

Moving to XML: The Investment

B. Tommie Usdin

Mulberry Technologies Inc.

17 West Jefferson St.

Suite 207

Rockville MD 20850

Phone: 301/315-9631

Fax: 301/315-8285

info@mulberrytech.com

<http://www.mulberrytech.com>

Version 1.0 (January 2006)

©2006 Mulberry Technologies, Inc.



Moving to XML: The Investment

| | |
|--|-----------|
| Moving to XML: The Investment..... | 1 |
| Content Providers and Vendors Investment..... | 1 |
| XML Gives Both Content Providers and Vendors..... | 2 |
| Content Provider Investment in XML..... | 2 |
| XML Will Change the Way Publishers Work..... | 3 |
| Added Value Means Added Work..... | 3 |
| Costs and Benefits Not Equitable..... | 4 |
| Service Vendor Investment in XML..... | 4 |
| So Service Providers Need to..... | 5 |
| Vendors Must..... | 5 |
| Placing XML in the Workflow..... | 6 |
| When is the Data Tagged..... | 6 |
| XML Workflow Means Changes in Staffing/Jobs..... | 7 |
| Warning: XML Does <i>Not</i> Reduce Staff..... | 8 |
| What it Takes to Make XML happen..... | 8 |
| Select (or Create) a Model..... | 9 |
| Basic XML Skills For Everyone..... | 9 |
| Tagging Knowledge Varies for Different Models..... | 10 |
| XML Management Training..... | 10 |
| XML Wrangler Training..... | 11 |
| Prepress Skill Set Includes..... | 11 |
| DTD and Schema Skills..... | 12 |
| <i>New</i> Data Manipulation Skills..... | 12 |
| Manipulating XML is Prepress Life Blood..... | 13 |
| Parsing and Delivering “Clean” XML..... | 13 |
| Publishers Can Use Transformation Skills Too..... | 14 |
| New Tools and Tool Training..... | 14 |
| XML Validation Tools..... | 15 |
| XML-Aware Editors..... | 15 |
| Transformation Language and Processors..... | 16 |

Moving to XML: The Investment

A New Option: Making Print Directly From XML.....16
Repositories/Content Management Systems.....17

Conclusion

The Bad News: There is no Free Lunch.....17
The Good News: You Can Do XML and Benefit.....18
Content Providers Get the Benefits They Want.....18
Conversion Vendors Would Make a Killing.....18
There are Real XML Benefits for Page Production.....19
Colophon.....19

Moving to XML: The Investment

slide 1

Moving to XML: The Investment

- Many benefits to XML
- Doesn't come for free
- Changes to
 - Work flow
 - Staff skills
 - Balance of power
- New opportunities

slide 2

Two (Different) Investments in XML

- “Moving to XML” means very different things to
 - Content owner / content provider (publisher)
 - Service vendor (printer, typesetter, data conversion vendor)
- The two viewpoints have
 - very different responsibilities / roles
 - different first steps and workflow
 - different software (some overlap)
 - similar validation strategies/techniques

XML Gives Both Content Providers and Vendors

- Some similarity in production advantages/gains
- Mostly different problems
- Similar skill sets/training requirements (some overlap)
- New ways to do what they always did
- New things to do

Content Provider Investment in XML

Publisher must decide:

- What the XML *is for*
 - business case (advantages)
 - ROI
 - what to do with the XML
- What does the XML look like (tag set and validation rules)
- How XML will be created
- How to store/manage the XML
- Where XML falls in the production workflow

XML Will Change the Way Publishers Work

- Line between authoring, content edit, and copy edit shifts/blurs
- Proofing and checking changes
 - level changes (less worry about transposed letter and more about missing structures)
 - specifics change (don't look for comma, look for <author>)
 - new QA possibilities (like false-color proofs)
 - one error may show many times but still be just *one* error
- Multiple vendors may use one file
- Machines do some work once done by people
People do some work never done before

Added Value Means Added Work

- Added metadata— Repository storage and rights/permissions may need metadata that does not print or display
- Semantic tagging— Value-added tags for searching may require subject knowledge
- Index terms — Adding inline indexing is the same huge effort as a back-of-the-book index
- Writing changes — If structures are to be used in parallel ways; they must be parallel

Costs and Benefits Not Equitable

(Warning: Costs in budget of Department A, while benefits accrue to Departments C and D or to the company as a whole)

- Some groups will have increased workload, for others production is faster and easier
- Expense and hard work are immediate (group-specific)
- Most benefits
 - Increase the bottom line for a different group
 - Are long term
 - Are corporate-wide or related to sales
 - Are cumulative; One job takes 2 months longer, but other jobs are done in half the time, are much cheaper, or are only possible now)

Service Vendor Investment in XML

(it's not your data, it's your service)

- If Content Provider has XML content, they will already have DTDs or schemas (not necessarily appropriate for publication)
- They will expect you to
 - accept XML source (minimally)
 - work in XML (better)
 - work in *their* XML
- They will expect a compositor/typesetter to Provide round-trip services (even when it is not clear *if* that is possible)

So Service Providers Need to

- Be able to produce backend XML
 - from a backfile of hardcopy pages
 - from a backfile of computer files
 - in addition to pages or websites they make
- Be able to receive customer-specific XML and
 - produce pages
 - produce websites
 - return it (updated and *uninjured*) when through (delivery of cleanly parsing, correctly tagged material)

Vendors Must

- Work with a variety of XML tag sets
 - customers have varying requirements / XML tagsets
 - vendors must use many and transform more
- Work with good, bad and indifferent tagging, therefore know
 - the difference
 - when to ask for changes
 - when to offer help
 - how to fix problems

Conclusion: Vendors need better XML skills than content providers

Placing XML in the Workflow

(Both Publisher and Vendor may be Involved)

- How XML will be created
 - when in the workflow
 - by whom (inhouse or outsourced to vendor)
 - how will newly created XML be validated
- How to store/manage the XML
 - content management system
 - database (XML native or not)
 - file system to manage documents
- Using XML for production and QA
- What products made from the XML (and how)

When is the Data Tagged

- XML on the front end
 - During creation/authoring
 - Between author and any edit
- XML as a production step
 - as part of editing
 - before page production
 - as part of page production
- Post Production (XML on the back end)
 - as extension to process (for example by compositor before return to publisher)
 - when placed into repository
 - long after production, possibly by years

XML Brings New Proofing/QA Possibilities

More intelligent proofing than “Does it look right?”

- List any element / combination
- Print content of reference next to place where reference is made
- Bibliographic references checked and made live before publication
- False-color proofs
 - Add color to make things stand out
 - Add numbering
 - Add links to what needs to be checked

XML Workflow Means Changes in Staffing/Jobs

- Much rekeying eliminated
- Proofing/rekeying cycles shortened or eliminated (exception *math!* and *chemistry*)
- Changes the nature of the grunt work
- More time devoted to content instead of format
- Transformation becomes a major activity
 - More post processors
 - More potential output products

Warning: XML Does *Not* Reduce Staff

- XML can increase
 - Accuracy
 - Opportunities in sales and new products
 - Customer satisfaction
 - Timeliness (reduce time to market)
- But not likely to be done with fewer people

(Not likely to increase staff dramatically either, the few can do more!)

What it Takes to Make XML happen

- Tag set (DTDs and schemas)
- Skills and training
 - basic
 - wrangler
 - prepress
- Software

Select (or Create) a Model

(DTD or schemas)

- Sources of tagsets/models
 - Off-the-shelf
 - Publicly-available
 - Government, industry, interest area, corporate groups, standards bodies
 - Altered-to-fit
 - Started with off-the-shelf model
 - Modified to meet organization needs
 - Designed and developed for an organization's specific requirements

(Good documentation is the key to success)

Basic XML Training for All XML Users

- Principles of generic markup and document modeling
- Tagging basics
 - Format-specific vs. generic
 - Structure vs. content
 - Logical structure of documents
- How to handle XML
 - how to modify and create new tagged data
 - how to apply a document model
 - use of XML software tools
 - parsing and other validation error resolution

Tagging Knowledge Varies for Different Models

- For all taggers and validators, basic XML concepts
- For structural tagging
 - good language skills (in the appropriate language)
 - knowledge of how your documents work
- For bibliographic tagging, reference knowledge (how to tell a journal from a book from a thesis)
- For subject tagging (legal, medical, pharmaceutical)
 - subject-matter expertise
 - user expectations
 - publisher budget

XML Management Training

For anyone approving budgets, buying tools, developing production schedules or goals, or managing the process

- Basic concepts and vocabulary
- Costs and scheduling
- Risk and opportunities
- Production concerns and personnel factors
- An understanding of how capacity and opportunity have changed

XML Wrangler Training

- XML authoring
- Tagging or conversion from word-processing
- Checking and fixing XML documents
- Distinguishing *tagset* problems from *tagging* problems (and fixing both)
- DTD/schema skills
 - read XML models
 - teach models to vendors and new employees
 - maintain tagset and documentation
- Reading and writing typesetting and conversion specifications

Prepress Skill Set Includes

- Produce customer-specific XML
- Check and fix XML documents, models
- Flow XML into current tools
 - QuarkXPress, InDesign, etc.
 - Microsoft Word and other word-processors
- Extract XML from current tools
 - QuarkXPress, InDesign, etc.
 - Microsoft Word and other word-processors
 - and then that XML *real* XML
- Handle customer expectations of round-tripping

DTD and Schema Skills

- Identify conflicts between the tagset and the content
- Modify a DTD or schema
 - to solve production problems
 - without killing downstream processing
 - versioning and configuration management
- Cheat the markup to resolve production problems (during a production crisis)

New Data Manipulation Skills

XML customers believe the XML hype

- Assume that if you have tags you can work wonders
- Assume you have skills
 - beyond theirs
 - that your old customers did not assume you had

Manipulating XML is Prepress Life Blood

- The tasks
 - XML extraction from other files
 - Transforming XML tags to typesetting driver code (multiple outputs)
 - Rearranging content (into publication order)
 - Inserting generated text (chapter numbers, footnote symbols, automatic Table of Contents)
 - Adding metadata from other sources
 - Vendors do QA too (they get paid for quality)
- The skill set
 - XML transformation
 - string manipulation

Parsing and Delivering “Clean” XML

- You would think is basic
- But it may not be possible
 - document model requires elements you cannot know
 - publication date
 - DOI
 - full name of author (you have initials)
- Welcome to creative tag abuse
 - `<doi>XXX DOI here when assigned XXX</doi>`
 - `<pub-date>XXXXXX</pub-date>`
- These *will* happen so work out preferred solutions with your publisher (*in advance!*)

Publishers Can Use Transformation Skills Too

- Reports, checklists, false-color proofs for QA
 - which figures do not have captions?
 - what are the key terms in each section?
 - any footnotes *not* referenced?
 - make all the surname pink (for human checking)
 - collect all abbreviated journal titles (software check with authority file)
- Almost-instant galleys or editors' proofs
- Generated text is managed separately:

(And publishers will need one or two XML “gurus”)

New Tools and Tool Training

(software costs are not the big issue*)

- Validation tools
- Creation and editing applications
- Transformation language and engine(s)
- Software to make print pages and webpages
- New delivery applications

*Except maybe Content Management Systems

XML Validation Tools

(validation is central to all human-created XML processing)

- XML “parser”
 - checks an XML document against a DTD or schema
 - don't *buy* these, they are free or built in
- Schematron processors are free
- Validation beyond DTDs and schemas is typically programming (Perl, C++, .Net, javascript, etc.)

(Remember that an XML document is just an editable file)

XML-Aware Editors

- Like an XML word-processor
- Understands tags and knows your model
- Typical features
 - styled representation (font, colors, forms-entry)
 - choose to view tags or not
 - DTDs and schemas guide authoring:
 - context-sensitive tag choices
 - validation
 - attribute editing
 - pick lists and shortcuts

Transformation Language and Processors

- The need is to
 - pour XML into desktop packages
 - select, excise, rearrange for reuse and repurposing
 - rework XML for QA
 - make HTML for delivery
- Solution = a transformation language and software
 - standalone transformer (XSLT engines are free)
 - built into developers environments, XML editors, GUI application-generators
 - commercial packages (OmniMark)

A New Option: Making Print Directly From XML

Using XSL-FO (Extensible Stylesheet Language-Formatting Objects)

- A way to make PDF, Postscript, RTF, TeX, etc. directly from XML
- Application-independent vocabulary for specifying
 - text formatting
 - page-layout
- Special-purpose XSL-FO rendering software
 - takes XML as input
 - gives PDF, Postscript (RTF, MIF, etc.) as output

(Typesetters understand XSL-FO better than do most people who work with XML)

Repositories/Content Management Systems

- Will you need one
 - publishers probably/possibly
 - prepress vendors probably not
- There are many things called CMS (Content Management Systems)
- Free advice:
 - start without one
 - determine your real checkin, checkout, workflow, routing, reuse requirements
 - with your requirements known, then look at a lots of them

Conclusion

The Bad News: There is no Free Lunch

- Training is required
- Added value (that's *richer tagging*) means more work
- Just because it's XML doesn't mean it's good
- The benefits of XML are available only to those who
 - Plan
 - Play by the rules
 - Work hard

The Good News: You Can Do XML and Benefit

- Training and books are available
- Many good tools are free, many others are affordable
- Service vendors rapidly becoming XML savvy
- The XML marketplace is growing fast
- Users of electronic products more sophisticated every day, demanding better electronic products

Content Providers Get the Benefits They Want

- Re-use and repurpose content
- Internationalization, localization, customization
- Quality checking and consistency
 - does every chapter have a title
 - are all the cited works referenced
- Print and screen publications publish at the same time
- Vendor and software independence

Conversion Vendors Would Make a Killing

If there were not so many of them

There are Real XML Benefits for Page Production

- Content is in a single, familiar form:
 - few structural surprises
 - no unique and one-time-only styling
- Content is cleaner when typesetting starts
- Multiple outputs produced simultaneously from one source
- Pre-processing reports on what the document contains:
 - do we have a graphic for every map?
 - how many tables are there?
 - are there equations?
- Generated content can be error-free

Colophon

- Slides and handouts created from single XML source
- Slides projected from HTML which was created from XML using XSLT
- Handouts created from XML:
 - Source XML transformed to Open Office XML
 - Open Office XML opened in Open Office
 - Pagination normally adjusted
 - Saved as PDF
- Slideshow materials available at:
<http://www.mulberrytech.com/slideshow>