

## **Conversion of Word files to S1000D-conformant XML files**

S1000D is an international specification for the procurement and production of technical publications. It is an SGML/XML standard for preparing, managing, and using equipment maintenance and operations information.

Graham Every Ltd has developed a process that enables the automatic conversion of Word documents into XML files that conform to the S1000D standard. The automated nature of the process means that conversion takes place more quickly and is therefore more time- and cost-effective.

Naturally, discretion is paramount when handling the types of files needing conversion to XML files that conform to the S1000D standard. Graham has considerable experience of working with such sensitive data. He has also signed the Official Secrets Act.

### **The conversion process**

The first step is to create the conversion script that takes the files from Word format to XML files that conform to the S100D standard. Graham Every has over 20 years' experience in this field and his conversion scripts are written to minimise the level of (costly and time-consuming) manual post-conversion clean-up required.

Using the conversion script, the Word files are converted to well-formed XML. If required, these files can be further split into separate files that match section levels in the Word file (i.e. section 1.1, section 1.2 etc). These files can be named according to the section levels in the document (i.e. section 1.1 etc) or according to DMC naming conventions, if these have been supplied electronically. Following the completion of any post-conversion manual clean-up required, the resulting split files will be parsing XML files that are conformant to the S1000D standard.

Graphics in the Word document are handled in the conversion as follows. For every graphic encountered, a graphic element is placed in the resultant XML file and an entity created. As with the document text, if DMC naming conventions for the images have been supplied electronically, these can be incorporated into the XML file. Graphics can be extracted from the Word file, but they will be at a resolution of 72 dpi, which will affect their quality and suitability for press. However, should the graphics be supplied separately, together with an electronic document mapping the source file name to the target file name, then the graphics can be renamed accordingly during the conversion process.

On completion of the conversion, the client is supplied:

- an XML file for each component part of the original Word document(s), assigned the relevant DMC number if available
- a file that maps the component sections in the Word file to the outputted file name
- a file that details the graphics in the documents. If the source figures do not have DMC numbers, the file simply lists the figures; if the source figures do have a DMC number, the file maps the figure names to the appropriate DMC number.