

How to check that a Brownfield Land Register csv file is valid

Version 1
Date 11/08/2017

Contents

What is a CSV file?	1
What can be wrong with a CSV file?	2
Does it follow the rules for being a good csv file?	2
Does it contain the right information?	2
Validating a CSV file	2
Validating a Brownfield Land Register	2
Using the LGA's validator	3
Other LGA validator features	4
Using the ODI's validator	5
Document Change History	9

What is a CSV file?

'csv' stands for 'comma separated values', which is a file format, in plain text, where

- each row is separated from the next by a new line;
- each column is separated from next by a comma.

The 'csv' format provides a way to publish spreadsheet style data which can be read by a wide range of both proprietary and open source tools.

The format is a simple text file, which could be written using a text editor such as notepad. Most popular spreadsheet or database programs have a facility to export to 'csv' format, so data is typically prepared in a spreadsheet form and exported, rather than being created and edited as a text file.

The 'csv' format is not precisely defined. We follow the advice from the Open Data Institute, and use the definitions of how to write good csv at

<http://csvlint.io/about>

What can be wrong with a CSV file?

Does it follow the rules for being a good csv file?

CSV files use a character to separate each column, typically a comma, and a new line to separate each row. Some fields can have content that needs special treatment. For example, what if the data contains a comma or a new line?

The Open Data Institute have recommend how these types of situation should be represented in a CSV file, which is also usually how software that exports to the csv format handles them too.

So, when using a tool like a spreadsheet to manage data and then exporting to the CSV format, these types of potential problems are normally handled without the user having to do anything.

Does it contain the right information?

CSV files normally start with a header row which gives a name to each column.

A separate file, known as a schema, can be provided that contains the rules defining

- what the column names must be;
- what type of data can appear in each columns;
- which columns must have a value on every row;
- the values that can appear in a column;

... and so on.

Using a schema file, it is then possible to check that a CSV file contains valid information. It can't of course check that the contents are accurate; that is still for the publisher to get right.

Validating a CSV file

Schema files are themselves written in a standard form, which means that there are a number of websites that can be used to validate a csv file.

We recommend services from

- The Local Government Association - <http://opendata.esd.org.uk/>
- The Open Data Institute - <http://csvlint.io>

Validating a Brownfield Land Register

iStandUK has helped DCLG to publish guidance to Local Authorities about the Brownfield Land Register standard. In that guidance, the meaning of each column is defined along with the validation rules. So for example

Column Number	11
Column Name	Hectares
Field Type	number
Cardinality	1
Mandatory / Optional?	Mandatory
Definition	The size of the land in hectares.
Validation	A decimal number with up to 2 decimal places.
Notes	

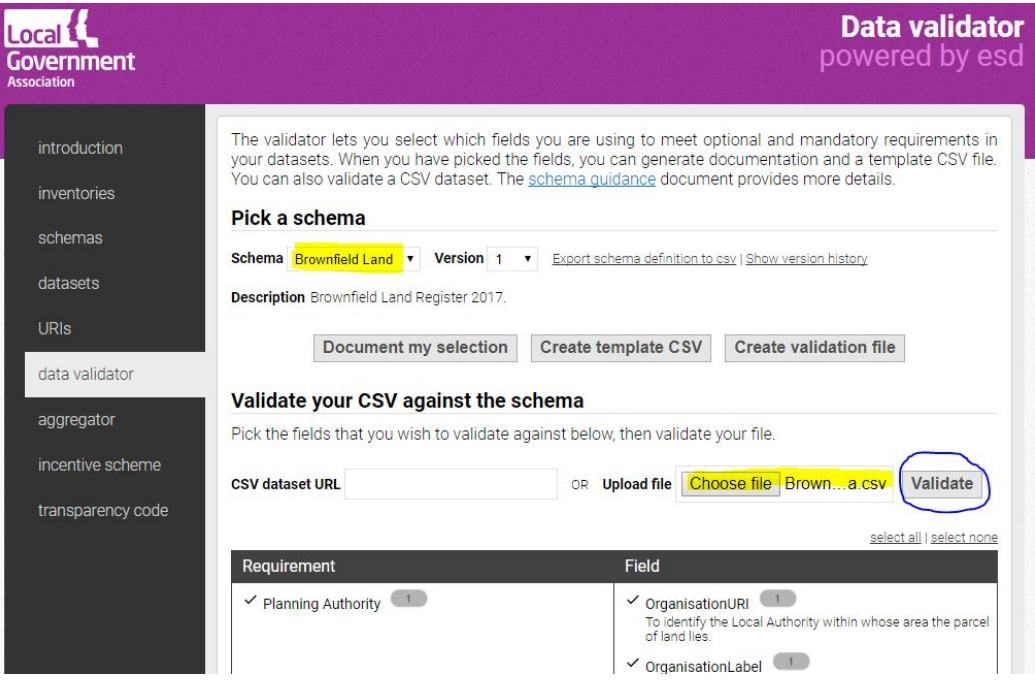
iStandUK has built these rules into schemas that can be used with the tools listed below, and other sites that use the same approach to validation.


Where a cell in a file is highlighted as an error, the tools do their best to explain what the problem is. If you can't work it out from that, you will be able to look at the DCLG guidance for that field and check the definition against the value in the file, to understand the fault.

Using the LGA's validator

The LGA's "data validator" is part of the LG inform Plus suite of open data tools which are available to use freely without signin. The validator offers a choice of schemas for common local datasets, including the Brownfield Land Register and datasets published according to the Local Government Transparency Code.

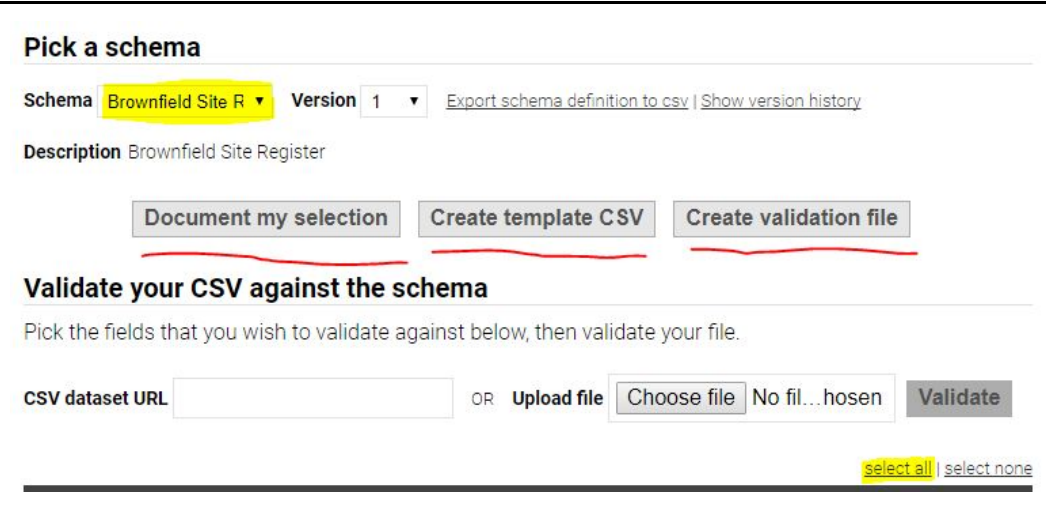
The validator will check compliance with the schema but will also allow, unchecked, any additional columns you choose to add to your csv.

<p>Visit the LGA's open data validator at http://validator.opendata.esd.org.uk and select "Brownfield Land Register</p> <p>or go straight to http://validator.opendata.esd.org.uk/brownfieldlandregister2017</p>	 <table border="1"> <thead> <tr> <th>Requirement</th> <th>Field</th> </tr> </thead> <tbody> <tr> <td>✓ Planning Authority 1</td> <td>✓ OrganisationURI 1 To identify the Local Authority within whose area the parcel of land lies.</td> </tr> <tr> <td></td> <td>✓ OrganisationLabel 1</td> </tr> </tbody> </table>	Requirement	Field	✓ Planning Authority 1	✓ OrganisationURI 1 To identify the Local Authority within whose area the parcel of land lies.		✓ OrganisationLabel 1
Requirement	Field						
✓ Planning Authority 1	✓ OrganisationURI 1 To identify the Local Authority within whose area the parcel of land lies.						
	✓ OrganisationLabel 1						
<p>Enter the URL of your csv file or upload it.</p>	<p>If your csv file is already online, you can enter its URL. Otherwise, choose the file from a drive/folder and upload it.</p>						
<p>Click the "Validate" button</p>	<p>When you press "Validate", the validator will check the contents of your csv for</p>						

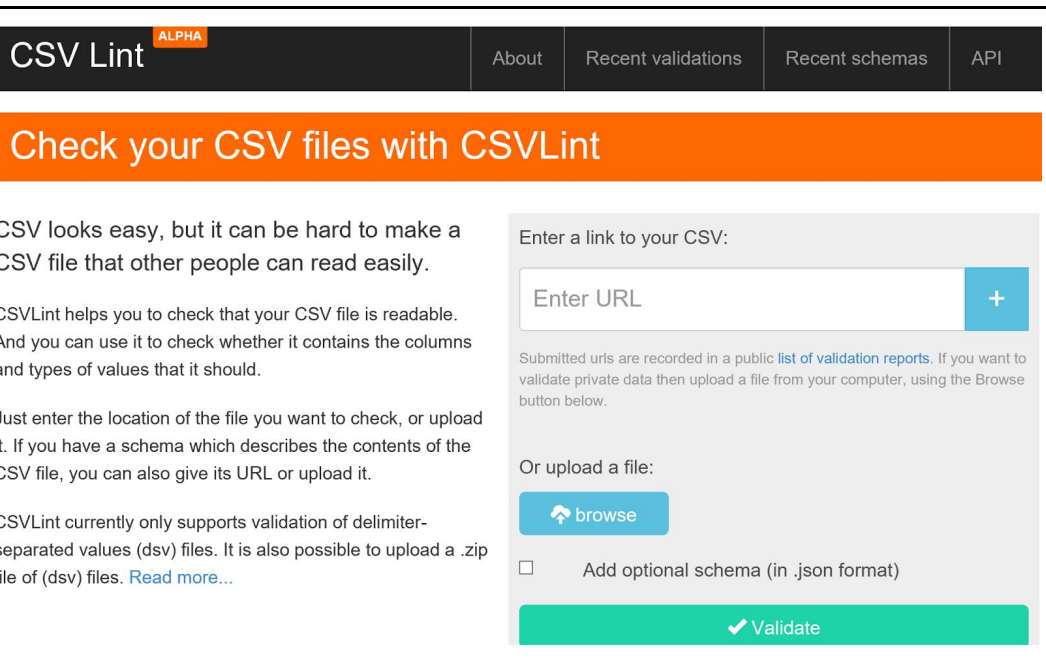
	<p>compliance with the schema and report any discrepancies. Columns with names other than those in the schema will be accepted without errors, so you can include local data columns if you wish.</p> <div> <h3>Validation errors/warnings</h3> <hr/> <p>11 unique errors OrganisationType cannot contain blank rows. GeoX cannot contain blank rows. LastUpdatedDate cannot contain blank rows. Field 'OrganisationURI' is not in proper format (1 times) Field 'SiteplanURL' is not in proper format (1 times) Field 'CoordinateReferenceSystem' is not in proper format (1 times) Field 'GeoY' is not in proper format (1 times) Field 'Hectares' is not in proper format (1 times) Field 'Deliverable' is not in proper format (1 times) Field 'FirstAddedDate' is not in proper format (7 times) Field 'LastUpdatedDate' is not in proper format (6 times)</p> <hr/> <p>2 unique warnings Please check all entries for OrganisationURI as the field includes invalid entries (1 times) e.g. http://opendatacommunities.org/id/metropolitan-district-council/bolton Please check all entries for SiteplanURL as the field includes invalid entries (1 times) e.g. www.planningpa.bolton.gov.uk/online-applications-17/search.do?action=simple&searchType=Application</p> </div> <p>The errors are explained below in the section on the ODI's validator which gives the precise row and column of each error.</p> <p>Note that the validator identifies each column from its column name, so it will accept columns in any order. DCLG however requests that you put columns in the order given by the schema.</p>
Correct for any errors and resubmit	<p>If the validator reports errors in your csv, correct them and repeat the above process until it passes with no errors.</p> <div>  Congratulations! Your file validates against the schema. </div> <p>You can contact support@esd.org.uk with any queries on using the validator.</p>




Other LGA validator features


The LGA's data validator performs other functions that you may find useful.

<p>Visit the LGA's open data validator at http://validator.opendata.eu.org.uk and select "Brownfield Land Register"</p> <p>or go straight to http://validator.opendata.eu.org.uk/brownfieldlandregister2017</p>	
<p>"Document my selection"</p>	<p>"Document my selection" will generate a new web page with each field, validation rules and a description. The page can be printed or its contents pasted into other documents.</p>
<p>"Create template CSV"</p>	<p>"Create template CSV" will download an empty csv file with just the first row of column names populated. Note that the csv file name just corresponds with the schema name. Your file name needs to be personalised for your authority as described in the schema guidance.</p>
<p>"Create validation file"</p>	<p>"Create validation file" generates a file in the json table schema format required by the Open Data Institute's CSV Lint tool to validate csv data. That tool is described in the following section of this document.</p> <p>As DCLG provides the validation file directly (see below), you should not need the LGA's "Create validation file" function.</p>

Using the ODI's validator

<p>Visit the web site at http://csvlint.io</p>	
<p>Click the 'browse' button</p>	<p>If the file has already been published to the web, then you can enter the URL to</p>

to search for the csv file on your machine, or network.	the file. However, that is unlikely in this scenario as you will want to check that your file is valid before publishing it.																
<p>When the file has been selected, it is then uploaded.</p> <p>You will see a progress bar as it loads, and the word '(completed)' when it is ready.</p> <p>Note, if the file has many rows, it may be sensible to try a short extract of the csv file first, to find any faults that may be reported on every row.</p>	<div><div>Or upload a file:</div><div><div> browse</div><div><div></div></div><div><ul style="list-style-type: none">Uploading BrownfieldExampleData.csv (completed)</div></div></div>																
<p>At this stage, you can do the first type of check; that is, does it meet the rules for a valid csv file.</p> <p>Click the Validate button</p>	<div><div> Validate</div></div>																
<p>Congratulations!</p> <p>If the file is a valid csv file, you will be congratulated.</p> <p>If there are errors, they will be listed, allowing you to fix them and retry the validation.</p>	<div><div><div>Validation Results<div>CSV valid</div><div><div>/validation/595df90933610a0004000024</div><div><div>Congratulations! Your CSV is valid!</div><div>Total Rows Processed = 7</div><div><div> Download Standardised CSV</div></div></div><div><table><tr><td></td><td>0 Errors</td><td>0 Warnings</td><td>1 Messages</td></tr><tr><td>Structure</td><td>0</td><td>0</td><td>1</td></tr><tr><td>Schema</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Context</td><td>0</td><td>0</td><td>0</td></tr></table></div><div><div>Structural problem: Assumed header</div><div>As there is no machine readable way to tell if your CSV has a header row, we have assumed that your CSV has one.</div></div></div></div></div></div>		0 Errors	0 Warnings	1 Messages	Structure	0	0	1	Schema	0	0	0	Context	0	0	0
	0 Errors	0 Warnings	1 Messages														
Structure	0	0	1														
Schema	0	0	0														
Context	0	0	0														
But ... that has not checked that the file meets the rules for the Brownfield Land Register ... so																	
Start the process again	<p>Get back to the csvlint home page.</p> <p>For example, click the CsvLint logo on the results page</p>																
Repeat the step to upload																	

the csv file.																	
Click the box 'Add Optional Schema', so that it is ticked.	<div><div><input checked="" type="checkbox"/></div><div>Add optional schema (in .json format)</div></div> <div><div>Enter optional schema URL</div><div>Or upload a file:</div><div><div> browse</div></div></div>																
We have provided a schema file that can be used to validate a Brownfield Land Register csv file.	https://raw.githubusercontent.com/iStandUK/BrownfieldLandRegister/master/schema/brownfieldlandregister2017.json																
<p>Copy the URL to the schema, and put it in the box that says 'Enter Optional Schema URL'</p> <p>!! Don't put this in the similarly looking box above it.</p> <p>Then click the Validate button, as before</p>	<div><div><input checked="" type="checkbox"/></div><div>Add optional schema (in .json format)</div></div> <div><div>ster/master/schema/brownfieldlandregister2017.json</div><div>X</div></div>																
<p>After a moment, the results are shown.</p> <p>For this example, I have deliberately put some errors in the file.</p>	<div><div>Validation Results</div><div>CSV invalid</div><div><div>/validation/598237220358d4000400001e</div><div><div>Sorry, your CSV did not pass validation. Please review the errors and warnings below:</div><div>Total Rows Processed = 1</div><div><div>Download Standardised CSV</div><div>Validate again</div></div></div><div><table><tr><td></td><td>9 Errors</td><td>3 Warnings</td><td>1 Messages</td></tr><tr><td>Structure</td><td>0</td><td>0</td><td>1</td></tr><tr><td>Schema</td><td>9</td><td>3</td><td>0</td></tr><tr><td>Context</td><td>0</td><td>0</td><td>0</td></tr></table></div></div></div>		9 Errors	3 Warnings	1 Messages	Structure	0	0	1	Schema	9	3	0	Context	0	0	0
	9 Errors	3 Warnings	1 Messages														
Structure	0	0	1														
Schema	9	3	0														
Context	0	0	0														
These errors are listed down the page.																	
Taking each of these errors in turn																	

<p>The SiteplanURL must start with http:// or https://</p>	<p>Schema problem: Value did not match pattern on row 2 in column 6</p> <div data-bbox="475 232 960 268"> <p>www.example.gov.uk/siteview.html?siteref=B123</p> </div> <p>The schema says that values in SiteplanURL must match <code>^(http https):\\\/</code>. Your value, www.example.gov.uk/siteview.html?siteref=B123, does not.</p>
<p>ETRS88 is not a valid value. Perhaps it should be ETRS89?</p>	<p>Schema problem: Value did not match pattern on row 2 in column 7</p> <div data-bbox="475 501 526 519"> <p>ETRS88</p> </div> <p>The schema says that values in CoordinateReferenceSystem must match <code>^(WGS84 OSGB36 ETRS89)\$</code>. Your value, ETRS88, does not.</p>
<p>GeoX is a mandatory column, but it is blank.</p>	<p>Schema problem: Missing value on row 2 in column 8</p> <p>The schema says that a value must be provided in the GeoX column. Make sure this column has values in all the rows.</p>
<p>GeoY has to be a decimal number, so 'abc' is invalid.</p>	<p>Schema problem: Value did not match pattern on row 2 in column 9</p> <div data-bbox="491 918 523 936"> <p>abc</p> </div> <p>The schema says that values in GeoY must match <code>^[+-]?\\d*\\.?\\d*\$</code>. Your value, abc, does not.</p>
<p>The Hectares field can have a maximum of 2 decimal places. This has 4.</p>	<p>Schema problem: Value did not match pattern on row 2 in column 10</p> <div data-bbox="481 1178 542 1200"> <p>0.4455</p> </div> <p>The schema says that values in Hectares must match <code>^\\d*(\\.\\d{1,2})?\$</code>. Your value, 0.4455, does not.</p>
<p>This field allows values of 'yes' or blank. It does not allow 'maybe'.</p>	<p>Schema problem: Value did not match pattern on row 2 in column 12</p> <div data-bbox="475 1420 529 1442"> <p>maybe</p> </div> <p>The schema says that values in Deliverable must match <code>^\$ ^\\(yes)\$</code>. Your value, maybe, does not.</p>
<p>The 29th February 2017 is not a valid date.</p>	<p>Schema problem: Invalid Type on row 2 in column 15</p> <div data-bbox="485 1684 616 1706"> <p>2017-02-29</p> </div>
<p>An error on the next row,, so back to the 1st column.</p> <p>The URI has been selected from the wrong set. URIs in this field must start with http://opendatacommunities.org/id/</p>	<p>Schema problem: Value did not match pattern on row 3 in column 1</p> <div data-bbox="469 1841 951 1863"> <p>http://nottherighturi.org/id/district-council/examplecouncil</p> </div> <p>The schema says that values in OrganisationURI must match <code>^http:\\\\opendatacommunities.org\\id\\</code>. Your value, http://nottherighturi.org/id/district-council/examplecouncil, does not.</p>

Fix the errors, and repeat the validation until all of the errors are gone.

Validation Results

CSV valid

</validation/595e029e6cd23d00040000a5>

Congratulations! Your CSV is valid!

Total Rows Processed = 7

[Download Standardised CSV](#)

[Validate again](#)

	0 Errors	0 Warnings	1 Messages
Structure	0	0	1
Schema	0	0	0
Context	0	0	0

Structural problem: **Assumed header**

As there is no machine readable way to tell if your CSV has a header row, we have assumed that your CSV has one.

Document Change History

Version	Published on	Notes
1	11/08/2017	