

Rochester Institute of Technology

RIT Digital Institutional Repository

Books

2009

The Open Publishing Guide: Creating a self-publishing website using open source content management tools

Patricia Albanese

Matthew Bernius

Sean Conklin

Alvin Crespo

Follow this and additional works at: <https://repository.rit.edu/books>

Recommended Citation

Albanese, Patricia; Bernius, Matthew; Conklin, Sean; and Crespo, Alvin, "The Open Publishing Guide: Creating a self-publishing website using open source content management tools" (2009). Accessed from <https://repository.rit.edu/books/23>

This Full-Length Book is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

The Open Publishing Guide: Creating a Self- Publishing Website Using Open Source Content Management Tools

By

Patricia Albanese

Gannett Distinguished Professor

School of Print Media

Matthew H. Bernius

Visiting Professor

School of Print Media

Sean Conklin

Alvin Crespo

Rachael Gootnick

Students

Rochester Institute of Technology

A Research Monograph of the
Printing Industry Center at RIT

No. PICRM-2009-03

The Open Publishing Guide: Creating a Self-Publishing Website Using Open Source Content Management Tools

By

Patricia Albanese
Gannett Distinguished Professor
School of Print Media
Rochester Institute of Technology

Matthew H. Bernius
Visiting Professor, School of Print Media
Rochester Institute of Technology

Sean Conklin
Alvin Crespo
Rachael Gootnick
Students and Open Publishing Lab Researchers
Rochester Institute of Technology



A Research Monograph of the
Printing Industry Center at RIT
Rochester, NY
January 2009

PICRM-2009-03

© 2009 Printing Industry Center at RIT— All rights reserved.

The research agenda of the Printing Industry Center at RIT and the publication of research findings are supported by the following organizations:



Table of Contents

Abstract.....	3
Introduction.....	4
Project Methodology	5
Defining the Project	5
Developing the Site Structure.....	7
Choosing a Technical Platform.....	9
Gathering and Authoring Content.....	10
Visual Design and Building.....	11
Launching, Tracking, and Maintenance	12
Findings	13
References.....	14

Abstract

The goal of the RIT Open Publishing Guide project is to create a web-based repository of information about self-publishing, including publishing tutorials. As this was primarily a production-based project as opposed to research, this report will focus on the process taken to develop the final website. It will discuss the steps necessary to produce a website of the scope and scale of the Open Publishing Guide, provide an overview of the production methodology, discuss the process by which content was acquired for the site, explore the web platform used to deliver the content, and, finally, outline the high level production workflow of the site. This report will be of use to those interested in deploying large scale, dynamic, websites and individuals interested in consumer-focused, Internet print-on-demand services.

Introduction

In recent years, advancements in desktop publishing and digital printing technology have made it possible to easily publish low cost personal books. During this same time, we have seen the rapid digitization of various content assets (text, images, audio, and video) and their release into the public domain. However, few people are aware of the wide range of tools and content that they have access to. Even fewer understand how to use all of these options to create and publish their own work. Thus there is an opportunity to create an education resource that remove barriers to access this technology and help empower people to self-publish.

At the 2007 Printing Industry Center Symposium and Planning Meeting, the Open Publishing Lab at the Rochester Institute of Technology proposed the development of a website that would serve a “beginners’ guide” to self-publishing. Dubbed “The Open Publishing Guide”,¹ or OPG for short, this repository site includes information about freely available, open source text and image assets and software programs. It also guides users through the publishing process, discussing each step of the publication process. This includes documenting how to connect various software offerings, how to sell books on Amazon.com, and understanding copyright and creative commons.

This report documents the production of the OPG. It first outlines the overarching methodology that was used to deliver the project. The report discusses how the objectives of the research project were converted into specific site requirements. Next, there is a discussion of tactics used to form strategic partnerships and gather the necessary content to populate the guide. From there, the report turns to content management systems, and the decisions that led to selecting the open source Drupal platform to deliver the OPG’s content. The report then outlines the overall production workflow used to create the site and how the site will be tracked. The final section of the report focuses on the findings that emerged from this research and suggests some potential avenues for future research.

The production-focused nature of this report will provide insights for those seeking to better understand the considerations that go into creating and maintaining a robust, web-based content repository. It is intended to be used to gain understanding of a general framework for website development using a content management platform, as opposed to an in-depth exploration of any specific stage of the web production workflow.

This report also contains insights into the depth and breadth of open source publishing applications and online self-publishing services. One will also find out about drivers and the challenges facing individuals who are interested in self-publishing. For readers interested in this information, the *Defining the Project* and the *Findings* sections will be of the most interest.

1 - URL: <http://opg.cias.rit.edu>

Production Methodology

At its core, any web development project is about balancing (1) the delivery of the best possible experience for the user with (2) the needs of the organization delivering the website and (3) the technological features and limitations of the software platform that the website will run on. The success of a project is therefore dependant on choosing a workflow that, from the start, will establish and a maintain the necessary equilibrium.

The workflow used to deliver the OPG was modeled off of the one laid out in Goto and Cotler's (2004) *Web ReDesign 2.0 : Workflow that Works*. Using their model, the production of the site went through five major stages:

1. Defining the Project
2. Developing the Site Structure
3. Designing the Visual Interface
4. Building & Integrating
5. Launching, Tracking, and Maintenance

Defining the Project

The goal of the first stage is to begin to understand both the needs of the users of the site and the needs of the organization. It is important to establish these first two areas *before* considering the technology that will be used to implement the site. This “needs based” focus, often referred to as *User Centered Design*,² helps avoid the temptation to design a solution around what technology is available or what “looks good” to a graphic designer.

The proposal for the OPG included the following specific organization goals:

- Defining the scope of the existing barriers for non-technical users to self-publish.
- Documenting current standards and trends in self-publishing and on-demand publishing.
- Developing a set of processes and options that provide solutions for the non-technical user to self-publish.
- Developing a common resource for public domain and freely available content to use in the self-publishing process.

In addition to those goals, the team also identified a number of other important require-

² - Note that in this case “design” means “problem solving.” So, as we will see, rather than just addressing the visual appearance of a website, user centered design touches every aspect of the website's production: from visuals, to writing, to the implementation of the content management system.

ments. The most significant of these was the goal to ensure that the OPG would be updated on a regular basis. Achieving this meant the website would need to be editable by anyone with the correct level of access, including individuals with limited technical skills. The OPG project manager captured these and all other requirements (branding, budget, etc.) in a requirements document.

The next step was to define the user requirements. In order to accomplish this, the team recruited a number of *Subject Matter Experts*, or *SMEs*. These individuals were familiar with the publishing process and regularly provided assistance to people seeking to self-publish. To fill these positions, the OPG team established a strategic partnership with RIT Libraries and the Open Publishing Lab at RIT, taking advantage of their organizational knowledge in this area. Additionally, the team also sought out self-publishing expert Frank Cost, Assistant Dean of the College of Imaging Arts and Sciences at RIT.

The OPG team began to meet regularly with the SMEs to identify the wide range of content, assets, and information that was available across the web. This led the team to an important decision: rather than attempting to replicate what was already available, the OPG would serve as a “golden thread”—a hub linking all of these resources together, and providing the content necessary to fill in the gaps.

During these meetings, the team began to establish general psychographic profiles for site visitors. It was decided that the average user to the site would have little to no experience with “industry standard” software (Photoshop, InDesign, etc.). They could be starting a project or already have some assets finished (text, photos, illustrations), and their projects could range from creating a family history, to a recipe book, writing a novel, or creating a photo book to record a child’s soccer season. Some would be coming to look for solutions to specific problems while others would be looking for guidance through the entire process. In the end, two main types of users emerged:

1. *Basic Users* – More interested in what goes on the page than how it gets there. These individuals are looking for “low pain” solutions and some friendly handholding and encouragement. They’re not power software or computer users, and are not looking to learn new software. They’re not necessarily interested in understanding resolution. Instead, they are just looking for someone to tell them what they should do to end up with a book.
2. *Advanced Users* – They care about the craft of what they are creating and want to have more control over the design of the final product. They like to learn new software packages and want to know more about publishing terms. They are interested in understanding all of the possibilities and then choosing the path that is right for their project. They also are interested in what other resources are available across the web to help them create the “perfect” book.

Based on these categories, the OPG team began to consider the content needs of each group. In particular, the team and SMEs focused on a core challenge: approximately 85% of the people who begin a self-publishing process do not complete it. The goal was

to populate the site with content and applications that would make it easier for individuals to successfully complete the publishing process.

A critical early decision was to focus on non-traditional publishing software in order to meet the needs of basic users. This meant that the team would need to document how commonly installed programs like Microsoft® PowerPoint could be used in lieu of programs such as Adobe® InDesign. The choice was also made to focus on freely available open source³ software whenever it was available, as it was assumed that both groups of users would not necessarily have the resources, or the desire, to purchase software. The team also assumed that both groups would be interested in having starter templates that they could use for their books.

The team also began to consider the various types of books that visitors to the site would be interested in creating. At first the group attempted to use traditional publishing categories: cookbooks, novels, picture books, etc. It was soon recognized that there were too many possibilities to easily list. Instead the decision was made to break the book projects into simpler options such as “book with pictures” and “book without pictures.”

Another important discovery process that began during this phase, and continued in various forms through the *Visual Design* and *Building* phases, was actively benchmarking what other sites were doing. This included surveying a wide range of sites dedicated to self-publishing as well as other informative and educational sites. This research allowed the team to identify best practices in those spaces. Also, in the case of print-on-demand sites, it led to a better understanding of the range of options that are available on the web today. Those learning will be discussed later in the *Findings* section of this report.

In web development, the requirements document falls into the category of a “living” document. This means that it will be updated throughout the production process as the project evolved. As the team’s understanding of its target requirements evolved, so did the requirements document.

Developing the Site Structure

As content was beginning to be identified, work began on establishing the overall structure of the site. The informational design of the site takes place long before any visual design work is done. The first step in this process is known as “bucketing” – taking each proposed piece of content and placing them with similar content (putting each group into a category “bucket”). This process allowed the team to see the different ways that the content⁴ could be grouped.

3 - Unlike copyrighted software developed at companies like Adobe and Microsoft, open source software applications (including editable source code) are freely available to anyone to download and use. There are numerous open source publishing software packages available today across the Internet.

4 - At this phase of the project, the content is not thought of in terms of “web pages.” This is because multiple pieces of content could all exist within the same web page, and individual pieces of content could also exist across multiple web pages.

Once general categories were agreed upon, the next step was to consider how each category would fit into the overall site structure. To accomplish this, the team wrote each category on individual index cards and then arranged the cards on a large table in a tree hierarchy. The cards were ordered and reordered to test different content arrangements. During this iterative process, various gaps in content were identified and then filled.

In order to test the different content arrangements, the team came up with a variety of tasks that individuals from each user type might try to complete. For example, an advanced user might be interested in finding out the steps necessary to get a book listed on Amazon, while a basic user could come looking for an explanation of what resolution means. Using the psychographic information, the team put themselves in “the shoes” of these users and imagined how they might go about looking for answers. This process was used to identify various site navigation paths, and also to think about how different sections might be named.

In the end, it was decided that the site would be divided into five sections:

1. *About* – Information about the guide and how to use it.
2. *Resources* – Links to content and assets across the web and a glossary of publishing terms.
3. *Community* – A section of the site that would encourage interaction between users, and also aggregate up-to-the-minute information about news developments within the world of self-publishing.
4. *Self-Publishing Advisor* – An interactive directory of online print-on-demand services.
5. *Publishing Step-By-Step* – A series of articles that would sequentially cover each stage of the publishing process (Plan, Create, Layout & Design, Publish).

Working with this process, the team not only established the overall navigation and linking structure, but also developed the content framework for each section of the site. At the end of this process, the project manager was able to deliver a list, or content map, of every piece of information that needed to be gathered or written for the website. The content map, in turn, is used to understand how “big” the website will be and what resources (time and personnel) will be needed to deliver the content.

Once that framework was in place, the project was almost ready to move from a linear production model (*A* needs to be accomplished before *B*, *B* before *C*, *C* before ...) to a parallel production model where work could be done on multiple aspects of the project at the same time. All that remained was to decide on what technologies would be used to deliver the site.

Choosing a Technical Platform

Now that the User and Organization requirements were documented, the team was finally in a position to make well grounded decisions about which web technologies to use. As documented earlier, an important organization requirement for the OPG was that it must be editable by people who have minimal web experience. Additionally, based on the content map, it was clear that the OPG would have dozens of pages, and could potentially grow into the hundreds. Thus, any technical solution also had to allow for the easy addition, management, and deletion of web pages. Finally, due to the size of the site, it was also a requirement that it must be easy to change the overall visual layout of the site without having to edit each page individually.

All of these requirements meant that, rather than creating pages by hand using Hypertext Markup Language (HTML), the OPG would employ a *content management system*, or *CMS*. The decision to use a CMS in turn generated additional organization and content gathering requirements. In particular, the CMS that was chosen needed to be one that was well supported and could be learned in a short time. Also, as will be documented later in the report, it also required that all of the content would have to be gathered and written to highly structured guidelines.

After evaluating a number of software packages, the OPG technical lead proposed that the project use *Drupal*,⁵ a popular open source content management platform. While Drupal has a steep initial learning curve for individuals who will be implementing its technical features, it is an extremely flexible and powerful program that allows for a significant amount of customization. Furthermore, the Drupal platform is supported by an active worldwide community of developers and users. This meant that when problems were encountered, it was easy to use the web to find solutions. This development community also constantly releases new open source plug-ins and extensions to extend Drupal's functionality, meaning that the OPG team didn't necessarily have to reinvent the wheel when seeking to extend the functionality of the platform. Finally, while the learning curve might be steep on the technical side, the content authoring interface resembles a word processor, making it easy to add, remove, and edit content.

In addition to the CMS, there were other technical decisions that were made to support the content. During the definition phase of the project it was decided that the OPG should contain an interactive tool to help users evaluate the different online print-on-demand services.⁶ After evaluating a number of methods to deliver that experience, it was decided that Adobe® Flash would provide the highest levels of interactivity to the widest range of visitors to the site. Unfortunately, Flash, originally created as an anima-

5 - <http://www.drupal.org>

6 - In many respects the chooser application, as it became known, represented a project within a project. Because of the size and relative complexity of the end product, the chooser proceeded on a parallel, though somewhat separate, track through each of the production stages the OPG went through. For example, a standalone requirements document was developed for this application, and then used to drive its technical implementation.

tion program, is not a robust programming environment. Luckily, during the course of the project, Adobe released a new product called Flex which enables great programming flexibility while keeping the final output in Flash format.

Finally, it was also decided during the definition phase that the site would benefit from the addition of tutorial videos. These videos would take visitors through all the steps of preparing a book to be published using the different software packages referenced on the site. To produce the movies, the team identified a software package that would record actions as they occurred on screen and sync them to a voice over track. In terms of distribution, the team made a somewhat unconventional decision. Rather than hosting the videos on the OPG website, the decision was made to use *Vimeo*,⁷ a third party hosting community. Choosing Vimeo ensured that the videos would be housed on a server specifically configured for the delivery of streaming video. Additionally, because it has a community feature similar to YouTube, Vimeo will allow OPG video content to be accessible to a wider audience across the web.

Once these decisions were made, the project moved into its production phase. Content gathering and authoring, visual design, and technical implementation began to happen in parallel.

Gathering and Authoring Content

Just as the user and organization requirements shaped the OPG's technology path, the technology path shaped how the content was collected. At the end of the Site Structure phase, the team had a content map that outlined every piece of content on the site. Each site section contained specific categories of content (definitions, articles, project templates, etc). The next step was to establish what made up each of those categories. For example, it was decided that a web resource would contain the following elements:

- A title,
- The URL link of the resource,
- A description of the resource (up to 100 words), and
- An optional image of the resource.

Establishing these elements ensured that content would be consistently gathered and added to the site. The project manager, working with the site designer, created online input templates for each type of content and provided them to the SMEs who then began the process of identifying and collecting content.

The exception to this process was gathering the content for the interactive guide to online publishing services, which came to be referred to during production as “the chooser.” Because the chooser needed to be programmed in Adobe® Flex, and the initial focus for the technical resource was setting up the Drupal environment, the research-

7 - <http://vimeo.com>

ers who collected data for the chooser initially used Microsoft® Word and Excel to save their results. This report will discuss details learned about the websites included in the chooser as part of the *Findings* section.

In addition to formatting the structure of the content, care was also taken to format its style. The intention was that the site should function as an advisor or guide rather than as a teacher. Therefore, the OPG's text would be written in an open and friendly style. Jargon was to be avoided if at all possible, and, where it could not be, it was to be accompanied by a definition or a link to the site glossary. All of these decisions were captured into a style guide to assist current and future SMEs in writing for the site.

Visual Design and Building

One of the great advantages to working with a CMS is that it enables parallel workflows. Once the technical lead implemented the content entry forms for the SMEs, they could begin to load text into the site even though the “display” aspect of the site did not yet exist. This style of workflow demonstrates the overall power in separating content (raw assets) from visual design. It provided the technical and design leads with a flexible environment to experiment with different designs.

The first step was to use the content map to establish different page templates. This step resembled the content categorization that took place in phase two. This exercise established the different types of web page templates that had to be created to meet the content needs of the site. Broadly speaking, the pages fell into two major categories: navigation (the home page, sub navigation pages) and content (articles, references, blog, etc). As with the content categories, the goal was to manage breadth with depth, ensuring that there were enough templates to optimally deliver the content without creating so many templates that they became difficult to manage.

Once the templates were approved by the OPG team, the next step was to create individual wireframes for each template. A wireframe is a “bare bones”, outline representation of a web page. The goal of the wireframe is to approximate where different individual pieces of content, or “chunks”, will be placed on the page and, relatively speaking, the amount of room each content piece will take. During the wireframing process, traditional aspects of visual design (color, typography, finished images) are not represented on the page.

After the wireframes for each template were established, the design lead began to “flesh” them out – creating visually designed mockups of each template using samples of the content that had already been uploaded to the site. The mockups incorporated all of the visual design requirements documented as part of phase one. As she was making the mockup, the design lead also developed a visual style guide to match the content guide. The design lead also developed a “shot list” of illustrations that needed to be created.

At the same time, the technical lead began to create the actual web page templates in Drupal using the wireframes. Through the use of Cascading Style Sheets (CSS), the

visual design remained largely separate from the placement of text within the template. This allowed the tech lead to engage in rapid prototyping of the templates and collaborate with the designer to correct issues that appeared in the migration from wireframes on paper to their execution on screen. Only after a final visual design had been approved did the technical lead begin to implement the final “look” of the site.

The iterative nature of this process extended to the writing and editing of content as well. Once the wireframe templates were created by the tech lead, they were immediately applied to the content that was being developed by the SMEs. The resulting “skeleton” site allowed the SMEs to begin to see how their writing would appear on the web. This in turn allowed them to begin to edit the text to fit together in a more cohesive fashion. It also identified content gaps that were not apparent during the *Site Structure* phase.

Beyond the writing, wireframing, and implementation of the visual design, there was one other important creation effort underway. In order to maximize site searchability, and, in particular, Google ranking, appropriate metadata was added to each page of the site.

Launching, Tracking, and Maintenance

This report is being written prior to the launch of the site in mid-November, prior to the 2008 Printing Industry Center Symposium. At that time, the site will be “soft” launched. There will be little initial publicity other than word-of-mouth/viral marketing. This will provide time to stabilize the environment and watch as individuals begin to interact with the content.

Google Analytics will be used to provide quantitative data about traffic patterns to and within the site. That information can be used to begin to understand what content is getting the most traffic, and therefore may be more popular. That data will be supplemented with other quantitative and qualitative data gathered through online user surveys and offline user testing. The goal of all of this will be to iron out the proverbial “kinks” in the system, and to test the content to ensure that visitors find it useful. As data comes in, the design, navigation, and site metadata will be altered appropriately.

After three months of iterative testing, the site will be “hard” launched. The three-month gap between the launches will also provide time to develop an appropriate PR campaign to ensure the widest reach.

Findings

In addition to the experience gained from implementing the site, this project yielded a number of valuable findings about the state of online publishing. The first was the wide and diverse range of open source publishing tools available online. Open source software is available for every single step of the publishing process, and there is an open source equivalent for every major software package used by industry—from Microsoft® Office to the Adobe® Creative Suite. Additionally, there is a huge selection of publicly available, royalty-free fonts available online. Simply put, without engaging in any form of software piracy, individuals can download, for free, tools of the trade that begin to approach (although, in some cases, they are not the equals of) those used in the professional space. The effect that this availability of tools may or may not have on the publishing industry over time is one that warrants further investigation.

Further, Frank Cost's work in illustrating various workflows demonstrates how most users can use non-traditional software packages such as Microsoft® Powerpoint to create books that are more than "good enough" for themselves. Once the OPG is launched, it will be important to track the popularity of those publishing tutorials relative to the ones that use more traditional publishing tools.

Finally, the process of assembling the chooser, or, as it appears on the final site, *The Self Publishing Advisor*, provided a wealth of information about the current state of online publishing services. What impressed the SMEs was the wide range of options available across all of the sites. In fact, arguably their greatest commonality was the diversity of their product and service offerings. This diversity is also reflected in the parameters that the Advisor uses to guide users to various services, including:

- Book types (e.g. photo books, text and photo book, text),
- If the site offers templates or authoring tools,
- Color vs. black & white printing,
- Binding options, and
- If the site provides editorial, marketing, and/or design services.

One thing that almost all of the sites offered were services to obtain an ISBN number for a customer's book. However it was noted by SMEs that, in some cases, it is indicated in the small print that the print-on-demand site retains ownership of the ISBN number. An interesting line of inquiry would be to survey users to see if they understand the implications of such agreements.

In general, the many points of differentiation are better understood as continuums. For example:

- The amount of publishing knowledge that the end user is expected to have varies greatly from site to site. Some are very easy for beginners (those sites feel very friendly, easygoing, and simple), while others appear much more formal and complicated, using publishing jargon without explanations. Some sites offer step-by-step instructions for publishing aspects, such as making PDFs and working with images, while others provide nothing more than specifications.
- Book sizes range from standard paperback to large-scale high quality photo books. No site offered all options. Echoing the previous point, some sites included pictures to help illustrate the size and scale of the books, while others relied on the user to understand the proportions of their book.
- The length of production runs vary across the sites, ranging from those that offer very short runs (one or two copies) to others that will only print in bulk.
- The range of services also varies greatly. Some sites offer one-on-one help with editing, designing, and marketing, while others offer virtually no help. In the case of one site, users can hire a publishing services assistant, at a premium price, who guides the user through the process, laying out and proofing the content for them.
- Some sites utilize social networking tools to interact with users and facilitate communication between users. Some feature groups that users can join to share work and ideas with other users who are producing work in similar topic/genre areas. Others provide tools that assist in the promotion of books.

Perhaps the best way to understand these differences is to chart of them on two axes. One axis charts the user's experience level with publishing, from neophyte to expert. The other axis represents the goal/the audience of the author. The products and services available range from those that appeal more towards individuals creating books for themselves and their immediate social circles to those that offer services for individuals seeking to niche or mass market their books. There is an opportunity to perform additional research into how the features of these sites fit inside the axes describe above.

References

Goto, K., & Cotler, E. (2004). *Web ReDesign 2.0: Workflow that Works* (2nd ed.). Berkeley, CA: Peachpit Press.



Rochester Institute of Technology
College of Imaging Arts and Sciences
55 Lomb Memorial Drive
Rochester, NY 14623
Phone: (585) 475-2733
<http://print.rit.edu>