



# **How much does XML publishing cost?**

**Small, medium, large,  
and enterprise scenarios**

Sarah O'Keefe

February 11, 2019

## Executive summary

There is no single answer to the question of how much XML-based publishing will cost. This white paper describes XML publishing based on small, medium, large, and enterprise scenarios.

The situations described here are fictional composites and provide only general estimates.

## Cost components

The estimates provided here are intended as general guidelines. They include items such as:

- Software licenses, such as a content management system, authoring tools, linguistic analysis, translation management software, and others
- Software installation and configuration
- Content conversion services (but not employees rewriting content internally)
- Content strategy development
- Information architecture and content modeling
- Content systems architecture and implementation
- Training

They do not take into account the following “soft” costs, which vary greatly for each organization:

- Employee time spent on managing the project, reviewing deliverables, researching options, and negotiating with vendors
- Content rewrites required to bring legacy information into alignment with new standards, such a new content model, topic-based authoring, or minimalist principles
- Lost productivity during the transition
- Costs from any staff turnover that might occur

They also do not include IT system costs:

- Hardware costs (that said, server costs are usually an insignificant fraction of the overall implementation budget)
- IT resources to install, configure, and maintain a new system
- For cloud-based systems especially, network infrastructure costs (network bandwidth or latency issues are an obstacle to successful use of cloud-based or distributed authoring systems)

## Small scenario

*Source control and an open-source standard for content storage in a single department*

This mid-sized organization has ten or fewer content creators who want to reduce the amount of time spent on formatting and increase their reuse percentage. They manage roughly 3,000

pages of content in English. Translation is required for seven common languages (French, Italian, German, Spanish, Chinese, Japanese, and Korean).

The company uses a source control system to manage file versioning. They choose an XML standard and use it “as is.” The reuse strategy is straightforward and mostly at the topic level. Conditional processing is needed for a few audience variants. The organization pushes output to PDF and HTML, and content is published and translated a few times a year.

The localization vendor provides some support for translation management efforts.

The company moved away from a help authoring tool or a desktop publishing tool, perhaps with single sourcing, because of increasing scalability problems. Over the next several years, the company expects to increase the number of languages that must be supported to more than 20.

## Costs for small scenario

- Authoring software: \$5,000
- Information architecture/reuse strategy: \$5,000
- Configuration of PDF and HTML output options: \$19,000
- Content migration: \$15,000
- Training: \$6,000

Estimated cost: \$50,000

## Medium scenario

### *Maximize reuse in inexpensive content management and translation management systems*

This organization has 20 content creators and two content production specialists spread across four offices in three countries (and two continents). Authors create content in English and French. Translation is required into over two dozen languages, including Russian, Arabic, and Thai.

The translation effort is costing several million dollars per year, and at least 30% of that effort is in reformatting work. Although there is a lot of reuse potential, small inconsistencies mean that reuse in translation is only about 10%. The goal is to increase the translation memory usage to around 30%. Industry benchmarks indicate that this number is conservative; similar companies are reporting over 50%.

The company implements (relatively) inexpensive content management and translation systems and a reuse strategy intended to maximize reuse down to the sentence level. They choose DITA as the content model and add a few elements and attributes to support company-specific content requirements. Output is PDF and mobile-friendly HTML. Both outputs are required for all languages, so the stylesheets must include support for all languages.

## Costs for medium scenario

- Content management systems, translation management systems, and authoring software: \$75,000 (annual cost for cloud-based software systems)
- Information architecture/reuse strategy: \$35,000
- PDF and HTML stylesheets: \$40,000
- Content migration: \$25,000
- Training: \$10,000

Estimated cost, year 1: \$185,000

Estimated cost, year 2 and onward: \$75,000

## Large scenario

### *Formalized content strategy and focus on authoring efficiency*

This organization has 50 content creators in half a dozen locations worldwide. Authors create content in English only. Translation is required for more than 30 languages.

The organization has over 50,000 pages of existing content, mostly delivered in PDF. As part of a new focus on user experience, the organization is asking what information customers really need and what information is just tradition. In addition, there are numerous information types that are managed in multiple departments, mostly in rickety spreadsheets with no clear accountability for the information.

The company begins with a content strategy assessment, which looks at the following issues:

- What content should be preserved from the existing documents?
- What is the best delivery format for each content type?
- How should information be linked?
- How can the organization ensure a single source of truth for each type of information?

Based on the recommendations from the content strategy team, the organization will minimize the amount of information delivered in traditional documents. Instead, the organization will develop decision support tools and configuration wizards that will guide users through complex buying, configuration, and troubleshooting decisions.

The content strategy document is extended with a formal business case that quantifies how the investment in content delivery will improve business results via better buying decisions, increased customer satisfaction, and reduced load on technical support channels.

The planning phase takes six months to develop the content strategy, business case, and implementation roadmap. Additionally, the organization formalizes vendor selection with a Request for Proposal process.

After extensive demos and vetting of software candidates, the company purchases content management and translation management systems, along with authoring support software. The cost of authoring support software is easily justified because of the large numbers of

authors. For example, a 10% increase in author efficiency is equivalent to 5 extra full-time employees, or roughly \$500,000 per year.

The company has unique content structure and metadata requirements, so the process of content modeling and information architecture takes a good bit of time.

Document conversion is minimized because some information is moved into configuration tools. In other cases, the company decides to rewrite information instead of attempting to convert the legacy content into the new system. The rewriting process is staged over a period of several years as documents are updated and transferred into the new systems.

Training costs are reduced by using a single delivery of a train-the-trainer class, along with live, web-based instruction instead of in-person classroom training.

The company is reducing its reliance on PDF, but does need basic PDF output, along with web content. The configuration tools require software development effort. Search functionality is a big concern because there is so much information available. The company also wants to ensure that related content is connected across the site.

- Content strategy, along with information architecture and reuse strategy: \$75,000
- Content management and translation management systems (including authoring software and linguistic support): \$500,000
- PDF, web, and other output formats: \$150,000
- Content migration: \$40,000
- Training: \$15,000

Estimated cost: \$780,000

Estimated ongoing cost: 20% of software cost for maintenance and 10% of services cost for support

## Enterprise scenario

### *Integrated content across disparate systems and departments*

The enterprise scenario is an expanded version of the large scenario: more authors, more languages, more content types, and more output requirements. There are several factors that increase the complexity and cost. They include:

- Combining XML content with data from other systems, such as product lifecycle management (PLM) and enterprise resource planning (ERP) software
- Aligning content strategy, authoring, and delivery across multiple departments or business units
- Integrating content originally authored in different languages
- Unifying technical content, product content, marketing, and other content types to produce a coherent customer experience

- Aligning all content types with the overall customer journey
- Migrating complex legacy content from hand-crafted formatting to structured markup

The wider the scope of the content project, the more expensive and challenging it becomes.

Estimated cost: unknown

## Conclusion

A transition to XML workflows requires alignment of many moving parts: content strategy, content modeling, information architecture, tools, delivery, conversion, and more. For larger investments, the planning and design phases will take many months.

Consider getting help from an experienced consultant such as Scriptorium to guide you through the process. [Contact us](#).

## About the author

*Sarah O'Keefe* is the chocoholic founder and CEO of Scriptorium Publishing and a content strategy consultant.

Scriptorium offers content strategy services to help customers solve business problems and achieve strategic goals. Since founding the organization in 1997, Sarah has worked with numerous organizations to assess content issues, develop solutions, and implement new technologies that unlock content value. She is the coauthor of [Content Strategy 101: Transform Technical Content into a Business Asset](#).



An experienced public speaker, Sarah is in demand at conferences worldwide. Sarah is fluent in German, reads voraciously, and enjoys swimming, kayaking, and other water sports along with knitting and college basketball. She has strong aversions to raw tomatoes, eggplant, and checked baggage.

You can follow Sarah on Twitter at [@sarahokeefe](#).