
**webMethods Integration Server
MS Excel**

**SimpleExcel Package User Guide
Version 3.1**

Prepared by
Igor Androsov.

December 11, 2002

Table of Contents

<i>Table of Contents</i>	2
Overview	4
Purpose	4
Architecture Overview	4
SimpleExcel Package Detailed Description	5
SimpleExcel Package Components	5
System Requirements	5
SimpleExcel Deployment Directory structure	6
SimpleExcel Configuration	6
Working with Excel File Data	7
Known issues and limitations	8

Date	Author	Rev	Change Description
Dec 11, 2002	Igor Androsov	1.0	Created initial draft.
Aug 20 2005	Igor Androsov	1.1	Add encoding for Cell handling Asian language, handle null rows and cells, addition to process formulas and Blank cells

Overview

The SimpleExcel utility package for webMethods Integration Server has been designed to enable processing of any standard Microsoft Excel documents. SimpleExcel package integrates as component into webMethods Integration Server, and provides functionality to read, write, parse and map Excel files at binary level similar to standard processing of XML and Flat file documents.

The SimpleExcel package is based on Apache POI project that provides most comprehensive support for Microsoft's OLE 2 Compound Document file formats such as MS Word and Excel files, allowing them to be processed in pure Java. The SimpleExcel package uses HSSF API of POI project and enables webMethods platform to manage Excel files virtually on any supported platform.

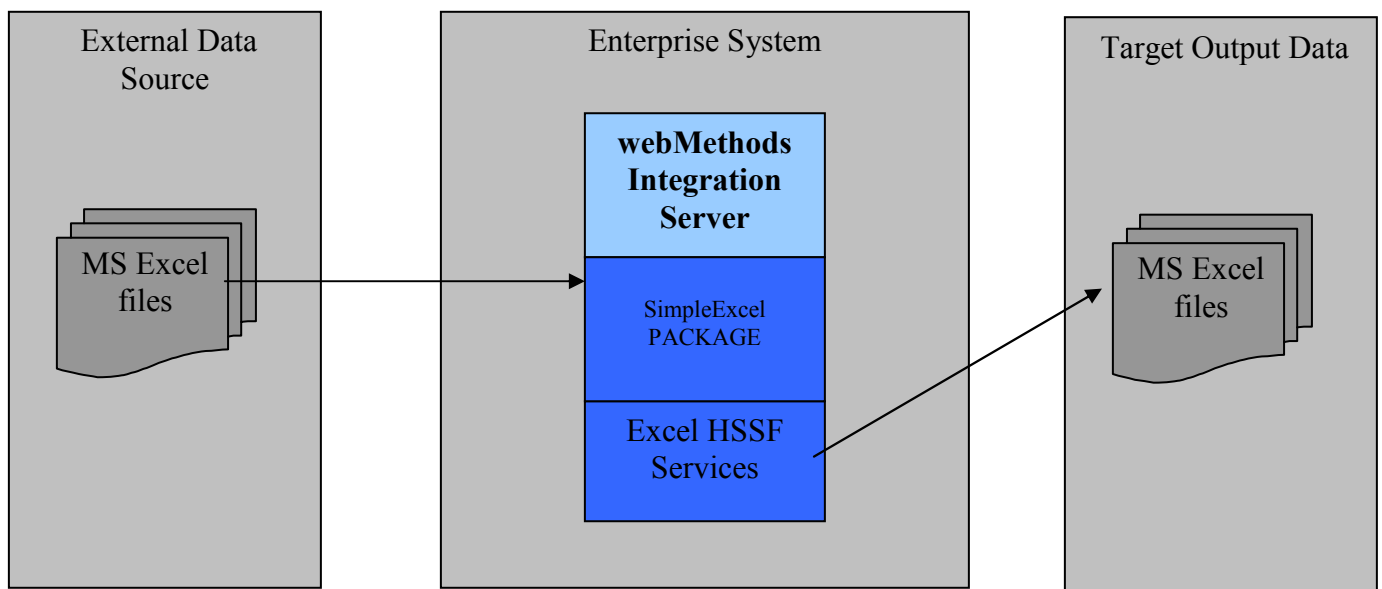
Purpose

This document describes the technical information for webMethods SimpleExcel utility package, and its deployment.

Architecture Overview

The SimpleExcel package is designed for webMethods Integration Server (IS) this package will allow IS to read, create, query and map MS Excel documents directly at native binary level.

This package provides several utility services and samples to handle MS Excel file content inside Integration Server.



SimpleExcel Package Detailed Description

SimpleExcel Package Components

This section specifies the components of Simple Excel package

SimpleExcel_RV_1_0.zip – Integration Server standard package enabling core functionality of Simple Excel.

readme.txt – provides latest updates and installation instructions

SimpleExcelUserGuide – User guide documentation

sample.xls – sample excel file

poi-bin-2.5.1-final-20040804.zip – complete distribution of Apache POI library and documentation. For information about Apache POI project and latest updates please reference official Apache www.apache.org web site.

System Requirements

This section specifies SimpleExcel Package's system requirements.

Installed and configured webMethods Integration Server 6.1

Memory – min 256MB max 1024MB (This may vary per installation and volume load)

Platform and Operating System	Supported JREs
Microsoft Windows 2000 Professional, Server, and Advanced Server	JRE 1.3.1 and 1.4.2
Microsoft Windows XP Professional	JRE 1.3.1 and 1.4.2
Microsoft Windows Server 2003 Standard Edition, Enterprise Edition (32-bit), and Datacenter Edition (32-bit)	JRE 1.3.1 and 1.4.2
Sun Solaris 8 and 9 (both 64-bit)	JRE 1.4.2
Hewlett-Packard HP-UX 11i v1 for PA-RISC(64-bit) and HP-UX 11i v2 for Itanium2	JRE 1.4.1
IBM AIX 5.1 and 5.2 (both 64-bit)	JRE 1.4.1
Red Hat Enterprise Linux AS 2.1 and ES 2.1	JRE 1.4.2

SimpleExcel Deployment Directory structure

This section specifies the SimpleExcel package's directory structure for deployment. The directory structure follows standard Integration Server package directory structure.

```
<Integration Server>
  packages
    SimpleExcel
      code
        classes
          SimpleExcel
        jars
        source
          SimpleExcel
      config
      doc
      lib
      ns
        SimpleExcel
          sample
          util
      pub
      resources
      templates
      web
```

SimpleExcel package is installed as standard Integration Server package.

SimpleExcel Configuration

This package does not require any configuration on the server.

Working with Excel File Data

This section describes the method of handling MS Excel data within webMethods Integration Server. The SimpleExcel package provides methods to read and create binary Excel documents. As noted earlier the MS Excel data is stored in binary format known as Microsoft's proprietary OLE 2 Compound Document file format. In order to make this data available to webMethods platform SimpleExcel package provides several utility services to convert Excel binary data into webMethods record or produce an Excel file from a webMethods record.

User may choose to create a custom flow or java service or use SimpleExcel package services to handle MS Excel binary data.

To read MS Excel file with webMethods platform SimpleExcel package provides a services *SimpleExcel.util:MSEExcelWorkSheetToRecord*. This service has several options of accepting Excel data, as fully qualified file path, input stream of data or binary array of data. Depending on the requirements user can read Excel file using built in webMethods service such as *pub.fil:getFile* or custom Java service.

To create excel data from webMethods record user may use service provided by SimpleExcel package: *SimpleExcel.util:RecordToMSEExcel*

Note: SimpleExcel package provides two sample flow services that demonstrate basic MS Excel file handling. The *SimpleExcel.sampletest_read* and *SimpleExcel.sample:test_write* services demonstrate how to convert Excel data into webMethods internal record format so it can be manipulated. How to create Excel from webMethods record and output data into file, stream or binary data array

SimpleExcel Services

This section provides details on utility services provided by this package.

Service Name	Description
SimpleExcel.util:RecordToMSEExcel	Convert WM record into excel binary data. Based on option binary excel data can be written directly to file, stream or binary array
SimpleExcel.util:MSEExcelWorkSheetToRecord	Convert excel data or file into WM record. Record structure is created dynamically based on MS Excel file Row/Cell tabular structure
SimpleExcel.util:query_cell	Return cell data based on worksheet, row and cell number.
SimpleExcel.sample:test_read	Flow service to convert a sample Excel file into webMethods dynamic record structure.
SimpleExcel.sample:test_create	Flow service creates set of records and then converts them into Excel data and writes it as a file on local C: drive
SimpleExcel.util:qry	Wrapper Flow service to query cell data inside MS Excel data file

Known issues and limitations

This section provides details on known issues and limitation in processing binary excel files. The Excel binary files that are corrupted or incomplete cannot be read by SimpleExcel package. In some special cases when excel files are generated by MS Excel application that has some issues the resulting file may be invalid even though it may be readable by another MS Excel installation. In these cases it is recommended to resave the file in a standard valid install of MS Excel software and retry processing.

The SimpleExcel package's MS Excel interpreter functionality is based on open source Apache POI project. And as such it will have all the limitation in processing MS Excel files as POI does at this point. At this time Apache PIO final release is known to be one of the most complete ports of Microsoft's OLE 2 Compound Document format using pure Java. For most current updates and issues with POI refer to www.apache.org/poi.

It has been noted that some older versions of MS Excel may have written some special headers that are incompatible with POI. Such files may have to be modified or saved by more recent version of MS Excel to function in this package.

File Size

File size is a common limitation when it comes to webMethods and any type of processing. The SimpleExcel provides standard processing and falls under similar file size limitations as does Integration Server. The issue is, excel files are binary data and unlike XML they must be read completely into memory in order to be processed. Therefore it is not possible to employ large file processing for Excel binary data. It is not possible to read 30% or 50% of a binary excel file and expect it to be translated. This will be considered as incomplete or invalid data.

The package reads data from a stream in buffered manner but it must get complete data to translate the file correctly. Normally in Integration Server there will be only single copy of the data read into memory. This may serve as guideline for sizing your excel file processing with Integration Server memory requirements.

MS Excel Formula Handling

Current version of SimpleExcel Package uses release 2.5 version of POI that has complete support for Excel Formulas. POI version 3.0 has added more complete Formula interpreter however it is still in Alpha stages.

Normally if MS Excel file containing formulas is read by SimpleExcel package it will interpret the formula and return the result value of type double or a string. However, in some cases POI may not be able to interpret the formula and therefore result values will return default string: NO IDEA SHARED FORMULA EXP PTG. This text comes from POI API and until formula processor is enhanced its support will be limited.