Drupal. e107 (http://www.e107.org) is another open source CMS licensed under the GPL and developed in PHP. Modules are called plugins and this appears to have the least available number of plugins of all the tools discussed here. Plone (http://plone.org/) is built upon the previously described CMF, Zope. Plone has many plugins and is the basis for several derived tools including the educational open courseware management system, eduCommons (http://cosl.usu.edu/projects/educommons/), Joomla (http://www.joomla.org/), the winner of the 2006 Packt CMS Award, is a descendant of another open source CMS: Mambo. Joomla supports modules, called addons, that can be developed in PHP and has automated component, module, and package install scripts, making administration of the CMS completely webpage driven. In our experience, Drupal and Joomla are well supported and are a worthy solution to a group with PHP experience, while Plone fits well for groups with Python experience.

e-Learning or Learning management systems (LMS) are tools for the development of and the teaching of courses. LMS applications manage the administration of courses and all related course content. A popular open source tool used for this purpose is Moodle (http://moodle.org/), a GPL-licensed course management system based on Linux, Apache and PHP. Moodle supports ‘activities modules’ and ‘resources’ to build courses. Activities include forums and chats, assignments, quizzes, surveys and others. Resources are content objects for a course, including text or webpages and external links. Moodle is well supported with (according to their website), over 150,000 registered users, more than 22,000 sites in 75 languages and 160 countries. Another commonly used, notable tool is A1tutor (http://www.autor.ca/), a GPL-licensed LMS that is, again, based on PHP and MySQL. As an aside, it is worth noting that the primary corporate provider of LMS tools, Blackboard, has patent claims on many aspects of LMS tools. They have recently pledged not to enforce their patents against open source LMS projects (http://blackboard.com/) [7].
With so many choices and so many unique problems to be solved, it is difficult to choose a CMS appropriate for the application, particularly when multiple options exist in the preferred programming language. Table 2 summarises the open source content systems described in here.

### CMS use in biomedicine

CMS use in biomedical informatics is growing. Typically, these software systems are used as intranet tools, collaboration tools, conference websites and lab websites. Although not often used for managing biomedical data, the CMS architecture shows potential as a solution to managing shared, accessible data. Some examples of CMS tools in action are listed below.

#### Collaborative websites

Collaboration communities of researchers and educators are finding value in CMS systems. The Bioinformatics Curriculum Collaboration (BCC) currently uses KnowledgeTree to manage curriculum materials (Figure 1). Mentioned earlier, eduCommons, is a Plone-based system for managing materials developed for OpenCourseWare projects. Another important collaborative effort lies in the curation of knowledge. One example of this is WikiGene (http://andromeda.gsf.de/wiki/index.php/WikiGene), a gene annotation and curation system built on a wiki [8]. Use of these applications appears to be catching on in scientific research. A recent report in *Nature* described a company that is being developed based on community-driven projects using wiki-like tools [2].

#### Conference websites

Another common application of CMS tools is in managing and hosting conference or meeting websites. The Automated Function Prediction SIG (http://biofunctionprediction.org/) uses Plone and Drupal for its home page and the 2007 ISMB SIG [9]. Tools like Drupal provide out-of-the-box authentication, scheduling and calendaring, the ability to have a committee develop announcements, and extending the application to have registration functionality is relatively easy with intermediate PHP experience.

#### Databases of content

CMSs are often used in development of databases of links with metadata. The Health Sciences Library at the University of Pittsburgh maintains a comprehensive database of bioinformatics tools and resources called the ‘Online Bioinformatics Resources Collection’ [10]. This application, built upon Zope, manages an array of bioinformatic tools and resources.

#### Laboratory intranets

Finally, group intranet tools used in the Center for Computational Biology and Bioinformatics at Indiana University provide website support, a MediaWiki-based wiki for intranet site

---

**Table 2:** Commonly used open source website and content management tools listing their homepages

<table>
<thead>
<tr>
<th>Tool</th>
<th>Type</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zope</td>
<td>CMF</td>
<td><a href="http://www.zope.org/">http://www.zope.org/</a></td>
</tr>
<tr>
<td>Ruby on Rails</td>
<td>CMF</td>
<td><a href="http://www.rubyonrails.org/">http://www.rubyonrails.org/</a></td>
</tr>
<tr>
<td>CakePHP</td>
<td>CMF</td>
<td><a href="http://www.cakephp.org/">http://www.cakephp.org/</a></td>
</tr>
<tr>
<td>MediaWiki</td>
<td>Wiki</td>
<td><a href="http://www">http://www</a> mediawiki.org/wiki/MediaWiki</td>
</tr>
<tr>
<td>KnowledgeTree</td>
<td>DMS</td>
<td><a href="http://www.knowledgetree.com/">http://www.knowledgetree.com/</a></td>
</tr>
<tr>
<td>Drupal</td>
<td>CMS</td>
<td><a href="http://www.drupal.org/">http://www.drupal.org/</a></td>
</tr>
<tr>
<td>Joomla</td>
<td>CMS</td>
<td><a href="http://www.joomla.org/">http://www.joomla.org/</a></td>
</tr>
<tr>
<td>e07</td>
<td>CMS</td>
<td><a href="http://www.e07.org/">http://www.e07.org/</a></td>
</tr>
<tr>
<td>Xoops</td>
<td>CMS</td>
<td><a href="http://www.xoops.org/">http://www.xoops.org/</a></td>
</tr>
<tr>
<td>Plone</td>
<td>CMS</td>
<td><a href="http://www.plone.org/">http://www.plone.org/</a></td>
</tr>
<tr>
<td>Moodle</td>
<td>e-Learning</td>
<td><a href="http://www.moodle.org/">http://www.moodle.org/</a></td>
</tr>
<tr>
<td>Atutor</td>
<td>e-Learning</td>
<td><a href="http://www.atutor.ca/">http://www.atutor.ca/</a></td>
</tr>
<tr>
<td>eduCommons</td>
<td>e-Learning</td>
<td><a href="http://cosl.usu.edu/projects/educommons/">http://cosl.usu.edu/projects/educommons/</a></td>
</tr>
</tbody>
</table>

The types described here are as follows, a CMF is a content management framework; a Wiki is a community or group editable webpage; a DMS is a document management system; a CMS is a content management system or portal; and tools of type e-Learning are curricula and course management solutions.

---

**Figure 1:** The Bioinformatics Curriculum Collaboration (BCC) Database based on the KnowledgeTree software. KnowledgeTree manages documents in an easy-to-use web-based file browser. Here, the BCC database stores curriculum materials.
development, and KnowledgeTree for management of manuscripts, posters, presentations and grants. Figure 2 shows an example of a MediaWiki page.

**How to choose a tool?**

Choosing the most appropriate tool can be a laborious process. First and foremost, security should be evaluated as well as the aptitude, availability and coverage of the user support community. Next, be sure the system you choose will work with minimal extension development. One useful site, OpenSourceCMS (http://www.opensourcecms.com/), provides demo websites of many open source CMS tools. These demo sites refresh every two hours, enabling use of these systems in their default install state (or close to it). Also to be considered is the number of supported, externally developed modules available. Larry Wall, creator of the programming language Perl, is attributed with saying, ‘easy things should be easy and hard things should be possible’[11]. The same applies to evaluating CMS systems: setup should be easy and developing new extensions should be possible.

**CONCLUSIONS AND RECOMMENDATIONS**

CMS are becoming widely used in managing web content. These tools provide a platform for construction of web-based tools that are easier to manage, more scalable, and more flexible than traditional home-built web tools. It is difficult to make recommendations on which tool is the best or the most useful for a given application. A user’s development language and platform preference is generally the biggest determinant of which tool to use. If the group is experienced in the Python programming language, Zope would appear to be a good choice for developing an application de novo while Plone would be useful for extending an existing CMS. If the group is experienced in PHP, Drupal or Joomla are well supported and have active communities. Additionally, Ruby on Rails is well supported and has an active community. For specific applications, KnowledgeTree is an outstanding DMS, MediaWiki is an excellent, well-supported wiki tool and Moodle is a well-supported e-Learning tool. All of these are built upon PHP. Hopefully, this introduction provides a useful overview of the most popular CMS tools, describes their common features, and suggests some ways a CMS can be applied in biomedical research. If our community of researchers would support extension development for one or a few well-supported CMS tools, the benefit to the community would be great, and would reduce the cost and time required for developing web applications.

**Key Points**

- Open source content management systems are a powerful source of website and web content management tools.
- The use of CMS tools in biomedical informatics is growing; however, the landscape of available tools is complicated and diverse.
- Use and development of CMS application plugins enables customisation of a CMS.

**Acknowledgements**

We would like to thank Noel Southall and Brandon Peters for helpful comments. Funding provided by K22LM009135 (PI: Mooney) and INGEN. The Indiana Genomics Initiative (INGEN) is funded in part by a grant from the Lilly Endowment.

**References**


