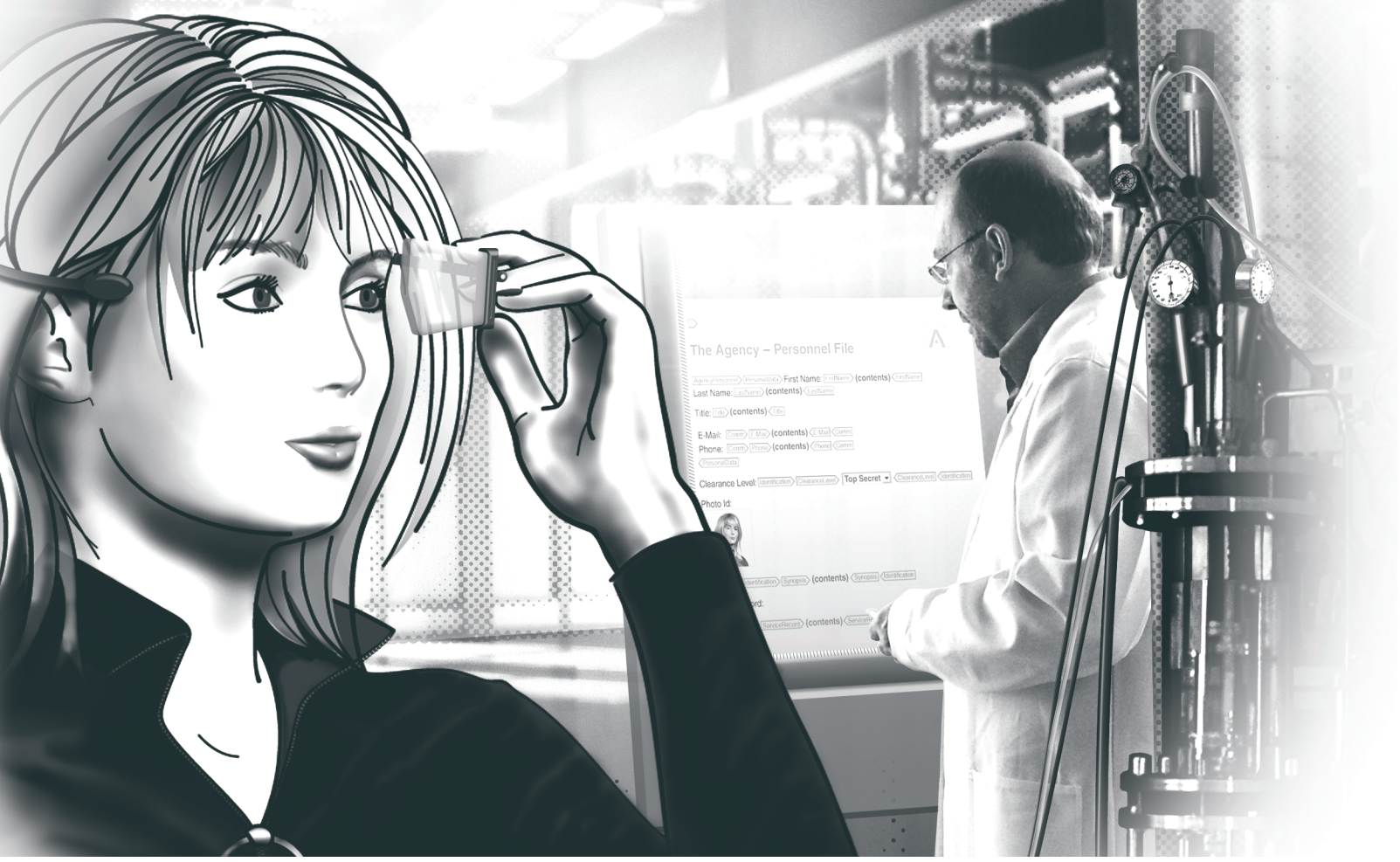
ALTOVA®

WhitePaper

Getting More from Your Content with Single Source Publishing:

Scoping out the benefits, obstacles, and a unique starting point, Altova StyleVision® 2005



Altova, Inc. · 900 Cummings Center, Suite 314-T · Beverly, MA, 01915-6181, USA · Tel: 978-816-1600 · Fax: 978-816-1606 · Web: www.altova.com

Executive Summary

The ability to offer customers business content in multiple formats, such as Web pages, email messages, Word documents, PDF documents, etc., is vital for meeting customers' needs and expectations in today's hyper-competitive Internet economy. However, this task presents a significant challenge since the popular formats in use today are generally not interoperable. XML technologies provide a framework for implementing single source publishing to produce multiple output formats from one source of content, but this still requires the time-consuming development of multiple intricate stylesheets. In addition, many XML-based single source publishing implementations do not address the need to publish content stored in the relational database, which is the prevalent data storage mechanism in today's enterprise.

Only Altova StyleVision 2005 delivers on the promise of single source publishing by simultaneously generating multiple stylesheets based on XML or databases, providing tremendous cost and time savings.

Introduction

In the not so distant past, the dominant medium for communicating with customers was the hard copy, printed page. Today, however, the pace of business is such that printed materials often become outdated as soon as they come off the presses, and advances in communication technology have exploded the number of publishing options available.

Contemporary methods for communicating important business information include multiple media options, such as email, PDF (Portable Document Format) documents, Web pages, word processor documents, and so on. This wide range of formats for representing business content has allowed organizations to customize messages to meet the needs of a variety of consumers - and at the same time it's raised the bar when it comes to consumer expectations. Today's savvy customers expect real-time, up-to-date information available in their preferred format. With the global marketplace online at their fingertips, customers who don't find the information they want from you in a snap will just as quickly look for it from a competitor.

Companies, therefore, face the awesome challenge of producing the same content, such as product descriptions, marketing materials, purchase orders, etc., in multiple formats, while at the same time ensuring accuracy and timeliness. Given that each medium has unique requirements and technical specifications, publishing this content in multiple formats often consumes vast amounts of financial and human resources.

XML technologies offer the ability to implement single source publishing by separating content from style. This inherent ability of XML allows you to publish a single source of content into multiple formats, such as Web pages, printed documents, and so on. However, though XML technologies lay the foundation for single source publishing, only Altova StyleVision 2005 overcomes the significant challenges required to make it a reality.

Multiplicity of Media

In many companies, the process for tackling publishing in multiple media is based on legacy methods for producing printed materials. However, today the number of steps and people involved has multiplied. Content authors generally write content using their preferred word processing program and send files to Web publishers, graphic designers, etc., for publication. For instance, a common chain of events takes place when a marketing representative needs to add content to the company Web site. He/she writes the new content in a word processing application and sends the file to the Web developer, who then has to code the text in HTML (Hypertext Markup Language), add formatting in the form of a stylesheet, add Web-specific features such as hyperlinks and navigation information, and so on. Because the content from the word processor application is incompatible with Web formatting, the marketing rep and Web developer are doing much of the same work twice.

Next, the marketing rep sends that very same content to the graphic designer, who transforms it to PDF format. Though the same content is being published, PDF formatting requires the addition of headers, footers, page numbers, and other layout considerations.

Depending on the number of formats in which each document needs to be replicated, this process snowballs rapidly.

To further complicate matters, this multi-tiered process is repeated each time the original content is updated. Whether an update is required to simply change a word, correct an error, or add new information, each and every occurrence of that information must be tracked down and updated. This becomes quite a daunting task when many versions of each file and different types of file formats have accumulated. After just a few iterations, the same content exists in multiple file formats, multiple versions, and multiple places.

The amount of time and money spent getting customers the information they need - when, where, and how they want it - can quickly eclipse even the largest company's budget restraints. What is needed is an efficient method for producing documents in a variety of media quickly, easily, and free of errors. XML technologies offer a solution by providing the basis for single source publishing.

XML for Single Source Publishing

The eXtensible Markup Language (XML) is a universal language for specifying a standard syntax to describe a document's content. It's a clearly defined way to structure, describe, and interchange data that separates the actual content from its style, i.e., the way the content is presented on a screen, on paper, etc.

Traditional tools for writing and displaying documents such as word processor applications embed formatting in the content. This makes translating content for use in multiple media difficult, because changing the formatting ultimately changes the content, and vice versa.

XML, in contrast, separates content from formatting and layout information to allow effective content re-use. It's this inherent ability of XML to store content and style separately that promises to unify corporate information and lessen the chasm between disparate document formats. XML technologies are ideally suited to solve content management and publication challenges because they are both platform and programming language neutral, inherently transformable into other formats, easily stored and searched, and easily transmitted via XML-based Web services.

XML single source publishing makes use of XML-based XSL technologies. The eXtensible Stylesheet Language (XSL) consists of, among other things eXtensible Stylesheet Language Transformations (XSLT) and eXtensible Stylesheet Language Formatting Objects (XSL:FO). These two stylesheet languages are used to apply style to XML documents. XSLT is a standardized language for transforming XML documents to simple output forms such as text, HTML, or RTF (Rich Text Format), which is a common document format used by applications such as Microsoft® Word®. XSL:FO is a language for expressing advanced document layouts such as those used by PDF and PostScript formats. XSL is the style component that corresponds to the XML content. Using XML with XSL allows you to change a document's style without affecting the content to, for example, publish a product catalog in HTML, Word, and PDF formats. Conversely, it allows you to change a document's content while preserving the style, which is necessary when publishing that same product catalog in multiple languages for international distribution.

Transforming a document to produce another data format involves processing an XML file and a stylesheet in an XSL processor to generate a new output document. This way, a single source of XML content can be published to a wide variety of formats by simply applying the appropriate stylesheet. At the same time, you can modify the XML data itself without changing the existing stylesheet(s) to customize information for different users, update changing content, or publish a particular document in multiple languages.

Using XML-based single source publishing in this way also enables the re-use of standardized, company-approved information. For example, saving small pieces of information in XML allows reuse of commonly published information such as product descriptions, price lists, company overviews, etc. Because an XML file is used to contain and describe these nuggets of data, consistency and accuracy are ensured, and updating the information once means that the change will be reflected everywhere that content is used. This ability is especially important for product information, for example, which in many industries changes very frequently. Representing a product description in XML means that a required change can be made once - and be automatically reflected in every document and Web page that contains that content.

Once an organization establishes an underlying single source publishing framework, most publishing tasks can occur automatically, since the documents and stylesheets conform to the defined XML model.

The potential of single source publishing for saving time and money is tremendous. In fact, as analyst Ronald Schmelzer notes in the "XML in the Content Lifecycle Report[1]," "XML-based content management and single-source publishing can reduce up to 75% of total publishing costs."

Consider the advantages of being able to publish (and later update) the items listed below in Web, print, etc., formats, as well as in multiple languages, simultaneously:

- Product and services descriptions
- Product catalogs
- User manuals
- Price lists
- Articles and papers
- Direct-mail correspondence
- eBooks and hard-copy books
- Presentations
- Marketing collateral: brochures, data sheets, Web pages, etc.

[1] Source: ZapThink, LLC - "XML in the Content Lifecycle Report" - http://www.zapthink.com/report.html?id=ZTR-CL100

Single Source Publishing Challenges

Despite the countless advantages of the XML single source publishing model, challenges remain. Each output format (HTML, PDF, RTF, etc.) has its own associated XSL stylesheet. For example, an XML document that is to be rendered in HTML requires the application of one XSLT stylesheet, and a different XSLT stylesheet is necessary for RTF output. Producing that same content in PDF requires an XSL:FO stylesheet. If additional output formats are desired, additional stylesheets must be developed.

Because of the need to have different stylesheets for each format, the process of developing these files becomes extremely complex and time consuming. Many designers have little if any experience in this process, because though one might be skilled at developing XSLT stylesheets for HTML output, for instance, he/she may still not know the intricacies of working with XSL:FO or the many other stylesheet variations in use today.

Adding to the pressure is that the XSLT language is a programming language of its own, and a single stylesheet error can prevent the generation of any useful output. HTML output for use on the Web often requires the coding of dynamic forms that change based on user-input as well as complex tables with multiple rows and columns. Writing XSL stylesheets to support these functions by hand can take a considerable amount of time and is an error-prone process.

To help alleviate the complexity of stylesheet design, visual data mapping tools have been created to accelerate the development of XSLT stylesheets. These tools, however, are generally restricted to generating one particular stylesheet format.

Another significant challenge is getting content into XML to begin with. Though XML syntax is designed to be simple, XML documents contain both content and markup tags, and therefore are generally not a user-friendly media for content creation and editing. Non-technical business users, who generally fulfill the content authoring role, usually do not have the technical experience required to write XML documents, and simply saving word processing documents as XML, for instance, does not automatically produce meaningful output that is truly reusable. So, an extra step remains for which a developer must convert the authors' documents to an XML format.

Though XML provides a solution for rendering content created by subject matter experts in multiple formats, the majority of critical business information is stored in relational databases. A common requirement is producing information from databases for display on Web pages and other media to, for example, provide a customer with a history of his/her transactions, extract pricing data for use in a purchase order, and so on. Though stand-alone, proprietary applications for mining databases and generating database reports are available, they are separate from the XML single source publishing model and therefore require additional cost and labor expenditures. An effective single source publishing solution must provide an efficient means for accessing data stored in corporate databases.
All these challenges contribute to a long felt need for solutions that can be used to efficiently and accurately design stylesheets based on XML documents and databases to render multiple output formats. Only then will single source publishing solutions deliver on their promise to simplify and accelerate the process of publishing accurate, up-to-date content in multiple formats, giving your customers the information they'll use to make purchasing decisions.

Altova StyleVision® 2005 for True Single Source Publishing

Altova designed its StyleVision 2005 to enable true single source publishing, that is, generating multiple output formats from a single source in just one step. StyleVision 2005 generates multiple stylesheets and multiple output formats at once from a single design. Using StyleVision 2005, developers or designers create a StyleVision Power Stylesheet (SPS) visually using an intuitive drag-and-drop interface and intelligent design functions. From the SPS, StyleVision 2005 auto-generates standards-conformant XSLT 1.0/2.0 and XSL:FO stylesheets for creating the most popular types of output requested today: HTML for Web pages and email communications, RTF documents used by word processors such as Microsoft® Word®, and PDF for tamper resistant electronic documents that can be viewed and printed. Each SPS also creates an Authentic Form, which allows non-technical users to enter and update content in XML documents and relational databases without being exposed to the underlying technology, using Altova Authentic® 2005.

StyleVision 2005 allows you to base your SPS on any existing XML Schema or DTD, and it also allows you to use a database as a source. When you design an SPS based on a database, StyleVision 2005 automatically processes the database data in XML behind the scenes. This makes designing a database report with output in HTML, Word/RTF, and PDF as easy as designing any SPS.

Finally, StyleVision 2005 allows you to convert existing HTML content visually for re-use in an XML single source publishing implementation. This feature adds extensibility to your existing HTML content, which can now be used in XML applications and rendered in PDF, Word/RTF, and Authentic Form formats, as well.

Authentic Forms

The foundation for any document is the content itself, and in single source publishing, the first step is to get content into XML. StyleVision 2005 also gives organizations the unique ability to extend the single source publishing model to include relational databases, so content may be stored in this format as well.

Authentic Forms created with StyleVision 2005 allow non-technical knowledge workers to enter content directly into XML and databases. Authentic Forms are used with Altova Authentic® 2005 software, which is available free-of-charge in desktop and browser editions. Authentic 2005 is a user-friendly, word processor-like electronic forms interface for entering content to be saved in XML or databases. Using Authentic Forms created in StyleVision 2005, authors enter their content directly into an XML document or a database. This saves the time previously spent by the developer copying and pasting or using other complex conversion processes to get the content author's document into an XML or database format. Authentic Forms also eliminate the need to purchase special applications to allow non-technical users to update database data, thereby saving the time and money required to train users on these applications.

Once data is in the proper format, it's easy to publish it. The same SPS used to create the Authentic Form simultaneously produces industry standard XSLT and XSL:FO stylesheets for HTML, Word/RTF and PDF, so these can be applied with one click to produce the desired output. This innovative approach to content editing unlocks corporate knowledge and increases overall efficiency.

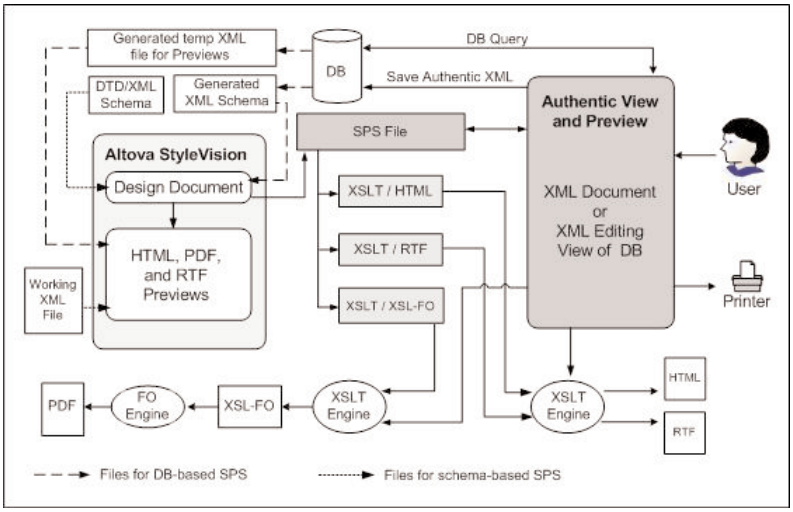


Figure 1: The StyleVision 2005 single source publishing model incorporating the free Authentic 2005 content editor

More information about Authentic Forms is included in the Designing Authentic Forms section below.

In addition, StyleVision 2005 includes ready-to-use templates for editing XML documents based on the most important industry standard vocabularies such as the DocBook DTD for technical documents, the Darwin Information Typing Architecture (DITA) for creating documentation, NewsML for representing electronic news items, and more.

SPS - the Meta Stylesheet

The technology that makes auto-generation of Authentic Forms and transformation stylesheets possible is the Stylevision Power Stylesheet (SPS).

The SPS is designed using visual drag-and-drop functionality in an intuitive user interface. As you are building the SPS in the design window, StyleVision 2005 automatically generates the corresponding Authentic Form, XSLT 1.0 or 2.0 stylesheets, and XSL:FO stylesheet. It's in this way that the SPS functions as a meta-stylesheet, i.e., a stylesheet that generates stylesheets. This SPS meta-stylesheet approach is much more powerful than editing multiple XSLT stylesheets by hand, because it uses one unified design to create different manifestations of stylesheets for multiple output formats.

Because StyleVision 2005 automates the process of writing complex stylesheets, the occurrence of errors is virtually eliminated, and your XSLT 1.0/2.0 and XSL:FO stylesheets are inherently standards-conformant. Designing an SPS in StyleVision 2005 allows you to produce intricate, effective stylesheets for multiple output formats in minutes - instead of hours or days.

Creating Database Reports

Because StyleVision 2005 allows you to base your SPS on a database as well as a DTD or XML Schema, you can create database reports with eye-catching output to HTML, RTF, or PDF quickly and easily. Supported databases include:

- Microsoft Access®
- Microsoft SQL Server
- MySQL®
- Oracle®
- IBM DB2®
- Sybase®
- Any database via ADO connectivity

To design a database report, simply connect to a database using StyleVision 2005. StyleVision 2005 automatically recognizes relationships in database data, and you can filter the information in your report according to any criteria. Once you've specified the filtering options, StyleVision 2005 transparently creates an XML Schema and an XML instance document that contains the data in the database. Then, you design your SPS based on that schema using the same process used to build SPS files for transforming XML documents (this process is explained in the Designing an SPS section below). At any time during the design process you can preview the resulting stylesheets and HTML, RTF, and PDF-formatted database reports that are generated based on your design and the database data.

Using StyleVision 2005 for database reporting not only allows you to create eye-catching HTML, RTF and PDF files representing database data, it also automatically supplies you with the corresponding standards-conformant XML Schema and XML instance document, which you may use as desired in your XML applications. After saving your SPS file, you can save the XML Schema, XML instance document, XSLT and XSL:FO stylesheets, HTML page, RTF document, and PDF file from within the StyleVision 2005 interface.

In addition, you can call StyleVision 2005 from the command line to generate any or all of these files. The command line features allow you to call the stylevision database reporting functionality from within your own applications or backend systems - StyleVision 2005 will run silently (i.e., without the GUI being opened), generate the required files, and close.

In one simple step, StyleVision 2005 allows you to generate HTML, RTF, and PDF reports, plus the corresponding XML files and stylesheets, using the same user-friendly process used to transform XML documents. Because StyleVision 2005 converts your database data to XML behind the scenes, information stored in relational databases becomes more accessible for publishing, and organizations save the time and money previously spent procuring and using standalone database reporting tools.

Designing an SPS

The StyleVision 2005 user interface allows you to design an SPS using visual functions, which simplifies and accelerates the process significantly. As shown in the screenshot in Figure 2 below, you simply load your content model (DTD, XML Schema, or database), into StyleVision 2005, and the hierarchical structure of the schema or database is displayed in the left-hand panel. To define the formatting and layout of your documents, simply drag and drop elements listed in the tree on to the design window. When you let go of the mouse button, a menu opens prompting you to select how you'd like that element displayed, as shown below. You can create contents, paragraphs, tables, bullets, input fields, and more.

As you design your SPS, StyleVision 2005 automatically generates:

- An Authentic Form
- An XSLT 1.0 or 2.0 stylesheet for HTML output
- The corresponding HTML preview
- An XSLT 1.0 or 2.0 stylesheet for RTF output
- The corresponding RTF/Word preview
- An XSL:FO stylesheet for PDF output
- The corresponding PDF preview

You can view any of these stylesheets or output previews by simply clicking on the appropriate tab at the bottom of the design window pictured below in Figure 2.

StyleVision 2005 helps you create eye-catching output quickly and easily by allowing you to apply rich formatting options to your design using text styles and block styles. Text formatting options include font sizes, weights, and colors. Block formatting can be applied to components that have been defined as blocks, and includes background colors, borders, spacing around the block, etc. These options are available in two palette windows on the left-hand panel, as shown in the screenshot in Figure 2 above.

Global templates for your document are also displayed in a helper menu, shown below in Figure 3. StyleVision 2005 populates the global templates list with the global elements in your XML Schema or all the elements in a DTD, and you can define properties and rules for processing each global element in a highly reusable manner. This saves significant time because you don't have to recreate these rules at each occurrence of an element.

Advanced Stylesheet Functions

StyleVision 2005 also includes support for designing advanced stylesheet features such as complex tables, dynamic layouts, conditional templates, auto-calculation, and more. Coding these features by hand is a complicated, time consuming process. StyleVision 2005 allows you to define complex functions in your SPS using the same visual design paradigm, saving time and increasing the sophistication of your stylesheets and the quality of your output.

XSLT 1.0 and 2.0 Stylesheets

Based on your SPS design, StyleVision 2005 auto-generates XSLT 1.0 or 2.0 stylesheets for transforming database and XML data to HTML and RTF formats. StyleVision 2005 offers full support for XSLT 1.0 and XPath 1.0, and it includes the industry's first production-ready implementation of the powerful new XSLT 2.0 and XPath 2.0 specifications from the W3C. These new specifications include significant improvements and new features that increase code stability and reusability, as well as an extensive array of new functions and operators. In all, XSLT 2.0 and XPath 2.0 allow you create stylesheets that are much more powerful and highly efficient. You can select whether you wish to use XSLT 1.0 or XSLT 2.0 for individual designs. Depending on your selection, the appropriate functionality will automatically be made available in the design interface, and StyleVision 2005 will create the required XPath 1.0 or 2.0 selectors. Transformations will be made with the corresponding award-winning Altova XSLT processor (XSLT 1.0 or XSLT 2.0), which comes built-in with StyleVision.

Once you select XSLT 1.0 or XSLT 2.0 and begin designing your SPS, you can preview the XSLT stylesheets that are auto-generated for each format using the XSLT-HTML tab and the XSLT-RTF tab, as shown in Figure 4 below. StyleVision 2005 automatically creates standards-compliant, error-free stylesheets, which saves a significant amount of time when compared to hand coding XSLT. Furthermore, since you design your SPS using visual tools, you don't need to know all the technical details of the XSLT 1.0/2.0 and XPath 1.0/2.0 languages used to create the stylesheet. In fact, since StyleVision 2005 can auto-generate standards conformant XSLT 2.0 stylesheets complete with XPath 2.0 selectors, it can be a valuable learning tool as you come up to speed with these important new specifications.

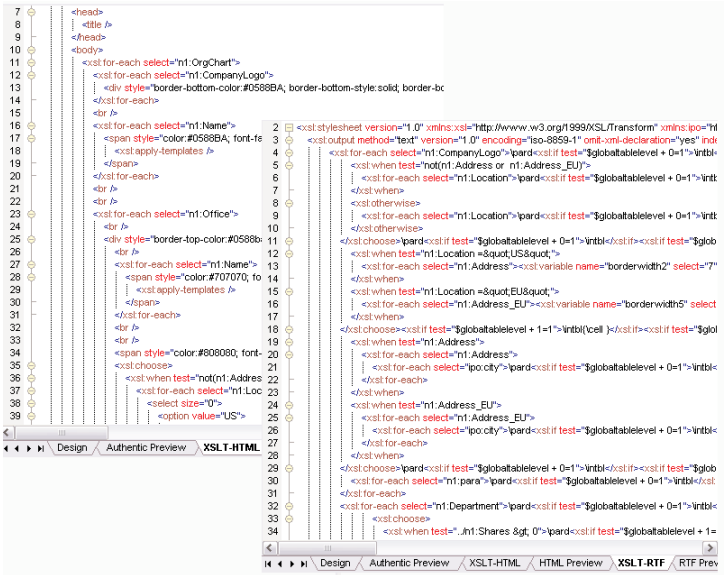


Figure 2: Previewing the auto-generated XSLT stylesheets

In addition to previewing the XSLT stylesheets, if you've supplied content in the form of an XML instance document (this is automatically supplied if your content model is a database), you can preview the actual HTML and RTF/Word output to test your design immediately. A screenshot of the HTML preview tab is shown in Figure 3 below. Figure 4 illustrates the RTF preview tab, which uses Microsoft Word running inside StyleVision 2005 to produce a preview of the output document.

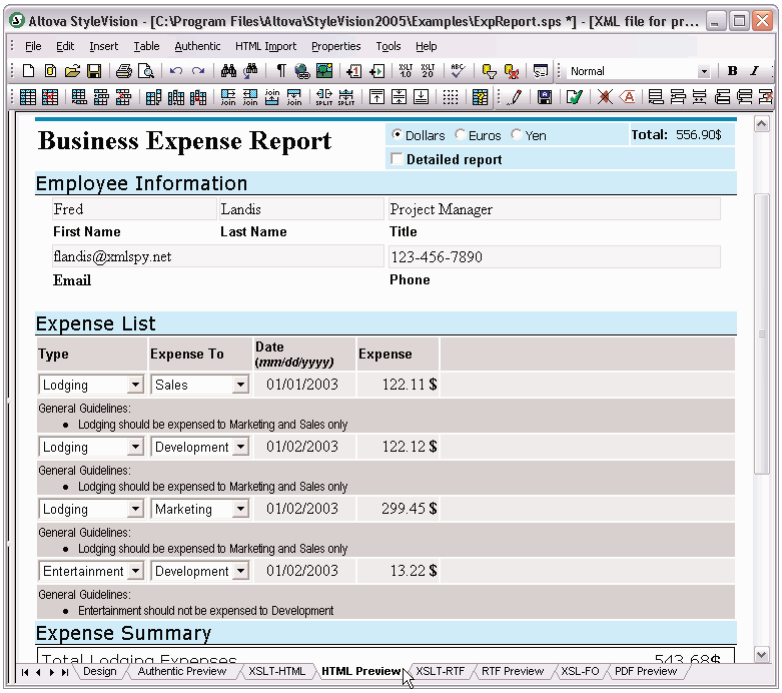


Figure 3: Previewing HTML output

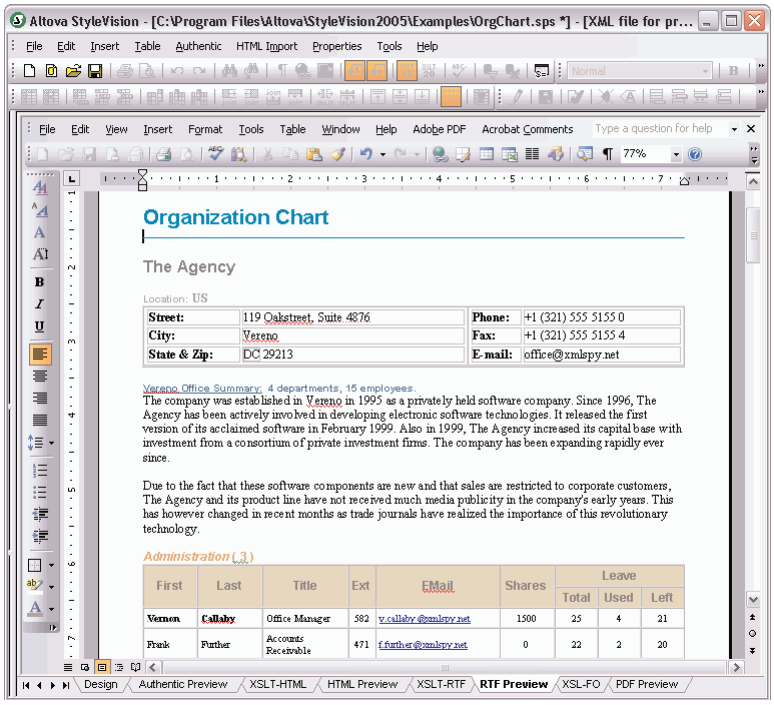


Figure 4: Previewing RTF output

After you've previewed the XSLT stylesheets and the HTML and RTF/Word output files, you can save any of the files as required. These files are generated according to the respective standards specifications and can be used in any application.

XSL:FO Stylesheets

As with XSLT stylesheets, an XSL:FO stylesheet is auto-generated based on the SPS you design, and you can view it by clicking the XSL:FO tab, as shown in Figure 6.

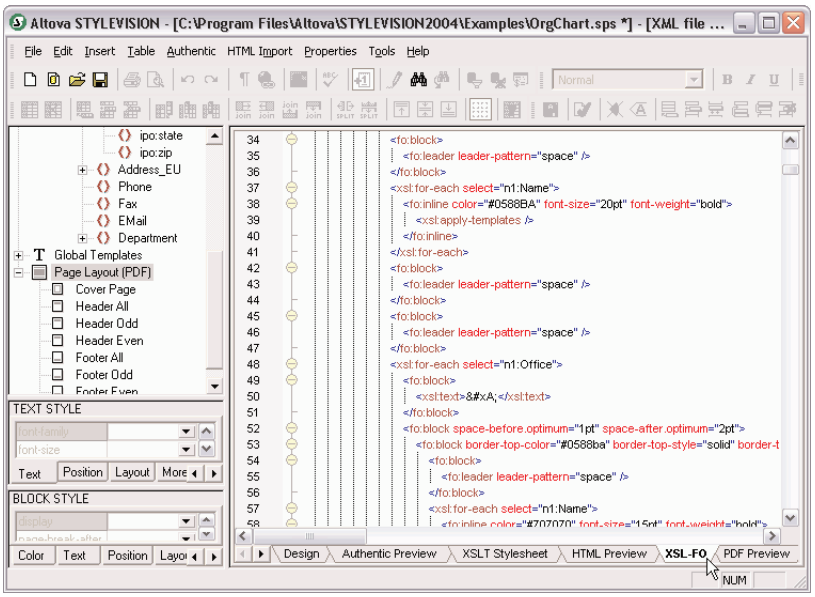


Figure 6: Previewing the auto-generated XSL:FO stylesheet

PDF files often include formatting features such as headers and footers, page numbers, cover pages, etc. StyleVision 2005 makes it easy to apply these PDF-specific formatting options, which are available in the left-hand panel of the design window.

The PDF preview tab allows you to save the generated PDF and to run Adobe Acrobat directly inside StyleVision to check the results of your transformation. An example of an expense report converted to PDF in StyleVision 2005 is shown in Figure 7.

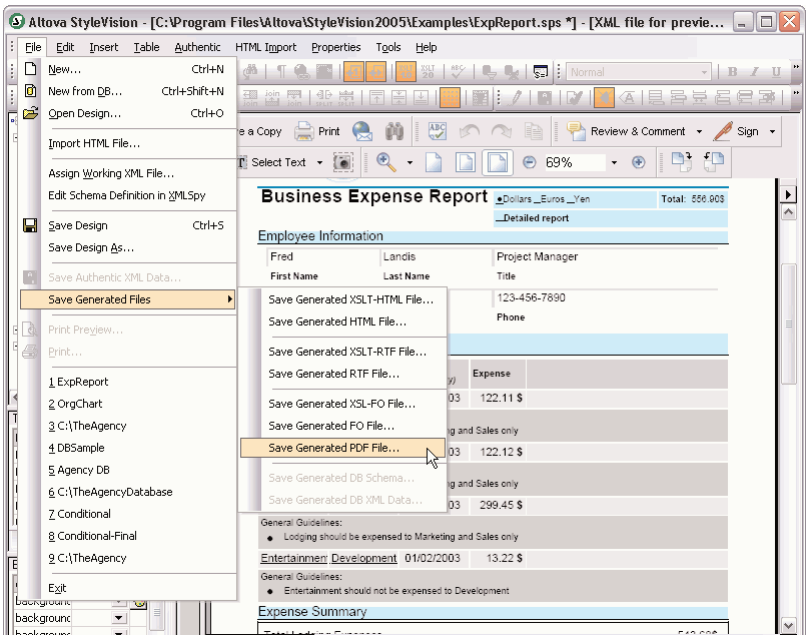


Figure 7: Previewing and saving PDF output

Designing Authentic Forms

In addition to XSLT 1.0/2.0 and XSL:FO stylesheets, your SPS design automatically generates an Authentic Form. Authentic Forms include rich editing features and strong validation capabilities that allow non-technical users to enter data in XML and databases easily and efficiently using the complimentary Altova Authentic 2005 application.

StyleVision 2005 allows you to add Authentic-specific features to your electronic form. For instance, in addition to schema-based validation, which Authentic performs implicitly, Authentic Forms support business logic validation. For every input element in an Authentic Form, you can specify any number of additional validation rules based on XPath expressions that will be automatically enforced when the user enters data in Authentic 2005.

Authentic Forms also support user-friendly time/date controls, the ability to populate combo box values at runtime based on data in the XML file or database, complex tables, and other advanced features. An example of an Authentic Form for editing data in a database is shown in Figure 8.

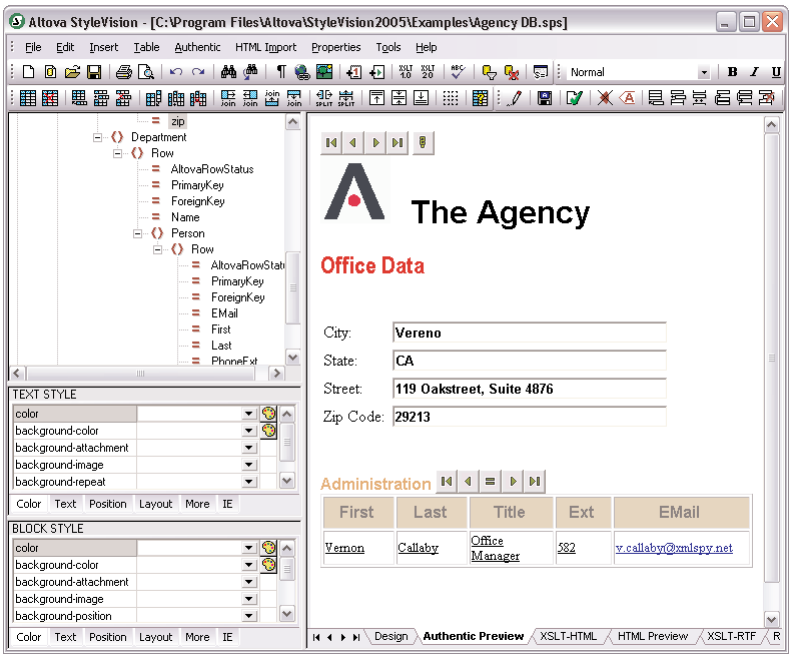


Figure 8: Previewing an Authentic Form for database editing

To bring the single source publishing process full circle, after business users enter data in the Authentic Form, you can output the file to HTML, RTF, or PDF instantly using the stylesheets that StyleVision 2005 auto-generated when you created the Authentic Form. Because your SPS design simultaneously produces the XSLT and XSL:FO stylesheets and an Authentic Form, the time and money previously spent on publishing are vastly reduced.

HTML Conversion

While the above model for single source publishing is effective for transforming newly created content and database reports, there still exists a significant amount of data in formats such as HTML Web pages. To access the data stored in HTML while preserving a document's style, StyleVision 2005 allows you to convert existing "legacy" HTML pages to XML as shown in Figure 9. The conversion results in an XML Schema that reflects the content model, an XML instance document that holds the actual content, and an XSLT stylesheet with the presentation style information.

In the past, conversion from HTML to XML was a complex, time consuming process. StyleVision 2005 allows you to convert using the same visual drag and drop process used to design an SPS, allowing you to separate HTML content and style quickly and easily for use in your single source publishing implementation.

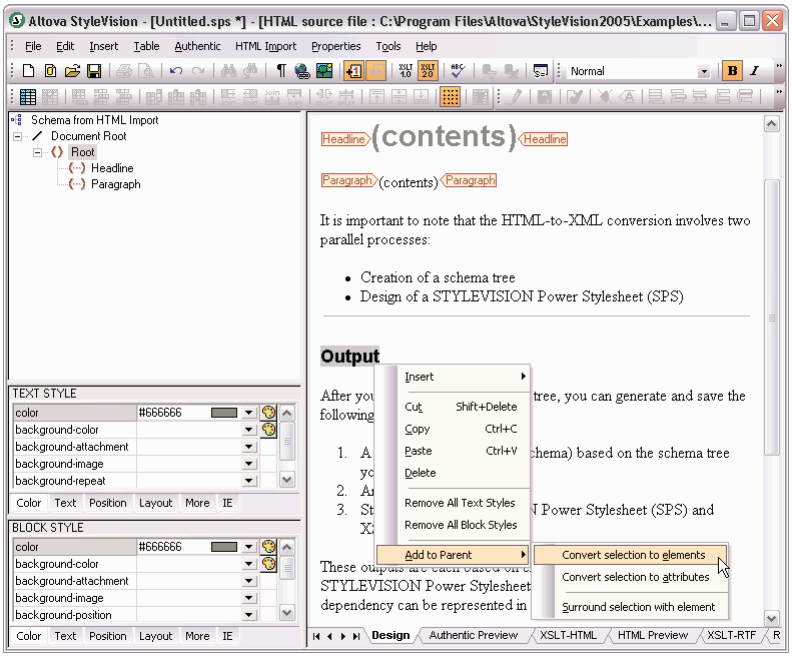


Figure 9: Converting HTML to XML

Conclusion

Getting a handle on the multiplicity of media to meet demands for accurate, up-to-date information in a variety of formats is a daunting - yet extremely critical - task. Maintaining competitive advantage means delivering customers the information they need quickly and effectively, whether it's in a printed catalog, on the company Web site, via email, via direct mail, or through some other method.

XML-based single source publishing solutions allow you to separate content from style to effectively and efficiently publish a single source of content to multiple media. However, despite its advantages, the single source publishing model presents many challenges, the most significant being the amount of time and skill required to produce multiple stylesheets to render the various output formats required. In addition, XML-based single source publishing often leaves out an important source of corporate data: the relational database.

Altova StyleVision 2005 delivers on the promises of true single source publishing by allowing you to create one design to output data stored XML or databases in multiple formats. In addition to XSLT 1.0 stylesheets, XSLT 2.0 stylesheets, and XSL:FO stylesheets, the SPS meta-stylesheet also generates an Authentic Form. This further simplifies publishing by allowing business users to enter content directly into databases and XML, which can then be published directly in the industry's most popular formats: HTML, RTF, and PDF. Stylesheets auto-generated in StyleVision 2005 are fully standards-conformant and error-free, and you can preview the resulting transformations to check your progress as you work.

With the additional capability of HTML-to-XML conversion for leveraging existing work, StyleVision 2005 provides the ultimate solution for true single source publishing in today's enterprise.

You can download a free 30-day trial of StyleVision 2005 at www.altova.com/download.

About Altova

Altova accelerates development and integration projects with software, services, and solutions that enhance productivity and maximize results. As an innovative, customer-focused company and the creator of leading software development tools - including XMLSpy, the pre-eminent product in its class - Altova products are the choice of over 1.5 million clients worldwide and virtually every Fortune 500 company. With customers ranging from vast development teams in the world's largest organizations to progressive one-person shops, Altova's line of software applications and custom-tailored professional and educational services fulfills a broad spectrum of business needs. Altova is an active member of the World Wide Web Consortium (W3C) and is committed to delivering standards-based platform-independent development solutions that are powerful, affordable, and easy to use. Altova was founded in 1992 and has headquarters in Beverly, Massachusetts and Vienna, Austria.

Visit Altova on the Web at www.altova.com.