Mitchell! What I Learned When Converting a WWII Pilot Manual to DITA

KEITH SCHENGILI-ROBERTS

PRECISION CONTENT.com
Making information easier to use

Agenda

1. Who is this guy?
2. What is the need for DITA samples?
3. Some of the better code examples available
4. How and why we choose the Pilot's manual for the B-25 Mitchell
5. Converting the text to DITA
6. Steps in the DITA conversion process
7. Deficiencies with DITA 1.3
8. Conclusion

Who is this guy?

Keith Schengili-Roberts

- Recently joined Precision Content as a Content Specialist and Information Architect
- Member of OASIS DITA Technical Committee
- Chair of DITA Adoption Committee
- Professor of Information Architecture at University of Toronto
- The “Writer” behind DITAWriter.com
- Keith@PrecisionContent.com


ConVEx
Virtual Experience
CIDM
"Build trust in your content."
PRECISION CONTENT.com
Make your information easier to use
Your partner in ...
Structured Authoring
Writing
Conversion
Training
Content Strategy
Information Architecture
Consulting & Support
Your partner in ...
Content Management
PRECISION CONTENT.com
PRECISION CONTENT.com
Make your information easier to use
Your partner in ...
Publishing Technologies
DITAinPrint
GET THE WORD OUT
DITAonPoint
ENTERPRISE PUBLISHING
WittyDITA
SHARE YOUR CONTENT
PRECISION CONTENT.com
@PCASinc


ICBC
Insurance Corporation of British Columbia
xerox
coastcapital SAVINGS
Westoba CREDIT UNION LTD.
Genetec™
Lilly
Syneos Health
SASB SUSTAINABILITY ACCOUNTING STANDARDS BOARD
TSYS®
BCAA CAA
LE CORDON BLEU 1895 PARIS
NCCI
REDSEAL
conexus Credit Union
KPMG
shopify
Bell
CHRISTIE®
eviCore healthcare innovative solutions
AJCC American Joint Committee on Cancer Validating science. Improving patient care.
Ultimate SOFTWARE People first.
Celestica™
MERCK
mastercard
BARCLAYS
Moneris®/MD
HYDRAFORCE
Children's℠ Healthcare of Atlanta

Our clients

Why the need for DITA code samples?

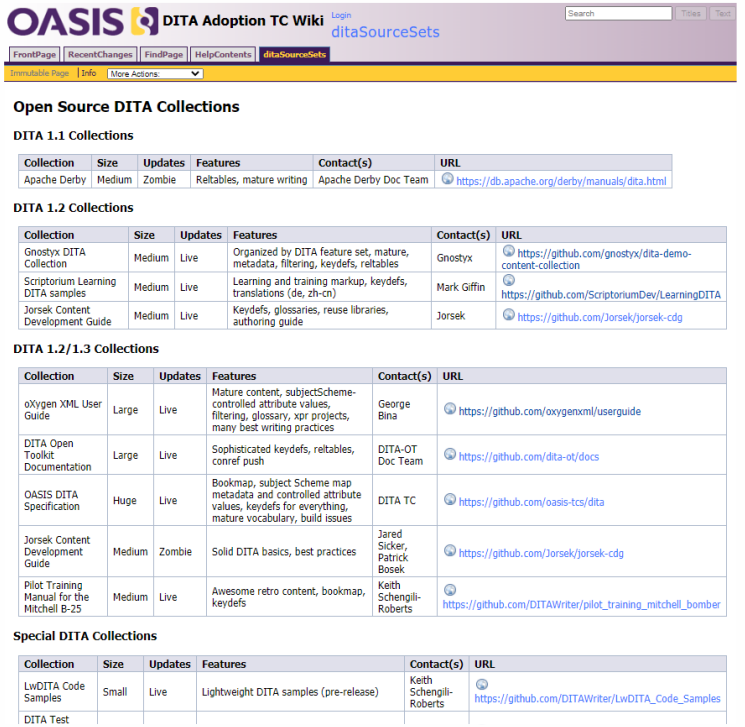
One of the ways people learn is from looking at the examples provided by others

- It's one thing to read a specification, or to work your way through a lesson, and looking at code samples can help demonstrate how things are done in practice
- Code samples can also serve as a good reference when you need to double-check how a feature is supposed to work

Existing DITA code examples

Stan Doherty (also of the DITA Adoption Committee) has compiled a list at: wiki.oasis-open.org/dita-adoption/ditaSourceSets (bit.ly/3aeqHL2)

- Includes information as to what version of DITA is used, who made it, whether it is still being updated, and where to find it



OASIS DITA Adoption TC Wiki — ditaSourceSets

FrontPage | RecentChanges | FindPage | HelpContents | ditaSourceSets

Open Source DITA Collections

DITA 1.1 Collections

Collection	Size	Updates	Features	Contact(s)	URL
Apache Derby	Medium	Zombie	Reltables, mature writing	Apache Derby Doc Team	https://db.apache.org/derby/manuals/dita.html

DITA 1.2 Collections

Collection	Size	Updates	Features	Contact(s)	URL
Gnostyx DITA Collection	Medium	Live	Organized by DITA feature set, mature, metadata, filtering, keydefs, reltables	Gnostyx	https://github.com/gnostyx/dita-demo-content-collection
Scriptorium Learning DITA samples	Medium	Live	Learning and training markup, keydefs, translations (de, zh-cn)	Mark Giffin	https://github.com/ScriptoriumDev/LearningDITA
Jorsek Content Development Guide	Medium	Live	Keydefs, glossaries, reuse libraries, authoring guide	Jorsek	https://github.com/Jorsek/jorsek-cdg

DITA 1.2/1.3 Collections

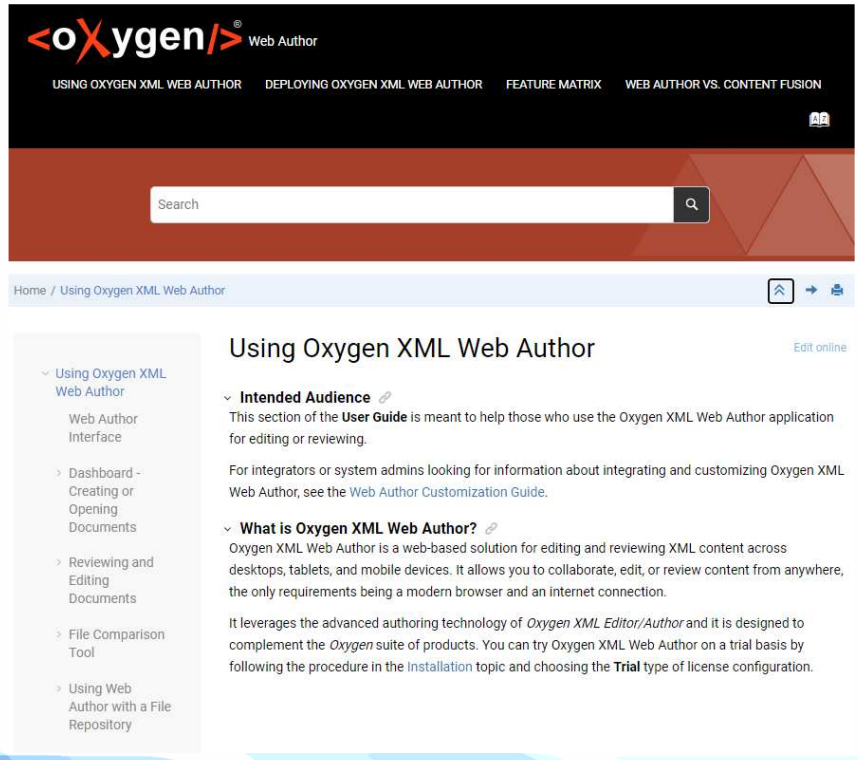
Collection	Size	Updates	Features	Contact(s)	URL
oXygen XML User Guide	Large	Live	Mature content, subjectScheme-controlled attribute values, filtering, glossary, xpr projects, many best writing practices	George Bina	https://github.com/oxygenxml/userguide
DITA Open Toolkit Documentation	Large	Live	Sophisticated keydefs, reltables, conref push	DITA-OT Doc Team	https://github.com/dita-ot/docs
OASIS DITA Specification	Huge	Live	Bookmap, subject Scheme map metadata and controlled attribute values, keydefs for everything, mature vocabulary, build issues	DITA TC	https://github.com/oasis-tcs/dita
Jorsek Content Development Guide	Medium	Zombie	Solid DITA basics, best practices	Jared Sicker, Patrick Bosek	https://github.com/Jorsek/jorsek-cdg
Pilot Training Manual for the Mitchell B-25	Medium	Live	Awesome retro content, bookmap, keydefs	Keith Schengili-Roberts	https://github.com/DITAWriter/pilot_training_mitchell_bomber

Special DITA Collections

Collection	Size	Updates	Features	Contact(s)	URL
LwDITA Code Samples	Small	Live	Lightweight DITA samples (pre-release)	Keith Schengili-Roberts	https://github.com/DITAWriter/LwDITA_Code_Samples
DITA Test					

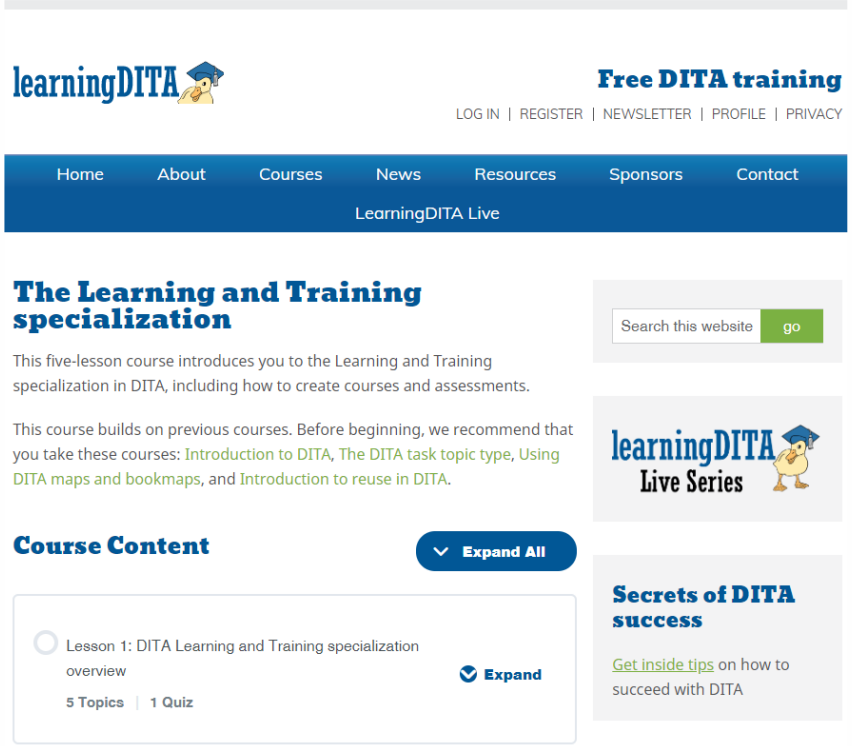
The Better Examples: Oxygen User Guide

- Syncro-soft (the makers of Oxygen) have their user documentation available as source files on GitHub at: github.com/oxygenxml/userguide
- The most extensive example available, containing almost 2,000 topics, ditavals, images, maps, keys, glossaries, and more
- Perfect example of a real-world use of DITA in-action



The Better Examples: Learning DITA

- All lessons for the LearningDITA.com website are written in the Learning & Training specialization, available as code from GitHub at: github.com/ScriptoriumDev/LearningDITA
- Content is translated into Chinese and German, showing localization techniques in practice
- Good example of keydefs as well



The Better Examples: Thunderbird

- A deliberately-created set of example software manuals crafted by Joe Gollner and Eliot Kimber, available from GitHub at: github.com/gnostyx/dita-demo-content-collection
- Particularly good for learning key usage, providing two different ways they can be implemented
- Relatively small and easy-to-grasp

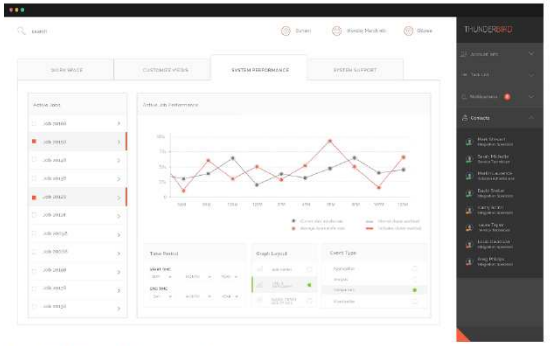


Figure 3: *MobileView* Performance Monitor

Monitoring **System Performance**

The **System Performance** tab provides the ability to monitor the performance of selected cluster hosts and specific processing jobs. In order to facilitate this, the **System Performance** tab provides simple mechanisms for selecting **active jobs** (tasks), setting the **time period** within which analysis and reporting will occur, and the type of **graphical layout** and **event type** will be applied.

- **Active Jobs** Analysis
  - Based on the **Active Clusters** selected in the **Workspace** view, the subordinate active jobs will be displayed.
  - The data stream in the data view will be updated to reflect the active jobs selected.
  - These settings will be maintained until changed.
  - These settings can be saved for later use or shared with colleagues.
- **Time Periods**
  - Use the **Time Period** controls in the **System Performance** view to constrain the analysis being performed to a specific time period.
  - Use the date controls to set a start and end time for the analysis window.
    - When you select a date control, a **Calendar View** will open.
    - Within the **Calendar View**, start and end dates can be chosen.
    - The **Calendar View** provides fields for designating specific start and end times.
    - The **Calendar View** provides the option to define rules that will be used to automatically select date and time ranges.
  - The data stream in the data view will be updated to reflect the time period selected.
  - These settings will be maintained until changed.
  - These settings can be saved for later use or shared with colleagues.

But There is a Need for More Code Examples

- Almost all of the available examples focus on software, and one of the requests that came out of the DITA Listening Sessions was for more code examples, especially for other industries

There's Also a Family-related Reason

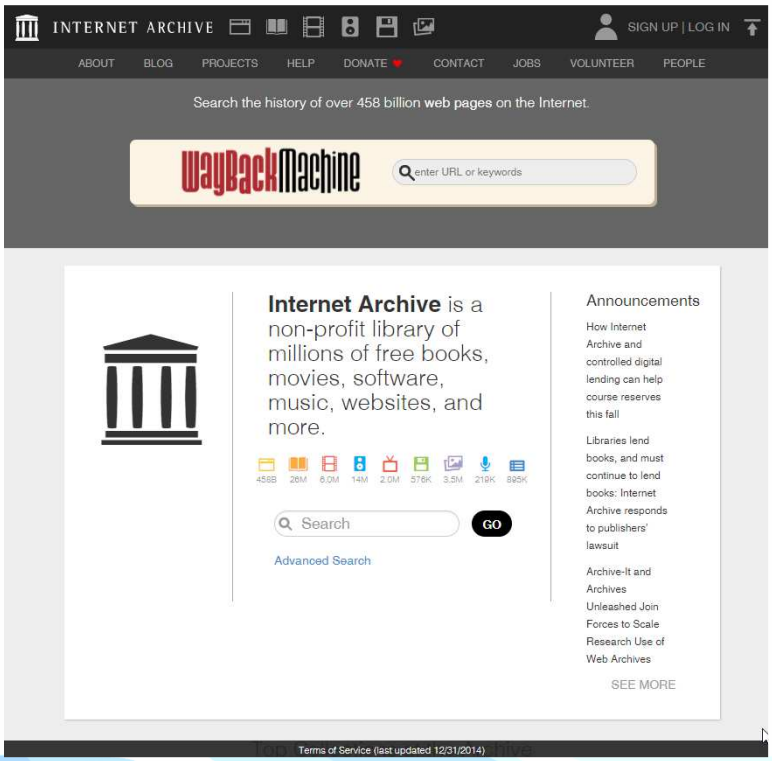
- Both my Grand-Father and Father served in the RAF, and I thought converting an old pilots manual would be a good way to honour their memory
- And others who served in their respective air forces in WWII

Francis Okill Roberts with his Father, WWI

A. Henry Roberts in Egypt, WWII

Finding the Right Manual to Convert

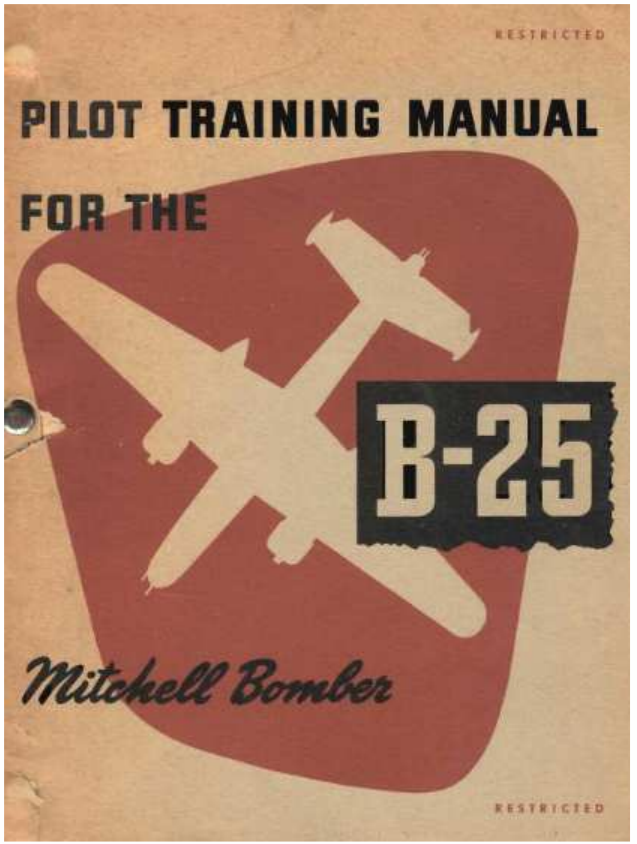
- Scott Hudson (then working for Boeing) suggested something in the aviation domain
- But where to get source material?
- Enter archive.org's extensive collection of out-of-copyright aviation material



What We Chose

In the end, we decided upon this manual:

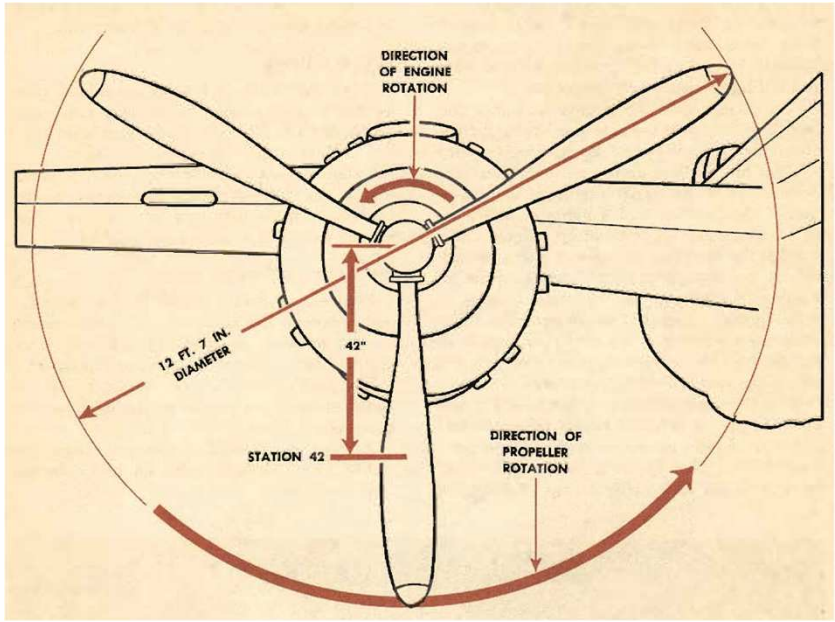
- An instruction manual focusing on the basics of the airplane and its flight characteristics
- A surprisingly good fit, as it was relatively well-written and structured
- Made for an interesting read (which helps when you are converting content)



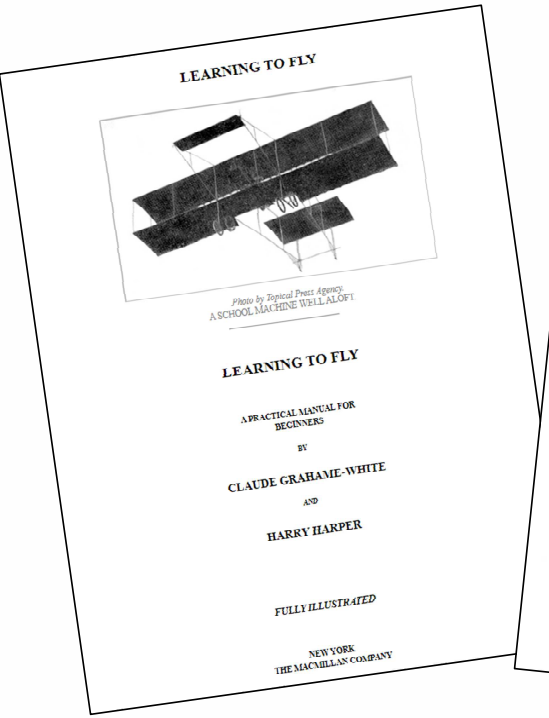
A Comment from Scott

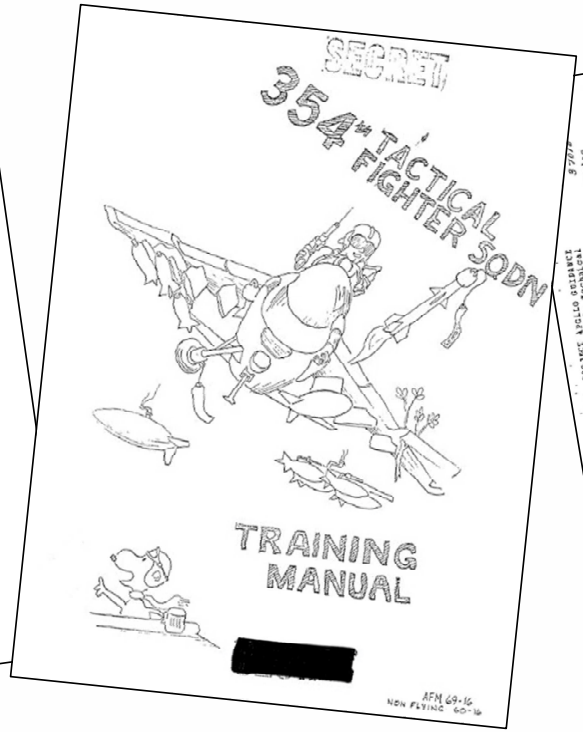
"As it would happen, the DITA Adoption TC was also discussing how to collect 'live' content samples across various industries. One obvious gap was aviation. Keith and I volunteered to take on the challenge of finding a candidate example to convert. I didn't have access, nor permission, to distribute existing Boeing content, so Keith and I found a pilot reference manual for the Mitchell bomber in open source. We found that the manual contained various tasks, illustrations, terms and more. It was a perfect fit for our challenge!"

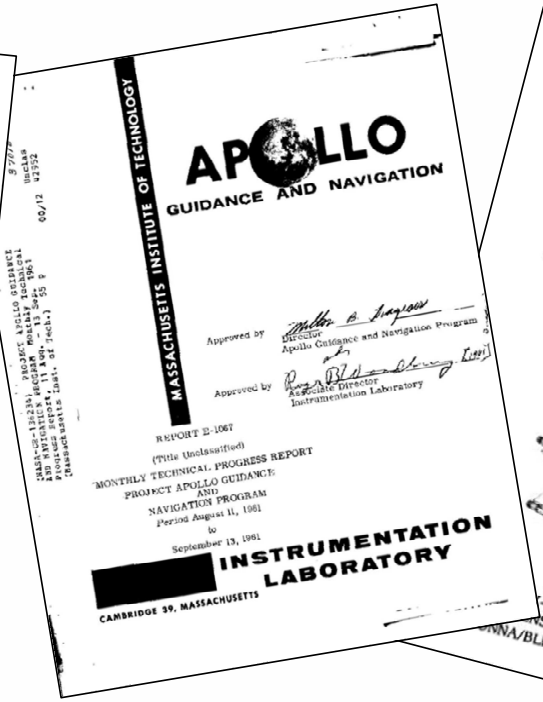
- Scott Hudson

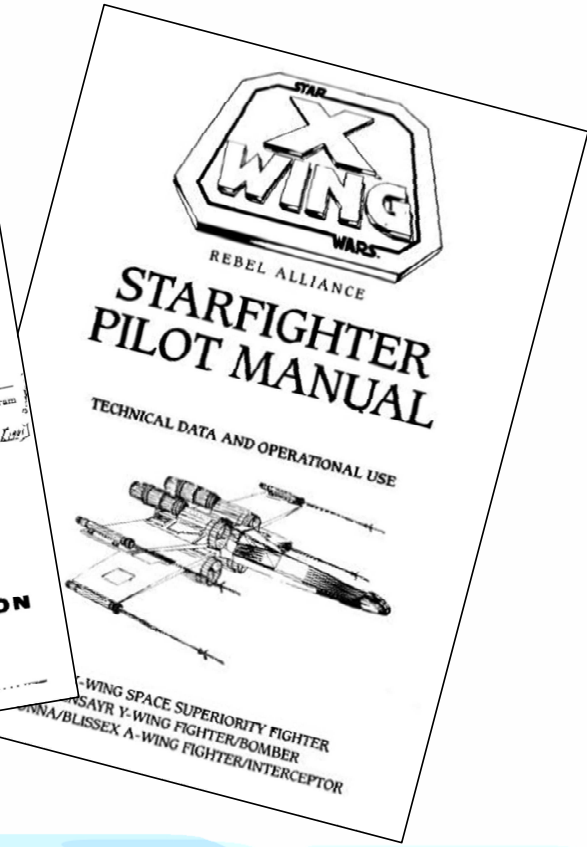


Some of the Other Options…









Some History

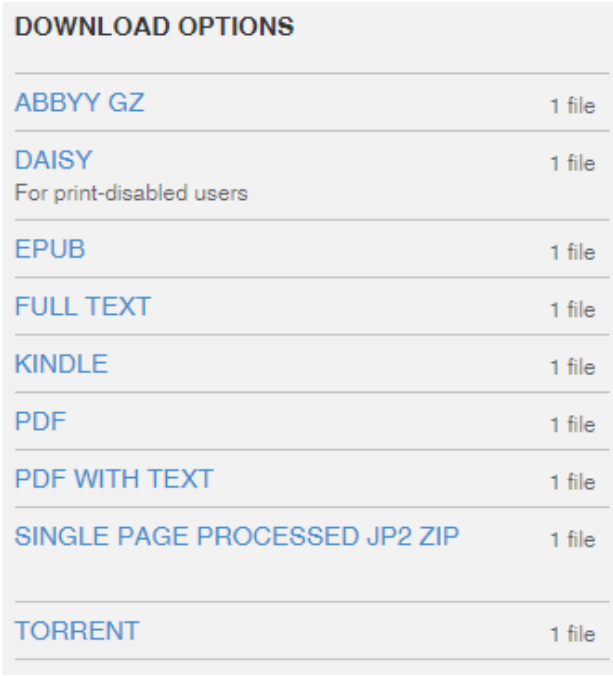
North American B-25 Mitchell

- A medium bomber introduced in 1941
- Nearly 10,000 were built, in 10 major variants (B-25A – B-25J)
- Was the plane chosen in the famous Doolittle Raid
- Used not just by the USAF, but by the RAF, RCAF, RAAF, Dutch Air Force, Soviet Air Force (lend-lease) and more

Starting the Conversion Process

The Internet Archive provides a lot of options in terms of content types

- In the end, a combination of “Full Text” and “PDF with Text” were chosen
- OCR was used on the original text, and having the PDF was handy in order to correct the many mistakes in the text



The Need for “Clean Text”

- Before doing any DITA-related work, we needed to clean up the text that was available
- On the whole it was pretty good, but had problems with faint or low contrast text
- Dumped contents into Word for spell-check pass
- Notepad++ using regex to do search and replace routines

PDF ➡

These are some of the practical questions which you as airplane commander must be able to answer.

OCR ➡

```
These are some of tb;^ pr^^riiciii questions
which you as airplane cornm^ind":r must be
able to answer.
```

PDF ➡

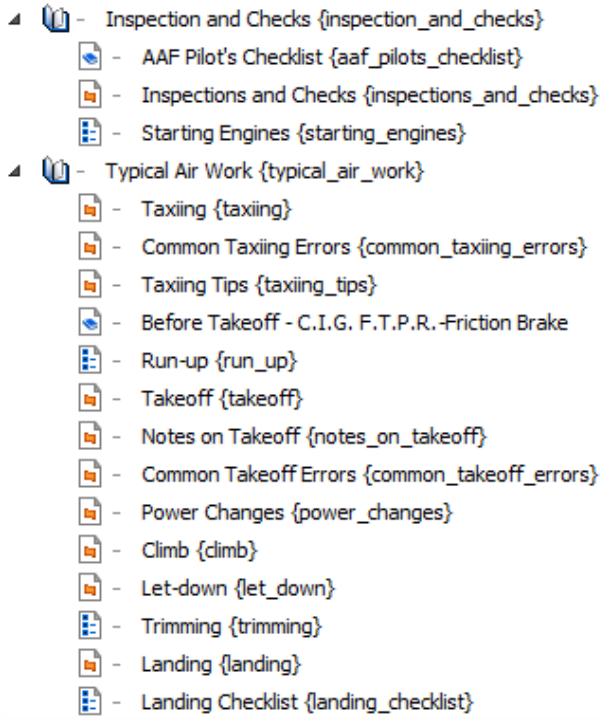
★ The Mitchell medium bomber is a high-speed, mid-wing land monoplane. Positive dihedral in the inner and negative dihedral in the outer wing panels give the plane a gull-wing appearance, while adding control and maneuverability. ★ A twin tail section with large rudders increases stability and maneuverability and allows a greater concentration of firepower to the rear.

OCR ⬇

```
^ Tha Mitchell medium bomber Is q high-speedi mid-witig land monoplcine, PosiHve dihedrai in the
inner and negotive dihedra! in the outer wrng panels give fhe plane a gull-w>ng appearance, wKfle
odding control and matieuverabHity, ^ A twin toil section with large rudders increases stability and
maneuverobility ond allows a greater concentration of firepower to the rear«
```

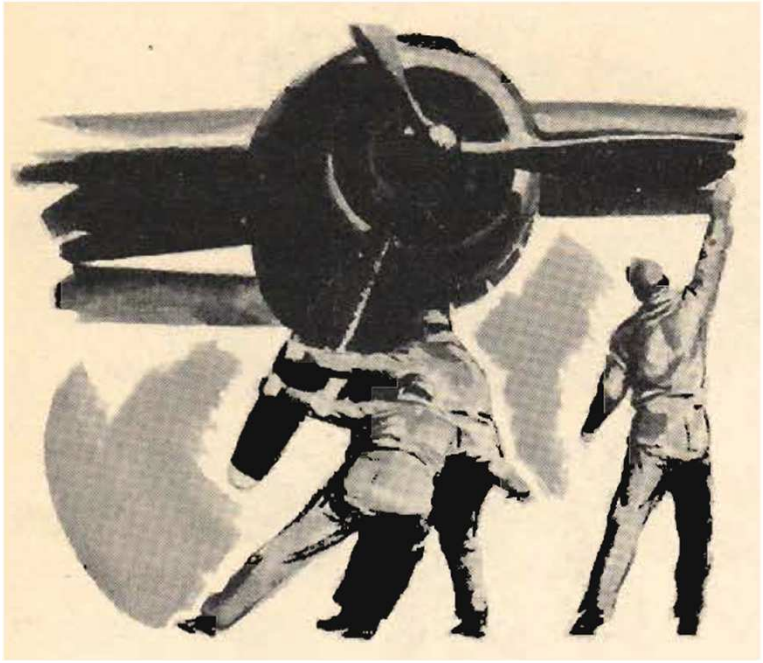
DITA Work Part 1: Topic Typing

- It was relatively straightforward to turn everything into XML and turn the major sections within each chapter into a generic topic
- From there, we went through and typed all of content into concepts, references, and tasks
  - Oxygen came into its own here, making it easy to create typed topics that we could port our content into and add to the growing map



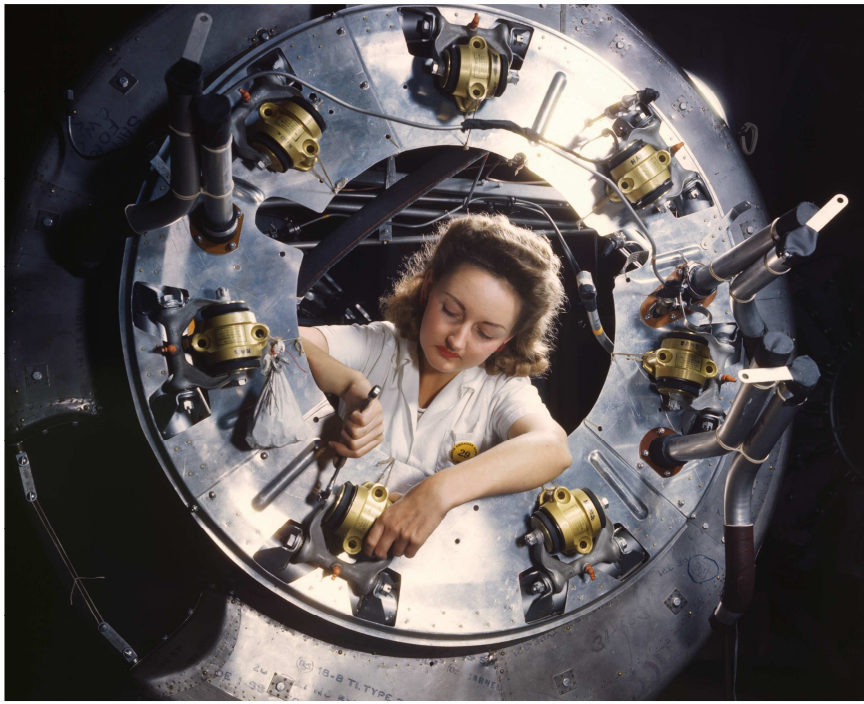
DITA Work Part 2: Adding Images via Keys

- Eliot Kimber donated his time to create a “keystore” of all of the images in the manual, which could easily be referred to within each topic
  - There are 233 images that needed to be “keyed” and then referenced within individual topics (Scott Hudson did most of this work)



DITA Work Part 3: Adding Keys for Content

- In aviation documentation, “designator” is used to cover the series number of the airplane, which in this case is “B-25”
- Used as a key, it prevents incorrect variants, such as “B 25”, “B25”, or “B-52”
- If DITA existed back in the 1940s, writers probably would have used this mechanism for the subsequent B-26 “Marauder” pilot manual

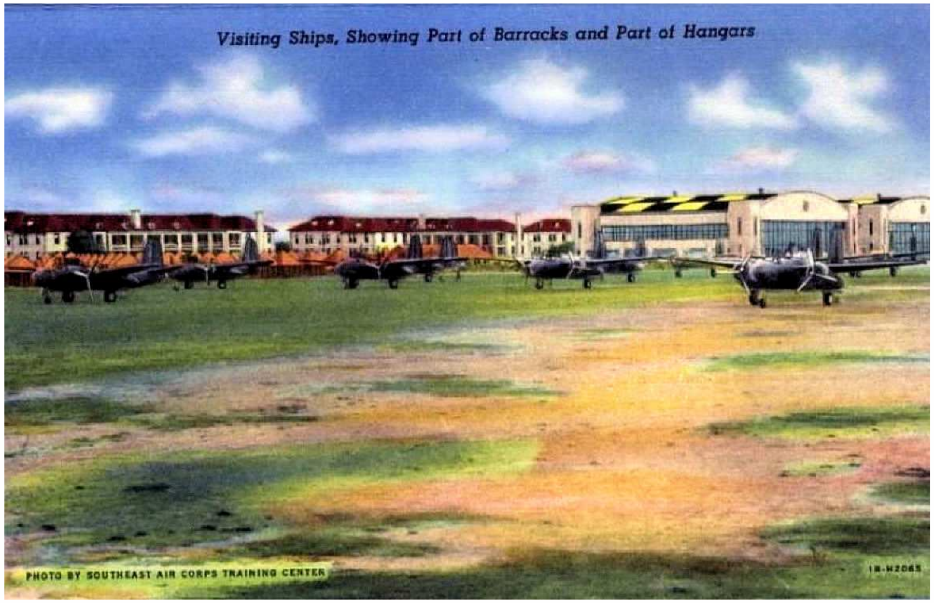


DITA Work Part 4: Adding Glossary Entries

- Very few (if any) other DITA code examples demonstrate how to use the glossentry topic type
  - While what's there is not extensive, it's there and it works
  - This work was also done by Scott Hudson
  - We may expand upon this in the future

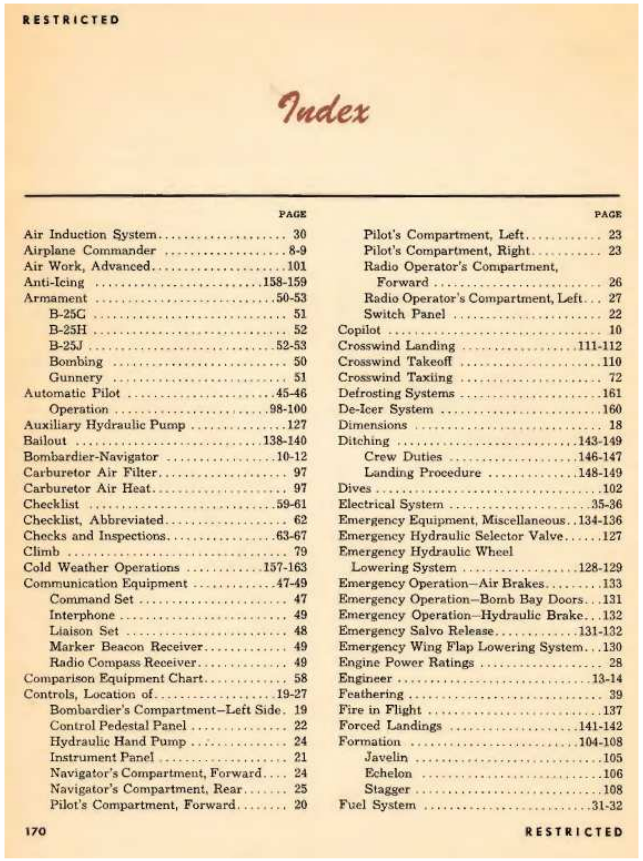
DITA Work Part 5: Short Descriptions

- This needed to be created for every topic
- Following best practices, there are unique, hand-crafted short descriptions for every topic in the manual (156 topics)



DITA Work Part 6: Index Pass

- Where it made sense to do so, we emulated the original indexing as much as possible
- Search-and-replace routines in Oxygen helped make this possible
- Again, this is something rarely done in most DITA examples out there



RESTRICTED

Index

DITA Work Part 7: Accessibility Issues

- All images were given descriptive alt text
  - Few of the images came with captions, so many had to be created
- The tables in the document were also made accessible, providing additional information for screen readers

DITA Work Part 8: Relationship Tables

- Again, this is something not often demonstrated in DITA code examples, so we wanted to do it here
  - We went for a simple approach, adding reltable links to every topic within a chapter, so each topic within a chapter links to its peer
  - May revisit and refine this in the future

DITA Work Part 9: Multimedia Added

- Training films were made back in the day, and these are available on YouTube
  - Some were clearly based on the pilot's manual
- Used this material to demonstrate how and when to use multimedia in a presentation with Leigh White at DITA NA 2019

1. Fly a normal traffic pattern, except that you do not lower landing gear until you are on the approach.

Make *[gl_GUMPS]*GUMPS check here.

2. On the approach, lower and check the gear, check brake and hydraulic pressure, set props at 2400 rpm.

3. Lower half flaps — use the rest of the flaps and the power to increase the accuracy of the approach.

4. Never let the airspeed fall below 140 mph on the approach until you are positive the landing can be made.

5. Reduce power and reduce trim. Never allow the power reduction to get ahead of the trim. The ideal approach is one where the power is reduced. Then the trim, then power, then trim — and so on throughout the approach.

6. Maintain 140 mph until the landing is in the bag, then lower the remaining flaps and reduce power to make an accurate landing.

Landing a [designator]B-25 on one engine

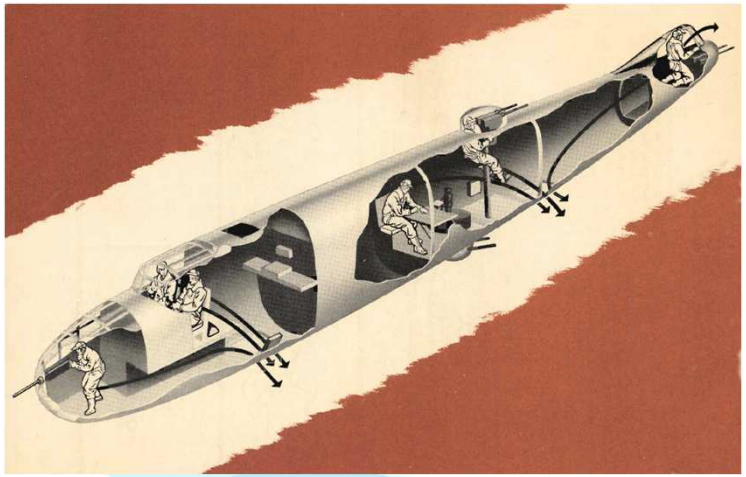
DITA Work Part 10: Converting to LWDITA

- Scott Hudson devised a conversion routine to make this happen
- For obvious reasons, this had to be done at the end of the overall (regular) DITA conversion process
- The LWDITA files are also available from the same GitHub repository

A Comment from Scott

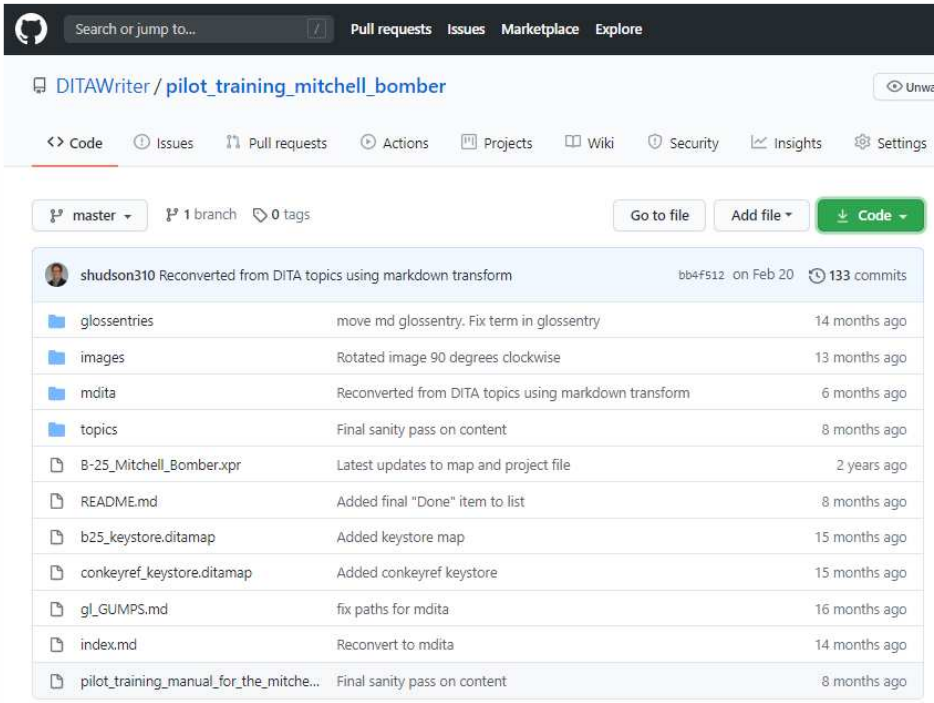
“Not only did we want to validate DITA in an aviation setting, but we also wanted to test the Lightweight DITA specification as well. Once we converted the base content to DITA and cleaned up some of the tagging, we ran a transform to convert the content to HTML and then to MDITA. This was one of the most enjoyable conversion experiences I’ve had in a long time, and the illustrations we came across were just beautiful!”

- Scott Hudson



Using GitHub as a Repository

- Currently available under DITAWriter, but will be moved to a DITA Adoption GitHub repository
  - GitHub is great at storing files and showing where the updates are
  - Not to be confused with a CCMS; definitely not designed for searching for content for example



Where They Can Be Found

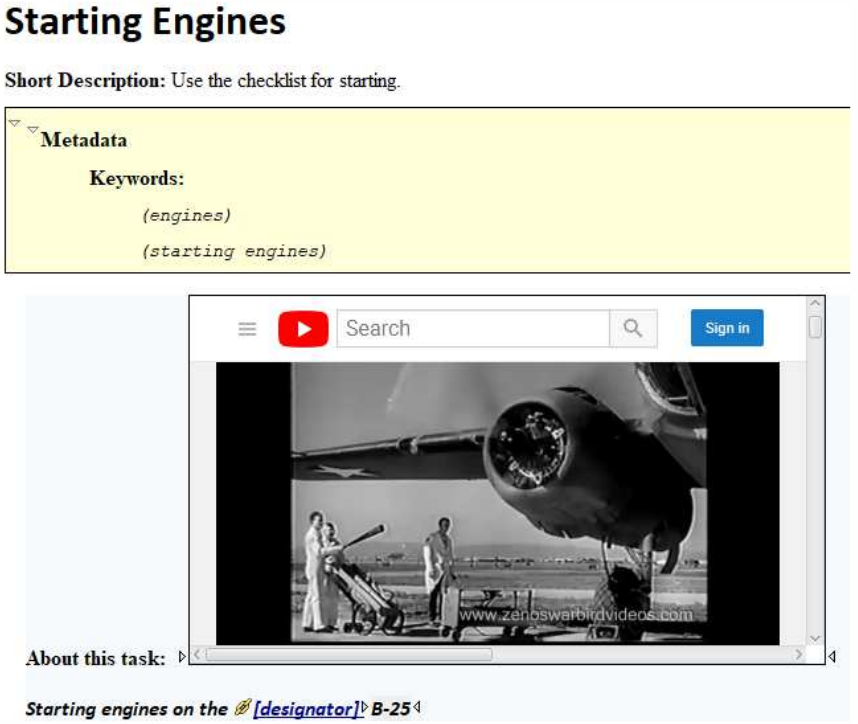
"Full" DITA version:
github.com/DITAWriter/pilot_training_mitchell_bomber

MDITA version:
github.com/DITAWriter/pilot_training_mitchell_bomber/tree/master/mdita

(They will be ported to an OASIS GitHub repository soon)

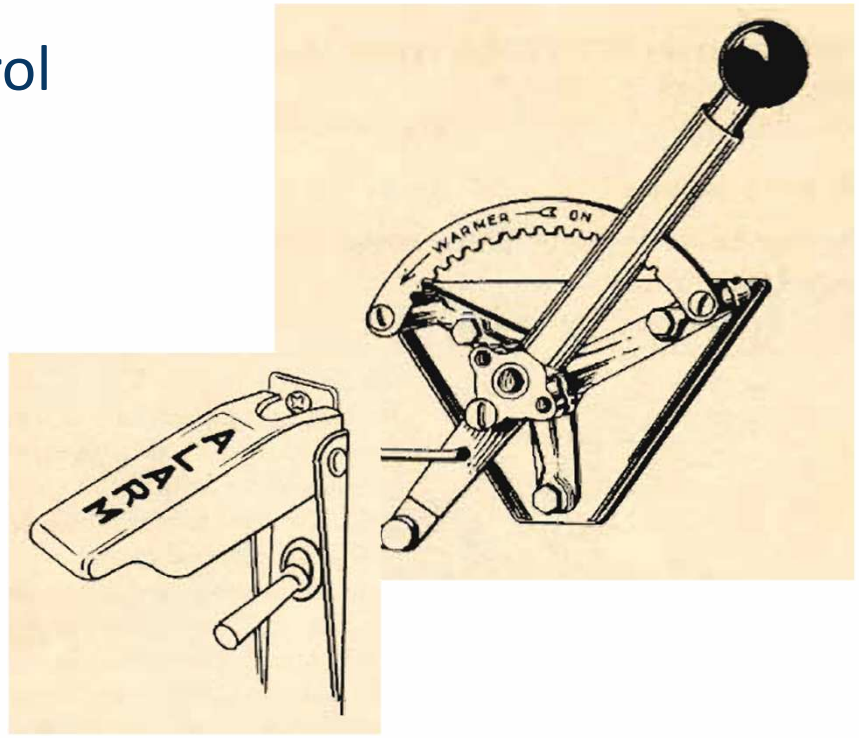
Deficiencies with DITA 1.3: Multimedia

- The presentation done back at DITA NA 2019 illustrated the deficiencies with using multimedia in current DITA
  - Video autostarts by default
  - No fallback image possible
  - Much tweaking necessary in HTML to make it more usable
- This will be fixed in DITA 2.0



Deficiencies with DITA 1.3: Hardware Controls

- It was very odd to use software control elements like `uicontrol` when describing the action for a hardware control, like pulling a lever, pushing a button, or twisting a knob
- Highlighted the need for a hardware-based, non-software controls
- This is also coming in DITA 2.0



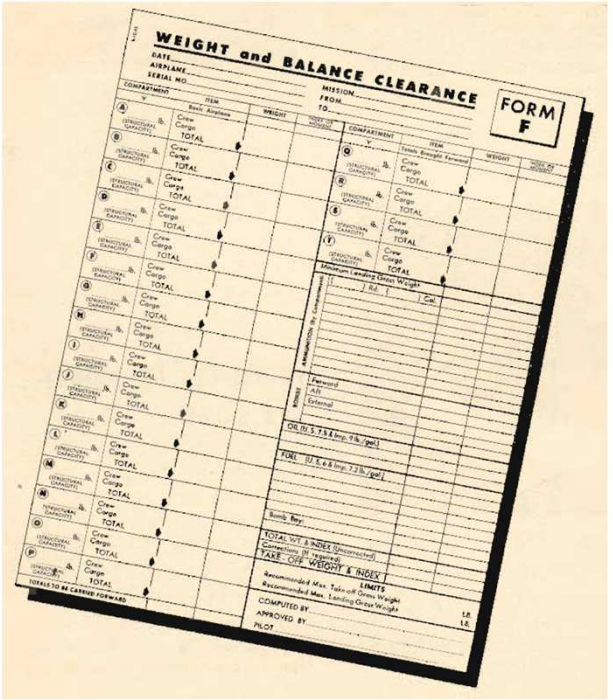
It Was a Great Learning Experience

"While working at Jeppesen, I was looking for an opportunity to expand the use of DITA across the Boeing enterprise, since they make so many different types of aircraft. I wanted to see if there was a way to apply the DITA standard to aviation content in general, which would include manufacturing and assembly instructions, maintenance manuals, safety manuals, and operations guides. Would aviation content require a specialization to support the industry? Inquiring minds want to know! One way to test my theory of applying DITA was to find an existing manual and try to convert it!"

- Scott Hudson

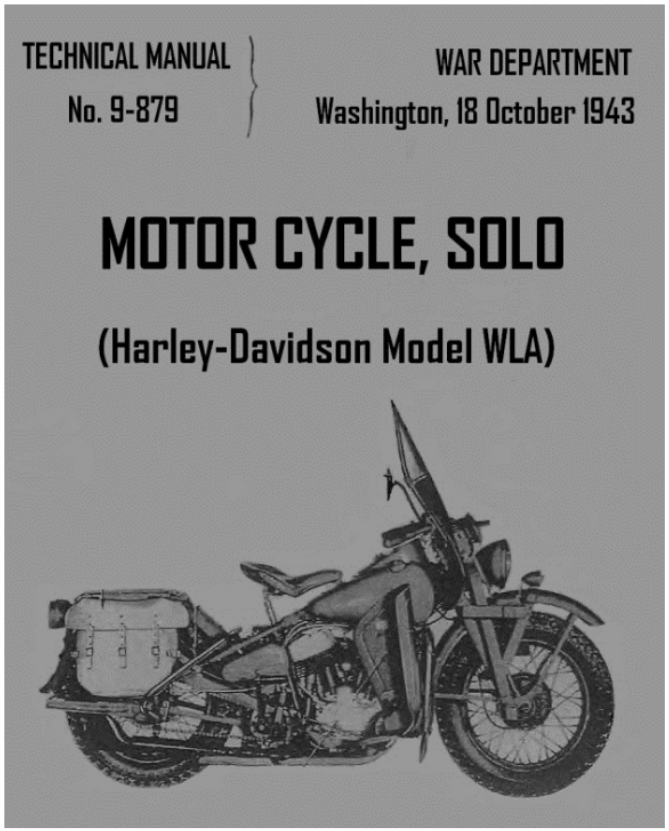
What Would We Do Differently Now?

- I think it makes sense to have a specialization for the many checklists that appear in the manual
  - This may be a future goal
- Plan is to keep refining and tweaking existing content
  - DITA 2.0 and LWDITA updates



What's Next?

- I have been working on a conversion of the following out-of-copyright manual for a little while now…
- If there are any volunteers who want to help, just let me know!



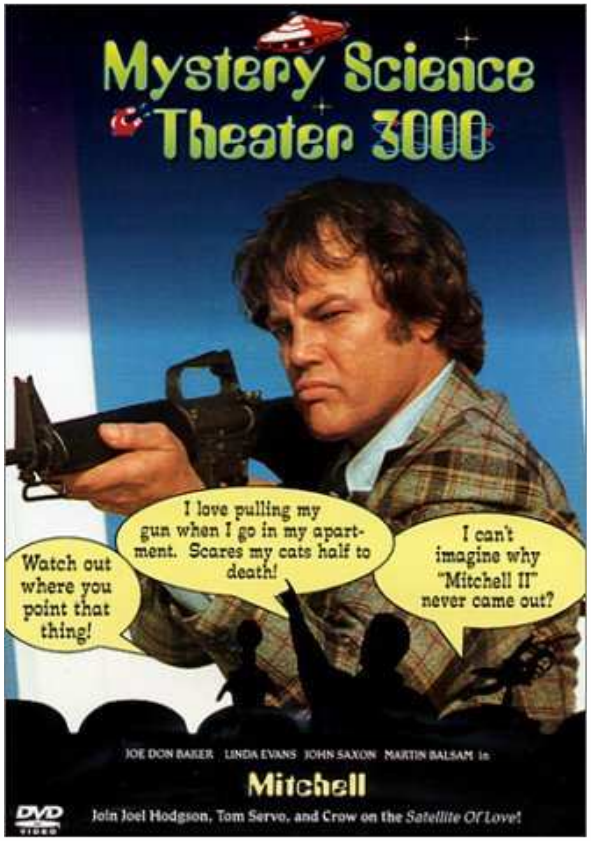
About the OASIS DITA Adoption TC

- OASIS DITA Adoption Technical Committee https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=dita-adoption (or try: goo.gl/5BySBc)
- Twitter: @dita_adoption
- SlideShare: www.slideshare.net/DITA_Adoption
- OASIS DITA Adoption Committee articles: look under “Expository Work Produced by the Committee” on our OASIS webpage

- My thanks to Precision Content for allowing me dedicated time devoted to DITA Adoption efforts

More “Mitchell!”

- Liz Fraley and I are planning to do a riff of an old DITA webinar of mine à la *Mystery Science Theatre 3000*-style later during this conference
- Keep a look out for it in the schedule!



Thank you!

Keith Schengili-Roberts

Information Architect

keith@precisioncontent.com

Precision Content

www.precisioncontent.com

Ask me about …

- Information Architecture
- Content Strategy
- Content Transformation
- Author Training
- Publishing Technologies

DITAinPrint
GET THE WORD OUT

PRECISION CONTENT.com

@PCASinc