

Once we decided what to do (convert to XML) and how to do it (use the Cocoon publishing framework), we were ready to begin the actual task of doing it. Since much of our Web site consists of documents (reports, project results, presentations), we began by converting one of these documents, **Opening Gateways: A Practical Guide for Designing Electronic Records Access Programs**, into an XML format. We used this 50-page guide as a prototype to test the viability of converting our Web site while it enabled us to learn XML and the Cocoon environment.

The first step was to convert the existing document. Like most of our publications, the final version of the **Gateways Guide** was in a PDF format. Our goal was to convert this PDF into an XML format that could serve as our final version, single-source file. We also wanted to automate this process as much as possible to reduce the time involved and ensure consistency in the results. To reach this goal (from PDF to RTF to XML), we used a four-step process with two software conversion programs and two "clean-ups" after each conversion as illustrated in Figure 6.

Our resulting XML document conformed to the DocBook standard which is a widely used document type definition standard containing a popular set of tags for describing books and articles (such as <book>, <chapter>, <title>, etc.). Within XML, it's important to use standard, widely accepted markup tags to describe your data so that you can use and share this data over time and in a variety of applications. While nothing within XML prohibits you from creating your own markup tags, it is not good practice because it potentially isolates your content and limits the flexibility of stylesheets to transform and present that content. If like items (such as chapter titles) are referenced in like ways, then one XSL stylesheet, for example, could transform an unlimited number of different XML documents.

This conversion process worked very well for us and allowed us to convert a document within a few hours. This does not mean that we recommend the same conversion process and tools in all cases. Every environment has different needs. Fortunately, several conversion tools and methods are available for a variety of situations. (See our list of references at the end of this document for additional information on conversion tools.)

Figure 6. Transforming Documents from PDF to RTF to XML