WordXML+ Extraction and Transformation Tool

User Guide

Software Version Release 1.3.3. Lead developer / support: dawud.Rahman@aqovia.com / ioannis.selimas@pearson.com

5th October 2015

Contents

- Introduction
- Understand Word-XML format
- Getting started with Word-XML+
 - 1 Login to the application
 - 2 Dashboard
 - 3 Upload your Microsoft documents
 - 4 Select Template Type
 - 5 -Starting transformations
 - 6 Viewing new transformed documents
 - 7 Download XML documents
 - 8 Preview as XML and HTML
 - 9 Settings page Define Specification Unit templates
 - 10 User Management

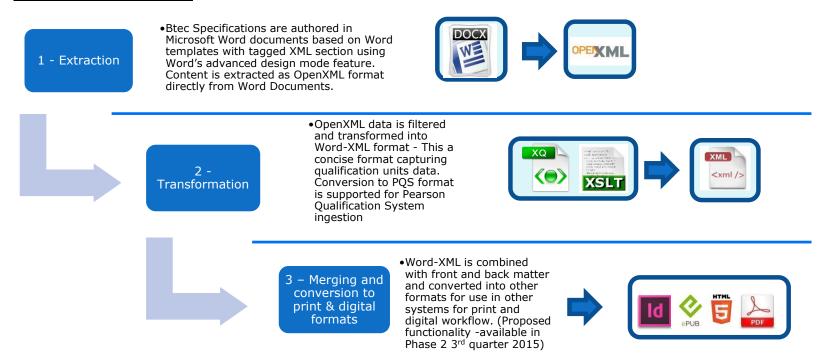
Introduction

WordXML+ is a Cloud software application which enables Pearson Qualification developers and writers to extract Btec Nationals and Btec Level 2 specification data authored in Microsoft Word documents as XML which is then ingested into the MyBTEC product. The application is built on the open source DocX4J library and executes Xslt and Xquery scripts to generate IQS formatted XML required for the Pearson Edexcel course accreditation system (PQS).

What it does

This application replaces the manual outsourced process handled by Aptara to convert Btec specification MS Word documents into IQS XML. In absence of a long term strategic system this application is an interim solution to address this need for automated and bulk extraction, which represent a significant cost saving to Pearson Qualification department.

How Word-XML+ Works



Understanding the Word-XML format

users of u

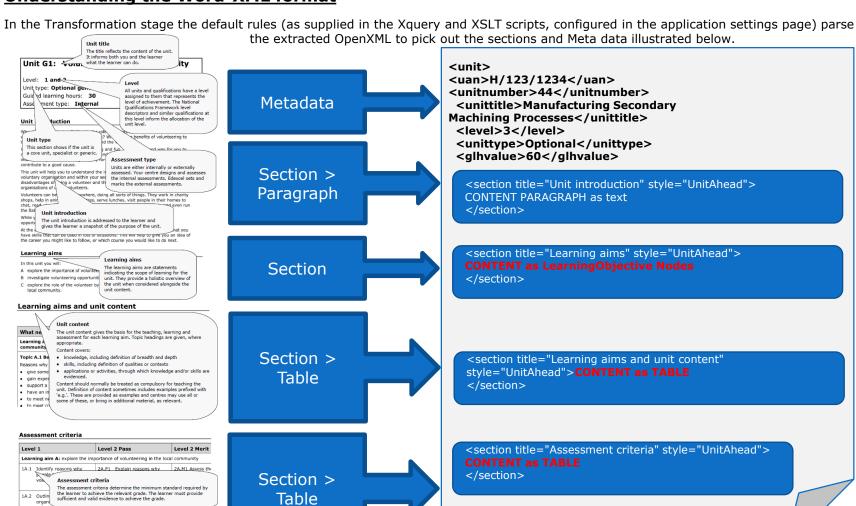
Learning aim B: investigate volunteering opportunities and organisations in the local

18.3 Identify the aims of two

28.P3 Describe the aims of

28.M3 Compare

ALWAYS LEARNING



PEARSON

Continued....

Understanding the Word-XML format (Continued...)

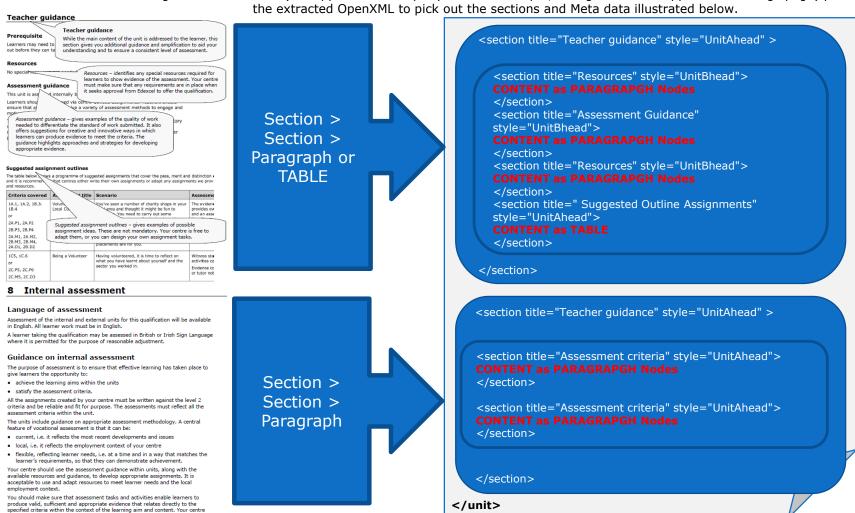
should plan to include a range of 'fit for purpose' approaches in line with the vocational nature of the qualification. Learners can produce evidence in a variety of

audio-visual recordings of portfolio, sketchbook, a working logbook etc

different forms, including:

written reports, graphs, posters
projects, project plans
performance-observation records
time-constrained practical assessments

In the Transformation stage the default rules (as supplied in the Xquery and XSLT scripts, configured in the application settings page) parse



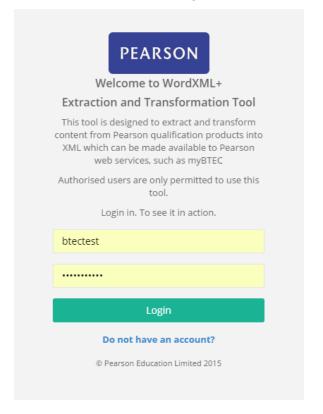
Getting started with Word-XML+

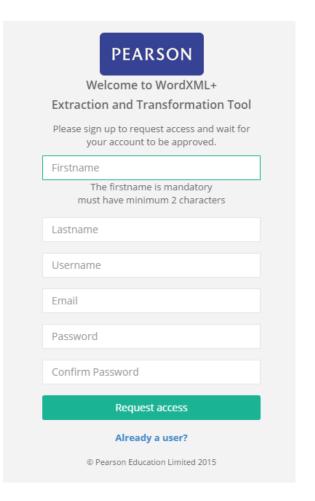
1 Login to the application

The application is host on the cloud:

http://wordxmltool.elasticbeanstalk.com/public/login.html

- Sign up for a login account if you don't already have an account setup.
- Once approved by your Qualification Development Team leader, sign in.
- For Administrator see the section guide on User Management.





PEARSON

Getting started with Word-XML+

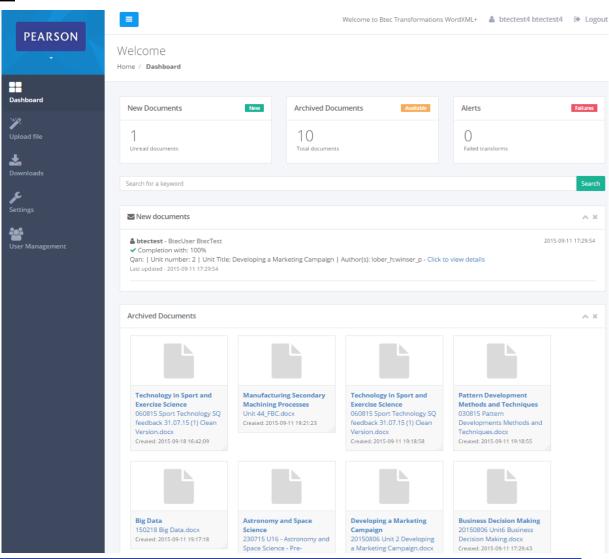
2 Dashboard

Upon login, this is the main page showing:

- New unread documents that have been recently transformed and status.
- Transformed Word-XML documents.
- Summary statistics of transformed documents.

Left Hand Navigation is the common way to use the application:

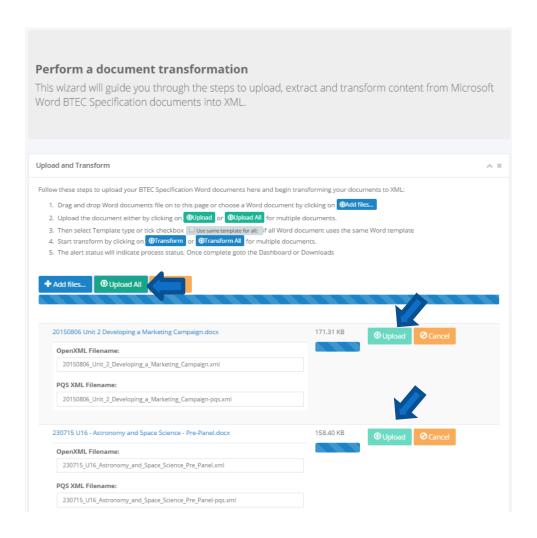
- Dashboard this page
- Upload file upload and perform Word to XML file transformation
- Downloads completed transform documents are listed here
- Settings add Word document validation and xml tag filter rules
- User Management Administrators can manage users here.



PEARSON

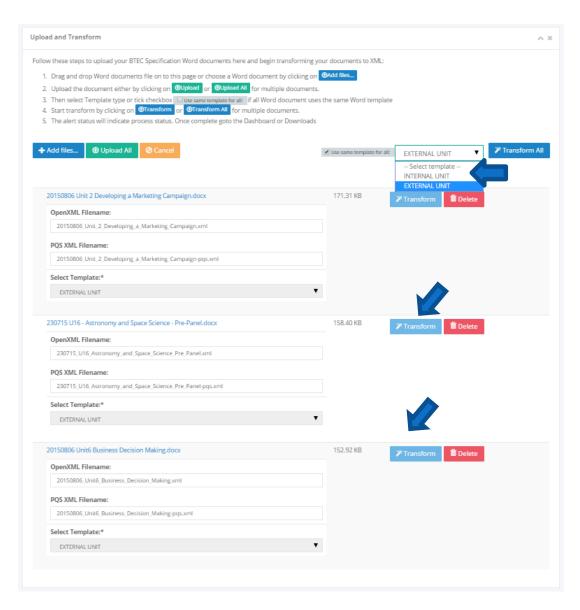
3 - Upload your Microsoft Word documents

- Before uploading your Word documents finalise your content by accepting all changes and turn change-tracking in Word.
- Click on 'Upload file' tab in the left hand navigation to go to the file upload wizard.
- Drag and drop your files onto the page or click on 'Add files' and add them to the list.
- Click on 'Upload ALL' or 'Upload' to upload the files to the server. Wait until the progress bars have completed.



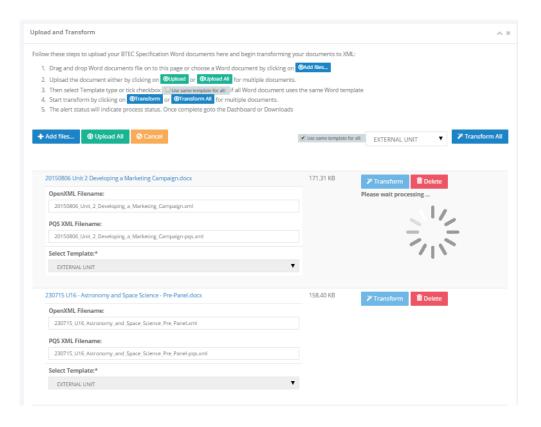
4 - Select Template Type

- When file is uploaded is the 'Transform' button will be enabled.
- Select the Template type for all the documents and enter UAN number and XML document filename.
- Click on 'Transform ALL' or 'Transform' to upload the files to the server. The status message will appear when the document has transformed and the document will be cleared from the list.



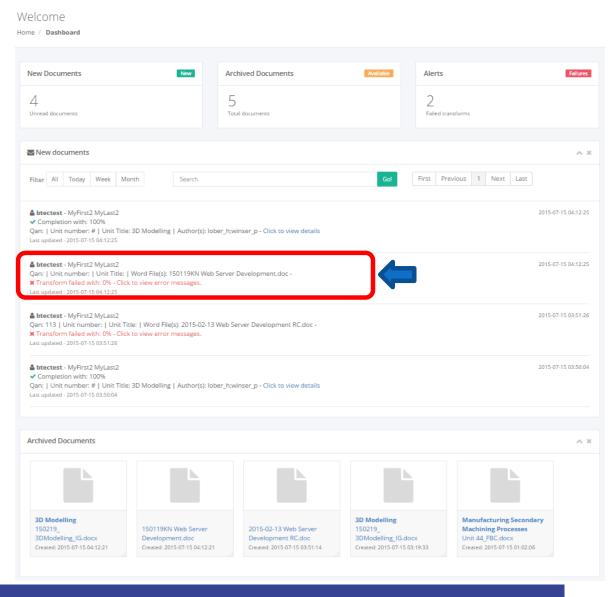
5 – Starting transformations

- Each valid work document will be processed sequentially.
- The process will come to a halt if there are errors. For failed documents, click 'Delete' and upload the correct Word document again.
- Failed transformations will display with an error message. This could be due to either :-
 - Change tracking is present in document
 - Incorrect template type selected or document template has been modified.
 - Word file is corrupt or is not Word Docx format.
- Check and correct the Word document and upload and transform again.

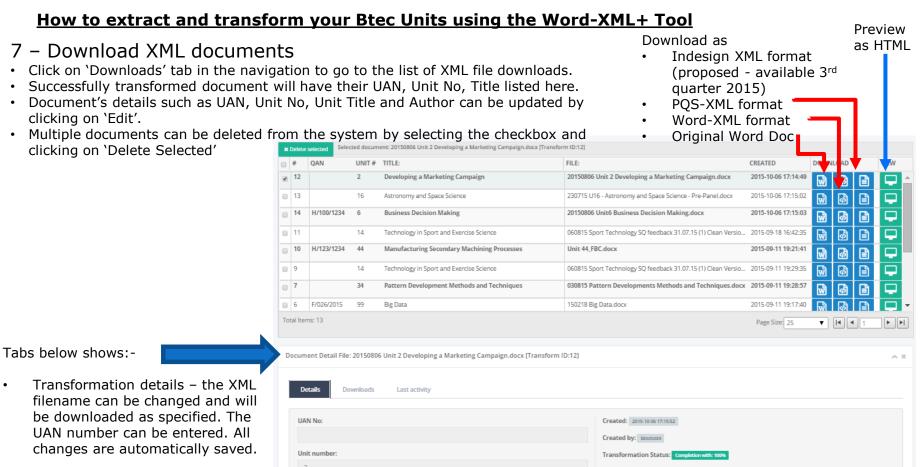


6 – Viewing new transformed documents

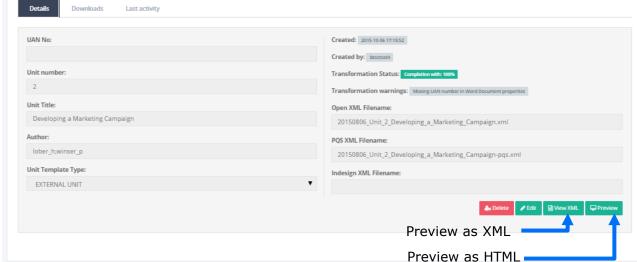
- Click on the 'Dashboard' tab in the navigation to view the resulting transformed document.
- For failed documents, go to the 'Downloads' tab and delete the selected document. Then go to the upload correct Word document again.
- Recently transformed document will appear here, click on the link to view details of the transform. Once you view the document its status will change to 'READ' and will show up in the Archived documents.



PEARSON



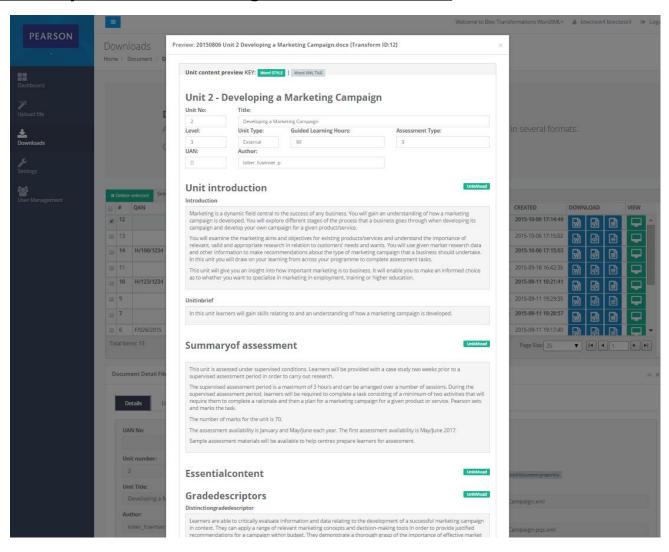
- Transformation details the XML filename can be changed and will be downloaded as specified. The UAN number can be entered. All changes are automatically saved.
- List of available downloads will list all the formats. The XML document will instantly generate upon download.
- Last Activity this maintain an audit trail of edits, reviews, comments made to the XML content via the editor (available 3rd quarter 2015)



8 – Preview as XML and HTML

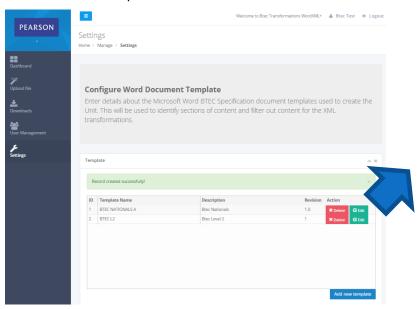
 On 'Downloads' page in the 'Details' tab and list is the 'Preview' button. This will open a page to display a HTML rendered preview and an XML preview of the document.

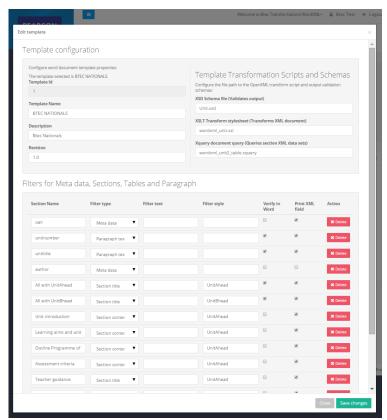




9 – Settings page - Define Specification Unit templates

- Click on 'Settings' tab in the navigation to go to the setting pages.
- Word-XML tool needs to know what XML tagged content is expect in the Word document. This populates the dropdown list for Template type in the Upload file page.
- To add a template click on 'Add new template' or 'Edit'. Add the Sections that you would like to validate on and extract to the Word-XML output file. Here's what the Template field means :-
 - Section name define a name for this content rule.
 - Filter type state what type of section this is.
 - Filter text search for a text i.e Title
 - Filter style search for Word formatting style i.e UnitAHead, UnitBHead
 - Verify in Word validates this field in Word document.
 - Print XML field if none are select then all the XML tagged content will be output to the Word-XML file.





10 – User Management – for users with Admin Role only

- · Click on 'User Management' tab in the navigation to go to the user management pages.
- New users are created from the sign-up link on the login page.
- The new user account will need to be approved by the Qualification Development Team leader and assigned a Role. This can be either Admin, Editor, Author or Viewer only.
- Once assigned a role the user can login. The login account is different from the Pearson cloud login. This is so that temporary accounts for freelancers and temporary staff can be created quickly.
- Each user account will maintain an audit trail which is recorded in the Cloud server logs for monitoring the applications and identifying causes of system errors.

