

Data File Converter for AB1, FSA and SDS files – Release Notes

Introduction

The Data File Converter is a console application that enables you to export data from the following data files into text and XML formats:

- .ab1 (sequencing)
- .fsa (fragment analysis)
- .sds (real-time PCR)

The converter exports all the data elements inside these data files, along with their data types and corresponding values. The converter has a number of command line options to support various types of operations.

Installation

All the components of the Data File Converter are delivered in a single compressed archive. Extract the contents of the archive to a folder on the system where you want to use it. The archive contains one executable, DataFileConverter.exe, plus some required DLLs. There is also this readme document and a *Test* folder.

The *Test* folder contains a few sample AB data files (AB1, FSA and SDS), three example XSL stylesheets and one batch file, test.bat. You can use test.bat to verify proper Data File Converter operations. Double click on it to launch, and after it terminates examine the Test folder. There should now be one text and one xml files for each sample data file. Use your web browser to open and inspect the XML files. Use Notepad or another text editor to display the text files.

General Usage

In general, the Data File Converter can be launched from a console window, or from any external script or process that can run a console application. When you use the Data File Converter, you can specify various options to help you convert one or more files, with more or less details, in one or more output formats. Multiple options must be separated by a space. Order is not relevant. The general usage syntax is:

```
DataFileConverter -i InputFile [-o OutputFolder] [-y StyleSheet] [-v] [-s] [-d] [-t] [-x] [-c] [-u] [-h]
```

Files and folder names with spaces must be qualified using double quote e.g.

```
DataFileConverter -i "c:\My Data\In\*.ab1" -x -o "c:\My Data\Out" -d -s
```

The input file to be converted must not be locked by other application. If the output file already exists it will be overwritten.

Output Files

The output files can be tab delimited text or XML. The content of the exported files consists of one or more headers, followed by a series of data structures. For AB1 and FSA files, this is a series of data elements or tags, each with the same format:

- Tag, ID, Description, Type, Elements, Size(bytes), Data.

For SDS files, the header is followed by one or more different structures, if they are present in the data:

- Plate: Name, Description, Column, Row, InUse, Detector(s), Task(s), DetectorUse(s), Quantity(s)
- Detectors: ID, Detector, Reporter, Quencher, Description, Comment, Owner, Locus, Color, Start, End, Threshold, AutoCT_Quality, AutoCT_Baseline, AutoCT_Threshold
- Markers: ID, Marker, Description, Comment, AlleleX, AlleleY, Color, CV, Owner, AutoCT, Cluster
- Calibration: WellNumber, Constant, ChannelCount, SpectralValues

All remaining data elements are output as groups of tags, each with format: Component, Tag, Size, Type, Value.

DataFileConverter.exe command line options

Multiple options can be specified on the command line as from the list detailed below. Multiple output types can be specified, the example below creates both text and xml output files (-x -t)

```
DataFileConverter -i "c:\\My Data\\In\\my file.ab1\" -v -d -x -t
```

The table below describes the Data File Converter options.

OPTION	DESCRIPTION	COMMENTS
-i	Input Filename	<ul style="list-style-type: none"> This is the only required option. Specifies a valid path to the source file(s) to convert. With no other option, DataFileConverter.exe opens the specified file(s) and displays the conversion progress on the standard output. Can use wildcards to match and process multiple files. If filename or folder contains whitespaces, must use double quotes, e.g. <ul style="list-style-type: none"> o -i Data001_*.ab? o -i "C:\My Data\Data"
-o	Output Folder	<ul style="list-style-type: none"> Specifies a valid name of an existing folder where DataFileConverter.exe creates output file(s). If not specified, output files are placed in same folder as source file(s). Folder must exist, wildcards are not allowed.
-y	XSL style sheet	<ul style="list-style-type: none"> Optional, specifies a valid Extensible Stylesheet Language (XSL) document filename. A web browser can display XML documents based on an appropriate XSL document. When this option is specified, DataFileConverter.exe adds the stylesheet name in the output XML file (requires -x) There are three examples of stylesheets provided for different types of AB source documents. <p>Note: Additional information on XSL specifications is available on the World Wide Web Consortium (W3C) Internet site.</p>
-v	Verbose	When specified, the description of ABIF and FSA tags is added to the output file(s). Not used for SDS and SDT files.
-s	Silent	When specified, DataFileConverter.exe does not display progress information on the standard output (console). Errors and warnings are still displayed.
-d	Includes Header	When specified, DataFileConverter.exe adds file header information at the beginning of the output file(s). In the XML output there is a header element name.
-t	Tab Delimited Text Output	Creates an output text file in tab delimited format. The file has the name <code>inputfilename.xyz.txt</code> , where <code>inputfilename.xyz</code> is the name of the input file specified by the -i option. If -o is also specified, the final filename is going to be <code>OutputFolder\inputfilename.xyz.txt</code>
-x	XML Output	Creates an output file in XML format. The file has the name <code>inputfilename.xyz.xml</code> , where <code>inputfilename.xyz</code> is the name of the input file specified by the -i option. . If -o is also specified, the final filename is going to be <code>OutputFolder\inputfilename.xyz.xml</code>

-c	Prints character array as string	When selected, DataFileConverter prints an array of characters as a string. Non-printable characters in the array are printed as dot (.). This option only affects tab delimited text output.
-u	Checks for Data File Converter updates on AB website	Creates a web browser window and navigates to the Applied Biosystems product web page where users can check for the most current program information and updates.
-h	Shows help	Shows a brief help text

Version History:

Version 1.0 – July 2006

AB1 and FSA files are exported as a sequential dump of all tags. SDS and SDT files are exported as a sequential series of partially-organized components. They include Plate Setup, Detectors, Markers, Dyes, Calibration and Thermocycle. Unknown components are exported as sequential dump of tags.

Known issues:

- Data File Converter attempts to detect XML-reserved characters in the input source, such as <> % and &, and replaces them with a dot (.) character. If for example the source file has a comment that says “Method A&B”, it will be replaced in the output XML file as “Method A.B”. This is required to avoid creating invalid (not well formed) XML files. No replacement is done in the tab delimited text output.
- This version does not support the conversion of data files generated by the Applied Biosystems 7900HT Fast Real-Time PCR System. The Data File Converter will process the 7900HT data files but report incorrect values for a number of fields such as Calibration and the “Quantity(s)” for Plate Setup and Detectors components.