



How to Get the Most Out of Content Migration to DITA

X-Pubs Conference

June 2008

DCL Information:

Data Conversion Laboratory

61-18 190th St., Suite 205
Fresh Meadows, NY 11365
Telephone: (718) 357-8700
Fax: (718) 357-8776
Web: <http://www.dclab.com>

Mikhail Vaysbukh, Sr. Project Manager

mvaysbukh@dclab.com
718-357-8700 Ext.# 228

Don Bridges, DITA Sales

dbridges@dclab.com
505-275-2223



Goals for Today

- ✓ Quick Introduction to DITA
- ✓ Why is DITA Important?
- ✓ Is Conversion Worth It?
- ✓ Enhancing the Migratability of your Content





What's DITA and What's New About It?

- Pre-packaged XML – ready to use (sort of)
- Infrastructure included - taxonomy (dtd & schema); printing stylesheets; lots of tools
- Extensible for specialized application as you grow



New and Different

- Stand-alone *topic* paradigm
- Increased emphasis on content reuse
- Information flows across models vs down models
- Architecture constraints not present in DTP (Desktop Publishing), DOCBOOK, etc. – It's important to understand what these constraints are in order to develop robust migration strategy.



“Getting there using DITA is ... like building with prefabricated modular components that can be quickly assembled into a suitable structure.”

*– Doug Henschen
intelligententerprise.com*



DITA's Impact On Business Issues

- DITA will likely offer significant ROI if...
 - You have multiple output requirements (single source)
 - You have similar product lines (content reuse)
 - You translate to multiple languages (content reuse)



Building A ROI Case

- Review other ROI analysis & Case Studies
 - www.ptc.com/WCMS/files/35544/en/35544en_file1.pdf
 - <http://dita.xml.org/resource-directory?filter0=95>
- Research available tools and services
 - <http://dita.xml.org/services>
- Inputting an accurate estimate for reuse
 - Content Reuse Assessment can be micro or macro



A Few Thought About Legacy Content

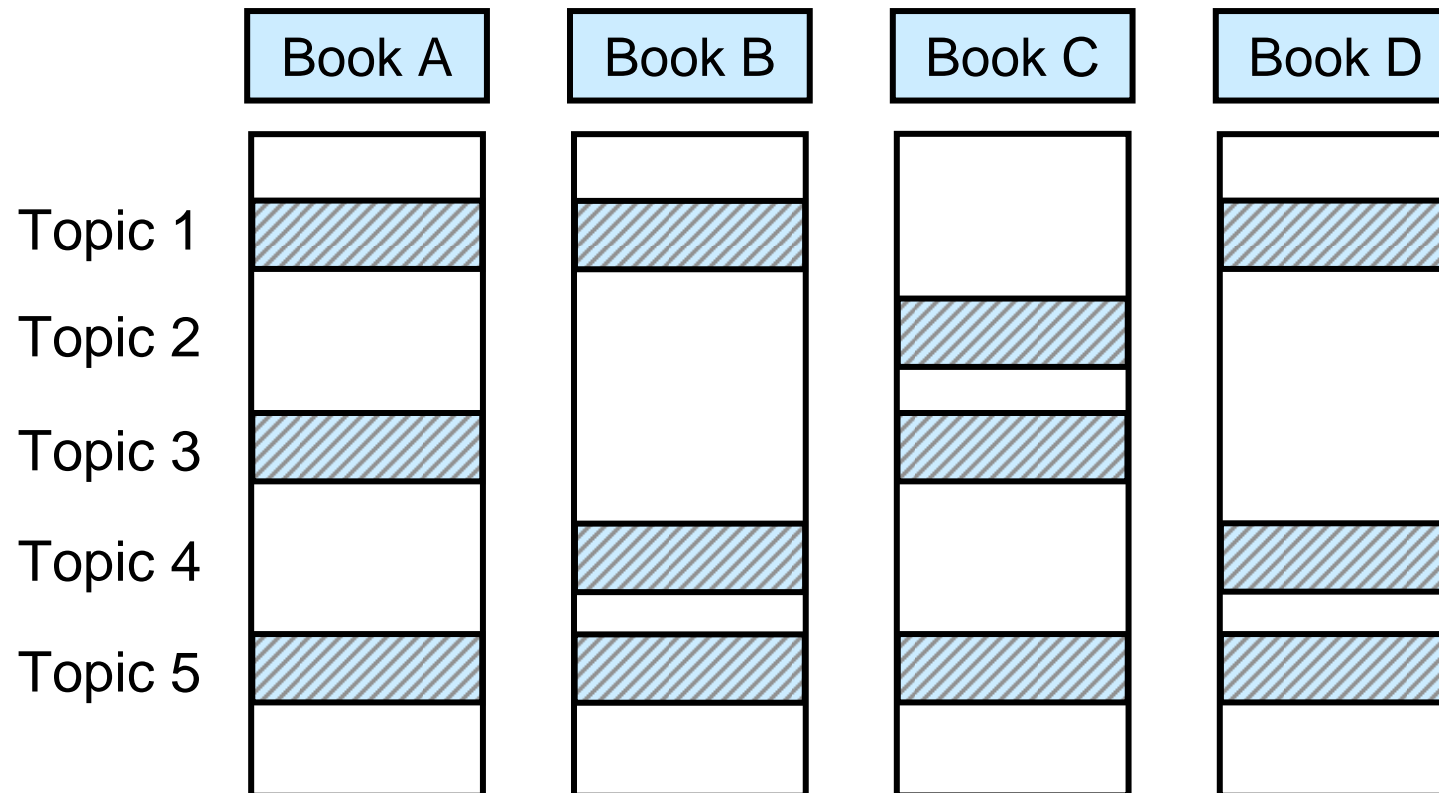
- There's valuable content locked up in your older documents. How much?; And how badly do you need it? What is the remaining life span of the documentation?
- Few writers have the *clairvoyance to* author content thinking it will be converted in the future.
- It's more fun and less trouble to author anew, but not necessarily faster or cheaper. It may cost \$50 to create new documentation page Vs. \$3-5 for converting and re-using legacy page.



Conversion Decision Factors

- What Makes DITA Conversion different?
- Is it worth Converting
 - Do you have live materials that's worth using. What's the 'best' approach?
- In-house, Out-source, or hybrid?
- How much preparing should you do?
 - Enhancing the 'migratability' of your content

DITA from a Conversion Perspective



- Conversion can be topic sourced or book sourced



DITA Topic Sourced Conversion (the best, but at a cost)

- Analyzing 'similar' topics that exists in many 'books' and re-writing
- Edit as a 'stand alone' topic that holds meaning on it's own
- Manual process
- Maximizes reuse



DITA Book Sourced Conversion

- Designation of topics by type
 - Task
 - Concept
 - Reference
- Usually based on a Heading level
- Automated process
- Typically 80%+ of the output topics don't need re-authoring and those that do – don't need it right away. (as legacy presentation of a topic is still maintained)
- Limited native reuse 'out of the box'
 - Re-writing for Reuse happens as a follow-on step



Harmonize Content to Leverage Reuse

- DITA Topics that express the same concept with different words (even a single different word or punctuation mark) are considered different topics.
- Perform a manual or automated audit to have as much reusable content as possible.



DITA Topic vs. Book Sourced Conversion

	Topic Based	Book Based
Conversion	Manual	Automated
Reuse	Maximized	Native Initially
Implementation Time	Very Long	Very Fast
Cost	Upfront	Distributed



Which Way Should You Go?

In-house Approach if...

- Data is so sensitive that it must stay internal
- Paradigm is consistent
- Your schedule is flexible
- Materials are not complex
 - Clean-up requirements may vary widely
- Budget is flexible & in-house resources available
 - Tools are relatively cheap
- Project is small





Which Way Should You Go?

Out-sourced Approach if...

- Paradigm is changing
- Meeting schedule is critical
- Materials are complex
 - Demonstrate expected results while there is still time to make modifications
- Budget is tight
 - Understanding the project costs and the trade-offs
- Project is large
 - Process scales as big as needed



Background of Data Conversion Laboratory

- Established in 1981
- Pioneer in defining and developing the emerging data conversion industry (founding member of SGML User Group)
- Expertise in large complex conversion projects
- Substantial experience in managing multiple vendors for large-scale projects, with automated tracking and reporting of data throughout
- Established sophisticated multi-step conversion processes with multiple passes of automated and human quality control steps
- Wrote the data conversion chapters in Charles Goldfarb's *The XML Handbook* and in Bill Kasdorf's *Columbia guide to Digital Publishing*
- Produces a Monthly Newsletter devoted to SGML/XML and Electronic Publishing topics with a subscriber base of 7,000
- Recognized speaker on SGML/XML conversion (AIA Tri-services, Marines Tri-services, IADS Users, AFEI, Pentagon, SGML/XML Users, XML World, Society for Technical Communications, Booktech, Online World, Internet World, Society for Scholarly Publishing, American Medical Publishers Association, XyUsers Group, FrameMaker Users Group)



DCL DITA Experience

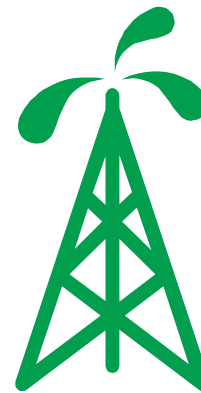
DITA conversions completed for over a dozen different customers in these vertical markets:



Manufacturing



Life Sciences



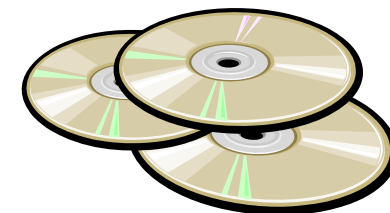
Oil & Gas



Telecom



Aerospace



Software




Enhancing the Migratability of Content

- DITA has some constraints that you should be aware of.
 - Examples are listed on the following pages
- Native Reuse can typically be improved
 - fixing cases of ‘same concept – different words’
 - Fixing typos
- It's usually easier to fix these *before* migrating



Enhancing the Migratability of Content (cont'd)

Notes Done as Tables

	Note: This is an example of a note inside a table.
---	--

- Problematic for software to decipher if this is a real table or a note. This would be better handled in the input file as textual note:

Note: This is an example of a note inside a table.

- Which would convert to:

<note type="note"><para>This is an example of a note inside a table.</para></note>



Enhancing the Migratability of Content (cont'd)

Multiple Steps in a Single Task Topic

Example Scenarios

Example 1

1. A CHECK Flag is received by the EEC unit
2. A match is found in the Look-up table (Table 2-1) and the response is Type 2
3. Required response is CHECK FAULT

Example 2

1. A CHECK Flag is received by the EEC unit
2. A match is found in the Look-up table (Table 2-1) and the response is Type 3
3. Required response is CLEAR FAULT



Enhancing the Migratability of Content (cont'd)

Multiple Steps in a Single Task Topic

- If “Example Scenarios” is a task, this would be a problem because a task is only allowed one set of <steps>. That being said, the steps under Example 1 would need to be tagged as a list, and the steps under Example 2 would be tagged as <steps>.
- ***Each section should be broken down into their own topics*** (i.e., *Example Scenarios: Example 1* and *Example Scenarios: Example 2*)



Enhancing the Migratability of Content (cont'd)

Procedure Authored as a Table

Step	Action
1	Check the phase sequence in the mains with the phase sequence indicator.
2	Check the direction of rotation of the impeller. If a wet installation is necessary, check the direction of rotation through the inlet elbow access cover.



Enhancing the Migratability of Content (cont'd)

Procedure Authored as a Table

- Cases with different variations of tasks or procedures authored as tables in the source present additional complexity to the conversion process in cases when they need to be deconstructed into <task>s with <steps>.
- *Tasks should be paragraph type flow.*



Enhancing the Migratability of Content (cont'd)

Untitled Topics and Pages References

- Legacy content is usually divided into DITA topics based on existing heading titles. Untitled topics increases a risk of topics being mis-identified and adds ambiguity to resolving cross references to the untitled topics.

Lost or forgotten password – Browse to the location of the private recovery key used for the project. (See [page 121](#) for information about creating a recovery key.)

- And on Page 121 we find:

To Set a Recovery Key:

You can specify the key for access to encrypted instances. If you specify password protection and want to reset the...



Enhancing the Migratability of Content (cont'd)

Having More Than Two Levels of Steps

- DITA only allows two levels of steps (<step> and <substep> below it.

2. Install the Outer Guard as follows:

- a. Spread the outer shin guard and place it over the inner shin guard.
- b. Install the outer guard cap screws based on your particular pump:

For engines with a motor saddle support bracket:

- i. Ensure the outer guard is straddling the support arm, and install but do not tighten the two remaining cap screws.

For engines without a motor saddle support bracket:

- ii. Insert the spacer washer between the holes located closest to the motor in the outer guard, and install but do not tighten the two remaining cap screws.

- c. Position the outer guard so it is centered around the shaft, and so there is less than a 1/4" of shaft exposed. Tighten the cap screws.

3. Close the cover.



Enhancing the Migratability of Content (cont'd)

Specialized In-line Elements

- If your content has specialized in-line elements\components such as uicontrol or apiname and you currently style them the same way as any non-specialized text strings, because they all should appear looking the same in the print output.
- Consider creating and using specialized style names to distinguish difference in meaning of these different components.
- For Example in the source:

Press the **OK** button. Then **REMEMBER** to check if lunch is ready. And if lunch is ready, add use **message.write** method to create a message output that you are at lunch.

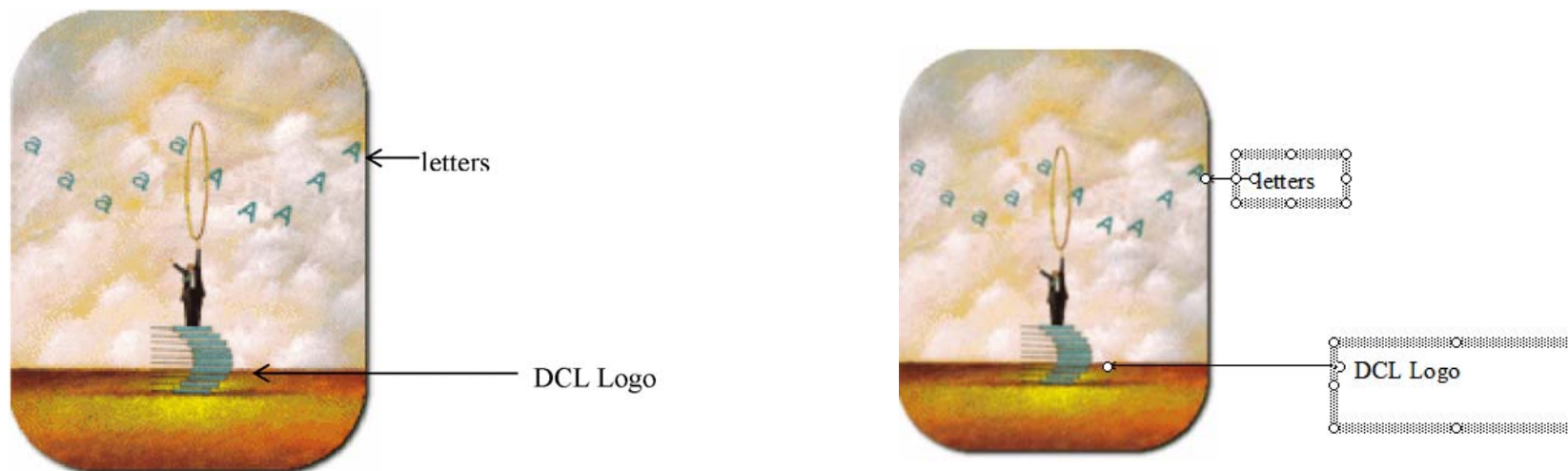
- Where in DITA it can be:

```
<p>Press the <uicontrol>OK</uicontrol> button. Then <b>REMEMBER</b> to  
check if lunch is ready. And if lunch is ready, add use <apiname>  
message.write</apiname> method to create a message output that you are at  
lunch.</p>
```

Enhancing the Migratability of Content (cont'd)

Think about strategy for illustrations with callouts

- In many source formats callouts are done as a separate layer and converted images are not “web ready” (as text does not stay in converted image).





Enhancing the Migratability of Content (cont'd)

Maximizing Reuse

- Macro analysis looks at legacy content on a paragraph level and quantifies/identifies.
 - Exact matches
 - Similar matches
- *Similar matches should be reviewed.*



Enhancing the Migratability of Content (cont'd)

Maximizing Reuse

aajq2115SECT16.RBW line: 152

Install avionics ram air diverter valve (21-21-22). **D** not do operational checkout.

Install avionics ram air diverter valve (21-21-22). **D** (**Do**) not do operational checkout.

Install avionics ram air diverter valve (21-21-22). ~~**D**~~ (**Do**) not do operational checkout.

aajq2115SECT17.RBW line: 204

aejn2126SECT16.RBW line: 178

acjq2114SECT17.RBW line: 213

acjq2114SECT16.RBW line: 191

spelling

aajq2115SECT13.RBW line: 323

Insulation blankets are fragile. **Be** careful when handling or damage may result.

Insulation blankets are fragile. (**fragile**,) **Be** (**be**) careful when handling or damage may result.

<Exact matches>

aajq2115SECT13.RBW line: 465

aajq2115SECT19.RBW line: 126

aajq2115SECT19.RBW line: 239

aejn2126SECT19.RBW line: 166

aejn2126SECT19.RBW line: 271

acjq2114SECT19.RBW line: 156

acjq2114SECT19.RBW line: 267

acjq2114SECT13.RBW line: 424

acjq2114SECT13.RBW line: 659

Insulation blankets are **fragile**. ~~**Be**~~ **fragile**, **be** careful when handling or damage may result.

aajq2115SECT18.RBW line: 755

aajq2115SECT18.RBW line: 786

aejn2126SECT18.RBW line: 838

aejn2126SECT18.RBW line: 865

acjq2114SECT18.RBW line: 870

acjq2114SECT18.RBW line: 899

capitalization & punctuation



In Summary

- DITA solves many business issues. It's not for everybody and everything but it's a very effective tool for some that are in the right situation.
- Once you're convinced, you'll need to convince management. Show potential ROI.
- You'll need to evaluate the approach that works best for your situation.
- Typically there are conditions in legacy content that need to be resolved.



Some useful resource

- ◆ DCL Technical Library – DITA Resources
<http://www.dclab.com/techlibrary1.asp?GRP=42>
- ◆ Join Converting Legacy Content to DITA
group on The Content Wrangler Community
<http://thecontentwrangler.ning.com/group/convertinglegacycontenttodita>
- ◆ Subscribe to DCLNews
http://www.dclab.com/request_subscription.asp



Some useful resource (con't)

- ◆ DITA vs. Docbook
http://www.dclab.com/dita_docbook.asp
- ◆ DITA-izing Your Documents: Five Issues to Think About When Converting Your Legacy Publications to DITA
http://www.dclab.com/dita_conversion_tips.asp
- ◆ Ann Rockley & Steve Manning on XML, DITA Conversions, and Dynamic Personalized Content: An Interview (Parts 1 and 2)
http://www.dclab.com/dita_xml_rockley.asp
http://www.dclab.com/dita_xml_rockley2.asp
- ◆ DITA Conversion – Fact Sheet
<http://www.dclab.com/dita.asp>



Questions...



& Answers

DCL Information:

Data Conversion Laboratory

61-18 190th St., 2nd Floor
Fresh Meadows, NY 11365
Telephone: (718) 357-8700
Fax: (718) 357-8776
Web: <http://www.dclab.com>

Mikhail Vaysbukh, Sr. Project Manager

mvaysbukh@dclab.com
718-357-8700 Ext.#228

Don Bridges, DITA Sales

dbridges@dclab.com
505-275-2223