c2x-wizard

Manual

(Regarding version 0.1.0)

Document history

Version	Date	Author	Reason/Changes
0.1	01/12/2012	J. Gödeke	Initial creation

Table of contents

1   csv2xml

1.1   Prologue

I'm always surprised how often CSV files have to be converted to XML. Thereby the first version of my converter csv2xml which I had created in 2002 was only a demo for a new CSV parser class which I had written in C++. I had released the software on my homepage and was amazed how popular it got. Soon the first change requests had reached me and favorable was the wish to use csv2xml on non-Windows systems. In 2003 I divided csv2xml in two tools. One was the command line converter csv2xml and the other was its calling graphical application csv2xml_win. It was the first time that csv2xml was also released for Linux. In 2007 I've renamed csv2xml into c2x-cmd, because in between its first release and 2007 many tools like csv2xml were published. c2x-cmd was the first version that had been compiled for popular other operating systems like AIX and MAC OS X. The problem was the replacement of csv2xml_win. This took about 4 years, cause my family and my job didn't offer me enough time for that. The other problem was that I had to find a suitable programming language which would enable me to write a graphical application that works on each operating system that was supported by c2x-cmd and its ansi c++ code. In the end I decided do use Java and Oracles (former SUN) integrated development environment Netbeans. I hope that the resulting application is impressive enough for you that you are willing to install an Java Runtime Environment if you haven't already done that.

1.2   License

c2x-cmd & c2x-wizard are subject of a FREEWARE License. That means that you could use these applications for free whether in a commercial or non-commercial purpose. You must not sell or change these applications. I don't give any warranty for damages that happen by using these applications.

1.3   Contact

If you have further questions, change requests or error reports don't hesitate to get in contact with me.
`Jens Gödeke`

Please use my contact page to get in contact with me.
http://www.jens-goedeke.eu/contact

You'll find the newest version of the applications or the documentation on the homepage of c2x-cmd :
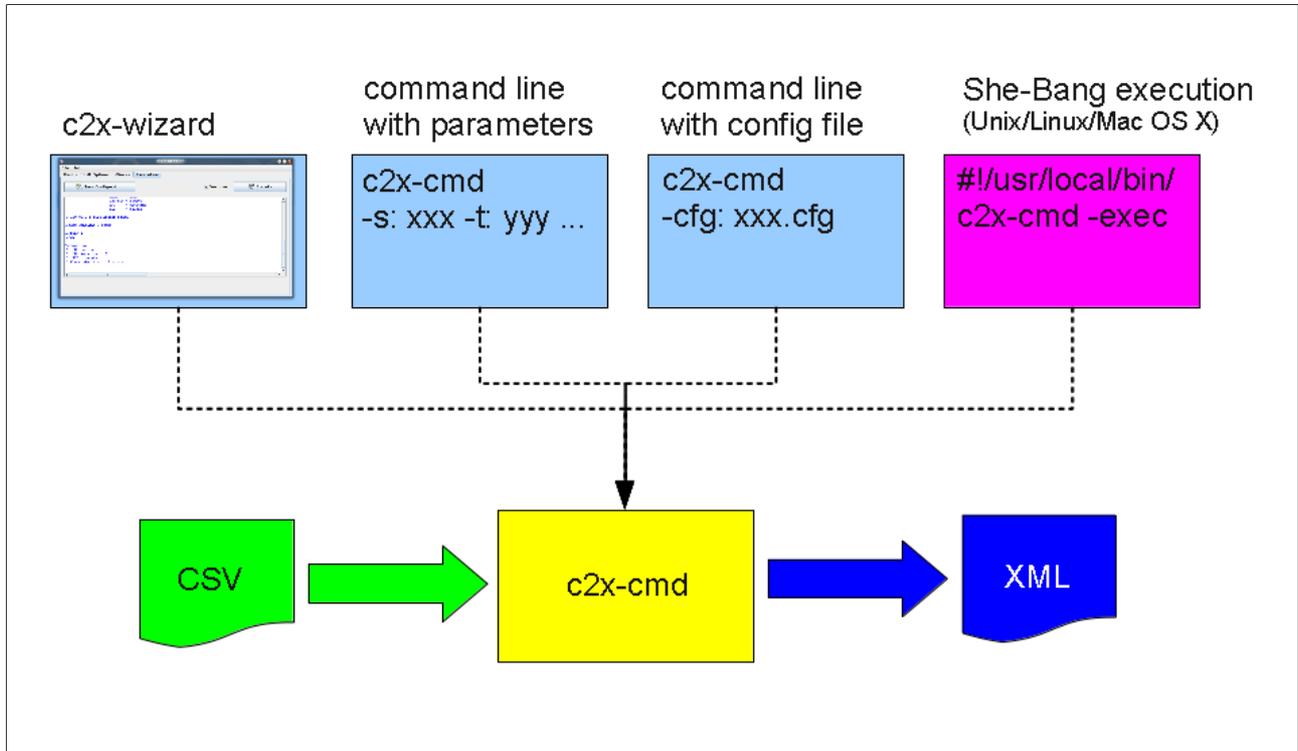**http://www.jens-goedeke.eu/tools/csv2xml**

1.4   FAQ

Not yet.

2   Overview

c2x-cmd is the command line converter which could be started in four different ways
Hint : Only the first three ways are supported by Windows.

2.1   Starting c2x-cmd from a command line

The advantage of a direct call of c2x-cmd from a command line is that you may run c2x-cmd in an automatic process like a shell or a batch script.
The execution over a command line works:

- if you give the necessary parameters as a sequence
- by giving a configuration file (maybe generated with c2x-wizard) or

unfortunately only under OS which are based on Linux or Unix:

- by using She-Bang[1]-Execution

You find more details on how to use c2x-cmd in its own documentation. See c2x-cmd manual.

2.2   Starting c2x-cmd from c2x-wizard

c2x-wizard is the generator tool which should help the user to set the right options with an graphical application. c2x-wizard has the following tasks:

- Supporting the user so that he is able to set the c2x-cmd options correctly
- Capability to save the user settings as configuration files
- Capability to send the user settings directly to c2x-cmd and execute c2x-cmd immediately afterward.

---

[1] see http://en.wikipedia.org/wiki/Shebang_%28Unix%29

2.3   Installation and Execution

2.3.1   Windows

You may download the ZIP file from my homepage and extract its content. If you want to start c2x-cmd from a batch file or another script type you have to add c2x-cmd to your search path. You can get more information on how to do this in the document "c2x-cmd manual".

If you want to execute c2x-wizard you may create a shortcut on your desktop. To create this shortcut you may navigate to your desktop and click your right mouse button. A pop up menu appears and then you should click on the menu item "new" -> "shortcut" to create the new shortcut. Select the file "c2x-wizard.bat" and go further through the setup process.

The application should start after a double click on the new created shortcut. An error has occurred if not c2x-wizard but the internet explorer starts showing a list of possible errors. The command line window shows the current error behind the message "Error : runJava ends with an error code :" (see red arrow in the following figure). The most probably error is a non installed Java Runtime Environment (JRE).
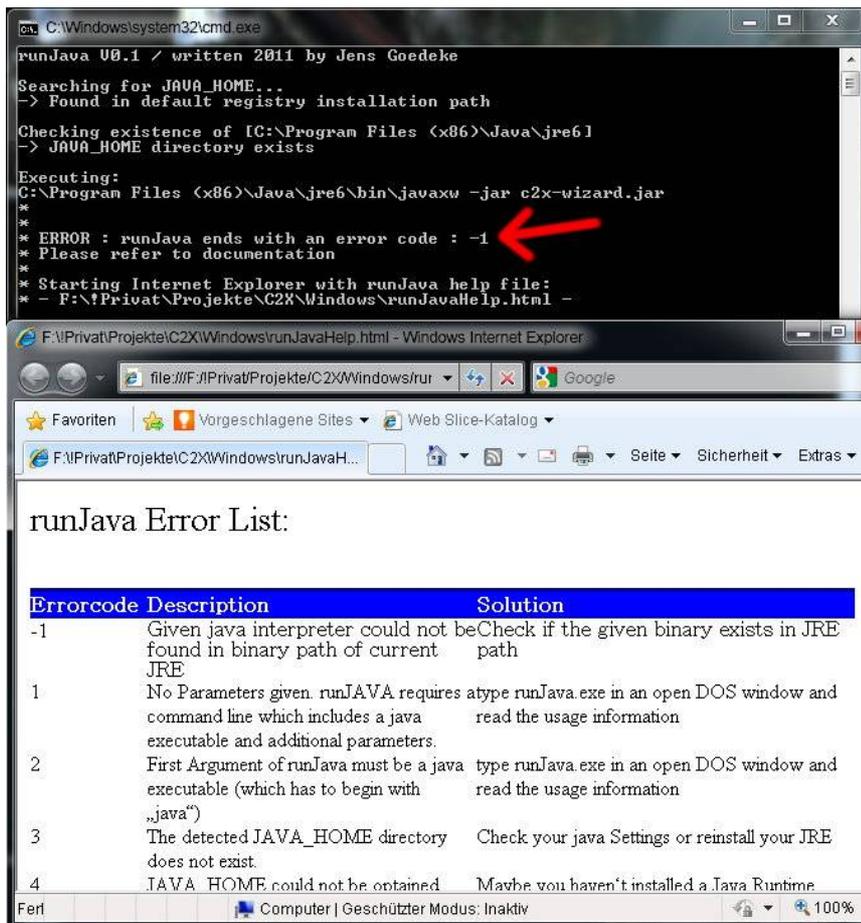
If the reason for the failed execution of c2x-wizard was the non installed JRE you may click on the link in the error list in order to get to a detailed description how to install a JRE under Windows.

Unlike Unix/Linux based operating systems Windows (Win95 – Win7) does not support one single script language that allows to download documents directly over your internet connection. Therefore the tool *get_manuals.bat* asks only for your preferred language and leads you to my homepage afterward. You could go directly to my homepage, if you click on the following link:
http://www.jens-goedeke.eu/tools/csv2xml

2.3.2 All operating systems based on Unix-/Linux except Mac OS X

You may download the tar.gz file from my homepage and extract its content. Unlike the Windows ZIP file c2x-cmd is not included in the tar.gz package. Download the c2x-cmd converter which is suitable for your OS from my homepage:
http://www.jens-goedeke.eu/tools/csv2xml
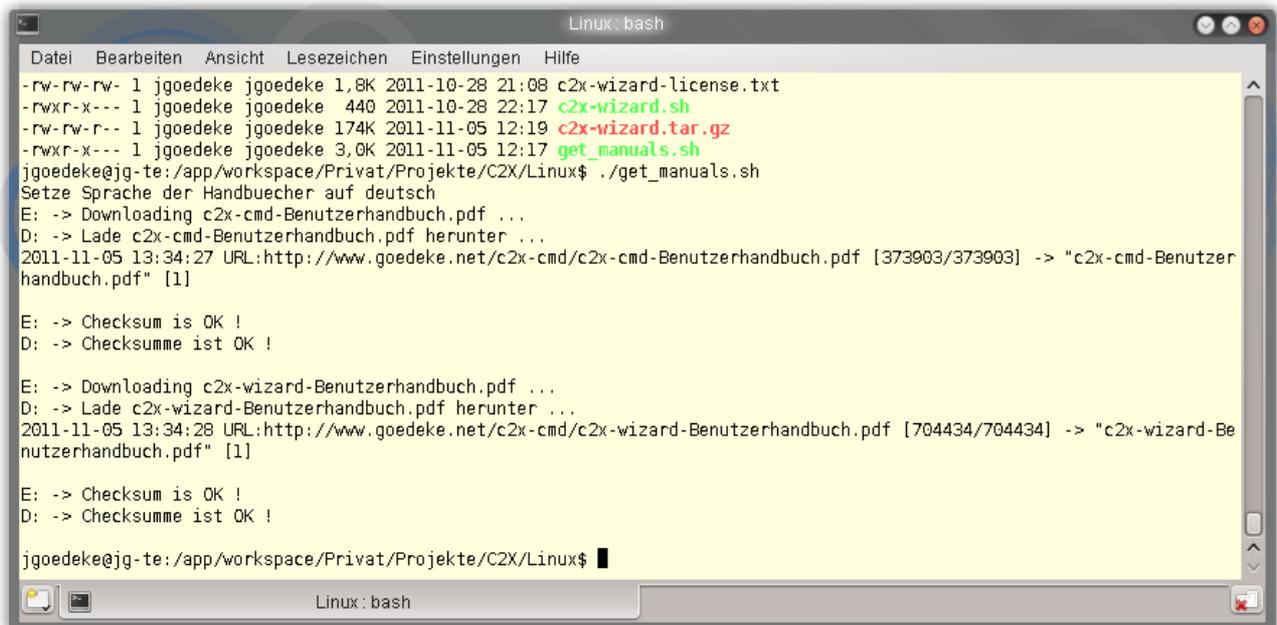If your OS is not mentioned you may send me an email.

After you got c2x-cmd from my homepage you should extract it into the same directory in which c2x-wizard is extracted or choose a directory of your search path (e.g. „/usr/local/bin" etc.) if you have the necessary rights to do that. The extraction into one directory of your search path is useful if you plan to integrate c2x-cmd into your scripts.

If you want to start c2x-wizard you might call the shell script „c2x-wizard.sh" which needs a Bourne-Shell compatible shell like SH, KSH, BASH etc.. Otherwise you can call it directly from a command line if you type:

```
java -jar ./c2x-wizard.jar
```

The start script „c2x-wizard.sh" checks first and foremost if a java runtime environment (JRE) is installed. If no JRE is installed you will see an error message. A JRE is installed on most Linux systems by default. C2x-wizard requires a JRE 1.6 or higher and maybe you have to update your system for this. This could be done with the standard package installation tools like dpkg, apt, aptitude, emerge, yast2 etc. If you are only a normal user (without root rights) you may ask your administrator to update your system.

To get the newest version of the c2x manuals you can use the bash script *get_manuals.sh*. This script uses the additional tools „wget" and „md5sum" which should be installed on your system by default. Otherwise proceed as mentioned before with your installation tool.

2.3.3 Mac OS X

You may create a new folder under your program or your document folder and extract the ZIP archive into this new folder. Unlike the Windows ZIP file c2x-cmd is not included in the tar.gz package. The suitable converter *c2x-cmd-macosx-ppc-x86.tar.gz* can be downloaded from my homepage: http://www.jens-goedeke.eu/tools/csv2xml

Extract the contents of the archive into the same folder where you have extracted c2x-wizard into. Like under Linux/Unix you can start the wizards by starting the shell script „c2x-wizard.sh". Unlike any other unix based os Mac OS X does not allow the direct execution of scripts from its file browser (finder). Therefore you have to start a terminal window. This could be found under *Programs → Tools → Terminal*. Go to the directory where c2x-wizard is extracted to and start c2x-wizard by typing:

```
./c2x-wizard.sh
```

```
Terminal — bash — 92×19
localhost:c2x-wizard jgoedeke$ ls -la
total 2968
drwx------@ 14 jgoedeke  staff     476  5 Nov 11:39 .
drwx------+  9 jgoedeke  staff     306  5 Nov 11:39 ..
-rw-------@  1 jgoedeke  staff    6148  5 Nov 11:40 .DS_Store
-rwxr-xr-x@  1 jgoedeke  staff  196296 19 Dez  2007 c2x-cmd
-rw-r--r--   1 jgoedeke  staff  373903 27 Okt 15:57 c2x-cmd-Benutzerhandbuch.pdf
-rw-r--r--   1 jgoedeke  staff  704434 29 Okt 02:27 c2x-wizard-Benutzerhandbuch.pdf
-rw-r--r--@  1 jgoedeke  staff  206358 28 Okt 23:47 c2x-wizard.jar
-rwxr-x---   1 jgoedeke  staff     440 28 Okt 22:17 c2x-wizard.sh
-rwxr-xr-x@  1 jgoedeke  staff     380 16 Dez  2007 csv2xml
drwxr-xr-x@  6 jgoedeke  staff     204 16 Dez  2007 examples
-rwxr-x---@  1 jgoedeke  staff    2974  5 Nov 11:22 get_manuals.sh
-rw-r--r--@  1 jgoedeke  staff    1823 28 Okt 21:08 license.txt
-rwxr-x---@  1 jgoedeke  staff    3679 19 Dez  2007 release_history.txt
-rwxr-x---@  1 jgoedeke  staff    3547 19 Dez  2007 usage.txt
localhost:c2x-wizard jgoedeke$ ./c2x-wizard.sh
localhost:c2x-wizard jgoedeke$ 
```

In order to download newer manuals automatically from my homepage you need the gnu tool *wget*. This tool can be downloaded in a compiled version from the following homepage:
http://www.merenbach.com/software/wget

If you have installed wget, you can call the bash script *get_manuals.sh* to check for newer documentation.

3   c2x-wizard

The c2x-wizard supports almost all options of c2x-cmd. This options are explained in this chapter while the non-supported options are explained in the document "c2x-cmd-manual.pdf". There are only a few mandatory options which have to be set in order to run a successful conversion. This options are marked with a tailing „(m)" (m= Mandatory). Front end elements (like buttons etc.) are printed this way : Front end elements

The Front end of c2x-wizard is divided into a simple menu and 4 tabs on which all functions are accessible.

3.1   The menu
The file menu includes the options Save as (which is additionally placed as a button called Save Configuration on tab Execution) and Exit. With Exit you are able to exit the program. Under Help you'll find a menu item called About on which the current version information is shown.

3.2   The tab Basics
As already realised in its predecessor csv2xml_win some of the parameters are preselected. The minimal necessary steps to run a conversion are to set the Source filename and the Target filename or its automatic calculation option Get target filename from source filename.
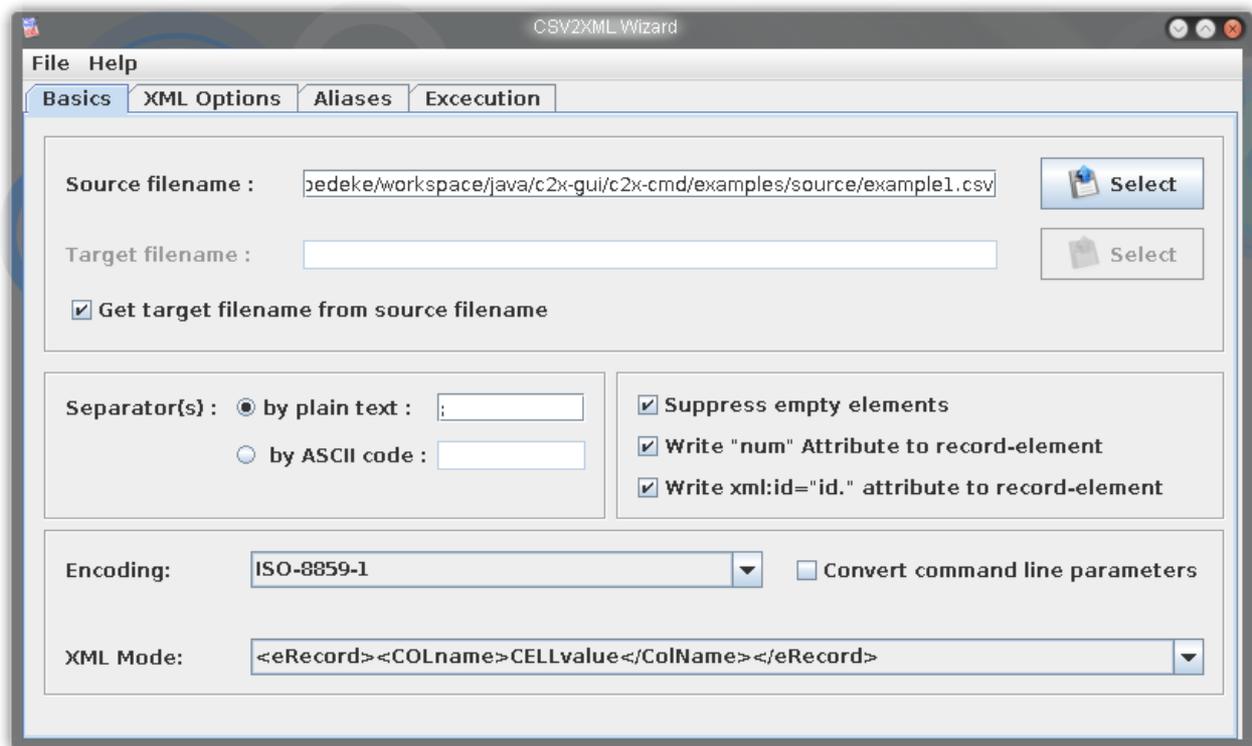
3.2.1   Option : Source filename (m)
This option defines the path of the csv source file. Its possible to enter this path directly or to obtain it from a file selection dialog which appears if you click the button called Select.

Within each of these file selection dialogs (see the following screenshot) there is a predefined file filter. If you want to see all files you can do this by select another file filter from the combo box Files of Type

.
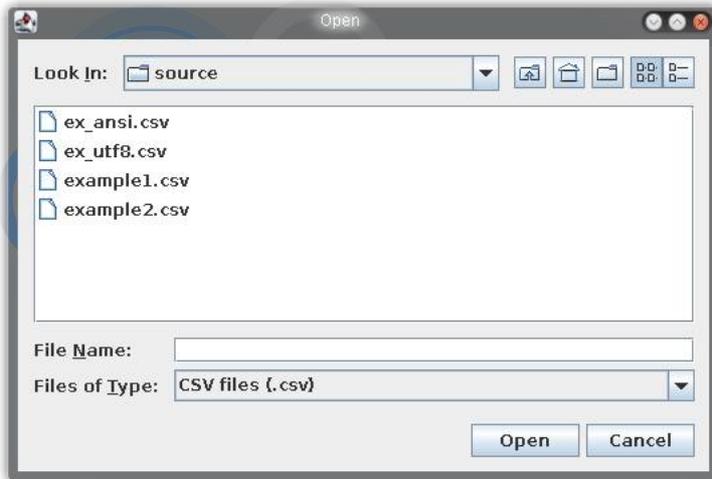
3.2.2 Option : Target filename (m)
This option defines the path of the target xml file. Its possible to enter this path directly or to obtain it from a file selection dialog which appears if you click the button called Select.

3.2.3 Option : Get target filename from source filename
If the target file name should be the same as the source file name except the extension you can choose this option. If this option is checked the Target filename option is disabled.

3.2.4 Option : Separators by plain text
This option defines one or more separators which have to be used by c2x-cmd to split the columns of the csv source file. The default value is a semicolon.

3.2.5 Option : Separator by ASCII code
This option defines a single separator by its ASCII code. This enables you to set separators which could not be given by command line directly. For example: If you want to set a tabulator as the separator you have to choose the ASCII value „9". A space has the ASCII code „32" and so on.

The complete list of ASCII characters and its codes can be obtained from this URL:
http://en.wikipedia.org/wiki/Ascii

3.2.6 Option : Encoding
This option defines the encoding of the target xml file.
You can choose between two different encoding strings:

- ISO-8859-1
  This is nearly equivalent to the Windows character set „1252". Each character uses exactly one byte.
- UTF-8
  This is the default character set of most Linux distributions. Language specific characters use two bytes for each character.

Most programs that are able to handle UTF-8 files write a so called BOM (Byte-Order-Marker) to the file header. c2x-cmd is able to detect this BOM. So c2x-cmd can convert source files automatically if the source file encoding differs from the target file encoding.

3.2.7 Option : Convert command line parameters

This option defines if the parameters of the command line should be converted into another encoding. If you have selected an encoding of UTF-8 and are calling c2x-cmd (or c2x-wizard) under Windows you should check this option if for example your Aliases include special characters.

3.2.8 Option : XML Mode

C2x-cmd supports 5 different xml output modes. To explain this modes I used the following csv source file content:

```
Name;Type;Year-1987;Year-1995;Year-2001;Year-2006
Berlin;Cnty;BE;3.260.000;3.471.418;3.388.434;3.404.037
```

All other options are set to default.

Hint:

Identifiers which a leading small character (like *eRecord*) indicate that this identifiers can be renamed. The default values are printed afterward in square brackets. Example: *eRecord [record]*. More information regarding this renaming capabilities you might refer chapter 3.4.

3.2.8.1 Mode 1 : <*eRecord*><COLname>CELLvalue</ColName></*eRecord*>

Hint:

This mode was former labeled as „Elements with Element Content (default)".

Output:

```
<record>
    <Name>Berlin</Name>
    <Type>Cnty</Type>
    <Year-1987>BE</Year-1987>
    <Year-1995>3.260.000</Year-1995>
    <Year-2001>3.471.418</Year-2001>
    <Year-2006>3.388.434</Year-2006>
</record>
```

3.2.8.2 Mode 2 : <*eRecord*><COLname *aValue*="CELLvalue"/></*eRecord*>

Hint:

This mode was former labeled as „Elements with Attributes Values".

Output:

```
<record>
    <Name value="Berlin" />
    <Type value="Cnty" />
    <Year-1987 value="BE" />
    <Year-1995 value="3.260.000" />
    <Year-2001 value="3.471.418" />
    <Year-2006 value="3.388.434" />
</record>
```

3.2.8.3 Mode 3 : <*eRecord*><*eItem aName*="COLname">CELLvalue</*eItem*></*eRecord*>

Hint:

This mode was former labeled as „Attributes with Element Content".

Output:

```
    <record>
        <item name="Name">Berlin</item>
        <item name="Type">Cnty</item>
        <item name="Year-1987">BE</item>
        <item name="Year-1995">3.260.000</item>
        <item name="Year-2001">3.471.418</item>
        <item name="Year-2006">3.388.434</item>
    </record>
```

3.2.8.4 Mode 4 : *<eRecord><eItem aName*="COLname" *aValue*="CELLvalue"/></*eRecord*>
Hint:
This mode was former labeled as „Attributes with Attribute Values".

Output:

```
    <record>
        <item name="Name" value="Berlin" />
        <item name="Type" value="Cnty" />
        <item name="Year-1987" value="BE" />
        <item name="Year-1995" value="3.260.000" />
        <item name="Year-2001" value="3.471.418" />
        <item name="Year-2006" value="3.388.434" />
    </record>
```

3.2.8.5 Mode 5 : *<eRecord* COLname="CELLvalue"/>
Hint:
This mode was former labeled as „Attributes with Values".

Output:

```
    <record Name="Berlin" Type="Cnty" Year-1987="BE" Year-1995="3.260.000"
Year-2001="3.471.418" Year-2006="3.388.434" />
```

3.2.9 Option : suppress empty elements
Empty elements or attributes are suppressed if this option is set.
Note: For this example I've deleted the values of the fields „Type" & „1987":

```
Name;Type;Year-1987;Year-1995;Year-2001;Year-2006
Berlin;;;3.260.000;3.471.418;3.388.434;3.404.037
```

Output: Extract of the target file with non checked option:

```
<record Name="Berlin" Type="" Year-1987="" Year-1995="3.260.000" Year-
2001="3.471.418" Year-2006="3.388.434" />
```

Output : Extract of the target file with checked option:

```
<record Name="Berlin" Year-1995="3.260.000" Year-2001="3.471.418" Year-
2006="3.388.434" />
```

3.2.10 Option : write „num" attribute to record-element
This option defines if all records get an attribute which is containing its record number. The numeration starts with „1".

Output : Extract of the target file with non checked option:
```
<record>
...
</record>
```

Output : Extract of the target file with checked option:
```
<record num="1">
...
</record>
```

3.2.11 Option : write xml:id="id." attribute to record-element
This option defines if all records get an attribute which is containing its record number. Unlike the former option (3.2.10) this option uses the XML-ID attribute. The numeration starts with „1".
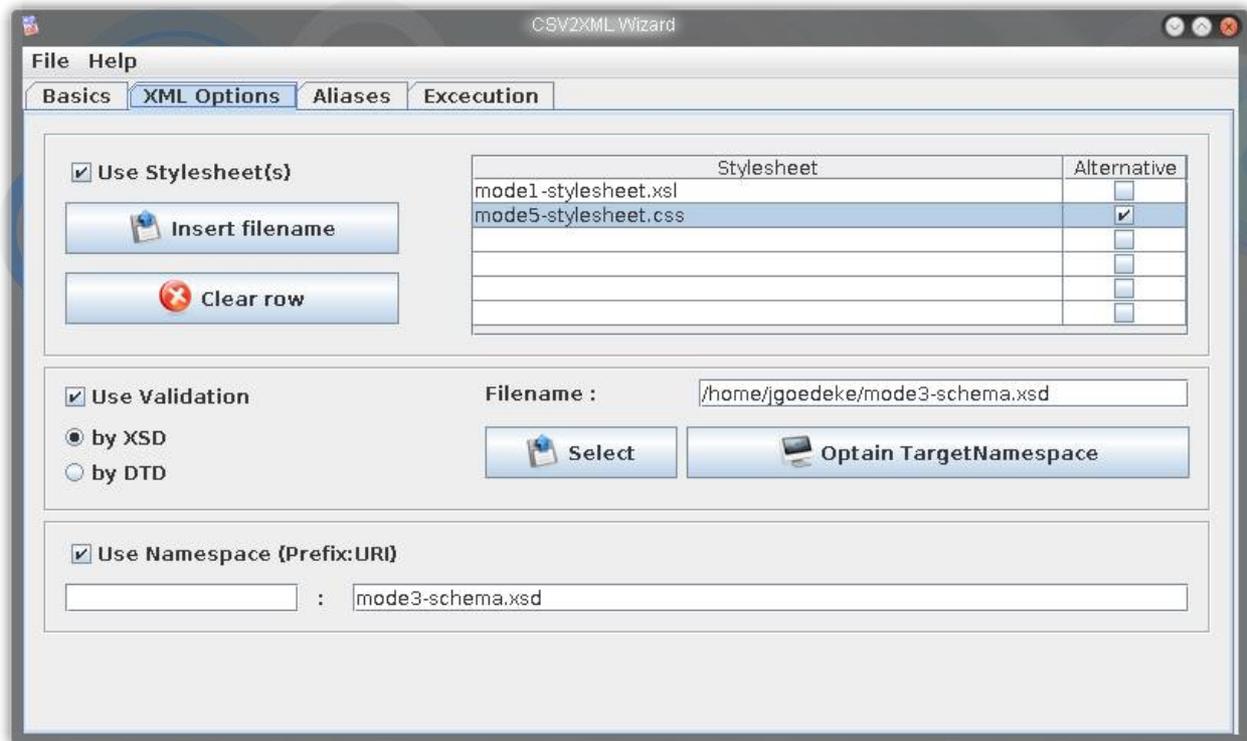
Output : Extract of the target file with non checked option:
```
<record>
...
</record>
```

Output : Extract of the target file with checked option:
```
<record xml:id="id.1">
...
</record>
```

3.3 The tab „XML Options"

3.3.1 Option : Use Stylesheet(s)

With this options it's possible to write one or more style sheet references into the header of the target xml file. This only an embedding – no transformation is done by c2x-cmd. To use this option the button Use Stylesheet(s) has to be pressed.

Two different kinds of stylesheets are supported:
- CSS – Cascading Stylesheets
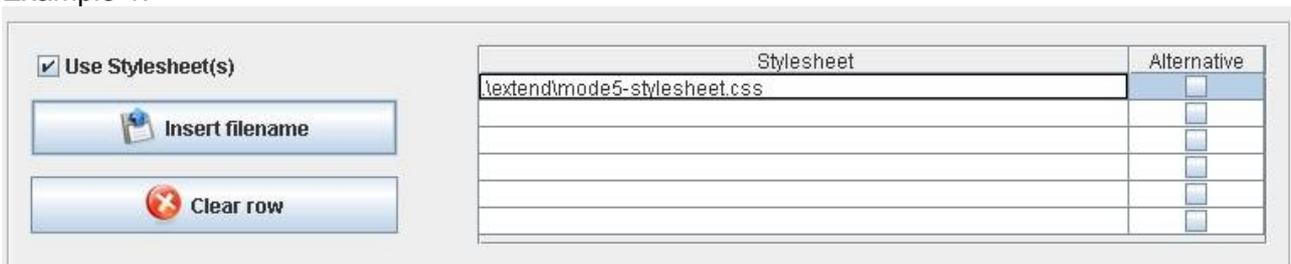- XSL – XML Stylesheets

Get more information about CSS files from : http://en.wikipedia.org/wiki/Css and
more information about XSL files from: http://en.wikipedia.org/wiki/Xsl

Usage:
In order to add a stylesheet you have to click on an empty cell of the column "Stylesheet". After that the buttons Insert filename and Clear row will be enabled. You can edit the content of a cell directly by double click with your left mouse button on it, or you simply press F2. If you want to obtain the file name from a file selection dialog you might click on the button Insert filename. If you want to remove one style sheet you have to click on Clear row.
Hint: Empty lines will be ignored in the xml target file.

Example 1:

Output:

```
<?xml-stylesheet type="text/css" href="./extend/mode5-stylesheet.css"
title="mode5-stylesheet" ?>
```

Example 2:

Output:

```
<?xml-stylesheet type="text/css" href=".\extend\mode5-stylesheet.css"
title="mode5-stylesheet" ?>
<?xml-stylesheet type="text/xsl" href=".\extend\mode1-stylesheet.xsl"
alternate="yes"  title="mode1-stylesheet" ?>
```

3.3.2   Option : Use Validation

With this option it's possible to add a reference to a Document-Type-Definition (DTD) or a XML-Schema-Definition (XSD). The file name can be entered manually into the field Filename or if you want to obtain the file name from a file selection dialog you might click on the button Select.
In case of a given xsd file the button Optain TargetNamespace will be enabled. If you click this button c2x-wizard tries to extract the value of the attribute „TargetNamespace" from the given file and puts its value into the field URI of the option Use Namespace.
Hint: This overwrites a possibly entered URI !

Example:

Output:

```
...
<csv_data_records xsi:schemaLocation="mode3-schema.xsd .\extend\mode3-
schema.xsd"
                  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                  xmlns="mode3-schema.xsd"
...>
```

3.3.3   Option : Use Namespace

With this option a name space URI and an optional name space prefix can be defined.
Example 1:

Output:

```
...
<csv_data_records xmlns="MyURI"
...
```

Example 2:

Output:

```
...
<MyPrefix:csv_data_records xmlns:MyPrefix="MyURI">
    <MyPrefix:record>
        <MyPrefix:Name>Aachen</MyPrefix:Name>
...
```
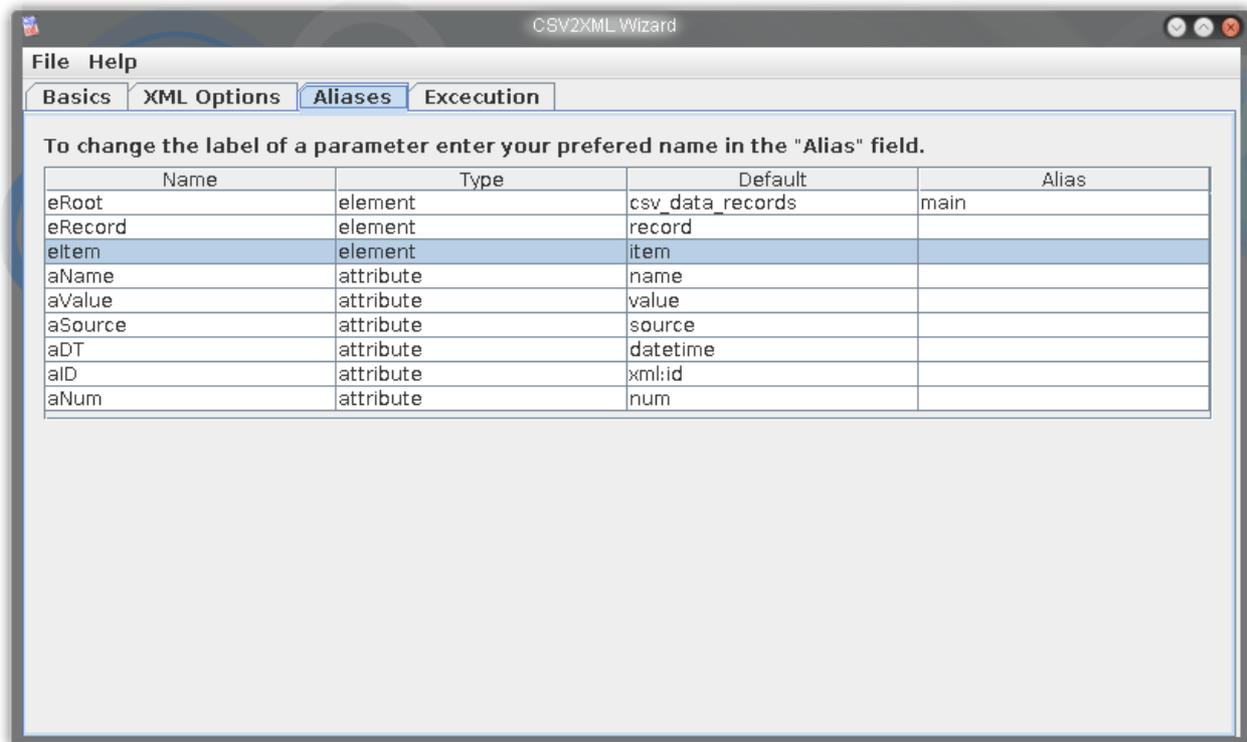
3.4 The tab „Aliases"

If you would like to rename element or attributes names you can do this with this option. The alias which has to be taken instead of the original name (see column "Default") has to be inserted into the field Alias. If this given aliases include non allowed characters c2x-cmd will automatically correct this by replacing the wrong characters by underscores ("_"). All alias parameters are listed in chapter 3.2.8.

Example:

To change the label of a parameter enter your prefered name in the "Alias" field.

Name	Type	Default	Alias
eRoot	element	csv_data_records	WurzelElement
eRecord	element	record	Datensatz
eItem	element	item	
aName	attribute	name	
aValue	attribute	value	
aSource	attribute	source	Quelle
aDT	attribute	datetime	Zeitstempel
aID	attribute	xml:id	
aNum	attribute	num	

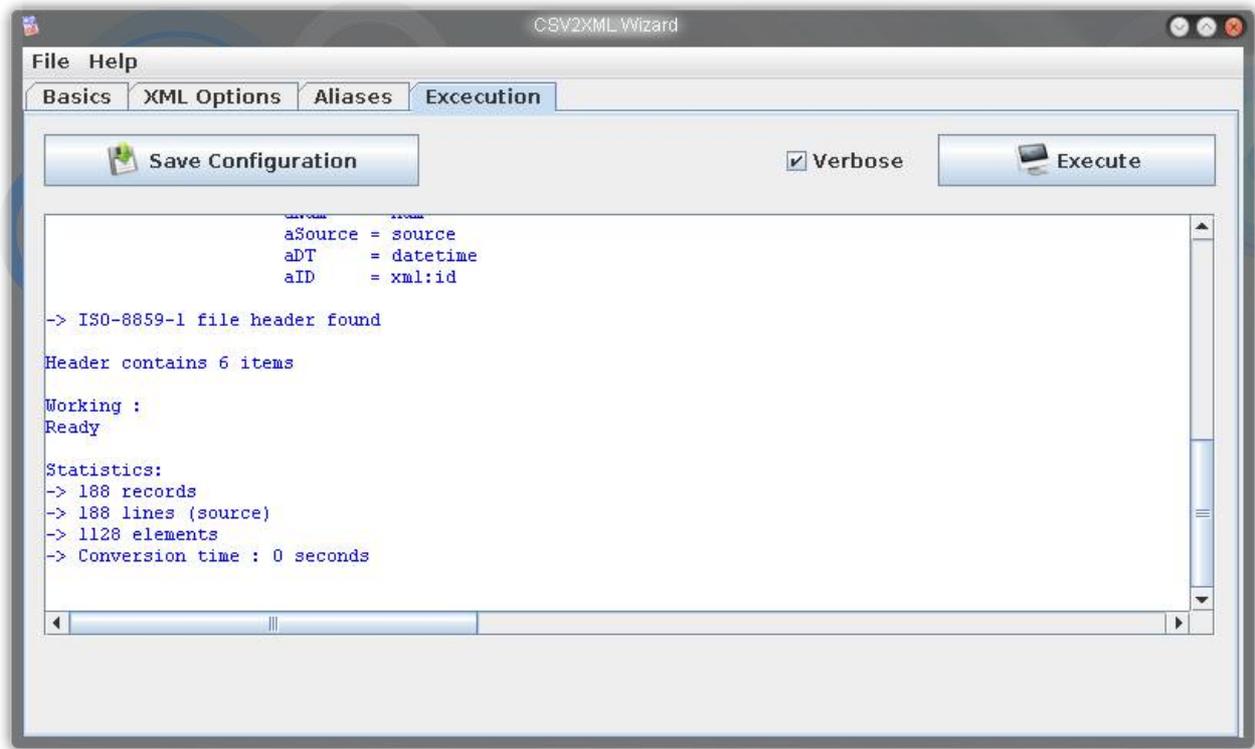Output:

```
...
<WurzelElement Quelle="example1.csv" Zeitstempel="2007-12-16T19:54:22+01:00" >
     <Datensatz>
       ...
```
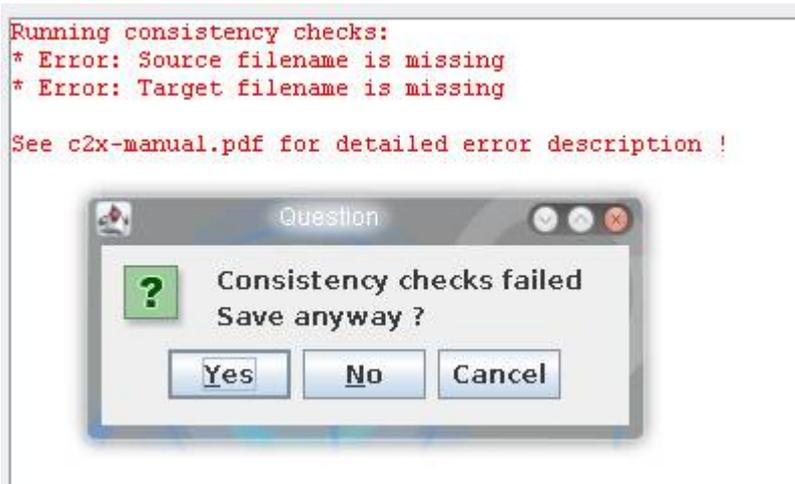
3.5   The tab „Execute"

3.5.1   Button: Save Configuration

With this button it's possible to store the settings of c2x-wizard into a c2x-cmd configuration file.
Right before the storing process c2x-wizard will do some consistency checks. The results of these checks are shown in the information area under the button. If errors occur the font color switches to red (otherwise the font color is blue):

```
Running consistency checks:
* Error: Source filename is missing
* Error: Target filename is missing

See c2x-manual.pdf for detailed error description !
```

In case of errors you can decide if you ignore these and save the file anyway (click on button Yes), or if you would prefer to correct the errors before storing the file (click on button No or Cancel).

If no errors occur or if you have pressed the button Yes a file selection dialog appears.
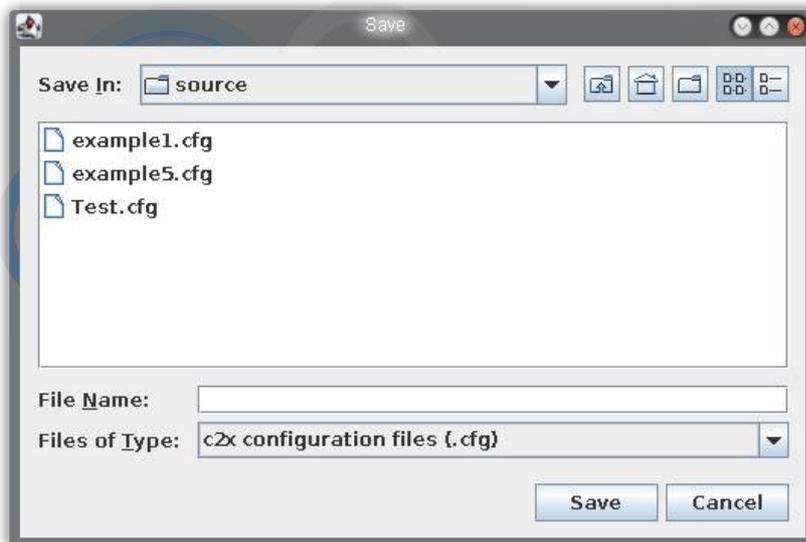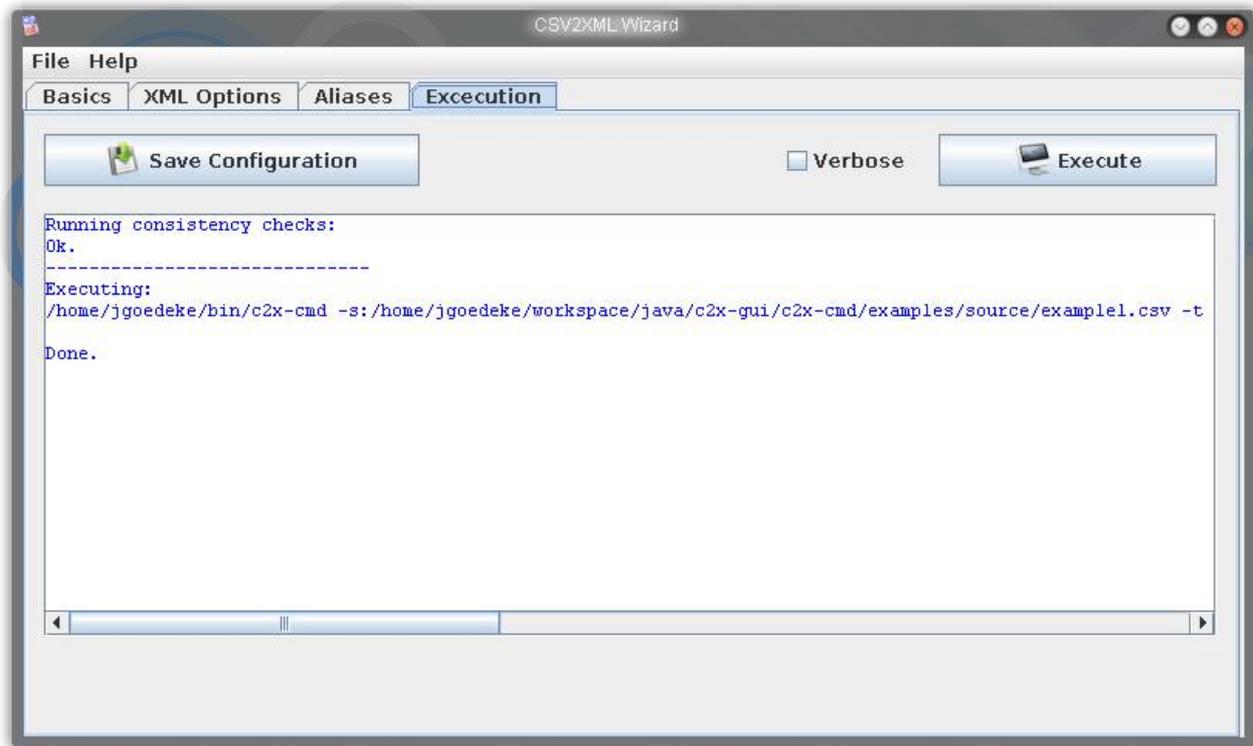
Now you have to enter a file name and after pressing the button Save you get an information whether the save progress was successful or not:
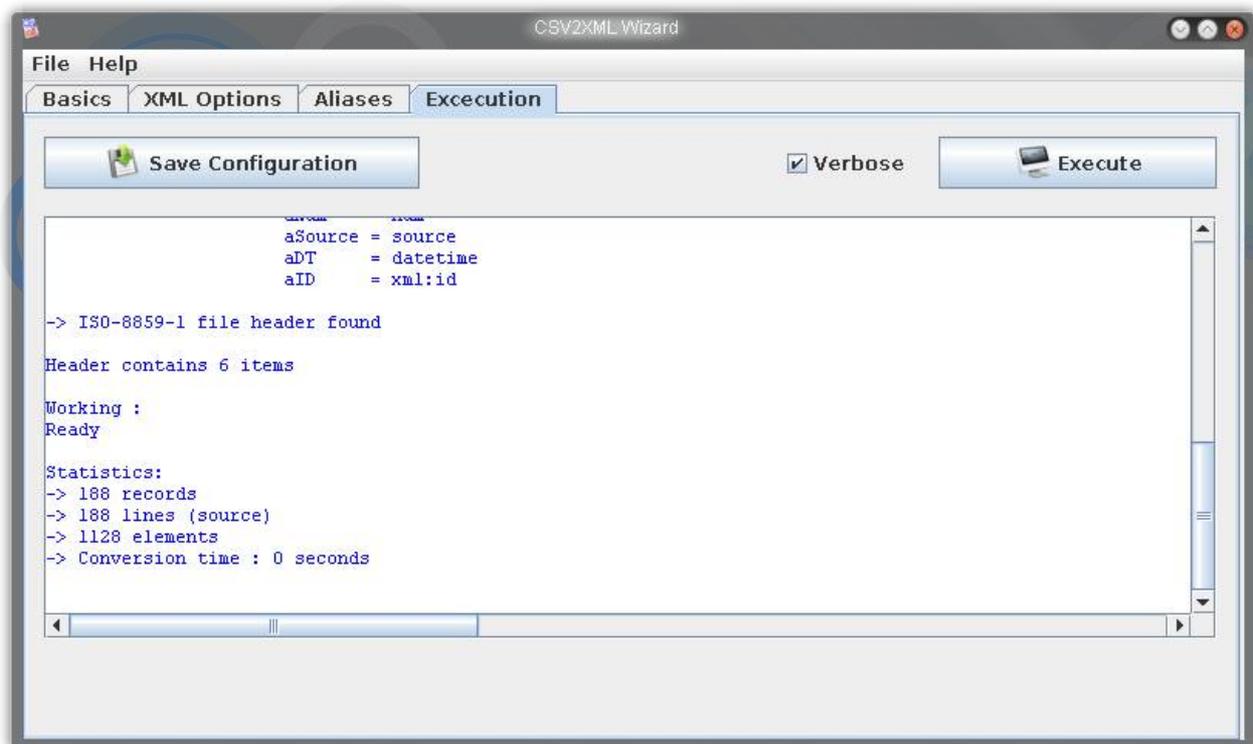
3.5.2 Option : Verbose
This options defines if c2x-cmd is started in verbose mode or not.
Example 1: (Non-Verbose)

In this case only the call of c2x-cmd followed by a "Done." is shown.

Example 2: (Verbose mode)

In this case the information area shows a lot of information delivered by c2x-cmd.

3.5.3 Button : Execute

The button Execute starts c2x-cmd with the settings made before. Right before the execution process c2x-wizard will do some consistency checks. The results of these checks are shown in

the information area under the button. If errors occur the font color switches to red (otherwise the font color is blue):

The most important error message is this one:

```
Running consistency checks:
* Error: executable [c2x-cmd.exe] could not be found in search path
```

Hint:
The message does only include the extension „.exe" if c2x-wizard runs under Windows.

This error means that c2x-wizard is not able to find an instance of the converter c2x-cmd in the current or the systems search path. If you want to know how to add c2x-cmd to the systems search path please read the document c2x-cmd manual.

Unlike as the Save Configuration function (see 3.5.1) c2x-wizard doesn't allow any execution if the consistency checks fail. In this case an error message appears. After confirming this message by clicking the OK button you are able to remove the errors.

If no errors were detected c2x-cmd converts the source file into the target file with the given options and you are able to do some further processing on the target files.

4   Appendix

4.1   References

Verweis	URL
c2x-wizard manual	http://www.goedeke.net/c2x-cmd/c2x-wizard-manual.pdf
c2x-cmd manual	http://www.goedeke.net/c2x-cmd/c2x-cmd-manual.pdf
c2x-wizard Symbols „Web Application Icons Set"	http://www.webappers.com/2008/02/12/webappers-released-free-web-application-icons-set/