

QGIS Application - Bug report #8385

QGIS does not deal with shapefiles with two fields with the same name

2013-07-30 04:59 AM - Pedro Venâncio

Status: Closed	
Priority: Normal	
Assignee:	
Category: Data Provider/OGR	
Affected QGIS version: master	Regression?: No
Operating System:	Easy fix?: No
Pull Request or Patch supplied:	Resolution:
Crashes QGIS or corrupts data:	Copied to github as #: 17158

Description

To reproduce, try to delete one column from the dataset I include attached. While saving, depending on the column that you try to delete, it shows one of these errors:

Could not commit changes to layer vegetacao

Errors: SUCCESS: 1 attribute(s) deleted.

ERROR: the count of fields is incorrect after addition/removal of fields!

Could not commit changes to layer vegetacao

Errors: SUCCESS: 1 attribute(s) deleted.

ERROR: field with index xx is not the same!

Another test that can be done is, in the attribute table, start editing, open the Field Calculator, select one of the fields [CLASSE_ELE, NOME_CELUL, DESCRICAO, INFO_LINK] in "Fields and Values" and "Load all unique values" (attached image). The information does not match with the contents of the attribute table.

The problem is in the shapefile, specifically in the attribute table that has two columns with the same name (Z83_PONTOS), and QGIS master "shuffle up" with this situation. Earlier versions (1.8, 1.7.4) were able to handle the situation and show the two columns, so it seems a regression.

I attached a set of images that illustrate this situation.

Thanks!

Associated revisions

Revision 54140310 - 2013-07-30 04:52 PM - Matthias Kuhn

[ogr] Rename columns with non-unique name (Fix #8385)

History

#1 - 2013-07-30 05:03 AM - Matthias Kuhn

- Assignee deleted (Matthias Kuhn)

#2 - 2013-07-30 06:53 AM - Matthias Kuhn

QgsFields does not accept more than one field with the same name.

- If this is changed, what behavior should be expected of e.g. feat['attributeA']? To return just a random/first of the two attributes?
- How many data providers do even support this? Is this a shapefile only problem?
- Should rather be considered appending '_{nr}' to the field (e.g. 'attributeA', 'attributeA_1' etc.)

#3 - 2013-07-30 06:58 AM - Nathan Woodrow

I'm not in favour of adding any work around in the code to handle having fields with the same name. I would just say we don't support it anymore and leave it as that. In most (all) cases it doesn't make sense to have two columns with the same name and it would make working with the data hard.

#4 - 2013-07-30 07:01 AM - Matthias Kuhn

But the current behavior is bad: opening a layer and silently half-dropping a column because it doesn't meet some silent QGIS standard.

I'd say we declare this a standard and defer handling to the provider level. So the Shapefile (and CSV and whatever) provider have to take care of it and provide a wrapper level or ask the user to rename the column...

#5 - 2013-07-30 07:04 AM - Nathan Woodrow

At the provider level should be fine. I think it would only be the ogr provider because most of the database types don't support columns with the same name.

#6 - 2013-07-30 07:15 AM - Pedro Venâncio

I don't know if this involves a lot of change in the code, but this would be, in my opinion, the best solution.

- *Should rather be considered appending '_{nr}' to the field (e.g. 'attributeA', 'attributeA_1' etc.)*

If you look at the attached image, libreoffice also make this change.

#7 - 2013-07-30 07:28 AM - Pedro Venâncio

The problem is definitely of the data, but this behavior

- *But the current behavior is bad: opening a layer and silently half-dropping a column because it doesn't meet some silent QGIS standard.*

is not good. It took me some time to realize what was the source of the problem when I tried to delete a column from the dataset attached.

#8 - 2013-07-30 07:54 AM - Matthias Kuhn

- *Status changed from Open to Closed*

Fixed in changeset commit:"54140310b4878c69395895ada0b70636123b47e3".

#9 - 2013-07-30 03:09 PM - Pedro Venâncio

- File corrupted_record.png added
- Status changed from Closed to Reopened

Thank you very much for the quick fix Matthias!

However, it seems that there are still some problems.

1) In the case of trying to eliminate the first column with the name repeated, while saving I get:

Could not commit changes to layer vegetacao

Errors: SUCCESS: 1 attribute(s) deleted.

ERROR: field with index 2 is not the same!

2) Eliminating any of the other columns, generates no error but corrupts the last record of the attributes table (see the image).

3) In the Field Calculator, if I "Load all unique values" in that first column repeated, nothing is shown. In the remaining fields, now everything works as expected, except in the second column with the name repeated, where the "Load" shows the values, but on Terminal I get "ERROR 1: Unrecognised Z83_PONTOS_1 field name in ORDER BY:".

#10 - 2013-08-01 01:16 AM - Matthias Kuhn

- Category changed from Vectors to Data Provider/OGR

#11 - 2013-08-16 04:35 AM - Jürgen Fischer

- Priority changed from Severe/Regression to Normal

I suppose the fieldname should be unique. As such it's an edge case and shouldn't be considered a blocker.

#12 - 2013-08-16 04:41 AM - Pedro Venâncio

Hi Jurgen,

| *I suppose the fieldname should be unique. As such it's an edge case and shouldn't be considered a blocker.*

I agree, but the most serious problem here is that the data may be corrupted.

#13 - 2013-08-16 04:41 AM - Giovanni Manghi

Jürgen Fischer wrote:

| *I suppose the fieldname should be unique. As such it's an edge case and shouldn't be considered a blocker.*

While I agree that is an edge case I guess that the point here was that previous qgis releases handled such case in a better way. Other software also handles this case but I would not mind if qgis does not, but I guess than that such vectors should be rejected on load.

#14 - 2013-08-16 04:53 AM - Jürgen Fischer

Giovanni Manghi wrote:

While I agree that is an edge case I guess that the point here was that previous qgis releases handled such case in a better way. Other software also handles this case but I would not mind if qgis does not, but I guess than that such vectors should be rejected on load.

Does that mean you still consider this an very important issue? .oO(apparently I'm still not convinced that any regression how minor and "edgecasey" it might be, should be considered a blocker)

#15 - 2013-08-16 04:54 AM - Giovanni Manghi

Jürgen Fischer wrote:

Giovanni Manghi wrote:

While I agree that is an edge case I guess that the point here was that previous qgis releases handled such case in a better way. Other software also handles this case but I would not mind if qgis does not, but I guess than that such vectors should be rejected on load.

Does that mean you still consider this an very important issue?

not for me

#16 - 2013-08-16 05:18 AM - Giovanni Manghi

Jürgen Fischer wrote:

apparently I'm still not convinced that any regression how minor and "edgecasey" it might be, should be considered a blocker

probably you are right. My point of view is that if/when possible (time, money, etc) then all regressions should be squashed, if not possible then minor/edge cases can be left open and a "know regressions" list added in the release notes.

#17 - 2013-10-01 07:12 AM - Matthias Kuhn

- Status changed from Reopened to Closed

Fixed in changeset commit:"54140310b4878c69395895ada0b70636123b47e3".

Files

attribute_table_calc.png	33.7 KB	2013-07-30	Pedro Venâncio
dataset.zip	958 Bytes	2013-07-30	Pedro Venâncio
field_calculator_master.png	144 KB	2013-07-30	Pedro Venâncio
field_calculator_1.7.4.png	46.6 KB	2013-07-30	Pedro Venâncio
fields_attribute_table_master.png	137 KB	2013-07-30	Pedro Venâncio
fields_attribute_table_1.7.4.png	59.1 KB	2013-07-30	Pedro Venâncio
corrupted_record.png	23.6 KB	2013-07-30	Pedro Venâncio