StocksAndFisheriesKB - Support #23175

HTTP 500 error while publishing some records

Apr 14, 2022 09:48 AM - Yannis Marketakis

Status:	Closed		**Start date:**	May 09, 2022
Priority:	Normal		**Due date:**	
Assignee:	Francesco Mangiacrapa		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
VREName:	GRSF_PRE			

Description

We've started refreshing GRSF and republishing the new contents on GRSF PRE VRE (
https://blue-cloud.d4science.org/group/grsf_pre/)
During the publishing of legacy records, most of them were published without any problems.
However for some of them (approximately 150 in number) the following HTTP 500 error is thrown

```
{"id":null,
 "knowledge_base_id":null,
 "product_url":null,
 "error":"Error while performing a POST! Request url is:https://ckan-grsf-pre.d4science.org/api/3/
action/package_create
        CkanResponse{error=Ckan error of type: Search Index Error message:Unable to add package
to search index: Solr returned an error:
        (u'Solr responded with an error (HTTP 500): [Reason: Exception writing document id 46fc8
a9d19a48d445c18da70f41d870f to the index; possible analysis error.]',)
        Other fields:{}, success=false, help=https://ckan-grsf-pre.d4science.org/api/3/action/he
lp_show?name=package_create}
        CkanClient{catalogURL=https://ckan-grsf-pre.d4science.org, ckanToken=*****MASKED_TOKEN**
*****}"
}
```

Attached you will find two indicative records (a stock and a fishery record) that fail to publish.

Subtasks:

History

#1 - Apr 26, 2022 02:43 PM - Luca Frosini

- Status changed from New to In Progress

Hi Yannis,

sorry for the late answer.
I revisited the service and the documentation see #23167. This review should solve the latest issues encountered and accomplish the new requirements such as #21995
I'm testing right now the new version in the preprod infrastructure. If everything will properly works I'll update the instance you are using.
I'll give you the green light.

#2 - Apr 26, 2022 02:49 PM - Luca Frosini

Regarding the issue I just tested the two attached records with the new version and I get 500 with both for the following reason: **"error": "refers_to cannot be null/empty"**

```
{
    "id": null,
    "knowledge_base_id": null,
    "product_url": null,
    "error": "refers_to cannot be null/empty"
}
```

If you think this is a gsrf-publisher error just tell and I'll fix the behaviuor.

#3 - Apr 26, 2022 02:51 PM - Luca Frosini

```
{"id":null,
 "knowledge_base_id":null,
 "product_url":null,
 "error":"Error while performing a POST! Request url is:https://ckan-grsf-pre.d4science.org/api/3/action/p
ackage_create
         CkanResponse{error=Ckan error of type: Search Index Error message:Unable to add package to searc
h index: Solr returned an error:
         (u'Solr responded with an error (HTTP 500): [Reason: Exception writing document id 46fc8a9d19a48
d445c18da70f41d870f to the index; possible analysis error.]',)
         Other fields:{}, success=false, help=https://ckan-grsf-pre.d4science.org/api/3/action/help_show?
name=package_create}
         CkanClient{catalogURL=https://ckan-grsf-pre.d4science.org, ckanToken=*****MASKED_TOKEN*******}"
}
```

Regarding this error, it should be a Solr issue @francesco.mangiacrapa@isti.cnr.it please check it

#4 - Apr 27, 2022 10:13 AM - Yannis Marketakis

Hi @luca.frosini@isti.cnr.it
Thanks for your reply.
I'm not sure I understand how to deal with the service reply message that you mention (refers_to cannot be null/empty). We do not use any attribute with that name (refers_to) while publishing legacy records.

Luca Frosini wrote in #note-2:

> Regarding the issue I just tested the two attached records with the new version and I get 500 with both for the following reason: **"error": "refers_to cannot be null/empty"**
>
> ```
> {
>     "id": null,
>     "knowledge_base_id": null,
>     "product_url": null,
>     "error": "refers_to cannot be null/empty"
> }
> ```
>
> If you think this is a gsrf-publisher error just tell and I'll fix the behaviuor.

#5 - Apr 27, 2022 10:17 AM - Luca Frosini

Yannis Marketakis wrote in #note-4:

> Hi @luca.frosini@isti.cnr.it
> Thanks for your reply.
> I'm not sure I understand how to deal with the service reply message that you mention (refers_to cannot be null/empty). We do not use any attribute with that name (refers_to) while publishing legacy records.
>
> Luca Frosini wrote in #note-2:
>
>> Regarding the issue I just tested the two attached records with the new version and I get 500 with both for the following reason: **"error": "refers_to cannot be null/empty"**
>>
>> ```
>> {
>>     "id": null,
>>     "knowledge_base_id": null,
>>     "product_url": null,
>>     "error": "refers_to cannot be null/empty"
>> }
>> ```
>>
>> If you think this is a gsrf-publisher error just tell and I'll fix the behaviuor.

Ok, thanks I'll investigate if the changes created a bug.

#6 - Apr 27, 2022 04:00 PM - Luca Frosini

Yannis Marketakis wrote in #note-4:

> Hi @luca.frosini@isti.cnr.it
> Thanks for your reply.
> I'm not sure I understand how to deal with the service reply message that you mention (refers_to cannot be null/empty). We do not use any attribute with that name (refers_to) while publishing legacy records.

Luca Frosini wrote in #note-2:

> Regarding the issue I just tested the two attached records with the new version and I get 500 with both for the following reason: **"error": "refers_to cannot be null/empty"**
>
> ```
> {
>     "id": null,
>     "knowledge_base_id": null,
>     "product_url": null,
>     "error": "refers_to cannot be null/empty"
> }
> ```
>
> If you think this is a gsrf-publisher error just tell and I'll fix the behaviour.

The error was related to my fault in service invocation for the tests. I used GRSF as source in URL path in place of Fishsource.

#7 - Apr 28, 2022 02:45 PM - Luca Frosini

I just replicated the issue in the preproduction infrastructure.
Trying to reduce the complexity of the polygon in "spatial" field I don't get the error.
@francesco.mangiacrapa@isti.cnr.it is going to investigate if in Solr.

@marketak@ics.forth.gr was this record already published with such a polygon or is it a new one with a more complex polygon?

#8 - Apr 28, 2022 03:16 PM - Yannis Marketakis

Hi @luca.frosini@isti.cnr.it

Thanks for investigating the issue. The provided record (technically all the records) are new and were not published before. I thought that the lengthy polygon might be an issue, however, I noticed that other records with such lengthy polygon values (i.e. https://data.d4science.org/ctlg/GRSF_Pre/cdd7c484-d005-3b2a-8078-f9d137b6bbba) were published without errors.

#9 - Apr 28, 2022 03:52 PM - Luca Frosini

Yannis Marketakis wrote in #note-8:

> Hi @luca.frosini@isti.cnr.it
>
> Thanks for investigating the issue. The provided record (technically all the records) are new and were not published before. I thought that the lengthy polygon might be an issue, however, I noticed that other records with such lengthy polygon values (i.e. https://data.d4science.org/ctlg/GRSF_Pre/cdd7c484-d005-3b2a-8078-f9d137b6bbba) were published without errors.

@marketak@ics.forth.gr thanks a lot for your support.
I compared the two records the one you linked has 404 points, instead, the attached fishery record has 6126 points.

Anyway, the real error is related to the maximum accepted length for a single field. This is the printed error in Solr.

```
bytes can be at most 32766 in length; got 231415
```

@francesco.mangiacrapa@isti.cnr.it is investigating how to overcome the issue.

#10 - Apr 28, 2022 04:47 PM - Luca Frosini

@marketak@ics.forth.gr can you provide me with the JSON of 5 legacy stock and fishery records for each source (i.e. RAM, FishSource, FIRMS) and 5 stock and fishery GRSF records?

I'll use them to test the new version of the service in the preproduction infrastructure.

#11 - Apr 28, 2022 06:06 PM - Francesco Mangiacrapa

- File SOLR_Exception_with_Large_Text.txt added

I attached the file containing the log of the SOLR exception thrown by publishing one of the records (stock/fishery) attached by @marketak@ics.forth.gr. In fact, the problem is the dimension of "(geo)spatial" (i.e. the polygon) field but more in general large "text" fields on SOLR. Unfortunately, it seems related to a bug (https://issues.apache.org/jira/browse/SOLR-8495) that affects the SOLR vesion (i.e. 4.10.X) used by our CKAN (v2.6.x).

To be sure of this, my suggestion is to try publishing of (at least one) record attached (to current ticket) in another environment (e.g. GRSF_PRE of PRODUCTION?) and check the result.

However, I'm going to open an enhancement ticket to *System Engineers* for future investigation and upgrading (if feasible) of SOLR version used by our CKAN.

#12 - Apr 28, 2022 06:25 PM - Luca Frosini

Francesco Mangiacrapa wrote in #note-11:

> To be sure of this, my suggestion is to try publishing of (at least one) record attached (to current ticket) in another environment (e.g. GRSF_PRE of PRODUCTION?) and check the result.

Just to clarify to @francesco.mangiacrapa@isti.cnr.it that the records attached by @marketak@ics.forth.gr have been published in GRSF_PRE of PRODUCTION and raises the same error I get in GRSF_PRE of PRE-PRODUCTION.

#14 - Apr 28, 2022 06:32 PM - Francesco Mangiacrapa

Luca Frosini wrote in #note-12:

> Francesco Mangiacrapa wrote in #note-11:
>
> > To be sure of this, my suggestion is to try publishing of (at least one) record attached (to current ticket) in another environment (e.g. GRSF_PRE of PRODUCTION?) and check the result.
>
> Just to clarify to @francesco.mangiacrapa@isti.cnr.it that the records attached by @marketak@ics.forth.gr have been published in GRSF_PRE of PRODUCTION and raises the same error I get in GRSF_PRE of PRE-PRODUCTION.

OK, as I suspected. Thanks @luca.frosini@isti.cnr.it

#15 - Apr 29, 2022 08:16 AM - Yannis Marketakis

- File GRSF Records.zip added

Thank you both @luca.frosini@isti.cnr.it and @francesco.mangiacrapa@isti.cnr.it for working on this.
Since the problem seems to be related with the polygon length, I suggest that I remove those polygons for now (in order to proceed with GRSF Refresh) and see if it is fixed in the next iteration (in the next GRSF refresh).

@luca.frosini@isti.cnr.it attached you will find the requested records (10 of each type and source)

Are you planning to work with the GRSF Publisher right now? Can I proceed with records publishing as suggested above?

#16 - Apr 29, 2022 09:30 AM - Luca Frosini

Sorry @marketak@ics.forth.gr I'm sick at home.

Of you can wait the next week so that se Will upgrade the service wirh new functionalities.

I ask to
@roberto.cirillo@isti.cnr.it to give high priority to this

#17 - Apr 29, 2022 10:12 AM - Yannis Marketakis

Hi @luca.frosini@isti.cnr.it . Sorry to hear that you are sick. I wish you to get well soon and a quick recovery.

No worries about the new service. Since you found out that the problem is with the polygons I can remove them for now (it is really not a big deal), so that @aureliano.gentile@fao.org can have some time to inspect the new GRSF contents.

As soon as the service is updated we can test it by recovering the polygons.

#18 - Apr 29, 2022 12:05 PM - Aureliano Gentile

Thanks to all for the kind support, please let me know when it is time to review the GRSF PRE with both new data and the updated CKan interface (labels, groups, tags etc.).

#19 - May 09, 2022 03:03 PM - Francesco Mangiacrapa

- Status changed from In Progress to Feedback

- Assignee changed from Luca Frosini to Francesco Mangiacrapa

- % Done changed from 0 to 100

Hi all,

about the issue (SOLR vs large text fields) reported in the #23175#note-11, I have just applied a patch (on the GRSF_PRE of PRODUCTION i.e. https://blue-cloud.d4science.org/group/grsf_pre/data-catalogue) that should fix the issue.
Feel free to check if GRSF records with large polygons, now, can be published without problems.

If the patch works as expected, it will also be applied to *GRSF* and *GRSF_ADMIN*

#20 - May 10, 2022 10:33 AM - Yannis Marketakis

Thanks @francesco.mangiacrapa@isti.cnr.it

I confirm that the records are published successfully i.e. https://data.d4science.org/ctlg/GRSF_Pre/b9a5ac80-a001-3c20-99ec-6ce387365823

Thanks

#21 - May 10, 2022 11:14 AM - Francesco Mangiacrapa

- Status changed from Feedback to Resolved

Files

stock-c67e427e-4cef-3b00-ace9-d3bc528757a8.json	47.6 KB	Apr 14, 2022	Yannis Marketakis
fishery-42986b8a-3296-3195-a298-5b61b19d06e5.json	236 KB	Apr 14, 2022	Yannis Marketakis
SOLR_Exception_with_Large_Text.txt	9.8 KB	Apr 28, 2022	Francesco Mangiacrapa
GRSF Records.zip	175 KB	Apr 29, 2022	Yannis Marketakis