DITA and the Challenges of Single-Source Article Publishing

Tony Self
Swinburne University of Technology
aself@groupwise.swin.edu.au

Tony Self has worked as a technical communicator for almost 30 years, with the last 20 of those years specifically in the areas of online help systems, computer-based training, and electronic documents. In 1993, Tony founded HyperWrite, a hypertext and technical documentation company based in Melbourne, Australia. The majority of his work involves providing online and Internet strategy advice, innovative solutions and specialised training for customers in Australia and other parts of the world. Tony also lectures in the Technical Communication programme at Swinburne University in Melbourne, and holds a Graduate Certificate in Teaching and Learning and a Graduate Diploma in Technical Communication. He is a Fellow of the Institute of Scientific and Technical Communicators (UK), member of the OASIS DITA Technical Committee, and chair of the OASIS DITA Help Subcommittee.

Abstract

DITA is an emerging writing methodology built around XML, and promises major improvements in efficiencies of writing, editing, maintaining, publishing and delivering textual information (Day et al, 2005). Although it is currently predominantly used for IT documentation, DITA can be applied in most fields of publishing, including Web site content management and magazine article publishing, to reduce costs and increase efficiency. DITA is built upon the principle of separation of content and form, which attempts to completely abstract the message of written communication (the content) from the reading format in which it is presented to the reader (the form). This separation paradigm makes redundant many traditional methods of composing, editing, changing, laying out, printing and publishing. The benefits of a change in approach to writing are numerous. One benefit is the ability to single-source, where content source can be maintained in a delivery-agnostic format (DITA), and automatically published to many different types of publications.

The approach taken in this paper is to introduce a case study, explain the concept of the separation of content and form, and to then discuss the challenges that a DITA workflow brings to editing in single-sourced publications, to content re-use, and to single-sourcing with different style requirements.

This paper concludes that for single-sourcing to be feasible for article and whitepaper publishing:

- An open source content style manual needs to be developed to provide agreed trans-organisational mechanical editing standards.
- A greater emphasis must be placed on editing in the developmental stages of an article.
- The presentational format of an article can be largely automated at an organisational level, leading to improved efficiency and consistency within the publishing workflow.
DITA

The Darwin Information Typing Architecture (DITA) is an open source, XML-based architecture and methodology for designing, writing, managing, and publishing many kinds of information in print and on the Web. The “Darwin” in the name pays homage to Charles Darwin, the famed naturalist most responsible for the theory of evolution. DITA incorporates principles of specialisation, adaption and inheritance in document structures that are reminiscent of Darwinian theory. “Information Typing” refers to the focus on categorisation of information. “Architecture” indicates that DITA is not just an XML standard; it is an approach, a workflow, a methodology, and a philosophy. Although the first version of the DITA standard was only published in 2005, DITA is already being adopted by technical publication departments in large corporations, including Sun, IBM, Adobe and Nokia.

DITA is lauded as a highly flexible and capable single-source publishing architecture, allowing a range of content re-use options. DITA proponents claim that topics written in DITA by a supplier can be incorporated into a manufacturer’s technical publications. On initial examination, it also appears that a magazine article, written as a DITA topic, can be published to a number of Web sites and in a number of different printed publications such as magazines and newsletters.

Single-sourcing is an attractive ideal for publishers, because it can dramatically reduce the cost of writing, and maintaining the currency of, information content. An individual piece of text may be used in a dozen different publishing contexts (on a Web site, in a Help system, in a User Guide, in an Administrator Guide, in a product brochure, etc), but any edits to that text source need occur only once, with the changes flowing to all output publications. The principle in DITA which enables single-sourcing is that of separation of content and form through semantic mark-up.

In this article, we will test the single-sourcing claim in the context of magazine articles and whitepapers by introducing a case study, then identifying the challenges in achieving single-sourcing, and finally recommending an approach to editing to maximise the chances of effective re-use for future XML-based magazine article publishing.
Case Study

In November 2008, an article written by the author entitled *What if Readers Can't Read?* (Self, 2008), intended for publication in technical communication society publications and on industry Web sites. The article, based on a keynote presentation delivered at a Technical Communicators Association of New Zealand (TCANZ) event in New Zealand, was accepted for publication in the following online and traditional publications:

- *Southern Communicator* magazine (Quarter 1, 2009)
- HyperWrite Web Site (www.hyperwrite.com)
- WritersUA Web Site (www.writersua.com)
- STC *Intecom* magazine (February, 2009)
- STC France Web site (www.stcfrance.org)

The base DITA *content model* has three information types: concept, task and reference. The article was written as a DITA concept information type, in one topic. Prior to submission for publication, the article underwent a number of edits, including a third-party review.

*Intecom* magazine, a publication of the Society for Technical Communication (STC), requested that the article be submitted in Microsoft Word(R) format. The Word document was generated from the DITA concept topic using the DITA Open Toolkit (DITA OT). The DITA OT is an open source collection of software tools used by new DITA adopters. DITA adopters tend to migrate from the DITA OT to commercial tools as their documentation projects grow larger.

The Word document was reviewed by an STC Editor, whose changes are summarised in the following table.

<table>
<thead>
<tr>
<th>DITA Mark-up</th>
<th>Word(R)Output</th>
<th>STC Edits</th>
<th>Nature of Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;title&gt;What if readers can't read?&lt;/title&gt;</td>
<td>What if readers can't read?</td>
<td>What If Readers Can't Read?</td>
<td>Capitalisation</td>
</tr>
<tr>
<td>&lt;shortdesc&gt;This topic discusses...</td>
<td>This topic discusses...</td>
<td>This article discusses...</td>
<td>Wording</td>
</tr>
<tr>
<td>&lt;shortdesc&gt;... workers, and...</td>
<td>workers, and</td>
<td>workers and</td>
<td>Serial comma</td>
</tr>
<tr>
<td>resounding &lt;b&gt;Yes&lt;/b&gt;</td>
<td>Resounding Yes</td>
<td>Resounding &quot;Yes&quot;</td>
<td>Formatting</td>
</tr>
<tr>
<td>DITA Mark-up</td>
<td>Word®Output</td>
<td>STC Edits</td>
<td>Nature of Edit</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>&lt;p&gt;What evidence?...</td>
<td>What evidence?</td>
<td>What evidence do we have?</td>
<td>Writing Style</td>
</tr>
<tr>
<td>...with a Web edition...</td>
<td>With a Web edition</td>
<td>With a web edition</td>
<td>Capitalisation</td>
</tr>
<tr>
<td>A study by &lt;xref...&gt;Springer...</td>
<td>A study by Springer</td>
<td>A study by Springer (<a href="http://www">www</a>...)</td>
<td>Externalising links</td>
</tr>
<tr>
<td>...mentioned teaching...</td>
<td>mentioned teaching</td>
<td>mentioned using eBooks for teaching</td>
<td>Writing style</td>
</tr>
<tr>
<td>&lt;tm&gt;Facebook&lt;/tm&gt;</td>
<td>Facebook™</td>
<td>Facebook</td>
<td>House style</td>
</tr>
<tr>
<td>...wrote:/p&lt;/q&gt;Grammar rules used to be...</td>
<td>...wrote:Grammar rules used to be:</td>
<td>...wrote: &quot;grammar rules used to be...&quot;</td>
<td>Writing style, Punctuation</td>
</tr>
<tr>
<td>11%</td>
<td>11%</td>
<td>11 percent</td>
<td>House style</td>
</tr>
<tr>
<td>neighbours</td>
<td>neighbours</td>
<td>neighbors</td>
<td>AU to US spelling</td>
</tr>
<tr>
<td>texting. 2,000 messages...</td>
<td>texting. 2,000 messages...</td>
<td>texting. Two thousand messages...</td>
<td>House style</td>
</tr>
<tr>
<td>to realise that...</td>
<td>to realise that...</td>
<td>to realize that...</td>
<td>AU to US spelling</td>
</tr>
<tr>
<td>our audience &lt;term&gt;LCD&lt;/term&gt;(lowest common denominator)...</td>
<td>our audience LCD (lowest common denominator)...</td>
<td>our LCD audience...</td>
<td>Writing style</td>
</tr>
<tr>
<td>an article titled &lt;cite&gt;Hype Alert...</td>
<td>an article titled Hype Alert...</td>
<td>An article titled &quot;Hype Alert...&quot;</td>
<td>Formatting style</td>
</tr>
<tr>
<td>&quot;To google&quot; is now a verb.</td>
<td>&quot;To google&quot; is now a verb.</td>
<td>To &quot;google&quot; is now a verb.</td>
<td>Writing style</td>
</tr>
<tr>
<td>habits. The ABC...</td>
<td>habits. The ABC...</td>
<td>habits. ABC...</td>
<td>AU to US usage, and acronym usage</td>
</tr>
<tr>
<td>In one 3 minute and 51</td>
<td>In one 3 minute and</td>
<td>In one short (under 4</td>
<td>Writing style</td>
</tr>
<tr>
<td>DITA Mark-up</td>
<td>Word® Output</td>
<td>STC Edits</td>
<td>Nature of Edit</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>second video...</td>
<td>51 second video...</td>
<td>minutes) video...</td>
<td></td>
</tr>
</tbody>
</table>

**Separation of Content and Form**

The use of semantic mark-up in DITA, where text elements are marked up based on their meaning, allows the content to be completely separated from its rendition and display to the reader. For example, a term is marked up as a `<term>` and a citation as a `<cite>`, and no information about how those elements will be displayed is stored in the content. Stylistic (display) rules are applied when the DITA content is transformed into a reading format, such as HTML or paper. In a DITA workflow, documents are created as collections of modular, re-usable topic files, and mechanisms allow not only the format to be separated from the content, but the context also. The same topic may be a section in the context of one publication, but a sub-section in the context of another. The intermingling of content, format and context in a style-based document workflow essentially eliminates the possibility of re-use. Once a paragraph is styled as having a 13 cm left margin, it cannot be used on paper 12 cm wide. A phrase marked up in italic won't render as italic on a reading device that doesn't support italic. But a citation identified as a citation in a DITA topic can be processed to italic by one transformation process, to bold red by a different transformation process, and to synthesised voice by another transformation process.

Most popular styles guides haphazardly mix content (or *editorial*) style rules (eg, "use active voice for instructions") with presentational (*form*) style instructions (eg, "use italics for product names"). The pie-charts in *Figure 1 - Breakdown of Content (Editorial) and Presentational Style Rules in Four Style Guides* show a simplistic calculation of the proportion of content and presentation style rules in three common publishing industry style guides, and a contrasting newspaper style guide, based on the number of pages devoted to each type of rule.
Figure 1. Breakdown of Content (Editorial) and Presentational Style Rules in Four Style Guides

Unlike article publishing organisations such as STC, the separation of content and form in the newspaper publishing process, particularly at News Limited, is complete. Journalists submit stories to the news content management system without any presentational style information. Those stories (which are modular in nature) are single-sourced to Web editions of the paper, traditional paper editions, syndicated newspapers, database archives, and RSS XML feeds.

The Challenges of Editing in Single-Sourced Publications

According to the Freelance Editorial Association (n.d.), there are four different categories of editing:

- developmental (editing during the content creation process)
- substantive (improvements after the writer has completed the manuscript)
- copy (correcting errors and enforcing style manual rules)
- proofreading (checking the final version for typographical and layout errors).

Other sources categorise developmental and substantial as being synonyms, and some prefer the term line editing to copy editing. Some forms of editing, such as photo editing, largely involve the selection and rejection of alternative options.
The Chicago Manual of Style (1982) divides the editorial process into mechanical editing and substantive editing. Mechanical editing involves such matters as capitalisation, spelling, agreement of subjects and verbs, punctuation, and use of numbers. Substantive editing involves re-writing, re-organising, and reworking of the writing style to conform with the guidelines in a style manual.

The respected JPL publication The Levels of Edit (Van Buren and Buehler, 1980), has a more comprehensive editing typology, which provides more subtle distinctions in nine types of edits:

- Co-ordination: planning and preparing a manuscript for further editing
- Policy: ensuring the manuscript conforms to company policy (and document framework)
- Integrity: ensuring cross-references, page numbers and paragraph and list numbering is correct
- Screening: checking spelling, punctuation, missing material, and subject/verb matching
- Copy Clarification: identifying hyphenation points and marking the manuscript to denote typesetting requirements
- Format: specifying typographical and layout requirements
- Mechanical Style: ensuring the manuscript conforms to style manual rules
- Language: checking the spelling, expression, punctuation, syntax, fluency, parallelism, terminology and appropriateness of titles
- Substantive: deals with the coherence of the manuscript and the meaningfulness of the content.

Other editing functions not categorised may include reducing the length of a manuscript.

Because of its authority, for the purposes of this paper, we will use the Van Buren and Buehler editing typology: co-ordination, policy, integrity, screening, copy clarification, format, mechanical style, language and substantive. The STC edits can be organised into the editing types as follows.

Table 2. Organisation of STC Intecom Edits into Editing Typology Types

<table>
<thead>
<tr>
<th>Nature of Edit</th>
<th>Editing Typology Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capitalisation of headings</td>
<td>Mechanical Style</td>
</tr>
<tr>
<td>Nature of Edit</td>
<td>Editing Typology Types</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Wording</td>
<td>Substantive</td>
</tr>
<tr>
<td>Serial comma</td>
<td>Language</td>
</tr>
<tr>
<td>Formatting style</td>
<td>Format</td>
</tr>
<tr>
<td>Writing style</td>
<td>Substantive</td>
</tr>
<tr>
<td>Externalising links</td>
<td>Mechanical Style</td>
</tr>
<tr>
<td>House style</td>
<td>Mechanical Style</td>
</tr>
<tr>
<td>Punctuation</td>
<td>Language</td>
</tr>
<tr>
<td>AU to US spelling</td>
<td>Language</td>
</tr>
<tr>
<td>Acronym usage</td>
<td>Language</td>
</tr>
</tbody>
</table>

The edits fitted most comfortably into just four of the nine types: language, mechanical style, substantive, and format.

The Challenge of Content Re-use

In a modular DITA documentation workflow, documents are assembled from a pool of standalone topics, and those topics can be re-used in many different publications. *Write once and once only* (WOOO) is an underlying principle of modular documents; this approach results in lower creation and maintenance costs, and reduces the time required to produce different publications.
When DITA is used for magazine article or whitepaper content, the context suits a single topic rather than a multi-topic structure. Single-sourcing is therefore limited to producing different variations of output deliverable from the same source topic, as shown in Figure 3 – Non-Modular Single-Sourcing Schematic.

Figure 2. Modular Single-Sourcing Schematic

Figure 3. Non-Modular Single-Sourcing Schematic
If the main benefit of single-sourcing is the reduced cost of maintaining the content, does it really matter that some publication-specific changes are made prior to print publication? The answer to this question is probably "no, it doesn't matter", but the fact that virtually all printed publications are also published online does complicate the matter. Print journals such as *Intecom* are snapshots of what a document was at a point in time, while online content is often continually published, in that there is not a fixed publishing date, and the content may be updated at any time. Perhaps this was the wrong question, in any case, because DITA represents the future of publishing, whereas snapshot print publishing represents the past.

Future publishing models are likely to include continuous publishing to paper media, through devices such as the Espresso (*DADirect*, n.d.), and to non-paper media including E-Ink (*E-Ink Corporation Web Site*, n.d.).

The Challenge of Single-Sourcing with Different Style Requirements

A DITA workflow effectively separates the content of a document from its presentation, format and delivery ("the separation of content and form"). Authoring mark-up of the document is based on the semantics of the content elements, rather than how the content is to be presented to the reader. Publishing is largely an automated process, where DITA semantic mark-up is transformed, using software style rules, to a reading format.

The editing process in a DITA workflow is therefore simpler than that of a (traditional) style-based document workflow. However, the editing task that confirms that a document conforms to the requirements of a company's policies, or to the requirements of a style guide, is still required.

The delivery publications for the *What if Readers Can't Read* article have different requirements. For example, the *HyperWrite Web site* requires its content to be in Australian English, while the STC *Intecom* magazine requires its content to be in US English. A number of conflicting requirements of style between those two deliverables are shown in the table below.

<table>
<thead>
<tr>
<th><em>HyperWrite Web Site</em> Requirements</th>
<th><em>STC Intecom Magazine</em> Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence case for section headings</td>
<td>Title (<em>maximal</em>) case for section headings</td>
</tr>
<tr>
<td>Australian English spelling and words</td>
<td>US English spelling and words</td>
</tr>
<tr>
<td><strong>HyperWrite Web Site Requirements</strong></td>
<td><strong>STC Intecom Magazine Requirements</strong></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Inclusion of serial comma</td>
<td>Avoidance of serial comma</td>
</tr>
<tr>
<td>Sentence can start with numeric form of large number</td>
<td>Sentence should start with spelled form of large number</td>
</tr>
<tr>
<td>Percentage symbol used to express percentage fractions</td>
<td>&quot;percent&quot; (spelled) used to express percentage fractions</td>
</tr>
</tbody>
</table>

Some edits were subjective, in that the editor might prefer, for example, the phrasing "in one short (under 4 minutes) video" over the original "in one 3 minute and 51 second video". Other edits reflected the different requirements of online and print media, such as the need for hyperlink targets to be separated from the link text for print media.

The changes suggested by the Intecom editor make the article as published the Intecom different to the version published on the HyperWrite Web site. True single-sourcing would see the one article source master being used for all publications, not modified versions or derivations of that single source. Once an article is edited from its original source, any changes or updates will have to be repeated throughout any derivations, thus denying the cost-saving efficiencies of true single sourcing.

**Minimising Publication-Specific Edits**

For single-sourcing to be practised, the edits specific to a particular publication must be eliminated, or at best, reduced.

The Intecom edits made to the *What if Reader's Can't Read* article fitted into the following four editing typology types:

- language (spelling and punctuation changes)
- mechanical style (changes to conform to policies and style guides)
- substantive (improvements to meaningfulness, often subjective)
- format (typographical and layout presentational changes - media-specific).

Language edits are usually contentious only when the differences are cultural or dialectical, and such differences can be handled through either DITA or complementary XML technologies such as XML Localization Interchange File Format (XLIFF) (see Raya, 2004). Mechanical style edits are only problematic when different style guides are in conflict; this conflict may be overcome by the introduction of a common style guide, developed in an open...
source ideological framework. Ideally, subjective edits (that is, changes to suit the personal preferences of an individual editor) can be rejected, although any such diminishment of an editor's role is likely to invoke resistance. Format edit types for an individual article are not applicable to an automated publishing workflow, so such changes would not arise in a wholly DITA environment.

**Language (Spelling) Differences in a Single-Source Environment**

DITA includes a mechanism for marking up words or phrases (*inline elements*) for special processing in the publishing phase. For example, the word "realise" could be marked up as `<ph translate="yes">`; and then extracted (along with any other Australian English phrases) into a list of words to be translated into US English ("realize"). The translated phrases could then be re-injected into the document skeleton to form a translated document. (The XML technology makes this sort of extraction and re-injection of phrases a technically trivial matter.) For this approach to work, the document would need to pass through a mini-translation process, and the author would have to be sufficiently disciplined to remember to mark up such phrases. The publisher would also need XML extraction and translation software.

Currently, the labour cost of the mark-up and translation would probably exceed the cost of manually editing the content for each publishing output. Feature-rich, DITA-aware XML tools are not yet readily available, but when such tools reach the market, the reduced mark-up costs would likely make this mini-translation approach cost-effective.

Another option would be to use DITA’s content referencing mechanism to place contentious words in a separate file, *transclude* those phrases into the article topic, and then provide two different versions of the separate file for different language versions of the article.
As well as spelling changes, colloquial wording changes might be necessary for different readerships. For example, in the USA, the octothorpe symbol (#) is colloquially described as a pound sign, while in New Zealand the same symbol is described as a hash symbol, and in the UK as the square. It would be optimistic to expect an author to be familiar with the colloquial nature of many words. In some cases, word use changes from region to region within the same country, such as the Queenslander Australian use of the word "scallop" to mean what Victorian Australians refer to as a "potato cake". Dialectical differences are not a new problem, however, and controlled language (such as Simplified Technical English, ASD-STE100) is normally used to minimise confusion in terms in technical documents (Muegge, n.d.).

Mechanical Style (Writing and Wording Style) Differences in a Single-Source Environment

Different style manual rules often conflict, so an article written to one style guide may break the rules of another.

For example, the Manual of Style (Snooks & all) calls for the separator in numbers greater than three digits to be a space. The What if Readers Can't Read article uses a comma as a separator in the number "2,000". The Style Book (Wallace and Hughes, 1995) suggests (but does not directly mandate) the use of a comma as a separator. The Chicago Manual of Style (1982) dictates a comma separator for general numbers, but a space separator for "scientific copy". Although it is possible that numbers could be semantically identified in DITA, and then processed according to the particular style guide rules, this is likely to be impractical.

It seems, therefore, that the conflicting rules of style manuals represent an obstacle on the path to DITA single-sourcing across organisations with different style manuals. If all publications used a common style manual, the obstacle would dissolve.

Substantive Edits (Subjective Wording Changes) in a Single-Source Environment

If one editor prefers a particular phrasing of a sentence, but another prefers an alternative, single-sourcing cannot practically occur. A poor workaround is to hold off subjective changes until the latest possible point in the publication process, and then repeat those changes if or whenever the source document changes.
Editors are likely to resist any restrictions on their ability to make subjective changes. Interestingly, the problem is not with subjective edits as such, but with conflicting subjective edits.

**Format Edits (Media-Specific Changes) in a Single-Source Environment**

Any special requirements of a particular media can be easily catered for by changes in the transformation process that generates deliverable output documents from DITA source files. The issues of differing media requirements can be addressed by improving the transformation processing. The *What if Readers Can't Read* article was transformed to Word format for submission to the STC *Intecom* magazine using the standard DITA OT DITA-to-RTF rules. These rules could have been customised to format DITA cross-references so that the text of the link was unadorned (no underline, no colour) and the link target (a Web URL) written out after the link text and enclosed in brackets.

**The Same Songsheet**

It might be stating the obvious, but for single-sourcing with DITA to be effective, all parties to the single-sourcing have to be DITA or XML environments. Currently, very few article publishing organisations are using DITA, although many are using XML. The *Australian Journal of Communication* is a typical publisher of academic papers. Its submission requirements are:

“contributions should conform to current APA documentation style and to guidelines for inclusive language. Contributions should also be accompanied by a 50-100 word abstract. Contributors whose papers are accepted for publication will be required to submit a final copy of their paper, in hard copy and via e-mail to <e-mail address>. Diagrams and tables should be submitted in a separate MS Word document.”

Most of the efficiencies in moving to a DITA workflow will be lost if Microsoft Word(R) is a key part of the publishing cycle. Single-sourcing in a Word or similar non-structured, style-focussed technology is largely hypothetical, because a non-structured document technology does not lend itself to content re-use through the separation of content and form. The importance of single-sourcing will arise in the future when publishing is universally XML-based (see Kulik and Nigloschy, 2003). (This paper is concerned with those future scenarios.)
The separation of content and form in a DITA process must be matched by a separation of content rules and form rules that define a publication's style.

Assuming that all parties in a future publishing scenario were working in an XML environment, or a DITA environment in particular, interchangeability will grow in importance. At the centre of interchangeability practice is standards; it is logical to conclude that single-sourcing in a DITA environment will rely on a number of standards, including editing and writing style standards.

DITA is not yet common in a publishing workflow, but paper-centric tools are. Technological changes such as E-Ink devices, on-demand paper publishing and printing (through equipment such as the Espresso "book ATM"), digital distribution of books and magazines, and multimedia e-books (see Elgan, n.d.) will drive the demand for smarter and more appropriate publishing solutions, such as DITA-based workflows.

**Recommendations**

In the case study, four types of edits that impede the realisation of single-sourcing in a future DITA-based publishing environment were identified. The impediments can be overcome through the following changes to editing practice.

Table 4. Recommendations for Changes to Editing Practice

<table>
<thead>
<tr>
<th>Edit Type</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Edits</td>
<td>Use XLIFF or transclusion to cater for different spelling and word usage in different outputs. The option of forcing a controlled vocabulary on authors of magazine articles should not be adopted, as it is likely to unreasonably constrict creativity and result in bland, uninteresting documents.</td>
</tr>
<tr>
<td>Mechanical Style Edits</td>
<td>Eliminate publication-specific changes by adopting a standard, open source, trans-organisational content style manual, and restrict corporate style guides to issues relating to formatting presentation. (Refer to Figure 6 – Possible Post-DITA Style Guide Cascade.)</td>
</tr>
<tr>
<td>Edit Type</td>
<td>Recommendation</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Substantive Edits</td>
<td>Remove the substantive editing task and place a greater emphasis on the developmental editing stage. Abandon post-developmental subjective editing changes.</td>
</tr>
<tr>
<td>Format Edits</td>
<td>Address format edit changes at the transformation rules stage, at an organisational level rather than at the article or publication level.</td>
</tr>
</tbody>
</table>

Publishers in a DITA workflow should use custom XSL transformation routines, rather than the default DITA OT transformers, to ensure that format *edit types* are not required beyond a one-off verification of the publication style rules after the transformer is created. This would address issues such as hyperlink references in printed output. Provided the link target and the link text are semantically identified in the source, references can be created to suit the publication's individual formatting and layout preferences without affecting the XHTML output.

The most important and significant of the above recommendations is to adopt a standard, open source, trans-organisational, international content style manual.

**Open Source Standard Content Style Manual**

An editing process uses a number of authorities for establishing the rules for the wording, structure and presentation of documents. Rules are built through a *cascade* of authorities. For example, a typical style ruleset might require spelling to conform to the *Macquarie Dictionary*, except when a different spelling is recommended by the *Microsoft Manual of Style, 2004*. A typical style rules cascade is depicted in Figure 5 - Typical Existing Style Guide Cascade.
As shown in Figure 1 - Breakdown of Content (Editorial) and Presentational Style Rules in Four Style Guides, most style manuals used in article publishing mix form (presentational style) rules with content (writing and editorial style) rules, making them difficult to interpret in a semantic mark-up authoring environment.

To improve the prospects of content re-use, an agreed standard or style manual for formalising editing decision-making is vital. Open source development is a proven approach to the creation of trans-organisational and international standards (Weber: 2005). An open source style manual might fit into a style rules cascade as shown in Figure 6 – Possible Post-DITA Style Guide Cascade. The standard style manual should be concerned only with writing style (ie, content), and not with presentational style (ie, form).
Necessary Changes to Editing Typology

The separation of content and form in the publishing process, the adoption of an open source style manual, the increasing importance of continuous publishing, and the use of automated document generation processes will require a re-thinking of the Van Buren and Buehler editing typology. The schematic diagram Figure 7 - Editing in Structured and Unstructured Publishing Models illustrates a comparison between the editing types in a DITA publishing workflow and existing style-based publishing workflow.
Any attempt to standardise punctuation (such as serial commas) and even spelling is bound to meet dogged resistance, as it is natural for people to protect their own language use and conventions. The creation of a standard style manual that crosses organisational culture and national cultural boundaries is not going to be without major challenges. It is important, therefore, that provision for dialectical and spelling differences through a mini-translation XLIFF process be developed in parallel with a standard style manual.

**Further Research**

Further research is required to analyse the acceptance of changes in editing practice amongst editing professionals, and what steps would need to be taken to develop and manage an open source style manual for single-source publishing.
References


