

CSV to Catalog converter user guide



This document contains instructions for running application CSV to Catalog converter within the EPISODES Platform. The application is a tool for conversion of a CSV (Comma-Separated Values) file into a Catalog in EPISODES Platform Matlab-based format.

To obtain more general information about working with applications within the Platform, see [Applications Quick Start Guide](#).

CATEGORY Converters

KEYWORDS Format conversion

CITATION If you use the results or visualizations retrieved from this application in a publication, then you must cite the data source as follows: Orlecka-Sikora, B., Lasocki, S., Kocot, J. et al. (2020) An open data infrastructure for the study of anthropogenic hazards linked to georesource exploitation., Sci Data 7, 89, doi: [10.1038/s41597-020-0429-3](https://doi.org/10.1038/s41597-020-0429-3).

Requirements for input file

The input file to be converted to the EPISODES Platform Matlab-based format should be a valid CSV file, with values separated by commas (',') - see also CSV format specifications, e.g., <https://tools.ietf.org/html/rfc4180>. The file has to include a header line with names of the columns in the catalog. The numbers within the file should use English locale, the text, if containing special characters (including comma, used for values separation) should be enclosed in double quotes.

Example input file

The excerpt below contains a sample content of a valid input CSV file. It defines 8 catalog columns with two entries. Each entry contains a value for each column, if the value is not specified, it should be left empty (note the double comma in the last line of *Excerpt 1*). The most popular column names and formats are described in [this guide](#), however, custom columns are also allowed. The time should be written as a single text value with any format, however, this format has to be later specified within the application form (see the [Filling form values section](#)).

```
ID,Time,ML,Mw,Depth,RMS time residual,Hypocenter quality index,Comments
Id001,2014-01-01 00:05:02.017,3,3.5,0.4,1.123,2,"mainshock"
Id002,2014-01-08 00:40:09.136,,4.3,0.5,2.234,1,"aftershock"
```

Excerpt 1. Sample content of a CSV input file. Note, the empty value (third in the last row) and text values surrounded with double quotes.

Input file specification

The application requires single file of type *Catalog in CSV* - see also [previous section](#) for the input file requirements.

Figure 1. Application input file specification

Filling form values

The application form is generated based on the specific input file - the column names from header line are displayed in the first column (see *Figure 2*). By default all columns will be present in the result catalog, however, to exclude any of them uncheck the box near to the column name (as marked with (1) in *Figure 2*). To read the file content correctly, it is required to set the column content type (marked with (2) in *Figure 2*), so that the program knows how to interpret the subsequent values. If the content type is *Date and time*, it is also required to set the time format (marked with (3) in *Figure 2*) to allow correct reading of the time fields. It is also possible to specify a different name for the column in the resulting catalog (by default the name is the same as in the header) or add a description or unit. To set a specific column type (e.g. to specify that a column is a magnitude), use the *Column type* drop-down list. Additionally, a display format might be added to specify a custom mode of display (e.g. engineering notation), with a help of a wizard (accessible with icon marked with (4) in *Figure 2* and options visible in *Figure 3*).

If the column name from the file header was defined among the [standard catalog column names](#), all of the above properties will be already filled with defaults by the system (rows from *ID* to *Depth* in *Figure 2*). For other values, they have to be specified by the user.

Include field	CSV content type	CSV content format	New name	Description	Column type	Unit	Display format
<input checked="" type="checkbox"/> ID	Text		ID	Event ID	-		3
<input checked="" type="checkbox"/> Time	Date and time	yyyy-MM-dd HH:mm:ss.SSS	Time	Event origin time	-		5
<input checked="" type="checkbox"/> ML	Real number		ML	Local magnitude	Magnitude		4
<input checked="" type="checkbox"/> Mw	Real number		Mw	Moment magnitude	Magnitude		4
<input checked="" type="checkbox"/> Depth	Real number		Depth	Hypocenter depth measured from the ground l	-	km	13
<input checked="" type="checkbox"/> RMS time residual	Real number		RMS time residual		-		
<input checked="" type="checkbox"/> Hypocenter quality index	Real number		Hypocenter quality i		-		
<input type="checkbox"/> Comments	Text		Comments	Particularities of the respective event (main sh	-		3

1 Detailed information on column data formats and types, please consult [User Guide](#).

Sort output catalog file by time ☐ 5

Figure 2. Application form with most important elements marked.

Text
Time
Integer
Real number
Real number, format: 0
Real number, format: 0.0
Real number, format: 0.00
Real number, format: 0.000
Real number, format: 00
Real number, format: 00.0000
Real number, format: 000
Real number, format: +00
Real number, format: +000
Real number, format: +0.00E00
Real number, format: +0.00E0

Figure 3. Wizard used for display format specification

It is advisable that the catalog is always sorted by time, therefore, the form also offers an option to sort the resulting catalog (marked with (5) in Figure 2).

ID column

The catalog in the [EPISODES Platform format](#) has to contain an ID column, in order to identify each event correctly (e.g. to be able to match it to registrations from a station or miniSEED files). If your CSV catalog does not contain such a column, or you do not want to include it, you can choose the ID to be generated by the system. In such case, you only specify the ID prefix, and a unique ID is generated for each event with the given prefix.

Include field	CSV content type	CSV content format	New name	Description	Column type	Unit	Display format
<input type="checkbox"/> ID	Text		ID	Event ID	-		3

Figure 4. ID prefix field when the original ID column from CSV is chosen not to be included.

Produced output

The result file will have the same name as the input, with only the extension changed to .mat, and it will be visualized within the application outputs.

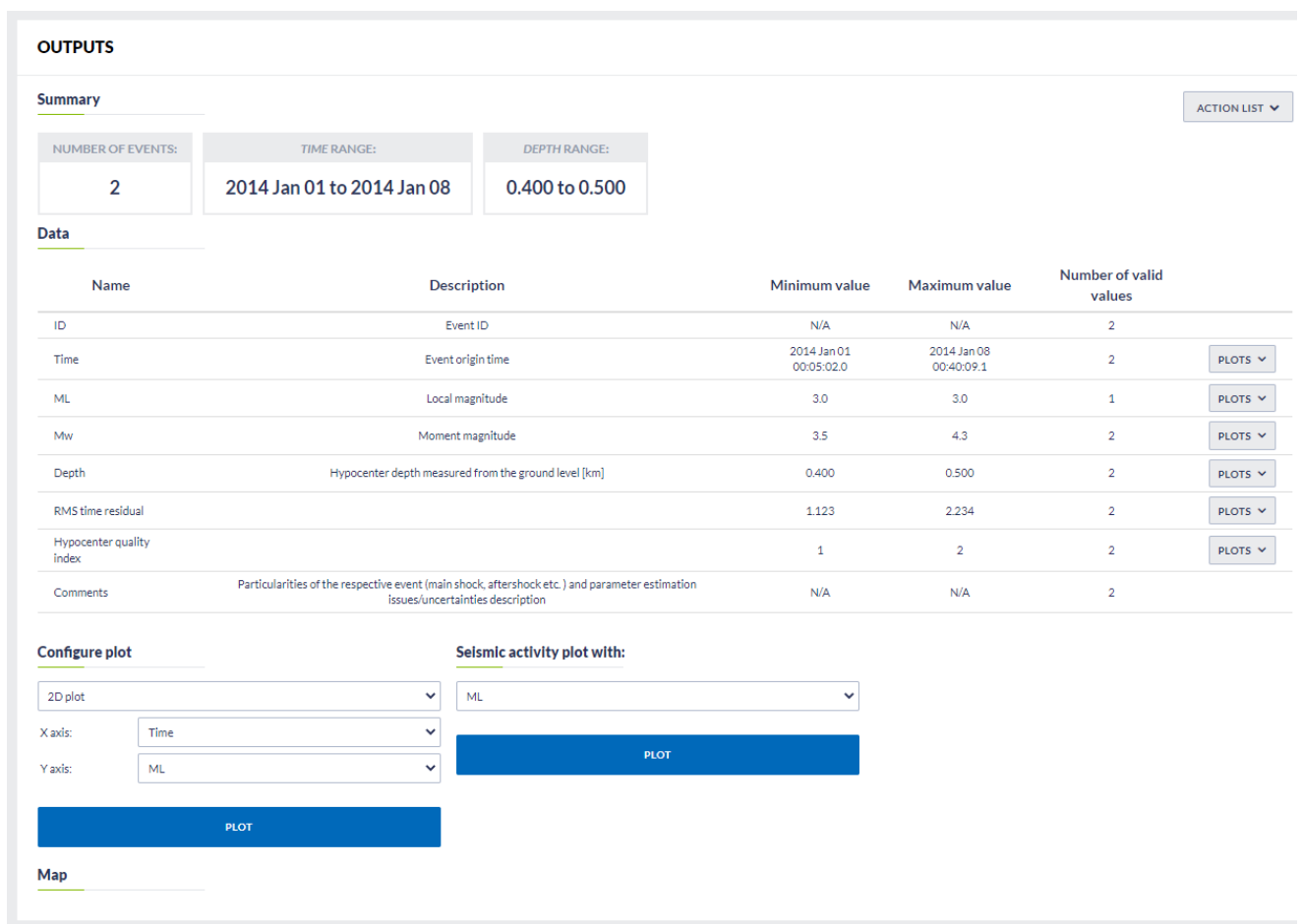


Figure 5. Result file visualization

By browsing the catalog content (*Action list* menu > *Catalog preview*) you can check that it has the same content as the initial CSV file (compare *Figure 6* with [sample input file](#)).

Catalog contents								
Search for text			Search in fields		Time	SEARCH		Switch to advanced filtering
No.	ID	Time	ML	Mw	Depth [km]	RMS time residual	Hypocenter quality index	Comments
1	Id001	2014 Jan 01 00:05:02.0	3.0	3.5	0.400	1.123	2	mainshock
2	Id002	2014 Jan 08 00:40:09.1	-	4.3	0.500	2.234	1	aftershock

Figure 6. Preview of the result catalog. The preview shows the result produced from the [sample input file](#) provided in the first section.

Troubleshooting

The most common errors that might be spotted when running the application are caused by incorrect specification of the input format - either format of the time field or incorrect type of the field. If the value in the date/time field does not match the time format from the form (field marked with (3) in *Figure 2*), the application fails with an error that it cannot parse the field, showing also the content of the field on which the parsing failed - this is illustrated in *Figure 7*. In this example the time field has value 2014-01-01 00:05:02.017 (as in [sample input file](#)), therefore, the format of the field is YYYY-MM-dd HH:mm:ss.SSS. However, the content of the format column within the application form was specified as YYY MMM dd HH:mm:ss.SSS (the month written as three letter abbreviation - in this format the date value should be 2014 Jan 01 00:05:02.017) and the system could not read the value due to that. It is also important that all the dates/times within one CSV column have the same format.

Include field	CSV content type	CSV content format	New name	Description	Column type	Unit	Display format
<input checked="" type="checkbox"/> Time	Date and time	yyyy MMM dd HH:mm:ss.SSS	Time	Event origin time	-		5

Status: Finished with error

Invalid input time format specification or CSV content to parse (2014-01-01 00:05:02.017)

Figure 7. Application error in case of incorrect time format specification.

Figure 8 illustrates another common mistake when filling the application form - incorrect type of data inside the CSV column. In this example the data in *Comments* column (see the [sample input file](#)) is text, however, in the form it was specified as a real number. As the system cannot parse the text as the number, it returns an error also informing at which value the error occurred (in *Figure 8*: "mainshock").

<input checked="" type="checkbox"/> Comments	Real number	Comments	Particularities of the respective event (main shock, aftershock et	-		3
--	-------------	----------	--	---	--	---

Status: Finished with error

Cannot parse number input (type: REAL): mainshock

Figure 8. Application error in case of incorrect content type specification.

[Back to top](#)

Related Documents

- [CSV to Catalog converter user guide](#)
- [SEED to Network Inventory converters user guide](#)
- [Parameters Catalog builder user guide](#)
- [Time Series Builder user guide](#)
- [Seed converters user guide](#)
- [GDF to Vectors converter user guide](#)
- [GDF to CSV converter user guide](#)
- [Catalog to Vectors converter user guide](#)
- [Catalog to CSV converter user guide](#)