

## StocksAndFisheriesKB - Bug #19017

### Error occurred while calculating the union of bounding boxes

Apr 06, 2020 07:56 AM - Yannis Marketakis

<b>Status:</b>	In Progress	<b>Start date:</b>	Apr 06, 2020
<b>Priority:</b>	High	<b>Due date:</b>	
<b>Assignee:</b>	Emmanuel Blondel	<b>% Done:</b>	10%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	GRSF		
<b>Description</b>			
An error occurred while calculating the union of the bounding boxes of the following records:			
<ul style="list-style-type: none"><li><a href="http://data.d4science.org/ctlg/GRSF_Admin/65aca22f-9e75-3e37-809d-8f9852e12d9d">http://data.d4science.org/ctlg/GRSF_Admin/65aca22f-9e75-3e37-809d-8f9852e12d9d</a></li><li><a href="http://data.d4science.org/ctlg/GRSF_Admin/cb1b972a-3acc-395e-911c-7db6f662b666">http://data.d4science.org/ctlg/GRSF_Admin/cb1b972a-3acc-395e-911c-7db6f662b666</a></li><li><a href="http://data.d4science.org/ctlg/GRSF_Admin/03a34f39-f4f4-3a0a-a00d-6abe399f1c21">http://data.d4science.org/ctlg/GRSF_Admin/03a34f39-f4f4-3a0a-a00d-6abe399f1c21</a></li></ul>			
The geometries are:			
<ul style="list-style-type: none"><li>{'type':'Polygon','coordinates':<a href="#">[-12.1,60.5],[-12.1,85.1],[69.1,85.1],[69.1,60.5],[-12.1,60.5]</a>}</li><li>{'type':'Polygon','coordinates':<a href="#">[-11.61,9959],[-11.89,99],[68.5,89.99],[68.5,61.9959],[-11.61,9959]</a>}</li><li>{'type':'Polygon','coordinates':<a href="#">[4.733,62.4116],[4.733,78.6565],[45.2158,78.6565],[45.2158,62.4116],[4.733,62.4116]</a>}</li></ul>			
The result reports that the geometries are invalid. Find attached the detailed log.			
<b>Related issues:</b>			
Related to StocksAndFisheriesKB - Task #18313: Approve-Merge-Annotate GRSF re...			<b>Closed</b> <b>Dec 17, 2019</b>

### History

#### #1 - Apr 06, 2020 07:57 AM - Yannis Marketakis

- Related to Task #18313: Approve-Merge-Annotate GRSF records (RAM vs FishSource) added

#### #2 - Apr 06, 2020 08:55 AM - Emmanuel Blondel

Not sure which geoservice you used for that, but if you used the simple one (with geometry strings as inputs), it doesn't expect a geojson string but a simple string "minx,miny,maxx,maxy" eg -180,-90,180,90

If you can send me the inputs you used for the algorithm i'll start from there.

Thanks

#### #3 - Apr 06, 2020 09:10 AM - Yannis Marketakis

We are using the following service:

- SERVICE\_ENDPOINT="<http://dataminer-prototypes.d4science.org/wps/WebProcessingService>";
- SERVICE\_WPS="WPS";
- SERVICE\_UNION\_IDENTIFIER="org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.GEOGRSF\_BBOX UNION";

The particular call that fails is the following:

[http://dataminer-prototypes.d4science.org/wps/WebProcessingService?request=Execute&service=WPS&Version=1.0.0&gcube-token=\\*\\*\\*\\*&lang=en-US&Identifier=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.GEOGRSF\\_BBOXUNION&DataInputs=geom1%3D%7B%27type%27%3A%27Polygon%27%2C%27coordinates%27%3A%5B%5B%5B-11%2C61.9959%5D%2C%5B-11%2C89.99%5D%2C%5B68.5%2C89.99%5D%2C%5B68.5%2C61.9959%5D%2C%5B-12.1%2C60.5%5D%2C%5B-12.1%2C85.1%5D%2C%5B69.1%2C85.1%5D%2C%5B69.1%2C60.5%5D%2C%5B-12.1%2C60.5%5D%5D%5D%7D%3Bgeom2%3D%7B%27type%27%3A%27Polygon%27%2C%27coordinates%27%3A%5B%5B%5B-12.1%2C60.5%5D%2C%5B-12.1%2C85.1%5D%2C%5B69.1%2C85.1%5D%2C%5B69.1%2C60.5%5D%2C%5B-12.1%2C60.5%5D%5D%5D%7D](http://dataminer-prototypes.d4science.org/wps/WebProcessingService?request=Execute&service=WPS&Version=1.0.0&gcube-token=****&lang=en-US&Identifier=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.GEOGRSF_BBOXUNION&DataInputs=geom1%3D%7B%27type%27%3A%27Polygon%27%2C%27coordinates%27%3A%5B%5B%5B-11%2C61.9959%5D%2C%5B-11%2C89.99%5D%2C%5B68.5%2C89.99%5D%2C%5B68.5%2C61.9959%5D%2C%5B-12.1%2C60.5%5D%2C%5B-12.1%2C85.1%5D%2C%5B69.1%2C85.1%5D%2C%5B69.1%2C60.5%5D%2C%5B-12.1%2C60.5%5D%5D%5D%7D%3Bgeom2%3D%7B%27type%27%3A%27Polygon%27%2C%27coordinates%27%3A%5B%5B%5B-12.1%2C60.5%5D%2C%5B-12.1%2C85.1%5D%2C%5B69.1%2C85.1%5D%2C%5B69.1%2C60.5%5D%2C%5B-12.1%2C60.5%5D%5D%5D%7D)

This is the first time the above is failing (we've used it for more than 100 merges so far)

#### #4 - Apr 06, 2020 09:14 AM - Emmanuel Blondel

Ok i was looking at intersect service, sorry. Indeed this service has nothing special and should work. I look at it right now.

#### #5 - Apr 06, 2020 10:16 AM - Emmanuel Blondel

- % Done changed from 0 to 10

The log is clear: isFALSE R function is not available, which is normal because from the logs Dataminer runs in R 3.4.4 and isFALSE function is only available from R 3.5.0

I had a check of R package functions used in the algorithm and based on recent package versions, none is using isFALSE. I'm waiting for <https://support.d4science.org/issues/19018> to be solved Once I can repack the algorithm I will further inspect package versions in use and see if isFALSE is used, although I doubt of that, since in general R package developers know that using isFALSE is not recommended for backward compatibility with R < 3.5.x

In the meantime, I'd suggest to CNR team to have a look to R algorithm wrappers (SAI/Dataminer) and see if isFALSE was somehow introduced (maybe to deal with algorithm inputs check/control). If it's the case, it should be substituted by !isTRUE (safe), or R should be upgrade at least to 3.5

#### #6 - Apr 06, 2020 11:55 AM - Aureliano Gentile

Dear colleagues, let me step in as I feel I miss some knowledge here.

Let's take the case a FIRMS (sourced) stock record is asked to be merged with a FishSource stock record. FIRMS record is "dominant" therefore the species and area codes are retained in the new merged records.

If the above is clear, why do we need to compute union of bounding boxes?

(I am surely missing something for which I say thank you in advance for your clarifications.)

#### #7 - Apr 06, 2020 12:04 PM - Yannis Marketakis

Your comment is correct, However I see that is part of the agreed merging workflow (documented in [https://support.d4science.org/projects/stocksandfisherieskb/wiki/18-11-16-GRSF\\_validation#Union-of-Polygons-bboxes-in-case-of-merge-records](https://support.d4science.org/projects/stocksandfisherieskb/wiki/18-11-16-GRSF_validation#Union-of-Polygons-bboxes-in-case-of-merge-records)).

Let me know if you wish to change that and stick to the polygon of the dominant record.

#### #8 - Apr 06, 2020 12:31 PM - Emmanuel Blondel

- Priority changed from Normal to High

@marketak@ics.forth.gr i've further checked the algorithm and package versions.

The responsible of this bug (isFALSE function) is called in a dependent R package geojsonio (or depends of it). I confirm it's not an issue with Dataminer/e-infra, but likely a bad use of isFALSE in some package. Somehow the package was upgraded (logically) on the e-infra while we stucked using R 3.4.4, leading to the incompatibility.

I'm investigating further the issue source, and see if I can apply a patch for us.

---

#### Package versions

- sp: 1.4.1
- jsonlite: 1.6.1
- geojsonio: 0.9.0
- rgeos: 0.5.2

#### #9 - Apr 06, 2020 12:41 PM - Aureliano Gentile

Thanks @marketak@ics.forth.gr , and thanks for having found the relevant discussion on this matter.

I suspect that both replies (combine/not combine bboxes) will raise issues in any case.

But I think you are already applying the logic of the dominant record for building the semantic identifiers and other identity metadata in the GRSF record.

So it would make sense to apply the same logic to the geospatial coordinates, i.e. inherit only those from the dominant record and ignore the others.

As we do not know today if for any case it is better to join, do you think is viable to keep both and for the time being return to the record only the dominant one and not the combination?

Do you see any issue or overwhelming complexity?

#### #10 - Apr 06, 2020 12:54 PM - Emmanuel Blondel

- Status changed from New to In Progress

Tracking deep in R, issue was in sf package. For this we need to wait it's fixed and published on CRAN, hopefully soon, afterwhat the fix will be transparently applied by e-infra nightly R package installs.

See FYI: <https://github.com/r-spatial/sf/commit/f8f52beb82262eb74a907c158513e604d7a47328>

#### #11 - Apr 06, 2020 01:10 PM - Yannis Marketakis

Thank you both for your answers.

@aureliano.gentile@fao.org the rationale behind the union of polygons was decided before proposing the dominant record, and that's the reason for having it in place in the merging workflow. I am OK with your suggestion (about using the polygon of the dominant record (if it exists). I'll just have to do some small modifications in the workflow.

#12 - Apr 06, 2020 01:13 PM - Aureliano Gentile

With many thanks, we keep going learning in this process...

#13 - Apr 06, 2020 02:38 PM - Yannis Marketakis

So, @emmanuel.blondel1@gmail.com feel free to close this issue and postpone the updates. We'll move on using the polygon of the dominant record instead of calculating the union.

Files

1586151666533_6ad4db24-9b91-42b7-bc57-bcf5085ab95b_output_report_20200406_10663920	82 Bytes	Apr 06, 2020	Yannis Marketakis
1586151666532_RLOGdba58bf9-79a4-4499-b8bc-83416ee0a2ac.txt	2.23 KB	Apr 06, 2020	Yannis Marketakis