XML: Your Base Solution for Print and e-Calendars
Presenters: Jocelyn Raymond, Information Systems and Marjorie Morris, Communications
University of Alberta

A major problem in the information industry is how to present one document in many formats. This problem arises from the fact that we usually mix the content of the document with the layout or presentation attributes. In order to achieve this separation of concerns, we need to understand what is involved and how it can be accomplished.

Some Background
Before we start, I would like to give you a little background information on the University of Alberta that will give you some context for the complexity of our University Calendar:
• we have more than 34,000 students, including 2100 students from 146 countries
• there are 17 degree-granting faculties and 65 departments
• we offer more than 200 undergraduate degree programs and 170 graduate degree programs, and more than 5,000 courses

At the U of A, we have a strong Communications Unit and a full service Information Systems Unit capable of supporting new endeavors of the Registrar's Office in addition to providing ongoing systems support.

What we want to share with you today is our experience--the ups and the downs—of our new business process that has streamlined publishing the Calendar by producing both print and online Calendar version from the same document source.

The Importance of Teamwork

- **Close working relationship critical**
  Our Communications Unit always relies on Information Systems for general technical support for our initiatives. However, on this project we needed a different kind of working relationship. We had to form a working team with joint responsibilities critical to the success of the project.

- **Commitment and persistence**
  The team requires commitment--not just to the goals of the project--but real commitment to working as a united team from beginning to end.

- **Drawing lines of responsibility**
  The Communications Unit was responsible for clearly communicating the process we go through in preparing and editing the Calendar. This included the need to produce two drafts for Faculty review before we moved to final, and to articulate specific requirements in formatting, numbering and indexing.

  The Information Systems Unit was responsible for the technology used and the implementation of the solution needed to meet Communications' requirements. It was also very important for us to understand the process of the Calendar production and see its complexities.

Old Calendar Process

Our Calendar is a very complex document.

For many years PageMaker was the primary tool for producing the Calendar—it was used both for editing and page layout. It was the base document and all changes were made using that page layout software. This had the advantage or our always being able to see the changes as they were being made. We then exported the finished Calendar to HTML. We have been producing an online HTML version of the Calendar for about eight years.

Challenges Driving Change

- **Tightened deadlines for online Calendar**
  One of our driving needs was to get the Calendar online faster. Because Information Systems couldn't start their part of the process until we had finalized the document for print, it was a challenge to get the online version up and tested in less than three weeks to a month.

- **Calendar ever-changing, fluid document**
  The Calendar changes continually—this impacted the Information Systems Unit because they couldn't just write one program that would last from year to year.

- **Complex legacy System**
  We had to maintain a legacy system every year which was not, to say the least, the easiest. Several changes and verification were needed each year.

- **Not meeting today's needs**
  Our legacy system was using old technology. Compared to what the technology offers today and the need for increased flexibility to provide services, a change was obviously needed at our University.

Why XML?

- **Extensible Markup Language**
  Extensible is the key word here. XML is not a fixed format like HTML

- **Not a proprietary development of any company**
  XML is a project of the World Wide Web Consortium

- **Functionality and flexibility**
  XML is designed to be flexible, and because of all the related XML technologies, it can provide a large amount of functionality to a system. Even the publishing industry today is adding XML to their main line of products.

- **Not just for the web**
  XML is not just for Web pages--it can be used to store any kind of structured information. In other words, you can make any kind of XML application for any kind of needs.

Publishing Tools

In the Communications Unit, we currently use three different publishing tools: PageMaker, Quark, and InDesign in our various projects. We explored these to see if any were a fit for the new Calendar process.

Both Quark and InDesign support XML so that narrowed the field. However, after testing, we found that Quark was not as user friendly—not as fluid as using InDesign.
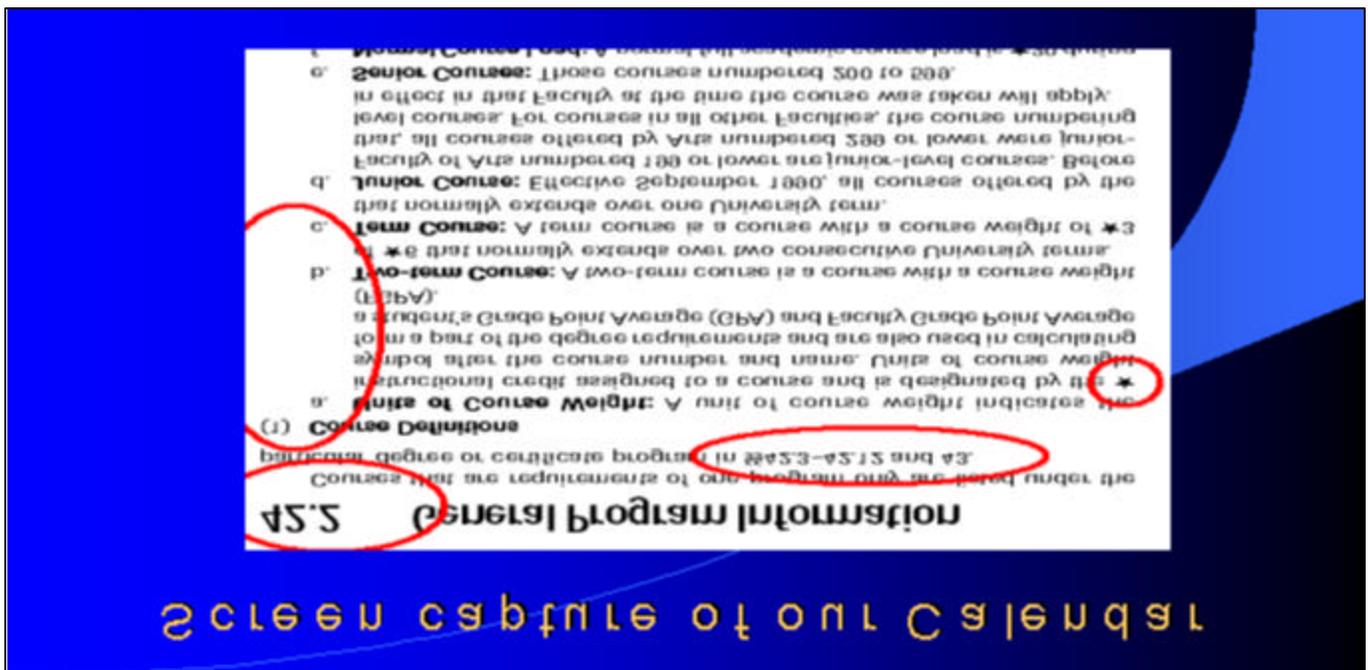
University Calendar: The Problem

The problem that our process was having:

- **Production of the online Calendar** – the inability to produce the online Calendar version until the print version was finished—the online Calendar was always delayed by several weeks
- **Cross references** – were also a problem—moving a section meant to go over the entire Calendar to correct any references to other sections
- **Section numbering and list numbering** – had to be done manually. These were all prone to human error which increased proportionally with the size of the Calendar (those problems apply to both the online and printed version of the Calendar).

Screen capture of our Calendar

The screen capture of our printed calendar contains a Section number, numbered list and cross references. It also follows a specific layout such as list number may be in parenthesis or can be of lower-alpha type.

Here we show only one particular "column" of text. Most of our calendar's text flows in two columns which is part of the layout of our printed calendar.

University Calendar: Our Dream World

If we could have it all, what would we want?

We certainly wanted the ability to produce the HTML version of the Calendar whenever necessary.
We also wanted to have all cross references automatically managed such that when sections are moving around, the cross references re-adjust.
We also wanted the Calendar section numbers to be produced automatically as well as list items.

There are move things we could add to this list such as the ability to automatically integrate addenda, revision control, database integration, etc…. This is our dream world!

The Fundamental Problem

Publishing tools tie the content and the presentation together

One of the first steps toward our dream world was the need to pinpoint the fundamental problems with our existing process because none of the existing publishing tools could provide our dream world.

Previously our entire Calendar was kept in PageMaker, which is the publishing tool.
The main problem is that we have been editing and doing the presentation (laying the content) of the calendar using PageMaker. When I talk about "presentation", I am talking about how the document looks—how the content is "arranged" on piece of paper or on a display. When you think about it, a publishing tool's role is actually more concentrated on "presenting" than "editing". And it is because of this tight coupling between the content and presentation that we cannot produce the online Calendar whenever needed.

Therefore, during that time our content was tied to only one presentation: the print presentation.

Thus, we needed a different solution or process that would separate those concerns.

XML

This is where XML comes in:
- XML has the ability to separate the content from the presentation. In fact, there is no other way.
- XML has the ability to produce an immense amount of different formats including HTML and Publishing Tool format such as InDesign
- XML can provide all the wishes of our dream world:
  - immediate production of HTML
  - automatic cross referencing
  - automatic section and list numbering
  - support for index creation
  - export to publishing tool
  - much more….

XML was the perfect technology choice for producing both online and print Calendars!

How does it work?
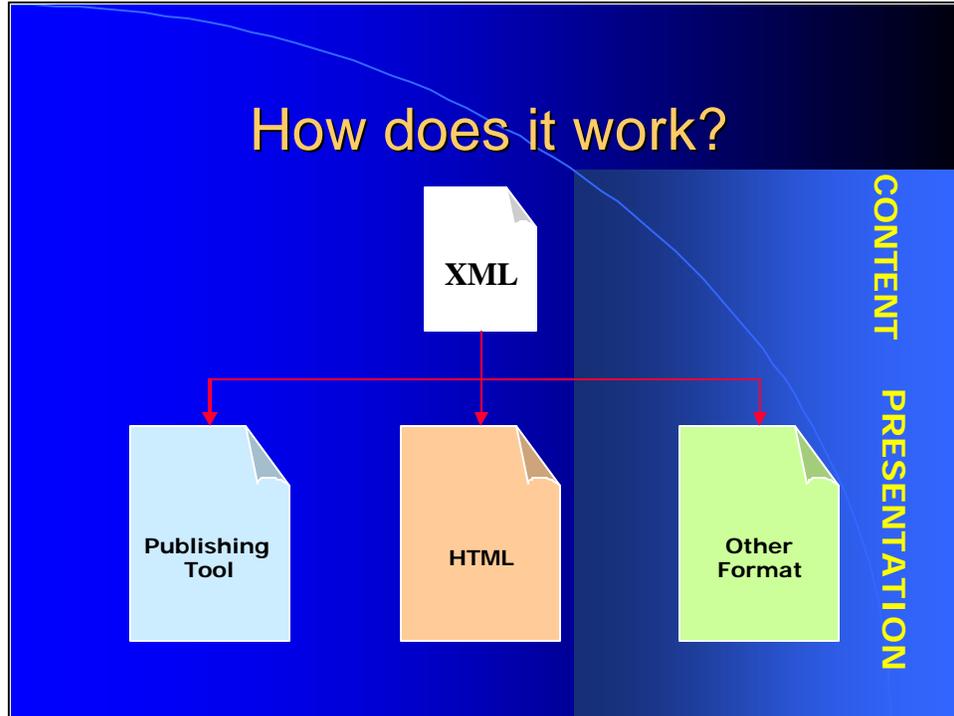
The Calendar would be in an XML format, and the editing of the Calendar would therefore occur in that format. Therefore, XML is where the Calendar 'lives'.

Then, in turn, this format can be transformed instantly into other formats such as the Publishing Tool (InDesign), HTML and other formats such as PDF.

Therefore, XML offers the power of having once source document with many different presentations. If the source changes, all presentations change as well. This gave us the consistency that we also needed.

XML presents other problems

- **Editing the content is not the same as with a conventional text editor**
  XML can present a challenge for staff initially. Nevertheless, training and persistence can easily overcome that.

- **Exporting to publishing tool**
  The fact that the content and the presentation is separated can cause inconvenience such that any little changes need to be applied into the XML document first, then exported to the Publishing Tool.

  If a change is discovered during the layout function (using the publishing tool), the user has to go back to the XML and apply the changes, then export again back to the publishing tool.

Does this mean that all the layout that was done previously needs to be done again? No. We found with InDesign that we can import the content any number of times and it will keep the layout intact.

Despite these inconveniences, there is still a large gain using XML. And it is why we have gone this road. Now let me show you some inconvenience that the Communications Unit had to go through using XML.

Editing in oXygen

This is a product called Oxygen—an XML Editor

As you can see, our staff not only edited the content but also edited the XML structure (the XML tags). Therefore, they were doing two jobs during this time which was not as easy as it looks.

What we needed was an WYSIWYG Editor—What You See Is What You Get!—where there is no need to manage all those XML tags.

Third party WYSIWYG XML Editor

We investigated third party WYSIWYG XML editors and found a few that had potential:
- Xopus is a PC based product
- XML Mind works on all platforms, and
- Morphon – also works on all platforms

We did investigate a few more but those are the ones that stood out. Our Communications Unit uses the Macintosh platform. Therefore, we chose XML Mind as it has great flexibility and can be enhanced way beyond the base product. Their standard edition is also free of charge, which allowed us to try it out at our own convenience.

XML Mind
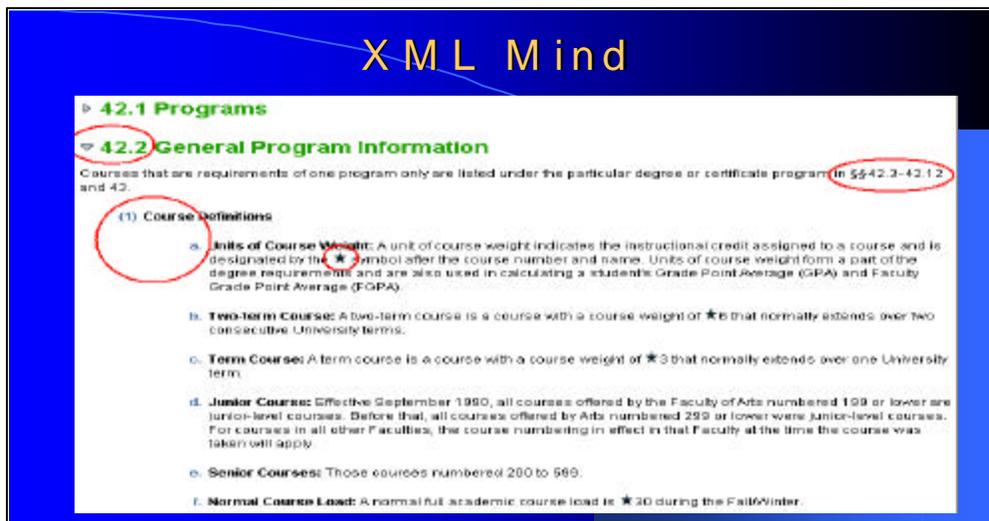
This is what XML Mind looks like. This is an actual screen shot of our content in XML Mind minus the "Buttons" of XML Mind. But this is actually how XML Mind displays our XML Calendar.

XML mind has several advantages over the other editors especially in extending the editor which was very important for implementing some essential requirements.

Note that it seems that some "formatting" or "layout" has been applied to the content, which kind of defies what I have been saying about separating content with layout.

However, XML Mind does provide some formatting as you can see, in order to facilitate the editorial process without affecting the content. This formatting and layout were developed using the extensibility of XML Mind to meet the requirements that were previously identified by the Communications Unit.

Once you are in XML Mind, the layout is "fixed" and cannot be changed by the user.   Therefore, the user never actually formats the "presentation" but only modifies the content.

So this is our only source of the Calendar content.

It supports automatic section numbering, automatic list numbering and formatting. The user cannot edit those numbers. But in the case of list numbers, the users have the ability to choose if it should be alpha-numeric or lower alpha for example.

It also has the ability to provide a "decoration" on the list numbers such as here we "decorate" the ONE with a parenthesis before and after the number. This decoration can be anything you want. For instance, the "a" is decorated with a dot. During our analysis phase, we decided it was important to not have those numbers editable.

We defined that even though those visual cues are seen as part of the content, they are actually not part of the document's content.

Rather we say that those are "derived values" which is very important to understand. Thus if the user inserts a new header or list item, it automatically renumbers accordingly. The user does not have to worry about managing those numbers.

Same thing for cross references. These references to any place in the whole calendar are adjusted automatically if any changes happen in those section (for example if a section number has been changed).

So again, this is the "base" source for this section content. This is actually XML but the XML tags are hidden.

Notice the section here is 42.2 and I will show you now, two different "presentations" of this content: HTML and print.

First The HTML presentation

The content that you just saw in the previous slide is shown here.

What appears above is a small table of contents that is generated from the XML source. The table of contents is not actually stored in the XML but rather is derived from it. As you can see, we are in the Undergraduate Programs, Faculty of Arts and Section 42 – Program of Study. Those subsections are actually links. The ones in gray are disabled because this is what is being displayed. So XML gives us the power to do such things on HTML.

Now this represents the actual content of our XML document. You can see section 42.2 General Program Information. Also all the list numbering is kept intact as are the cross references. Remember that those values are not part of the content but derived from it. Therefore, we can produce the online calendar version which contains the same content plus additional supporting features to make it more functional specifically to the online presentation. You can also see that anything that made sense of being a link, are actually links such as the cross references. **And finally, the list numbering** is kept as desired.

Now lets look at the same section but in the print presentation.

So this is a screen capture of the same section, 42.2, which again comes from the same source content: the xml file from XML-Mind. **Again you** can see that the section number, the list numbering as well as the cross references are kept intact. **Now, if a user make some modifications** in the XML source, say adding a header plus new cross-references, both, the web and print calendar presentations will reflect those changes accordingly.
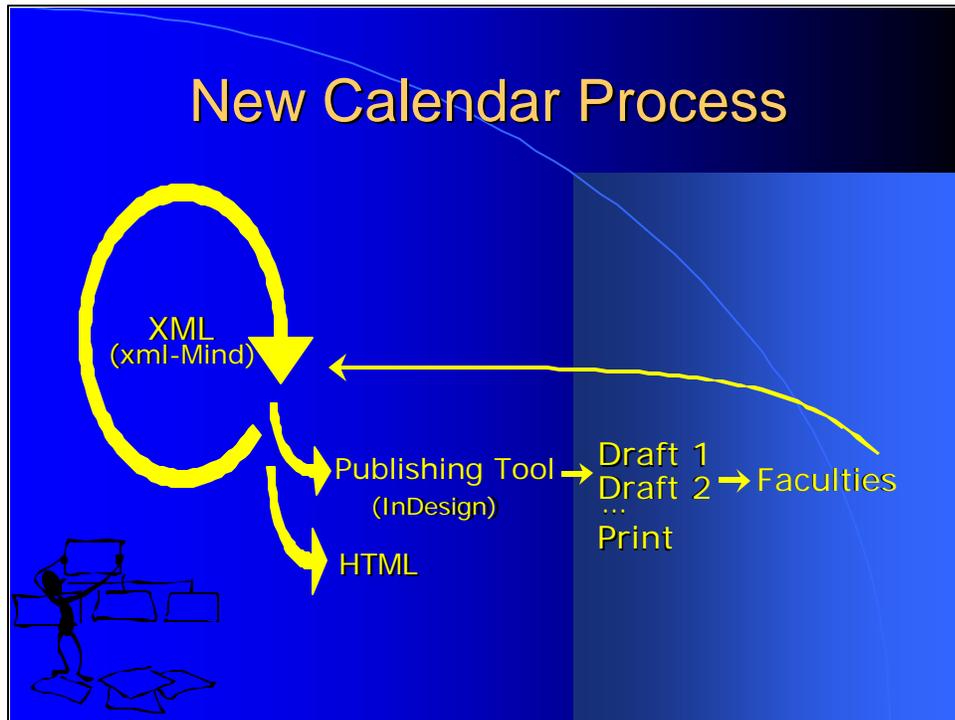
So our new process goes as follows:
- **First** the content is edited accordingly in XML Mind until the first draft is completed.
- **It is then transformed** to InDesign, our publishing tool, where all the layout of the content for the print calendar is done.
- **The drafts** then go to Faculties for review and feedback and returns to our Communication unit, where additional modifications and approved additions are carried out.
- **So the cycle starts again**. Each time it is transformed into InDesign the user does not need to re-do the whole presentation again—the original layout is kept with the difference that it contains the modified content.
- **Those steps** are repeated until we are ready for the print version.
- **Once we have the print version**, we can transform the same content for the web and thereby having a web version instantly.
- **This completes our new process**. Now we will share with you some of our experiences during the development of this process.

Not without Hiccups

- **REALLY understanding the lines of responsibility**
  The process was not always smooth. While we thought we knew which area was responsible for what, sometimes the lines were blurred.

The Communications Unit had always had complete control of the Calendar editing and production process, and it was hard to give up control and let the Systems area do the development part of the process.

There were several hiccups that were due to other factors:

- First, we were new to XML in the Information Systems area. We had a lot of learning to do and thus were slow to come up with best XML practices.
- Because the nature of this project was so tightly coupled with the Communications Unit (due to XML) it was hard to completely separate the "implementation" of the solution without having them involved. Usually the implementation is left up to the software engineer.
- Also, because of the nature of XML and this tight coupling with Communications, we knew that some technology implementation was best suited for the Calendar but difficult to explain why it should be done this way.

- **Building special needs into the development process**
  As we went through the process, we discovered additional requirements. For example, we need a star symbol to identify course credits, and accents were needed in the French section of the Calendar. U of A has a French Faculty—Faculté Saint-Jean.

- **Adjusting to differences in editing tools**
  There has been a learning curve on how to use the XML Editor. Editing in an XML Editor is different from editing in a more conventional Text Editor such as Microsoft Word of even InDesign.

- **Differences between platforms**
  Our office is mainly PC based with the exception of the Communications Unit which uses MACs. The difference in platforms did give us some difficulties.

What did we Learn?

- **Building vs Buying an Editor**
  In our initial analysis, we thought of building our own XML 'WYSIWYG' Editor. Although XML structure is simple, we quickly found that an implementation of such an editor of magnitude was beyond our time frame.

  This made us investigate the possibility of a third party XML Editor. The usual problem with third party software is that you usually lock yourself into a proprietary solution with limited extensibility. However, XML Mind had proven to us that this product is fully extensible.

- **Need to formalize the development process at the beginning**
  We also learned that despite the nature of this project, a more formal process should have been understood by both units at the initial stage of the development.

- **There are Curves in the road**
  We were moving into totally new territory in Calendar production. We were unable to find someone who had used XML to produce a University Calendar.

  We learned there is no straight road in a new development process—and, because of the size of the project, it was hard to conceptualize everything we would run into from A to Z.

- **Additional Challenges**
  During the development process which involved identifying requirements, reviewing what Systems was developing, testing the various products being considered—the Communications Unit was also in the midst of editing and producing the 2004/2005 Calendar.

The Timeline
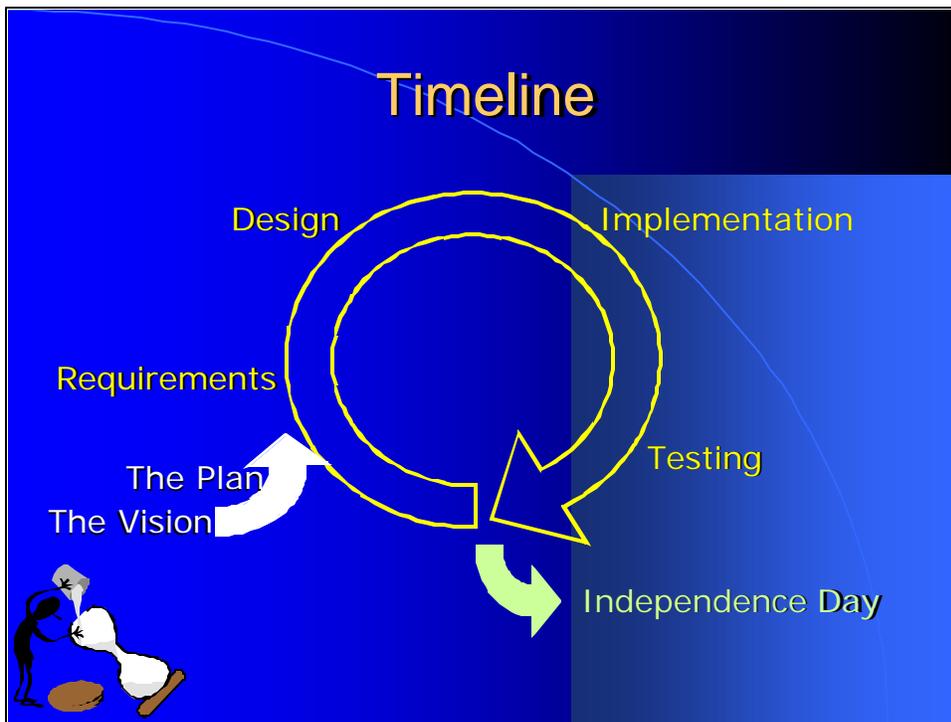
Despite all that, we did try to follow a certain process.

Initially, there was the vision of improving the Calendar process. This was done in conjunction with both the Communication Unit and the Information System Unit. Then we worked out the idea, we identified problems and checked feasibility of the idea.

Then we tried to provide a small iteration of development. The idea was to gather the main requirements. Those requirements were defined by our Communications Unit. The Information Systems Unit then went through the design and implementation. The testing was performed by the Communication Unit—and this completed one development cycle.

Then the cycle started again where the Communications Unit added or modified requirements. Information System designed and implemented the new requirements which were then tested by our Communication Unit.

This cycle can be repeated as many times as necessary. Our goal is to deliver at the end, a solution where staff who are responsible for editing the Calendar become independent of the Information System Unit. If the Communications Unit requires minimal support from us, then we will have succeeded. However, it is a long road to this "Independence Day" and more work is needed.

We are about to start the new Calendar cycle, and this time, the Communications Unit will be utilizing XML Mind for editing, and InDesign as a Publishing Tool. Testing on these tools were done earlier in June so we are not anticipating any surprises.

Key Benefits

- **Successfully produced a 700 page document for print and web**

- **Produced a newly formatted online Calendar**
  With the increased functionality of XML, we were able to completely reformat the online Calendar to make it more user-friendly, easier to navigate, and much more visually appealing. The Table of Contents on the side provides direct links to content—students feel this makes it much easier to find the information they need. Once they choose a section, the sub-sections are indicated above the content so they can see what their information options are within that section. (See www.registrar.ualberta.ca/calendar for our new online version)

  This has received such a positive response from students that we will be able to significantly reduce the number of print Calendars. This year we printed 38,000 Calendars, but next year we will reduce the number produced for continuing students by about half. We still provide a print Calendar to all newly admitted students, as well as for student advisors in Faculties and for administrative staff that require a print version in their job.

- **Not locked into a proprietary solution**
  Another key benefit is that an XML solution does not lock our department into a proprietary solution. In the future, if a better XML Editor comes along, we can switch to the new editor without affecting or modifying our Calendar content.

- **Ability to support future needs**
  Addenda

  At the U of A, we have chosen to keep the online Calendar version the same as the print version. However, XML lets us highlight the placement of any addenda or errata within the online version, without changing the content. For example:

  - **A visual cue (tag)** can be inserted into the online Calendar section wherever an update has been approved
  - **Text that has changed** can be highlighted in color
  - **A link is provided** to the revised text on a Calendar Addenda file

  On the Calendar Addenda file, we may provide more of the changed section than just the actual text change (it may have been just a sentence) to provide context. The actual revised portion would appear in a different color. **This addenda process ensures that any changes are totally transparent for users.**

  An Addenda List is also provided on the Online Calendar mainpage and each item is linked to the Calendar Addenda file.

  Another key benefit is the ability to incorporate any addenda into next year's base document. Therefore, staff can start with a merged copy of the previous year's Calendar.

  - **Database support**
    XML is flexible enough to allow database integration. This is important. For instance, we have various dates for various needs stored in our main database system at the University. Those dates are also edited in the Calendar. XML has the flexibility to get this information directly from our database system instead of being managed manually in the Calendar.

Resources

XML
- www.w3c.org

XML Editor
- www.xmlmind.com
- www.xopus.com
- www.morphon.com

U of A online Calendar: www.registrar.ualberta.ca/calendar

Contacts: marjorie.morris@ualberta.ca
                 jocelyn.raymond@ualberta.ca