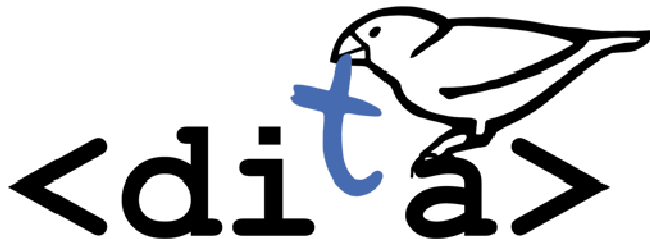




l10n

+



Localization and DITA: What You Need to Know

Keith Schengili-Roberts, IXIASOFT
DITA Specialist

Agenda

- Introduction
- Quick overview of key features of DITA
- Where DITA adoption is today
- DITA Localization best practices
- How and why our clients use DITA for localization
- Q/A

Who's This Guy?



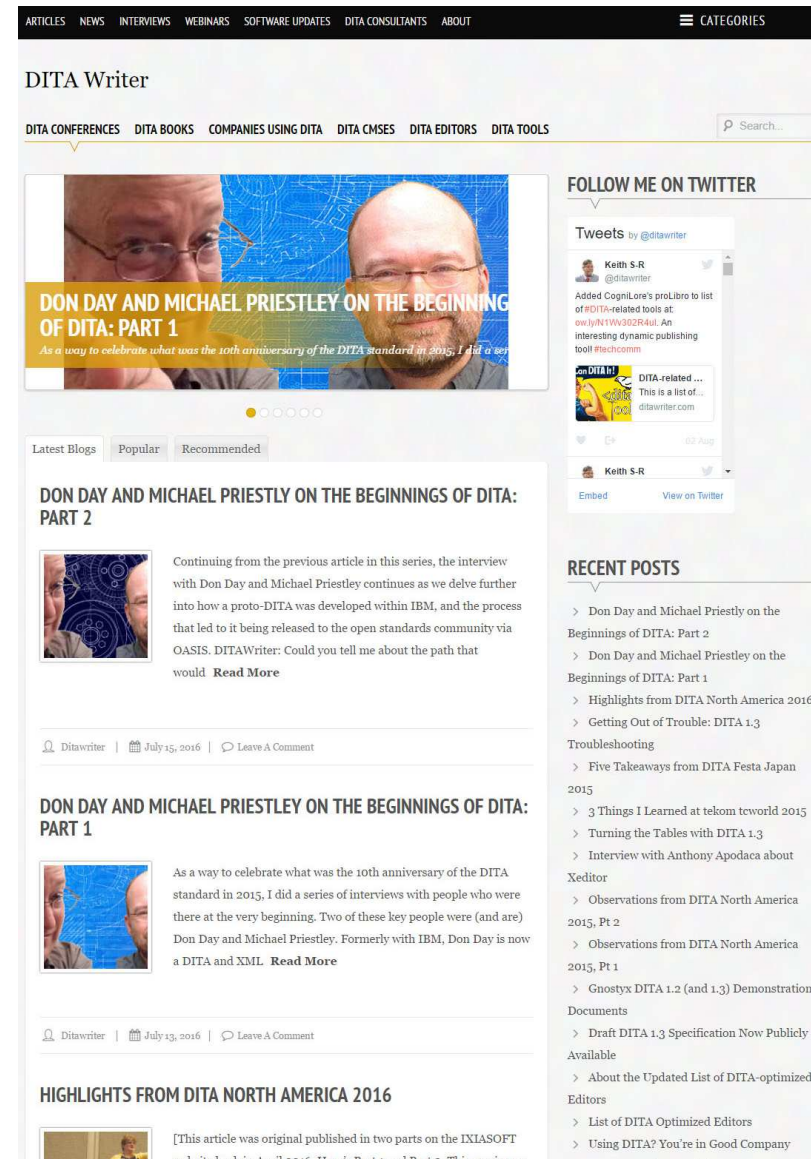
Keith Schengili-Roberts,
IXIASOFT DITA Evangelist and
Market Researcher

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
- Formerly Documentation & Localization Manager at AMD

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 CMSes
 - 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 site title "DITA Writer" is displayed, followed by a secondary navigation bar with links for DITA CONFERENCES, DITA BOOKS, COMPANIES USING DITA, DITA CMSes, DITA EDITORS, and DITA TOOLS. The main content area features a large featured image with the text "DON DAY AND MICHAEL PRIESTLEY ON THE BEGINNING OF DITA: PART 1" and a sub-headline "As a way to celebrate what was the 10th anniversary of the DITA standard in 2015, I did a series of interviews with people who were there at the very beginning. Two of these key people were (and are) Don Day and Michael Priestley. Formerly with IBM, Don Day is now a DITA and XML." Below this, there are sections for "Latest Blogs" (with tabs for Popular and Recommended), "DON DAY AND MICHAEL PRIESTLEY ON THE BEGINNINGS OF DITA: PART 2", and "DON DAY AND MICHAEL PRIESTLEY ON THE BEGINNINGS OF DITA: PART 1". The "PART 1" article includes a small image and a "Read More" link. At the bottom, there is a section for "HIGHLIGHTS FROM DITA NORTH AMERICA 2016" with a "Read More" link. On the right side, there is a "FOLLOW ME ON TWITTER" section with a tweet from Keith S-R (@ditawriter) and a "RECENT POSTS" section listing various articles.

My Own Experiences with L10N

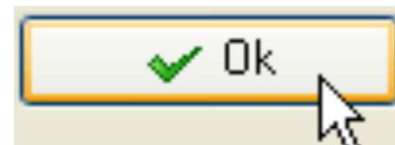
- Was in charge of AMD's localization budget spend for technical communications
- Helped rationalize localization spend across the company
- Key decision in choice of DITA CCMS was that it used Unicode (to get rid of code page problems)
- Cost of shift to DITA was underwritten largely by expected localization cost savings
- Continual pressure to do more with less

Localization Gone Wrong

Examples I've personally encountered:

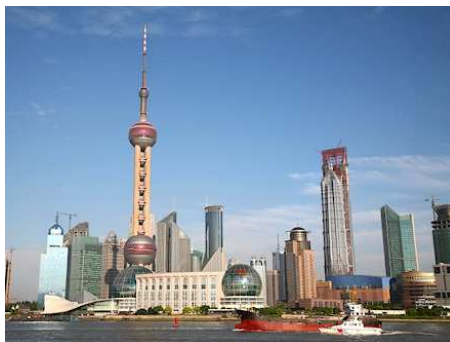
- “Senior Thermal Engineer” was translated as: “Senior Keep-Warm Engineer”
- “Click OK to finish.” translated to: “Click OK to [become] Finnish.”

Bad translation affects the perception of brand quality



How I Learned the Difference Between Localization and Translation

- When we were testing localization quality on engineering documentation at AMD, our Chinese immigrant engineers in Markham made lots of corrections
 - This made us question the quality of the work being done
- We then asked colleagues in the Shanghai office to double check the work
 - We discovered that the Chinese speaking engineers in Markham were using long out of date engineering terms, and our Shanghai colleagues confirmed that LSP was properly localizing the terms



≠



The appeal of DITA

QUICK OVERVIEW OF KEY FEATURES OF DITA

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 appendices, 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 

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 for clients
 - Also removes DTP charges from LSPs...

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

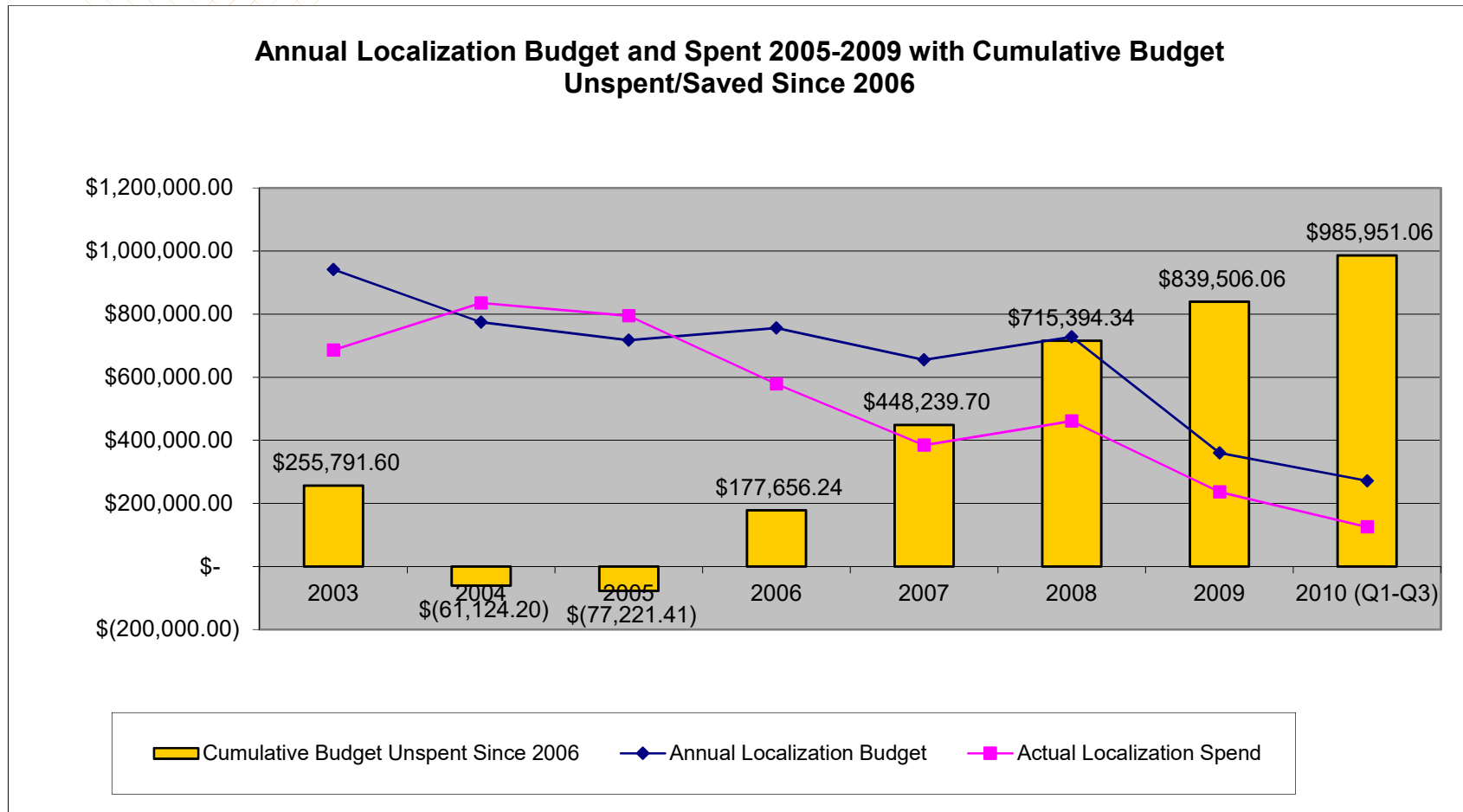


Reduced Localization Costs with DITA

- Content reuse in English = localization savings
- If many target languages, ROI argument for move to DITA is easier



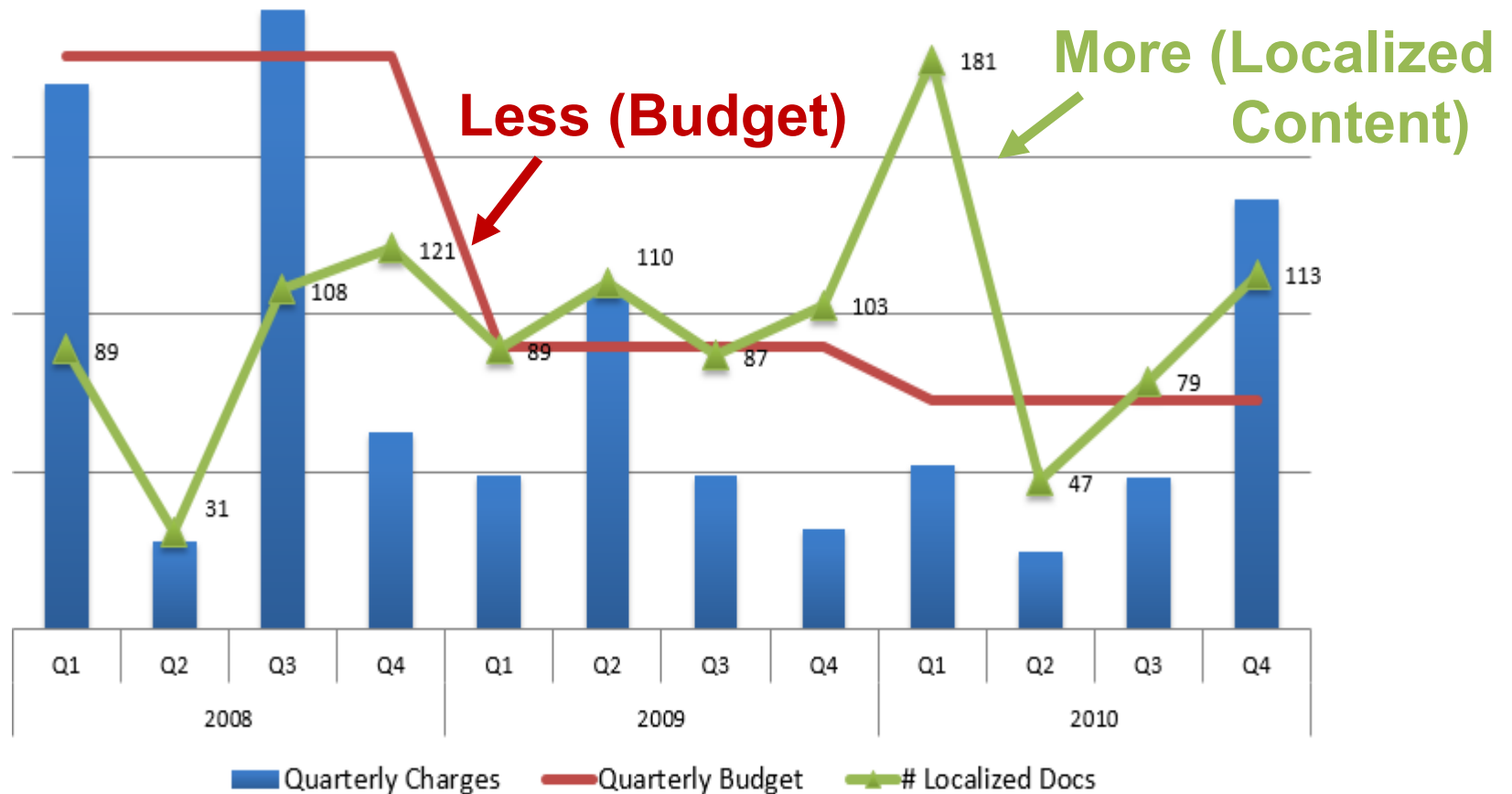
An Example of Cumulative Localization Cost Savings



- This is the type of cost savings that is used to convince C-level execs of cost-savings to expect when switching to DITA

Doing More with Less

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



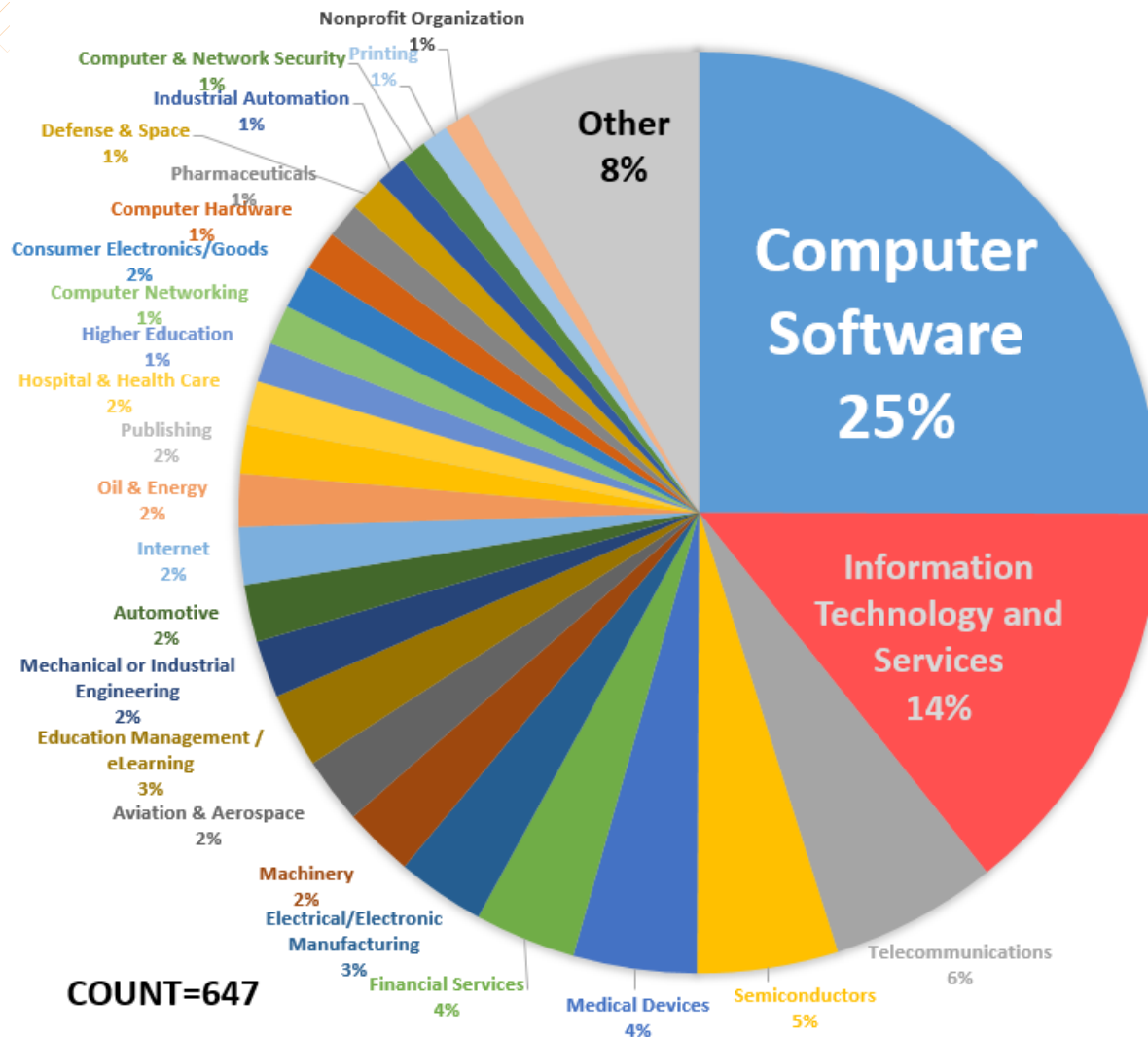


The Companies and People Using DITA

WHERE DITA ADOPTION IS TODAY

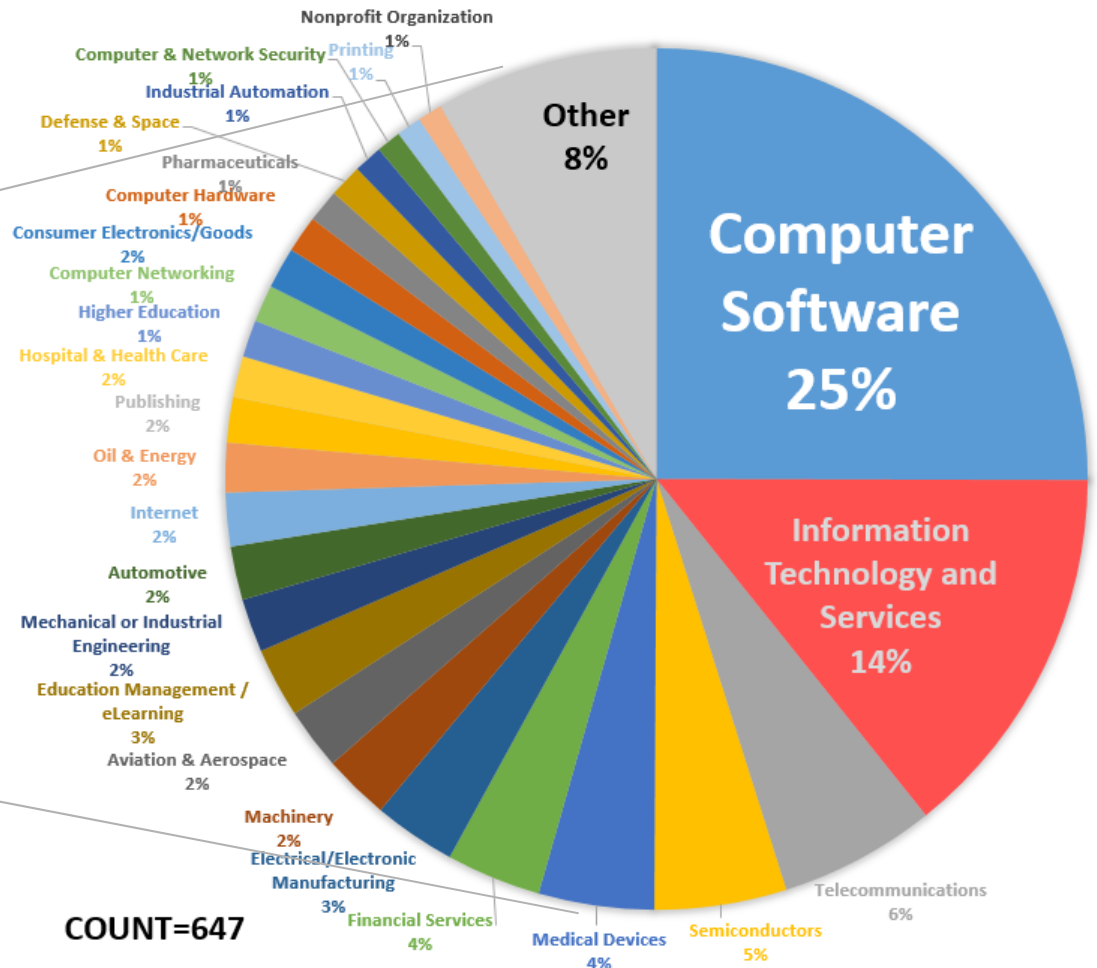
Current DITA Usage by Industry Sector

DITA USAGE BY INDUSTRY SECTOR, Q3 2016

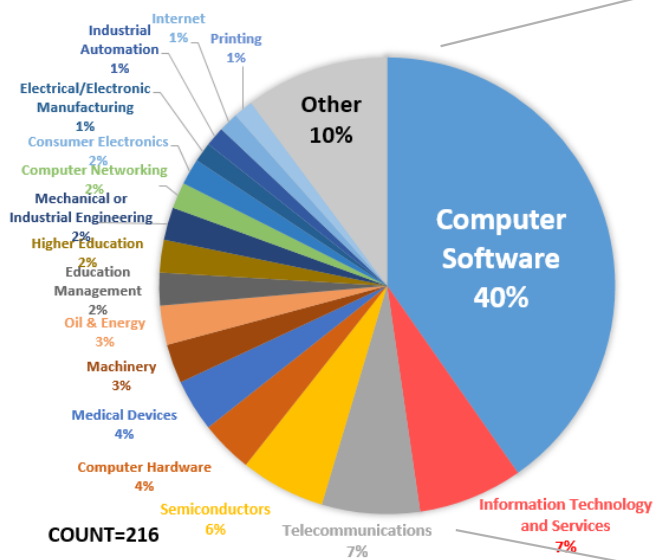


Sector Growth Over Time

DITA USAGE BY INDUSTRY SECTOR, Q3 2016



DITA USAGE BY INDUSTRY SECTOR, Q1 2012

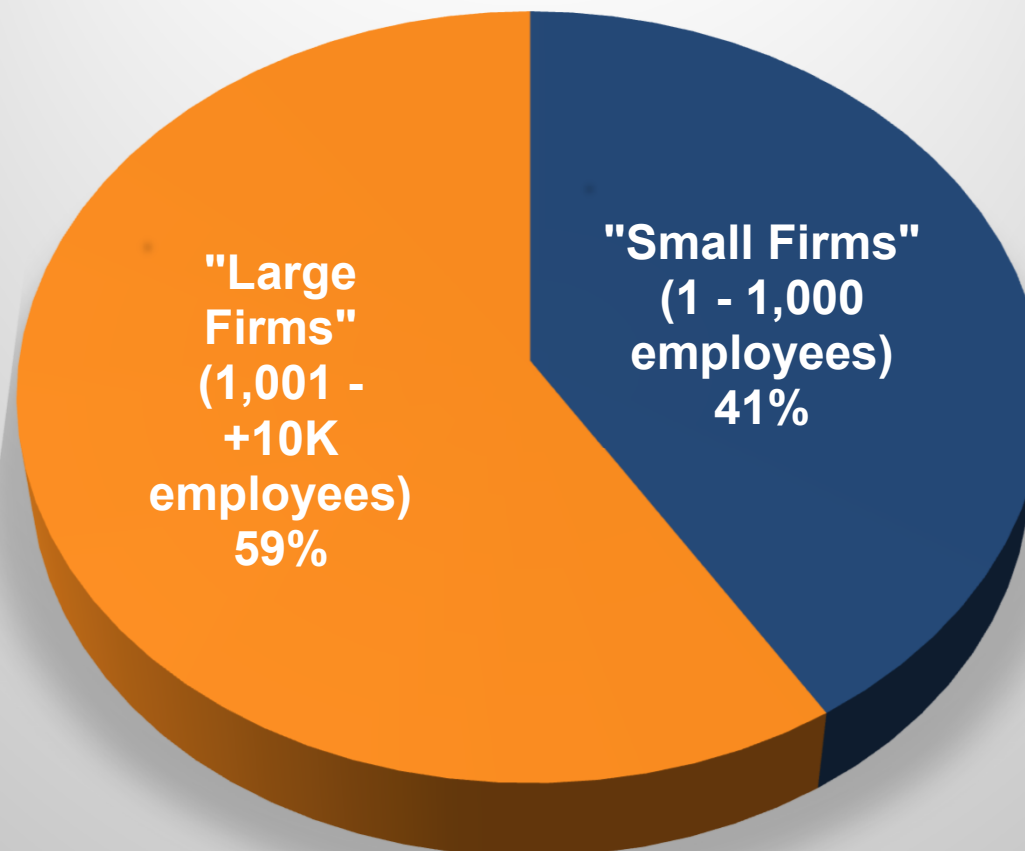


COUNT=647

COUNT=216

More Large Firms than Small Use DITA

**Size of Companies Using DITA
(By Number of employees) as of Q3 2016**



Count: 631 Firms

Sector Breakdown of Firms Using DITA

100% of the top 10 software companies are using DITA.

1. Microsoft
2. Oracle
3. IBM
4. SAP
5. Symantec
6. EMC
7. Hewlett-Packard
8. VMWare
9. CA Technologies
10. Salesforce.com

100%

Ranking data from Investopedia

At least 60% of the top 10 medical device firms globally are using DITA.

1. **Johnson & Johnson**
2. **General Electric**
3. **Medtronic**
4. **Siemens AG**
5. Baxter International
6. Fresenius Medical Care
7. **Philips**
8. Cardinal Health
9. **Novartis AG**
10. Covidien

60%

Top 10 info based on info from MDDI for 2014

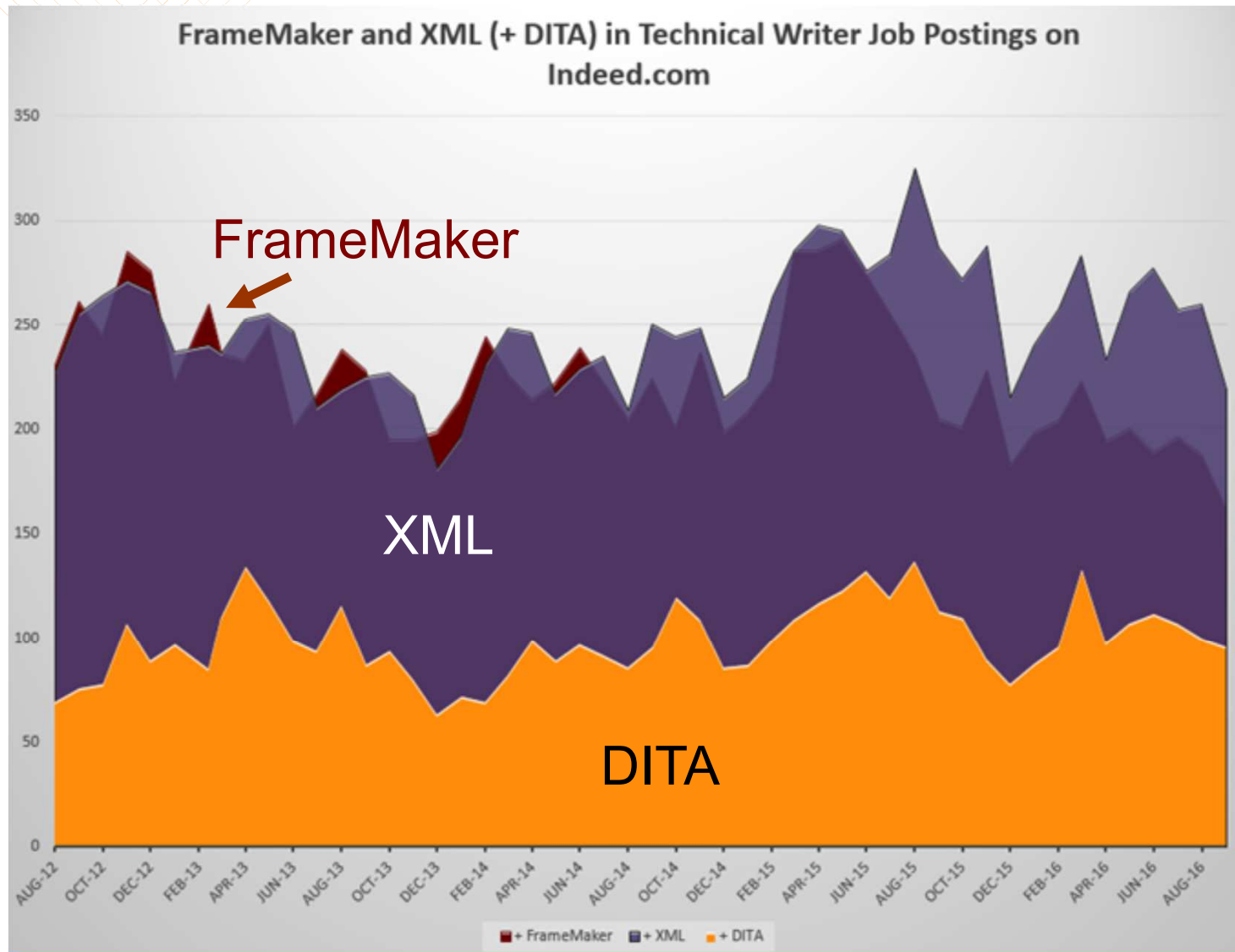
Sector Breakdown of Firms Using DITA

At least **70%** of the top 20 semiconductor companies globally are using DITA.

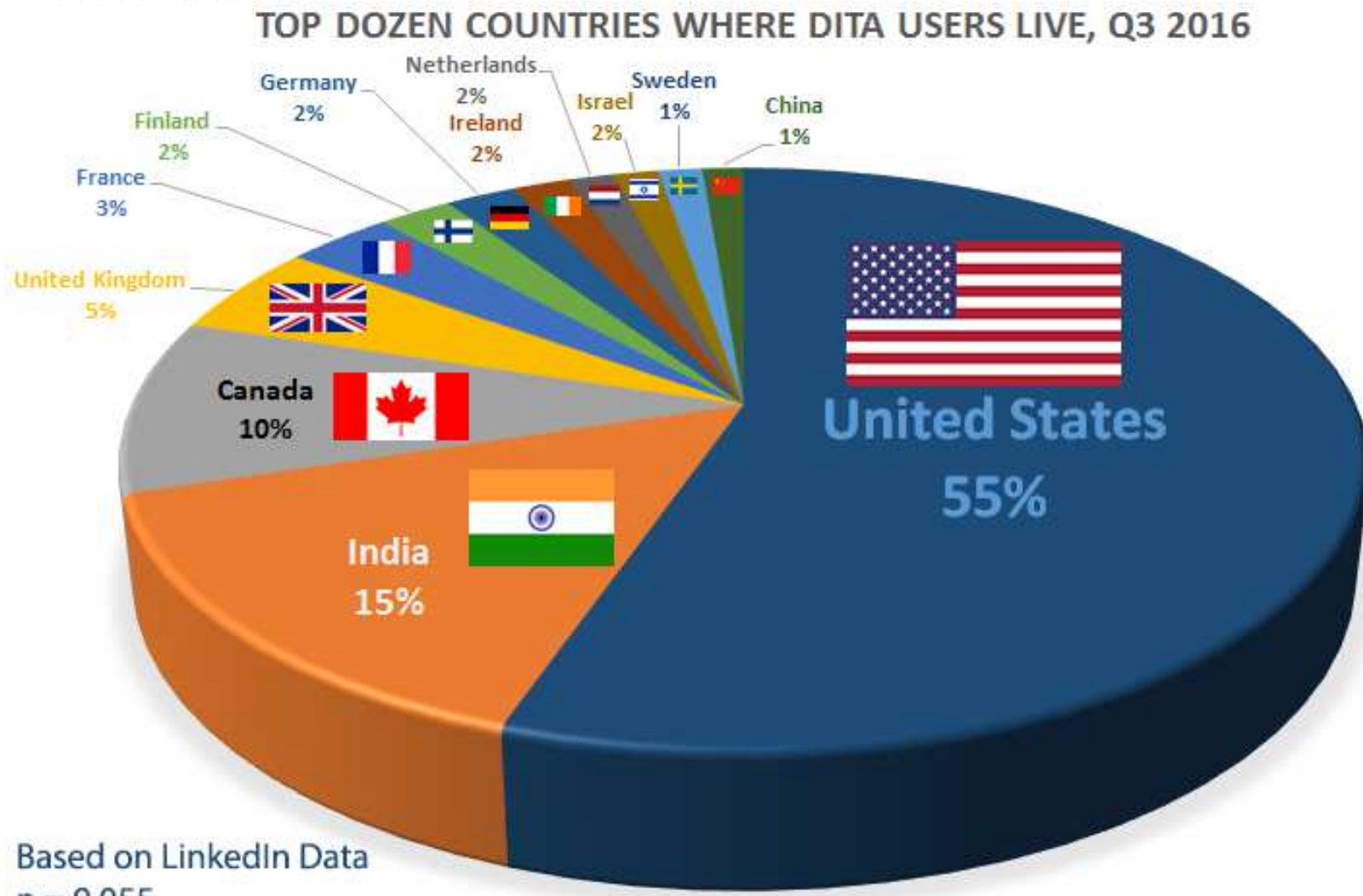
- | | |
|--------------------|----------------------------|
| 1. Intel | 11. MediaTek |
| 2. Samsung | 12. Infineon |
| 3. TSMC | 13. STMicro |
| 4. Broadcom | 14. Apple |
| 5. Qualcomm | 15. GlobalFoundries |
| 6. SK Hynix | 16. Renesas |
| 7. TI | 17. Nvidia |
| 8. Micron | 18. Sony |
| 9. Toshiba | 19. UMC |
| 10. NXP | 20. AMD |

*Top 20 List from
IC Insights for 1H16*

XML Has Eclipsed FrameMaker in US Job Posts



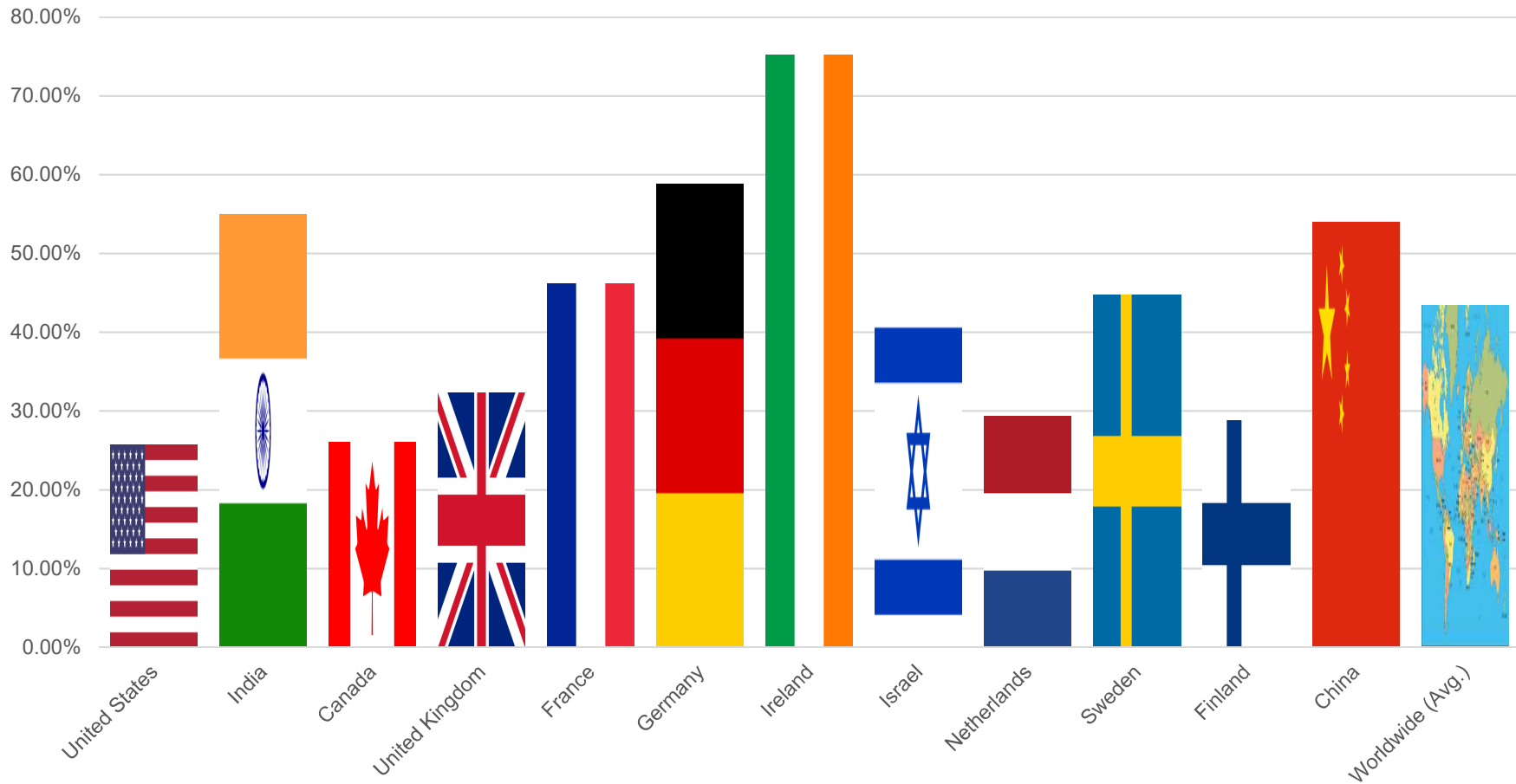
Where DITA Users are Located



Based on LinkedIn Data
n = 9,055

Percentage Growth of DITA in Past Two Years

Percentage Growth of People Claiming DITA XML Experience on LinkedIn per Country Between 2014 and 2016



DITA Localization Volume from a Major LSP #1

- 15% of their customers use DITA, tends to be their larger accounts
- DITA localization work represents 20-25% of what they do

- Q: Are clients coming to you specifically because of your firm's expertise with XML and DITA content?
- A: Yes, most definitely!



*Shamus Dermody
XML International
(at Booth 08)*

DITA Localization Volume from a Major LSP #2

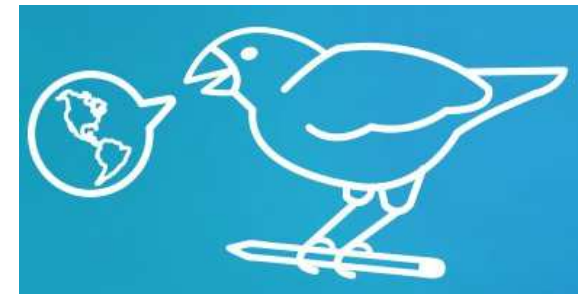
- “We are seeing more customers adapting DITA. The pace is steady. I would say the percentage of our customers using DITA is about 40%.”
- Q: Are clients coming to you specifically because of your firm's expertise with XML and DITA content?
- A: “Yes, we have clients coming to CSOFT because of our XML and DITA experience. We have engineering resources that can work in creative ways to process the XML files to ensure the translation process is smooth and that the final translated files are trouble free for importing into the client system.”



*Carl Yao
Executive Vice President,
Global Strategy
CSOFT*

DITA Localization Volume from a Major LSP #3

- Over 50% of their customers use DITA, with over 75% of that comprised of new customers that started with them since January 2015
- DITA localization work represents 50% of what they do, and this continues to grow
- Q: Are clients coming to you specifically because of your firm's expertise with XML and DITA content?
- A: “Yes, over 95% of them, since this is our main focus. They are referred to us by DITA consultants, DITA software suppliers, other DITA customers, and DITA info architects moving to new companies.”



*Dominique Trouche
CEO of WhP*

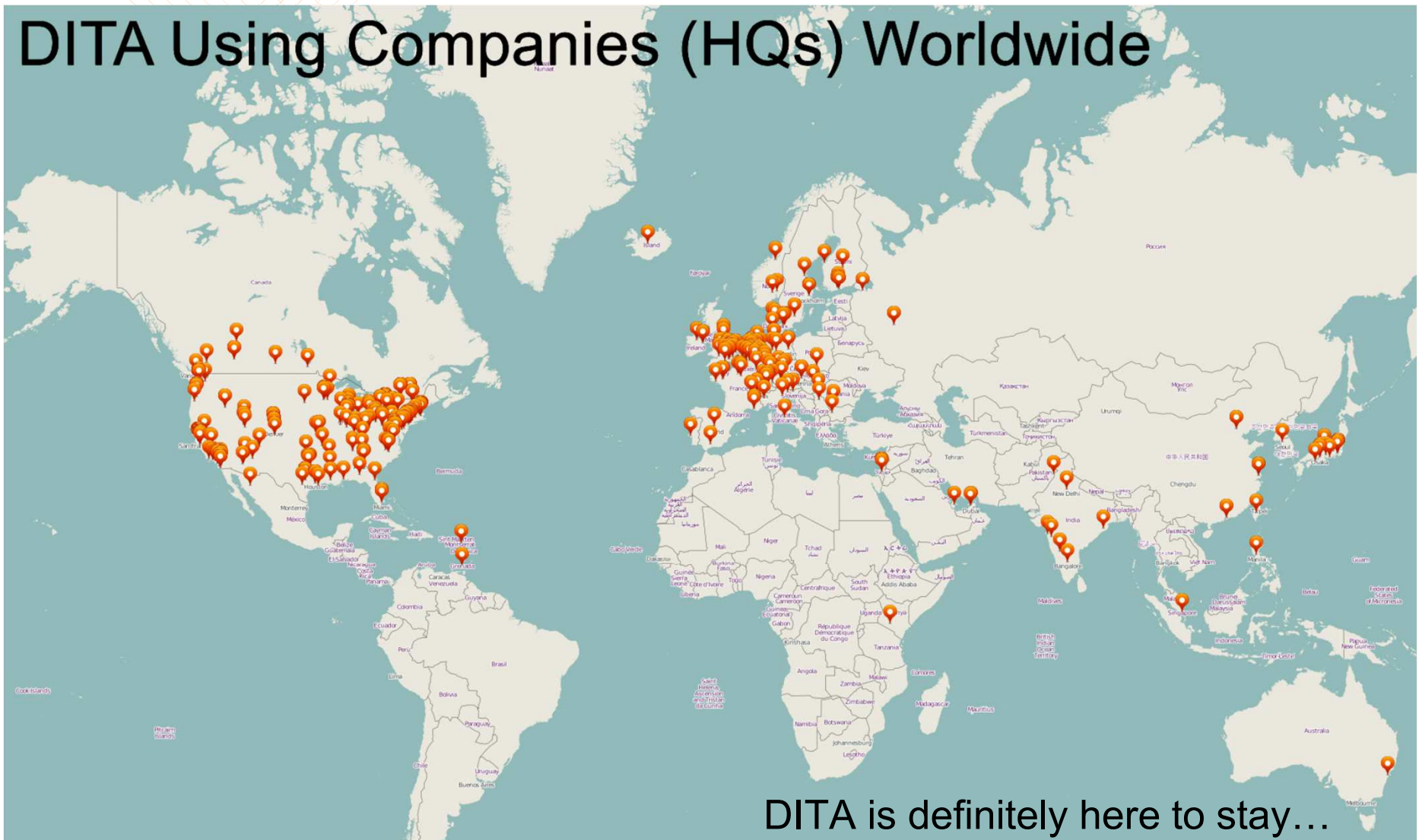
DITA Adoption Now Crosses All Sectors

- 50+ industry sectors using DITA
- Wide range of firms
- DITA is used to document:
 - Industrial egg beating machines
 - Underwater exploration equipment
 - Radiology scalpels and more!




DITA is in Use Worldwide

DITA Using Companies (HQs) Worldwide



DITA is definitely here to stay...



Authoring & publishing
CCMS
Language Service Providers
Translators and editors

CHALLENGES RELATED TO NATURE OF DITA XML

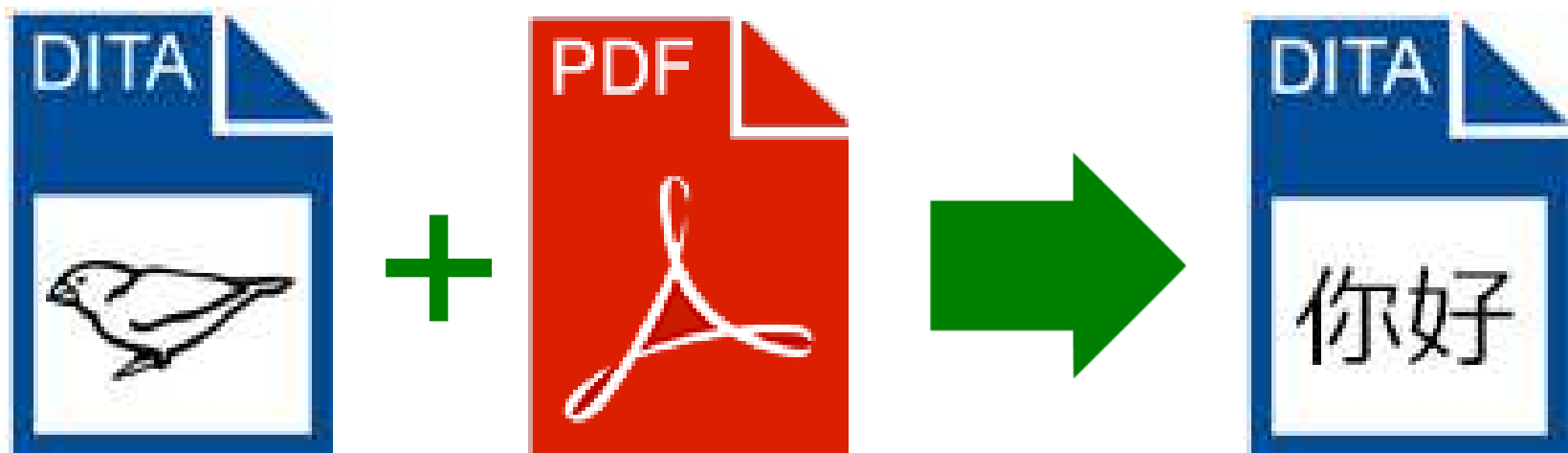
Challenges for LSPs with DITA-using Clients

- May get incomplete translation packages
- Challenges of working with just-in-time workflows (i.e. client sends only a few topics infrequently)
- Depending on how content is reused, may be hard to understand context for translation



Incomplete Translation Packages/ IXIASOFT Topic-by-topic Delivery

- It is not out of line to ask for context for the content to be translated
 - Ask for a PDF or the full DITA topics in order to get the context you need

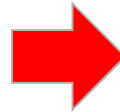


Reuse with Conrefs

- Conrefs pose a risk to translation because the reusable content is translated without the context of the topics in which it is used
- Translators cannot easily process sentences that are segmented by phrase-level conrefs
 - Translators must move the parts of a sentence to different locations depending on the language
 - Moving text so that it can be translated is more difficult if translators must contend with text in phrase-level elements that is used as a conref.
- Best practice is to conref only the following:
 - Keywords (such as product names)
 - Block-level elements and NOT parts of blocks

Best Practice for Conref-ing Comparison

Bad
(only a phrase
is conref-ed)



`<p>`If you revert a referable-content topic that's reached Localization:done, then all the topics that reference that topic will have their status set to their initial Localization state. `<ph conref="topic123.xml#topic123/abc ">`These topics will be included in the localization kit,`</ph>` their content will remain unchanged and they need not be sent to the translation team.`</p>`

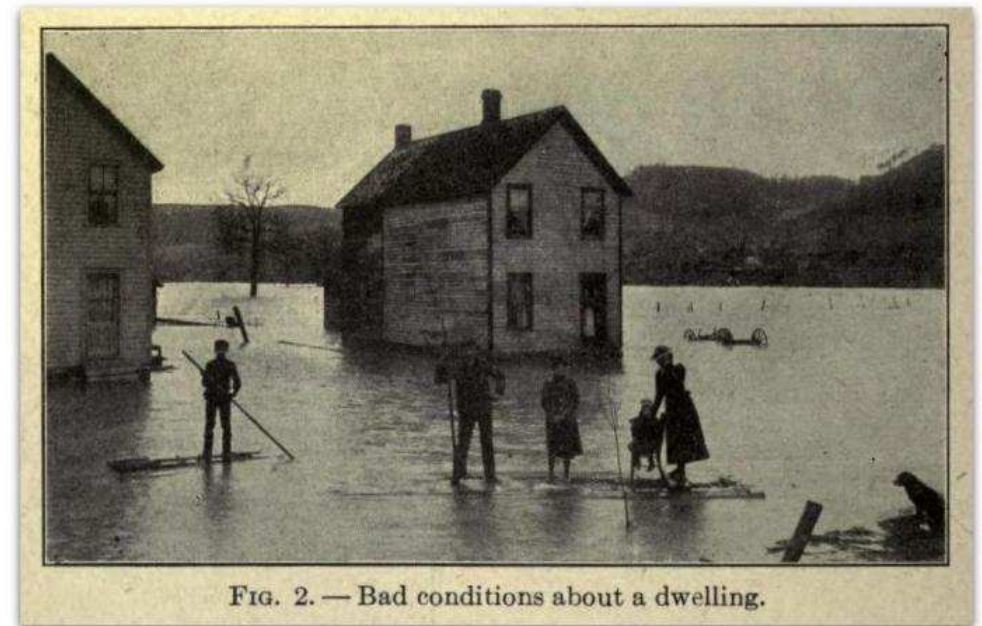
Good
(whole block
Is conref-ed)



`<p conref="topic123.xml#topic123/abc">`If you revert a referable-content topic that's reached Localization:done, then all the topics that reference that topic will have their status set to their initial Localization state. These topics will be included in the localization kit, their content will remain unchanged and they need not be sent to the translation team.`</p>`

Reuse with Conditions

- If part of a sentence in the source language file is conditionalized, it makes it difficult for the translator to rewrite the sentence for the syntax of the translation language
- As you'll see, this is a similar condition to the previous example



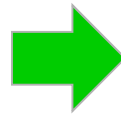
Best Practice for Using Conditions

Bad
(only a phrase
is conditionalized)



`<p>If you revert a referable-content topic that's reached Localization:done, then all the topics that reference that topic will have their status set to their initial Localization state. <ph audience=" expert ">These topics will be included in the localization kit,</ph> their content will remain unchanged and they need not be sent to the translation team.</p>`

Good
(whole block
Is conditionalized)



`<p audience="expert">If you revert a referable-content topic that's reached Localization:done, then all the topics that reference that topic will have their status set to their initial Localization state. These topics will be included in the localization kit, their content will remain unchanged and they need not be sent to the translation team.</p>`

Conditionalizing for Markets

- If some markets/locations require additional content, then you have to plan for that content in the source files

- Use conditionals to exclude the bullet for the localized language, such as `<li @audience="ja-jp">Content`

Result: When the Japanese content is imported in the localization process, the empty English will be replaced.

- Tags should be locked so that translator cannot change them.

Sorting Indexes

- Sorting order depends on language
- DITA 1.3 offers new `<sort-as>` element to help processors sort content where the base method for sorting is not obvious, such as for CJK character sets
- A content creator can now add the phonetic sound of a word and have the processor sort things like glossary
 - The Chinese word for "hello" could be sorted in a glossary by its sound ("ni hao") as: `<glossterm><sort-as value="ni hao"/>你好</glossterm>`, placing alongside other Chinese words starting with an "n" sound.

Localization Options for DITA Content

Do you need to localize the following?

- metadata, keywords, index terms, terms
- user interface terms
- images, ALT text, videos
- variables and navtitle
- Acronyms
- Clients need to markup content *consistently*
- Clients need to mark content not to be translated with translate="no"

Writing for Better Translation

- A good working relationship between an LSP and a client ought to involve good two-way communication so both can do their work effectively
- Communicate concerns and best practices back to client
 - Suggest to your client that they train their tech writers on internationalization
- A translation style guide can let writers know what to do and not to do for localization purposes



IXIASOFT DITA CMS and Localization

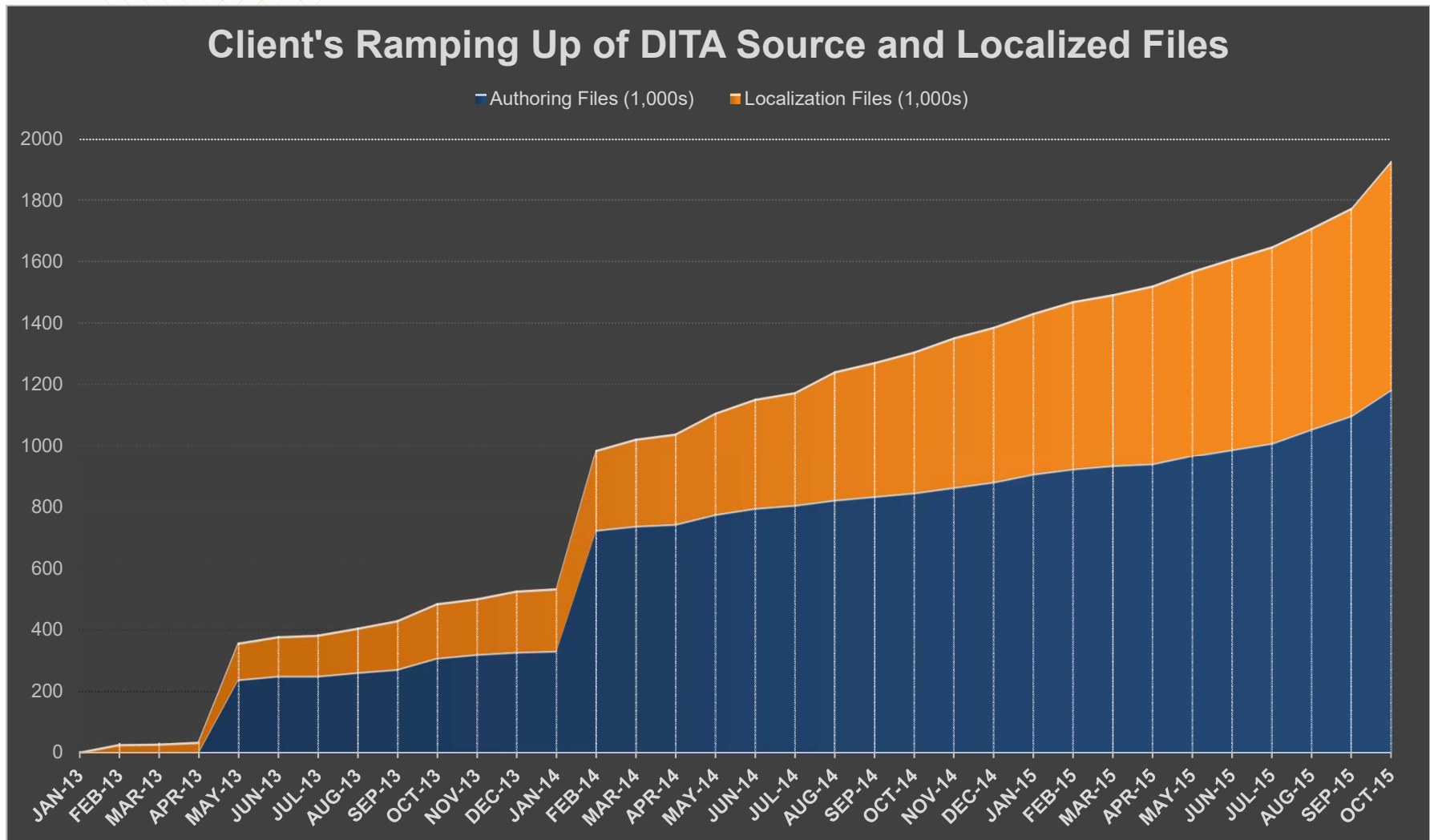
HOW AND WHY OUR CLIENTS USE DITA FOR LOCALIZATION

Scalability + Automating Content Processes

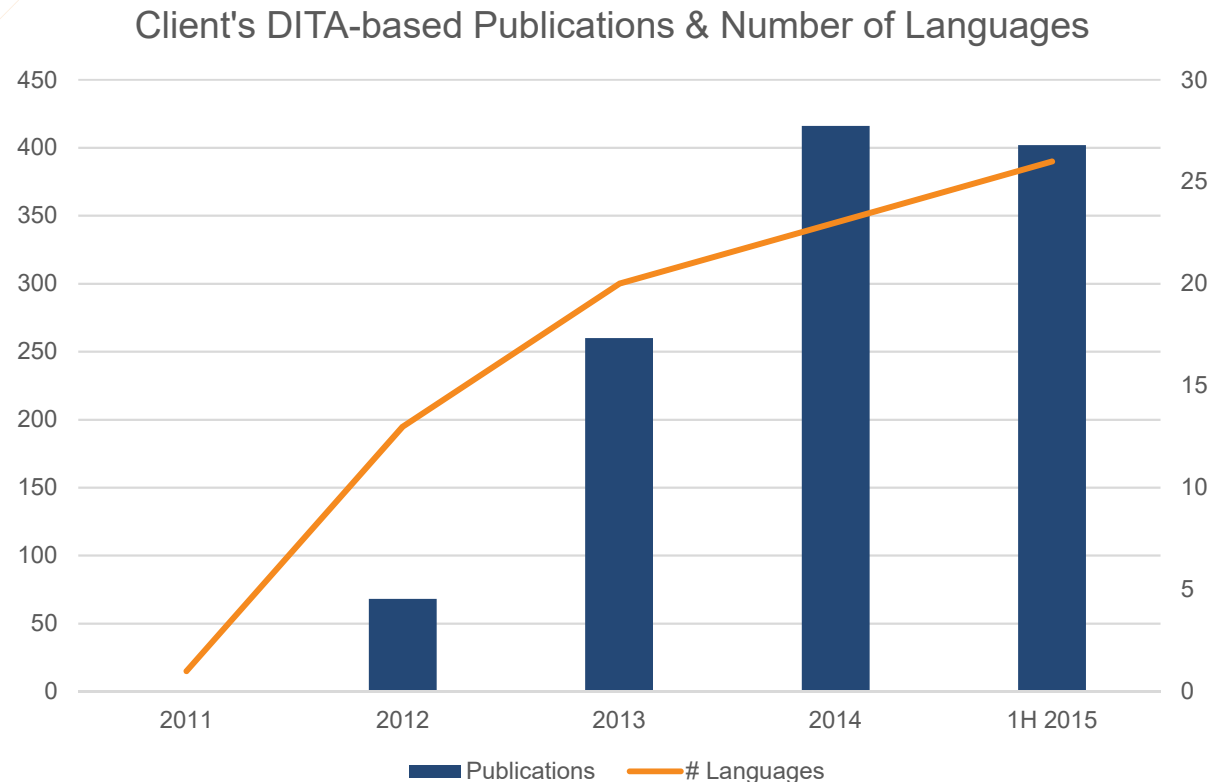
- With IXIASOFT DITA CMS a client started making nightly content builds of their documentation
 - Select customers now able to obtain early release content
- Firm localized content in 30+ languages
 - Were able to plug IXIASOFT DITA CMS into existing process; translation firms can now get hourly drops of content to work on



DITA Enables Client to Localize More Content



DITA Enables Client to Localize More Content



- A different client, showing their ramp-up of DITA-based publications and the number of languages they were able to do

Using Pre-localization

A pre-localization kit lets our customer take a snapshot of content that's under development and send it to their internal translation team so that they can prepare their translation memory

- Used solely for new product deliverables:
 - All new content
 - Allows them to stay on track with their project timelines in order to launch with required languages
 - It is easy to generate and can be done upon request when internal translators have time in their schedule

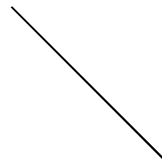
Greater Content Consistency

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

CONSISTENCY
IS 

Advice for Clients Seeking an LSP IXIASOFT that Works with DITA

“Ideally the LSP should have experience with handling DITA and be using the right technology to ensure they can process DITA proficiently. ***A DITA file is not just another XML filetype.*** Without the technology project management can become an excessively complex and a large overhead cost.”



Shamus Dermody
Client Solutions Director
XTM International

QA

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



IXIASOFT