talend

Talend delivers consistent documentation through metadata-driven publishing

A discussion with Elisabeth Sabot,
Senior Director of Technical Communication at Talend.

Elisabeth, can you please describe Talend and your role?

Talend is a global software vendor of data integration and integrity solutions. Catering historically to technical/IT users, Talend's solutions now also target business users and data scientists. I have been at Talend since day one, I was actually their first tech writer! I am now in charge of the technical communication department, overseeing the documentation, curriculum development & community teams.

What was the context of your dynamic delivery project?

Basically, our Confluence-based system (a wiki by Atlassian) had run out of steam. In order to publish new or updated content, even just a small hotfix, we had to upload the entire knowledge base - a complicated process that would take at least 48 hours and require IT involvement.

And for our users, any navigation started with a sequence of pages through which they had to select a product, then a version, then a type of task… it would take 4 or 5 clicks before reaching any useful content. The inability to search across sources, the lack of facets, and an overall rigidity of the system, finished to convince us that we needed to explore other options.

So we listed our pain points and our users' pain points. We designed our dream system, and looked at several possible paths: build our own, or buy a delivery platform we could customize to our needs. Great thing is that we had support from our executive team – they are a bunch of pragmatic folks and are very open to change. But we needed to act quickly.

Fluid**Topics**

One major thing we had to take into account: we have pretty advanced taxonomies at Talend, and we have designed a strategy where access to content is highly contextual and personalized, entirely driven by metadata. The system we would implement needed to integrate with our metadata management.

How are these taxonomies managed?

Talend uses a specialized product to create, develop and maintain corporate taxonomies. This company-wide initiative expands beyond the technical communication department. The end goal of these taxonomies is for them to apply to all content produced at Talend, regardless of who creates it: tech writers, consulting, support, presales, marketing, product, etc.

From a practical standpoint, we use them to describe the various products and packages Talend offers, the licenses, the use cases, etc. Every piece of content is tagged at a fine-grained level: when it comes to DITA content for example, we actually tag at the topic level, not just entire books. This helps to add context for when a piece of content applies, seen from different perspectives.

> **Access to content is highly contextual and personalized, entirely metadata-driven.**

For the end user, it means that the documentation they access expands beyond a traditional, monolithic user manual. Any dimension in the taxonomies is both an entry point, and a guide through a wide variety of content. The reader, depending on their role, technical expertise, and needs, is able to fully leverage all the content and to navigate the knowledge repository through tailored paths.

Who are the consumers of your documentation?

The first group comprises personas who buy Talend's solutions. They are pretty diverse: developers, consultants, business users, data scientists. The share of non-technical users is poised to increase, as we continue to deploy our Cloud services that remove a layer of complexity. Our documentation must be suitable for these new users.

We also address internal users: anyone who implements, or helps to implement our solutions. This includes presales engineers, support experts, professional services, project managers. These are advanced users, who know our products well. They really like to use taxonomies when they look for information.

📖 Definition

Taxonomy: set of concepts and things organized in a hierarchy (tree structure) to describe a specific domain.

For example: classification of plants

What drove the selection of Fluid Topics?

The guiding principle of our choice for a delivery platform was to offer to our end users the best possible experience with the documentation.

To summarize our key requirements, we were looking for:

- Native support of DITA.
- Centralized and consistent support of all our metadata.
- Real-time integration (via RDF/XML) with taxonomies – created and maintained outside of the delivery platform.
- Advanced faceted search.

Fluid Topics provides access to all content, regardless of its type, at the highest level of granularity. Users can navigate content using all the dimensions of our taxonomies. Search is ultra efficient, powered by a comprehensive filtering system.

Facets are actually very important. They are key to a positive search experience. Most users understand that, and use them extensively. But a handful of users don't. In these cases, we need to adapt to their behavior, understand the context of the search, and still return relevant results.

💡 About Talend

Founded in 2005 to modernize data integration, Talend (NASDAQ:TLND) is a leader in cloud data integration and integrity solutions, liberating data from legacy infrastructure. Its Talend Data Fabric is a suite of apps for data integration and integrity across public, private, and hybrid cloud, as well as on-premises environments, and facilitates greater collaboration between IT and business teams. Over 3,000 global enterprise customers have chosen Talend's solutions.

📄 https://www.talend.com/

How fast was the deployment?

Once we had selected Fluid Topics, it took us less than six months to go live with a portal that aggregates several generations of technical documentation (from legacy DocBook to modern DITA), that uses all our metadata, and provides efficient access to the entire content corpus. Fluid Topics makes a huge difference because the metadata is consistent across the different data types with the possibility to apply our taxonomies to any content, regardless of its source.

> *Fluid Topics is uniquely capable of leveraging our fine-grained metadata to deliver an efficient search experience.*

In parallel, we were working on the migration of all our source content to DITA and transitioning to a new content management system. We made the conscious choice to deploy the delivery portal first, and then handle this migration, because Fluid Topics would insulate our users from the changes this process brought. There was no disruption for documentation consumers.

Which contents are published through Fluid Topics?

As of today, we are integrating several types of content in the portal:

- Technical documentation topics, which includes legacy DocBook content and all new content created in DITA.
- Markdown content written by developers.
- HTML content provided by partners.
- Knowledge base articles, created and managed in Lithium.
- Community forums, hosted and moderated in Lithium.

The latter two types of publications are automatically tagged with a simplified version of our taxonomies, based on which board or category it has been created in.

We are also adding more sources. One area that is of special interest for us is to tap our subject matter experts: presales engineers, consultants, developers, software engineers – all have deep expertise in their respective fields, but no knowledge of technical writing. The content they produce is not always immediately suitable for publication.

" Expert thoughts

François Violette, Information Architect at Talend, explains how Talend's taxonomies were created.

Our taxonomies aim at describing all the products and services that Talend offers – not a simple feat by any means! From the client's point of view, a Talend product is very modular. It consists of an assembly of modules, add-ons, deployment styles (cloud or on-premises), server resources, etc.

It also describes the tasks a user can perform with the product. This defines a business-driven mapping, linked to usage. The overall mapping was created with help from project managers and technical experts who have inventoried use cases with a deep level of granularity.

For example:

- The user is navigating a knowledge base, reading a user guide,
- The target environment is Salesforce.com,
- The goal is to automate tasks, to schedule and monitor them.

This set of conditions defines a specific context that will lead to appropriate content.

One has to keep in mind that the taxonomies are defined and managed externally to the content. Which means that we don't tag with keywords, but with identifiers that correspond to the concept nodes of the taxonomies. Concepts are organized independently from languages, with localized labels. We have defined translation workflows for this metadata, and how to inject taxonomies into the documentation during its production by tech writers.

For the reader, navigation through topics is seamless but at the same time, a content piece that only exist in another language (for example, English) will still be suggested if relevant. Furthermore, we can rearrange the structure of the taxonomies or modify the labels without having to modify the content. This gives our product and marketing teams great flexibility.

We designed content workflows and publication mechanisms that combine reactivity with content quality assurance, to be able to properly leverage these contributions and publish them quickly. We also let subject matter experts write in formats they already know such as Markdown. And we give them ways to easily tag their content with our taxonomies.

How are you using taxonomies for content tagging?

All users don't master taxonomy at the same level (if at all). Therefore we have implemented different ways of tagging content, depending on the type of content and who produces it.

For tech doc topics, metadata are selected by the technical writer in the full metadata repository.

In Lithium, community contributors are driven toward a specific area of the site, through a hierarchy based on a simplified version of the tasks taxonomy. Any publication is then automatically tagged with metadata based on where it was created. Contributors can add other metadata but only from a proposed set, also controlled by taxonomies.

Subject matter experts (internal to Talend) receive help inside the publication workflow to select metadata from authoritative lists, without having to master the hierarchies of our taxonomies.

Aside from what you've mentioned before, anything else important about Fluid Topics?

Fluid Topics is natively multi-lingual. As a global company, we produce content in several

> **Quick Facts**
>
> - Tech doc team: **20 people** (including an information architect, 14 tech writers and 2 translators) based on **3 sites in 2 countries** (France and China)
> - Content: **DITA, DocBook** (legacy), **Markdown, AsciiDoc**
> - CMS: **oXygen** with **GitHub**, migration underway to **IXIASOFT** (DITA CMS)
> - Other sources: **Lithium** community sites, contributions by subject matter experts (no format consistency)

languages – English, Chinese, French, etc. Even if everything exists in English, it may not be the case for every language. Thanks to our language-independent taxonomies, users can still discover content that doesn't exist in their language.

Probably the most valuable aspect is the ability to deliver through the same portal, contents produced by the tech doc team and by the community – external and internal. Combine disparate sources of information with uneven levels of quality and authoritativeness, reconcile them, match heterogeneous contents, are all complicated endeavors. The only way to do this in a scalable way was to natively integrate our taxonomies, which Fluid Topics was able to do out of the box.

> *Fluid Topics combines out-of-the-box disparate sources of information that have uneven levels of refinement.*

Check out the Talend Documentation Portal: https://help.talend.com