

From harried to harmonious the conference proceedings story

Chrisdon Moffett

Arbortext, Inc., Ann Arbor, USA

cmoffett@arbortext.com

<http://www.arbortext.com>

Abstract:

This paper describes the processes and tools used to produce the proceedings for the XML Europe 2000 Conference held in Paris, France.

The request from GCA

In February of 2000, Arbortext was asked to publish the proceedings of the XML Europe 2000 conference. We were asked to publish the works of this year's authors in multiple media types. The GCA (Graphic Communication Association) requested that we produce HTML for CD and paper. We were also asked to provide multiple authoring approaches since not all authors would be using the same software to produce their papers.

The history of Arbortext and the proceedings

The history of Arbortext and the GCA conference proceedings is of significant impact on our acceptance to produce this year's proceedings. Arbortext was first approached in 1996 to produce the printed materials for the SGML '96 in Boston, MA. We rushed for the opportunity to produce these materials. The previous year's proceedings had been produced by the king of the block at that time. We saw this opportunity as a chance to show our publishing capabilities. So, we went about creating our formatted output specification instance for the proceedings. Little did we know at the time that there was no document type definition for paper. GCA had one but it really did not support print as we were asked to produce it.

We worked hard that year with the GCA team to make modifications to support the required output. That effort went smoothly until the papers started to arrive. We found ourselves, along with the GCA team, a few weeks away from the show with papers that did not parse. Papers that had graphics that were incompatible with our product at that time. And, id and cross references that all used the same values. For example: figure1, figure2, figure3, etc. We even had people change the DTD to meet their unique needs. Those of you from the past know what that did to your SGML environment. We ended up with empty required elements and duplicate cross references. So, the files did not parse when we loaded them.

These data issues caused us great pain as we worked our way through each and every paper. We expended close to 120 hours on the FOSI, 40 hours on the DTD, and over 300 hours cleaning up the papers so we could produce the printed versions.

The GCA Europe team also contacted us to do the SGML/XML Europe '98 conference proceedings. We thought that we had learned our lessons about receiving data from multiple sources. Guidelines were provided for the authors assuming that would eliminate a lot of the rework. The guidelines helped but not as much as we had hoped. We still ended up with papers that had all of the problems we experienced in our first attempt. Over half of the papers were close to perfect.

We repeated this process for XML Europe '99. We improved our guidelines for submitting papers. However, we still received papers that required clean up.

Preparing for this year's proceedings

When GCA approached us, we knew that we could not do it the way we had done it in the past. Our consulting team is was too busy to lose weeks reworking papers. In addition, GCA requested that we produce the proceedings on CD-ROM.

We decided that the best way to reduce the overhead of reworking the papers was to make our product available to all of the presenters. However, this presented a few issues for ATI. One, providing our full products to the presenters seemed like overkill for the purpose of authoring the papers. It was also possible, but somewhat surprising to us, that someone may not want to use our product. We also had to deal with presenters being all over the world.

We decided that the best method would be for us to provide multiple paths for the presenters. We decided we would provide our Epic Editor product. We also decided we had to provide a simple method for authors to use Microsoft Word.

We also chose Word because of the Word Interchange option in Epic Editor. This provided us a mechanism to quickly import the Word authored documents into XML. This required us to build a template that the presenters could use as a guideline for authoring their papers.

Once these selections were made, we assisted GCA in making a few changes to DTD. These changes were to support some additional capabilities and information (profiling) that GCA wanted to use this year.

At this time, ATI was also pursuing the ability to provide a remote authoring to option to our customers along the lines of an application service provider (ASP). This is also known as thin client technology.

We had just proven that we could provide an environment that would allow access to a central server to allow authoring of XML content in our product suite via a web browser. We felt that this would be one additional option that could improve the quality (parsing of the XML) for this year's proceedings.

We then reviewed our options with GCA as to the many ways we could work this year's proceedings. We decided that we would use our Epic Editor/Publishing product so the attendees could profile the data based on the technical implementers, business implementers or by day.

The last step to prepare for the creation of the data was to document the authoring procedures to avoid the duplication of data. We also needed this to assist the Word users so they would use the templates to ease the import process.

Upon review and acceptance from the GCA team, GCA and Arbortext produced a package that contained instructions for this year's papers. We also produced a special CD with our Epic Editor product and the appropriate Word templates. The presenters then were able to pick the tool of choice. They could use Epic Editor (local or thin client), the Word template, or any other tool of their choice.

GCA took care of distributing the packettes to all of the presenters. This process took less than 4 weeks from request to the presenters hands.

The submissions

The completed papers were submitted to the GCA for proofing and final review before being passed to Arbortext for printing and production of the CD. The GCA team took on the effort to clean the papers up so that the Word conversion process would go smoothly. The GCA team also took responsibility for verifying all of the issues that we had experienced in the past. Arbortext provided GCA with cleanup instructions from past experiences. They had to preview and print each paper, scale graphics, check for oversets, catch missing punctuation, make sure ids were unique, etc.

There were other issues with the submissions received that used the Word templates. The majority of these issues were the results of the user changing the templates. These changes to the templates resulted in commented data in the XML because the Word Import mapper did not know how to map these styles. The GCA team spent a number of hours working through these documents re-assigning the appropriate Word styles to support the import process.

A number of the XML submissions, not written with Epic Editor, needed some cleanup. Most of these documents were incorporated with minor effort.

A few of the XML submissions written with Epic Editor also needed some cleanup. Again, most of these documents were incorporated with minor effort.

The submissions were distributed across the products as follows:

- 45 Microsoft Word (35%)
- 83 XML Submissions (64%)
- 2 Microsoft Powerpoint or other formats (1%)

Approximately 20 of the XML documents were produced using our ASP server approach. It is difficult, of course, to determine how many used Epic Editor to produce their papers versus other products. This paper was produced using Epic Editor.

The return rate by the due date of 31 March 2000 was 130% higher than the return rate at that same point in 1999. We believe this was due to the pre-work performed in building the Word templates, providing Epic Editor, and having well documented guidelines for the authoring process. The procedures for installing the software and the guidelines made it very easy for the author to write their paper.

The publishing

Once all of the papers had been submitted, cleaned up, and compiled, we went about the process of producing the print and CD versions of the proceedings. We used Epic Publisher for this task. Epic allowed us to use the print FOSI we had developed 3 years ago to produce the hardcopy book provided to all attendees of this conference.

In order to perform the print, we did have to assemble the papers into one master book. We used a customization layer for this process. The customization layer allowed multiple papers and addressed the front and back matter of the printed proceedings.

Epic Publisher also allowed us to produce the CD version of the proceedings. There were two methods for producing these versions of the file. One was to develop an XSL style sheet. The other was to use the FOSI that was used for print. We decided to use the print FOSI since it met the needs of these proceedings.

Of course, at the time of writing this paper the publishing stage of the process has not been completed. The final version of this document will describe the effort that was required.

The Web version of the 2000 proceedings were produced by the Trinity College Dublin.

The products

This section provides a summary of the features that we used to produce the proceedings by product.

Microsoft Word

Microsoft Word templates and styles were used to provide authors a guided method for building a paper that could be transformed to XML easily. Samples of the Word template and support instructions can be found at <http://www.arbortext.com/xmlEurope2000.html>.

Even with the templates in place, some authors took liberties with the styles. Some created new styles and a few did not follow the template. These documents were manually moved back into the appropriate Word structure and then imported into Epic.

Epic Editor

Epic Editor is Arbortext's premier XML authoring tool. The primary features that were utilized by the process was the editing capabilities. We provided two methods for access for each presenter for editing. Each presenter had the option to load Epic Editor to their PC. Or, they could log into a server at our office in Ann Arbor, Michigan, via their web browser in our ASP environment.

We included the GCA DTD and a template for the presenters. The authoring directions described the key elements and their usage. We also identified which elements could be used for cross referencing. We also provided a list of elements that generate text so the presenters would not type that information. For example: the element <abstract> has generated text of ABSTRACT. Presenters did not need to type in the word 'abstract'.

If the presenter used our ASP, they had the option of storing their paper to their local drive or leave it on our server. The benefit to the ASP was that the user did not have to install a product. They only needed to access our site and click an icon. The entire product was available to them via their web browser.

Epic Publisher

Epic Publisher is the engine we used for producing the print and CD versions of the XML document. Epic Publisher was only available to the GCA team during the clean up. The GCA team also used the Epic Word Interchange option to import the Word documents.

Once all of the documents were ready, the GCA team assembled the master document. Epic Publisher was then used to produce the printed version of the proceedings. Arbortext is doing a final review of the papers and assembling the master document..

Epic's Publish to CD was used to produce the CD-ROM version of the proceedings. The CD master was burned from the resulting file. The features provided by our Publish to CD are as follows:

- Produces images that can be used for the Web server and the distribution CD-ROM

- Provides full text search in context of identified fields

- Republishes without obsolescence by linking the user to an updated website or a local disk area. This can be done via Web updates from the Internet.

Citrix Metaframe

Citrix Metaframe was the product we used to establish our ASP version of Epic Editor. Citrix provided an Internet-ready desktop interface for fast and reliable access to Epic on our local server.

Utilizing a Citrix front-end eliminated the need for presenters to load Epic Editor. It also allowed non-PC users to access the NT version of Epic Editor. Presenters with Mac or Unix based equipment could still access our Epic Editor via their web browser. They only needed to download a plug-in is less than 25K.

Citrix also provided the security required to protect each presenters papers.

The conclusion

This process demonstrated that multiple sources of data in multiple formats can be pulled together and then published to multiple media in a very small timeframe. The key to the success is to provide clear guidelines on what you expect from the contributors. You must provide templates and guidelines on how to use the templates.

You must also be prepared to do some level of clean up. You will always have a few contributors that do not adhere to your directions. Some because they do not have the appropriate hardware/software configurations. Others because they have a favorite product or style that does not work with the guidelines and products you have provided.

The GCA team and the Arbortext staff did a tremendous job pulling all of this year's proceedings together. The process started in January with a simple request from GCA. By March, there were templates, authoring guidelines, and Epic Editor on a CD on each presenters desk with GCA's 'gcapaper' doctype. Within 8 weeks, all papers had been submitted, cleaned, and sent to paper, HTML web and CD.

The final version of this paper, the authoring guidelines, the Word and XML templates, and the GCA doctypes can all be found at <http://www.arbortext.com/xmlEurope2000.html>

Author

Chrisdon Moffett

Director of Professional Services
Arbortext, Inc.
Postal Address:
1000 Victors Way, Suite 100
48108 Ann Arbor
MI
USA
Telephon: 734.997.0200
Fax: 734.623.6126
E-mail: cmoffett@arbortext.com
Web: www.arbortext.com

Chris Moffett - Chris Moffett is the Director of Professional Services for Arbortext, Inc., a leading provider of XML-based e-Content solutions. He has been with Arbortext since 1993. He is responsible for the training, consulting, and technical product support services at Arbortext. Chris graduated from The Ohio State University and received a Bachelors of Business Administration.