Determination whether DITA is good for you

C. Canart, M. Garnier, H. Morceau

Paris Diderot University, August 30th 2019

Since it was released to the public in 2005, DITA has been adopted by an increasing number of firms, and it has become one of the most popular standards for technical documentation. The DITA standard was designed with the intent to meet the needs of any firm and is sometimes thought of as a universal solution, but is this truly the case? When considering a DITA implementation, one should make sure that DITA is really suited to their particular situation and needs, and that it can be implemented successfully within the company. In this paper, we tried to understand how one can determine whether DITA is the right solution for their company before taking the leap and undertaking a content migration. For the purpose of this paper, we gathered data through a survey aimed at both DITA and non-DITA users, a review of existing resources, and several interviews of DITA experts and representatives of technical writing solution firms. Whether DITA is right for a company doesn’t seem to be wholly dependant on the size of the company, as one might think at first, but has more to do with the technical know-how of the personnel and the budget allocated to the documentation team.

Keywords: Darwinian Information Typing Architecture, DocBook, Multi-channel Publishing, Multi-format publishing
Acknowledgements

We would like to thank Camille Begnis, Alexia Daquin, Marie Girard, Rory Keeley, Nolwenn Kerzreho, Lionel Moizeau and Elise Roscillo for their support.
Introduction

Many standards, solutions and tools are available to technical writers and content professionals nowadays, and it may be hard to make a choice when trying to find a solution that truly fits your needs. Among these many options, the DITA standard is a very popular one, especially in the United States. Some technical writers only swear by DITA, and sing the praises of the Darwinian Information Typing Architecture. When choosing a long-term solution for their company, a writer might wonder whether DITA is suitable in their particular context.

DITA is an XML standard for authoring and publishing, which was originally developed by IBM for internal use. DITA was donated in 2004 to the Organization for Advancement of Structured Information Standards (OASIS) which is in charge of defining and maintaining the now open standard. Since it was released to the public in 2005, DITA has become very popular among the technical writing community. DITA is sometimes said to be a universal solution, as it was designed to “support many products, brands, companies, styles, and delivery methods” (Priestley and Schell, 2005). This is why the main topic types and their specializations were designed to fit all information and document types, and to allow infinite possibilities of combinations. Adding to this the possibility to create customized DTDs, the potential of DITA is impressive. This all seems very promising, but content migration takes a lot of time, effort and funds. Before taking the leap, it is legitimate to wonder if DITA truly is the best fit for your company. How can you determine if DITA really is suited to your particular situation and needs?

We wondered in which situations is it advantageous for a company to implement DITA, and when is it not advisable to do so? In this article, we tried to establish the contexts in which DITA is successfully implemented and lay out the benefits brought about by a DITA implementation, while also trying to understand the obstacles and pitfalls that could make implementing DITA a bad idea.

Methods

For this study, we gathered data through different means:

- a literature review of existing resources;
- a survey of technical writing professionals;
- interviews of both DITA experts and representatives of technical writing solution firms.

The survey, created using Google Forms, consisted of 27 questions. Any technical writing professional was able to participate, whether they used DITA or not, but the questions of the survey would differ depending on their answer regarding their use of DITA.

The participants answered 18 common questions regarding their company, team size, content quantity and reuse, as well as their publishing practices. There were also 6 additional questions aimed specifically at participants who used DITA and 3 questions aimed specifically at participants who did not use DITA.

The survey was shared through the Paris VII University network and released on the Technical Writing board of the website Reddit.com.
For the purpose of this paper, we conducted a total of 4 interviews with experts and professionals. The first interview was conducted in early April in the offices of a technical writing company that offers its services to a variety of clients from a large range of business sectors. During the interview, our hosts, Mrs. E.R and A.D showed us the deliverables they produced for a railway company and discussed the tools they used for this specific client and the reasons that motivated their choice.

The second interview was conducted at the end of April and consisted of a phone call with the CEO of a Saas technical writing solution, who told us about the solution his company developed.

The third interview was conducted through a video call in May with a DITA expert, L.M, who shared his experience and that of his colleagues. He explained the situations in which DITA can be especially useful for companies, but also described situations where trying to implement DITA would lead to failure.

The fourth and final interview was conducted through an exchange of emails with a representative of a DITA CCMS. She told us in which situations DITA is particularly useful, and what are the sectors that use DITA the most. She also told us what aspects of DITA could be improved.

We also reviewed existing literature on the topic of DITA. The complete list of the resources we consulted can be found in the bibliography section of this article.

Results

A total of 28 individual participants took the survey. The data presented in the results section of the study is based on these 28 participants’ answers and on the interviews we conducted.

Who uses DITA?

Among the participants of our survey, the majority of DITA users work in software or other computer sciences related sectors (Fig.1). The second most prominent sector is healthcare, with 18.2%, and then Industry.
Our contact within a well-known and popular DITA CCMS company shared with us the business sectors of their clients. Among them, 55% come from the software and telecom sectors, 25% from industry, 15% from the semiconductor industry and 5% from the training and education sector.

Similarly, a survey was conducted in 2013, led by Keith Schengili-Roberts, to identify the companies who are using DITA for their technical documentation. The project identified close to 400 firms as using this documentation standard in a variety of business sectors (Fig.2).
We can note two fields that particularly stand out when it comes to the use of DITA: Computer Software (28% of firms using DITA) and Information Technology and Services (17%). Then come the fields of Telecommunications (5%) and Semiconductors (7%).

While there are some variations in our results, our three sources of information for this particular parameter seem to be rather consistent.
The survey also focused on the size of the different companies using DITA (Fig.3). We can notice that it is mostly larger companies that are using DITA: among companies using DITA, 24% have over 10,000 employees and 20% have over 5,000 employees. Nevertheless, some small companies also use it: 12% of companies that use DITA have between 51 and 200 employees.

Regarding the size of companies who use DITA, our own research showed us that the size of the technical writing team does not really have an impact on the use of DITA. Very big teams (more than 50 employees) as well as smaller teams (between one and ten employees) seem to be using the DITA standard (Fig.4).

![Pie chart showing the size of documentation teams](image)

*Figure 4: Size of Documentation Team (number of employees). Survey on 28 respondents, July 2019.*

**Why choose DITA?**

One question in the survey we conducted for this paper aimed to analyse the benefits of DITA in terms of time and cost saving strategies. The participants gave their impressions regarding the benefits of DITA they could observe within their company.
The results show the benefits of DITA perceived within the participants’ companies:

- harmonize content structure (82%)
- harmonize editorial style (73%)
- reduce review time (73%)
- facilitate updates (73%)
- harmonize terminology (45%)
- reduce translation costs (36%)

Regarding time management, respondents also noted that the use of DITA induced shorter production times and faster deliveries.

Survey respondents also provided the following reasons when asked why they chose to use DITA:

- Ease of reuse and data management
- Portability of a widespread international standard
- Ability to give more structure to the information
- Modular nature of DITA content
- Richness and extensibility of the schema
- Richness of pre-existing tools and free resources
- Possibility of multiple publishing outputs

This last point seemed particularly important to DITA users. Indeed, more and more companies are looking to publish their content through different channels and formats. Print and PDF now have a multitude of concurrents.

The results of our survey show that 81% of DITA users produce multiple publishing outputs.
One may wonder how content is being published nowadays. How do technical content producers deliver content? What are the most popular channels and formats?

The Data Conversion Laboratory (DCL) and The Center for Information-Development Management (CIDM) conducted a joint study in 2018 to gain insight on the trends in content delivery. The sixth and seventh questions of the survey focused on the content publishing practices of technical communicators. The results of the survey showed that the most popular channels were the web, embedded user assistance and print, and the most popular formats HTML and PDF.
When inquired about the type of content delivery their customers asked them to provide, respondents provided the following answers:

- more searchable content (68.37%),
- learning videos (55.44%),
- content available on mobile devices (45.58%),
- embedded help within software applications (36.39%),
- topic-based documentation on a company website (34.01%).

**What is the cost of DITA?**

DITA in itself will not cost you anything, however, you need tools to implement DITA properly: an XML editor and a Content Management System (CMS).

To write documentation in DITA, one will generally need to use an XML editor. The most used and popular XML editor currently available is oXygen, which can be used with any XML-based technology. The price of the license starts at $1,155 a year.

XMetaL Author Enterprise is also an XML editor. It allows you to author in DITA and also provides structure and automation to content creation. This tool is slightly cheaper than oXygen and costs $895 a year.

Adobe FrameMaker is another XML editor, which was released in 2015, and that offers a complete solution to edit XML content in a simplified environment. This editor costs $999 a year.
While an XML editor allows you to produce DITA content, a CMS helps you manage this content. A CMS is a piece of software that can be used simultaneously by several individuals and that allows you to dynamically design, manage and update your content. A CMS also helps a company to structure its content with categories, pages, articles, etc. The modular nature of DITA usually leads to a large number of files to manage, which makes a CMS especially useful. However, this kind of software can be quite expensive, and the range of prices is large. For example, you can find IXIASOFT DITA CMS with a cost of €2,355.00 per year and per user while easyDITA from Jorsek Software will cost you €35,400.00 annually.

<table>
<thead>
<tr>
<th>Content Management System</th>
<th>Company</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>DITA CMS</td>
<td>IXIASOFT</td>
<td>€2,355.00 per user, annually</td>
</tr>
<tr>
<td>DitaToo</td>
<td>iG5 Authoring Tools</td>
<td>€1,500.00 per user, annually</td>
</tr>
<tr>
<td>easyDITA</td>
<td>Jorsek Software</td>
<td>€35,400.00 annually (+€45 for each additional user)</td>
</tr>
</tbody>
</table>

*Table 1: Example of CMS prices, July 2019.*

If not DITA, what else?

Which software(s) or method(s) do you use to write and publish your content?

![Software pie chart](image)

Even if DITA is increasingly popular, there is a large range of other options available. And indeed, the results of the survey show that among the participants, non-DITA users work with a large variety of solutions and software to produce their content. The most used software is still Microsoft Word (14.3%). At the second place there is the open and shared workspace Confluence (10.7%) on equal terms with the cloud customer service platform Zendesk and the MadCap Flare content authoring tool that generates output in various formats, including HTML5, Microsoft Word, PDF and WebHelp. Calenco (7.1%) is also a
popular software that can be used for multichannel publishing and that works with any XML standard (DITA, DocBook, S100D, etc.).

However, it is important to keep in mind that 76.5% of the respondents to this survey are Francophones. Therefore the results must be treated cautiously, as they are not necessarily representative of the technical writing community as a whole, but rather of the French and French-speaking community. On a global scale, results to this question could vary greatly.

Which software or method did you use before switching to DITA?

![Pie chart showing software usage]

**Figure 9: Solutions used by users before implementing DITA. Survey on 12 respondents, July 2019.**

The above figure (fig.9) shows which solutions technical writers used before they implemented DITA. Among the participants of our survey, Microsoft Word was the most used software for unstructured writing (41.7%), followed by Adobe FrameMaker, both structured and unstructured (25%), and finally, Confluence and Indesign (both at 8.3%). Let us also note that some technical writers were writing plain text without using any particular tool or standard. It would appear that most technical writers seem to have moved directly from unstructured writing to the DITA standard.

**Discussion**

There are many reasons that could make implementing DITA a good solution for your company’s needs. There are however some elements to consider before making that choice in order to circumvent the more avoidable pitfalls which could turn implementing DITA into a really bad decision.

**Who should use DITA?**

Data shows that, while the software, telecom, and information services sectors make up the majority of DITA users, there is a wide range of business sectors and industries in which
DITA has been implemented. It also appears that the same can be said about the size of the company and documentation team. Very big companies as well as small teams seem to be using DITA. While DITA was designed by IBM and is consequently suited to big companies' needs, it would seem that size alone is not the deciding factor when it comes to using DITA. Two question that comes to mind: “When is it particularly advantageous to use DITA?” and then “When is implementing DITA not advisable!”.

**When is it advantageous to use DITA?**

As shown in the results of our survey, DITA can help harmonize content structure, terminology and editorial style. This is largely due to the possibility of creating libraries destined to host reusable content such as product names, technical terms and definitions. This is particularly interesting for large teams who need to ensure consistency for the content they produce. Since DITA is a very powerful and extendable standard, companies which have the means to develop their own schema may also choose to do so. They would thus benefit from a tailor-made solution fitting their particular needs that would also make it easier to implement writing guidelines and ensure consistency.

The single-sourcing aspect of DITA makes it particularly interesting for companies dealing with a lot of content and/or many translation languages. Updates are done in one place and applied wherever the topic - or chunk of reusable content - is referenced, and what has been translated once will not need to be translated again. The use of libraries for technical terms also prevents the same term from being translated in different ways by different translators, as it will be translated only once and then reused. This contributes to reducing translation costs and update time, which are both key concerns for companies with a lot of content to maintain.

The separation of context from content in the DITA standard allows users to arrange their topics as they wish within maps that have different combinations of content. Coupled with content profiling, it makes DITA a very powerful solution for companies with a lot of product versions and variants, shared functionalities, or user types. DITA offers a lot of flexibility for output combinations, and makes it possible to easily produce and maintain content pertaining to many product variants and user types.

In the “Information Age” (Richard Leghorn, 1960) in which we live today, consumers require easily searchable and accessible content that is tailored to their needs, while companies in search of visibility try to extend their reach across channels of publication. Providing a multichannel experience to your customers is a great strategy to fulfill the needs of different types of users and create customer loyalty. For it to be successful, however, companies need to keep a consistent user experience, regardless of format and platform. Consistency and a cohesive user experience across channels are the key to a multi- or omnichannel publishing strategy, because they help build trust with customers and maintain a uniform, reliable identity for the company.

DITA separates content and context, which facilitates reuse of content across channels of publication. Moreover, DITA’s extremely modular structure not only facilitates reuse of content: it also helps to improve retrievability - with very small chunks of content that can describe very specific situations and be tagged to come up in search engine results. As seen in the results of our survey, DITA also improves consistency in terminology, content structure and editorial style, which greatly help build and maintain a consistent brand.
identity. Lastly, the single sourcing approach to multichannel publishing ensures that updates are facilitated and guarantees the consistency and validity of information across channels.

The technical challenges of DITA

Even if DITA greatly facilitates the production of multiple output formats, the basic output options included with DITA Open Toolkit (DITA-OT) are rather limited and simplistic in terms of visual design. In order to give the output - especially print - a nicer, more unique appearance that properly fits the company's brand identity, the company will have to invest either time or money. The company will need to create templates, generally in XSL-FO, which is the type of style sheet used to format XML documents for print, and CSS, for the web output.

If there is no technical expert on the team who is able to create DITA-OT plugins and templates, the company will need to hire a consultant to do the job. This method proves rather costly but in the absence of someone who knows their way around XSL-FO and other formatting languages for XML output, this is the best way to obtain a quality output formatting suited to the needs of the company and reflecting its identity and graphic charter. The DITA expert we interviewed informed us that many companies hired a programmer or technical writer with programming skills to manage output formatting and automate processes.

In short, while DITA is an extremely powerful standard when it comes to multiformat and multichannel publishing, customizing the content output using DITA-OT takes either a lot of time and skill, or a substantial budget. This may be an impediment for small teams or teams who run on limited budgets, but some solutions do exist for the more determined to get started.

For those who would like an easy way to customize their print output, Jacquie Samuels suggested in a TechWhirl review that Tengwar Systems' DITA InPrint, while still rather expensive, is much less so than hiring a consultant or spending time trying to customize the DITA-OT. According to her, DITA InPrint could be an interesting solution for small teams: "It brings the cost of DITA into the realm of the possible for many small and medium-sized documentation teams."

For those who would like to teach themselves how to produce print output for DITA and keep the processes in-house, resources like Leigh White’s book, DITA for Print, can help you build the skills required to get started with DITA-OT output customization. The book DITA for Print is considered the go-to reference for building a print customization plugin for the DITA Open Toolkit.

When it comes to the cost of CMSes and XML editors one is faced with a choice. Once again the decisive factor will be either budget or technical skills. It is entirely possible to produce XML and thereby DITA content with open source solutions like Notepad++. However these require a lot of rigour and are not nearly as efficient as more expensive solutions, like Oxygen XML Editor, which come with a parser and autocheck functionalities to enforce a schema, and thus ensure consistency in content structure. It is also be possible, in theory, to use DITA without a CMS, but it would require a very high level of technical knowledge and skills to do so properly. Most CMSes are expensive but ensure an efficient, quality system that is easy to use.
About pitfalls and the implications of a DITA content migration

Even if DITA is a very powerful standard that comes with many benefits, it is important to keep in mind that a DITA implementation also has its pitfalls. To avoid them, you must be sure that DITA is the right solution for you and plan ahead carefully. Implementing DITA implies undertaking a content migration. It is something that has to be prepared properly to avoid a major failure.

First of all, it is important to remember to look ahead in the process. You have to keep in mind what you want to do and where you are going. Thanks to the interviews we conducted for this study, we have been able to identify several points one needs to consider before implementing DITA:

- **The structure of previous and current documentation**
  
  Examining the structure of previous and current documents is a good way to evaluate the amount of changes that need to be made in order to have them fit the DITA standard. This process can help to determine which schema model or DTD will allow for minimal reworking of the source content during the migration.

- **The appearance and format of deliverables**
  
  We have already established that creating custom output with DITA-OT is no easy task for beginners. Therefore, the appearance and output formats of deliverables are an important point to consider when moving to DITA. The cost of producing customized templates can vary greatly depending on the technical profile of the team and the necessity to hire someone to do the job, as opposed to doing it in-house.

- **File storage and management**
  
  Another important aspect of a DITA migration is to decide how the topic files are going to be managed. Indeed, due to the modular nature of the standard, content production with DITA generates a much higher number of files than is common when dealing with traditional monolithic documents. This is why many technical writers choose to use a CMS to help them manage their files.

- **Documentation and content exchange between departments**
  
  If the documentation team often exchanges documents with other departments, like training or marketing, it is important to think ahead. Indeed, it is vital to know beforehand if the other services are going to move to DITA as well, and, if not, to establish how the exchange of documents should be handled.

- **Time constraints**
  
  It is important to know how long one can spend on the content migration. If there are tight time constraints, it might not be a good idea to migrate all the existing documentation to DITA: it is a long and tedious operation that must be done correctly from the beginning. One cannot rush the migration process or they might lose even more time in the future due to migration mistakes that will need to be fixed.

- **Training**
  
  DITA is a very technical tool. To implement it properly, it is important that the employees of the company be properly trained. If no one on the team has ever used DITA, it
might be necessary to hire an expert to help plan the migration and teach the team DITA best practices. All of this can require a substantial budget.

It is important that the project be well defined, and that a plan be developed for it, although this one might change or evolve during the process. It is also recommended that someone be responsible for the project from the beginning so they can be used as a primary point of contact.

One should keep in mind that, although the choice of software is important, it is not everything. There can still be some data issues and the team must be ready for it. Moreover, one should not rely on automation only during the migration process. Even if automation can be quite powerful, it is important to have human beings managing the data during the preparation and post-conversion.

Finally, it could be said that for such a project to be successful, one needs to understand - and be prepared for - all the implications of a DITA migration, and to make sure that the use of DITA really does align with the needs of the company and its documentation team.

**Why move away from DITA?**

Beyond the reasons that would prevent someone from moving to DITA, there are also reasons to move away from it. Some people have used DITA before deciding to move on to something else, because DITA no longer fit their needs, or they found something that worked even better for them. This is the case of Tom Johnson from the technical writing blog *I'd Rather Be Writing*.

Tom Johnson chose to move away from DITA for several reasons, among which:

1. The Information Typing model that enforces topic types can be too restrictive, and it is frustrating to use models that do not exactly fit what you're documenting.
2. Open source DITA solutions develop too slowly and innovation with DITA is too reliant on vendors.
3. The output is not easily customizable.
4. DITA doesn't integrate well with other non-DITA content, and this can create walls and silos between content from your service and that of other services like Marketing or Support.

**DITA's nor for me, what can I use?**

As we discussed it earlier, DITA, for all its potential and benefits, may not always be the best solution for everyone. Even in the realm of structured content, there are other solutions that might be more suited to different situations.

A team of technical writers we interviewed worked with a sizeable railway company, to help them build a documentation and training system for their technicians. The team worked on different deliverables for their client, producing output suited for two main user types working in different contexts. They successfully produced a web platform and interactive PDFs using single sourcing, but without the use of DITA. The team chose to use Docbook rather than DITA, because it is an easier place to start for companies that are unfamiliar with structured documentation standards: it is less costly, less technical, and easier to implement than DITA. While their client was a big company, DITA would have been too technical for the first experience of non-experts. This shows that, even if DITA was created by a big company and is generally more easily implemented in big companies (with both financial and human means), being a big company does not automatically make DITA the best solution.
there is. This particular project was conducted using Calenco, a SaaS solution for authoring and publishing content on different media. This solution allows writers to work collaboratively on projects, reuse content via single sourcing, automate page layout and graphical charter and also offers the possibility of multichannel, multiformat publishing.

Most of the professionals we interviewed suggested that with its less restrictive structure Docbook is easier for non-experts to get started on than DITA. If multiple teams and services - such as Training, Support and Marketing - are involved with the technical writing team and participate in content creation and exchange, it would be difficult to have them all move on to DITA. The training costs, among others, would be quite tremendous, and the other teams might not be amenable to moving to such a restrictive standard as DITA. So if other services regularly contribute to the documentation process, switching to DITA might not be a good idea, unless one is ready to train and convince all the involved personnel to use the DITA standard.

Another solution that was mentioned to us during one of our interviews is Paligo. Paligo is a relatively new CCMS designed in Sweden by a company that formerly worked in DITA consulting. Their goal was to come up with a solution that was easier to use and less expensive than DITA solutions. Paligo offers composability and component management without the significant content constraints on which DITA relies.

**Study limitations**

One of the study limitations was the small number of participants, with a total of 28 individual participants for our survey. Another limitation is the date of publication of some resources we used for our research, which can go back up to 2013, which is starting to become a bit old, especially in a sector where innovations come fast. Our study could have benefitted from a larger number of participants, and perhaps, more one-on-one interviews with professionals who use DITA in ongoing projects.

**Conclusion**

DITA is a very powerful tool if used properly. DITA can help you reuse, update and publish content in multiple formats, across multiple channels. It can help you to harmonize your team’s editorial style, content structure and terminology, as well as reduce translation costs for your company. If you are a big company with a lot of content to manage and update, if you need to account for many translation languages, if you reuse a lot of content for different product variants or user types, or if you wish to do multichannel publishing, then DITA might just be the right solution for you. However, DITA is also a technically complex standard, and not the easiest solution to implement.

One will need to have either the technical or the financial means to successfully implement DITA, and use it to its fullest capabilities. When considering whether to implement DITA, one should consider the following elements: do other services regularly get involved in content production? Will I manage to win them over to DITA? Is my team already familiar with XML and DITA practices? If not, do I have the financial means to train them? Do I have the budget for a DITA authoring software and a CMS? How will I customize my output? Do I already have an expert on the team, or will I need to hire a consultant? What are the time constraints for the migration?

If one has neither the technical knowledge nor the budget, implementing DITA might not be a good idea. There exist more accessible solutions which can replicate some of the
advantages of DITA, for a lower cost and with a lesser degree of technicity. Docbook-based solutions might be good to get started with XML among non experts or with a more limited budget.

Open-source DITA solutions develop slowly, but due to the enthusiasm and hype that DITA sparks among the community, new solutions might emerge in the following years to make DITA more accessible. We will need to be on the lookout for such advancements in the future.

References


