

# How and Why Are Companies Using XML?

**B. Tommie Usdin**  
**Mulberry Technologies, Inc.**  
17 West Jefferson Street, Suite 207  
Rockville, MD 20850  
Phone: 301/315-9631  
Fax: 301/315-8285  
btusdin@mulberrytech.com  
<http://www.mulberrytech.com>

Version 1.0 (January 2006)



**Mulberry  
Technologies, Inc.**

©2006 Mulberry Technologies, Inc.



# How and Why Are Companies Using XML?

<b>XML is Everywhere</b> .....	1
XML as Document Content is Different.....	1
Multi-Use is the Key.....	1
Printers, Compositors, Designers It's <i>Not</i> about <i>You!</i> .....	2
You Have Options.....	2
Publisher's View of an XML System.....	3
XML Features Appeal to Business Needs.....	3
<b>The Business Case</b>	
Why a Business Wants XML in Publishing.....	4
Benefits of XML (in general).....	4
Benefits for XML (for Documents).....	6
<b>The Dream</b> .....	6
The Dream can be Partially Realized.....	6
<b>Dream: Repurpose and Reuse</b> .....	7
We Still Print Textbooks.....	7
Textbooks May Have Instructor's Manuals.....	8
We also Want E-Textbooks.....	9
We Want to Make All Those Products.....	9
Case Study: Repurpose and Reuse.....	10
<b>Dream: New products — Mix and Match Existing Content</b> .....	10
Case Study: Central Repository.....	11
Case Study (continued).....	11
<b>Dream: Reduce Production Time and Improve Quality</b> .....	12
Case Study: Improve Quality.....	13
<b>Dream: Switch Software and Service Vendors Any Time</b> .....	13
This Dream Has Come True.....	14
<b>Dream: Pages — As Good as Current Pages — Automatically</b> .....	14
XML Can Feed Composition Systems.....	15
XSL-FO Pages Sometimes OK.....	15

***How and Why Are Companies Using XML?***

<b>Dream: Value-added Electronic Products — Automatically</b> .....	15
For the Printed Journal.....	16
For the Website.....	17
For an Abstracting/Indexing Service.....	18
Case Study: Select from Module Library.....	19
Case Study: Justification/background Hidden in XML.....	20
Case Study: Translation Means Translation (Not Typesetting).....	20
<b>The Promise of XML</b> .....	22
XML to A Service Provider.....	22
Colophon.....	22

# How and Why Are Companies Using XML?

---

*slide 1*

---

## XML is Everywhere

- In some circles, XML Web Services are all the rage
- Bank transactions are in XML
- e-Commerce and e-Business happen in XML
- Digital cameras create XML headers on images
- Printers use XML for job control
- State troopers record traffic warrants in XML

But that's not what we are talking about here

---

*slide 2*

---

## XML as Document Content is Different

- It isn't about format
- It isn't for the convenience of printers, compositors, or designers
- It doesn't make any (one) thing easier
- It makes many things more difficult

Content Creators and Publishers want it — for their own reasons

---

*slide 3*

---

## Multi-Use is the Key

There is always a

- cheaper
- faster
- easier

way to do **any one thing** than by using XML



## **Printers, Compositors, Designers It's *Not* about *You!***

It *is* about publishers

- they think it's "their" content
- they want
  - to use it, re-use it, slice it, and dice it
  - to own it and control it
  - to have access to it and be able to move it

## **You Have Options**

You can

- Provide the XML services more and more customers want, or
- Watch your customer base shrink

You can:

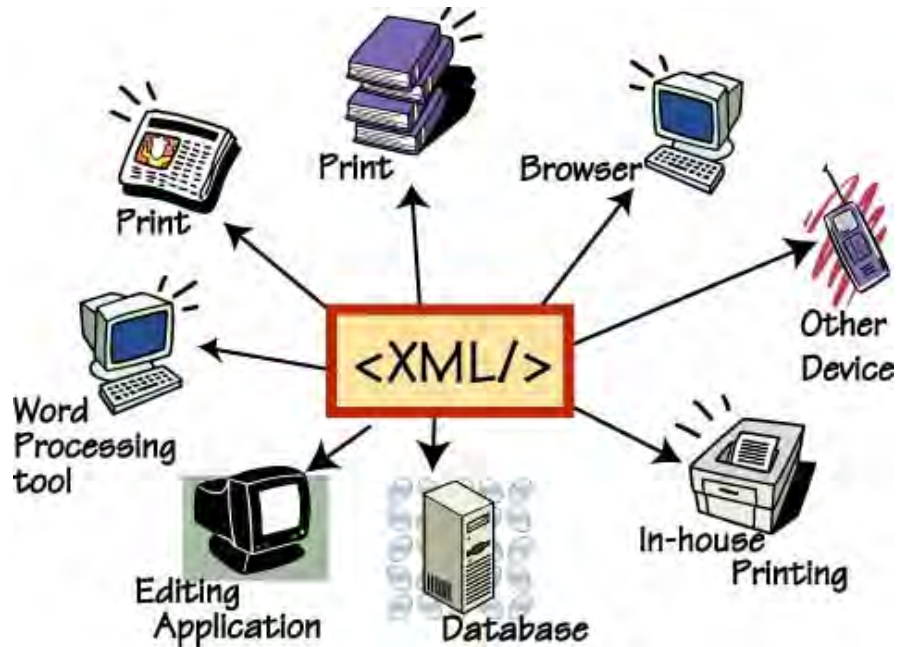
- Learn to work with XML smoothly and easily, or
- Fight XML tooth and nail

You can:

- Use XML content to make some of your processes easier
- Let XML be an added step, added expense, and continual nuisance

You can't make XML go away!

## Publisher's View of an XML System



## XML Features Appeal to Business Needs

- Platform- and vendor-independent
- ASCII/Unicode
- Public standard
- Control of the data format
- Separation of content from format
- Validation (document model)
- Computer-manipulable (and human readable)

---

## The Business Case

---

slide 8

### Why a Business Wants XML in Publishing

- Content re-use
  - repurposing and new products
  - customization and internationalization
  - multiple products from one source
- Smoother handoffs
- New processes
- Protect content investment

Flexible. Fluid. Single-source. Dynamic. Personalized. Interactive.

---

slide 9

### Benefits of XML (in general)

*(from Software AG, a vendor of XML tools)*

- ***Simplicity***  
Information coded in XML is easy to read and understand, plus it can be processed easily by computers.
- ***Openness***  
XML is a W3C standard, endorsed by software industry market leaders.
- ***Extensibility***  
There is no fixed set of tags.  
New tags can be created as they are needed.
- ***Self-description***  
XML documents can be stored without [schemas] because they contain meta data; any XML tag can possess an unlimited number of attributes such as author or version.



- ***Contains machine-readable context information***  
Tags, attributes and element structure provide context information ... opening up new possibilities for highly efficient search engines, intelligent data mining, agents, etc.
- ***Separates content from presentation***  
XML tags describe meaning not presentation.  
The look and feel of an XML document can be controlled by XSL style sheets, allowing the look of a document (or of a complete Web site) to be changed without touching the content of the document.  
Multiple views or presentations of the same content are easily rendered.
- ***Supports multilingual documents and Unicode***  
This is important for the internationalization of applications.
- ***Facilitates the comparison and aggregation of data***  
The tree structure of XML documents allows documents to be compared and aggregated efficiently element by element.
- ***Can embed multiple data types***  
XML documents can contain any possible data type — from multimedia data (image, sound, video) to active components (Java applets, ActiveX).
- ***Can embed existing data***  
Mapping existing data structures like file systems or relational databases to XML is simple....
- ***Provides a “one-server view” for distributed data***  
XML documents can consist of nested elements that are distributed over multiple remote servers. XML is currently the most sophisticated format for distributed data — the World Wide Web can be seen as one huge XML database.
- ***Rapid adoption by industry***  
Software AG, IBM, Sun, Microsoft, Netscape, DataChannel, SAP ...

## **Benefits for XML (for Documents)**

- Some of that list doesn't apply to documents
- Some of it CAN SOMETIMES apply to documents
- Some it is ABOUT documents

And there other advantages end users want from XML document content

---

## **The Dream**

- New products: mix and match existing content
- Reduce production time and improve quality
- Switch software and service vendors any time
- Automatically make pages — as good as current pages
- Value-added electronic products — automatically

without added cost, without substantial investment,  
without disrupting current processes

---

## **The Dream can be Partially Realized**

- There is no magic
- Getting benefits from XML requires investment
- XML doesn't replace skilled people

(Most of the promise is true for some organizations)

## Dream: Repurpose and Reuse

- Print is not enough any more
- Single-use data is too expensive
- Information is a corporate resource and must be managed accordingly
- If we can't get our data out, we don't want it in
- Web design and print design should be different

## We Still Print Textbooks

<b>Chapter 6: Classification</b>	<b>Page 55</b>
<b>6.9 Compounds</b>	
<b>compound</b>	A <b>compound</b> is a substance containing at least two elements combined chemically in definite proportions by mass. A compound can be chemically broken up into its constituent elements or simpler compounds. There are two types of compounds, <i>ionic</i> and <i>molecular</i> .
<b>ion</b>	An <b>ion</b> (pronounced <i>eye-on</i> ) is an atom or group of atoms that is positively or negatively charged. A negatively charged ion is an <b>anion</b> (pronounced <i>an-eye-on</i> ) while a positively charged ion is a <b>cation</b> (pronounced <i>cat-eye-on</i> ). An <b>ionic compound</b> is a compound that is held together by the attractive forces between positively and negatively charged ions.

## Textbooks May Have Instructor's Manuals

Jeremy's Chemistry  
Chapter 6: Classification

INSTRUCTOR GUIDE  
Page 55

### 6.9 Compounds

#### compound

A **compound** is a substance containing at least two elements combined chemically in definite proportions by mass. A compound can be chemically broken up into its constituent elements or simpler compounds. There are two types of compounds, *ionic* and *molecular*.

Testbank GDW  
6, 7, 9, 54

#### ion

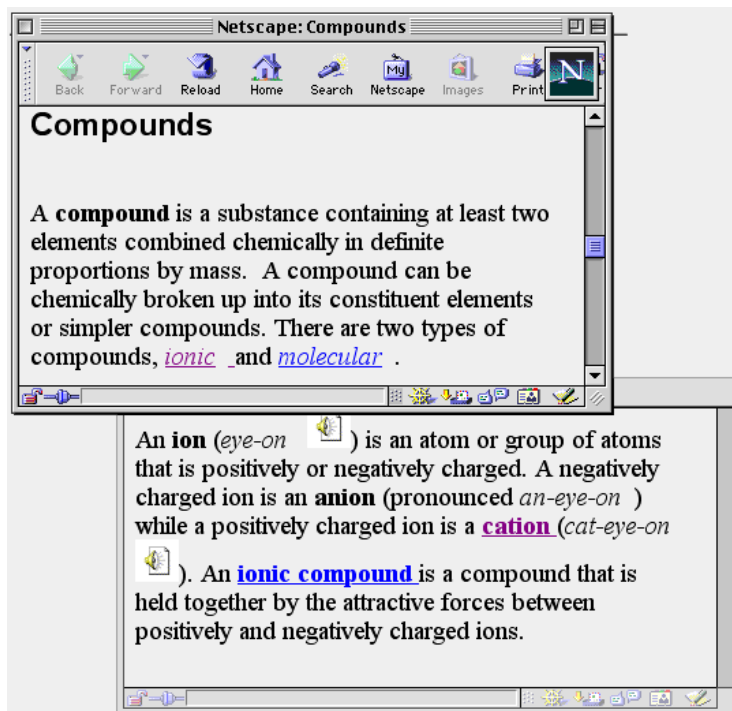
An **ion** (pronounced *eye-on*) is an atom or group of atoms that is positively or negatively charged. A negatively charged ion is an **anion** (pronounced *an-eye-on*) while a positively charged ion is a **cation** (pronounced *cat-eye-on*). An **ionic compound** is a compound that is held together by the attractive forces between positively and negatively charged ions.

Testbank GDW  
**ionic compounds**  
GDW 6, 7

**cations**  
GDW 9

**anions**  
GDW 25,26

## We also Want E-Textbooks



## We Want to Make All Those Products

- From the same source file (no parallel maintenance)
- By flowing content into pre-setup layouts (minimal designer slowdown)
- So that web does not lag print (or get out of sync)
- With the ability to add web-only features (like the pronunciations, animations, etc.)

## **Case Study: Repurpose and Reuse**

Publisher of Medical Reference Books

- Large, complex print publication
- Large, complex electronic products
- Syndication of varying amounts of content
- Subset publications
  - drugs/procedures relating to specific populations
  - drugs/procedures relating to specific diseases
  - information for specific groups
    - physicians
    - nurses
    - patients

**Dream:**

## **New products — Mix and Match Existing Content**

- Making coherent publications by slicing and dicing requires significant editorial preparation
- XML content can make it easy to
  - make anthologies by selecting from here and there
  - make sub-set publications
  - make alternative presentation formats (large print, voice synthesis, web, and print)
  - make improved navigation and discovery tools (RSS, ATOM, enhanced indexes, active ToCs, live references, post-publication references)
  - just-in-time merge from a form plus a database

## **Case Study: Central Repository**

A big North American publisher — one that has absorbed many of its competitors and now has many divisions, subdivisions, departments with overlapping product lines and markets.

- Content from many divisions/departments
- All content converted to same tag set (for repository)
- Search across all company content
- All company content available for re-use throughout company
- Many suppliers, many processes, all produce same XML

## **Case Study (continued)**

Based on XML repository, publisher sees:

- Some success in re-using content in other publications
- Some integrated "bookshelf" or "library" electronic products
- Significantly increased content syndication;  
sale of content to other publishers

**Dream:**

**Reduce Production Time and Improve Quality**

- Eliminate parallel creation and update
- Lights-out publishing (e.g., invoices, medical records, catalogs)
- Validation finds many surprises early
- Automate tedious and repetitive handwork
- New proofing and checking methods (lists, false color)
- Format automated, so
  - authors/copy-edit don't work on format, just content
  - consistency of formatting look and feel
  - virtually eliminates "check that every X is formatted as Y"
  - generated text (autonumbering, "Figure", punctuation)
  - cross references and citations checked



## **Case Study: Improve Quality**

A publisher of scholarly journals

- Quality requirements
  - discovery information — to allow scholars to find their articles
  - linking information — live links from references to cited works
  - forward citations — who is linking to our material
- Challenges to consistency
  - many editors, many citation styles, many different journal styles
  - many service vendors; printers in many countries
- Tools to ensure consistency and quality
  - XML vocabulary — DTD and detailed tag set documentation
  - validation tools
    - XML must be valid
    - additional checks for unlikely content
  - manual QA on random articles

Quality & discoverability enhance journal reputation, thus subscriptions

**Dream:**

## **Switch Software and Service Vendors Any Time**

- Content not tied into proprietary software
- Content moves at publisher's whim
- Content-provider investment in training carries over

## **This Dream Has Come True**

***(pretty much for all XML content)***

- Which *sounds* bad (they can leave at any time)
- But XML expertise is still a draw/selling point (they can switch *to you* at any time)
- Prove you have real XML expertise and contracts come to you
- Switching companies may entail
  - getting up to speed on the new tag set
  - writing new stylesheets or output specifications may be needed
  - new transforms

**Dream:**

## **Pages — As Good as Current Pages — Automatically**

- XML flows into publishing-system
  - XML tags matched to styles
  - pages mostly auto-styled, designer perfects
- XSLT transforms used to make XML into desk-top publishing driver codes
- Pages made automagically from XSL-FO

## XML Can Feed Composition Systems

- Semi-automagic preliminary pages
- Manual adjustments as needed
- High quality pages faster

## XSL-FO Pages Sometimes OK

- Lights-out pages (bills, statements, reports) often OK from XSL-FO
- Designed pages can start from XSL-FO
- XSL-FO tools weak when judgment needed page-by-page
- Example of pages from XML to PDF via XSL-FO:  
<http://www.mulberrytech.com/Extreme/Proceedings/xslfo-pdf/2001/Onizuka01/EML2001Onizuka01.pdf>
- Example of pages from XML to PDF with hand-work:  
<http://www.mulberrytech.com/Extreme/Proceedings/typeset-pdf/2001/Onizuka01/EML2001Onizuka01.pdf>

## Dream:

### Value-added Electronic Products — Automatically

- Web-based discovery and syndication: RSS and ATOM
- Real-time (or very frequent) updates
- Interactive publications
  - Equations and chemical reactions can be solved and tested
  - Tie incorrect test answers to specific text to reread
  - Add web-only features (animations, sound, interactive)
  - All bibliographic references live (and linked forward as well)
  - All cross-references are real links

## For the Printed Journal

### Raleigh's Discoveries in the New World

Tonia Renae Gaillard  
University of Kentucky\*

In 1584 Queen Elizabeth I charged Sir Walter Raleigh to discover and colonize lands in the New World on behalf of England. This article discusses the efforts and expeditions organized under Raleigh's patronage to found a colony in the current day North Carolina. Going beyond the chronological history of the doomed Roanoke colony, the author provides insights into relations between the colonists and their native counterparts, discusses the botanical riches of the region, and argues that Raleigh's efforts did not end in failure, as commonly might be thought.

#### Introduction

On March 25, 1584, Queen Elizabeth I of England charged Sir Walter Raleigh to discover, search, find out, and view such remote, heathen and barbarous lands, countries, and territories, not actually possessed of any Christian Prince, nor inhabited by Christian People . . . [Thorpe, 1997] That same year Raleigh sent two captains, Philip Amades and Arthur Barlowe, from England to Hispaniola and the Canary Islands; from there, the captains were instructed to scout the lands northeast of those already claimed by Spain, to wit, Florida. This land is now encompassing the Carolinas and Virginia is was claimed on behalf of England and named Virginia, in honor of the Virgin Queen.

#### Securing a Permanent Colony in the Claimed Lands

With land claimed in the New World, an expedition was mounted to establish a settlement. The first expedition failed. Led by Sir Richard Grenville in April 1585, it encompassed 600 men of which 105 remained in the colony while Grenville returned to England for additional provisions. (See Appendix 1.) However, when almost a year passed without Grenville's return, the remainder of the expeditionary force took advantage of Sir Francis Drake's arrival to seek return passage to England.<sup>1</sup>

The second expedition, organized by John White in 1587, fared better. It sailed with seven ships filled with Devon families intent upon establishing a colony in that part of Virginia called Roanoke, a word deriving from the speech of native peoples. (See Appendix 2.) Two years after founding the City of Raleigh, houses had been built for almost all families residing in the colony, and the colony had celebrated the birth of its first children

born in the New World. The first child, grandchild of John White and child of Ananias and Eleanor Dare, was named Virginia in honor of the sovereign.

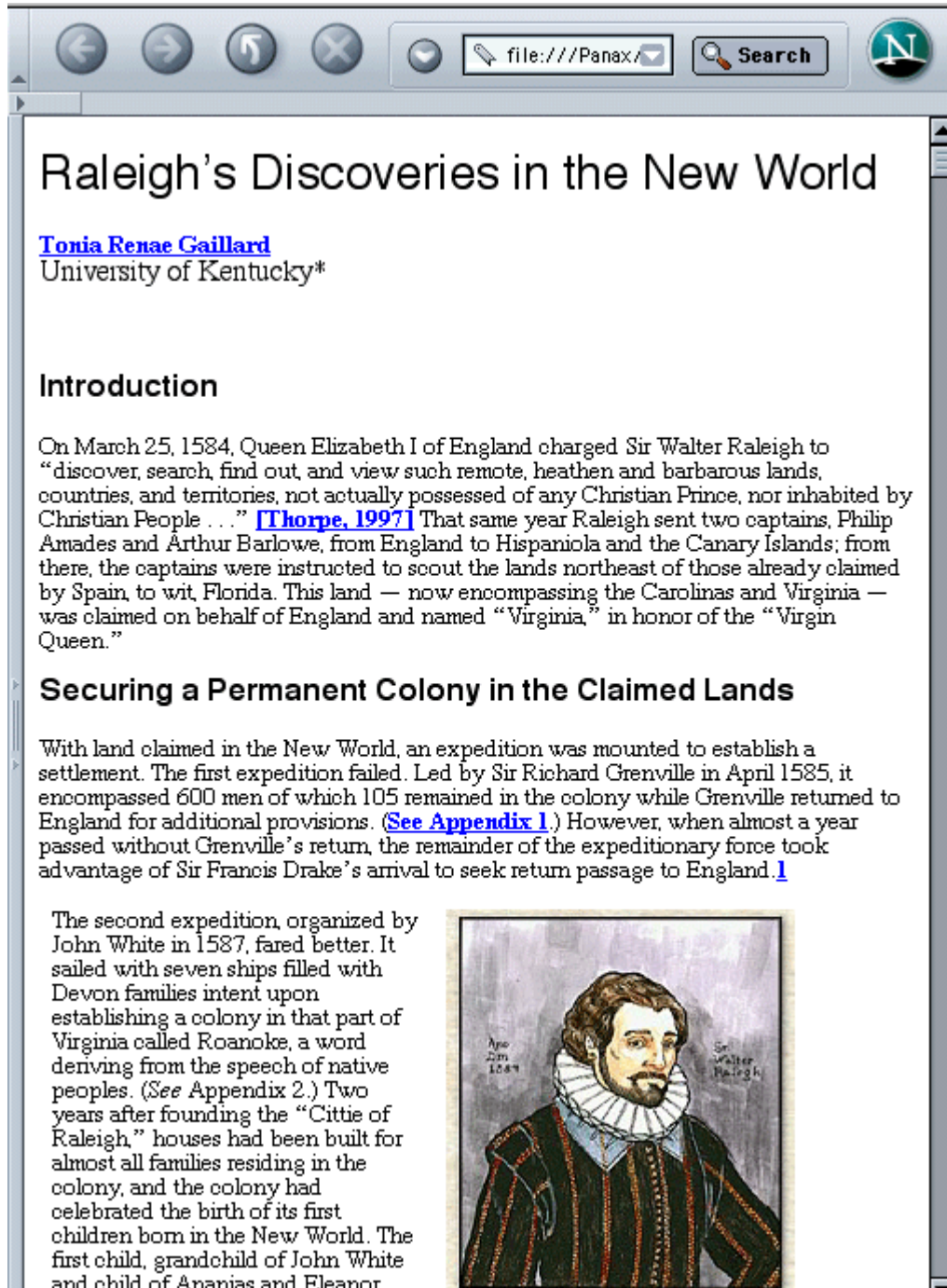


#### Government of the New Colony

Notwithstanding the initial military nature of the early expeditions, civil government was instituted in the colony; to wit, a subpatent was granted by Raleigh to several settlers. Under the authority granted by Her Majesty's charter, a council, comprised of a Governor and several assistants, was formed. Those persons named to this council were themselves planters, that is, men who each made investment in the colony and received 500 acres of land in Roanoke upon their arrival there. Each was charged to make laws and ordinances necessary to the success of the colony, except where those laws would circumvent the power of the queen and

History Revealed Volume 42 no 1  
©Mulberry Press Ltd All rights reserved.

## For the Website



The screenshot shows a web browser window with a title bar containing navigation buttons and a search box. The address bar shows 'file:///Panax/'. The main content area has the following text:

# Raleigh's Discoveries in the New World

[Tonia Renae Gaillard](#)  
University of Kentucky\*


## Introduction

On March 25, 1584, Queen Elizabeth I of England charged Sir Walter Raleigh to “discover, search, find out, and view such remote, heathen and barbarous lands, countries, and territories, not actually possessed of any Christian Prince, nor inhabited by Christian People . . .” [Thorpe, 1997] That same year Raleigh sent two captains, Philip Amades and Arthur Barlowe, from England to Hispaniola and the Canary Islands; from there, the captains were instructed to scout the lands northeast of those already claimed by Spain, to wit, Florida. This land — now encompassing the Carolinas and Virginia — was claimed on behalf of England and named “Virginia,” in honor of the “Virgin Queen.”

## Securing a Permanent Colony in the Claimed Lands

With land claimed in the New World, an expedition was mounted to establish a settlement. The first expedition failed. Led by Sir Richard Grenville in April 1585, it encompassed 600 men of which 105 remained in the colony while Grenville returned to England for additional provisions. (See Appendix 1.) However, when almost a year passed without Grenville's return, the remainder of the expeditionary force took advantage of Sir Francis Drake's arrival to seek return passage to England. 1

The second expedition, organized by John White in 1587, fared better. It sailed with seven ships filled with Devon families intent upon establishing a colony in that part of Virginia called Roanoke, a word deriving from the speech of native peoples. (See Appendix 2.) Two years after founding the “Cittie of Raleigh,” houses had been built for almost all families residing in the colony, and the colony had celebrated the birth of its first children born in the New World. The first child, grandchild of John White and child of Ananias and Eleanor



The portrait of Sir Walter Raleigh is a woodcut-style illustration. He is shown from the chest up, facing slightly to the right. He has dark, wavy hair and a full, dark beard and mustache. He is wearing a white ruff collar and a dark, patterned tunic with gold or red trim. The background is a plain, light color. There are small inscriptions on either side of the portrait: 'Ano 22m 1584' on the left and 'Sir Walter Raleigh' on the right.

## For an Abstracting/Indexing Service



## For an RSS Feed



## **Case Study: Select from Module Library**

A manufacturer of heavy equipment

- Produces many publications
  - many publications for each machine
    - owner's manuals
    - maintenance manuals
    - field engineer's manuals
    - marketing collateral
  - Much content the same in all manuals about same machine
  - Many components the same in several machines
- Manuals written as a library of modules
- Each manual is a list of which modules to select, and in what sequence

For new model, just write modules on new features. Call modules on unchanged systems, features, sub-assemblies.

The instructions on how to use the Model XX123 radio are the same if it is installed in a tractor, a truck, or a road grader!

## **Case Study: Justification/background Hidden in XML**

The US Internal Revenue Service taxpayer information books (the books all US taxpayers get every January)

- Text is marked up in XML
- Print and Web versions from XML
- Additional content embedded in XML, for internal use
  - for each number and form item
    - law it is based on
    - name of attorney who checked it
  - for each form and deduction
    - law it is based on
    - name of attorney who checked it
    - citations and summaries of relevant court cases
  - each time there is a change
    - effective date
    - law it is based on
    - name of attorney who checked it

## **Case Study: Translation Means Translation (Not Typesetting)**

A large manufacturer of cameras and digital equipment

- Has a need for
  - user manuals in 35–40 languages
  - warning brochures in over 150 languages



## Translation Case Continued:

### ***Solution: Text is marked up in XML***

- XML is sent to translators
  - who are told to translate words only
  - leave tags alone
  - return XML
- Simultaneously
  - stylesheets are written to make print, web, accessible versions
  - content needing no translation is made into print, web
  - special stylesheets are written for language direction, localization

## Translation Case Continued

### ***the payoff***

- As translations are finished and sent back (highly variable timeframes)
- Prepared stylesheets can be run to make web/print instantly
- Big company
  - controls look and feel
  - is no longer paying for 40 + 150 composition layouts

(Remember, companies can switch *to* your service)

### ***The Promise of XML (to Content Providers)***

- saves time and money
- platform independent
- vendor independent
- can be validated for QA
- can be made into other data formats

Moves control of content

### **XML to a Service Provider**

- Can be your competitive edge
- Provides opportunities as well as headaches
- Can also be used internally to speed operations
- Is no longer optional or in the far future

### **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>