Structured Content is Dead. Long Live Structured Content!

KEITH SCHENGILI-ROBERTS

OCTOBER 8, 2020

precision content

Agenda

Intro

Skeptics Make Poor Futurists

A Short History of Technical Writing

Where We Are Today

Some Likely Near Term Trends

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

About DITAWriter.com

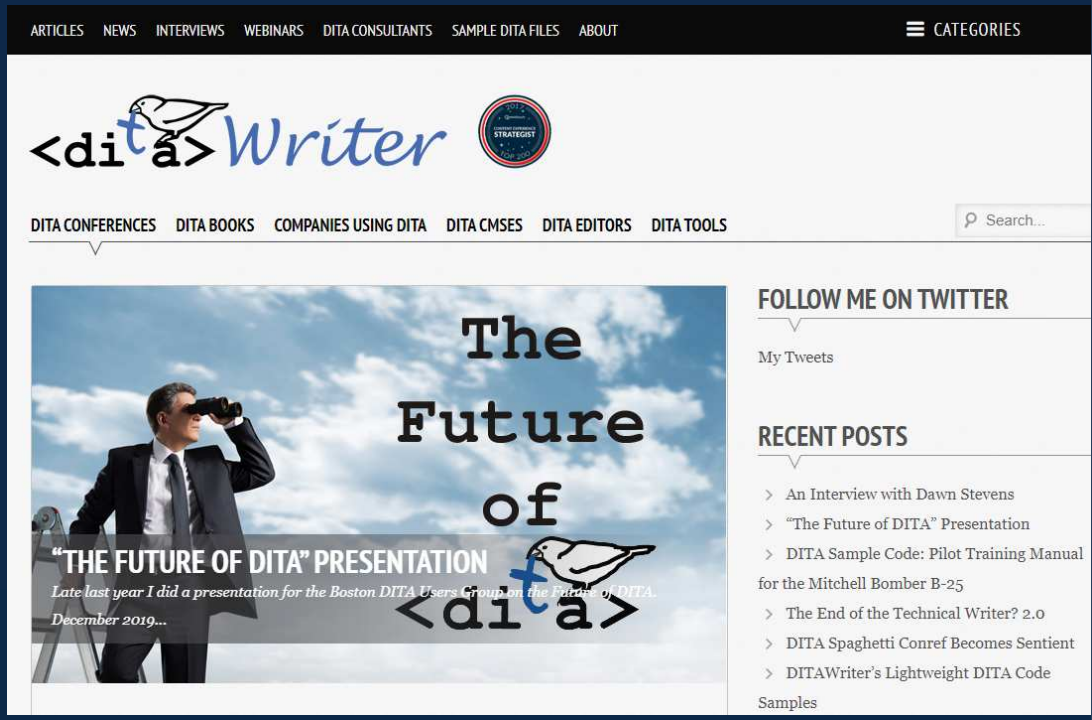
A nexus for DITA-related information,

Features interviews, helpful information for those just starting to work with the standard, my own thoughts about the current state of DITA, and much more

Over 376,000 hits

Listing of companies that are using DITA, designed to aid technical writers looking for DITA jobs close to them

Independent, no ads




EContent
2018 2019
100 COMPANIES
THAT MATTER MOST IN THE DIGITAL CONTENT INDUSTRY
2017 mindtouch CONTENT EXPERIENCE STRATEGIST TOP 200
mindtouch CONTENT EXPERIENCE INFLUENCER 20 TOP 25 17
precision content
PRECISION CONTENT.com
Make your information easier to use
"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
Your partner in ...
Publishing Technologies
DITAinPrint
GET THE WORD OUT
DITAonPoint
ENTERPRISE PUBLISHING
WittyDITA
SHARE YOUR CONTENT

Skeptics Make Poor Futurists

I was asked by Adobe to speak about the future of technical communications

Problem is, I am a skeptic by nature. I am probably not the person you want to choose to predict the future.

What I plan to do is survey what we currently know, and highlight some industry trends

And if 2020 has demonstrated nothing else, it's how unpredictable things can be

Explaining the Title for this Talk

I don’t think structured content is “dead”, but I do think that we are going to see some significant changes in how it is put to use

The same technologies being used on structured content are also being applied to unstructured content

And unstructured content is not necessarily as unstructured as you may think

DITA is ultimately a niche market, but what a niche!

Futurists Rarely Talk About the Past…

But I think it is important to know more about where we have come from in order to have a better idea as to where I think we may be going, and why

A Short History of Technical Writing

precision content

@PCASinc

The Early Days

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

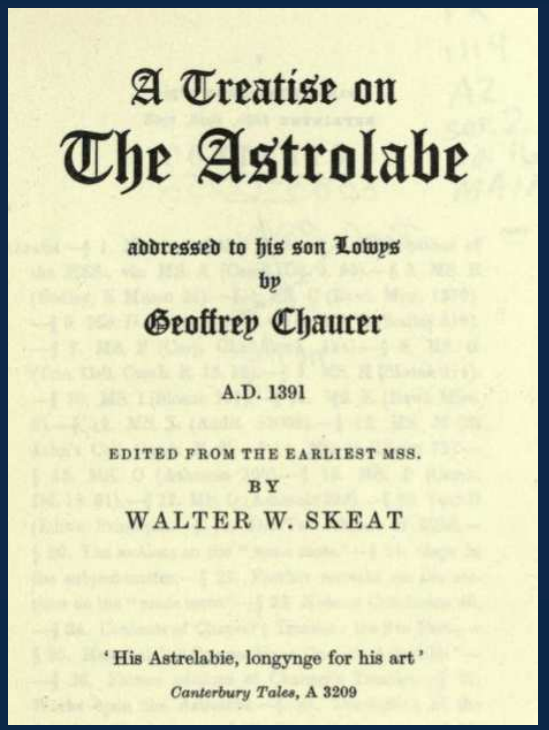
- Ancient Egyptian medical and mathematical texts

- 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"



A Treatise on
The Astrolabe

addressed to his son Lowys
by
Geoffrey Chaucer

A.D. 1391

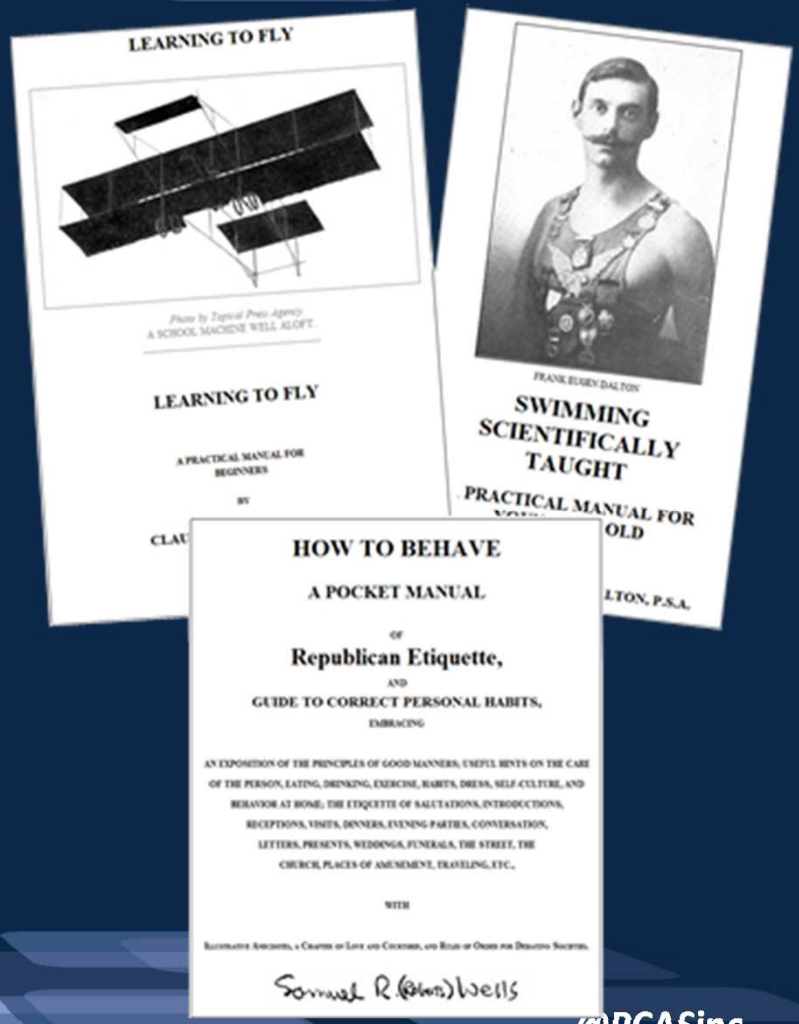
EDITED FROM THE EARLIEST MSS.
BY
WALTER W. SKEAT

'His Astrelabie, longynge for his art'
Canterbury Tales, A 3209

Up to Early 20th Century

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.



Beginnings of Modern Technical Writing

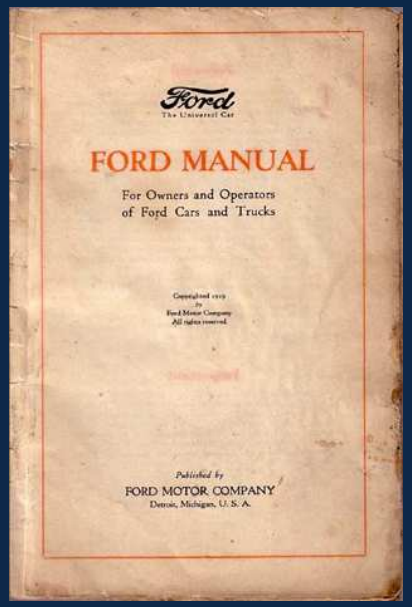
Technical writing as we know it started in early 20th century; focused on instructing people how to work with technology

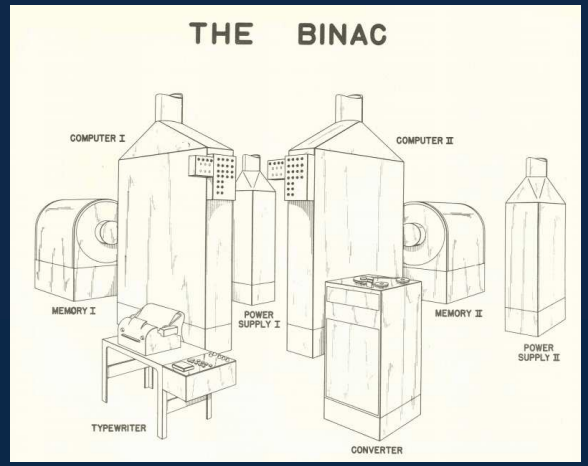
For Model T manual (1919) is a good example of this

Joseph Chapline wrote first computer manual for the BINAC (1949)

Was often done by a single person, often a Subject Matter Expert (SME)

They tended to be idiosyncratic and often were one-offs





Beginnings of Structured Authoring

Much of structured content 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.

Information Mapping®

Designed to provide quick and easy access to information at the right level of detail to readers

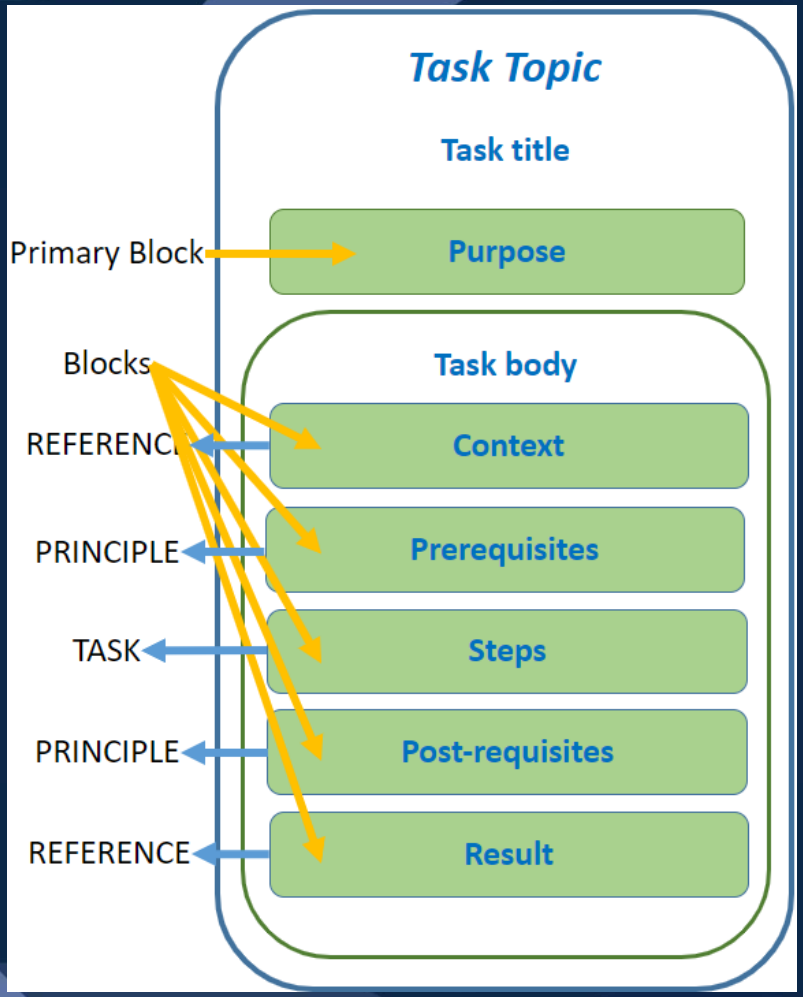
Introduced the idea of writing modular structure/content

- A block is limited to a single topic

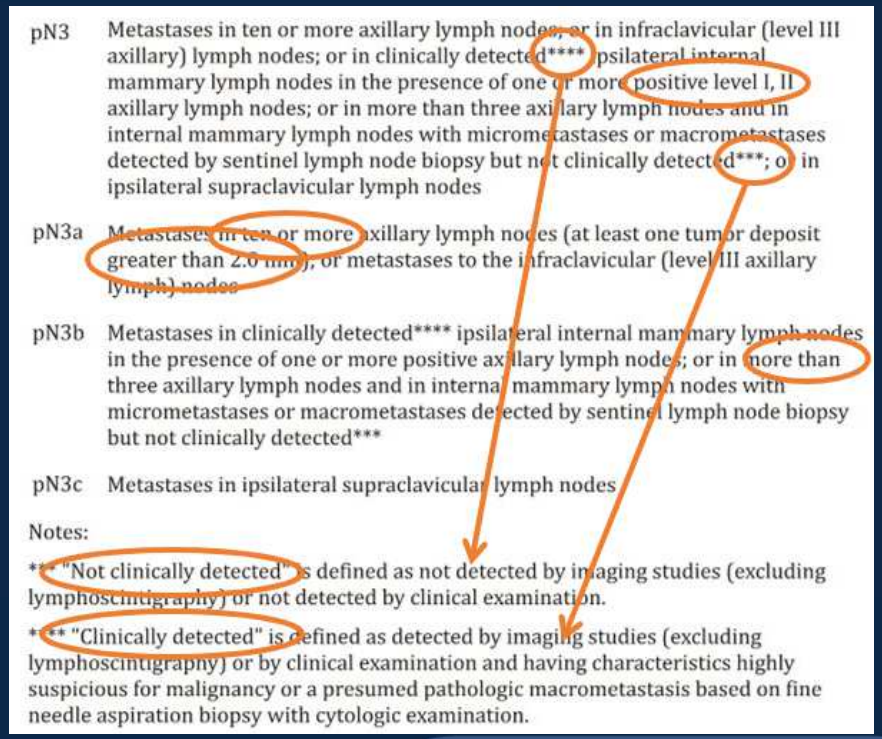
- Blocks are grouped into maps

Sound familiar? ☺

Precision Content is a modern adaptation of methods taught by Information Mapping International (though it is not affiliated or endorsed by it)



Precision Content Applied to Unstructured Content



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.

Strives for clarity, consistency, conciseness, and focuses on easily digestible blocks of content

From this

To this

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.

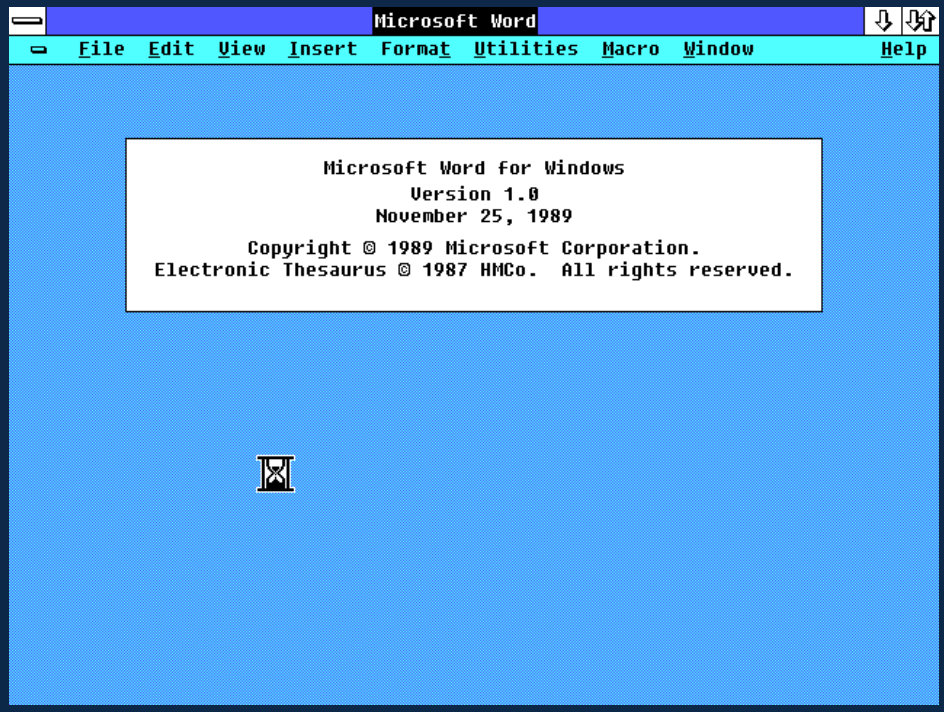
		• any number of ipsilateral internal mammary lymph nodes where micrometastases or macrometastases are detected by sentinel lymph node biopsy, and • 4 or more level I or II axillary lymph nodes.
	pN3c	Any number of ipsilateral supraclavicular lymph nodes.

The Rise of Desktop Publishing (and Unstructured Content) in 80s

Non-structured, but allowed for the creation of content quickly and easily

Over time, it was also possible to output content to various formats, including PDF and Web

This is the beginning of the unstructured vs. structured content dichotomy: ease-of-use + low overhead vs. significant value add but higher upfront processing



90s-era IBM Faced Two Documentation Issues

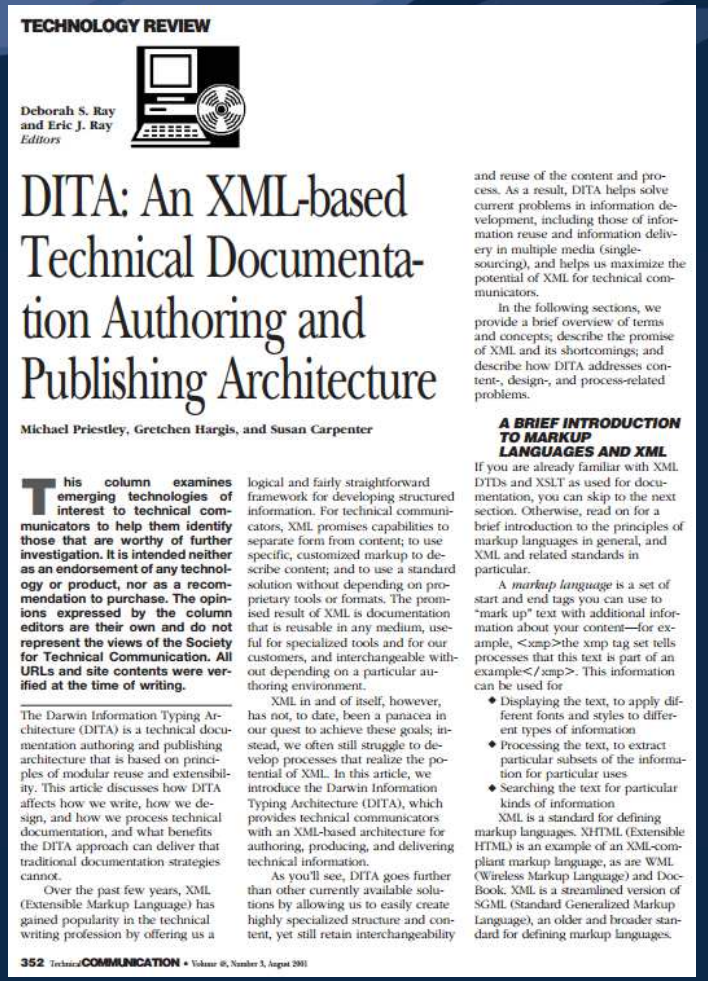
Back in the 90s, IBM documentation teams were facing two main issues:

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

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

DITA Open Toolkit opened up multi-channel publishing



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.

352 TechnicalCOMMUNICATION • Volume 48, Number 3, August 2001

DITA Has Evolved

Current version is 1.3, moving to DITA 2.0 likely next year

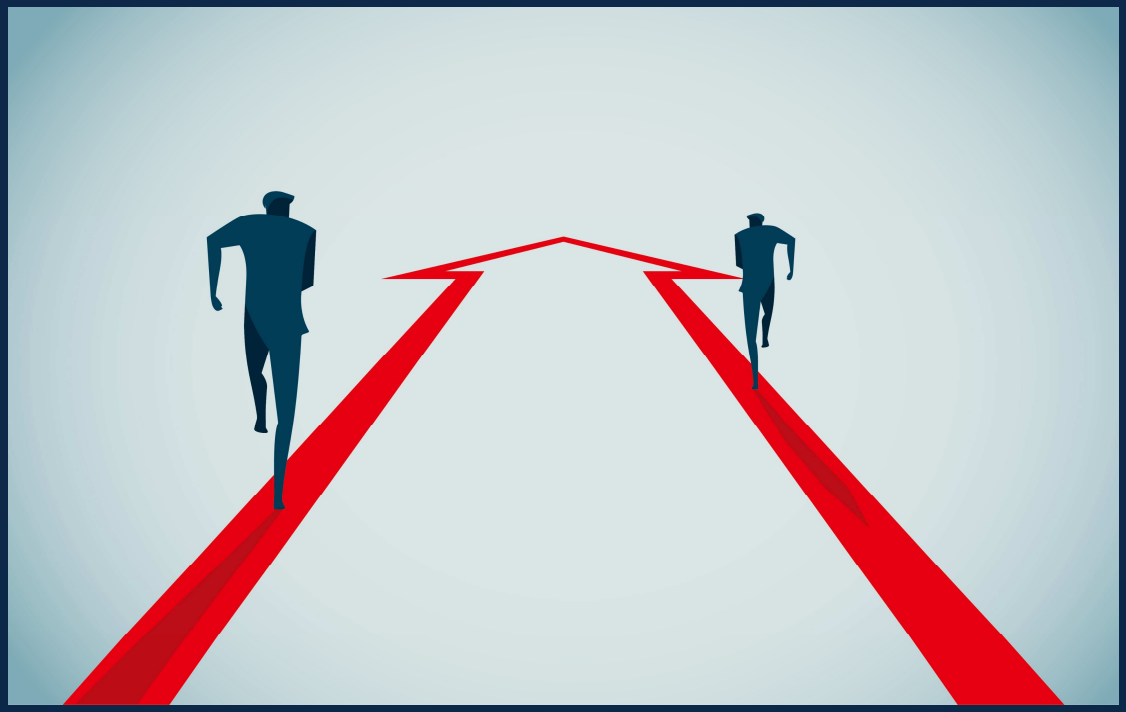
- Started out with three topic types (concept, task, reference)
- Precision Content adds two more: process and principle

There is also promise in Lightweight DITA

Information Type Examples

If the goal of the information is to …	Then use the information type …
• List the nutritional facts for cherry cola	Reference
• Explain what a soft drink is	Concept
• Instruct you on how to safely open your can of cola	Task
• Illustrate how cola is bottled	Process
• Warn you not to drop a Mentos in your cola bottle	Principle

Where We Are Today



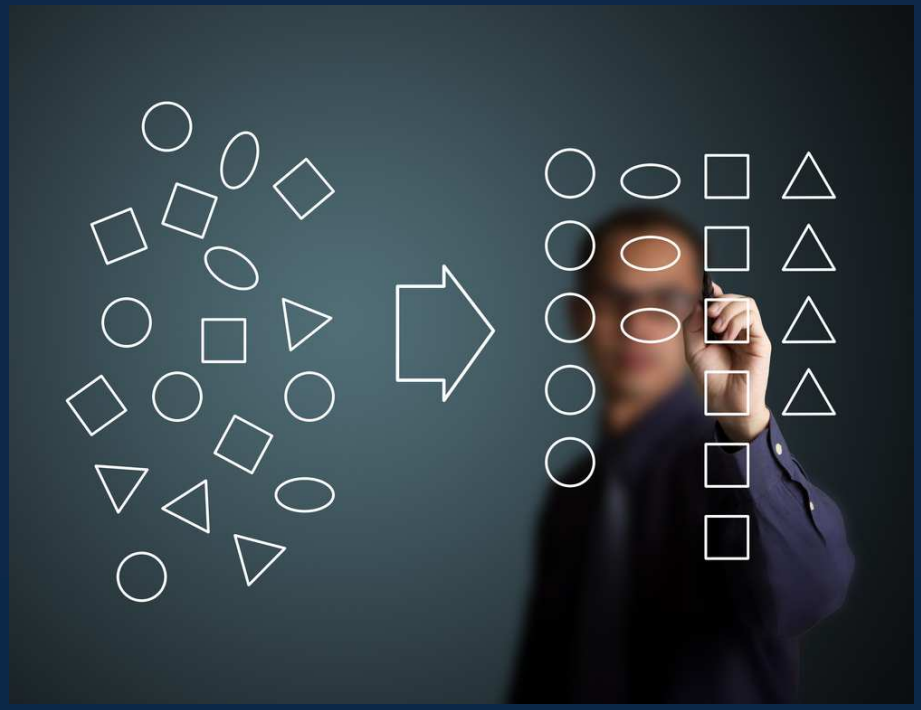
Most Content is Unstructured

Unstructured content is thought to make up roughly 80% or more of all enterprise data

And what is the world's most popular technical writing tool? Word.

So structuring content is the solution to making unstructured content more usable, right?

- The path is not so clear cut



Unstructured Content Has More "Structure" than You Might Think

MS Office applications have been using their own form of XML since at least 2003

- while it does not separate content from formatting per se, as you can see from this early example it does contain a semblance of structure

Similarly, image and video files come with a substantial amount of metadata content

- this provides useful information on when, where and who created it

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<?mso-application progid="Word.Document"?>
<w:wordDocument
   xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml"
   xmlns:wx="http://schemas.microsoft.com/office/word/2003/auxHint"
   xmlns:o="urn:schemas-microsoft-com:office:office"
   w:macrosPresent="no"
   w:embeddedObjPresent="no"
   w:ocxPresent="no"
   xml:space="preserve">
  <o:DocumentProperties>
    <o:Title>This is the title</o:Title>
    <o:Author>Darl McBride</o:Author>
    <o:LastAuthor>Bill Gates</o:LastAuthor>
    <o:Revision>1</o:Revision>
    <o:TotalTime>0</o:TotalTime>
    <o:Created>2007-03-15T23:05:00Z</o:Created>
    <o:LastSaved>2007-03-15T23:05:00Z</o:LastSaved>
    <o:Pages>1</o:Pages>
    <o:Words>6</o:Words>
    <o:Characters>40</o:Characters>
    <o:Company>SCO Group, Inc.</o:Company>
    <o:Lines>1</o:Lines>
    <o:Paragraphs>1</o:Paragraphs>
    <o:CharactersWithSpaces>45</o:CharactersWithSpaces>
    <o:Version>11.6359</o:Version>
  </o:DocumentProperties>
  <w:fonts>
    <w:defaultFonts
       w:ascii="Times New Roman"
       w:fareast="Times New Roman"
       w:h-ansi="Times New Roman"
       w:cs="Times New Roman" />
  </w:fonts>
```

Who is Using DITA?

This is using the latest DITA on the number of firms known to be using DITA

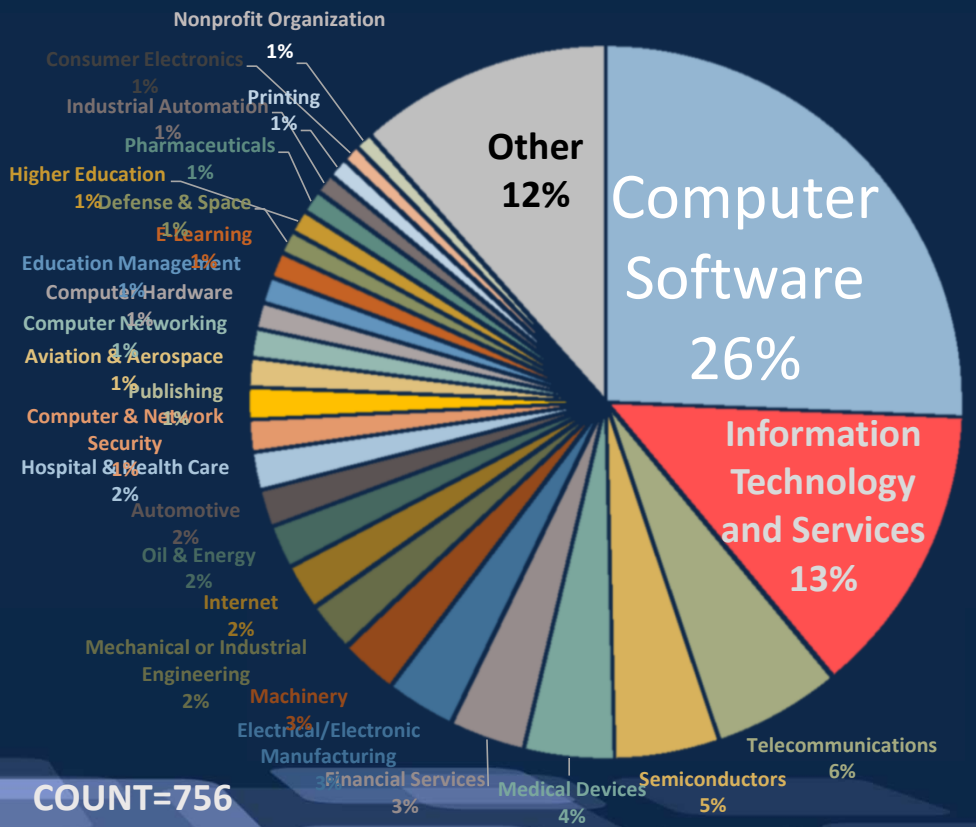
DITA started in the software sector, and it is still the single largest group of DITA users

- Worth noting that many of the other industries mentioned here are also using DITA when documenting their software

Despite that, DITA use is widespread, and still seeing overall growth

- "Other" in particular keeps growing, emphasizing widespread growth

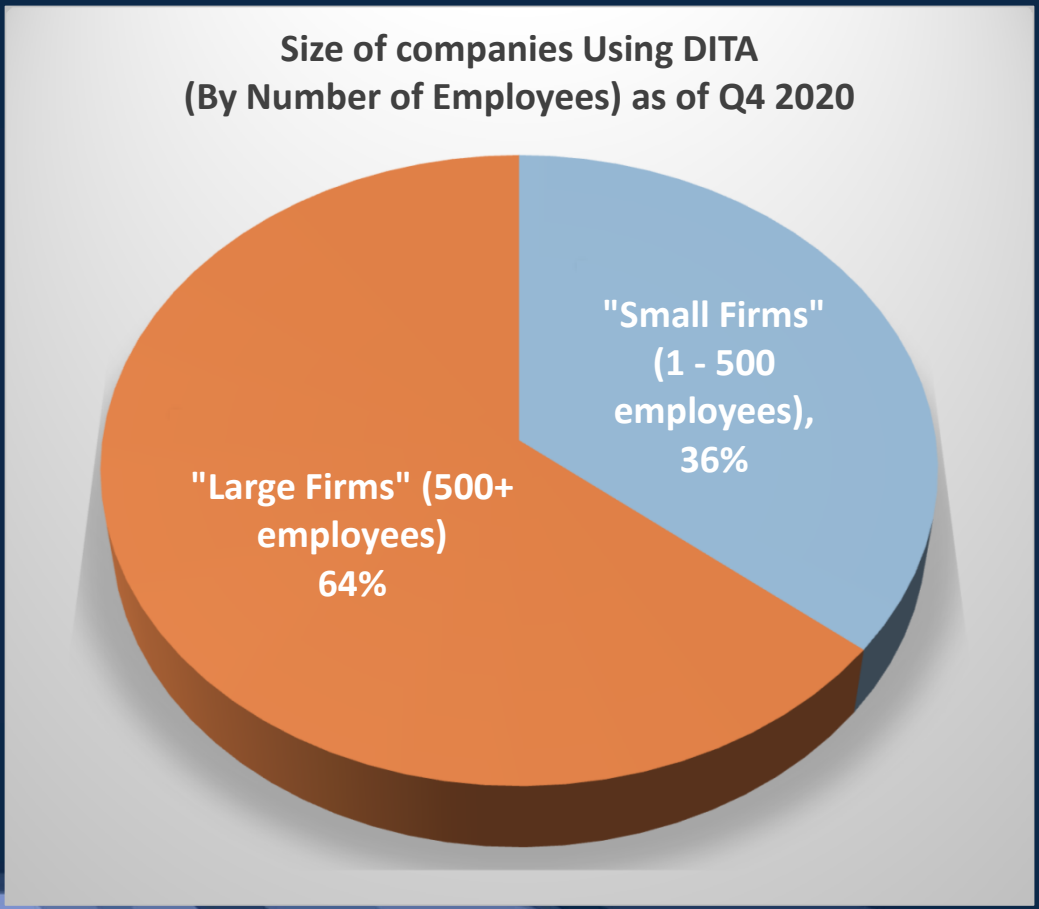
DITA USAGE BY INDUSTRY SECTOR, Q4 2020



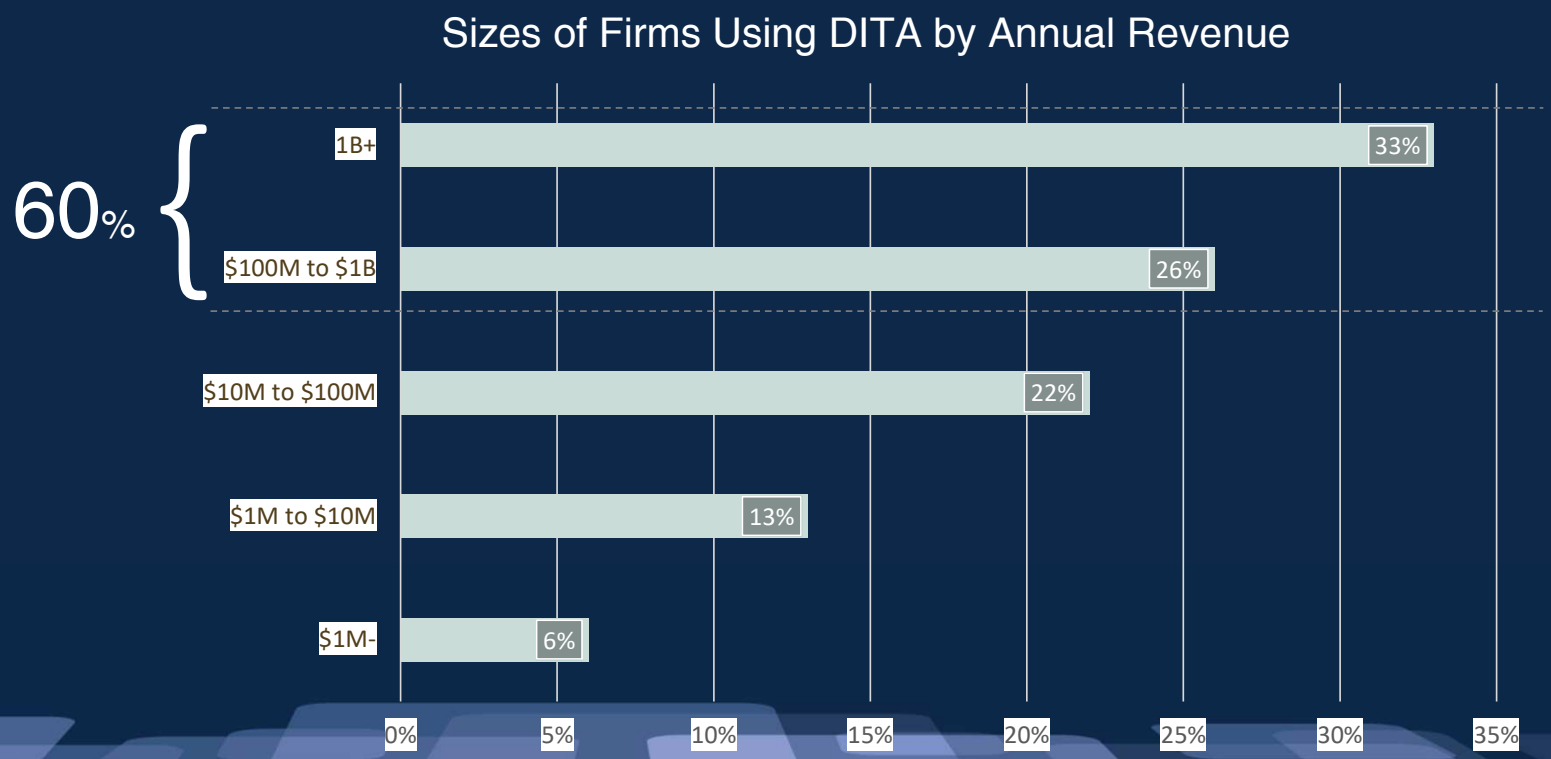
So Which Firms Get the Most Benefit from DITA?

It is large firms that can leverage the largest benefits from going down the path of DITA

The types of efficiencies that DITA brings work best at scale (more on that later)



How Large is Large?



Who is *Not* Using DITA?

Structured content requires an investment of time and money

- Benefits are worth it, but may not be apparent/obvious at first

Smaller firms

Many firms also adopting a hybrid approach

- Rare to find a large firm that is using DITA exclusively

DITA's Primary "Superpower": Content Reuse

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

It has enabled the business success of DITA

Chief Business Strengths of DITA

These are the reasons that most people with even a passing familiarity of DITA will be familiar with:

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

The Less Obvious Benefits of DITA

In addition to all the things DITA was designed for, when done right, it can also do the following:

- Advance product SEO
- Provide a better ~~user~~ customer experience
- Topic/semantic structure works well with emerging system, like chatbots

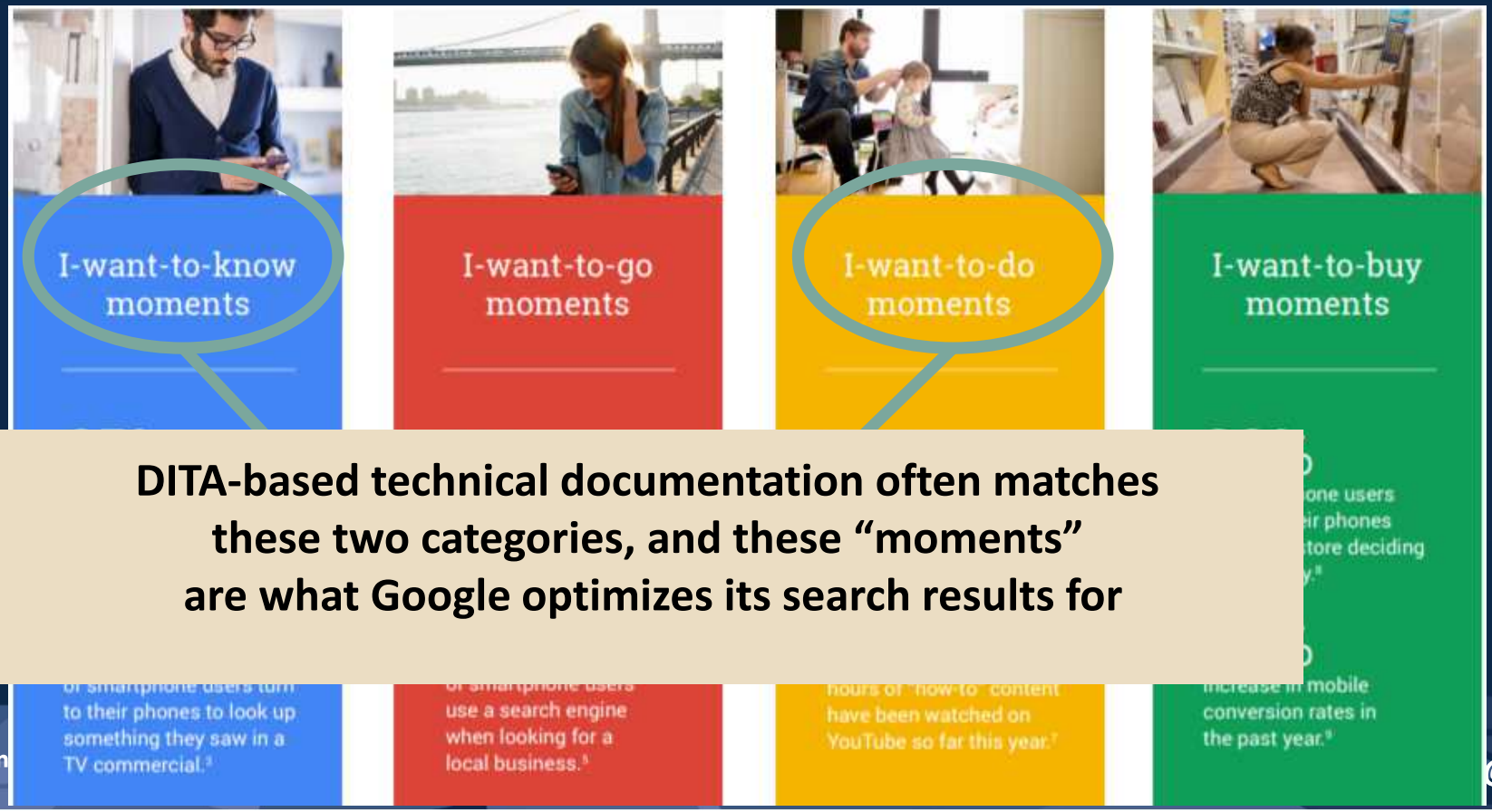
Customer Experience Has Become a Business Imperative

- Before the digital transformation, key business interactions with customers happened in the physical world

- But then our customers moved online. And we moved what had formerly been physical content online as well.

- Technical documentation has come to the fore, since it often contains targeted info customers are looking for

DITA Topics Target Two of Google's Four User "Moments"



DITA-based technical documentation often matches these two categories, and these "moments" are what Google optimizes its search results for

Enter: Chatbots and Artificial Intelligence

- This subject has been getting a lot of interest in marketing and technical documentation
- We are still in the early days, but it is likely to become another facet in the overall customer experience journey

How Does DITA-based Technical Documentation Fit In?

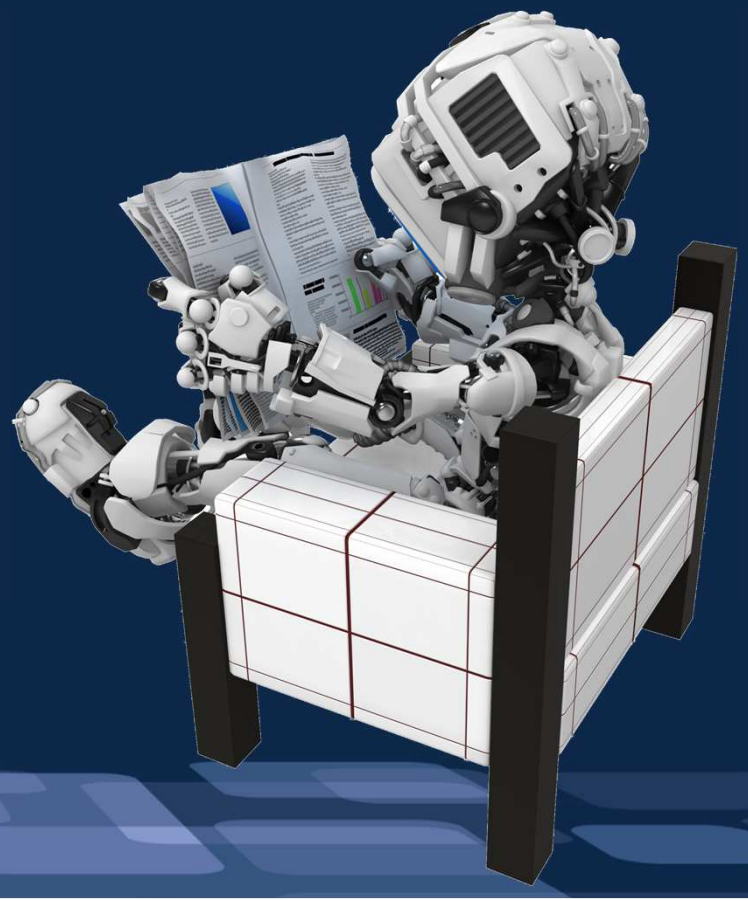
Chatbots take one of two paths when working with external content (including technical content):

1. Digests existing material which it incorporates directly within itself, or
2. It refers to appropriate material a user wants via metadata

Either way, having tech docs in a structured format (like DITA) makes it easier for chatbots to digest or find content to reference because well-defined content types + descriptive metadata make it easier for chatbots to use

Integrating DITA Content with Chatbots

- Think of chatbots as a new, complementary channel for content delivery that works best with structured content
- DITA definitely has a role here; not directly in the chatbot "conversation", but in providing technical content when it is needed

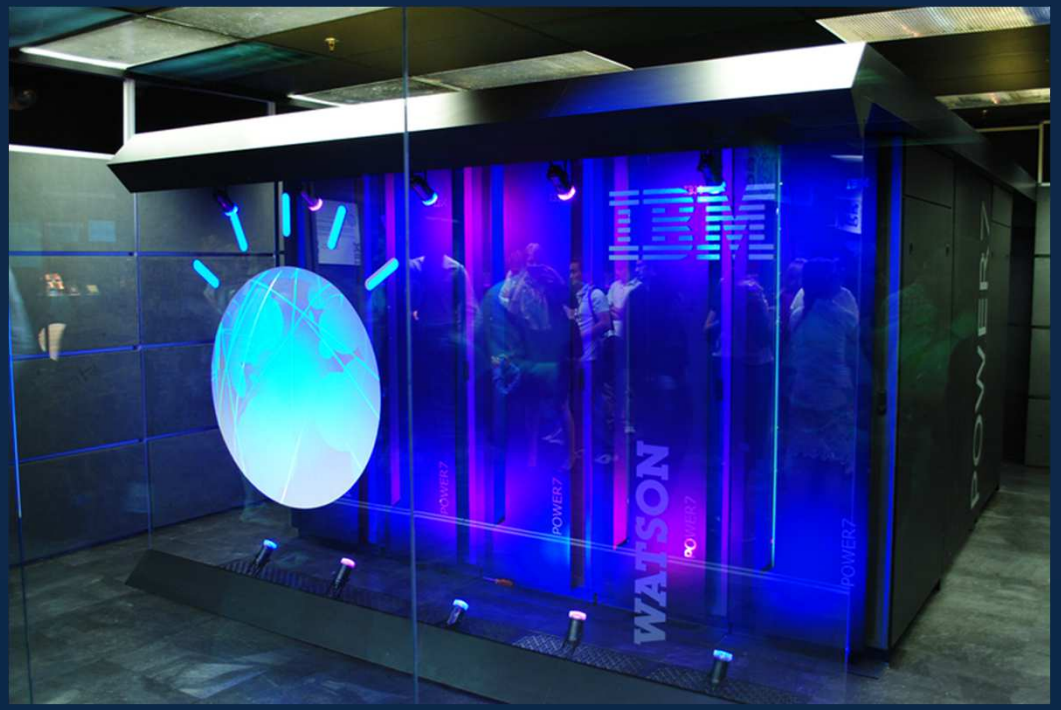


AI is Being Applied to Unstructured Content

IBM Watson can be applied to unstructured and structured content

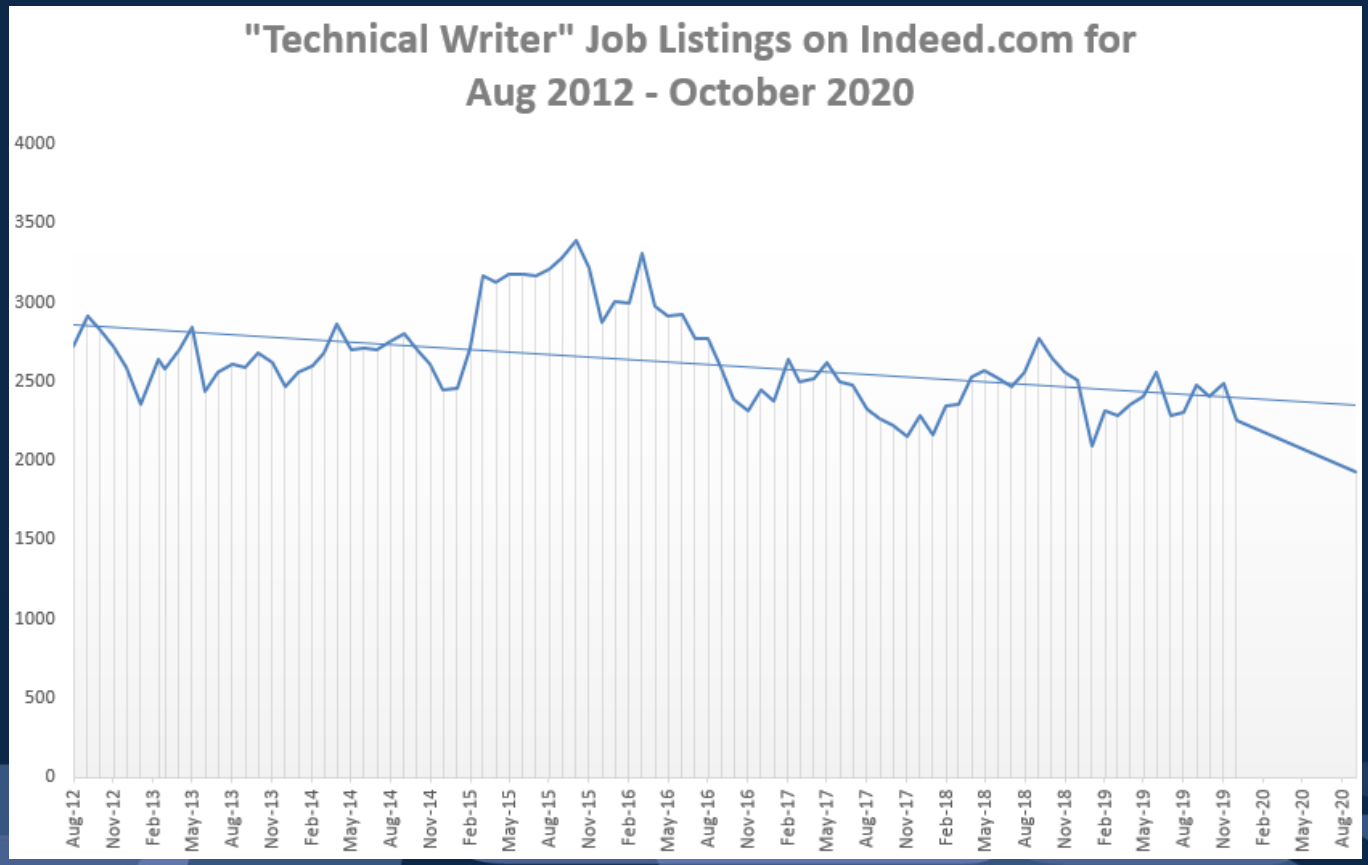
- IBM calls AI “the key to making unstructured data useful”
- Microsoft, Google, Amazon and others are all producing tools for analyzing unstructured content, images, and audio data

Still early days, but worth noting that AI (and chatbots) are not being used exclusively with structured content



Some Likely Near-term Trends

The “Technical Writer” Job Landscape is Changing…



One Reason: SMEs are Producing More Upstream Content

When it comes to API documentation, programmers are expected to provide much of the content. This is often framed and put into context by technical writers.

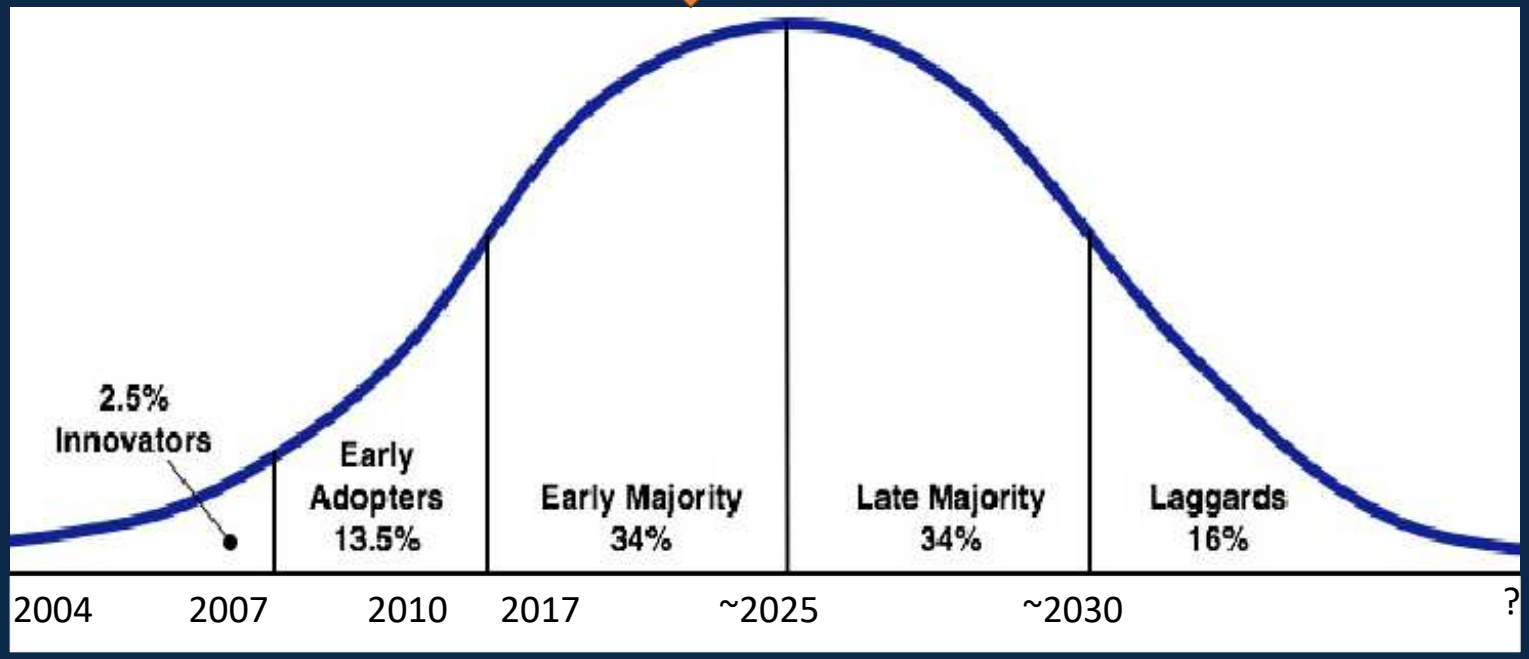
The advent of Agile documentation processes in small software development teams means that, in some circumstances, SMEs had to write content.

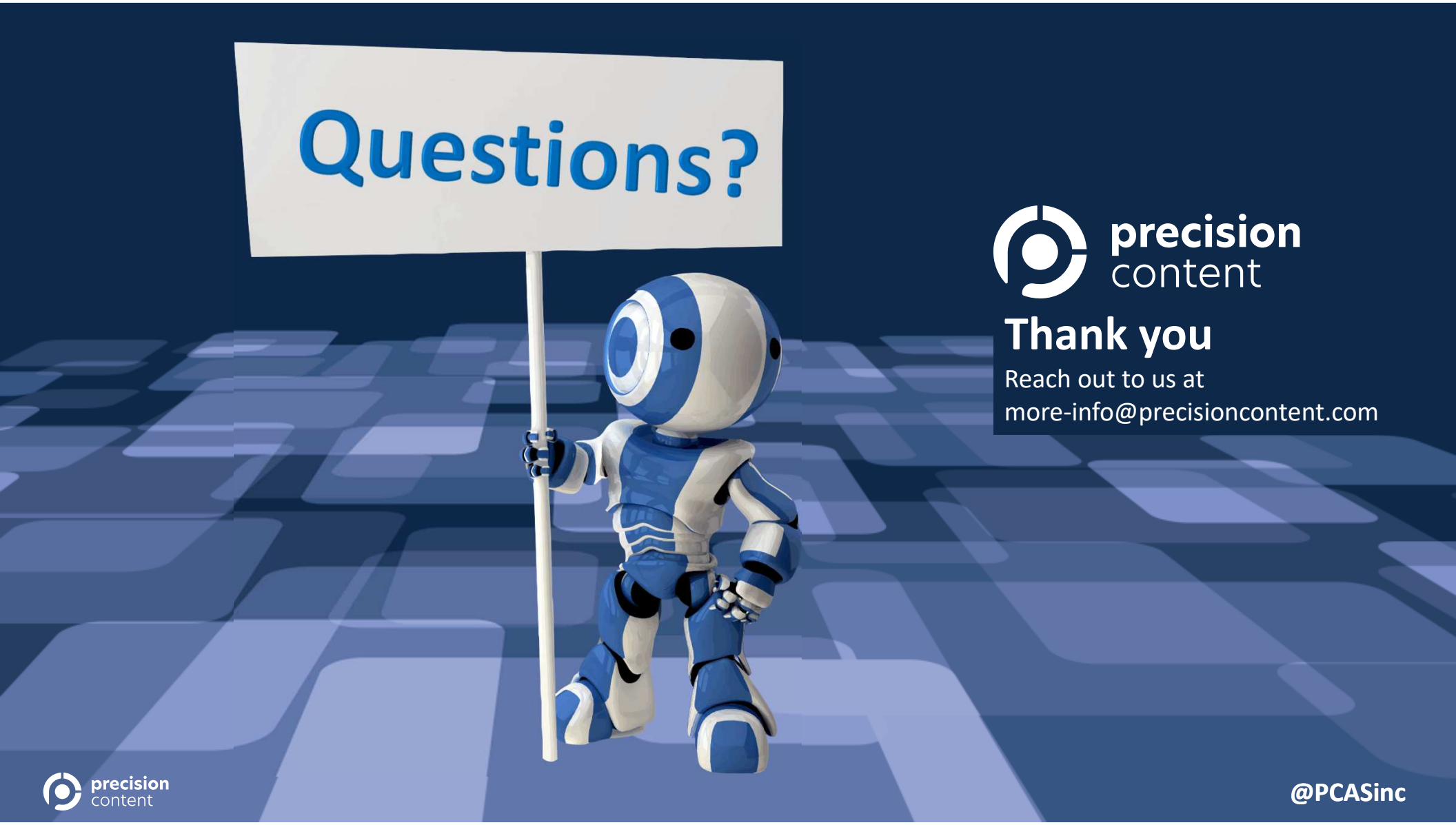
The Other Reason: The Role of the "Technical Writer" is Changing

In a survey I did of 1,500 LinkedIn profiles where people claimed to be using DITA, **66%** were not employed as traditional technical writers; some selected job titles:

- Applications Engineer
- Chief Information Architect, UX Analyst
- Consulting Content Strategist
- Content Architect
- Content Developer
- Content Management Specialist
- Content Strategist
- DITA Architect
- DITA Content Strategist
- Information Architect
- Information Developer
- Information Experience Manager
- Knowledge Architect
- Lead Information Developer
- Localization Program Manager
- Manager, XML CMS and L10n Systems
- Principal Content Experience Developer
- Principal Information Developer
- Project Manager and Documentation Engineer
- Senior Content Developer
- Senior Content Strategist
- Senior Documentation Tools Developer
- Staff Information Architect
- Team Leader Technical Documentation
- User Assistance Development Architect
- UX Designer
- XML/DITA Coordinator

Where Is DITA on the Technology Adoption Life Cycle?





precision content

Thank you

Reach out to us at
more-info@precisioncontent.com