



Open Source Web Content Management in Java

Release 1.0

February 2008

Open Source Web Content Management Options in Java

Seth Gottlieb

Version 1.0, Workgroup License
Copyright © 2007 Content Here, Inc.

License Agreement and Disclaimer

Workgroup License

This report is licensed under a "Workgroup License" that allows your company to make this report available to up to ten (10) staff members. It may not be shared with customers or copied, reproduced, altered, or re-transmitted in any form or by any means without prior written consent. Any rankings or scoring information may not be used in promotional materials.

Disclaimer

This report is intended to be an overview of the technologies described and not a recommendation or endorsement of a specific platform or technology strategy. The most appropriate platform for your use depends on the unique requirements, legacy architecture, and technical capabilities of your organization. Content Here, Inc., cannot ensure the accuracy of this information since projects, vendors, and market conditions change rapidly. Content Here, Inc., disclaims all warranties as to the accuracy or completeness of the information in this report and shall have no liability for errors, omissions, or inadequacies in the information presented.

All Rights Reserved. Not for Redistribution.

Acknowledgements

Thanks to the following people for reviewing sections of this report for accuracy

- Elie Auvray (Jahia)
- Kevin Cochrane (Alfresco)
- Arje Kahn (Hippo CMS)
- Alexander Kandzior (OpenCms)
- Boris Kraft (Magnolia)
- Steven Noels (Daisy)

Jennifer Gottlieb provided copyedit services and general encouragement to help me complete this report. Glenn Barnett customized the XSL style sheets used to format the report.

Cover Art

The photograph used on the cover was taken by Tan Quang Tuan [<http://www.flickr.com/photos/e8club/>] and published under the Creative Commons Attribution 2.0 License on Flickr.

Table of Contents

1. Introduction	1
The Demand for Open Source Java Web Content Management	1
The Need for This Report	2
Organization and Methodology	2
2. Open Source WCM Marketplace	5
State of the Industry: Web Content Management	5
Market Characteristics and Trends	5
Core Enterprise Requirements	8
Market Summary	12
Open Source Market Segmentation	13
Community Open Source	13
Institutional Open Source	16
Commercial Open Source	17
3. Product Evaluations	21
Informational Brochure	21
What Makes a Good Informational Brochure Platform?	21
Informational Brochure Platform Market Overview	26
Apache Lenya 2.0	27
Daisy 2.1	43
Magnolia 3.5 Enterprise	62
OpenCms 7.0.3	78
Informational Brochure Platform Summary	95
Web Content Management Framework	97
What Makes a Good WCM Framework?	98
WCM Framework Market Overview	102
Alfresco 2.2 WCM	104
Hippo CMS 6.05.02	123
Jahia Enterprise 5.0	139
WCM Framework Market Summary	154
Round Up	156
Comparing with Commercial Products	157
Selecting a CMS and Beyond	159
Glossary	161

List of Figures

3.1. Lenya Architecture Diagram: Use Case Framework	30
3.2. Lenya Screenshot: Edit Menu	32
3.3. Lenya Screenshot: BXE Editor	33
3.4. Lenya Screenshot: Kupu Editor	34
3.5. Lenya Screenshot: Editing Structured Content in BXE	35
3.6. Lenya Screenshot: Site Tab	35
3.7. Lenya Screenshot: Lenya Localization	36
3.8. Lenya Screenshot: Image Dialog	37
3.9. Lenya Screenshot: Workflow Syntax	38
3.10. Lenya Screenshot: Edit Permissions	39
3.11. Daisy Architecture Diagram: Daisy Architecture	46
3.12. Daisy Repository Server Architecture	48
3.13. Daisy Screenshot: Defining Field Types	50
3.14. Daisy Screenshot: Content Actions Menu	51
3.15. Daisy Screenshot: Link Builder	53
3.16. Daisy Screenshot: Editing a Navigation Document	54
3.17. Daisy Screenshot: Editing Image Properties	55
3.18. Daisy Screenshot: Daisy Diff	56
3.19. Daisy Screenshot: Defining ACLs	57
3.20. Daisy Screenshot: Faceted Browsing	58
3.21. Magnolia Screenshot: Configure Subscribers	66
3.22. Magnolia Screenshot: Browsing in AdminCentral	67
3.23. Magnolia Screenshot: Page Layout	68
3.24. Magnolia Screenshot: Edit Dialog	69
3.25. Magnolia Screenshot: Localized Edit Dialog	70
3.26. Magnolia Screenshot: Site Designer	73
3.27. Magnolia Screenshot: Configure Cache	74
3.28. OpenCms Screenshot: Editing Structured Content	81
3.29. OpenCms Screenshot: Configure Search Index	82
3.30. OpenCms Screenshot: Database Replication Module	83
3.31. OpenCms Screenshot: OpenCms Workplace Interface	84
3.32. OpenCms Screenshot: Editing XML Pages	85
3.33. OpenCms Screenshot: Link Checking	86
3.34. OpenCms Screenshot: Localizing Content	86
3.35. OpenCms Screenshot: Direct Edit Interface	87
3.36. OpenCms Screenshot: Insert Image	88
3.37. OpenCms Screenshot: OCEE LDAP Connector	89
3.38. OpenCms Screenshot: Content Tools	90
3.39. Architecture Diagram: Structured Publishing	100
3.40. Alfresco Architecture Diagram	106
3.41. Alfresco Architecture Diagram: Repository Services	108
3.42. Alfresco Screenshot: Web Content Properties	110
3.43. Alfresco Screenshot: Browse Site View	112
3.44. Alfresco Screenshot: Sand Boxes	113
3.45. Alfresco Screenshot: TinyMCE Formatting Buttons	114
3.46. Alfresco Screenshot: Image Position Dialog	114
3.47. Alfresco Screenshot: Workflow Dialog	116
3.48. Alfresco Screenshot: Managing Permissions	117
3.49. Alfresco Static Deploy Model Diagram	118
3.50. High Level Hippo Architecture Diagram	125
3.51. Hippo Architecture Diagram: Forms Generation Architecture	127

3.52. Hippo Screenshot: Taxonomy Browser	128
3.53. Hippo Screenshot: Document Browse Interface	130
3.54. Hippo Screenshot: Xopus Editor Integration	131
3.55. Hippo Screenshot: Managing Permissions	133
3.56. Hippo Screenshot: To Do List	134
3.57. Hippo Screenshot: JSF Repository Browser Demo	135
3.58. Jahia Architecture Diagram: Distributed Architecture	143
3.59. Jahia Code Sample: Content Type Definition	143
3.60. Jahia Screenshot: In-Context Content Management	144
3.61. Jahia Screenshot: Forms Based Editing	145
3.62. Jahia Screenshot: Advanced Search Form	146
3.63. Jahia Screenshot: Version Differences	147
3.64. Jahia Screenshot: Workflow Approval Page	148
3.65. Jahia Screenshot: Field Level Access Control	149
3.66. Jahia Screenshot: Personal Portal Page	150

List of Tables

1.1. High Level Summary of Products Reviewed	3
1.2. Scoring Key	4
2.1. How Community Projects are Governed	13
2.2. Commercial Open Source Revenue Sources	18
3.1. Informational Brochure Strengths and Weaknesses	26
3.2. Lenya Project Overview	28
3.3. Lenya 2.0 Summary	41
3.4. Daisy Project Overview	44
3.5. Daisy 2.1 Summary	60
3.6. Magnolia Project Overview	63
3.7. Magnolia 3.5 Enterprise Summary	76
3.8. OpenCms Project Overview	79
3.9. OpenCms 7.0.3 Summary	93
3.10. Informational Brochure Score Summary	96
3.11. Informational Brochure Strengths and Weaknesses	102
3.12. Alfresco Enterprise Project Overview	105
3.13. Alfresco 2.2 Summary	121
3.14. Hippo CMS Project Overview	124
3.15. Hippo 6.05.02 Summary	137
3.16. Jahia Enterprise Project Overview	140
3.17. Jahia 5.0.3 Summary	152
3.18. Informational Brochure Score Summary	155

Chapter 1. Introduction

The Demand for Open Source Java Web Content Management

Not long ago, companies looking for an open source Java web content management system (WCM) had limited options. While the open source content management system (CMS) community as a whole was thriving, most of the activity was on the PHP and Python stacks. The main Java options were Apache Lenya and OpenCMS. If you wanted a simple, widely used technology that your users would like, neither of these options looked very attractive. This state of the market was frustrating for many companies that had standardized on the Java platform and wanted to take advantages of the opportunities afforded by open source content technologies.

The building blocks have been available for a long time. The Java world is rich with frameworks that provide core services like persistence, access control, data validation, and presentation. Many companies have used these components to build custom systems that fit their needs. However, these homegrown systems tend to languish without a continuous commitment to maintenance and enhancement. Adding certain core content management features can be prohibitively complex. For example, adding versioning and/or localization to a data model that was not originally designed for it can disrupt the whole application. Furthermore many in house development teams building these systems do not have the wealth of subject matter experience that a dedicated content technology development team would have. Lessons learned can only be applied in the next release of the application - if there is one.

The state of the market is rapidly changing. More products are emerging and some of the older projects are seeing a resurgence. The momentum behind Java web content management (WCM) technologies started to surge in early 2006 when open source business applications began to get the attention of enterprise buyers who were having success with infrastructure products like Linux, Apache, and MySQL. Java was a natural requirement for large enterprises who had standardized on the language. At the same time, commercial open source vendors were starting to notch up their offerings and connect with these interested buyers. Many companies are reporting successful implementations using a new breed of Java WCM technologies. If you were disappointed the last time you looked for a Java web content management platform, it may be time to look again.

Companies that have successfully implemented solutions based on these platforms talk of lower project start-up costs and similar (not greater) integration and maintenance costs. Typically, they have strong development teams or rely on systems integrators to manage the systems for them. These same companies tend to have a history of frustration with commercial software because they do not feel that the value is commensurate with the licensing costs (because they spend so much time or money doing integration work) or they feel under-served by technical support and would like to be less dependent.

Companies have found the greatest leverage using open software to power basic informational web sites and also to provide content management services to highly dynamic, transactional or interactive web applications. As you will see in the pages of this report, the Java open source content management marketplace is rich with options in these categories. While the Java products still lag PHP and Python based systems in terms of social media oriented features and community size, they have good support for the more fundamental content

management functionality and several of these products offer the assurance of commercial support packages.

The Need for This Report

Now that there are a growing number of viable open source web content management choices available, a technology decision maker needs to be more informed than ever. Even if the best technology fit is a commercial product, the technology decision maker now needs to be able defend his choice of commercial software by demonstrating a knowledge of open source alternatives that were rejected. The answer "we looked at open source and it was all bad" is becoming weaker and weaker as a response to a challenge to consider open source. Alternatively, if an open source system is selected, it pays to have a deeper understanding of how the technology works than you can get away with commercial software. If the information is out there, there is no one to blame for your ignorance. Just like you don't want to announce your commercial software selection after the vendor has filed for bankruptcy, you don't want to select an open source project that has governance or process issues.

Despite the increasing relevance of open source software, traditional analyst firms have been slow to cover this sector of the market. Their standard evaluation processes do not work as well on open source communities or businesses with small or non-existent sales and marketing budgets. The buying practices of most customers are equally reliant on sales and marketing efforts by software vendors and rendered equally effective by the open source model.

It is not that open source projects are secretive. In fact *more* information is available because coordination and communication usually happen out in the open. It is just that the information is spread thinly across many sources and people. Compilation and interpretation take a lot of work and a different set of skills than your typical career analyst. In order to understand an open source application, you need to use it, configure it, and interact with the community (actively and passively). The source code itself also contains valuable information about the development standards and history of the project. It takes time to learn the personalities and group dynamics of the community. Not that it wouldn't be nice to know all this information about commercial software - it certainly would. It's just that commercial software doesn't allow you that access.

Organization and Methodology

This report evaluates seven open source Java WCM systems: Alfresco, Apache Lenya, Daisy, Hippo, Jahia, Magnolia, and OpenCMS. You can save yourself the trouble of leafing through the pages to find a "blue ribbon" winner or a magic quadrant. There is no universally superior product. Each has its strengths and weaknesses. The astute reader will notice that not all of the products reviewed in this report qualify under the Open Source Definition [<http://www.opensource.org/docs/osd>]. Many of the newer, fast moving entries in the open source marketplace are backed by companies (often venture-funded) that are exploring different models to build viable businesses out of free software. The upside to the customer is that there is more potential for accountability (and continuity) with a company than a nebulous community. The potential downside is that, if the business fails, the product will probably fail too. Potential buyers of these commercial products should pay close attention to the business model. The business should be viable but the pricing should be proportionate to the value that is provided. For the commercial open source products reviewed in this report, there will be a commentary on the business model and success of the company.

As in my report Content Management Problems and Open Source Solutions [http://contenthere.net/articles/optaros_cmsReport_012206_sgg.pdf], I take the approach of

positioning the each project in categories of use where it typically excels. In the above mentioned report, I used the categories: Informational Brochure Site, Online Periodical, Collaborative Workspace, and Online Community. The projects described in this report fall into two categories: Informational Brochure and Web Content Management (WCM) Platform. There is also a discussion of the overall web content management marketplace and how open source software fits in. Many of the products reviewed in this report are commercial open source - meaning that a software company develops the product as part of their business strategy. For these products, I discuss how the company makes money off the software: whether they sell a commercial version of the software that is better than the free version ("tiered product") or whether the revenue comes entirely from selling support services for the free version.

Table 1.1. High Level Summary of Products Reviewed

Platform	Version	Product Type	Started	Primary Use	Also Used
Alfresco WCM	2.2	Commercial: Tiered Product	2005	WCM Framework	Informational Brochure
Apache Lenya	2	Community	2002	Informational Brochure	Multi-site hosting
Daisy CMS	2.1	Commercial: Support	2003	Informational Brochure	Knowledge Base, Documentation Site
Hippo CMS	6	Commercial: Support	2000	WCM Framework	
Jahia Enterprise	5	Commercial: Tiered Product	1998	WCM Framework	Corporate Intranet
Magnolia Enterprise	3.5	Commercial: Tiered Product	2003	Informational Brochure	
OpenCms	7	Community/ Commercial	1999	Informational Brochure	

For each of the projects reviewed in this report, I have subscribed to the mailing list and monitored the volume and nature of the activity. I have talked to users of the software. I have built prototypes that involve defining content types, setting permissions, and developing layouts. To ensure factual accuracy, each evaluation has been reviewed by a project committer or company officer.

Within each evaluation, I discuss the architecture and integration potential, usability factors, the community, and how the project seems to be trending. For the business oriented reader, the content contribution and presentation sections describe how the application is used to manage content and what type of visitor facing functionality is possible. For the technical reader, the architecture and development sections describe how the product works behind the scenes and can be configured and integrated. Although, I do not give overall ratings of the product, I do rate each product along certain common criteria.

Table 1.2. Scoring Key

Score	Explanation
●	Not available
◐	Below average
○	Average
◑	Above average
●	Exceptional