Information Architects in Technical Communications and Content Strategy

Keith Schengili-Roberts,
Information Architect

precision content

Keith Schengili-Roberts

Information Architect with Precision Content

What I do:

- Work with clients on structured content IA-issues
- Liaison with OASIS; on DITA Adoption and Technical Committees
- Work with companies who are working with structured content, primarily DITA

Als
Inf
th

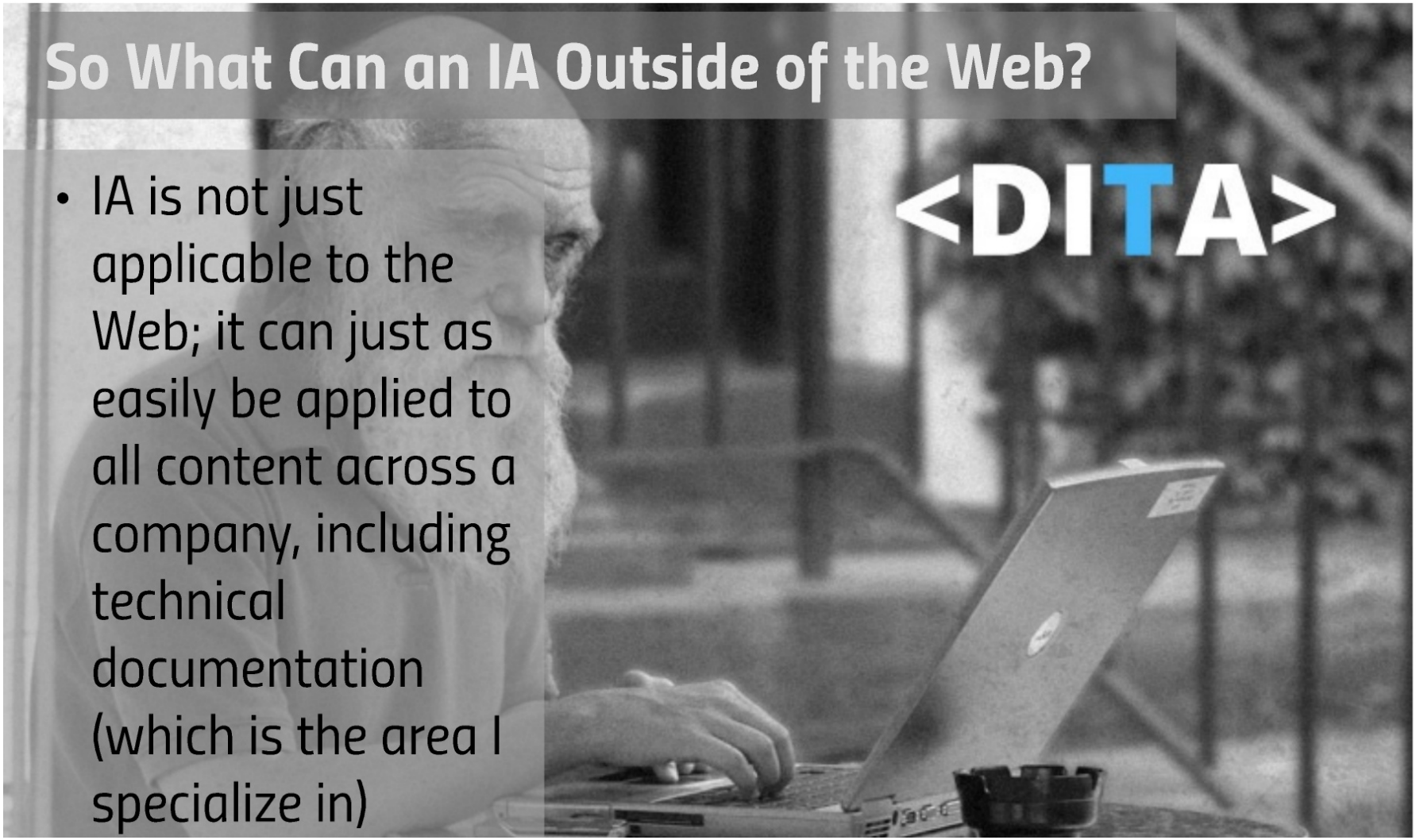
Also long-time lecturer of Information Architecture at the University of Toronto

UNIVERSITY OF TORONTO

- Have worked within firms as an Information Architect, and as an independent consultant

So What Can an IA Outside of the Web?

- IA is not just applicable to the Web; it can just as easily be applied to all content across a company, including technical documentation (which is the area I specialize in)



Content Strategy is based upon...

"A repeatable system that governs the management of content throughout the entire lifecycle."

- Rahel Bailie

Content Strategy and DITA

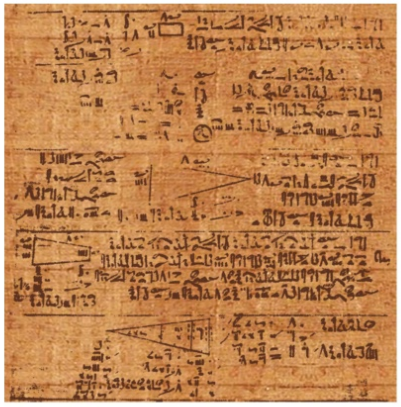
Business Communications	Technical Communications 
Information Architecture & Usability	Information Management

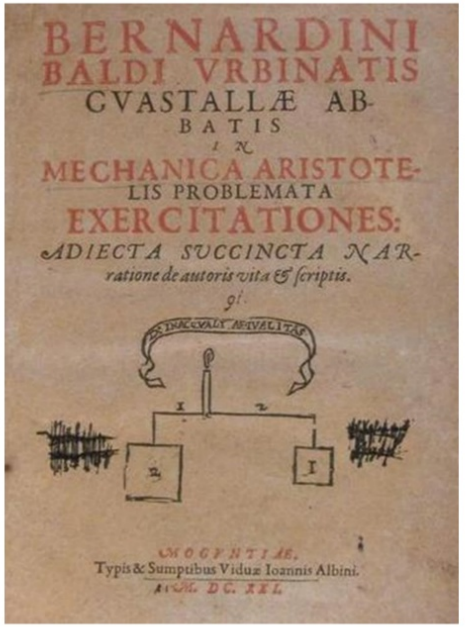
The Beginnings of Technical Writing

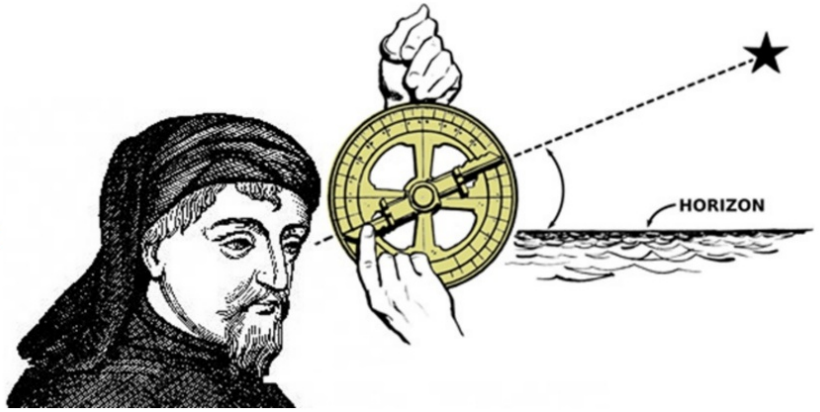
There has always been a need to instruct people on how to do practical things, including:

- Ancient Egyptian medical and mathematical texts
- Certain Ancient Greek philosophical works
- Roman engineering texts
- Medieval instructional publications

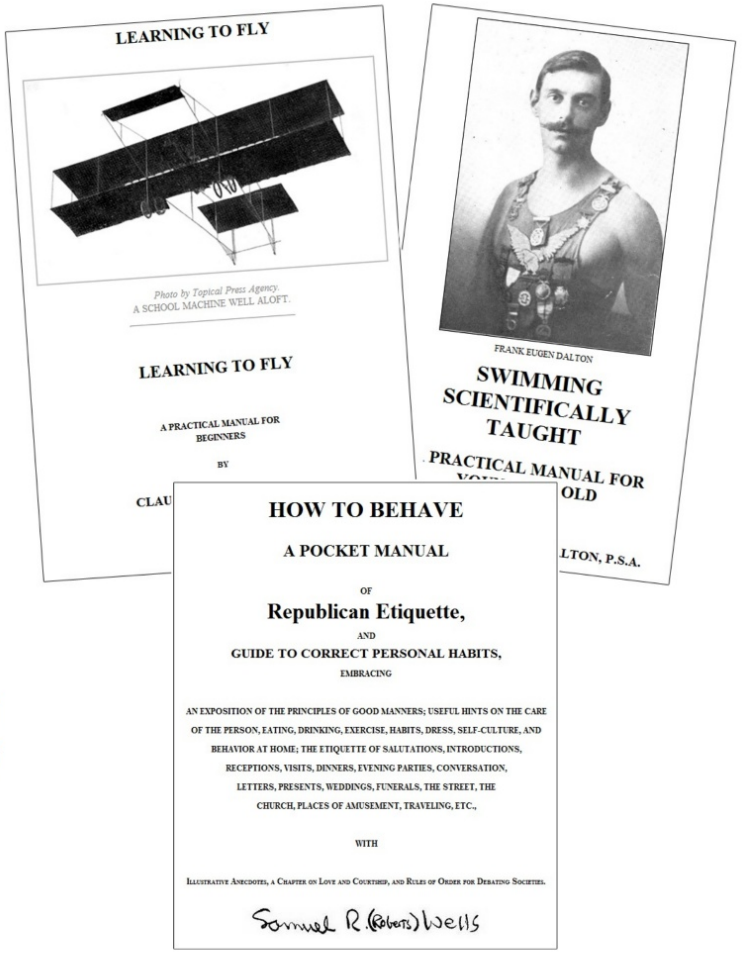
In English, the first instruction manual is considered to be Chaucer's "A Treatise on the Astrolabe"



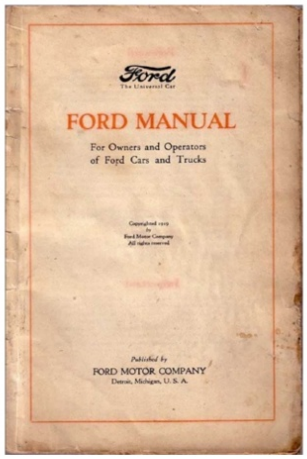


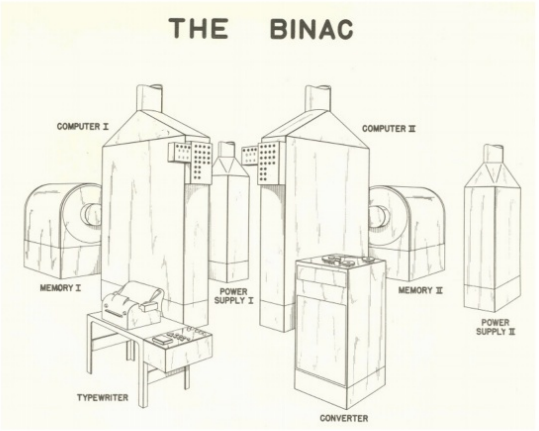


- From then until the early 20th century, manuals tended to focus on techniques, often including aspects of self-improvement
- Tended to contain lots of anecdotes, were often prescriptively moralistic, and came with suggestions on how to become a better person who could fight/ fly/swim/behave.

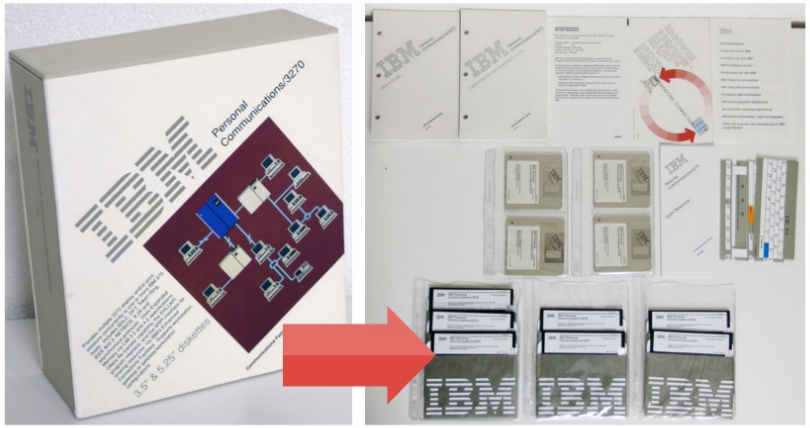


- Technical writing as we know it started in early 20th century; focused on instructing people how to work with technology
  - Ford Model T manual (1919) is a good example of this
  - Joseph Chapline wrote first computer manual for BINAC (1949)
- Was often done by a single person, often a Subject Matter Expert (SME)
- They tended to idiosyncratic and often one-offs





- Growth of technical writing developed alongside growth of software industry
- DeskTop Publication (DTP) programs made it possible to writers to create documents
- More documents were produced, often in volume
- Back in the day, software documentation was considered the "bulk" of the product; it made up most of the box that shipped with the product

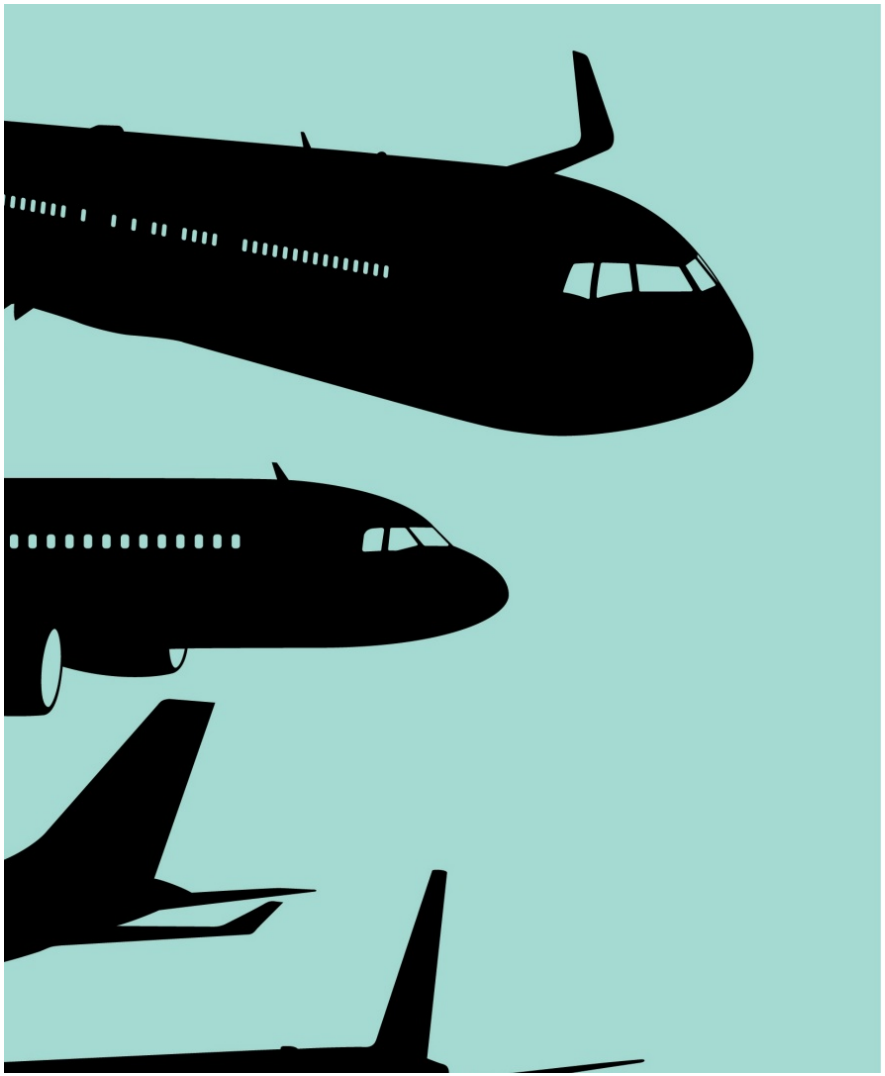


Reportedly, some documentation was delivered along with the software on a wooden pallet, filling the client's bookshelves with manuals


The physical weight of documentation became a real problem...

In the 1980s, A Boeing 747 required the following in terms of documentation:

- 129,000 parts, each requiring extensive documentation
- 31,000,000 pages of docs
- @ 10 minutes per page, would take a single technical writer 216,000 24hr days to write



- total weight of 31 Million pages of docs: 450 tons
- total weight of a Boeing 747: 437.5 tons
- with digitization, this same amount of information could fit on a few CD-ROMs
- however, there was still the task of organizing this info

While this solved the problem of document weight, creating and formatting this PDF output was still an ongoing issue

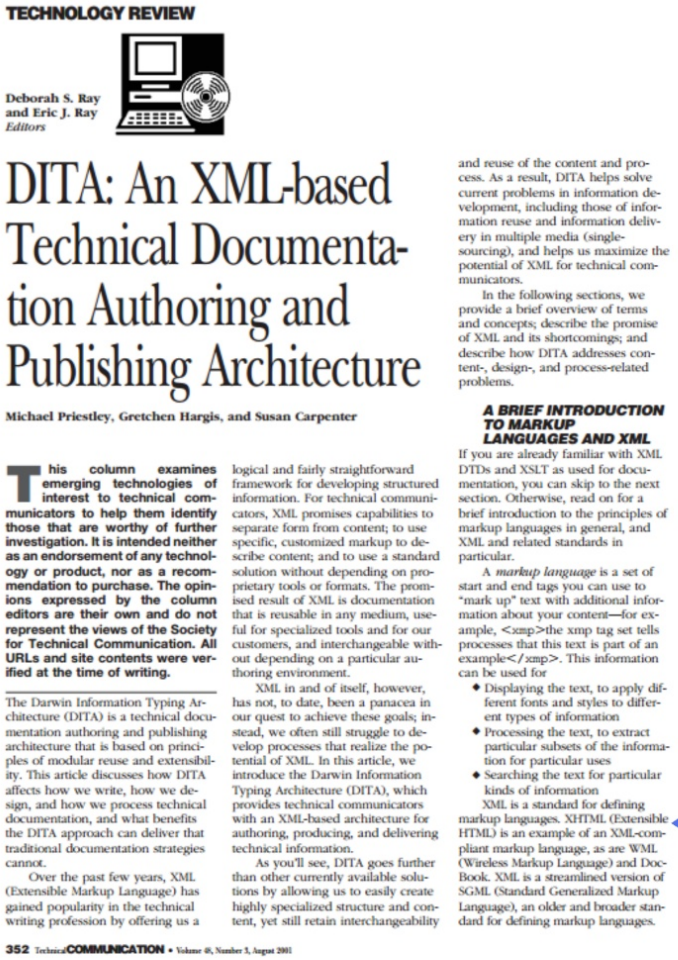
The Roots of DITA

IBM

Software and services

Back in the late 90s they were facing two problems:

- content needed to be tailored for the web
- too many output formats, needing lots of tools



TECHNOLOGY REVIEW

Deborah S. Ray and Eric J. Ray
Editors

DITA: An XML-based Technical Documentation Authoring and Publishing Architecture

Michael Priestley, Gretchen Hargis, and Susan Carpenter

This column examines emerging technologies of interest to technical communicators to help them identify those that are worthy of further investigation. It is intended neither as an endorsement of any technology or product, nor as a recommendation to purchase. The opinions expressed by the column editors are their own and do not represent the views of the Society for Technical Communication. All URLs and site contents were verified at the time of writing.

The Darwin Information Typing Architecture (DITA) is a technical documentation authoring and publishing architecture that is based on principles of modular reuse and extensibility. This article discusses how DITA affects how we write, how we design, and how we process technical documentation, and what benefits the DITA approach can deliver that traditional documentation strategies cannot.

Over the past few years, XML (Extensible Markup Language) has gained popularity in the technical writing profession by offering us a logical and fairly straightforward framework for developing structured information. For technical communicators, XML promises capabilities to separate form from content; to use specific, customized markup to describe content; and to use a standard solution without depending on proprietary tools or formats. The promised result of XML is documentation that is reusable in any medium, useful for specialized tools and for our customers, and interchangeable without depending on a particular authoring environment.

XML in and of itself, however, has not, to date, been a panacea in our quest to achieve these goals; instead, we often still struggle to develop processes that realize the potential of XML. In this article, we introduce the Darwin Information Typing Architecture (DITA), which provides technical communicators with an XML-based architecture for authoring, producing, and delivering technical information.

As you'll see, DITA goes further than other currently available solutions by allowing us to easily create highly specialized structure and content, yet still retain interchangeability and reuse of the content and process. As a result, DITA helps solve current problems in information development, including those of information reuse and information delivery in multiple media (single-sourcing), and helps us maximize the potential of XML for technical communicators.

In the following sections, we provide a brief overview of terms and concepts; describe the promise of XML and its shortcomings; and describe how DITA addresses content-, design-, and process-related problems.

A BRIEF INTRODUCTION TO MARKUP LANGUAGES AND XML

If you are already familiar with XML DTDs and XSLT as used for documentation, you can skip to the next section. Otherwise, read on for a brief introduction to the principles of markup languages in general, and XML and related standards in particular.

A *markup language* is a set of start and end tags you can use to "mark up" text with additional information about your content—for example, <xmp>the xmp tag set tells processes that this text is part of an example</xmp>. This information can be used for

- Displaying the text, to apply different fonts and styles to different types of information
- Processing the text, to extract particular subsets of the information for particular uses
- Searching the text for particular kinds of information

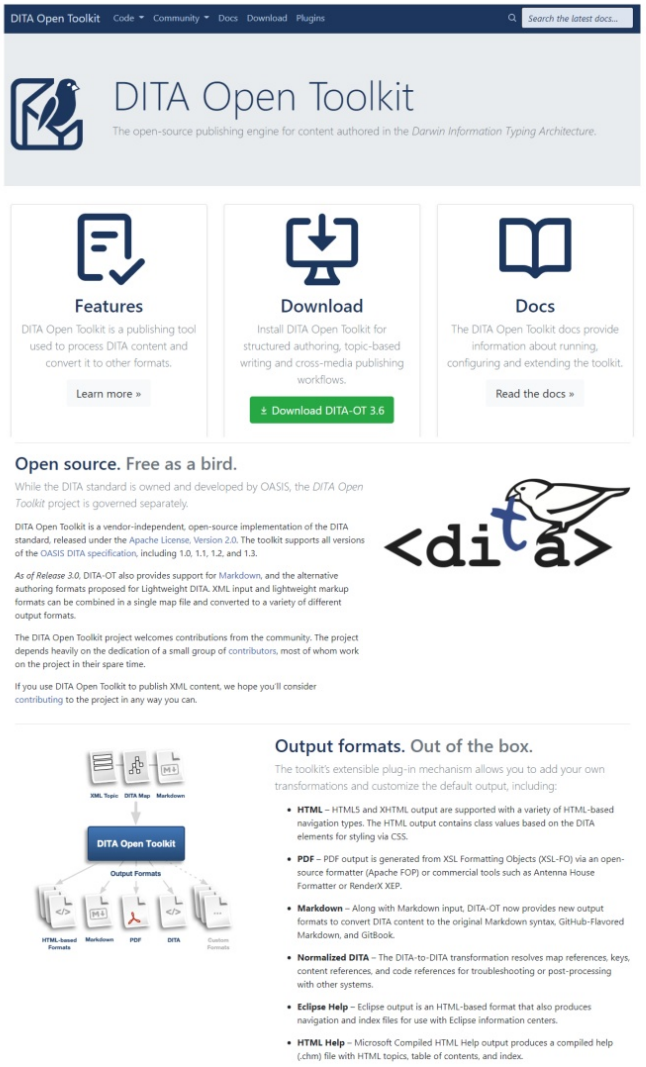
XML is a standard for defining markup languages. XHTML (Extensible HTML) is an example of an XML-compliant markup language, as are WML (Wireless Markup Language) and DocBook. XML is a streamlined version of SGML (Standard Generalized Markup Language), an older and broader standard for defining markup languages.

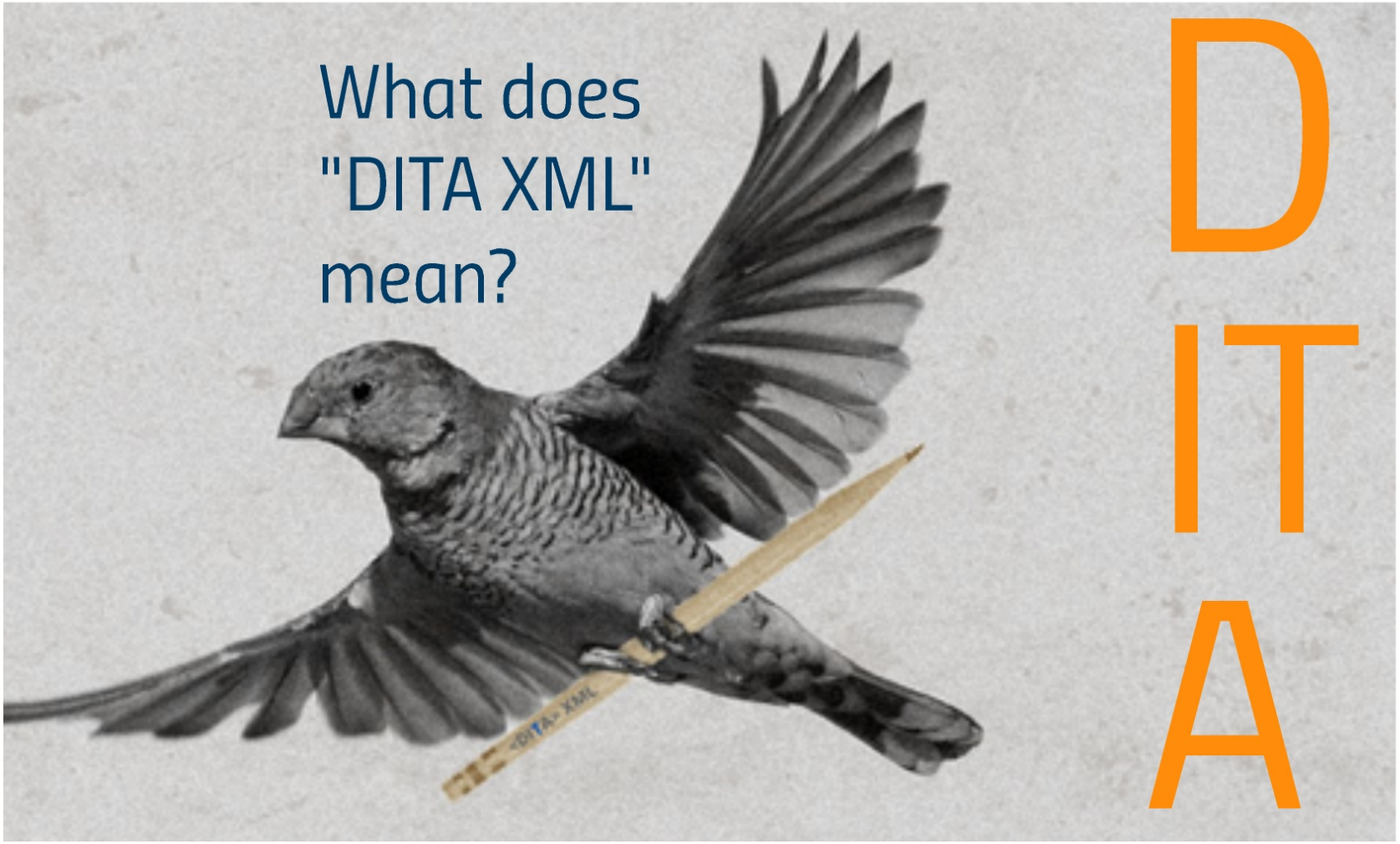
- IBM opted to create an XML-based, topic-typed documentation standard, called "DITA"
- 1 Topic = 1 Web page
- Made it an open standard so that they could benefit from sharing info with their business partners

The first public paper on DITA, published in August 2001

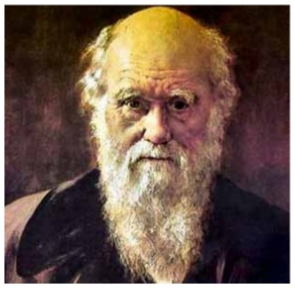
(one of its authors is speaking here later today)

Publishing was taken care of by creating the initial version of the DITA Open Toolkit, which took DITA content and published it to HTML5, HTML Help, PDF, XHTML, troff and more




What does
"DITA XML"
mean?
D
IT
A
is for D
Allows for
document
and specia
is for I
Provides an inf
for technical do
topic types of C
Task (and later,
Troubleshootin
is for
A model, to be
practices for b

D



is for Darwin

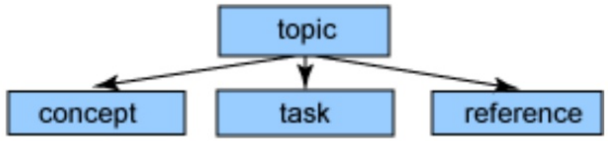
Allows for the natural evolution of document types through inheritance and specialization.

is for Darwin

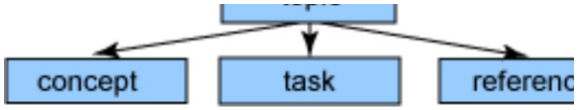
Allows for the natural evolution of document types through inheritance and specialization.

is for Information Typing

Provides an information architecture for technical documents with the base topic types of Concept, Reference and Task (and later, Glossary and Troubleshooting)



for technical documents with the base topic types of Concept, Reference and Task (and later, Glossary and Troubleshooting)



is for Architecture

A model, to be built upon, that encapsulates best practices for both design and the authoring process

XML

...is modular, non-narrative and designed with content reuse in mind

changingtheoil.xml* | conref-file2.dita* | hierarchy.ditamap*

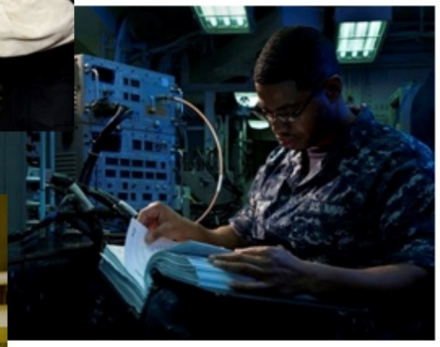
```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!DOCTYPE task PUBLIC "-//OASIS//DTD DITA Task//EN" "http://docs.oasis-open.org/dita/v1.1/OS/dtd/ta
3  <task id="conref-file">
4      <title>Conref file</title>
5      <taskbody>
6          <steps>
7              <step conaction="pushbefore">
8                  <cmd>Recyle the motor oil. In Durham, North Carolina, you can take it to the Waste
9                      Disposal and Recycling Center.</cmd>
10             </step>
11             <step conref="changingtheoil.xml#changeoil/drain-oil" conaction="mark">
12                 <cmd/>
13             </step>
14         </steps>
15     </taskbody>
```

It turned out that the design decision to make DITA content reusable ended up being of the standard's "superpowers"; none of the other major XML documentation standards (DocBook, S1000D) emphasize this

Much of DITA's design ultimately derives from studies done in the U.S. Army and Navy from the 1950's and 1960's seeking more effective documentation.

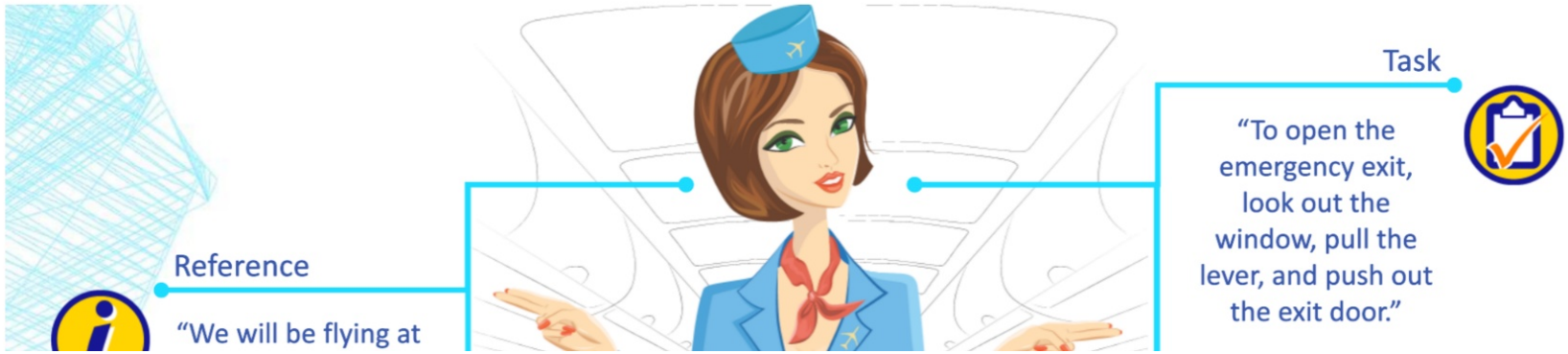
When seeking information quickly, users do not want a story (narrative), they want just the information they need to do the job.

Typed, modular information helps accomplish this



Topic typing enforces a kind of modularity to DITA, and provides focus for technical writers

- Concepts answer the question “What is it?”
- Tasks answer the question “How do I do this?”
- References answer the question “What can I do with this?”
- Process demonstrates to the reader how things work.
- Principle advises the reader about what they need to do or not do

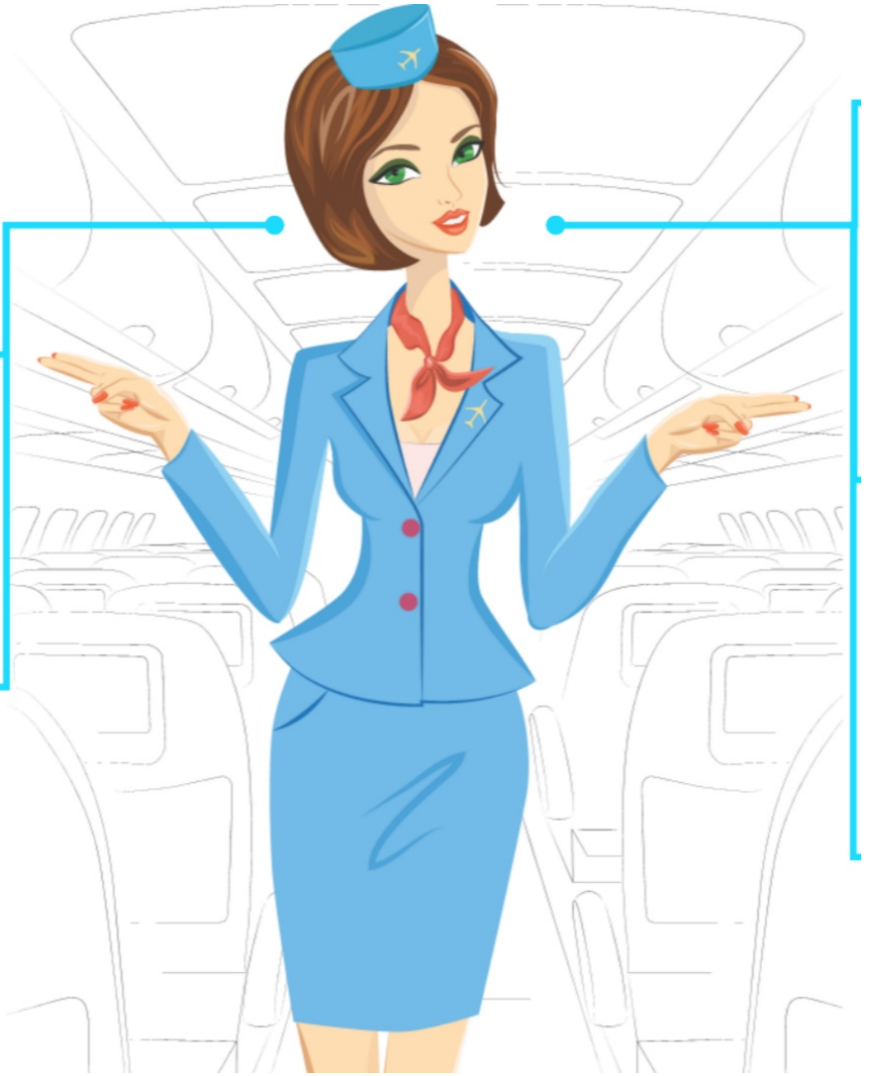


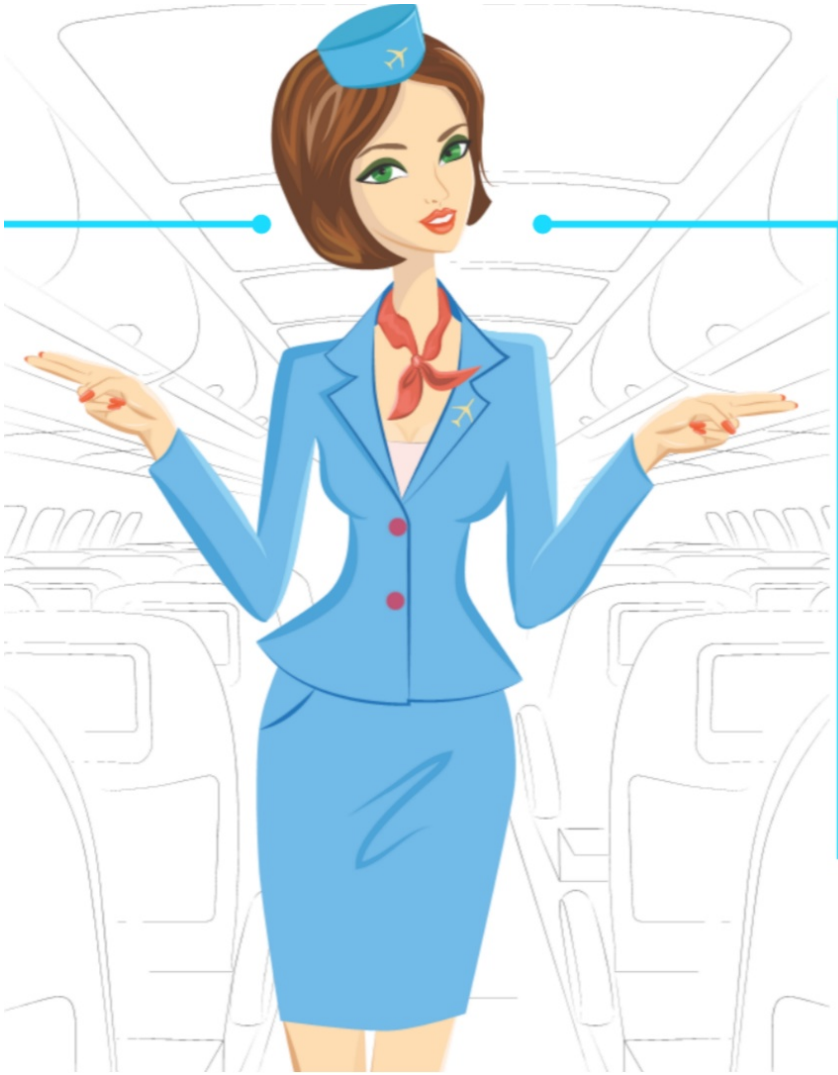
Reference

"We will be flying at an altitude of 35,000 feet."

Principle

"Always put on your oxygen mask before assisting other passengers."





Task

"To open the emergency exit, look out the window, pull the lever, and push out the exit door."

Process

"In the event of loss of cabin pressure, an oxygen mask will drop from the overhead compartment."

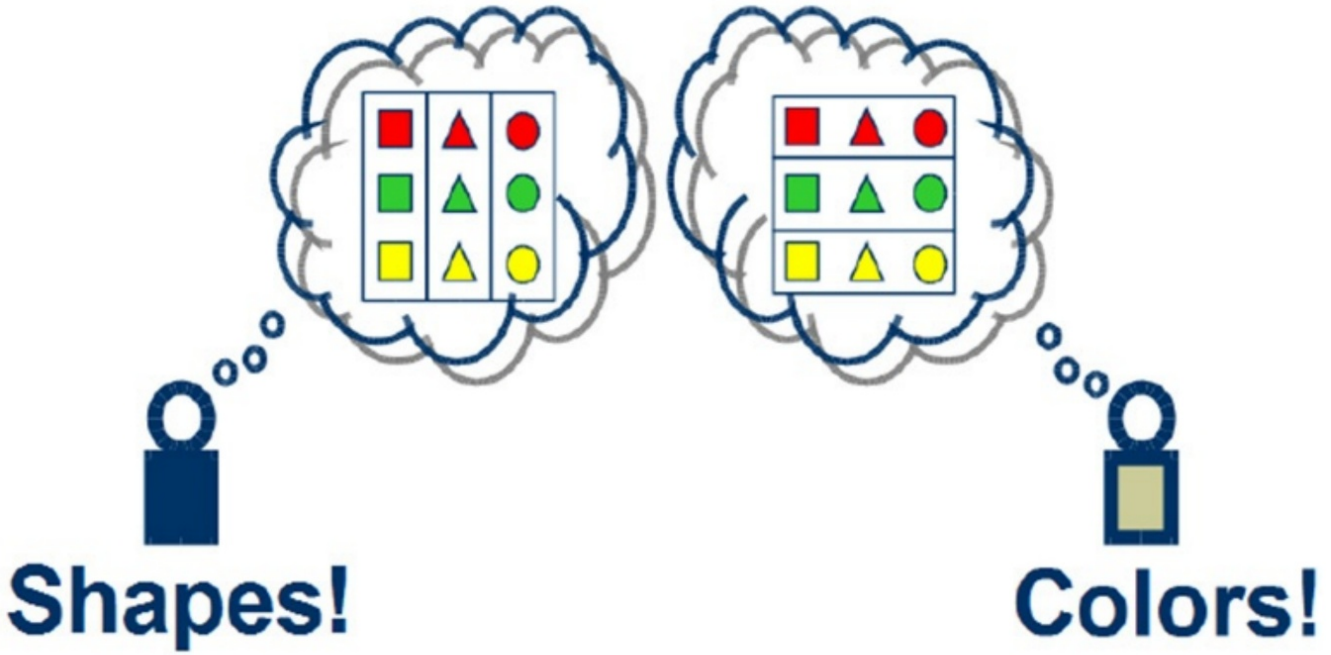
Concept

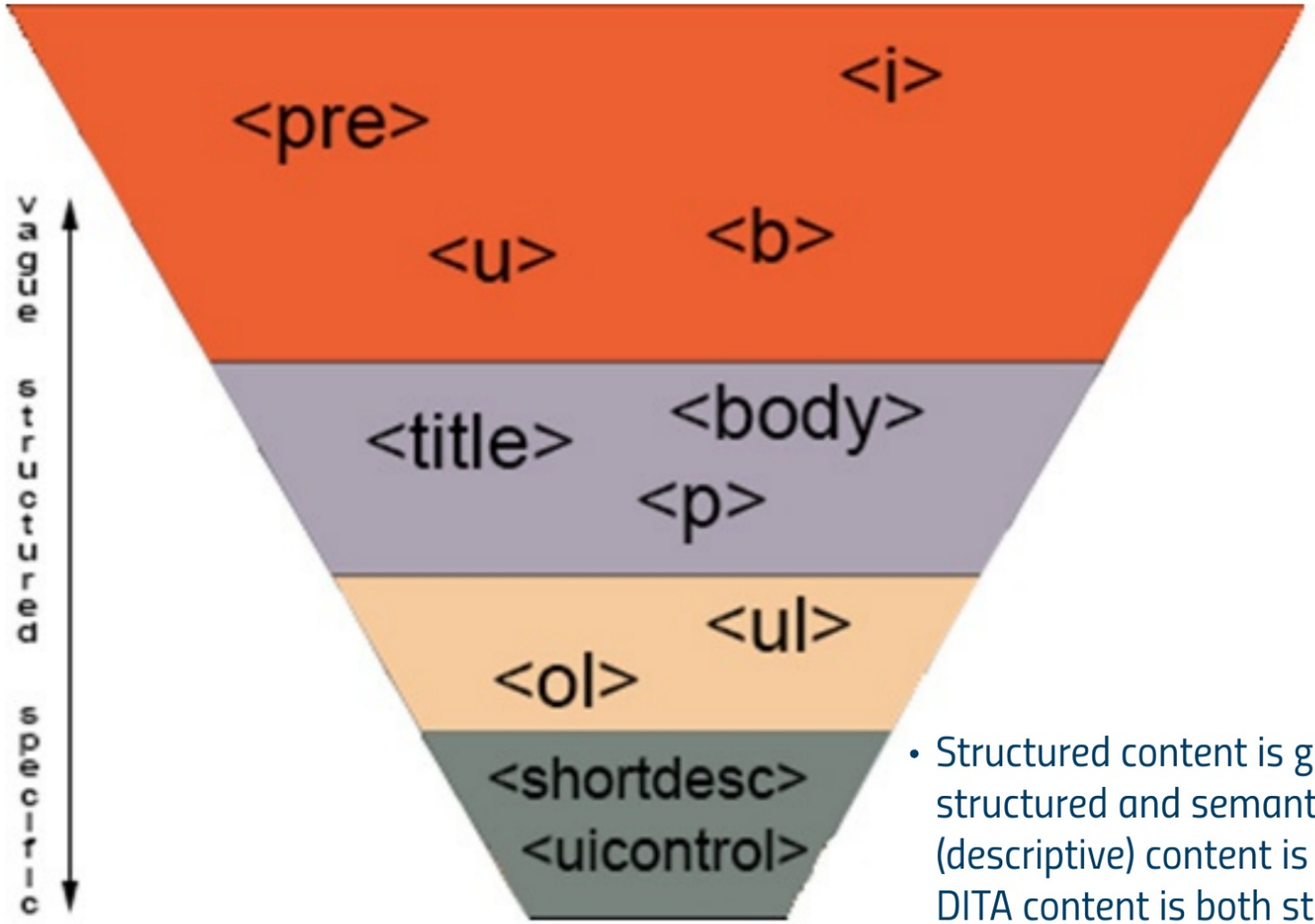
"On the left side of the plane you can see a typical example of a cumulonimbus cloud."

Its About Structuring Content

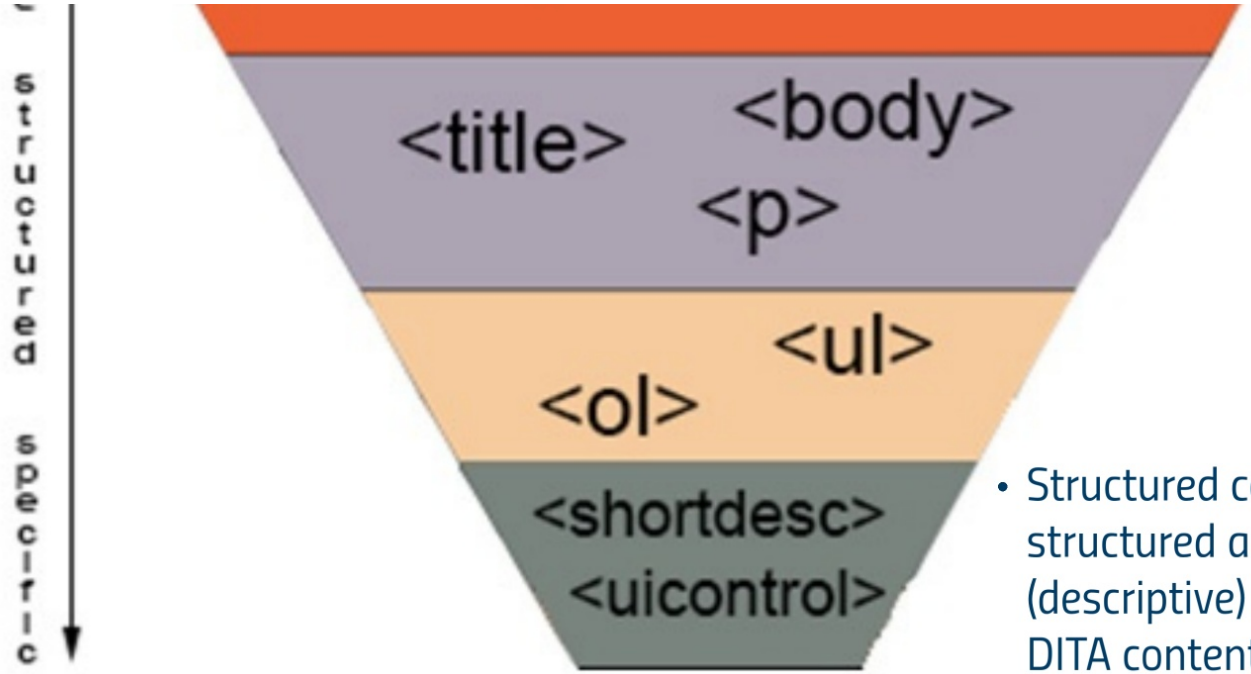
One area where IAs come into play is devising semantic structures for the content a firm produces using an IA's skillset to target the right audiences for content

Semantic tagging adds context to content





- Structured content is good, structured and semantic (descriptive) content is better; DITA content is both structured and semantically described



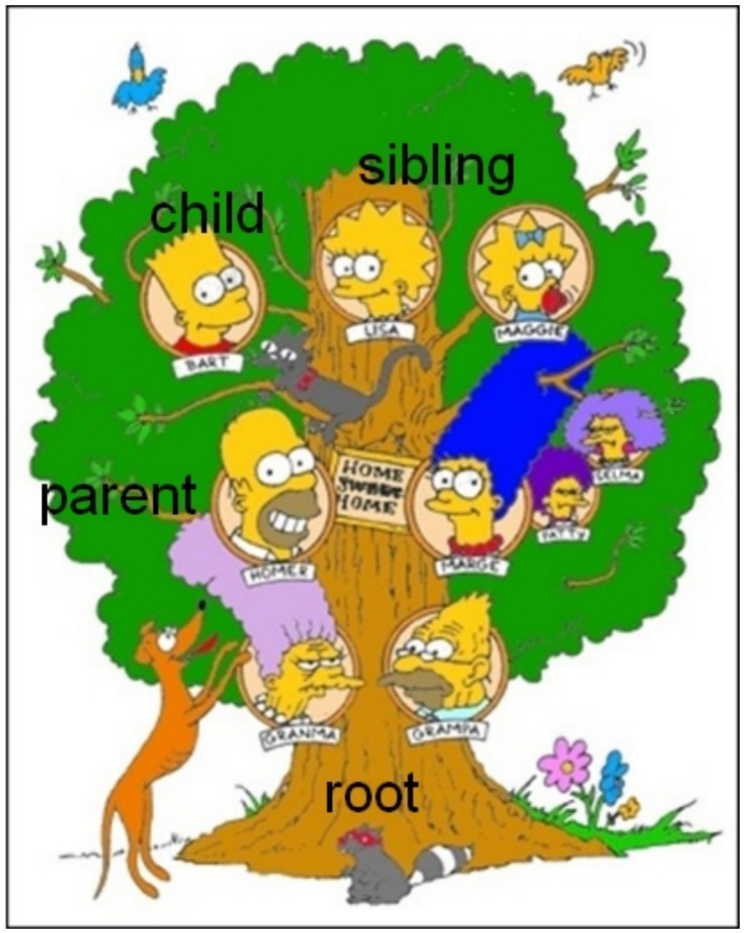
- Structured content is good, structured and semantic (descriptive) content is better; DITA content is both structured and semantically described

structural, descriptive

```
<menucascade>
    <uicontrol>Start</uicontrol>
    <uicontrol>Programs</uicontrol>
    <uicontrol>Accessories</uicontrol>
    <uicontrol>Notepad</uicontrol>
</menucascade>
```

specific, descriptive

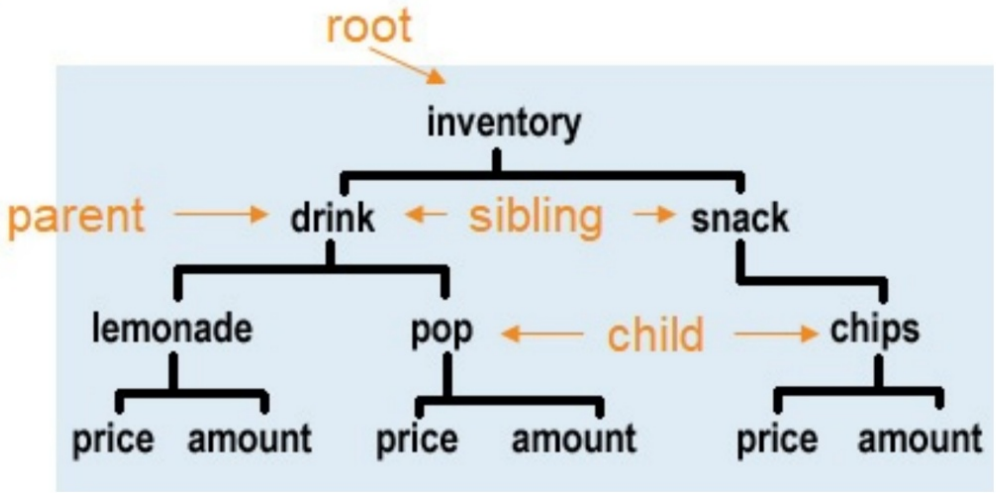
- XML works in a tree-like structure, similar to a family tree, which describing different levels and relationships
- Usually thought of as top-down rather than bottom-up (as here)
- Root
  - Parent
  - Sibling
    - Child



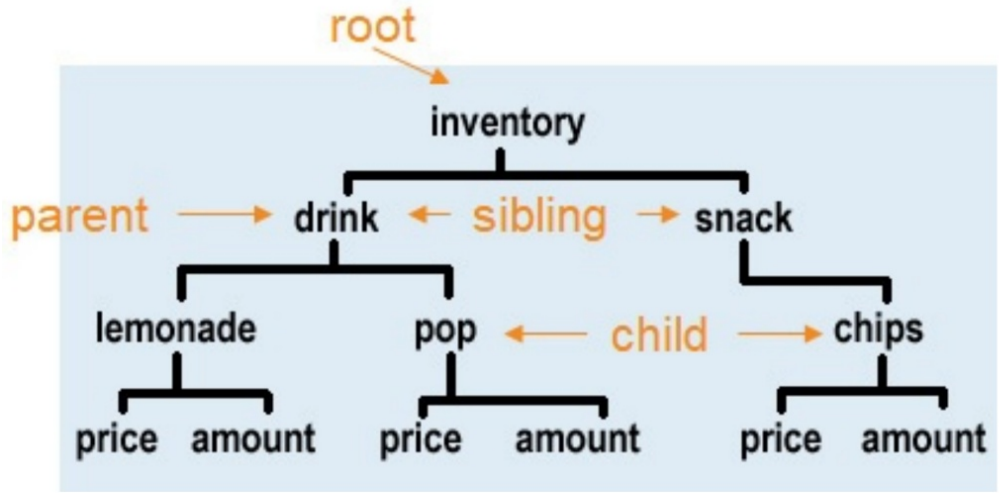
Sample XML Tree Structure (Not DITA)

```
<inventory>
        <drink>
                <lemonade>
                        <price>$2.50</price>
                        <amount>20</amount>
                </lemonade>
                <pop>
                        <price>$1.50</price>
                        <amount>10</amount>
                </pop>
        </drink>
        <snack>
                <chips>
                        <price>$4.50</price>
                        <amount>60</amount>
                </chips>
        </snack>
</inventory>
```

Example from:
www.tizag.com/xmlTutorial/xmltree.php



- "inventory" is root
- "drink" and "snack" are siblings
- "drink" is a parent of "pop" and "lemonade"
- "snack" is a parent of "chips"
- "lemonade" is also a parent to "price" and "amount"

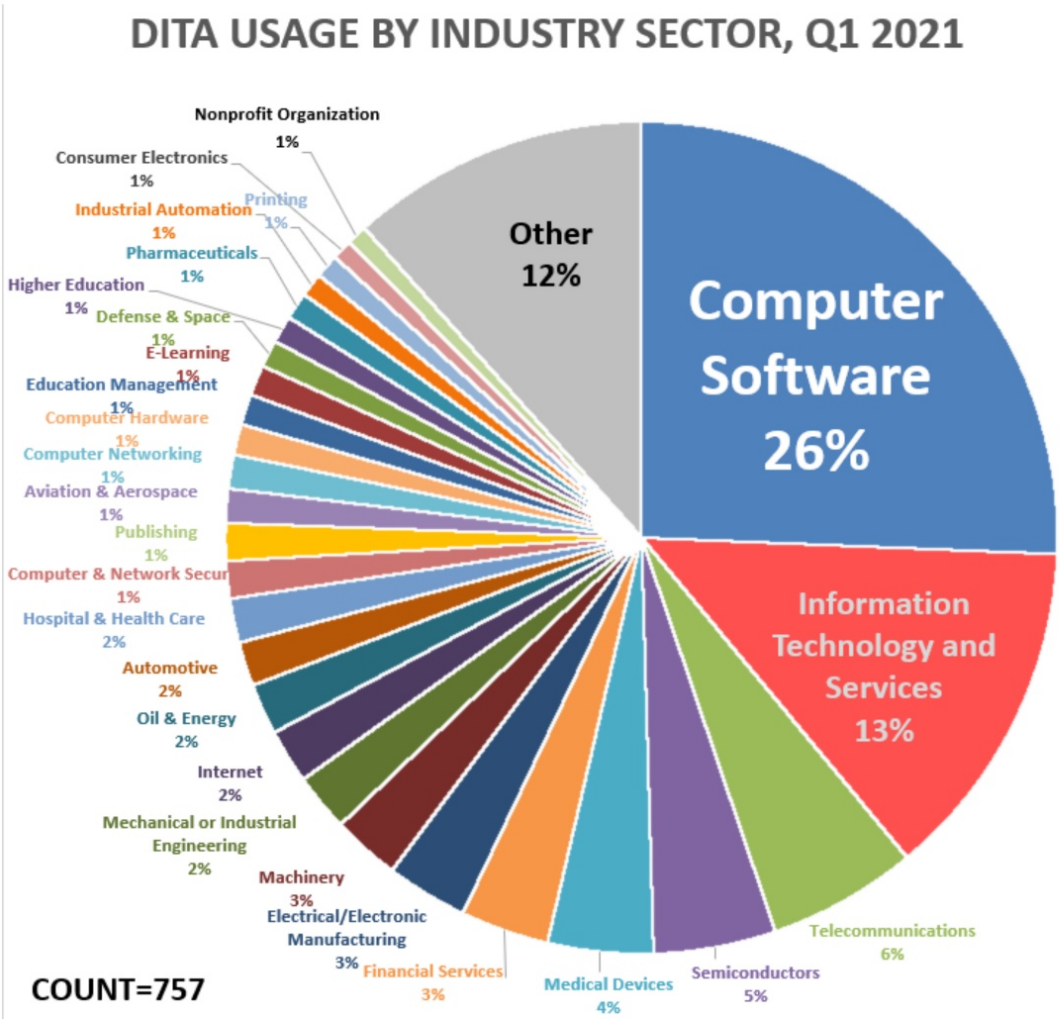


- "inventory" is root
- "drink" and "snack" are siblings
- "drink" is a parent of "pop" and "lemonade"
- "snack" is a parent of "chips"
- "lemonade" is also a parent to "price" and "amount"

- Certain child elements are only allowed by a certain parent; pop and lemonade belong to drink, but chips does not
  - This means that certain tags are available in the right context; a tag may be available in a DITA task topic but not in a DITA reference topic for example
- Structure is important: unlike HTML, XML must be valid
  - In other words, the correct tags must be used, and must be “closed” properly

So who uses DITA?

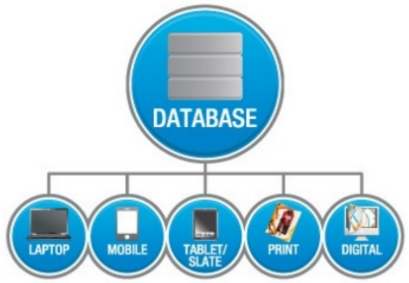
A selection of the 750+ firms using DITA worldwide


DITA USAGE BY INDUSTRY SECTOR, Q1 2021
Nonprofit Organization 1%
Consumer Electronics 1%
Industrial Automation 1%
Printing 1%
Pharmaceuticals 1%
Higher Education 1%
Defense & Space 1%
E-Learning 1%
Education Management 1%
Computer Hardware 1%
Computer Networking 1%
Aviation & Aerospace 1%
Publishing 1%
Computer & Network Secur 1%
Hospital & Health Care 2%
Automotive 2%
Oil & Energy 2%
Internet 2%
Mechanical or Industrial Engineering 2%
Machinery 3%
Electrical/Electronic Manufacturing 3%
Financial Services 3%
Medical Devices 4%
Semiconductors 5%
Telecommunications 6%
Information Technology and Services 13%
Computer Software 26%
Other 12%
COUNT=757

Some Reasons for Adopting DITA

The following are the key business reasons why companies want to work with forms of structured content like DITA:

- Content Reuse
- Lower localization costs
- Content/Formatting separation
- Multi-channel publishing



These are all very good reasons to move to DITA

An additional benefit…

- Content reuse also leads to the benefit of consistent content and messaging

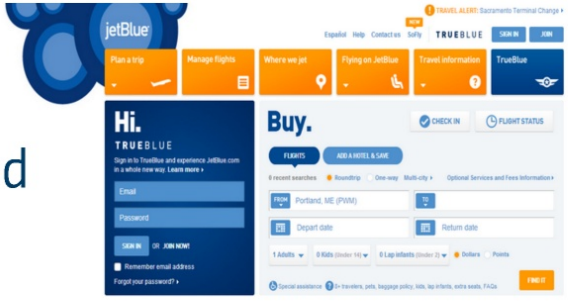
Consistency in the Cross-Channel Experience

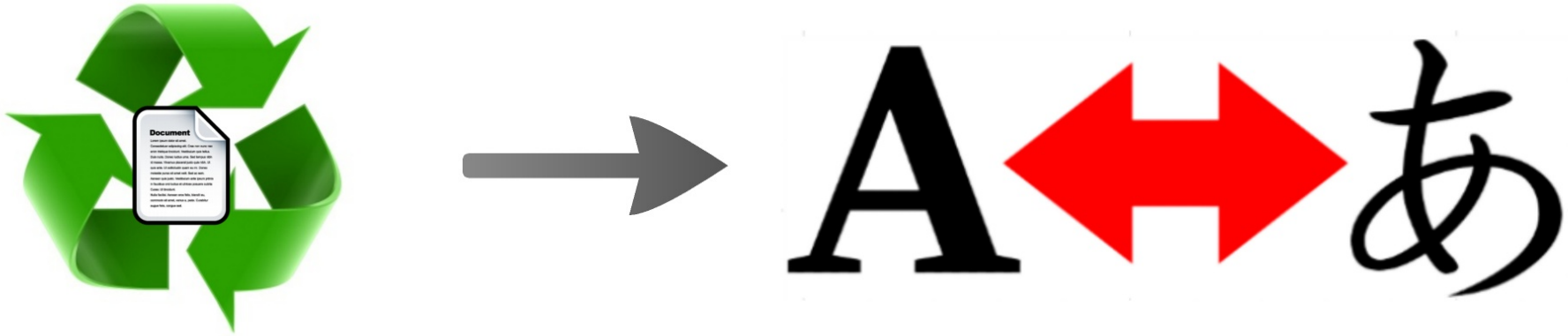
by **NIELSEN NORMAN GROUP** on October 27, 2013

Topics: Content Strategy E-commerce

Summary: A consistent user experience, regardless of platform, is one of the 4 key elements of a usable omnichannel experience. Consistency across channels helps build trust with customers.

Consistent content is the first plank of a consistent experience for users (along with being seamless, available and context-specific)





DITA Can Help Reduce Localization Costs

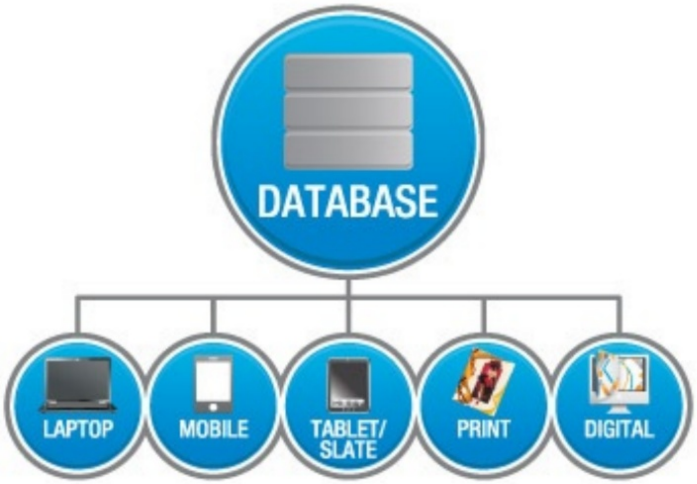
- Content reuse in English = localization savings
- If there are many languages to target, ROI argument for move to DITA (+ CMS) is easier

Localization as an ROI Factor for Moving to DITA

- One pharma client's ROI for DITA + CMS purchase was based on localization savings; translated Operations Manuals, Training materials and Interface Manuals into 14 languages
- Localization process with their Localization Service Provider was much shorter and cheaper than with the Desktop Publishing software they previously used
- Content consistency is greatly improved; PDF and HTML use same formatting template as English, so formatting consistency is ensured

Less Formatting = More Nimble Content Creation

- In a study I did prior to moving to DITA at AMD, found that fully half of the time spent using Desktop Publishing software was on formatting content
- Suggests considerable time savings, making content creators more nimble and able to concentrate on making additional content



DITA Was Built with Multi-channel Publishing in Mind

- This was a key design feature when DITA was still being devised at IBM; intent was to tame a multitude of output processes then in place with a single
- DITA-OT by default supports 11 output types, including PDF, RTF, HTML5/XHTML
- Other output types possible
- Strong vendor support (AntennaHouse and RenderX for PDF, Mekon DITAWeb, Zoomin Docs and FluidTopics for the Web)

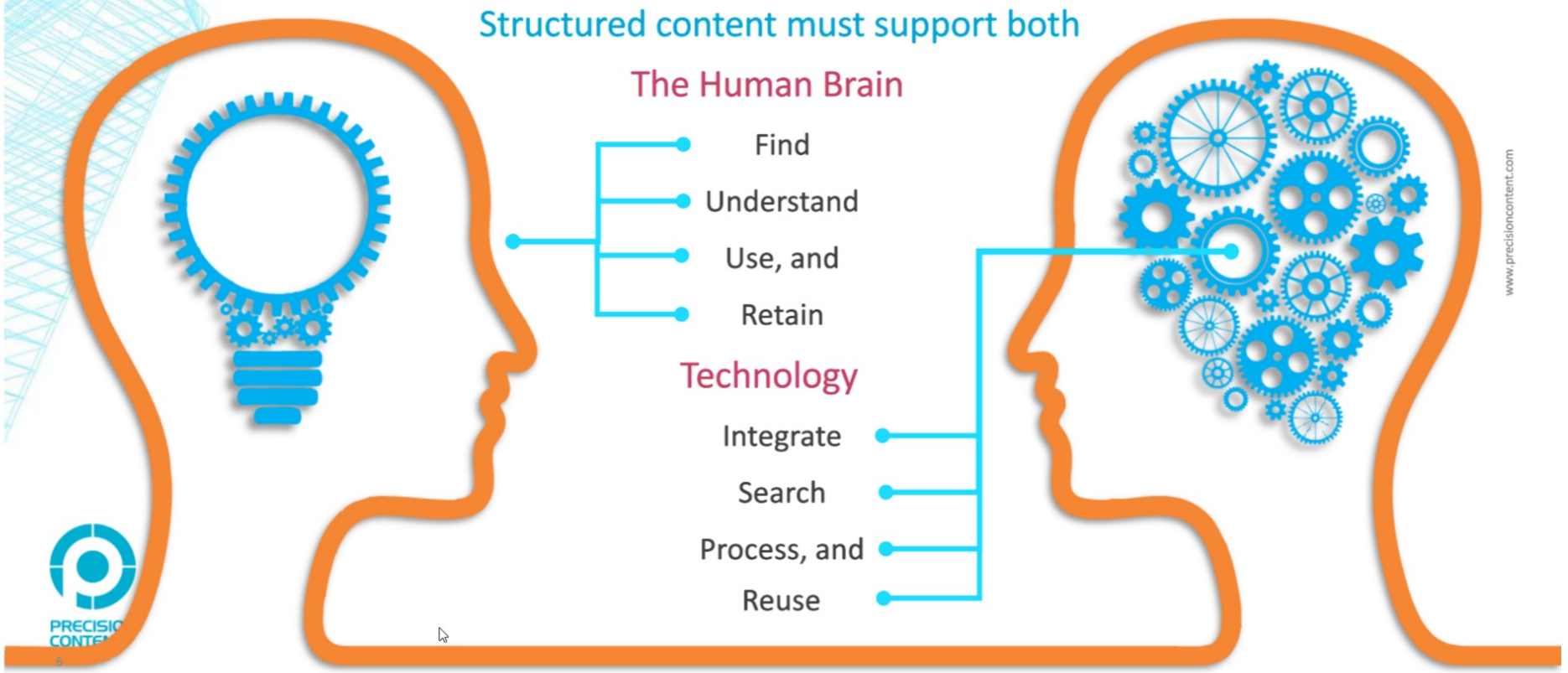
These are all very good reasons to move to DITA

So What Can an IA Do with Non-Web Content?

A lot of the same techniques that IAs use can be applied to non-Web content:

- Content audit
- Persona/Scenario development
- Accessibility
- User centric content
- Structure content (hierarchies)
- Create taxonomies describing content for finadability

IAs Need to Meet Two Sets of Fundamental Needs



An Excerpt from a Medical Journal

- pN3 description only closely mirrors descriptions for pN3a +pN3b + pN3c
- Use of footnotes confusing
- "Clinically detected" and "Not clinically detected" are not exact opposites
- Inconsistent enumeration of lymph nodes

pN3 Metastases in ten or more axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected**** ipsilateral internal mammary lymph nodes in the presence of one or more positive level I, II axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected***; or in ipsilateral supraclavicular lymph nodes

pN3a Metastases in ten or more axillary lymph nodes (at least one tumor deposit greater than 2.0 mm), or metastases to the infraclavicular (level III axillary lymph) nodes

pN3b Metastases in clinically detected**** ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected***

pN3c Metastases in ipsilateral supraclavicular lymph nodes

Notes:

*** "Not clinically detected" is defined as not detected by imaging studies (excluding lymphoscintigraphy) or not detected by clinical examination.

**** "Clinically detected" is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination and having characteristics highly suspicious for malignancy or a presumed pathologic macrometastasis based on fine needle aspiration biopsy with cytologic examination.

PRECISION CONTENT™

Same Content After Restructuring

- 44.2% reduction in word count
- 20% reduction in passive voice
- 18.4% increase in Flesch Reading Ease score
- 30% increase in white space
- Elimination of footnotes
- Addition of labels and visual elements

PRECISION CONTENT™

Metastases

Pathological node level 3 definitions

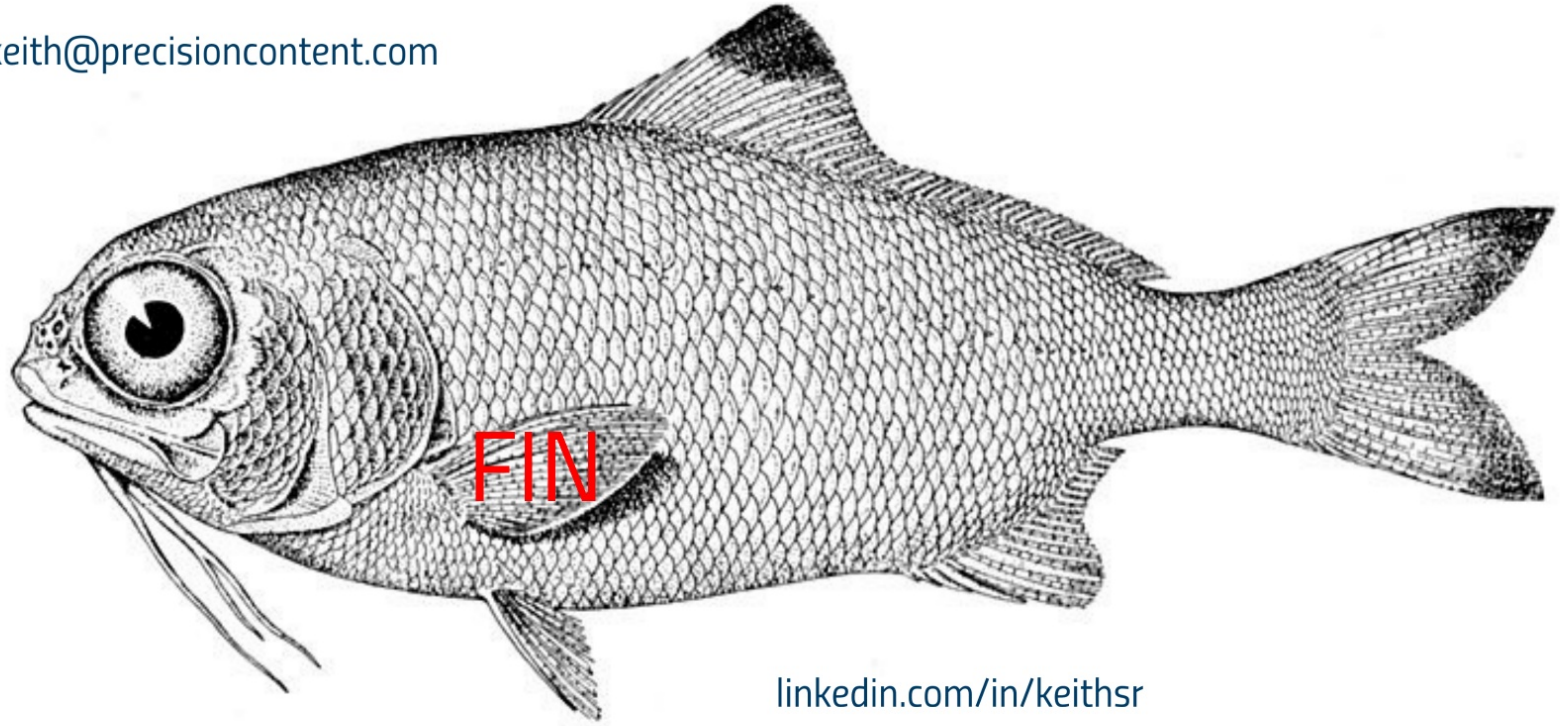
The following table is used by clinicians to classify metastases found in regional lymph nodes.

Node level...	Which includes...	Describes metastases found in...
pN3	pN3a	10 or more axillary lymph nodes where at least one deposit is greater than 2.0 mm.
		any number of infraclavicular (level III axillary) lymph nodes.

	pN3b	• any number of ipsilateral internal mammary lymph nodes detected by
		  ▪ clinical exam
		  ▪ fine needle aspiration biopsy, or
		  ▪ imaging study, and
		• any number of level I or II axillary lymph nodes.

		• any number of ipsilateral internal mammary lymph nodes where micrometastases or macrometastases are detected by sentinel lymph node biopsy, and • 4 or more level I or II axillary lymph nodes.
	pN3c	Any number of ipsilateral supraclavicular lymph nodes.

keith@precisioncontent.com



linkedin.com/in/keithsr