

<di**t**a>

+



# Improve Your Chances for Documentation Success with DITA and a CCMS

Keith Schengili-Roberts  
DITA Evangelist, IXIASOFT

# Agenda

- Introduction
- Overview of Key Features of DITA
- Four Chief Reasons for Wanting to Move to DITA + CCMS
- Four Things You Can Do with DITA and a CCMS
- Q/A

## Who's This Guy?



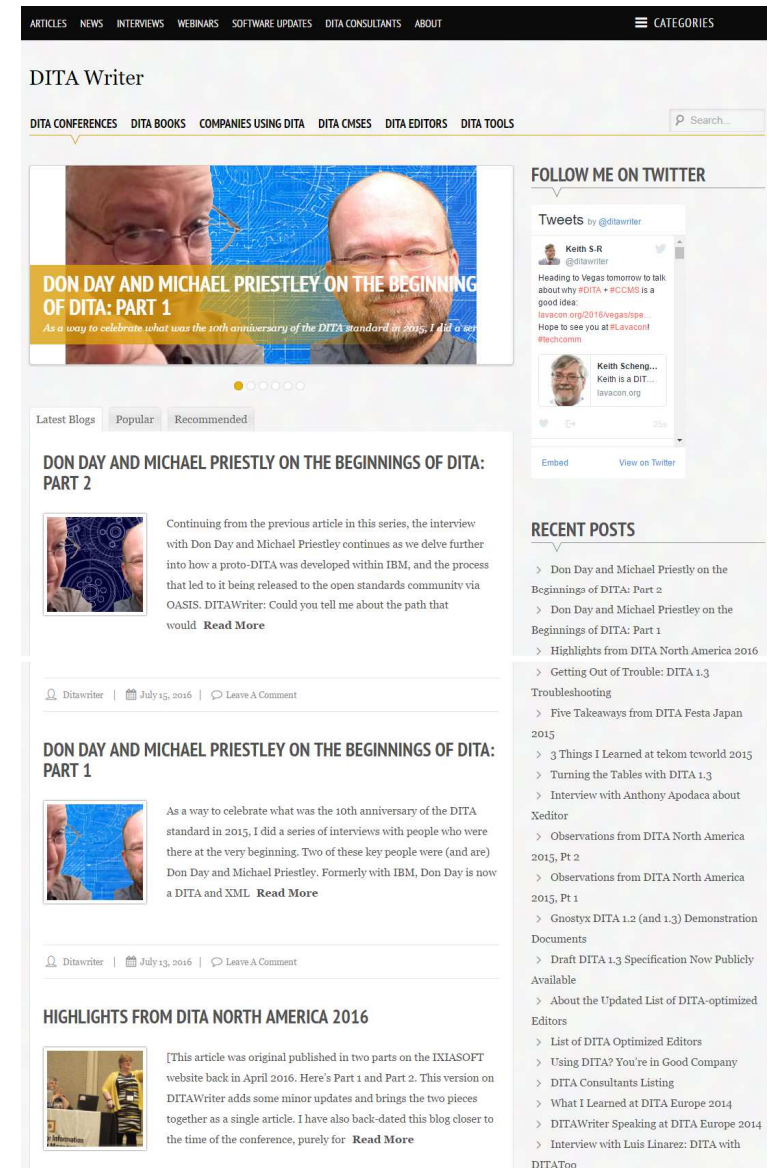
**Keith Schengili-Roberts**  
IXIASOFT DITA Evangelist  
and Market Researcher

### What I do:

- DITA evangelist
- Liaison with OASIS; on DITA Adoption and Technical Committees
- Industry researcher
- Lecturer on Information Architecture, University of Toronto
- 10+ Years of DITA XML experience

# Also Known As “DITAWriter”

- Industry blog started +5 years ago
- Just over 220,000 hits
- Regularly updated info on:
  - DITA Conferences
  - DITA Books
  - Companies Using DITA
  - DITA CMSs
  - DITA Editors
  - Other DITA Tools
  - DITA Consulting Firms
- News/views on DITA use
- Features interviews with those making a difference in the world of DITA



The screenshot shows the DITAWriter website interface. At the top, there is a navigation bar with links for ARTICLES, NEWS, INTERVIEWS, WEBINARS, SOFTWARE UPDATES, DITA CONSULTANTS, and ABOUT. A search bar is located on the right. Below the navigation, the page title is "DITA Writer". A main banner features a photo of Don Day and Michael Priestley with the text "DON DAY AND MICHAEL PRIESTLEY ON THE BEGINNING OF DITA: PART 1". Below this, there are tabs for "Latest Blogs", "Popular", and "Recommended". The main content area displays two blog posts, both titled "DON DAY AND MICHAEL PRIESTLEY ON THE BEGINNINGS OF DITA: PART 1" and "PART 2". Each post includes a small image and a brief description. To the right, there is a "FOLLOW ME ON TWITTER" section with a list of tweets and a "RECENT POSTS" section with a list of article titles and dates.

# Any STC Members Here?

- Much of this presentation is based on my article published in last month's STC magazine *Intercom*
- I plan to expand upon its points and go into more details than space allowed in the original article

## The Benefits of DITA Authoring WITHIN A Component Content Management System

BY KEITH SCHENGILI-ROBERTS

IT IS RARE for documentation teams to not use some form of database or repository for storing their technical content. A documentation team of any size will generate significant amounts of content over time that eventually outgrows shared network folders. This is doubly-so for any documentation team that moves to a topic-based writing environment like DITA, as the concept of topic reuse relies upon there being a significant amount of previous content to leverage.

What I have observed in documentation groups over the years is that there often comes a point where any further growth depends on moving away from a simple file folder structure or database to a content management system; and more specifically for DITA-based content, a component content management system (CCMS) where the "components" are individual topics.

When a documentation team is searching for a CCMS, they often do not know about all of the features that typical mature CCMSs have to offer. Moving to a CCMS is an opportunity for a documentation team to reinvent themselves and improve their processes, not to ossify existing and outmoded practices. When crafting a request for purchase of a CCMS, think about where you want your documentation team to be in five years and find a CCMS that can get you there. The following provides a brief survey of the types of features and benefits that can be gained by moving your DITA documentation processes over from a file system to a CCMS.

### Versioning of Content

Content versioning is a feature common to most content management systems, although not to most file systems.

Originally created for software development, versioning allows users to "check out" content for authoring and "check in" any changes. This type of practice is for programmers, as it allows them to revert to a previous version of their code to help root out bugs. Similarly, a typical CCMS will have versioning capabilities that ensures that only a single technical writer can work on a given topic at any one time, and will also register who made which edit and allow for a quick comparison between versions. Just like in a programming environment, any documentation "bugs" that are discovered in a later version of a topic can be reverted to an earlier version without the problem. A CCMS should also include the ability for authors to comment on the topic- and map-level changes they are checking into the system so that someone else can tell at a glance what changes occurred and when.

### Workflow

Another feature mature CCMSs provide is workflow. In addition to being able to check in/check out versioned content, authors can also notify the system when they consider their work on a topic to be complete. Depending on how the workflow is setup, the topic can then be routed automatically for someone else to review and approve, such as an editor or product manager. Automating workflow introduces many ways to improve documentation processes and quality. In an engineering environment, subject matter experts (SMEs) may be assigned within the system to write content which is then "polished" by a technical writer, and then sent to another SME for approval. This type of workflow implies that roles can also be assigned to contrib-

## What is this DITA Stuff?

- It's an open XML standard for creating structured content built primarily around the idea of content reuse
- Has been around for just over 12 years, has gained significant traction in technical authoring community
- DITA is used by ~5% of all technical writing groups in North America (and still growing!)



# Reuse is Built-in to DITA

- DITA was built around the idea of content reuse
  - This has helped make DITA the fastest growing XML-based technical communications standard



## 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.  
SIGDOC'01, October 21-24, 2001, Santa Fe, New Mexico, USA.  
Copyright 2001 ACM 1-58113-295-6/01/0010...\$5.00.

### 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

# DITA Content Reuse = Consistent Messaging

- Benefit of consistent content and messaging
- Consistent content means consistent user experience
  - Along with being seamless, available and context-specific



**CONSISTENCY**  
**IS** 

## Reduced Localization Costs with DITA

- Content reuse in English = localization savings
- If many target languages, ROI argument for moving to DITA (+ CCMS) is easier



More on this later...

## Separation of Form from Content

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

Less  
Time  
Spent  
Formatting

# 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



## Do You Need a CCMS?

- This was a question asked by Eliot Kimber's excellent presentation "DITA for Small Teams"
- I like to think of this presentation as a look as to the benefits a CCMS provides when the answer to this question is a definite "yes"

### DITA for Small Teams

Low cost solutions for DITA authoring, management, and production using low-cost, high-value tools

Eliot Kimber  
Contrext, LLC


DITA North America 2015

4/29/2015

DITA For Small Teams | DITA NA 2015 | Contrext, LLC

1





Typical pain points and how DITA + CCMS  
can help lessen them

## **FOUR CHIEF REASONS FOR WANTING TO MOVE TO DITA + CCMS**

## Pain Points that DITA + CCMS Can Solve

- Need to be more efficient/do more with less
- Have outgrown current toolset
- Escalating localization costs
- Requirement to verify content before it goes out the door
- In addition you also get to be one of the “cool kids” ;)



## Need to Be More Efficient / Do More with Less

- For many documentation teams it is not possible to get more headcount
- Need to reduce localization costs
- DITA + CCMS has a proven track record for improving overall costs + efficiency
  - Does not come for free however; there is a need to justify costs of moving to DITA + CCMS (stay tuned!)



## Have Outgrown Current Toolset

- Traditional Desktop Publishing tools cannot cope with growing need to produce content
- Hard for writers to find existing material for reuse in file folders or existing repositories
- Pressing need to automate existing processes
- Seeking a more mature process



## Escalating Localization Costs

- Ongoing need to produce localized content, but costs (especially with desktop publishing-based processes) have become prohibitive
- Additional costs are incurred for:
  - Formatting content
  - Fixing code-page errors (a glyph is transformed into a different character)
  - Sending more content than necessary for localization



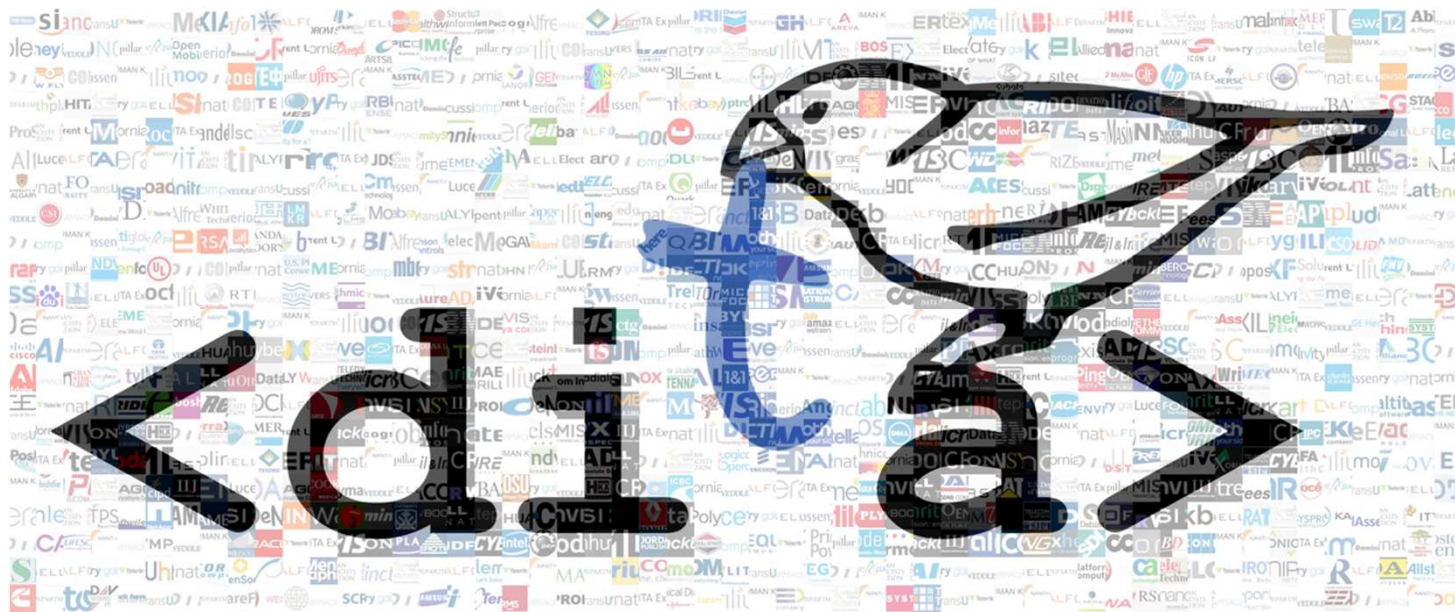
## Requirement to Verify Content Before it Goes Out the Door

- In regulated environments such as in the medical device sector, content must first be vetted by medical professionals
- Similar requirements in other sectors, such as legal publishing, heavy manufacturing, etc.



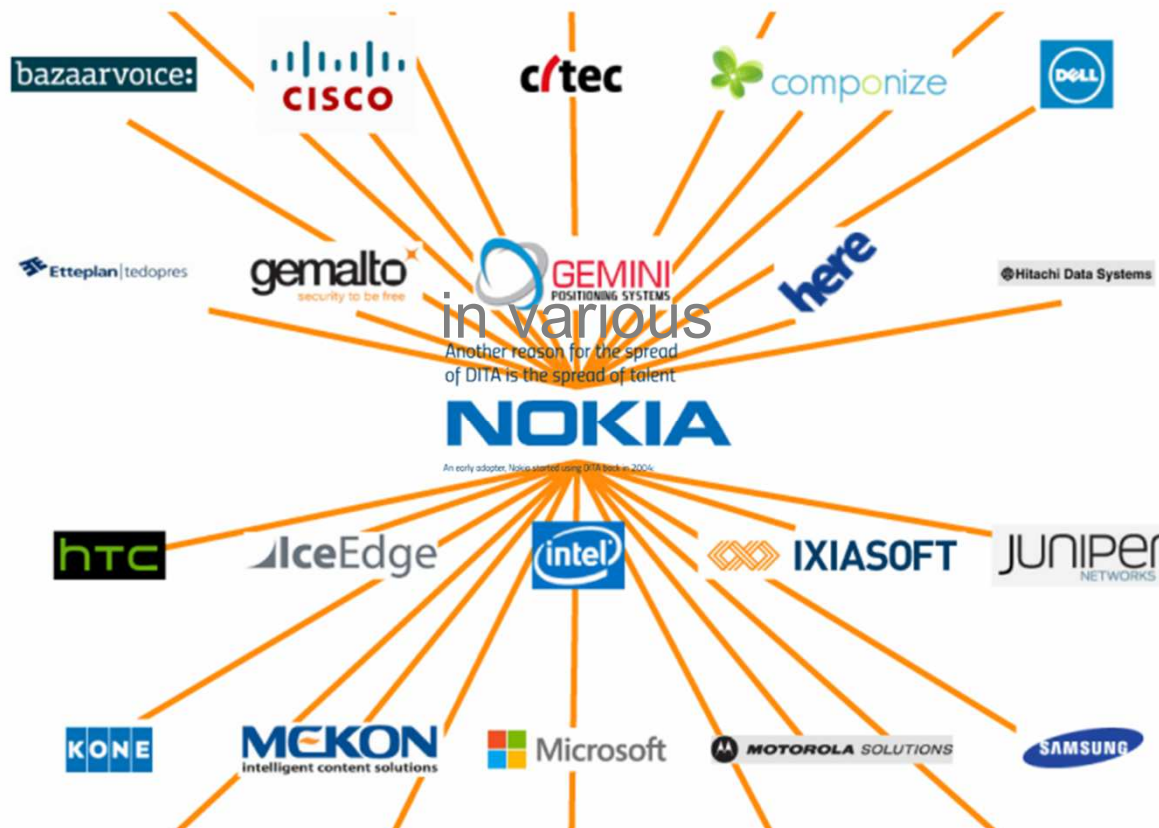
# Becoming One of the “Cool Kids”

- I actually wasn't kidding about this one...
- DITA has been adopted by documentations teams at over 650 companies worldwide



# This in Turn Adds to the Spread of DITA

DITA-using Technical Writers from Nokia have gone on to use DITA at all of these firms:

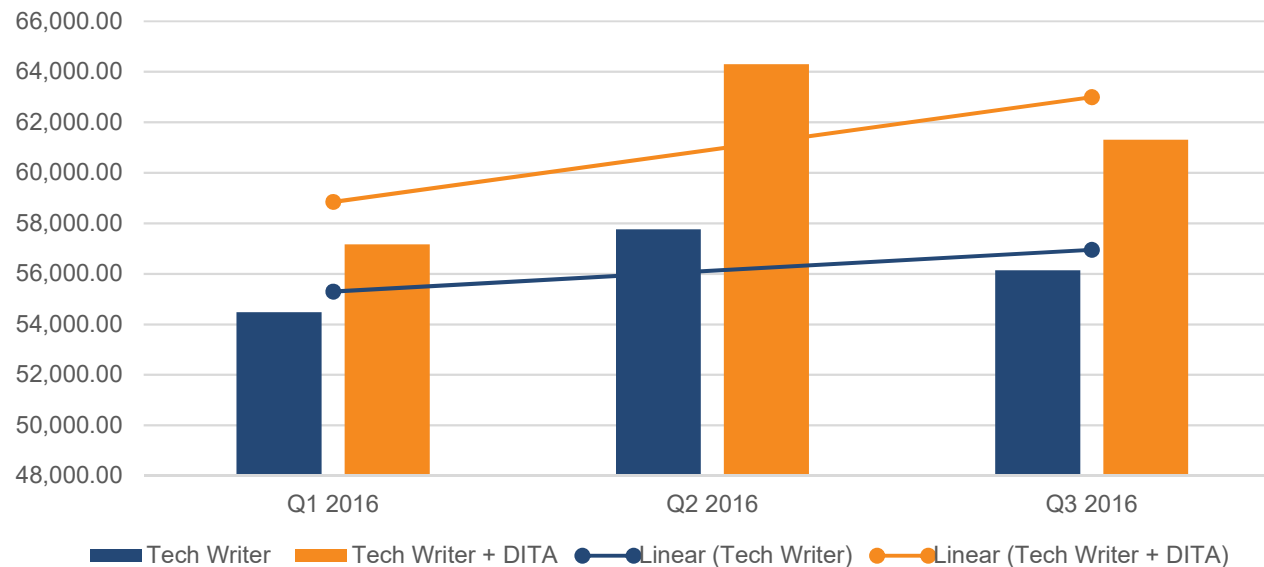


This is one of the ways in which DITA usage spreads from firm to firm

# On a More Personal Level...

- Those who learn DITA often go on to work at other firms where they are sometimes hired specifically for this skill

Avg. Starting Wage for Technical Writers on Indeed.com for 2016



## Some Other Possible Reasons for DITA + CCMS

- Need to produce additional deliverable types (for additional products, additional platforms, etc.)
- Need to produce filtered/custom deliverables
- Need to interchange with other groups using DITA
- Need to move from a custom XML tagset to DITA standard for greater scalability/maintainability outside support
  - In so doing, also get to join a growing community of people working with DITA



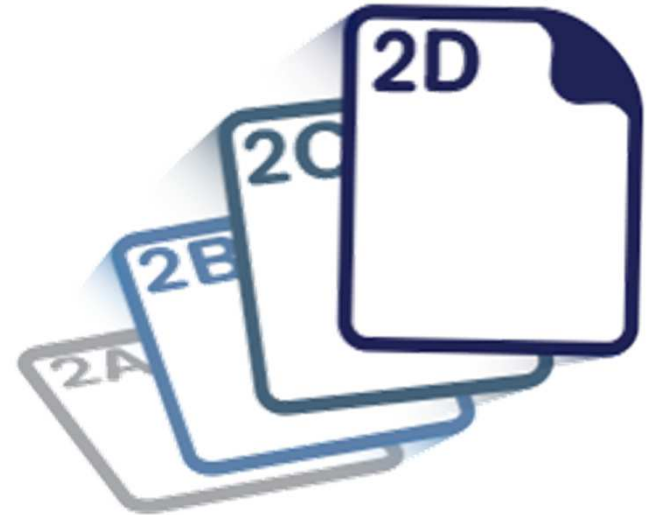


What a CCMS has to offer

# **FOUR THINGS YOU CAN DO WITH DITA AND A CCMS**

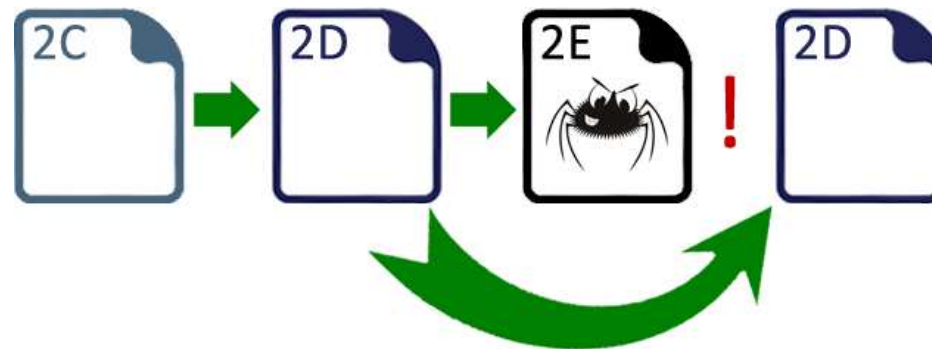
# #1: Versioning of Content

- A feature common to most content management systems, though not to most file systems
- Originally created for software development, versioning allows users to “check out” content for authoring and “check in” any changes



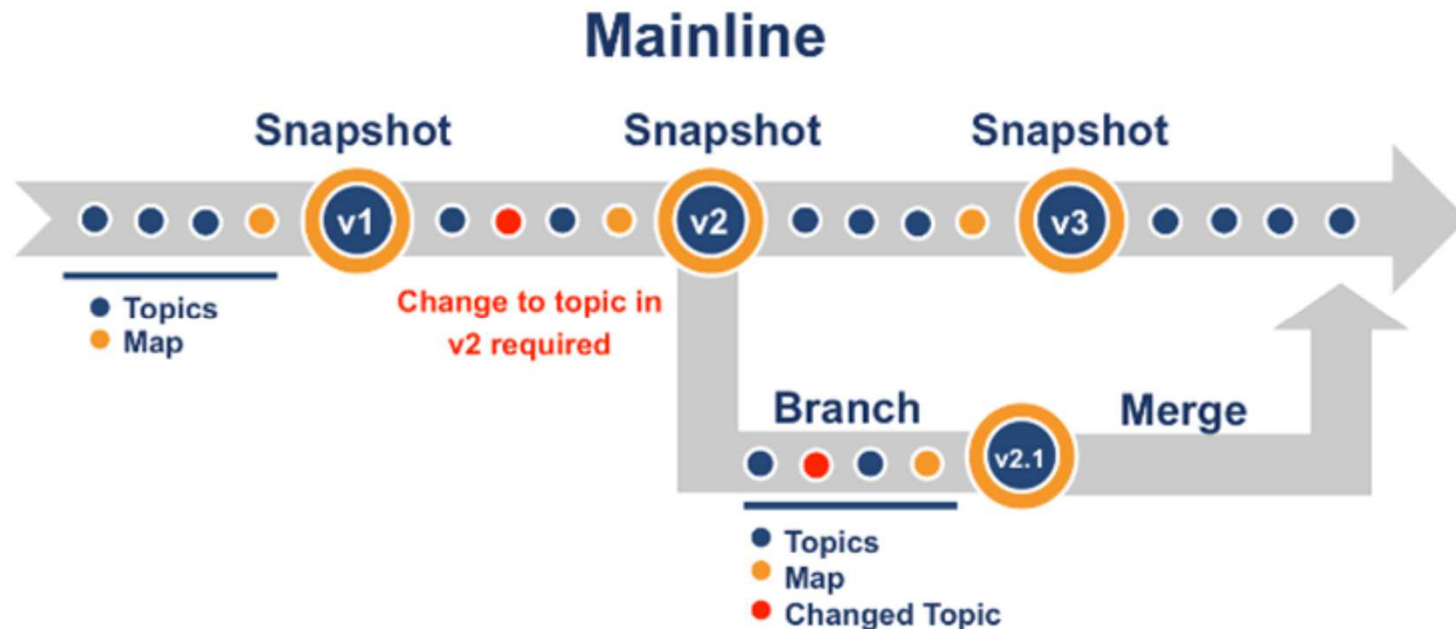
# CCMS and the Versioning of Content

- A typical CCMS will have versioning capabilities that ensures that only a single technical writer can work on a given topic at any one time, and will also register who made which edit and allow for a quick comparison between versions
- Any documentation “bugs” that are discovered in a later version of a topic can be reverted to an earlier version without the problem



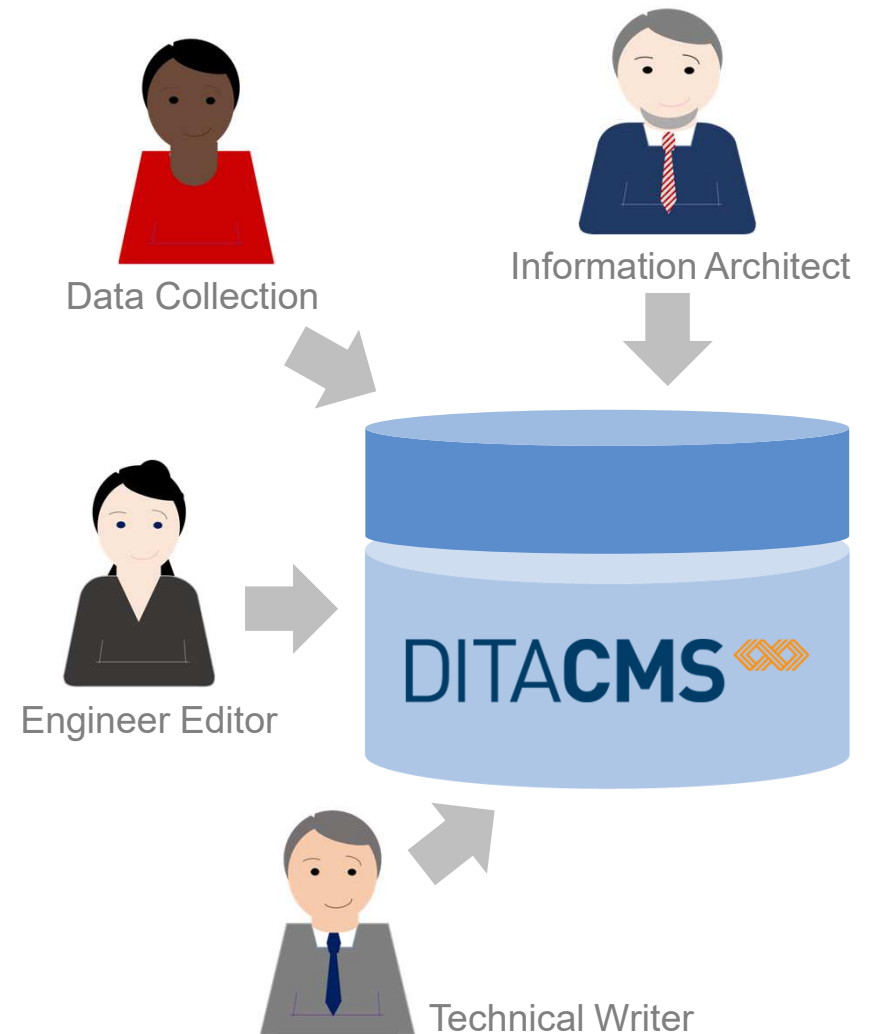
# Branching and Merging of Content

- This allows your writers to write on updates for different product versions in parallel
- IXIASOFT DITA CMS has “snapshot” feature which allows for easy branching and merging of content



## #2: Workflow

- Once a writer has completed writing a topic, can then be routed automatically
- Useful for when you need to have content edited/verified before it goes out the door
- Introduces ways to improve content quality



## Typical Workflow Examples

- An example: a Subject Matter Expert (SME) may be assigned within the system to write content which is then “polished” by a technical writer, and then sent to another SME for approval
- Or: technical writer creates content, is routed to another writer/editor who verifies quality



Engineer  
(SME)



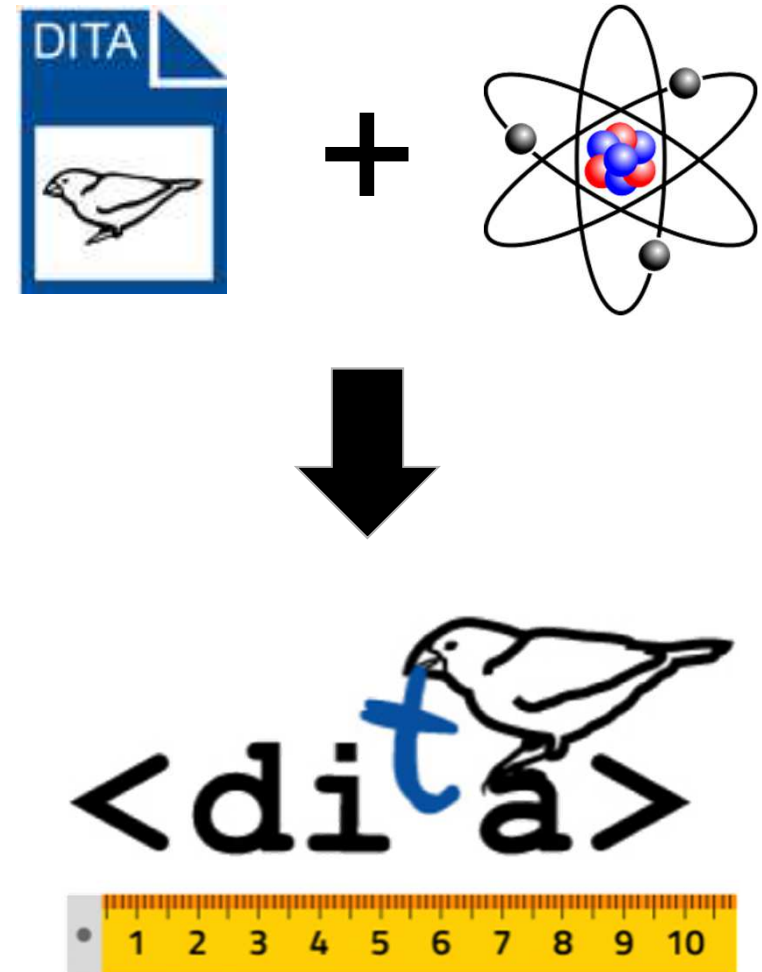
Technical  
Writer



Engineer  
Who Approves  
Content  
Accuracy

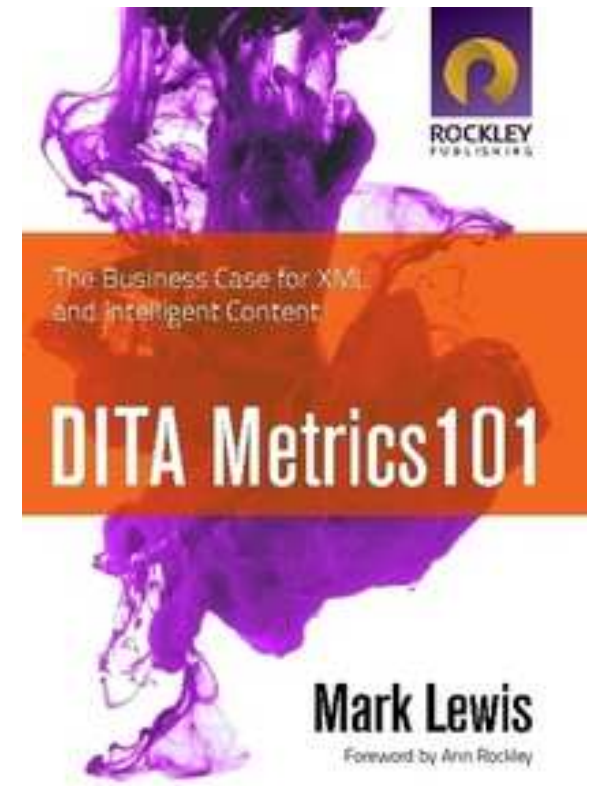
## #3: Measurement

- Having modular content opens up the possibility of measuring content production more accurately
- Not every topic will be the same size, but there will be an average value



# DITA and Return on Investment (ROI)

- This has been a primary focus of much work on DITA metrics
  - Topic-based nature of DITA lends itself to cost-based measures
- The book *DITA Metrics 101* (2013) looks at this aspect almost exclusively, focusing on justifying the cost of investing in DITA + CCMS































# DITA Metrics for Production Purposes

- But not everything is about ROI:
  - What if you have a mature DITA environment and have already established your ROI?
  - Or, are simply looking for ways to use DITA + metrics to measure things that are not cost-related?
- DITA metrics can be used to guide managers, information architects and writers on how to improve their content



# DITA Production Metrics *without* a CCMS








- DITA metrics outside of a CCMS are limited to information contained within the DITA files + file system
  - Can search for text strings within the XML, and also use date/time info from filename... and that's about it

| Name  | Date modified        | Type          |
|---|----------------------|---------------|
|  blowermotorswitchTraveller  | 2014-12-22 11:25 ... | DITA Document |
|  bonnetlock                  | 2014-12-22 11:25 ... | DITA Document |
|  CarBody                     | 2014-12-22 11:25 ... | DITA Document |
|  choke                       | 2014-12-22 11:25 ... | DITA Document |
|  CleaningandCare             | 2014-12-22 11:25 ... | DITA Document |
|  controls                    | 2014-12-22 11:25 ... | DITA Document |
|  controlsinstrumentsswitches | 2014-12-22 11:25 ... | DITA Document |
|  CoolingSystem               | 2014-12-22 11:25 ... | DITA Document |
|  directionindicators         | 2014-12-22 11:25 ... | DITA Document |
|  doorlocks                   | 2014-12-22 11:25 ... | DITA Document |
|  Driving                     | 2014-12-22 11:25 ... | DITA Document |
|  freshairheater              | 2014-12-22 11:25 ... | DITA Document |
|  fuelgauge                   | 2014-12-22 11:25 ... | DITA Document |
|  HeaterControlValveSaloon    | 2014-12-22 11:25 ... | DITA Document |
|  ignitionwarninglight        | 2014-12-22 11:25 ... | DITA Document |
|  instrumentlights          | 2014-12-22 11:25 ... | DITA Document |
|  instrumentsSaloon         | 2014-12-22 11:25 ... | DITA Document |
|  instrumentsTraveller      | 2014-12-22 11:25 ... | DITA Document |
|  Introduction              | 2014-12-22 11:25 ... | DITA Document |
|  JackingandWheelRemoval    | 2014-12-22 11:25 ... | DITA Document |
|  luggagecompartment        | 2014-12-22 11:25 ... | DITA Document |
|  MorrisMiniMinor           | 2014-12-22 11:25 ... | DITA Document |
|  oilpressuregauge          | 2014-12-22 11:25 ... | DITA Document |
|  oilpressurelight          | 2014-12-22 11:25 ... | DITA Document |
|  RecirculatoryHeaterSaloon | 2014-12-22 11:25 ... | DITA Document |
|  Running-inspeeds          | 2014-12-22 11:25 ... | DITA Document |
|  seatbelts                 | 2014-12-22 11:25 ... | DITA Document |
|  ServicingWarranty         | 2014-12-22 11:25 ... | DITA Document |

# DITA Production Metrics *with a CCMS*

- DITA metrics within a CCMS can capture additional information, which depending on the CCMS may include things like:
  - Author information
  - Workflow status
  - How many times a topic has been modified/versioned
  - Topic/map dependencies
- In general, can also expect CCMSs to have “canned” search routines that make the task of mining information from your DITA content easier

Results List page 1 of 2 ( 1,500/2,710 results found )

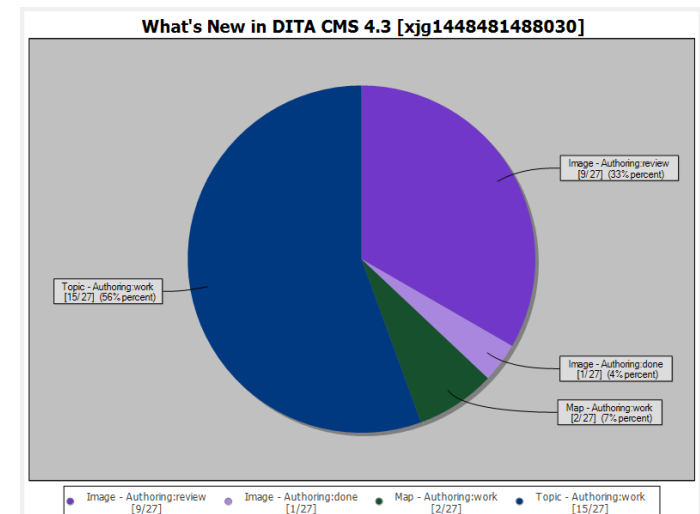
| Title   | Status         | Type           | Last Mod By                    | Last Mod Date       |
|---|----------------|----------------|--------------------------------|---------------------|
|  Integrate current PDF output in Output Generator    | Authoring:work | reference      | Jean-Francois Ameye [jean-f    | 2015-12-18 11:38:57 |
|  Exporting the TEXTML Index Definition document      | Authoring:work | task           | Kathleen Jette [kathleen.jette | 2015-12-18 11:24:53 |
|  Creating and managing index definitions             | Authoring:work | topic          | Kathleen Jette [kathleen.jette | 2015-12-18 11:14:36 |
|  Index Definitions                                   | Authoring:work | referable-cont | Kathleen Jette [kathleen.jette | 2015-12-18 11:12:04 |
|  CHM output enhancements                             | Authoring:work | reference      | Boris Aguilar [boris.aguilar@  | 2015-12-18 10:57:12 |
|  Attach EFS (Resource) files to PDF Outputs          | Authoring:work | reference      | Boris Aguilar [boris.aguilar@  | 2015-12-18 10:57:12 |
|  "Insert Variable" action in oXygen - In phase 1 mar | Authoring:work | reference      | Boris Aguilar [boris.aguilar@  | 2015-12-18 10:57:11 |

# Agile and DITA Metrics

- At Scrum meetings doc manager can report on topics assigned to their group and report on how “done” they are
- This is typically only feasible within a CCMS containing workflow metadata

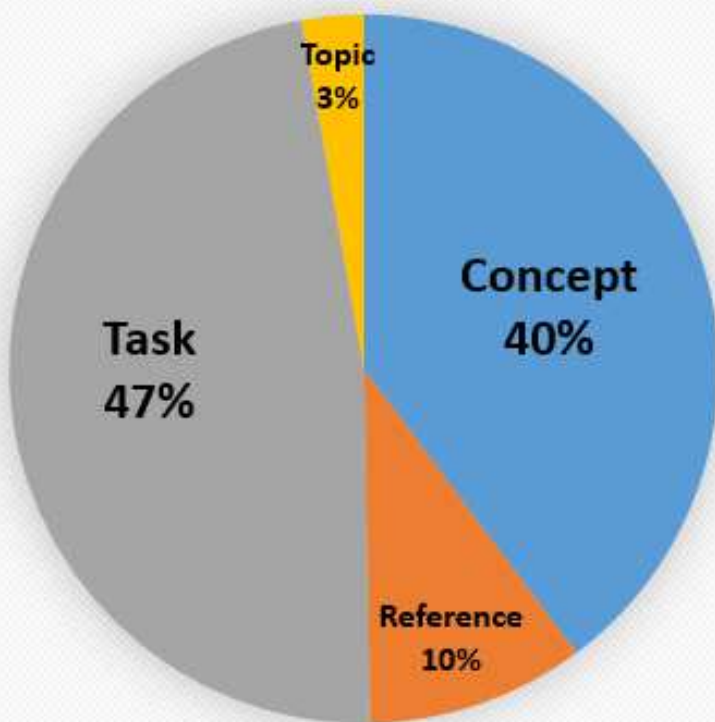
*“Our project managers could track progress of documentation deliverables within our DITA-based CCMS on a daily basis.”*

- Jason Owen



# Content Types within All Documents Over a Year

IXIASOFT Documentation DITA Topic  
Breakdown 2015

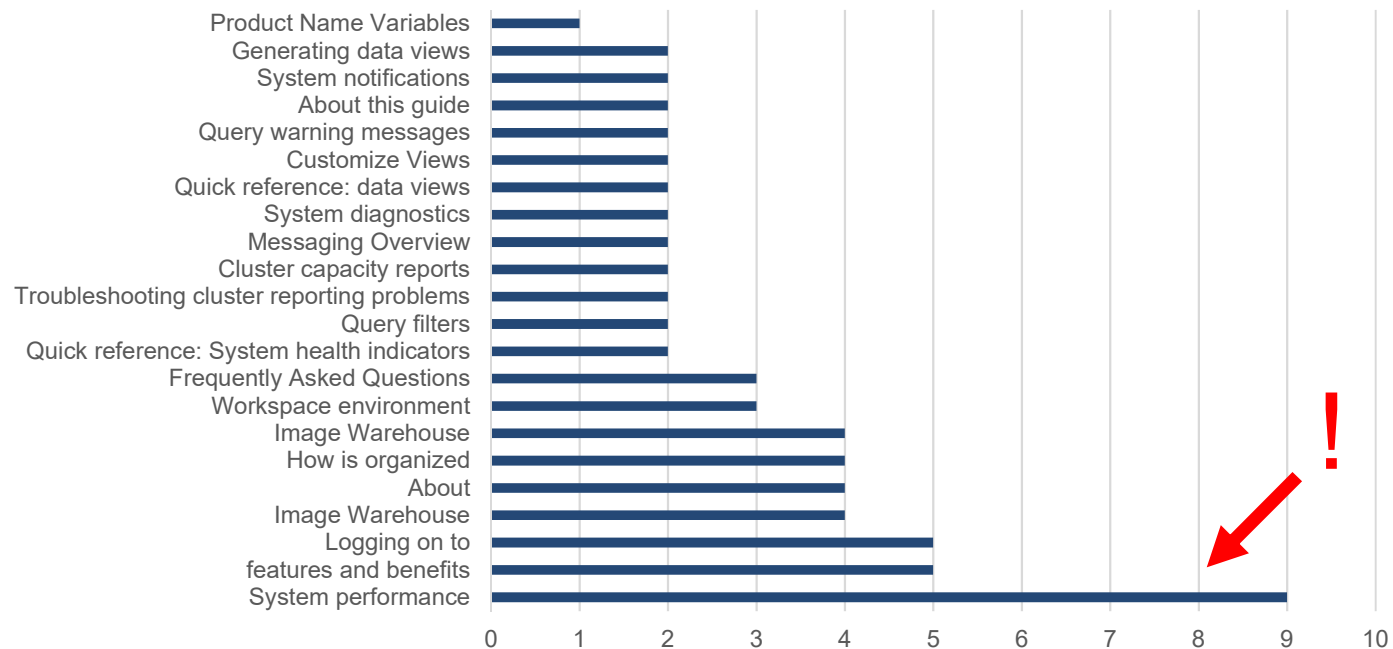


n = 1307

- This chart looks at the DITA topic breakdown for all documentation produced by IXIASOFT in 2015
- Documentation consists of User/Admin Guides for our DITA CMS and TEXTML software
- Good ratio of concept to task topics
- When I showed this to our Lead Tech Doc person, she immediately wanted to investigate the 3% of generic topic types
  - Nice practical example of how DITA metrics can improve quality!

# Can Easily Find Content that is Revised a Lot

Number of Revisions per Topic in the Thunderbird User Guide



- A topic that is often revised might be a sign that it needs to be investigated; perhaps too much tweaking is happening, or there is SME disagreement, or it should be multiple topics instead of just one

## DITA Reuse Metrics

- Arguably the most influential article on this topic is Bill Hackos' "Reuse of DITA Topics? What is the Best Metric to Measure the Success of Your Reuse of DITA Topics?" (<http://ow.ly/X7mzM>)



## DITA Reuse Metrics

- Bill Hackos proposed “Percent Repository Words Reused in Context” (PRWRC) where:

$$\text{PRWRC} = (\text{Words in All Produced Content} - \text{Words in the Repository}) / (\text{Words in the Repository})$$

- From his example:

- Document 1 – 25,413 words
- Document 2 – 23,069 words
- Document 3 – 26,366 words
- Total number of words in the produced documents – 74,848 words
- Total number of words in the repository – 40,060 words

$$\text{PRWRC} = (74,848 - 40,060) / 40,060 = 87\%$$

## Example Based on IXIASOFT DITA Documents

- Flesh tool can count words in publications
- Free tool called “Count Anything” can count total words in DITA files in your repository

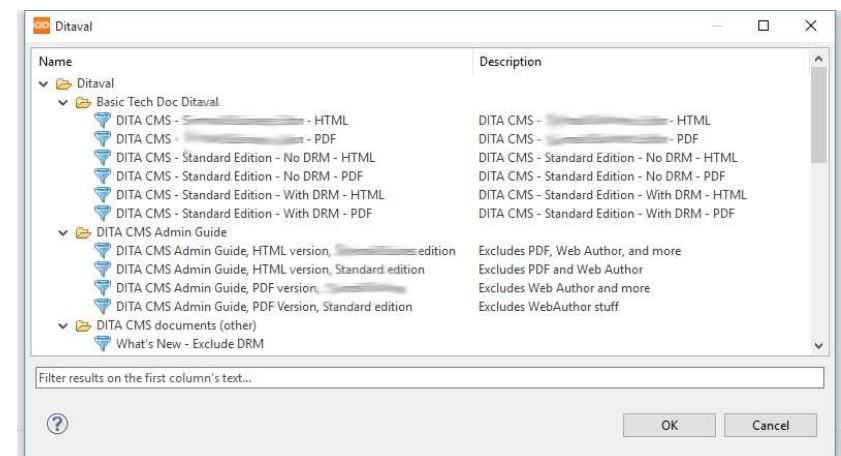
Based on 2015 numbers:

- Total number of words in the repository: 268,663
- Words in All Produced Content: 623,078
- $PRWRC = (623,078^* - 268,663) / 268,663$
- $PRWRC = 354,415 / 268,663 = \mathbf{132\%}$

\* Numbers based on available data

# How is a 100%+ Value Possible?

- Easy: DITAVAL
- Though DITAVAL is not mentioned in original article, Bill Hackos does talk about 100%+ values being entirely possible
- We have number of publications that are created based on a series of DITAVAL value, as much as 21 per bookmap
- Percentage value is the lower bound on our content reuse; could be considerably more but is *no less than 132%*

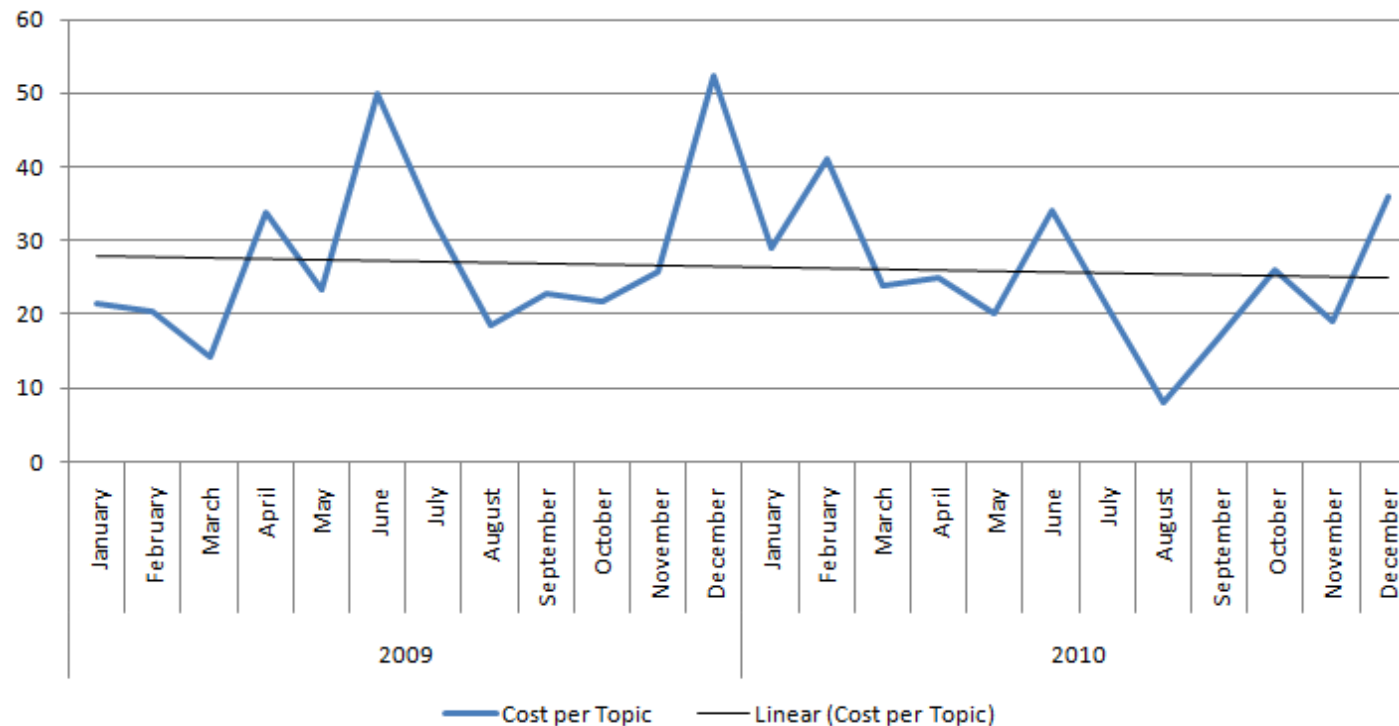


## Example of How CCMS Data Can Be Mixed with Outside Data: Cost-per-Topic

- Idea came from a CIDM *Best Practices* article by Mike Eleder (“The Illusive, Writing Productivity Metric: Making Unit Cost a Competitive Advantage”)
- Basically:
$$\text{cost per topic} = \frac{\text{monthly tech writer team cost}}{\text{topics produced monthly}}$$
- This is a unit cost measure, and it produces the monthly average cost for producing topics
  - Can estimate the cost of future work based on result

# Example Chart Showing COST-Per-Topic

**Cost per Topic for 2009-2010 with Trendline**

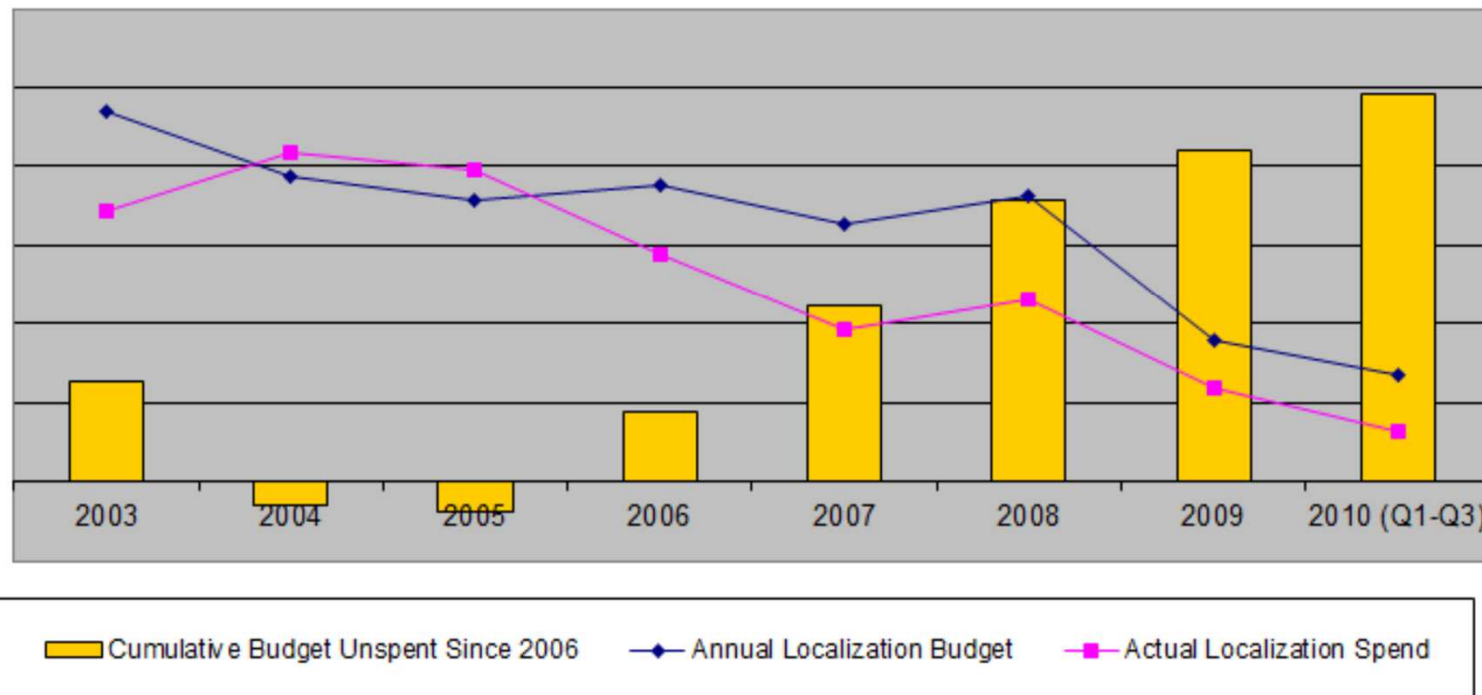


- The trend line records the *average* cost-per-topic over the time measured
- This shows an overall downward trend: ideally this is what you would expect over time
- Same cost measures can be applied to published documents, localized content and more

# An Example of This Using Localization Spend

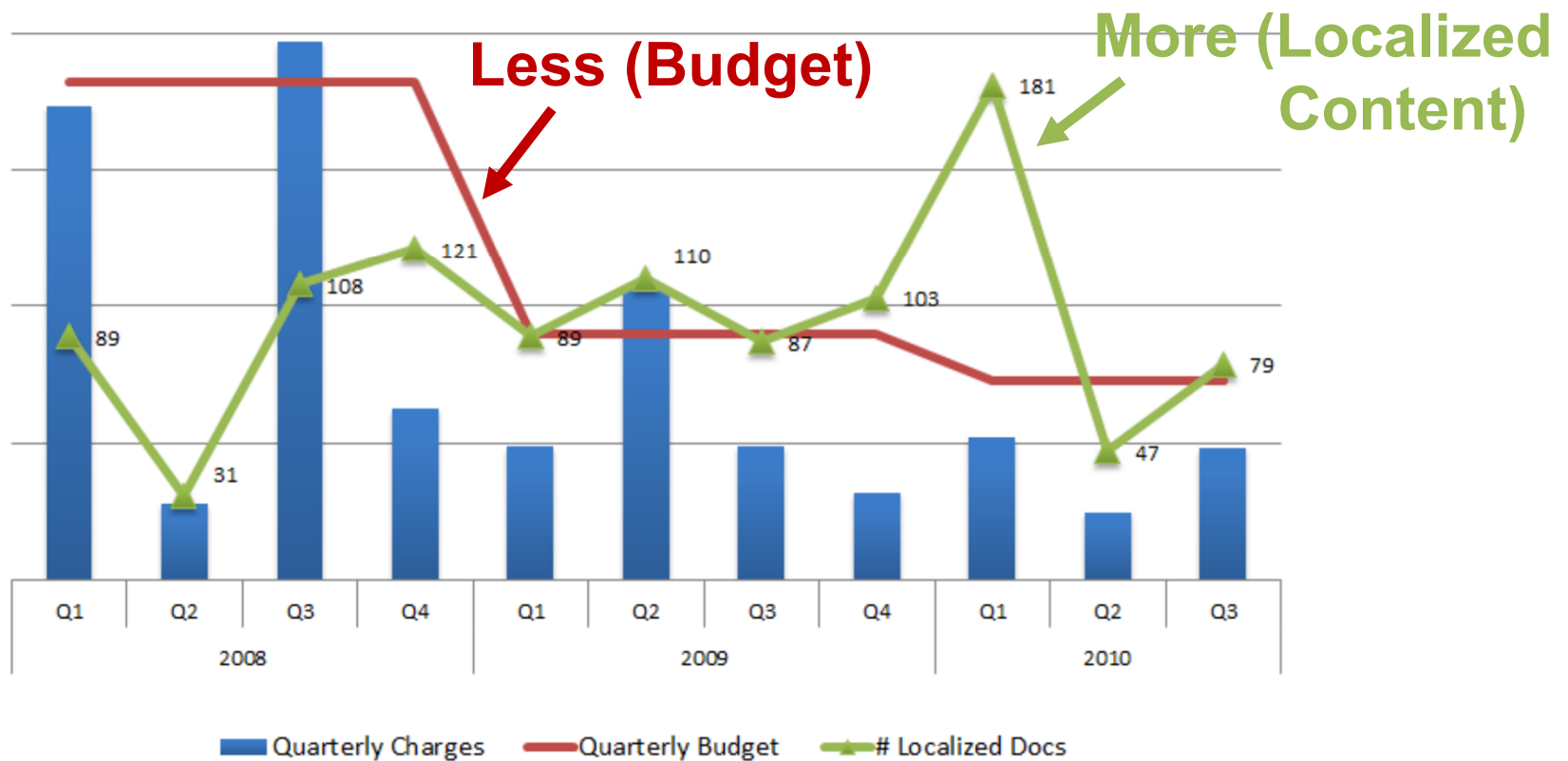
- ROI for move to DITA + CCMS justified in terms of unspent localization budget due to efficiencies (okay, sometimes it *can* be about ROI) ;)

**Annual Localization Budget and Spent 2005-2009 with Cumulative Budget Unspent/Saved Since 2006**



# Doing More with Less

Quarterly Localization Run-rate vs.  
Number of Localized Docs Published  
2008-Q3 2010



## #4: Localization

- Reduced localization costs come from DITA reuse
  - Content reuse in English = content reuse in other languages
- A CCMS provides additional benefits:
  - Improved packaging options
  - QA as part of localization workflow
  - Automation

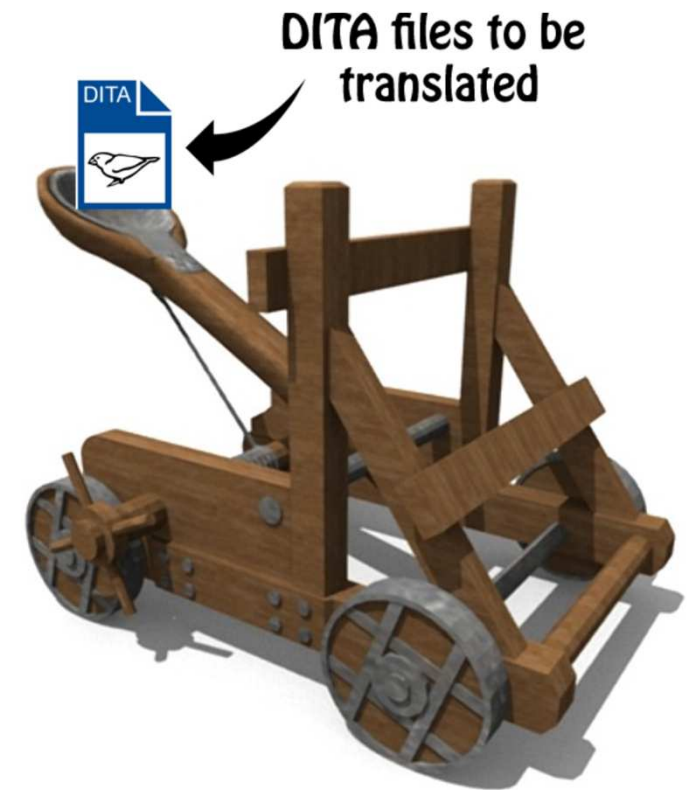
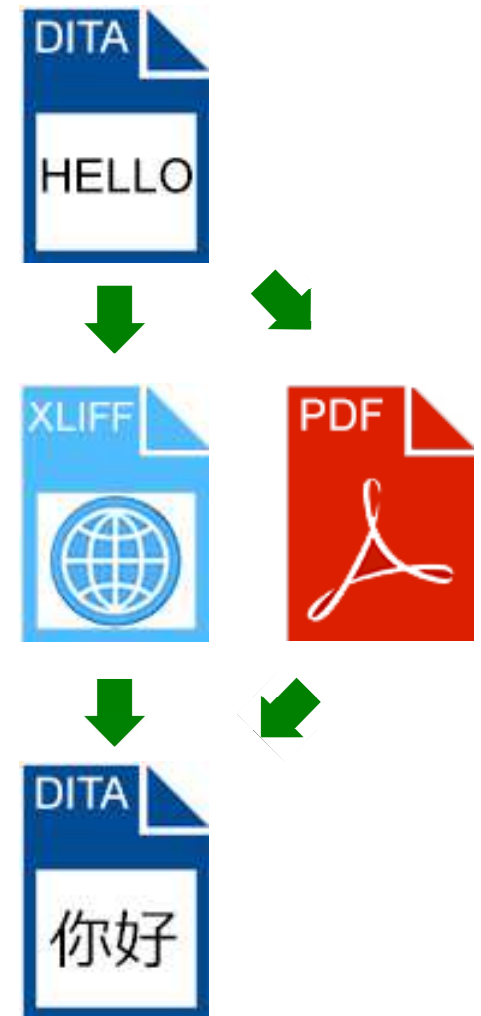


Illustration of how DITA localization without a CCMS typically works

## Improved Packaging Options

- In addition to sending content in DITA, also possible for CCMS to transform content into XLIFF
- Can also automate inclusion of PDF for context
- Also possible to send individual topics/changes for translation



## QA as Part of Localization Workflow

- A good DITA CCMS will include localization QA as part of the workflow for content
- Enables checking and verification of content prior to publication
- Can easily route “bad” localized content back to localization service provider



## Automation

- A capable DITA CCMS will enable a firm to automate localization processes
- Some of IXIASOFT's customers send individual topics out for localization while rest of document being worked on
  - New and changed content can be merged, enabling much faster turnaround times

Writing: Complete



Localization: Complete

## Other Things a DITA CCMS Can Do for You

- Provides a mechanism for SMEs to contribute directly to content
- Looking to move away from PDF? Can use DITA + CCMS with dynamic publication engines to create HTML5-based publications
- Metadata embedded in DITA maps and topics can aid with search engine optimization

## Upfront Costs vs. Opportunity Costs

- Yes, there are often significant costs associated with purchasing a DITA CCMS:
  - Conversion costs of legacy content
  - Change management
  - Cost of CCMS software
  - May involve separate purchase of repository software for content
- But can your organization support ongoing costs of current toolchain?
  - This is an opportunity for change and improvement...



OPPORTUNITY  
**COST**

# DITA + CCMS = The Road Ahead



- DITA + CCMS should be thought of as an opportunity to improve the quality of your content and your documentation processes
- Think of where you need to take your documentation team and what your users will expect of you five years from now

## QA

- Blog: [www.ixiasoft.com/en/news-and-events/blog](http://www.ixiasoft.com/en/news-and-events/blog)
- Twitter: @IXIASOFT (and @KeithIXIASOFT)
- IXIASOFT DITA CMS Users LinkedIn group: [www.linkedin.com/groups?gid=3820030](http://www.linkedin.com/groups?gid=3820030)
- OASIS DITA Adoption Committee articles
- Member of OASIS DITA Technical Committee