Dazzling DITA: 20 Years of Technical Writing's Best "OmniGadget"

Keith Schengili-Roberts – DITAWriter.com

ConVEx Pittsburgh - April 13, 2026

Dazzling DITA: 20 Years of Technical Writing's Best "Omnigadget"

Agenda

1. Introduction
2. DITA is Dazzling by Design
3. State of DITA 2026
4. DITA Dilemmas

Dazzling DITA: 20 Years of Technical Writing’s Best “Omnigadget”

Introduction

Who's This Guy?

- Keith Schengili-Roberts, Head DITAWriter of DITAWriter.com
- Started as a Technical Writer 30+ years ago, currently working as Senior Staff Content Engineer at Teradata
- Have worked primarily as a Content Strategist and Consultant since early 2000s
- DITAWriter.com is my consulting company, also an industry blog established 15 years ago
- I specialize in helping companies move from unstructured to structured (DITA) content
- You can contact me at: keith@ditawriter.com

“Greatest Hits” Approach + New Research

- Greatest Hits: The first part of this presentation is something of a retrospective, bringing together information from previous work I have done along with my own experience, synthesizing it into something new

- Research: Thanks to Stilo for letting me publicly talk about research I did for them about the DITA marketplace a few years ago
  - This is mixed in with wholly new/updated information from more recent survey

Intended Audience for this Presentation

- This presentation talks about the strengths and weaknesses of DITA; of interest to technical writers, managers, but also those who work as consultants and create tools for the market
- I think the future longevity of the standard rests on all of these groups understanding the current scope and practice of DITA

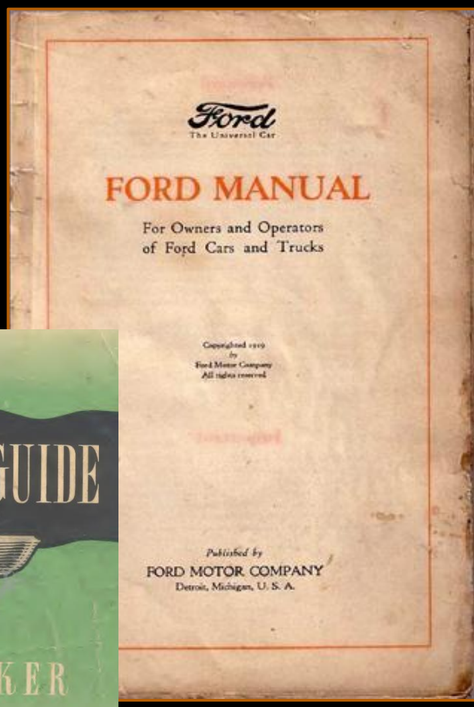
Dazzling DITA: 20 Years of Technical Writing's Best "Omnigadget"

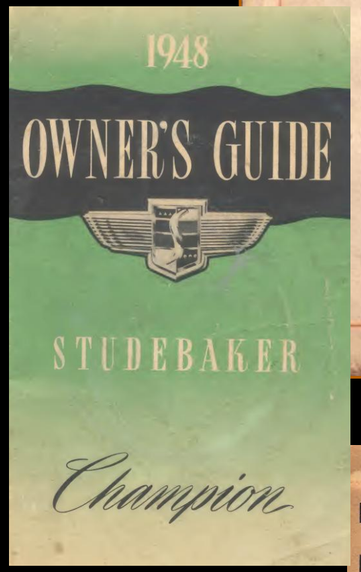
DITA is Dazzling by Design

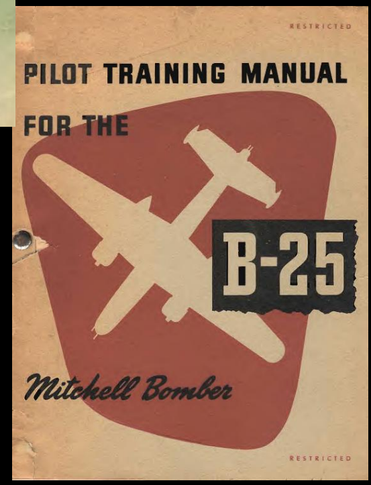
(…and by accident)

Some History: Beginnings of Technical Writing

- Technical writing as we know it started in early-mid 20th century; focused on instructing people how to work with technology
- Back when technical documentation could best be characterized as a cottage industry, manuals were typically created as one-offs, written by experts and usually aimed at experts
- This changed with the advent of consumer products and military equipment too complex to be understood at a glance.
- Non-SME writers began to enter the mix, and technical writing was recognized as a distinct profession during WWII.





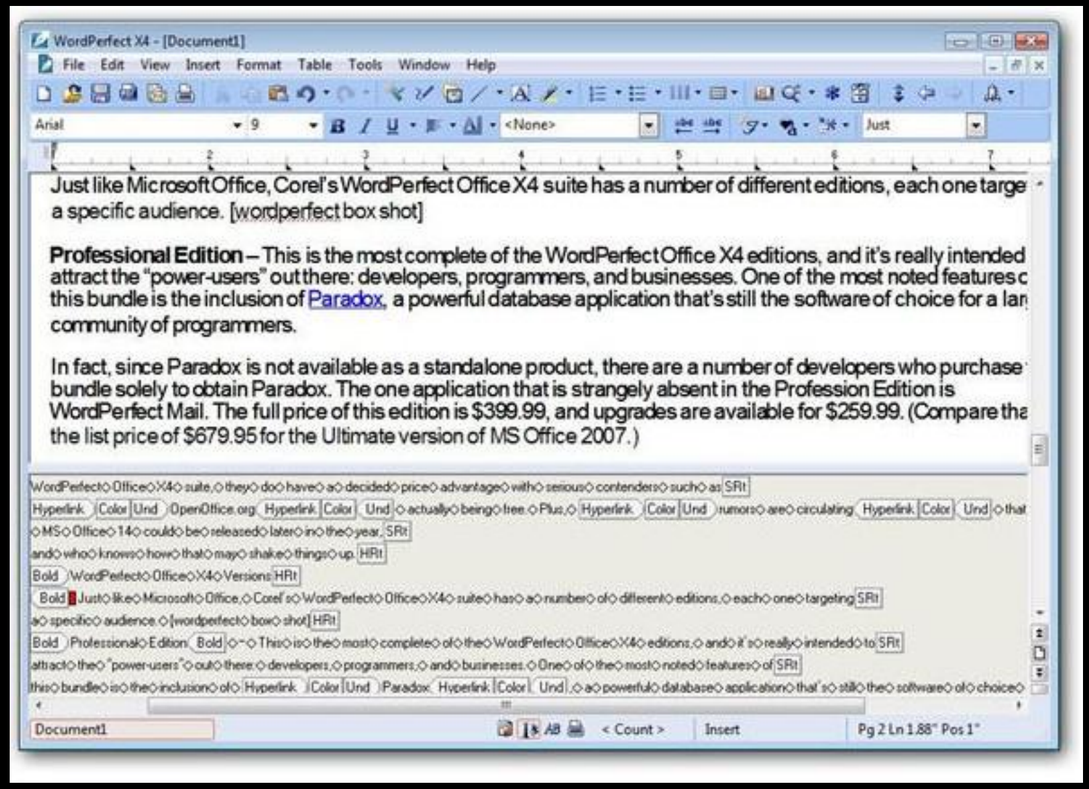


Beginnings of Structured Authoring

- 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.

Advent of Tag-based Authoring Processing

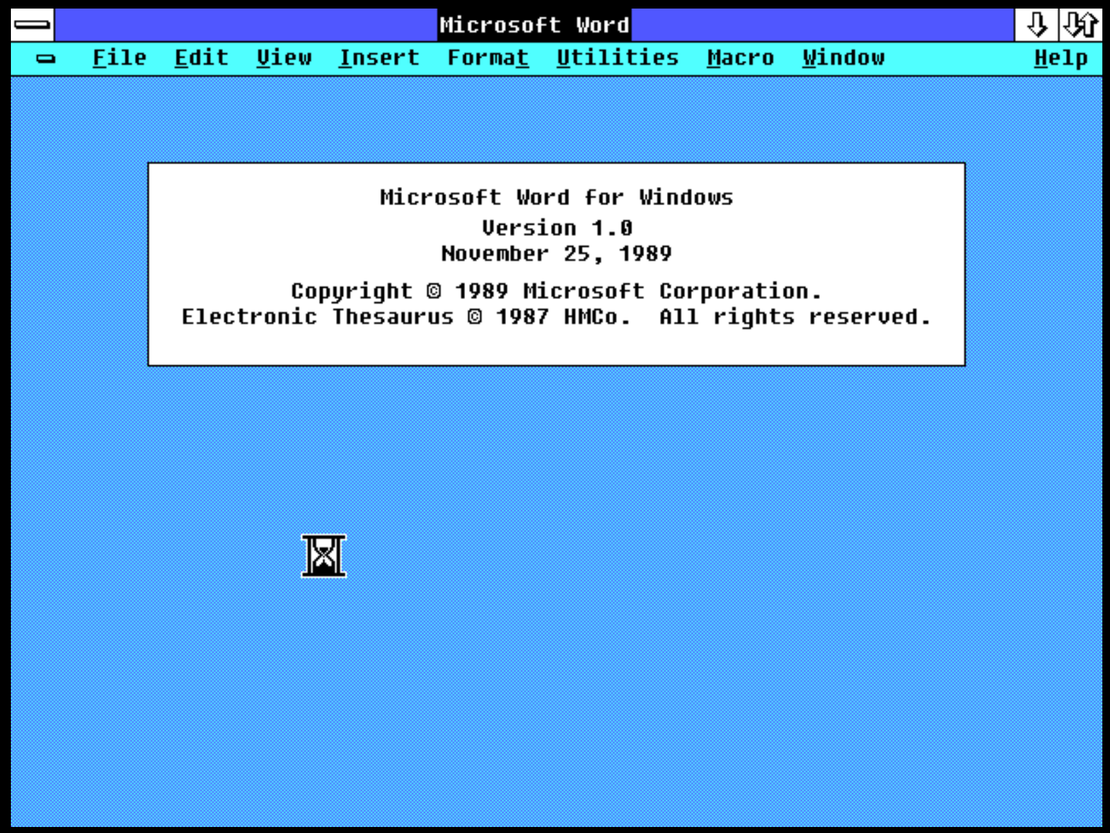
- While not the first word processing program, Word Perfect for DOS (launched November 1982) was how my generation first cut its teeth on the concept of using a computer to create content.
- A favorite feature was “reveal codes” which allowed for precise formatting
  - This is still definitely unstructured content, but anyone familiar with Oxygen in code mode will recognize that this is another form of “mark up”



WordPerfect in “Reveal Codes” Mode

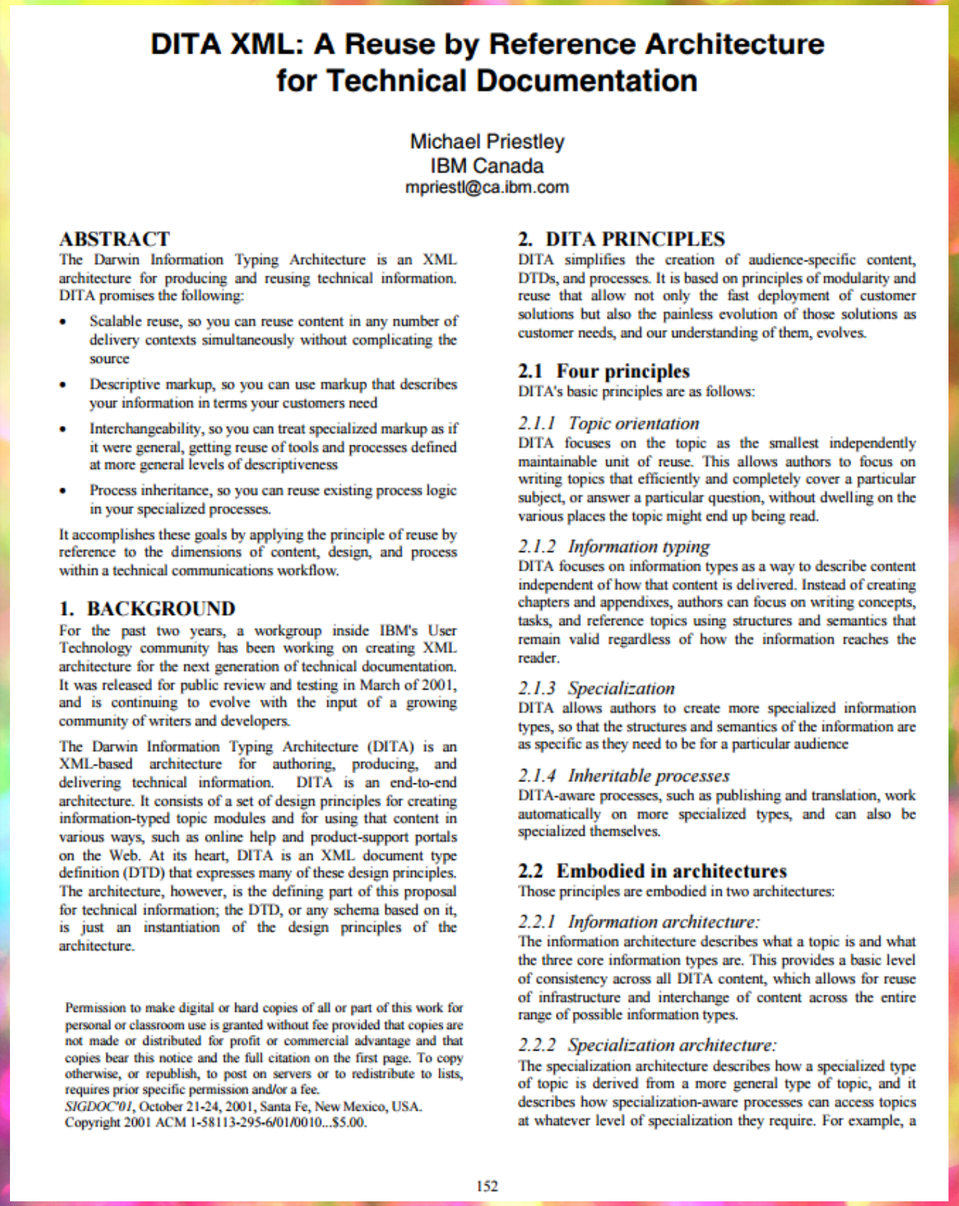
The Rise of Desktop Publishing (and Unstructured Content)

- Non-structured, but allowed for the creation of content quickly and easily
- Formatting and content went hand-in-hand
- This is the beginning of the unstructured vs. structured content dichotomy: ease-of-use + low overhead vs. significant value add but higher upfront processing



Enter DITA!

- Back in the mid- to late-90s, IBM documentation teams were facing two main issues:
  - Content needed to be tailored for the world wide 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



DITA XML: A Reuse by Reference Architecture for Technical Documentation

Michael Priestley
IBM Canada
mpriestl@ca.ibm.com

ABSTRACT
The Darwin Information Typing Architecture is an XML architecture for producing and reusing technical information. DITA promises the following:

- Scalable reuse, so you can reuse content in any number of delivery contexts simultaneously without complicating the source
- Descriptive markup, so you can use markup that describes your information in terms your customers need
- Interchangeability, so you can treat specialized markup as if it were general, getting reuse of tools and processes defined at more general levels of descriptiveness
- Process inheritance, so you can reuse existing process logic in your specialized processes.

It accomplishes these goals by applying the principle of reuse by reference to the dimensions of content, design, and process within a technical communications workflow.

1. BACKGROUND
For the past two years, a workgroup inside IBM's User Technology community has been working on creating XML architecture for the next generation of technical documentation. It was released for public review and testing in March of 2001, and is continuing to evolve with the input of a growing community of writers and developers.

The Darwin Information Typing Architecture (DITA) is an XML-based architecture for authoring, producing, and delivering technical information. DITA is an end-to-end architecture. It consists of a set of design principles for creating information-typed topic modules and for using that content in various ways, such as online help and product-support portals on the Web. At its heart, DITA is an XML document type definition (DTD) that expresses many of these design principles. The architecture, however, is the defining part of this proposal for technical information; the DTD, or any schema based on it, is just an instantiation of the design principles of the architecture.

SIGDOC'01, October 21-24, 2001, Santa Fe, New Mexico, USA.

2. DITA PRINCIPLES
DITA simplifies the creation of audience-specific content, DTDs, and processes. It is based on principles of modularity and reuse that allow not only the fast deployment of customer solutions but also the painless evolution of those solutions as customer needs, and our understanding of them, evolves.

2.1 Four principles
DITA's basic principles are as follows:

2.1.1 Topic orientation
DITA focuses on the topic as the smallest independently maintainable unit of reuse. This allows authors to focus on writing topics that efficiently and completely cover a particular subject, or answer a particular question, without dwelling on the various places the topic might end up being read.

2.1.2 Information typing
DITA focuses on information types as a way to describe content independent of how that content is delivered. Instead of creating chapters and appendixes, authors can focus on writing concepts, tasks, and reference topics using structures and semantics that remain valid regardless of how the information reaches the reader.

2.1.3 Specialization
DITA allows authors to create more specialized information types, so that the structures and semantics of the information are as specific as they need to be for a particular audience

2.1.4 Inheritable processes
DITA-aware processes, such as publishing and translation, work automatically on more specialized types, and can also be specialized themselves.

2.2 Embodied in architectures
Those principles are embodied in two architectures:

2.2.1 Information architecture:
The information architecture describes what a topic is and what the three core information types are. This provides a basic level of consistency across all DITA content, which allows for reuse of infrastructure and interchange of content across the entire range of possible information types.

2.2.2 Specialization architecture:
The specialization architecture describes how a specialized type of topic is derived from a more general type of topic, and it describes how specialization-aware processes can access topics at whatever level of specialization they require. For example, a

TechDocs at ATI in the Early 2000s

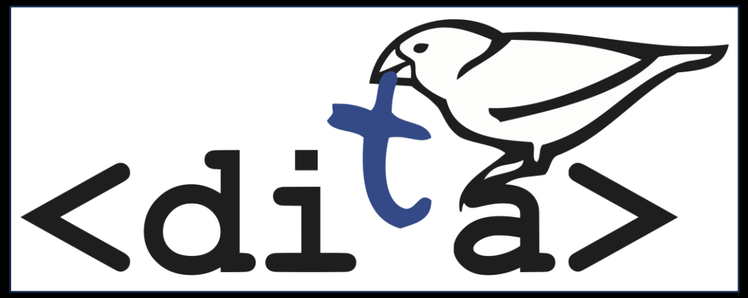
- ATI was where I was first exposed to DITA, and like IBM it had its own set of technical documentation problems to solve
- At the time, we translated select document types in up to 21 languages
  - This was very expensive; each document was tackled as a whole
- FrameMaker used extensively, but problematic; several hundred-plus page documents might take the better part of a day to publish, if it didn’t freeze
  - Freezing was a serious issue at the time, and sometimes delayed publications by several days
- Content was siloed, leading to inconsistent documentation
- Existing system was expensive and inefficient

“What’s this ‘DITA’ Thing?”

- TechDocs Management sought a solution that would make publishing more consistent, easier for the writers to work with, and reduce localization costs

- My then-Manager went to a CIDM conference in 2004 and learned about DITA XML from a joint IBM/Nokia presentation…
  - Its promise of improving the quality of content, reducing the cost of localization, and removing the task of formatting from authors was compelling

- Result: ATI started working with DITA prior to publishing of the official v. 1.0 standard
  - Created a DTD based on preliminary knowledge and had to make some best guesses as to what would come



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
- I’d argue that this has enabled the initial success of DITA from a business perspective

Chief Business Strengths of DITA

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

- Content Reuse – much of what is “dazzling” about DITA comes from this
- Lower localization costs
- Content/Formatting separation
- Multi-channel publishing
- Perfect fit for Agile production

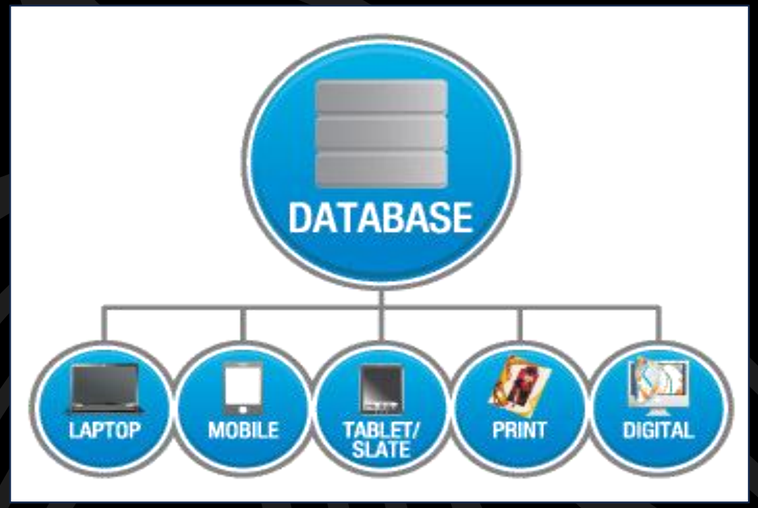
The Potential Return on Investment for Localization

- Content reuse has a direct impact when it comes to localization; reused content in the authoring language needs to be localized once and then reused subsequently*
- R.O.I. is compelling for any company that localizes to many languages; basically, the more languages, the better the R.O.I.
- Arguably it is larger companies who have worldwide distribution that get the most “bang for the buck”
- When I was managing the localization budget at AMD prior to implementing DITA, the yearly max. spend could be US $1M

Agile Works Well with DITA

- Content reuse: “write once, use many”
  - No need to re-write what already exists
  - Content consistency
  - Single-sourcing is built in
- Agile user stories map well to DITA task topics
  - Scrum-based Agile often take forms of various procedures that users will want to accomplish
- Epics are collections of related user stories comprising a complete workflow for a type of user
  - These can be used within DITA to help refine conditional processing for audiences

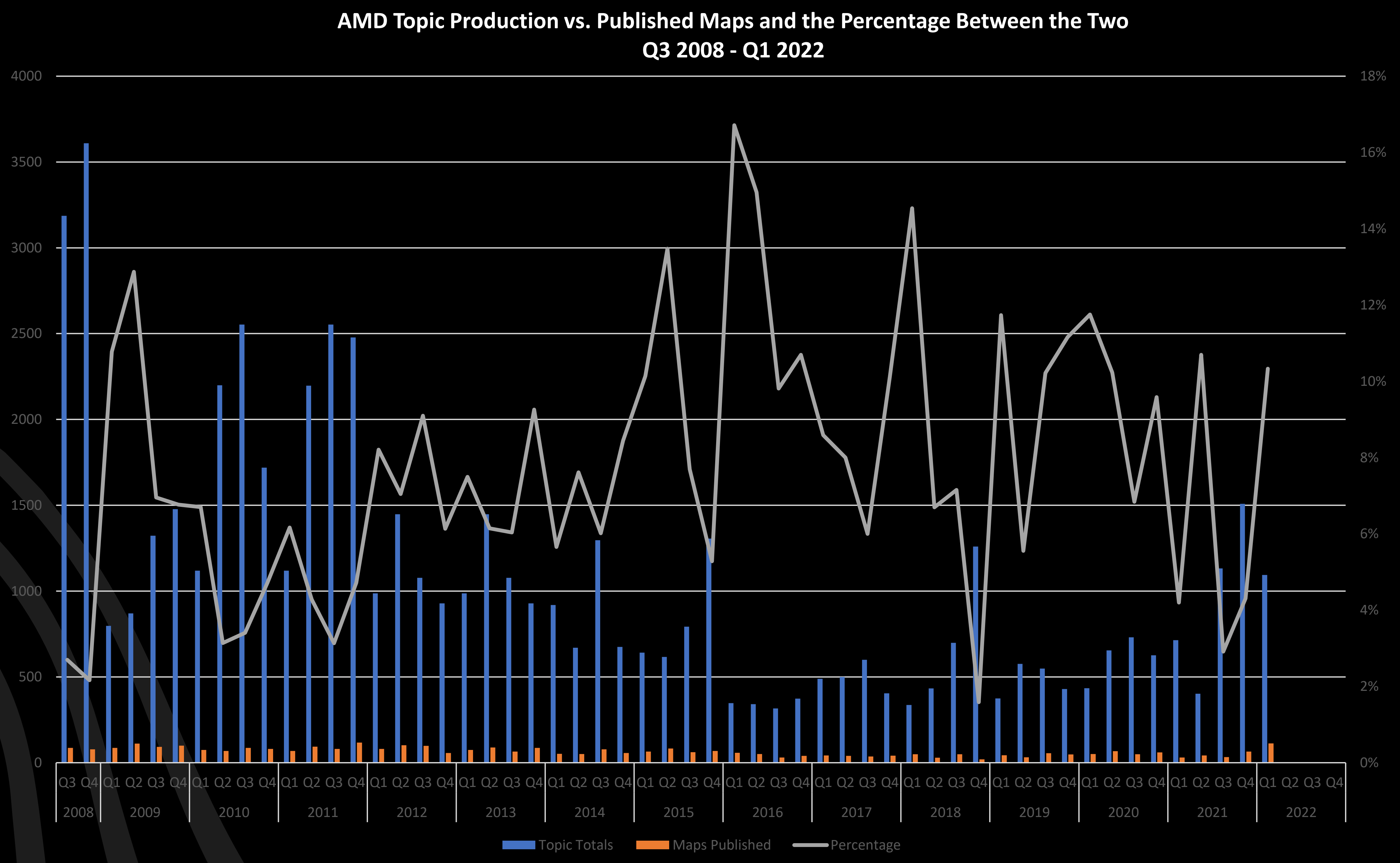
DITA Was Built with Multi-channel Publishing in Mind



- 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 just one
  - This became the DITA Open Toolkit (DITA-OT), the basis for “omnichannel” publishing
  - DITA-OT natively supports many core output formats (PDF, HTML5, XHTML, Markdown), and customized output types via plugins supports many more

DITA Enables Better Production Metrics

- Topic-based content creation means that it is much easier to extract production numbers
- It is relatively straightforward to determine rates of reuse, production efficiency, and predict when to expect surges in topic creation (due to product cycles, as seen here)



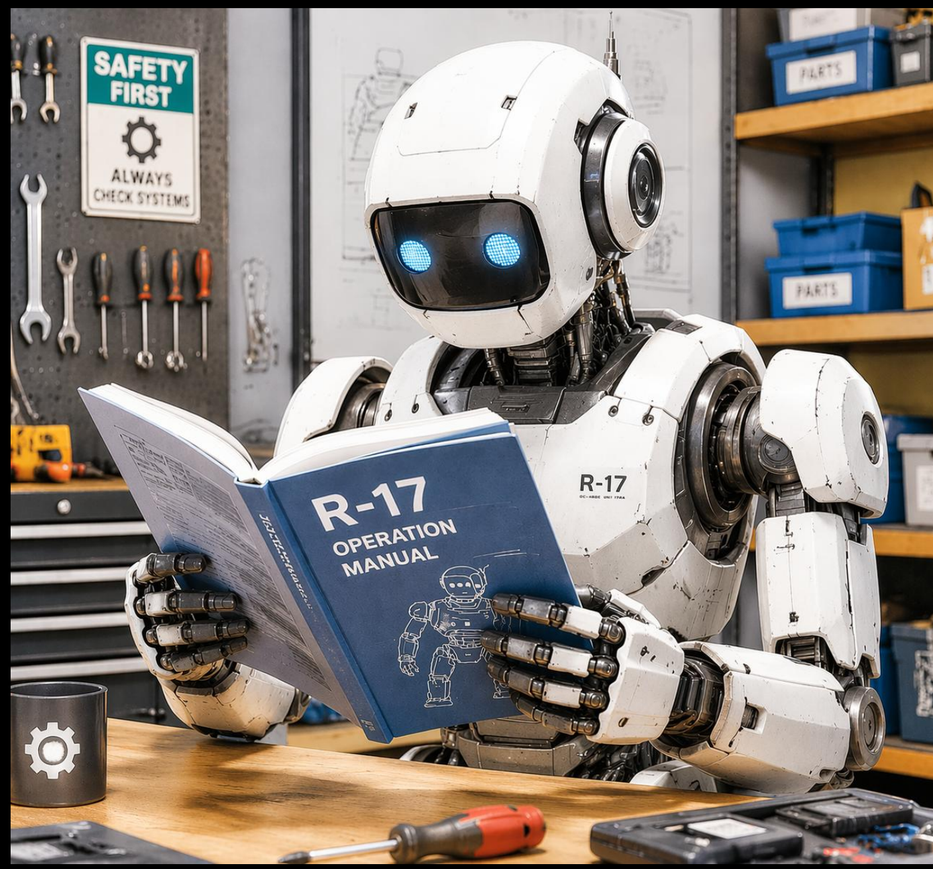
Other Not-So-Obvious Benefits of DITA

- Any of the previous scenarios make for a good budgetary argument to move to DITA
- Some teams are also swayed by how best practices are incorporated into DITA
  - Content reuse = consistent messaging
  - DITA short descriptions and metadata promotes findability = better Search Engine Optimization
  - Topic-based content easier for users to digest

  - DITA can provide context for Artificial Intelligence
    - Another “dazzling” aspect of DITA that could not have been anticipated when the specification was devised

DITA Provides Context to Artificial Intelligence

- DITA’s self-contained topics are easier for AI systems to understand, analyze, and retrieve.
- DITA elements (e.g., <steps>, <prereq>, <context>) represent specific semantic meanings that allow AI to understand the intent of the content
- DITA supports rich metadata, including audience, product, version, and region, ensuring the LLM understands the argetted audience
- Using Retrieval-Augmented Generation (RAG) to feed trusted, structured DITA content into AI models, companies ground the output in verified content, greatly reducing the AI from creating inaccurate or "hallucinated" answers.



Dazzling DITA: 20 Years of Technical Writing’s Best “Omnigadget”

State of DITA 2026

Stilo Survey from Late 2023

- I was commissioned by Stilo to investigate the size and potential growth of the DITA marketplace, and identify the key industry sectors where it was being used
  - This information was embargoed; they have given me permission to talk about the findings to a public audience
- Used LinkedIn information, covering information from over 2,400 people claiming DITA experience across the world
- Dun & Bradstreet Hoovers (D&B Hoovers) provided detailed information on company segments

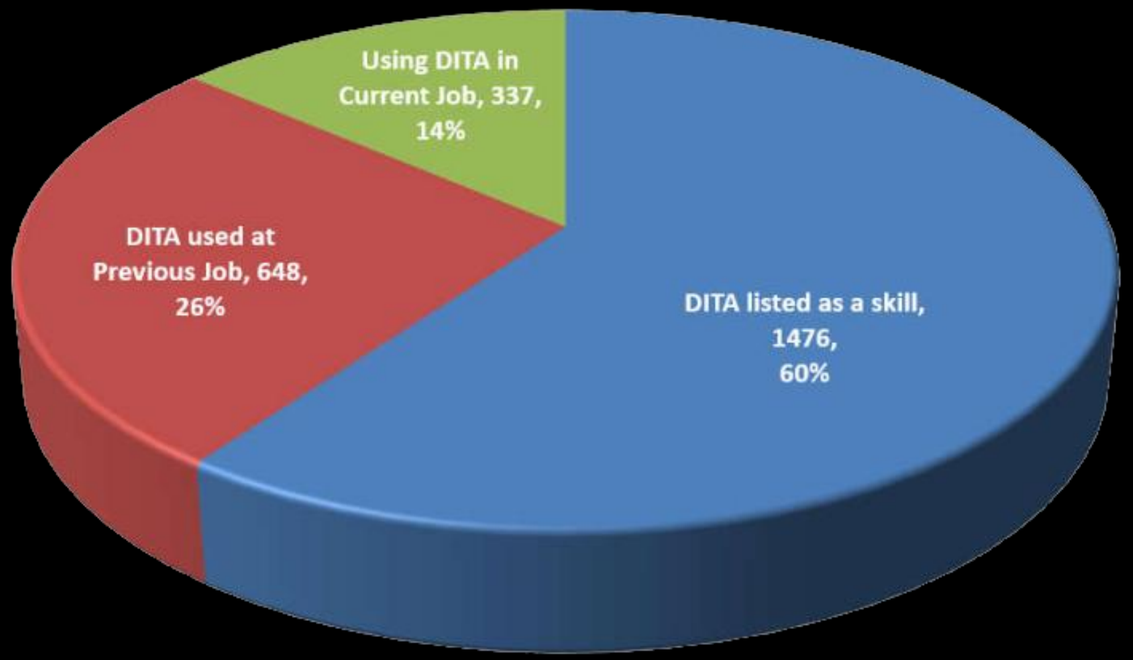
Breakdown of the Types of DITA Users Found in the Survey

Focus was on people who listed “DITA” or “DITA XML” in their LinkedIn profile

- 60% listed it as a skill, without any ties to a specific company
- 26% mentioned that DITA was used at a previous job
- Remaining 14% said that they were using DITA in their current job
- DITA experience was clearly widespread

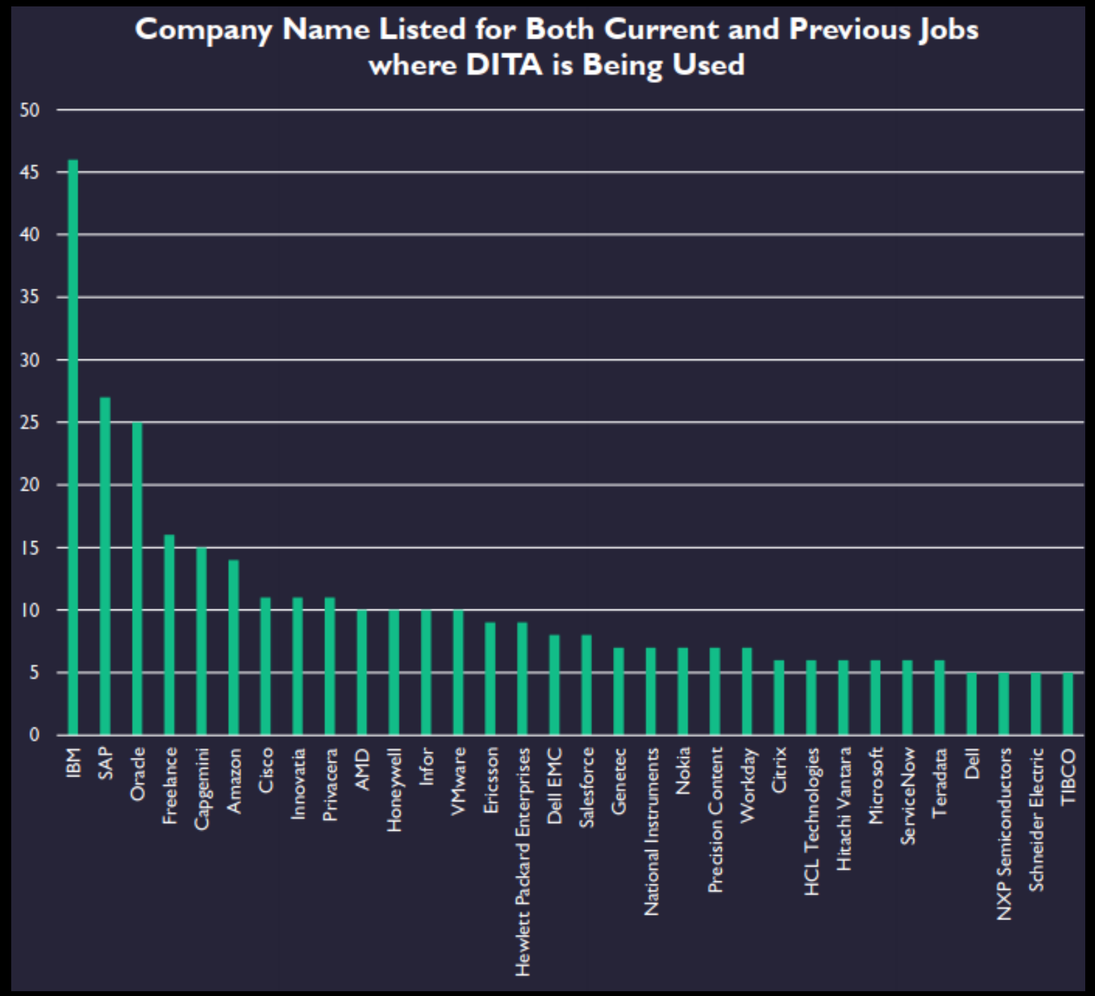
- My impression from a more recent survey indicates that “Using DITA in Current Job” has decreased, but the overall size of the pie has increased

Percentage of Users from Survey Claiming DITA Experience



Seed Companies for DITA Experience

- Several key companies (IBM, SAP, Oracle, Freescale, and others) were clearly the seed for ultimately spreading DITA experience
  - This was an important way for people to learn and transfer their DITA to the wider employment market
- There were an additional 53 instances where at least two people in this category mentioned that they had worked at a previous job at firm that used DITA



Diversity of Job Titles

- Primary search criteria was “Technical Writer” which came up 670 times, including common variants
- Some examples of job titles where DITA was referenced:
  - Agile Analysis and Technical Content Specialist
  - Certified Information Mapping Professional
  - Certified Professional Technical Communicator
  - Communications and Documentation Professional
  - Content and Information Design Specialist
  - Content Architect and Lead
  - Content Designer
  - Content Experience Designer
  - Content Strategist
  - Digital Strategist
  - Documentation Consultant
  - Documentation Engineer
  - Documentation Toolsmith
  - E-learning and Courseware Developer
  - Global Documentation Team Lead
  - Information Architect
  - Instructional Designer
  - Knowledge Manager
  - Lead Information Developer
  - Principal User Assistance Consultant
  - Product Development (Documentation)
  - Senior Content Creation Officer
  - User Assistance Developer
- My favorites were “Chaotic Good Documancer”, “40% passionate about writing, 60% into bothering engineers”, and “Happily Retired”
- Demonstrates that DITA is not just used by traditional technical writers, but a wide range of “knowledge engineers”

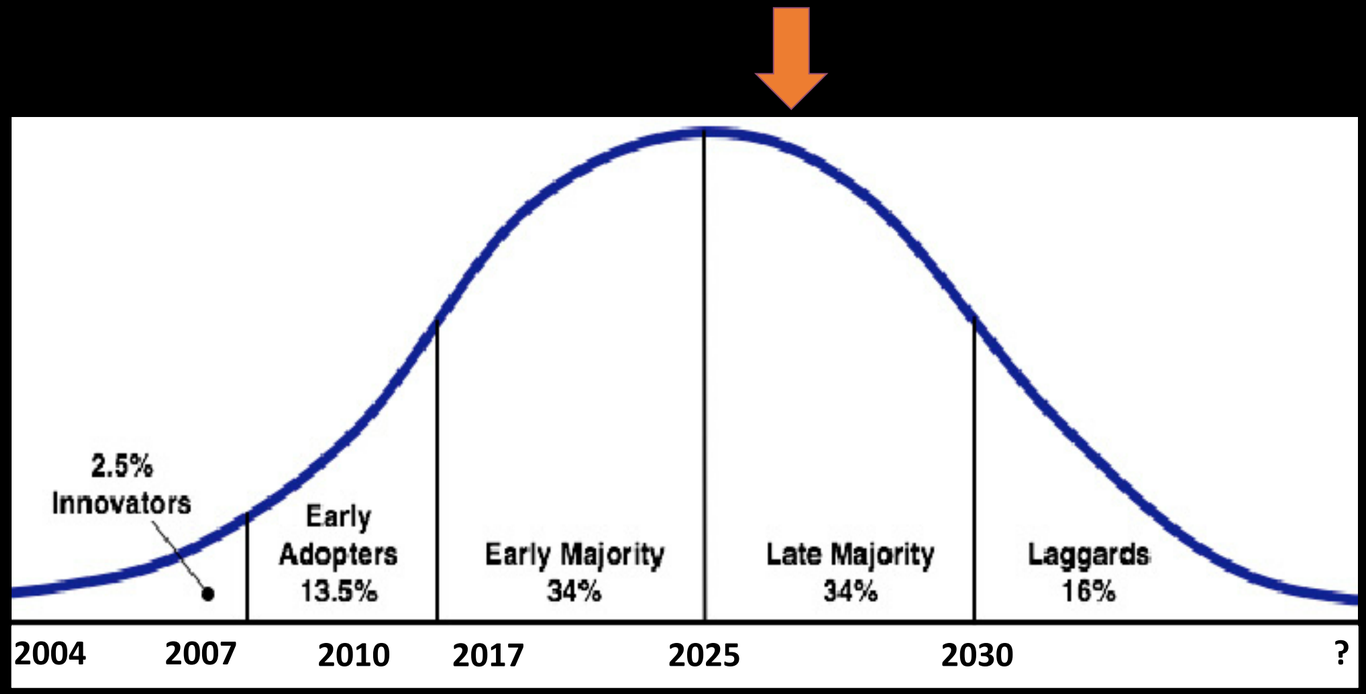
Top 20 Market Segments Using DITA (D&B Hoover Classification)

- The results largely correspond to where I have seen DITA being used according to LinkedIn's industry segmentation
- "Count" represents the number of DITA-using companies in that industry segment
- Top 5 from this list:
  - Software
  - Computer Programming
  - Computer System Design Services
  - Semiconductor Manufacturing
  - Computer and Peripheral Equipment

Hoover's Classification (Primary Industry)	Count
Software	134
Computer Programming	133
Computer System Design Services	61
Semiconductor and Other Electronic Component Manufacturing	40
Computer and Peripheral Equipment Manufacturing	35
Electromedical and Control Instruments Manufacturing	33
Miscellaneous Professional Services	32
Consulting Services	30
Architecture and Engineering	21
Medical Equipment and Supplies	19
Computer, Office Equipment and Software Merchant Wholesalers	16
Communications Equipment Manufacturing	15
Data Processing	15
Machinery Manufacturing	15
Colleges and Universities	12
Machinery and Equipment Manufacturing	12
Holding Companies	11
Research and Development Services	11
Computer Peripheral Equipment Manufacturing	10
Wireless Telecommunications Carriers	9

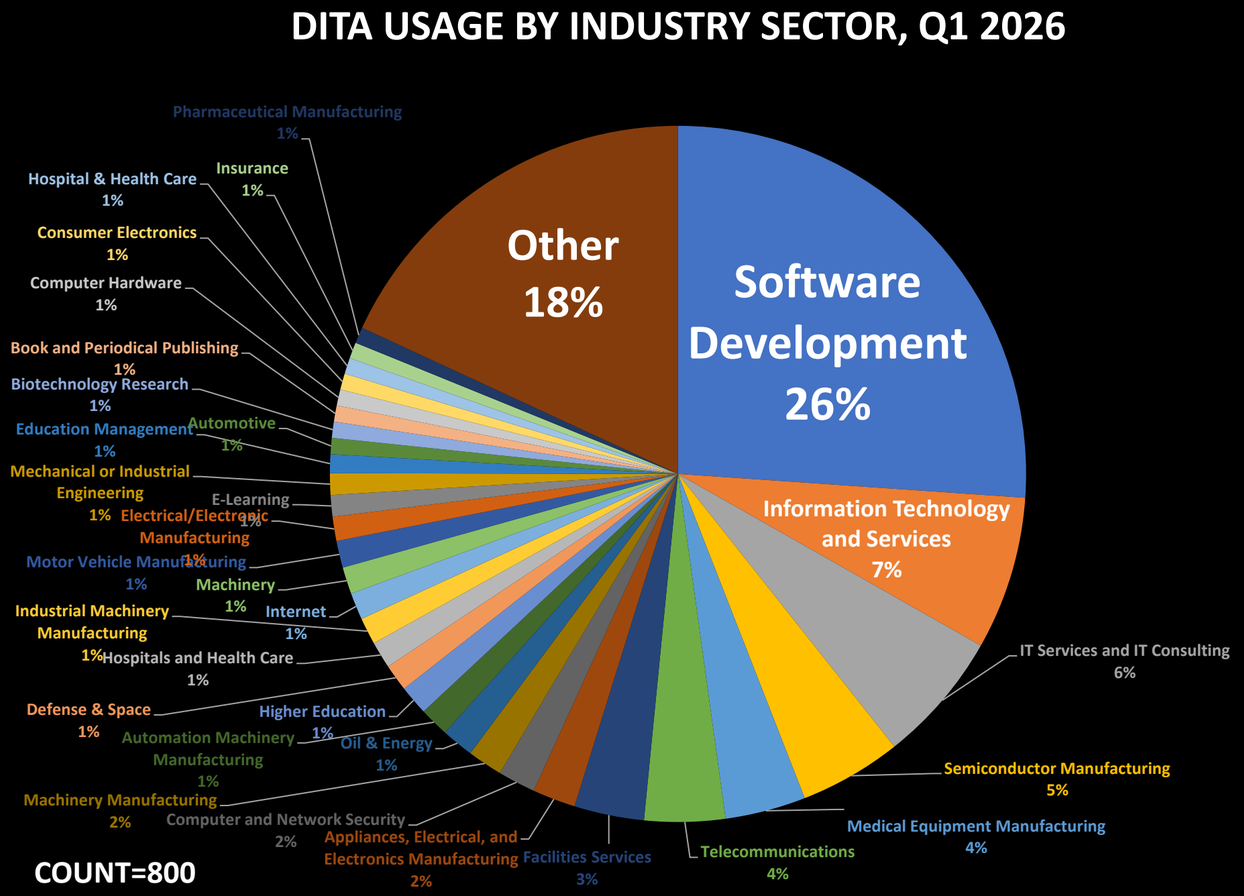
Where Is DITA on the Technology Adoption Life Cycle Today?

- In terms of the technology adoption cycle, we are most likely in the Late Majority stage
  - Not surprising since we are talking about a 20+ year old standard
- The survey I did for Stilo suggested that we were at the end of the Early Majority phase
- The real question at this point is: will the gentle curve continue?



Which Industry Sectors are Using DITA?

- This is using the latest LinkedIn info (Q1 ’26) on the number of firms known to be using DITA
- DITA started in the software sector, and it has consistently been 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
- The surprise in some ways is how diverse DITA usage is
  - “Other” in particular keeps growing, emphasizing widespread deployment of the DITA standard

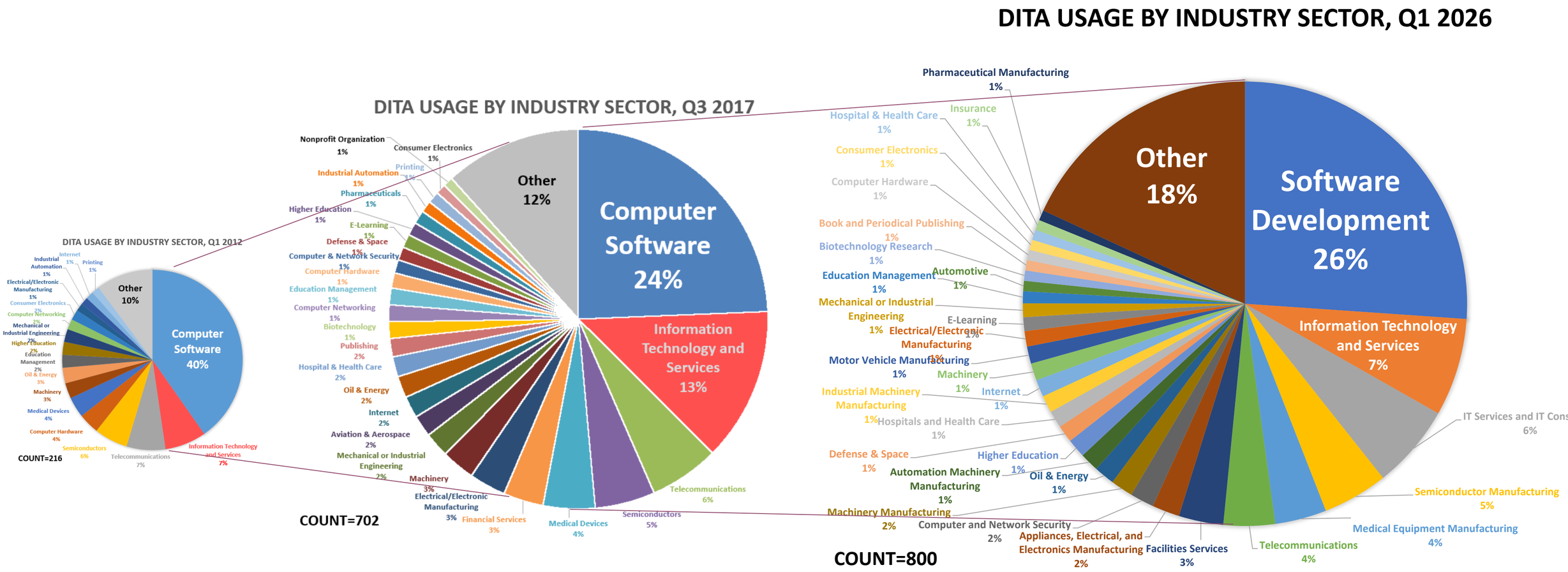


List of Firms in “Other” Category

- Accounting
- Advertising Services
- Airlines and Aviation
- Architecture & Planning
- Aviation & Aerospace
- Aviation and Aerospace Component Manufacturing
- Broadcast Media Production and Distribution
- Business Consulting and Services
- Capital Markets
- Chemical Manufacturing
- Civil Engineering
- Computer Hardware Manufacturing
- Computer Hardware Manufacturing
- Computer Networking
- Computers and Electronics Manufacturing
- Construction
- Consultancy
- Consumer Goods
- Data Infrastructure and Analytics
- Design
- Education
- Education Administration Programs
- Educational Institution
- Electric Power Transmission, Control, and Distribution
- Engineering Services
- Entertainment
- Environmental Services
- Freight and Package Transportation
- Gambling Facilities and Casinos
- Government Administration
- Ground Passenger Transportation
- Health, Wellness and Fitness
- Human Resources
- Human Resources Services
- Industrial Automation
- Information Services
- Manufacturing
- Maritime
- Measuring and Control Instrument Manufacturing
- Mining & Metals
- Nonprofit Organization
- Non-profit Organization Management
- Nuclear Electric Power Generation
- Printing
- Printing Services
- Professional Training & Coaching
- Public Safety
- Publishing
- Rail Transportation
- Railroad Manufacture
- Renewable Energy Equipment Manufacturing
- Renewables & Environment
- Research
- Retail
- Services for Renewable Energy
- Shipbuilding
- Staffing and Recruiting
- Technology, Information and Media
- Translation and Localization
- Transportation, Logistics, Supply Chain and Storage
- Transportation/ Trucking/ Railroad
- Utilities
- Wellness and Fitness Services
- Wireless
- Wireless Services
- Writing and Editing

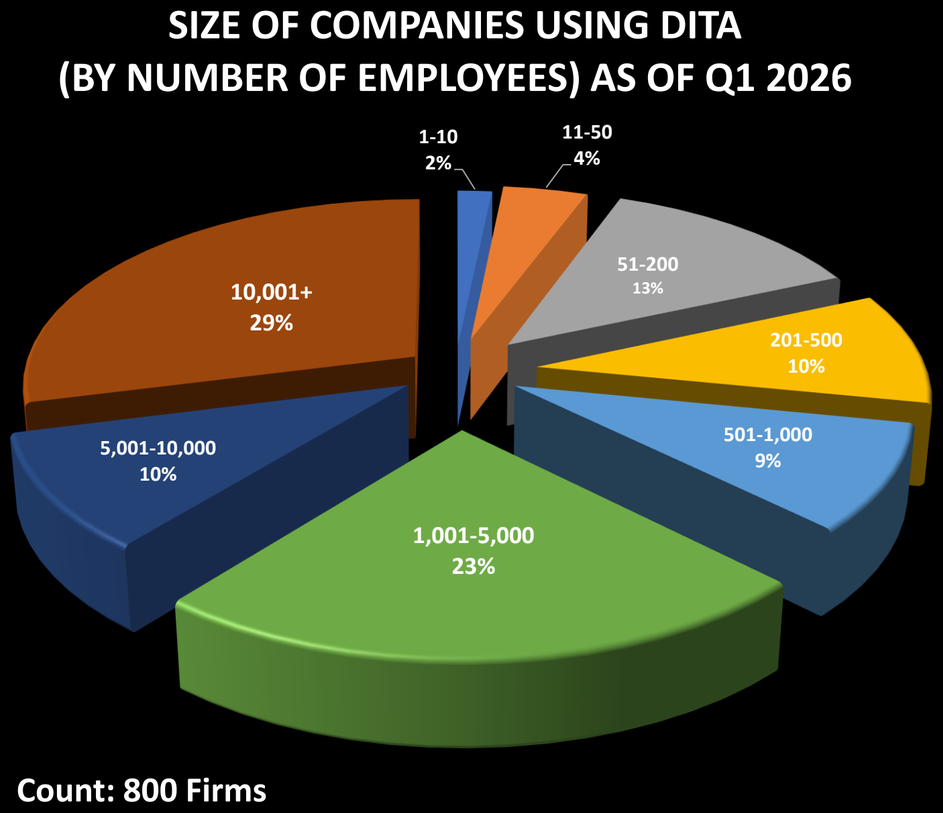
- Threshold >= 5
- A total of 66 types of firms in the “Other” category, representing 145 companies/organizations.
- An indication of the still-growing diversity DITA adoption

Relative Growth of DITA Usage from 2012 to Now



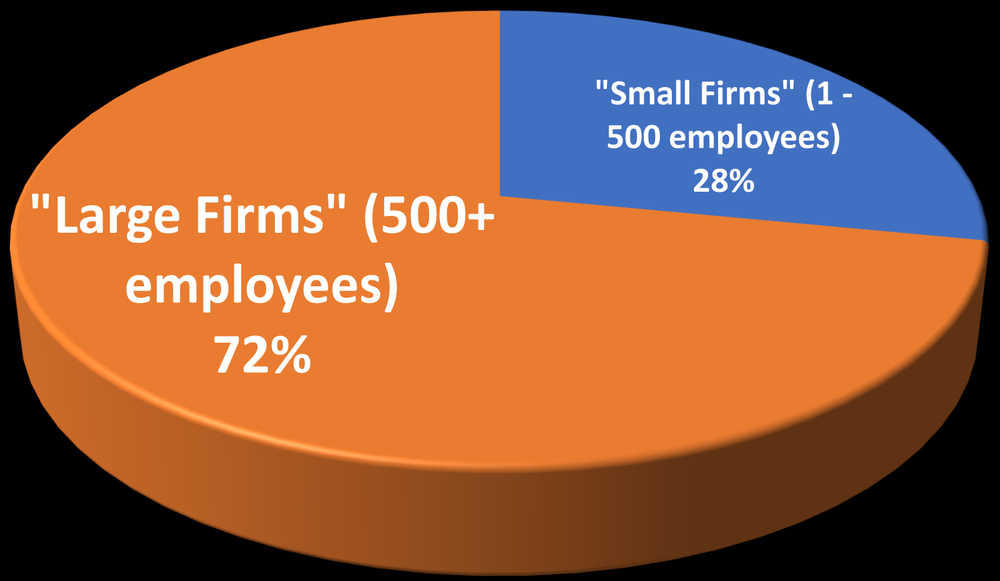
Advantages for DITA Appear to Come at Scale

- Here’s another look at the same data, but in this case looking at the size of the firms in terms of the number of people they employ
- While there are a lot of slices to this pie, the bigger slices represent larger companies



Advantages for DITA Appear to Come at Scale

- Here’s another look at the same data, but this time looking at the difference between “large” and “small” firms (at a cutoff point of 500 employees)
- The types of efficiencies that DITA brings work best at scale (more on that later)
- Back in 2021, “Large Firms” represented 64% of all companies using DITA, and “Small Firms” was 36%
  - Have seen a significant uptick of DITA use in large companies over small in that timeframe



Worldwide Distribution of DITA Using Firms

This chart tracks DITA use by HQ location:

- North America: 64%
- Europe: 26%
- Asia: 9%
- Australia: 1%
- South America and Africa: 0% (numbers don’t meet threshold)

- Percentage-wise most recent growth has been in Asia (primarily Japan, India, and China)

Dazzling DITA: 20 Years of Technical Writing’s Best “Omnigadget”

DITA Dilemmas

Who is Not Using DITA?

- Structured content requires an investment in time and money
- Benefits are definitely worth it, but may not be apparent/obvious at first
- Less obvious benefit to smaller firms
- Many firms adopting a hybrid approach
  - It is also rare to find a large firm that is using DITA exclusively

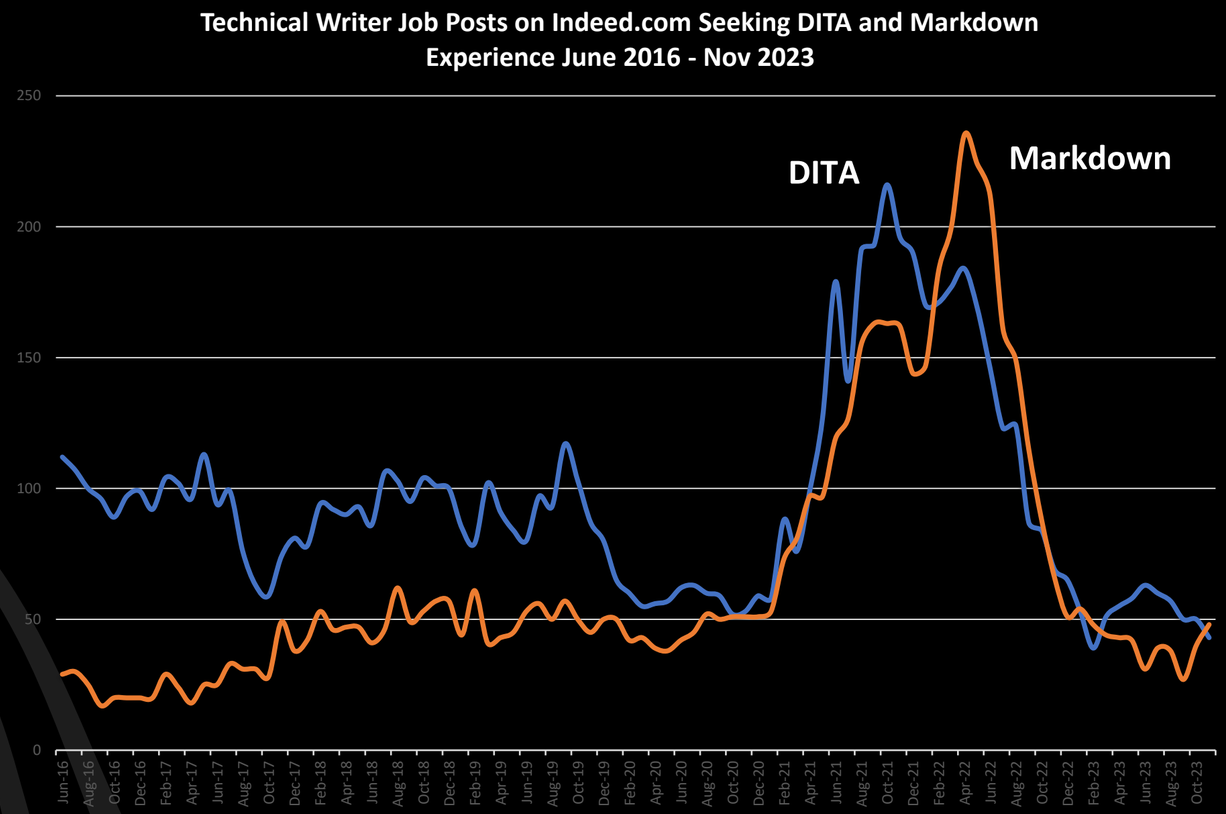
- Rise of alternatives to DITA where ROI for its use are reduced
  - No need for localization?
  - Situations where a docs-as-code approach is more compelling

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.
- And remember that software development is the single largest slice of DITA using companies

The Rise of Markdown

- While the Markdown standard is older than DITA, that is not the point
- It is what Computer Science programs are using to teach their students; XML is considered “last generation”
- It is also much easier to learn and use
- Content is not structured, but that may be offset by other benefits such as ease of use and flexibility



The Elephant in the Room: Artificial Intelligence

- What is the future of DITA in an AI world?
- Co-existence is most-likely long-term path
  - Tools such Oxygen AI Positron proves that it can be an aid to DITA technical writers
  - Documentation teams using AI to summarize existing content, create synopses of projected content, improve consistency of existing content, and to help enforce house style
- GitHub Copilot being used by developers to help them write content as they code

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?
  - Vast majority of Large Language Models (LLMs) have been trained on unstructured content; and for the most part, it works

Using AI to Add Structure to Unstructured Data

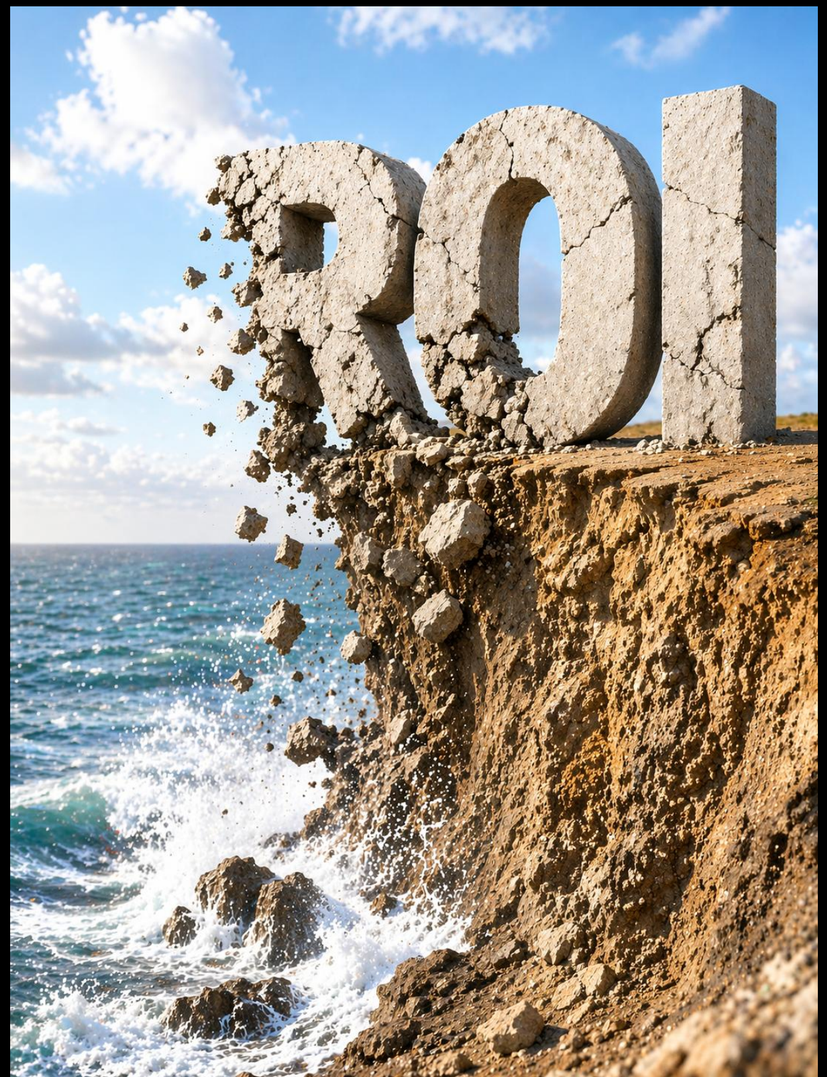
“I love unstructured data because we have been wrestling data into structure for years. That's been people's full-time jobs is to structure their data because that's what made it useful. That's how we were able to get those insights out of it. And AI is turning that on its head. Let’s take all of that unstructured data and put it to use.”

“Extracting structured data from messy documents with fewer manual corrections and faster processing is where AI actually drives impact.”

- Amanda Saunders, Director Enterprise Gen AI Product Marketing, NVIDIA

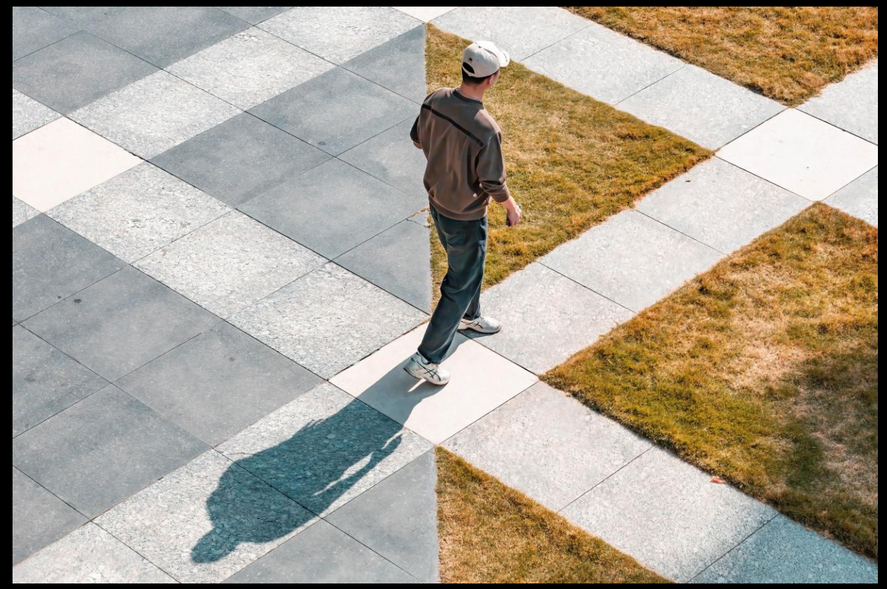
Factors Leading to Possible Erosion of DITA ROI

- AI is being used to significantly reduce localization costs by automating high-volume, repetitive tasks and shifting human efforts towards higher-value review roles
  - Localization tool vendor XTM claims using AI can reduce translation errors by 50%, and overall costs by 60% or more
- Taxonomic context can be added to Markdown (most typically through yaml)
  - While still not “structured”, it is a means to provide context



There are Now Issues About DITA *Retention*

I was asked to remove a software company from the Companies Using DITA listing as they were no longer using DITA.

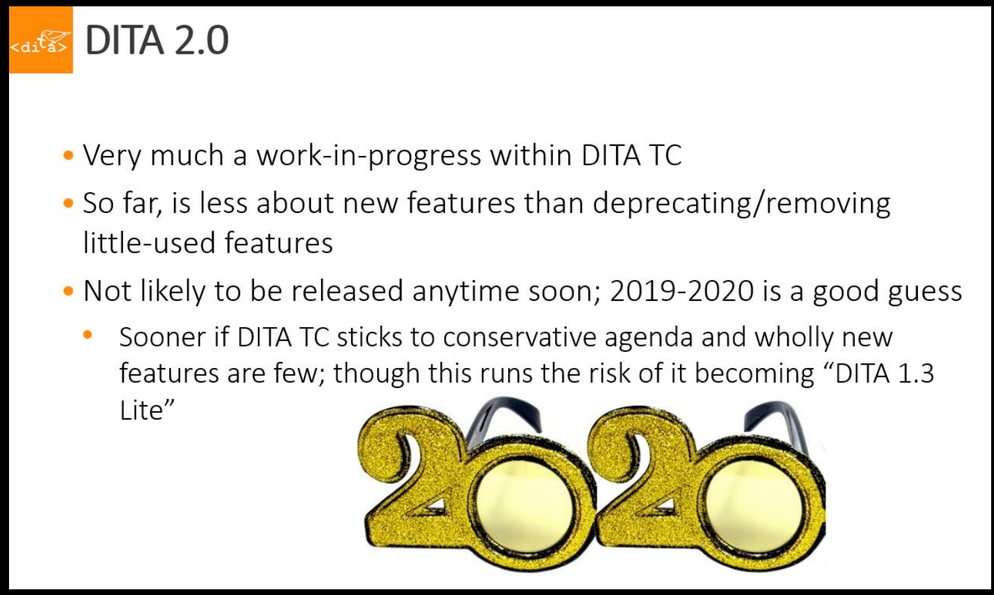


> “One of the main drivers was that we were using DITA for product documentation and a separate Markdown-based approach for API docs, and we wanted a single solution for both. Our internal systems for publishing DITA content to HTML and PDF were also fragile and required significant ongoing support. In addition, writers needed to publish updates quickly, which our DITA setup didn’t handle well. Since we don’t translate our product documentation, the ROI for DITA just wasn’t there for us.”

A Plea for the Status of DITA 2.0

- I have been talking about the release of DITA 2.0 since at least 2017
  - COVID, lack of financial funding, and the slow whittling away of active members (myself included) has contributed to this delay
  - This effort is entirely voluntary and unpaid; it is heroic work

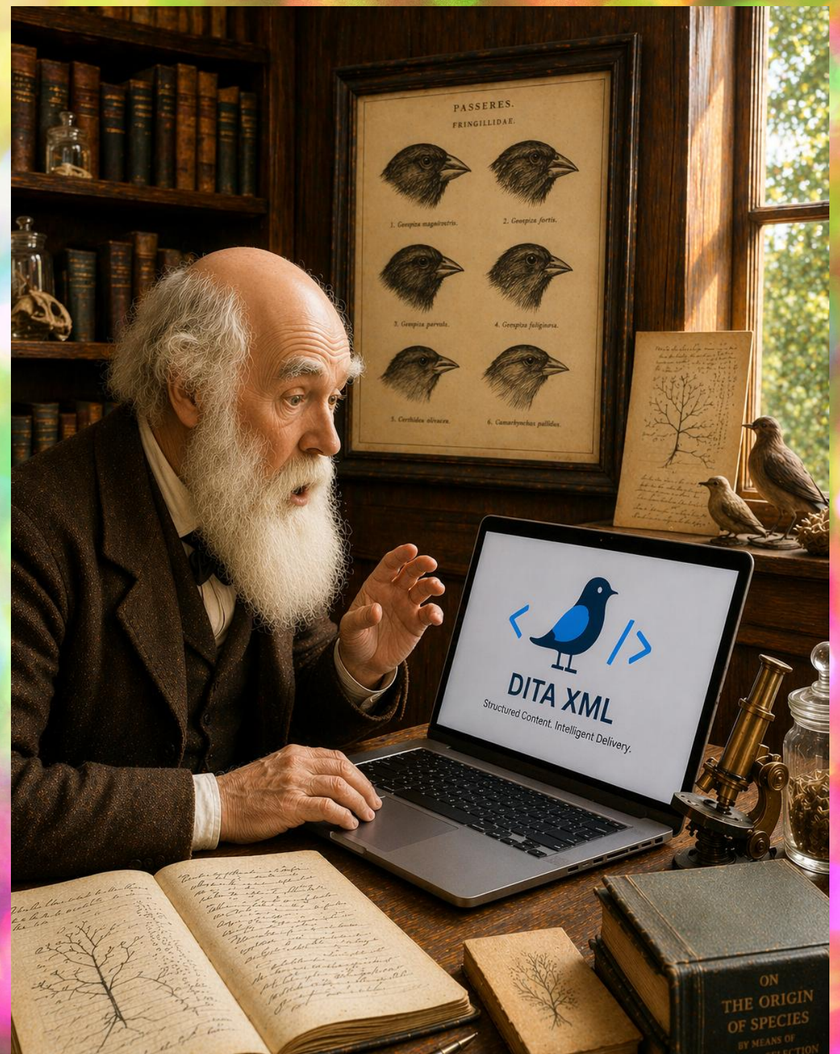
- Hoping for a committee note or statement from the Technical Committee to the wider community that the TC still has a heartbeat
  - I have been approached by some companies on the implementation parts of the upcoming DITA 2.0; problem is potential “drift” from what they implement to what is eventually released, much like what I experienced at ATI all those years ago



Slide from a Deck I Presented Back in 2017

Is There a Dazzling Future for DITA?

- Yes.
- The original design intent of the DITA standard has proven to be sound and resilient.
  - Despite significant technological change over the past 20 years it has proven to be adaptive, as the “D” for “Darwin” in its name would suggest
- It can definitely be thought of as an “omnigadget”, capable of being useful for purposes that were not foreseen when it was launched.
- DITA will not last forever, but it is likely that aspects of its core will be used in the next generation of technical content, whatever form that may take



Q&A

Thank You