Where DITA is Now and Where It is Going:
Lightweight DITA and DITA 2.0
Keith Schengili-Roberts – October 26, 2017
What We’ll Cover

A quick look at the projects DITA Technical Committee (TC) currently oversees:

• DITA Usage and the Changing Landscape of Technical Communications
• How DITA is Made (and by Whom)
• Lightweight DITA and Multimedia Additions to DITA 1.3
• DITA 2.0
DITA Usage and the Changing Landscape of Technical Communications
A Peek into My Research on DITA Usage

- Part of my role at IXIASOFT is to better understand who is using DITA XML and why
  - Has been very little independent research on the scope of the industry
- Major marketing research firms (Forrester, Gartner) likely consider DITA usage too small a niche; no significant research published from them for almost a decade
  - And a decade ago, DITA had only recently been released...
Where This Information Comes From

My main sources of information:

• People self-reporting DITA usage on resumes (LinkedIn, Indeed.com)
• Survey of presentations from technical communication conferences
• Published case studies / webinar presentations
• Technical writer job postings
• Information on individual companies and vertical market segmentation initially from LinkedIn, more recently from Hoover
• Published national labour statistics
Worldwide DITA Use Among Tech Writers and Companies

**Approximate number of companies currently using DITA:**

**700+**

**OF POTENTIAL COMPANIES WORLDWIDE**

**# USING DITA?**

**at least 1,400-3,000**

**Popular sectors:** SOFTWARE, TELECOM, SEMICONDUCTOR & MEDICAL DEVICE

**162,000**

TECHNICAL WRITERS ON LINKEDIN

**11,600+**

of these writers say they...

**KNOW DITA &**

**1,800**

say they are using DITA right now

~4.0% of U.S. tech writer jobs ask for DITA experience

**SOURCES**

Original research based on a 2015-2017 analysis of public data available on LinkedIn and Indeed.com.

© 2017 IXIASOFT Technologies. All rights reserved.
DITA is in Use Worldwide

DITA Using Companies (HQs) Worldwide

66%

26%

7%
“Traditional” DTP Tools Decline in Job Postings

FrameMaker and XML (+ DITA) in Technical Writer Job Postings on Indeed.com
Overall Decline in U.S. “Technical Writer” Jobs

"Technical Writer" Job Listings on Indeed.com for Q3 2012 - October 2017
I Think “Technical Writing” is Undergoing a Fundamental Shift

• If I do a non-restrictive search on Technical Writer (sans quotes) I get much higher figures from LinkedIn and Indeed.com
  • LinkedIn: 162K for “Technical Writer”; 190K without quotes
  • Indeed.com: 2K jobs for “Technical Writer”, 6.7K without quotes

• I strongly suspect “Technical Writer” no longer adequately describes what we do
Types of Jobs Asking for DITA Experience

- Instructional Designer/E-learning Developer
- Program Manager, Customer Service Content Strategist
- Content Specialist
- Senior Content Strategist
- Technical Editor
- Content Editor
- DITA-OT Programmer Intern
- Ontologist
- Content Management Administrator/Analyst
- Content Development Specialist
- Documentation Tools Developer

- Content Director
- Online Content Editor
- Manager, Knowledge Management
- Manager, Information Development
- User Assistance Developer 2
- UI Text Lead and Writer
- Digital Content Editor
- Sr Editorial Strategist
- Member of Technical Staff, Documentation
- Senior Consultant, Customer Technical Communications
- CMS Engineer
- Senior Staff Information Developer
Who Has a Corporate Need for an Ontologist?

- Larger firms with many products, help with efficient search
- From the job listing: “Develop and implement ontologies in consultation with stakeholders in teams dedicated to search, product management and product development.”

- I believe this is part of a shift of firms utilizing structured content
  - It’s no longer just about delivering a manual
Overall trend is that DITA adoption is firmly entrenched in tech-related sectors and is moving gradually into other sectors.
Sector Growth Over Time

DITA Usage by Industry Sector, Q3 2017

Computer Software: 40%
Information Technology and Services: 13%
Other: 12%

COUNT=702
Mainly Large and Very Large Firms Using DITA

• This is not too surprising, since the larger the firm, the more cost-effective DITA is when it comes to such things as:
  • Reduced localization costs
  • More efficient content production
  • Good fit with Agile processes
  • Other purposes requiring structured content
DITA Usage in Some Sectors is Very High

100% of the top 10 software companies are using DITA.

1. Microsoft
2. Oracle
3. IBM
4. SAP
5. Symantec
6. EMC
7. Hewlett-Packard
8. VMware
9. CA Technologies
10. Salesforce.com

At least 70% of the top 10 medical device firms globally are using DITA.

1. Medtronic
2. Johnson & Johnson
3. GE Healthcare
4. Fresenius (Medical Care)
5. Philips (Healthcare)
6. Siemens (Healthineers)
7. Becton Dickinson (BD)
8. Cardinal Health
9. Stryker Medical
10. Baxter International

Ranking data from Investopedia

Top 10 info from Proclinical.com for 2016
DITA Usage in Some Sectors is Very High (cont.)

At least 70% of the top 20 semiconductor companies globally are using DITA.

1. Intel
2. Samsung
3. TSMC
4. Broadcom
5. Qualcomm
6. SK Hynix
7. TI
8. Micron
9. Toshiba
10. NXP
11. MediaTek
12. Infineon
13. STMicro
14. Apple
15. Sony
16. Nvidia
17. Renesas
18. GlobalFoundaries
19. ON Semi
20. UMC

Top 20 List from IC Insights for 2016
DITA Ubiquity?

• No, but increasingly it appears as though large firms increasingly appreciate what can be done with structured content, including:
  • Better differentiating their products
  • Search Engine Optimization considerations
  • Working with chatbots and related interactive technologies
  • More efficient topic/content production
  • Faster localization processes and time-to-market for content
How DITA is Made (and by Whom)
The Role of the DITA Technical Committee

• The chief organizational body overseeing the development of the DITA standard within OASIS
• Chaired by Kris Eberlein
• Holds hour-long weekly meetings every Tuesday at 11AM EDT (5PM CEST)
• Members (which includes IXIASOFT) get a say + vote on new features/developments
• Non-members are invited to contribute ideas on the TC’s email list
While there are many more actual members, these are the members who currently (as of last week) have voting status, which effectively means they attend/contribute regularly.
Lightweight DITA: Process

- Currently being defined by a subcommittee (SC) belonging to the main DITA TC
- Co-chaired by Professor Carlos Evia and Michael Priestley (IBM)
- Status: draft Committee Note is currently under review by TC members
  - Committee Note will outline LwDITA elements and attributes
  - This will be followed by a full specification
  - A second Committee Note is planned; will cover template-based specialization
- First Committee Note to be released soon
LwDITA Highlights

• Fewer tags and attributes than “full” DITA 1.3; designed to be a “simpler DITA experience”
• Designed to be compatible with DITA 1.3; valid LwDITA code is also valid DITA 1.3 code *
• DITA is no longer necessarily bound to XML
• Three different “flavours” available:
  • XML-based XDITA
  • HTML5-based HDITA
  • Markdown-based MDITA

* Exception for the moment is the multimedia domain; more on this shortly...
LwDITA Audiences / Scenarios

- **XDITA**: tech writers wanting reduced/simpler tagset, environments where there is interchange between XML + Markdown / HTML5
- **HDITA**: technical marketing, software developers, trainers, bloggers
- **MDITA**: software developers, “individuals authoring content quickly that must be later refactored as structured content”
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>alt</td>
<td>shortdesc</td>
</tr>
<tr>
<td>body</td>
<td>simpletable</td>
</tr>
<tr>
<td>data</td>
<td>stentry</td>
</tr>
<tr>
<td>dd</td>
<td>sthead</td>
</tr>
<tr>
<td>dlentry</td>
<td>strow</td>
</tr>
<tr>
<td>dt</td>
<td>sub</td>
</tr>
<tr>
<td>dl</td>
<td>title</td>
</tr>
<tr>
<td>desc</td>
<td>topic</td>
</tr>
<tr>
<td>fig</td>
<td>topicmeta</td>
</tr>
<tr>
<td>fn</td>
<td>topicref</td>
</tr>
<tr>
<td>image</td>
<td>toc</td>
</tr>
<tr>
<td>i</td>
<td>ul</td>
</tr>
<tr>
<td>keydef</td>
<td>xref</td>
</tr>
<tr>
<td>linktext</td>
<td>media-autoplay</td>
</tr>
<tr>
<td>li</td>
<td>media-controls</td>
</tr>
<tr>
<td>map</td>
<td>media-loop</td>
</tr>
<tr>
<td>note</td>
<td>media-mutated</td>
</tr>
<tr>
<td>ol</td>
<td>media-source</td>
</tr>
<tr>
<td>p</td>
<td>media-track</td>
</tr>
<tr>
<td>navtitle</td>
<td>video</td>
</tr>
<tr>
<td>ph</td>
<td>video-poster</td>
</tr>
<tr>
<td>pre</td>
<td>48 Elements Total</td>
</tr>
<tr>
<td>prolog</td>
<td></td>
</tr>
<tr>
<td>section</td>
<td></td>
</tr>
</tbody>
</table>

433 Elements
MDITA Core and Extended

Two variations of MDITA: Core and Extended

- MDITA Core is based solely on GitHub-flavoured Markdown
  - Good for raw content exchange; little in the way of DITA features available
- MDITA Extended includes optional YAML-based header info, can also take HDITA (HTML-based) attributes and elements
Design Philosophy is Also Simplified

- Mixed content is not allowed. All text must be contained within `<p>`
- No CALS table elements (`<table>`, `<row>`, `<entry>`, etc.), just simple table
- No prolog metadata (everything is in `<data>`)  
- No related links
- Only highlighting domain is available, and only a subset of it (`<b>`, `<i>`, `<u>`, `<sup>`, `<sub>`)  
- Only generic topic is available; there is no concept, reference, task, troubleshooting or glossentry topic types
- Only map is available; there is no bookmap
- Only `@props` is available for filtering values

---

Sample LwDITA Topic Code (XDITA)

```xml
<topic id="install-and-setup">
  <title>Installing and Setting up Remote Lighting</title>
  <shortdesc>Installation of your lighting kit includes installing the light bulbs into light fixtures, preparing the remote control, and programming lighting groups.</shortdesc>
  <prolog>
    <data name="author" value="Kevin Lewis"/>
  </prolog>
  <body>
    <section>
      <title>Steps</title>
      <ul>
        <li><p>Install light bulbs.</p></li>
        <li><p>Prepare remote control.</p></li>
        <li><p>Program lighting groups.</p></li>
      </ul>
    </section>
  </body>
</topic>
```
(Sort of) Equivalent HDITA and MDITA Examples

```html
<!DOCTYPE html>
<html>
  <head>
    <title>Installing and Setting up Remote Lighting</title>
  </head>
  <body>
    <article id="install-and-setup">
      <h1>Installing and Setting up Remote Lighting</h1>
      Installation of your lighting kit includes installing the light bulbs into light fixtures, preparing the remote control, and programming lighting groups.

      <h2>Steps</h2>
      <ul>
        <li>Install light bulbs.</li>
        <li>Prepare remote control.</li>
        <li>Program lighting groups.</li>
      </ul>
    </article>
  </body>
</html>
```

# Installing and Setting up Remote Lighting
Installation of your lighting kit includes installing the light bulbs into light fixtures, preparing the remote control, and programming lighting groups.

## Steps
1. Install light bulbs.
2. Prepare remote control.
3. Program lighting groups.

### MDITA (Core)

```md
## Steps
1. Install light bulbs.
2. Prepare remote control.
3. Program lighting groups.
```

### MDITA (Extended)

```md
## Installing and Setting up Remote Lighting
Installation of your lighting kit includes installing the light bulbs into light fixtures, preparing the remote control, and programming lighting groups.

### Steps
1. Install light bulbs.
2. Prepare remote control.
3. Program lighting groups.
```
There Appears to Be an Immediate Need for MDITA

- Markdown is widely used by developers; is taught in more Computer Science programs these days than XML
  - Software firms are most-likely target for using Markdown
- Markdown and DITA are already being used
  - We have had at least two customers approach us seeking Markdown to DITA integration
- Presentation I did with Leigh White at the DITA North America conference earlier this year had standing-room only attendance
Four *Current* Scenarios for Using Markdown and DITA

- Stan Doherty works at Simplivity, a computer hardware/software firm, recently acquired by HPE
  1. Integrate Markdown topics as “blobs” (binary objects)
  2. Integrate Markdown content from API frameworks
  3. Development Markdown content that is DITA-aware
  4. Round-trip DITA for Markdown collaboration
Markdown to DITA Conversion in oXygen

- Open a Markdown file in oXygen, it is converted automatically to DITA and HTML
- Includes “Export to DITA” right-click function
Newly-available LwDITA Plugin for DITA OT 2.2

• Developed by Jarno Elovirta, available at: github.com/jelovirt/dita-ot-markdown

• Designed to work with HDITA and MDITA
Still to Come: LwDITA Template-based Specialization

- Draft model outlined at: markmail.org/message/pd4u5kfg44xp5x5c

1. Create an instance of the parent type
2. Annotate the instance to define model
3. Generate the specialized template
4. Reuse specialization using conref
How This Would Work

• New/repurposed set of elements and attributes designed for creating specializations:
  • `specmeta`: contains any number of `ph`, `data`, or `specatt` elements
  • `specatt`: contains text describing that portion of the specialization
  • `@specmodel`: determines whether a content model is sequence, choice or inherited (default)
  • `@specrole`: whether an element’s content should be generated by the transform, editable by the author (default), an editor prompt, documentation, or used for content modeling purposes only (or a mix of any the above)
  • `@outputclass`: provides the element-specific class value for the specialized element
Draft Example Specialization Template Code

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE topic PUBLIC "-//OASIS//DTD DITA Topic//EN" "topic.dtd">
<topic id="termdef_term" outputclass="tlotermtopic">
  <title outputclass="tloterm">Structured Content</title>
  <prolog outputclass="tlotermprolog">
    <data outputclass="tlotermauthor" specrole="prompt">author name here</data>
    <specmeta>
      <ph outputclass="tiophrase" specrole="doc">A new global phrase element</ph>
      <data outputclass="tlodata" specmodel="choice"
        specrole="modelonly">Simple text only for this global data specialization, but with a different specmodel you could do anything</data>
      <specatt outputclass="tloatt" specrole="doc">A conditional processing attribute called tloatt</specatt>
    </specmeta>
    <body outputclass="tlotermbody" specmodel="sequence">
      <section outputclass="tlowhat">
        <title specrole="generate">What is it?</title>
        <p>...</section>
      <section outputclass="tlowhy">
        <title specrole="generate">Why is it important?</title>
        <p>...</section>
      <section outputclass="tloessay">
        <title specrole="generate">Why does a technical writer need to know this?</title>
        <p>...</section>
      <section outputclass="tlosummary"
        collection-type="sequence">
        <title specrole="generate">Summary:</title>
        <p>...</section>
    </body>
  </prolog>
</topic>
```

- Draft code example created by Don Day
A Pre-release Version of the CN is Available Now

• All actions from OASIS are open and visible to the public, so you can view the latest draft of the Committee Note (v.21) at: www.oasis-open.org/committees/document.php?document_id=61809 (or: goo.gl/7Y3E67)

• A separate CN is planned to cover template-based specialization
Multimedia Additions to DITA 1.3

- Intention is to make available HTML5’s multimedia extensions to DITA
- Originally this was part of LwDITA proposal, DITA TC then realized that this was a serious oversight and ought to have been in DITA 1.3
- Plan is to incorporate this within DITA 1.3 as an add-on rather than wait for DITA 2.0
  - Will be released as a Committee Note, along with DTDs
  - This will also likely to be released before year’s end
## HTML5 Multimedia Elements

<table>
<thead>
<tr>
<th>Component</th>
<th>XDITA</th>
<th>HDITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td><code>&lt;audio&gt;</code></td>
<td><code>&lt;audio&gt;</code></td>
</tr>
<tr>
<td>Autoplay</td>
<td><code>&lt;media-autoplay&gt;</code></td>
<td><code>@autoplay in &lt;audio&gt; or &lt;video&gt;</code></td>
</tr>
<tr>
<td>Controls</td>
<td><code>&lt;media-controls&gt;</code></td>
<td><code>@controls in &lt;audio&gt; or &lt;video&gt;</code></td>
</tr>
<tr>
<td>Loop</td>
<td><code>&lt;media-loop&gt;</code></td>
<td><code>@loop in &lt;audio&gt; or &lt;video&gt;</code></td>
</tr>
<tr>
<td>Muted</td>
<td><code>&lt;media-muted&gt;</code></td>
<td><code>@muted in &lt;audio&gt; or &lt;video&gt;</code></td>
</tr>
<tr>
<td>Poster</td>
<td><code>&lt;video-poster&gt;</code></td>
<td><code>@poster in &lt;video&gt;</code></td>
</tr>
<tr>
<td>Source</td>
<td><code>&lt;media-source&gt;</code></td>
<td><code>&lt;source&gt;</code></td>
</tr>
<tr>
<td>Track</td>
<td><code>&lt;media-track&gt;</code></td>
<td><code>@track in &lt;audio&gt; or &lt;video&gt;</code></td>
</tr>
<tr>
<td>Video</td>
<td><code>&lt;video&gt;</code></td>
<td><code>&lt;video&gt;</code></td>
</tr>
</tbody>
</table>

- These are designed to match equivalent HTML5 elements
Ongoing Work: Errata for DITA 1.3

- An important step before moving fully to DITA 2.0
- **Goal**: fix descriptive errors, incorrect example code and typos in DITA 1.3 standard
- Started immediately after release of standard, still ongoing
- Released first corrected version of DITA 1.3 in Oct 2016
- Second errata should be out before end of this year
  - All changes must be non-substantive (i.e. the changes cannot fundamentally change how the standard works), and the fixes/improvements to sample code can only help
DITA 2.0

• Very much a work-in-progress within DITA TC
• So far, is less about new features than deprecating/removing little-used features
• Not likely to be released anytime soon; 2019-2021 is a good guess
  • Sooner if DITA TC sticks to conservative agenda and wholly new features are few; though this runs the risk of it becoming “DITA 1.3 Lite”
DITA 2.0 Triage Process

- [github.com/oasis-tcs/dita/projects/2](http://github.com/oasis-tcs/dita/projects/2)
Current DITA 2.0 Proposals Include:

- Separate L&T from “main” DITA
- Make @outputclass a universal attribute
- Remove topicset, topicsetref elements
- Deprecate or remove copy-to attribute
- Remove @xtrf and @xtrc
- Add titlealts element to maps
- Split base and technical content
- Deprecate note type="fastpath"
- Change name of @locktitle to something less ambiguous, and change default to "yes"
- Redesign hazard statement domain
- Modify bookmap design to allow <ditavalref> before front matter, as well as a <keydefs> container to hold key definitions (“publicationmap”)
Potential Future Challenges DITA 2.0 May Face

Following is based on *possible* changes/trends in the industry:

- New content or structure needed to support chatbot content
- Make map content interactive (using an if ... then structure)
- Support to help generate Schema.org SEO metadata
- Support for Intelligent Information Request and Delivery Standard (iiRDS)
- Other industry 4.0 considerations
In Summary

• LwDITA coming soon; breaking DITA away from XML is expected to bring new audiences to the standard
  • Easy DITA development for tech doc teams with XDITA, marketing departments with HDITA, software development with MDITA
  • Markdown + MDITA is getting the most interest from users
• DITA 1.3 Multimedia elements also coming soon
• DITA 2.0 is very much under development

• If you want to participate in the future of DITA, join OASIS!
Your opinion is important to us! Please tell us what you thought of the lecture. We look forward to your feedback via smartphone or tablet under http://ta27.honestly.de

or scan the QR code

The feedback tool will be available even after the conference!