

DITA and Drupal: The “Dynamic Duo” ...

...of content-rich, collaboratively produced publishing solutions



Anna van Raaphorst Johnson, Content Specialist
Richard H. (Dick) Johnson, Technology Specialist

Authors of the *DITAinformationcenter™* and *DITA Open Toolkit User Guide*

VR Communications, Inc.
www.vrcommunications.com

February 2012

What is a content-rich solution?

- An information collection that has a lot of text, images, or both
- For example, software or hardware documentation, procedures manuals, scientific data, product catalogs, or photo galleries
- Can be formal (“structured”) or informal (“unstructured”), or a combination of both
- Published to a web portal that is accessible by the target audience

What's important for users?

- Information appears as an “organized collection”
- They can easily browse, search, and query the information
- Information can be customized by document title or category, domain, user skill or interest, etc.
- If allowed and desired by the information owner, users can review, comment, edit, or contribute directly on the website

What's important for information producers?

- Published content will be seen, read, and appreciated by the target audience
- Information can be easily created, processed, published, and maintained
- Information owner can restrict access, as required

What are key publishing platform requirements?

- Portal that is popular with web users, builders, and developers
- Has a rich current function set (for example, workflow, views, metadata capabilities, easy-to-use native editor)
- Is easily extensible and customizable
- Allows for publishing and republishing of externally produced information
- Encourages contributions directly on the site, if desirable

What is our solution?

- Structured/unstructured “mashup” published to a Drupal 7 platform:
 - Structured information is DITA source that has been pre-published to XHTML prior to posting on the portal
 - Unstructured information is blogs, basic pages, announcements, etc. created with native Drupal functionality
 - The two information collections are organized, managed, and linked through a common vocabulary (metadata tags)

www.ditainfo.info

www.drupalinfo.info

www.edhinfo.info

DEMO WEBSITES

IS THIS SOLUTION FOR YOU?

Would our solution benefit your organization? Probably, if...(1)

- You are looking for a robust, popular, accessible portal to post your structured information
- You need a branding/styling solution that avoids the complexities of XSLT
- You need to avoid the high cost and extreme complexity of turnkey or highly customized solutions
- You have (legacy) content in Framemaker, MadCap Flare, WebWorks, Eclipse help etc. that you want to post now and convert later (or never)

Would our solution benefit your organization? Probably, if... (2)

- Your content solution is not complete without BOTH structured and unstructured information, and your informal information contributors can't or won't use DITA
- You need a web-based, interactive solution that includes capabilities like multi-person blogs and comments
- You want to keep the development of structured and unstructured content solutions separate
- The content for your structured/unstructured solutions is produced and assembled by multiple organizations that use different tools and operational procedures

HOW WE DID IT

How we built our demo sites

1. Created content in DITA 1.2 using XMLSpy and oXygen
2. Published to the XHTML output type using the DITA Open Toolkit
3. Posted to WordPress using a Python script (no plug-in required)
4. Also posted to Drupal 7 using a Python script that makes use of the APIs in Bulkpub, a contributed module that provides the required web services

Key DITA/Drupal features we use

- **DITA**
 - Hierarchical format
 - Keyword metadata
 - XHTML styling (no significant conflicts with CSS in Drupal)
- **Drupal 7**
 - Tagging automatically brought forward from DITA keywords
 - Search (across both structured and unstructured content)
 - Views (e.g. query-by-tag)
 - Themes (many are free)
 - CSS styling (easy to modify)
 - Book feature (hierarchical model)
 - Native editor for unstructured content (e.g. blogging)
 - Blogging clients (e.g. Firefox, Word) for laptop, phone, tablet

WORDPRESS VS DRUPAL

WordPress vs. Drupal (1)

- **Common characteristics**

- Both are open-source projects with a large developer and user community
- Both can be used to create a mashup on a single website
- Both can be customized using theming and configuration
- Both can be instrumented so you can see which pages are being viewed and who your users are
- Both platforms support remote blogging using XML-RPC

WordPress vs. Drupal (2)

- **WordPress**

- Easy to learn and maintain
- Mostly a blogging platform
- Only two built-in content types (page and post); not straightforward to define new types
- For us, CSS styling was problematic

WordPress vs. Drupal (3)

- **Drupal**

- Powerful and popular
- Steeper learning curve, but once you learn it, more extensible and customizable
- Easy to define custom content types and taxonomies for tags; good home for DITA content with topic metadata
- Drupal styling with default Drupal 7 Bartik theme works well with DITA XHTML content
- Once the DITA structured content has been published, easy to build views (see our Query-by-tag views)
- **Drupal is our favorite for DITA-based content**

STRUCTURED VS UNSTRUCTURED

Structured vs. unstructured

- **Structured information**

- Definitive, “official”
- Comprehensive in scope
- Formally architected and tagged
- Usually produced by professional writers
- Tightly controlled
- Small group of contributors

- **Unstructured information**

- Produced by (serious) amateurs or SMEs
- Large group of contributors
- Limited in scope
- Informally collected
- Loosely controlled

Structured/unstructured challenges

Structured Pros

- Controlled
- Accurate
- Essential
- Professional

Unstructured Pros

- Has popular “in-the-trenches” look and feel
- Easy, cheap
- Ubiquitous

Structured Cons

- Lacks “real-world” credibility
- Information silo
- Difficult, expensive

Unstructured Cons

- Uncontrolled, unreliable
- Lacks professional polish
- TOO ubiquitous



Solution:



**Collaboration!
Mashup!**

Structured
information

Unstructured
information

Structured/unstructured benefits

- **Best of both worlds:**
 - The information contributors (professional writers and SME contributors) can choose their favorite authoring and publishing tools
 - The structured collection is tightly controlled, single-sourced, and easily branded through theming on the portal
 - The unstructured information collection can be created and updated independently from the formally published documents, but are tagged with the same metadata

For more information



www.vrcommunications.com