

## StocksAndFisheriesKB - Task #16880

### Merge GRSF records (FIRMS - IOTC vs RAM)

Jun 05, 2019 04:05 PM - Aureliano Gentile

|                        |                   |                        |              |
|------------------------|-------------------|------------------------|--------------|
| <b>Status:</b>         | Closed            | <b>Start date:</b>     | Jun 05, 2019 |
| <b>Priority:</b>       | Normal            | <b>Due date:</b>       |              |
| <b>Assignee:</b>       | Yannis Marketakis | <b>% Done:</b>         | 100%         |
| <b>Category:</b>       |                   | <b>Estimated time:</b> | 0.00 hour    |
| <b>Target version:</b> | GRSF              |                        |              |

#### Description

Dear colleagues, to make progress on **GRSF records merge action** I make this ticket to request the merge of a few records we just spotted while reviewing FIRMS-IOTC records in GRSF.

Please to note that in the following list, the first items are sourced from FIRMS (and FishSource in some cases) and the second ones are from RAM. The RAM records are the one to be merged into the FIRMS records, so to keep the UUID originally assigned to FIRMS.

Once done the merges, we'll liaise with IOTC for the UUIDs validation as we completed the checks for all IOTC stocks.

#### • Skipjack tuna - Indian Ocean

Short Name: Skipjack tuna - Indian Ocean

GRSF Semantic identifier: asfis:SKJ+fao:51;fao:57

Record URL: <http://data.d4science.org/ctlg/GRSF/abcc951a-cb4f-3ec4-ad48-b4a1726c136f>

to be merged with

Short Name: Skipjack tuna Indian Ocean

GRSF Semantic identifier: asfis:SKJ+rfb:IOTC

Record URL: <http://data.d4science.org/ctlg/GRSF/f30ad382-16b7-35e2-a12b-a24d795fe093>

#### • Swordfish - Indian Ocean

Short Name: Swordfish - Indian Ocean

GRSF Semantic identifier: asfis:SWO+fao:51;fao:57

Record URL: <http://data.d4science.org/ctlg/GRSF/3419c394-9dba-33aa-afda-ddbhead0b1a2>

to be merged with

Short Name: Swordfish Indian Ocean

GRSF Semantic identifier: asfis:SWO+rfb:IOTC

Record URL: <http://data.d4science.org/ctlg/GRSF/1f21ca45-6ce8-3b5b-abd8-aaa2a4979c0a>

#### • Blue marlin - Indian Ocean

Short Name: Blue marlin - Indian Ocean

GRSF Semantic identifier: asfis:BUM+fao:51;fao:57

Record URL: <http://data.d4science.org/ctlg/GRSF/3080e139-54ad-348c-9246-2eda09c5270a>

to be merged with

Short Name: Indo-Pacific blue marlin Indian Ocean

GRSF Semantic identifier: asfis:BUM+rfb:IOTC

Record URL: <http://data.d4science.org/ctlg/GRSF/3a505ceb-10f9-3e6b-b36c-4359461b0961>

#### • Black marlin - Indian ocean

Short Name: Black Marlin - Indian Ocean

GRSF Semantic identifier: asfis:BLM+fao:51;fao:57

Record URL: <http://data.d4science.org/ctlg/GRSF/3a5cd3d5-9da0-3614-9aea-758355a047a2>

to be merged with

Short Name: Black marlin Indian Ocean  
GRSF Semantic identifier: asfis:BLM+rfb:IOTC  
Record URL: <http://data.d4science.org/ctlq/GRSF/7b09ee7c-7240-3c0a-bca3-3312876c0139>

- **Striped marlin - Indian Ocean**

Short Name: Striped Marlin - Indian Ocean  
GRSF Semantic identifier: asfis:MLS+fao:51;fao:57  
Record URL: <http://data.d4science.org/ctlq/GRSF/7ce28187-fa0f-397b-b2d2-f647a572d9c9>

to be merged with

Short Name: Striped marlin Indian Ocean  
GRSF Semantic identifier: asfis:MLS+rfb:IOTC  
Record URL: <http://data.d4science.org/ctlq/GRSF/04f1dbdf-bef9-3157-95db-6ee0eba6303c>

#### Related issues:

|  |               |                     |                     |
|--|---------------|---------------------|---------------------|
| Related to StocksAndFisheriesKB - Bug #17280: Errors in Abundance Level values   | <b>Closed</b> | <b>Jul 23, 2019</b> | <b>Jul 23, 2019</b> |
| Blocked by StocksAndFisheriesKB - Bug #16923: Cannot publish records in GRSF ... | <b>Closed</b> | <b>Jun 12, 2019</b> |                     |

#### History

##### #1 - Jun 05, 2019 04:07 PM - Aureliano Gentile

- Description updated

##### #2 - Jun 05, 2019 04:08 PM - Aureliano Gentile

- Description updated

##### #3 - Jun 06, 2019 11:58 AM - Aureliano Gentile

- Subject changed from Merge two records to Merge GRSF records (FIRMS vs RAM)

##### #4 - Jun 06, 2019 12:00 PM - Aureliano Gentile

- Description updated

##### #5 - Jun 06, 2019 12:05 PM - Aureliano Gentile

- Description updated

##### #6 - Jun 06, 2019 12:08 PM - Aureliano Gentile

- Description updated

##### #7 - Jun 06, 2019 12:10 PM - Aureliano Gentile

- Description updated

##### #8 - Jun 10, 2019 09:48 AM - Yannis Marketakis

Before moving on with the actual merging, please clarify what do you mean with UUID (I suspect you are referring to Semantic ID)?

For example for the record <http://data.d4science.org/ctlq/GRSF/abcc951a-cb4f-3ec4-ad48-b4a1726c136f>, the Semantic ID is **asfis:SKJ+fao:51;fao:57** while the UUID is **abcc951a-cb4f-3ec4-ad48-b4a1726c136f**.

The reason I am asking is because UUID are **NOT** preserved during the merging. In fact a new UUID is created for the new record to be created (i.e. the merged one)

##### #9 - Jun 10, 2019 09:50 AM - Yannis Marketakis

- Status changed from New to In Progress

##### #10 - Jun 10, 2019 05:21 PM - Aureliano Gentile

Thanks Yannis,

Yes you are right, let's try to clarify and to identify operational rules.

In case of merging two GRSF records:

1. the former GRSF UUIDS are deleted/lost

2. a new UUID is generated
3. the semantic ID for the new merged record is built upon GRSF standards and inherits the standard values of the reference source record vs. the other record which it is supposed to be merged with. Ideally, the system should offer the way to indicate which species and area, and in which coding systems. But considering that this is so far not available, we may stick to the rule of "what" is to be merged "with".

The requested merges in this ticket are all based on GRSF records sourced from FIRMS for which RAM records are to be merged with by adding the time dependent data and ignore their species and area codes. But indeed this is not the only possible case, we may encounter situations where the merge may take species from one record and area from another source record, but maybe this is less frequent case. To be verified.

Thanks in advance for your feedback on the above.

#### #11 - Jun 11, 2019 10:32 AM - Yannis Marketakis

Thanks for the confirmation Aureliano.

One more clarification. I ran the merging process for the first case shown in the description and the semantic ID that is created is the following "asfis:SKJ+fao:51;rfb:IOTC;fao:57".

This seems to be normal and compliant with the merging rules ([#12856](#)). Let me know what you think.

In addition, let me know if you wish to send you the full details of a merged record (one of them) so that you can inspect it.

#### #12 - Jun 11, 2019 02:40 PM - Aureliano Gentile

Thanks Yannis.

In the example you provided there is indeed one of the main issue of the merge process: the merge is suggested because FAO areas 51/57 are the equivalent of IOTC area of competence, and the FAO areas would be the correct way to describe that stock (as per IOTC definition <https://www.iotc.org/about-iotc/competence> and FIRMS implementation).

Therefore, the merge should produce a semantic identifier by retaining the identity data from the first record and discarding the data from the second record:

"asfis:SKJ+fao:51;fao:57"

How to do that, based on which criteria? I do not know, unless you provide an interface where it can be specified what should be retained and what discarded. This only for the GRSF identity and similar issue for the fishery records.

For the time dependent data we said simply to merge all the data provided it is a collation with any datum (e.g. each row) is properly referenced.

We can have a skype on that if you want.

#### #13 - Jun 11, 2019 04:12 PM - Yannis Marketakis

For the moment and as regards the records under merging that are described in this issue we could move on by manually editing the semantic IDs as needed.

However, we should find a more sustainable (and preferably automatic) solution for supporting the merging. It seems to me that merging requires human intervention after all (for resolving issues) like the one mentioned above. I see two potential solutions, which are described below.

## Solution 1:

Your proposal could be an option. Technically, it requires to implement: (a) the corresponding services for GRSF KB and (b) GUI enhancements in the catalog to enable user feedback.

## Solution 2

Another solution is to have association matrices that would map entities found in different systems (e.g. *rfb:IOTC equalsTo fao:51&fao:57*). Then we could have a prioritization on the standard codes to be used (i.e. in this case as regards assessment areas, *fao precedes rfb*) in order to construct the desired values

#### #14 - Jun 12, 2019 10:34 AM - Aureliano Gentile

Thanks, solution 2 is pretty logic but then other factors drive the choice. For example for iattc and iccat, in place of the equivalent fao areas the preferred standard is the rfb coding system. Therefore, I think the best would be option 1 where somehow is possible to indicate what should be merged with. Waiting for such advanced interface, would it be enough for the time being implying a rule that the first record selected for merging is the one holding the right data? So the identity elements of the second one are ignored. What do you think?

#### #15 - Jun 12, 2019 11:29 AM - Yannis Marketakis

Yes that would be an option.

I'll record this (in fact I'll create a wiki page with the workflow for the merging) and move on with the merging of these records.

#### #16 - Jun 12, 2019 11:42 AM - Yannis Marketakis

In addition:

- should we apply this rule (about the dominant record) only for the semantic ID, or we should apply this for the assessment areas as well (and as a result omit rfb coded area from the assessment areas of the merged record)?

- should we apply this rule (about the dominant record) for other fields as well?

#### #17 - Jun 12, 2019 12:06 PM - Aureliano Gentile

I think as a temporary rule, the "dominant" record is the one providing all and only the codes for the semantic identifier (stock or fishery). I am saying this to ensure a proper unique identification compatible with the competent authority (in this case IOTC), which need to be verified case by case. Hence this notion of "dominant record" sounds viable. While for the fields within the GRSF fact sheet, I understand your suggestion and indeed we could keep all the data (properly referenced) with the understanding of the above convention.

In conclusion, the dominant record is the one holding all the info for populating the semantic identifier.

If that is not the case, then a manual merge should be asked through a proper ticket. This as ad interim solution waiting for ad hoc services and interface.

#### #18 - Jun 12, 2019 03:34 PM - Yannis Marketakis

- Blocked by Bug #16923: Cannot publish records in GRSF Admin VRE (unauthorized) added

#### #19 - Jun 13, 2019 03:56 PM - Yannis Marketakis

- % Done changed from 0 to 10

I've triggered the workflow for merging the first two GRSF Stock records. The result is

<https://ckan-grsf-admin2.d4science.org/dataset/a7a7c506-dd40-33e7-bf05-5cba18181807>

Before moving on with the pending merges, @aureliano.gentile@fao.org could you please have a look at the new record (the merged one) to see if you spot any errors?

FYI the following things have happened:

- 2 GRSF Stock records have been removed from the GRSF KB and the catalog (the ones under merging)
- 1 GRSF Stock record has been created in the GRSF KB and published in the catalog (the merged one)
- 2 GRSF Stock records have been updated in the catalog (they have the merged one in their similarity list)
- 20 GRSF Fishery records have been updated in the catalog (they have the merged one in their exploited list)

The overall time for the above activities was approximately 5 minutes (particularly 4' 47")

#### #20 - Jun 20, 2019 12:03 PM - Aureliano Gentile

Hi Yannis, my apologies for the delayed answer but I was sick the last few days.

I checked <https://ckan-grsf-admin2.d4science.org/dataset/a7a7c506-dd40-33e7-bf05-5cba18181807> but it returns a 404 page, then I searched in the GRSF Admin vre by ""a7a7c506-dd40-33e7-bf05-5cba18181807" and I find 5 results, among those I guess the merged record is the following [https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807)

This being said I notice:

GRSF name Katsuwonus pelamis Indian Ocean, Western Indian Ocean, Eastern **OK**

Short Name: Skipjack tuna Indian Ocean ; Skipjack tuna - Indian Ocean **DUPLICATED, maybe we should apply the "dominant record rule" also here**

GRSF Semantic identifier: asfis:SKJ+fao:51;fao:57 **OK**

Record URL: [https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807)  
**OK, I understand it is a new URL**

I noticed some issues on the catch and biomass data time series, but not related to the merge rather to the importing process and we can handle that in other tickets.

#### #21 - Jun 21, 2019 03:42 PM - Yannis Marketakis

Hi Aureliano,

I hope you are better and have recovered.

Indeed, the URL of the record is

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807)

As regards the duplication in the short name, this has been derived from the concatenation of the records under merging. We have agreed to do it like this since the short name can be changed afterwards. However, I am perfectly fine in using the short name of the dominant record.

#### #22 - Jun 24, 2019 10:52 AM - Aureliano Gentile

thanks, the week-end contributed to the full recovery ;-)

Regarding the short name, the dominant record is the one having all the good and right information on the identity of the stock or the fishery.

Therefore, also considering from a maintenance view point, the identity data are all from the dominant and in case the reviewer can add more content afterwards.

**#23 - Jun 27, 2019 09:45 AM - Yannis Marketakis**

Thanks for the update Aureliano.

I've updated the short name of the merged record and the corresponding rule in the merging process, so that it will re-use the short name of the dominant record.

I will proceed with the merging of the other records.

**#24 - Jun 27, 2019 10:18 AM - Yannis Marketakis**

- % Done changed from 10 to 90

The following merged records have been created. @aureliano.gentile@fao.org please feel free to check them

| Record URL  | Record Name                   | Record Sem ID           |
|---|-------------------------------|-------------------------|
| <a href="https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807">https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807</a> | Skipjack tuna - Indian Ocean  | asfis:SKJ+fao:51;fao:57 |
| <a href="https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F5bfc64f2-a4f5-36f2-b7c2-84a1f37496fd">https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F5bfc64f2-a4f5-36f2-b7c2-84a1f37496fd</a> | Swordfish - Indian Ocean      | asfis:SWO+fao:51;fao:57 |
| <a href="https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F7febfe97-2952-3392-bde2-dfc48d754ff5">https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F7febfe97-2952-3392-bde2-dfc48d754ff5</a> | Blue marlin - Indian Ocean    | asfis:BUM+fao:51;fao:57 |
| <a href="https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F43c75b29-85b0-34aa-a725-f95b6d49090c">https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F43c75b29-85b0-34aa-a725-f95b6d49090c</a> | Black Marlin - Indian Ocean   | asfis:BLM+fao:51;fao:57 |
| <a href="https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F37f59a2f-8d43-3fa4-ad54-2c78318adf14">https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F37f59a2f-8d43-3fa4-ad54-2c78318adf14</a> | Striped Marlin - Indian Ocean | asfis:MLS+fao:51;fao:57 |

- point 1: Obviously, the initial GRSF records have been removed from GRSF\_Admin VRE. Since all of them were approved this means that they have been published in GRSF VRE. Just for your and mine inspection, I haven't removed them from GRSF VRE yet. As soon as we see that everything is OK, I will remove them from GRSF VRE as well.
- point 2: The merging of the above records have been carried out using the merging workflow we've implemented.

**#25 - Jun 27, 2019 10:23 AM - Yannis Marketakis**

- Status changed from In Progress to Feedback

- Assignee changed from Yannis Marketakis to Aureliano Gentile

**#26 - Jul 22, 2019 05:40 PM - Aureliano Gentile**

- File 43c75b29-85b0-34aa-a725-f95b6d49090c\_FIRMS-2015-AbundanceLevel.jpg added

Dear Yannis,

Sorry it took a while for this review.

The merges are fine and can go ahead. Please ensure the obsolete records are deleted in both environments.

However, we noticed an issue in the harvested data:

in this record **43c75b29-85b0-34aa-a725-f95b6d49090c** the abundance level value for FIRMS 2015 is wrong as per attached screenshot. It seems that the "-" is ignored sometime during the data harvest process. Please investigate as this might have occurred systematically to all records.

We are ready to submit new merges, please indicate if we should do it in a new ticket or in this one.

**#27 - Jul 22, 2019 05:51 PM - Pasquale Pagano**

- Assignee changed from Aureliano Gentile to Yannis Marketakis

**#28 - Jul 23, 2019 11:47 AM - Yannis Marketakis**

- Status changed from Feedback to In Progress

Thanks for the feedback Aureliano.

I will proceed with the removal of the corresponding records from GRSF VRE (they have already been removed from GRSF Admin VRE).

In addition, I will open a new ticket investigating the issue with the missing '-' from the timeseries of record **43c75b29-85b0-34aa-a725-f95b6d49090c**

As regards the merging of more records, please proceed as you desire (either in this ticket or in a new one).

**#29 - Jul 23, 2019 11:57 AM - Yannis Marketakis**

- Assignee changed from Yannis Marketakis to Aureliano Gentile

The corresponding 10 GRSF stock records have been removed from GRSF VRE

**#30 - Jul 23, 2019 11:58 AM - Yannis Marketakis**

- Status changed from In Progress to Feedback

**#31 - Jul 23, 2019 12:04 PM - Yannis Marketakis**

- Related to Bug #17280: Errors in Abundance Level values added

**#32 - Jul 24, 2019 12:28 PM - Aureliano Gentile**

Thanks, I have approved the above listed 5 stocks/assessment unit and I have referenced this ticket in the annotation.

Please note that I do not see these records updated in the GRSF VRE, is it still a manual process or will it be synchronized automatically sometime?

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2Fa7a7c506-dd40-33e7-bf05-5cba18181807)

Skipjack tuna - Indian Ocean

asfis:SKJ+fao:51;fao:57

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2F5bfc64f2-a4f5-36f2-b7c2-84a1f37496fd](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F5bfc64f2-a4f5-36f2-b7c2-84a1f37496fd)

Swordfish - Indian Ocean

asfis:SWO+fao:51;fao:57

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2F7febfe97-2952-3392-bde2-dfc48d754ff5](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F7febfe97-2952-3392-bde2-dfc48d754ff5)

Blue marlin - Indian Ocean

asfis:BUM+fao:51;fao:57

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2F37f59a2f-8d43-3fa4-ad54-2c78318adf14](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F37f59a2f-8d43-3fa4-ad54-2c78318adf14)

Striped Marlin - Indian Ocean

asfis:MLS+fao:51;fao:57

[https://bluebridge.d4science.org/group/grsf\\_admin/data-catalogue?path%3D%2Fdataset%2F43c75b29-85b0-34aa-a725-f95b6d49090c](https://bluebridge.d4science.org/group/grsf_admin/data-catalogue?path%3D%2Fdataset%2F43c75b29-85b0-34aa-a725-f95b6d49090c)

Black Marlin - Indian Ocean

asfis:BLM+fao:51;fao:57

**#33 - Jul 26, 2019 10:52 AM - Yannis Marketakis**

- Assignee changed from Aureliano Gentile to Yannis Marketakis

Thanks for the records Aureliano.

Indeed they are not published automatically in GRSF VRE (they should).

While investigating this, I will publish them manually.

**#34 - Aug 29, 2019 12:21 PM - Aureliano Gentile**

- Subject changed from Merge GRSF records (FIRMS vs RAM) to Merge GRSF records (FIRMS - IOTC vs RAM)

**#35 - Oct 22, 2019 05:22 PM - Aureliano Gentile**

- Status changed from Feedback to Closed

I see that these merged and approved records are now also in GRSF VRE.

**Files**

43c75b29-85b0-34aa-a725-f95b6d49090c\_FIRMS-2015-AbundanceLevel.jpg

Jul 22, 2019

Aureliano Gentile