Metadata annotation

David Gråbæk, Systime
Agenda

1. Background
   Single Source Publishing
   Levels of metadata
   From static products to continuous processes
2. Our challenge
3. What we did
   Metadata consolidation
4. How we use our metadata
5. Lessons learned
Single Source Publishing - Concept

Product A
- Content

Product B
- Content

Product C
- Content

Traditional production

Medium A

Medium B

Content

Future media

Single Source Publishing
Single Source Publishing - Technologies
Levels of metadata

We have metadata schemes at (at least) two levels:

**Title related metadata**

This deals mainly with Subject, Line of study, Student level etc.

These data are collected in our financial systems to be used across our shop, account-system, etc.

**Content metadata**

This deals with the content itself.

We use a semantic markup scheme that has been going through a lot of changes over the past few years.

In the past few years we have experienced a change in customer buying patterns away from single titles towards flat-rate solutions or collections of content. This has challenged our metadata schemes in several ways.
Our challenge

During the first years of producing books and digital publications this way, the number of content categories went from approx 40 to more than 200!

Most \( \left( \frac{3}{4} \right) \) were local ones only used in a single title.

- This made a lot of internal workflows difficult to manage.
- This made it difficult for us to achieve the level of service we wanted - each title had to be thought of as an independent entity - but our business focus was shifting towards selling collections of titles and/or subscription based models.
- The data was based on several different principles - eg. semantic markup, student-level, access restriction etc.
- The problem was growing fast!

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |  |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |  |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
What we did

- 200 content categories were consolidated into approx 50. This was a process that involved a lot of pseudo-philosophical discussion about the inherent differences between cases, examples, theorems etc.
- The product of this was a global list of categories into which all of our content should fit.
- All new categories were based on a principle of semantic markup.
- Old categories were converted to fit into the new scheme. This was a semiautomated process where we converted categories one title at a time using a self-developed tool.
- We produced easy to use documentation for all content categories, with descriptions, examples etc.
- We setup procedures for ongoing revision of the global categories.
Content markup principle

Separation between content and presentation

Visual representation vs Semantics

This content is an example

In this context examples are presented in a blue box

Example

This is a blue box containing an example

Example

Traditional production

Database publishing
How we use our metadata

- Metadata is used for layout in all of our products - digital and printed. Layout features can easily be reused from project to project.
- Metadata is used for filtering search results across titles - “Book!”. Customers can create their own custom filters based on our metadata.
- Metadata is used to find and compare products in our shop.
- Title related metadata is being used to measure the performance of each title. This is a useful management tool for evaluating a books performance and setting future goals and focus areas.
Lessons learned

- It can be difficult to anticipate which metadata schemes will work for you. A fair degree of experimentation is needed.
- Important to get your editors involved in the development.
- Documentation is - as always - important.
- No metadata scheme can be permanent. Allow your metadata to evolve over time.
- Metadata issues can snowball if they are not dealt with fast.

The right metadata allows you to

- Perform better - especially with digital products.
- Gives you more options for cross-referencing, search-filtering etc. This can really improve user-experiences.
- Content markup can be useful in versioning of your books - eg as ebooks.
Questions?

Comments, ideas, thoughts?