

# Implementing DITA

Production notes  
from a real-world DITA Open Toolkit project



Anna van Raaphorst  
Richard H. (Dick) Johnson  
VR Communications, Inc.  
[www.vrcommunications.com](http://www.vrcommunications.com)

July 12, 2006

# Contents

- **Demo:** Working environment, source and output files
- Overview of the *DITA Open Toolkit User Guide and Reference* project
- Selected production notes:
  - Evaluating DITA/DITA OT, and getting started
  - Demo: Document build
  - Processing (building) and troubleshooting
  - Migrating to DITA OT 1.3
- **DITA as a strategic solution** (notes from the May 2006 XTech conference)

# Demo

Working environment,  
source and output files

# Project Overview

# What is the project?

- *DITA Open Toolkit User Guide and Reference* (release 1.3)
- Volunteer contribution to OASIS and the DITA community
- Bookmap specialization, 20 chapters, 300 topics
- Target release date: August/September (may be staged to accommodate 1.3 development schedule)
- Drafts posted bi-weekly on Yahoo DITA user forum (files section)
- Comments and input welcome

# Document contents

- Release 1.3 information
- Getting started
- Getting information
- Evaluating DITA and DITA OT
- Installing, upgrading
- Setting up your working environment
- Processing and publishing to all 9 target environments
- Troubleshooting the build process
- Creating topics and maps, sample files
- Linking, customizing, distributing, reusing, migrating
- DITA core vocabulary (controlled vocabulary)

# Why We Did the Project

- Significant interest and prior involvement with structured writing, content management, scripting, and programming
- Way for us to gain knowledge and experience quickly
- Help the DITA community
- Opportunity to use our individual skill sets in a collaborative effort



# Key Challenges

- Learning and writing simultaneously – for example:
  - How to choose authoring tool
  - Unfamiliar build environment
  - Some gains, some losses in functional richness
- How to collaborate effectively
- Change of initial project focus





# Production Notes

# Evaluating, Getting Started

1. Be honest (can you answer “yes” to these questions?)
  - Do you have strong motivation and a champion with money?
  - Do you have the skills required (architecture, communication, technical)?
  - Can you collaborate effectively?
  - What are your key target outputs and will you be satisfied with the results (or will you need to customize)?
  - Do you have appropriate demo, prototype, and end-to-end projects in mind?
  - Do you have the time and patience to be successful?
2. Plan and architect your demo and prototype projects
  - Talk to others who have done similar projects
  - Get multiple people involved, including an information architect
3. Choose a (temporary) authoring tool
  - Start with with a free or inexpensive “DITA-aware” editor
  - Set it up to use the DTD catalog

# Getting started, continued

4. Install DITA OT and set up your working environment
  - Check the system requirements carefully (components must be compatible)
  - Separate your source and output files from the DITA OT files
5. Do a demo project
  - Single map, 12-20 files (include all information types), individual or two-person project
  - Add tables, definition lists, metadata, filtering, indexing, simple linking
  - Publish to all key target environments
6. Do a prototype project with real content
  - Multiple maps, 20-50 files, group project
  - Publish to all key target environments
  - Add complex linking in stages

# Demo

Processing (building) files,  
our debugging tool

# Why create debugging tools?

- OT error messages are sometimes cryptic and misleading, or completely missing
- Nice to have simple reporting tools in the absence of a content management system
- Relatively easy to do for experienced software engineers:
  - DITA files are well-formed and valid; by parsing XML, scripts can “walk” the tree
  - You can use SimpleXML for parsing the XPath queries
- Our tool was written in PHP, is available free, and is primarily for debugging

# Cryptic, misleading error messages

Toolkit error message:

```
[DOTJ013E][ERROR] Failed to parse the referenced file  
'installing\indows_settingenvvariables.dita'
```

Message from free version of our debugging tool:

```
Error, file  
C:\DITAOT_UGRef_SOURCE\installing\indows_settingenvvariables.dita  
Does not exist! Bad reference:  
C:\DITAOT_UGRef_SOURCE\installing\windows_installingant.dita  
-> indows_settingenvvariables.diat
```

# Name folding errors

No debugging help from the Toolkit (but some serious output errors)

Message from free version of our debugging tool:

Name folding error, file

C:\MY\_DITA\_SOURCE\samples\garage\tasks\../Concepts/oil.dita

Does not exist!

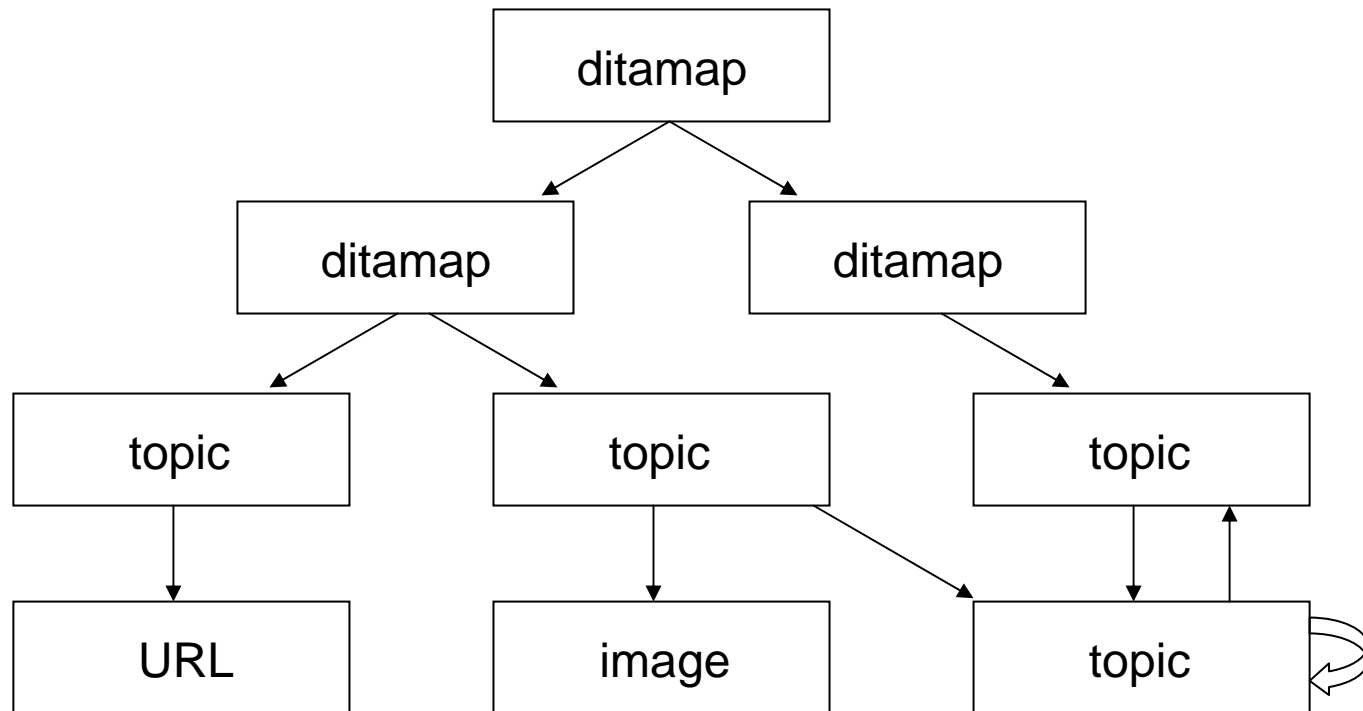
Bad reference:

C:\MY\_DITA\_SOURCE\samples\garage\tasks\changingtheoil.dita

->../Concepts/oil.dita

# DITA source tree: the key to debugging and reporting

Scripts can “walk the tree”





# URL errors

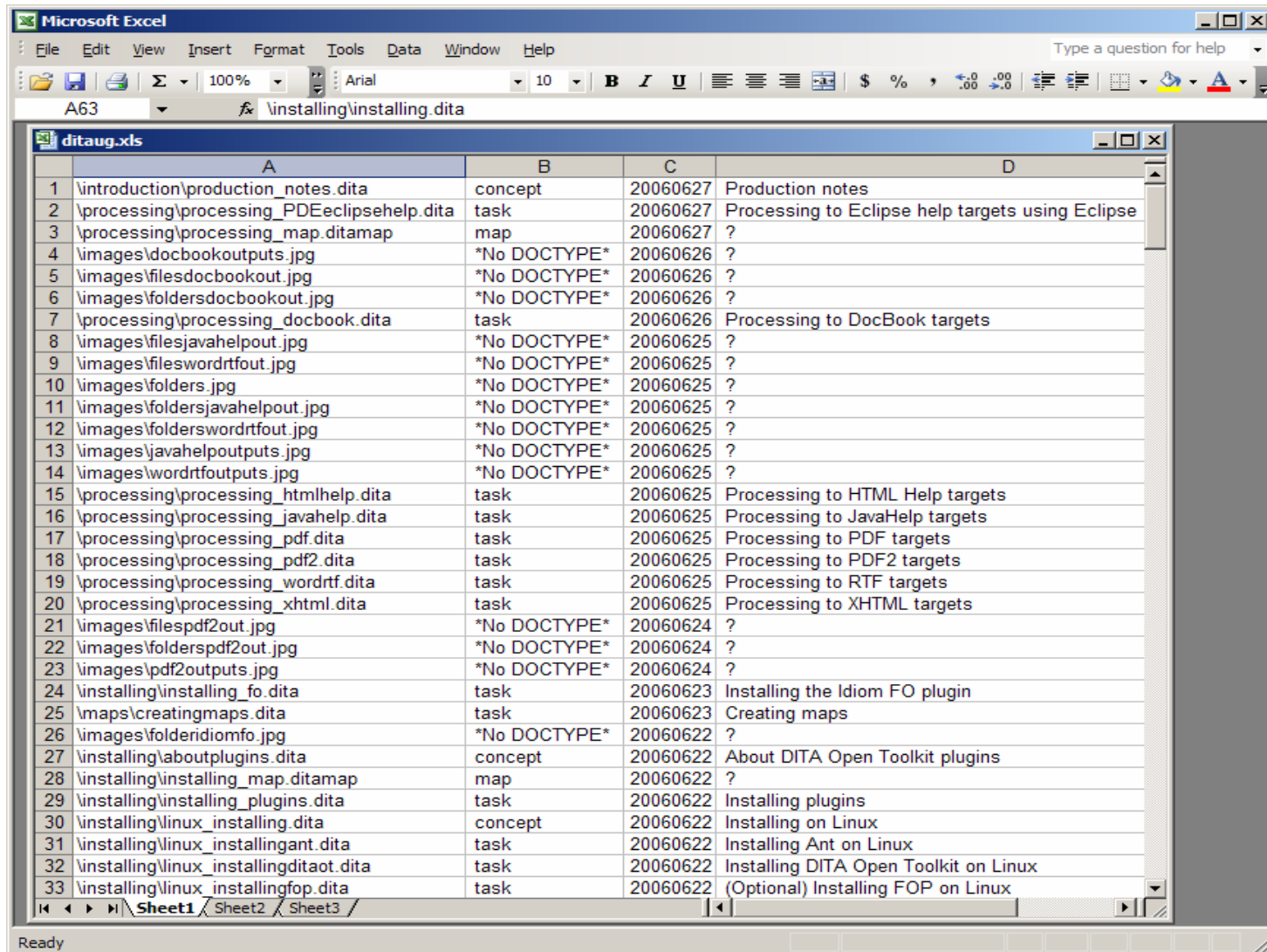
No debugging help from the Toolkit (but broken links)

Message from an extension to our debugging tool:

URL `https://sourceforge.net/forum/?group_id=132728` open failed.

Host `unt.apache.org` invalid for `http://unt.apache.org/bindownload.cgi`.

# Beyond debugging: list of changed files



The screenshot shows a Microsoft Excel spreadsheet titled 'ditaug.xls' with the following data:

	A	B	C	D
1	\introduction\production_notes.dita	concept	20060627	Production notes
2	\processing\processing_PDEclipsehelp.dita	task	20060627	Processing to Eclipse help targets using Eclipse
3	\processing\processing_map.ditamap	map	20060627	?
4	\images\docbookoutputs.jpg	*No DOCTYPE*	20060626	?
5	\images\filesdocbookout.jpg	*No DOCTYPE*	20060626	?
6	\images\foldersdocbookout.jpg	*No DOCTYPE*	20060626	?
7	\processing\processing_docbook.dita	task	20060626	Processing to DocBook targets
8	\images\filesjavahelpout.jpg	*No DOCTYPE*	20060625	?
9	\images\fileswordrtfout.jpg	*No DOCTYPE*	20060625	?
10	\images\folders.jpg	*No DOCTYPE*	20060625	?
11	\images\foldersjavahelpout.jpg	*No DOCTYPE*	20060625	?
12	\images\folderswordrtfout.jpg	*No DOCTYPE*	20060625	?
13	\images\javahelpoutputs.jpg	*No DOCTYPE*	20060625	?
14	\images\wordrtfoutputs.jpg	*No DOCTYPE*	20060625	?
15	\processing\processing_htmlhelp.dita	task	20060625	Processing to HTML Help targets
16	\processing\processing_javahelp.dita	task	20060625	Processing to JavaHelp targets
17	\processing\processing_pdf.dita	task	20060625	Processing to PDF targets
18	\processing\processing_pdf2.dita	task	20060625	Processing to PDF2 targets
19	\processing\processing_wordrtf.dita	task	20060625	Processing to RTF targets
20	\processing\processing_xhtml.dita	task	20060625	Processing to XHTML targets
21	\images\filespdf2out.jpg	*No DOCTYPE*	20060624	?
22	\images\folderspdf2out.jpg	*No DOCTYPE*	20060624	?
23	\images\pdf2outputs.jpg	*No DOCTYPE*	20060624	?
24	\installing\installing_fo.dita	task	20060623	Installing the Idiom FO plugin
25	\maps\creatingmaps.dita	task	20060623	Creating maps
26	\images\folderidiomfo.jpg	*No DOCTYPE*	20060622	?
27	\installing\aboutplugins.dita	concept	20060622	About DITA Open Toolkit plugins
28	\installing\installing_map.ditamap	map	20060622	?
29	\installing\installing_plugins.dita	task	20060622	Installing plugins
30	\installing\linux_installing.dita	concept	20060622	Installing on Linux
31	\installing\linux_installingant.dita	task	20060622	Installing Ant on Linux
32	\installing\linux_installingditaot.dita	task	20060622	Installing DITA Open Toolkit on Linux
33	\installing\linux_installingfop.dita	task	20060622	(Optional) Installing FOP on Linux

# Beyond debugging: reporting statistics

count copyright holder

=====

225 VR Communications, Inc.  
4 IBM Corporation  
20 Comtech Services, Inc.  
1 Publishing Smarter

count author

=====

179 Anna van Raaphorst  
3 IBM  
45 Richard Johnson  
20 Jen Linton  
20 JoAnn Hackos  
20 Kylene Bruski  
2 Deborah Pickett  
1 Bernard Aschwanden

# Migrating to DITA 1.3

- DITA 1.1 support
  - Bookmap
  - Abstract element (container for one or more shortdesc elements)
  - Indexing (*see, see also*; index ranges; sort order)
- Localized index
- Incremental build
- Topicmerge
- Ant refactoring (lesson: don't modify the system files!)

# DITA as a Strategic Solution

# XTech 2006 and RDF

- Multiple papers on Semantic Web, RDF, and SPARQL query language
- RDF:
  - Machine-readable metadata encoded in output files
  - Subject/predicate/object triples
  - Example: <thisdocument> <createdby> <marktwain>
  - Can be searched (crawled), browsed, or queried
- SPARQL: W3C candidate (April 2006)
- Sibersafe: RDF/OWL-based knowledge modeling technology
- Thesaurus plug-in
- Firefox plug-in: PiggyBank semantic Web browser
- Swoogle (U Maryland) search engine

# RDF, continued

Some RDF data included in DITA output by default:

```
<meta name="dc.creator"="Mark Twain"/>
```

Some RDF data can be easily added to the output:

```
Has RDF: <head profile=http://purl.org/NET/erdf/profile>
```

```
Namespace: <link rel="schema.dc" href=http://purl.org/dc/elements/1.1/>
```

Could be accomplished as a processing or post-processing step

# RDF, continued

- For more information:
  - Embedded RDF:  
<http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml>
  - Thesaurus:  
[http://sourceforge.net/project/showfiles.php?group\\_id=132728](http://sourceforge.net/project/showfiles.php?group_id=132728)
  - SPARQL:  
<http://www.w3.org/TR/rdf-sparql-query/>
  - SiberSafe (DITA edition):  
<http://www.siberlogic.com/dita/>



# (Last Slide) Implementing DITA

For more information and doc drafts:  
<http://groups.yahoo.com/group/dita-users/files>  
[www.vrcommunications.com](http://www.vrcommunications.com)

