StocksAndFisheriesKB - Task #24816

Task # 24611 (In Progress): Create the new GRSF Publisher service

Task # 25008 (Paused): Requested fields changes

Change the content of 'refers_to'

Mar 21, 2023 04:01 PM - Luca Frosini

Status:	Paused		**Start date:**	Mar 21, 2023
Priority:	Normal		**Due date:**	
Assignee:	Yannis Marketakis		**% Done:**	80%
Category:			**Estimated time:**	0.00 hour
Target version:	GRSF			

Description

Stock and Fishery records contain the fields refers_to.

As an example, please consider the following the Stock Record (Assessment Unit) from GRSF with GRSF UUID cb468fb0-fed5-370f-9bc6-08fa99c54f3c.

https://data.d4science.org/ctlg/GRSF_Admin/cb468fb0-fed5-370f-9bc6-08fa99c54f3c

Please find attached the json submitted by the KB in June 2021 (maybe it is not the latest version but this is not important to present the situation)

```
"refers_to" : [ {
    "id" : "2109694f-17d1-3206-815d-179bc52e5ec7",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/2109694f-17d1-3206-815d-179bc52e5ec7"
  }, {
    "id" : "c0d7d44f-71cb-3b28-9b65-ddf72c57d44f",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/c0d7d44f-71cb-3b28-9b65-ddf72c57d44f"
  }, {
    "id" : "1b3ce0de-1ed1-3ba8-8a80-f866b287617e",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/1b3ce0de-1ed1-3ba8-8a80-f866b287617e"
  } ]
```

Currently grsf-publisher-ws service uses refers_to to:

- Generate a Resource with as name the name of the source of the record. The source is discovered by lookup the resource pointed by the URL.

I'm wondering if we can avoid this lookup and have the name in the JSON in place of the id (which is useless).

refers_to will be something like:

```
"refers_to" : [ {
    "name" : "RAM",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/2109694f-17d1-3206-815d-179bc52e5ec7"
  }, {
    "name" : "FishSource",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/c0d7d44f-71cb-3b28-9b65-ddf72c57d44f"
  }, {
    "name" : "FIRMS",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/1b3ce0de-1ed1-3ba8-8a80-f866b287617e"
  } ]
```

Moreover, maybe the name of the source is not very useful, and we can use a more significant name as we are doing for referring_stock_record and referring_fishery_record in the traceability unit see #24610

Please note that if we agree on this, the change must be made when the new service will be available and not for the current implementation.

History

#1 - Mar 21, 2023 04:02 PM - Luca Frosini

- Description updated

#2 - Mar 21, 2023 04:07 PM - Luca Frosini

- Description updated

#3 - Mar 22, 2023 10:34 AM - Yannis Marketakis

Hi @luca.frosini@isti.cnr.it .

Your proposal is absolutely fine by me.

#4 - Mar 22, 2023 01:05 PM - Luca Frosini

As I anticipated in #24817 we can't use the source as name for the resource because it clashes with the name of the resource generated by database_sources.

My opinion is also that using the source of the record as name confuses the user.

The initial ideas for an alternative resource name have issues:

- source + id : e.g. RAM 2109694f-17d1-3206-815d-179bc52e5ec7 but I don't really like it.
- semantic identifier to the target record. We must be sure that any record has a sematinc identifier. Unfortunately I don't find the semantic identifier of the target record https://data.d4science.org/ctlg/GRSF_Admin/2109694f-17d1-3206-815d-179bc52e5ec7
- short name: unfortunately is not so short and the resource name has a maximum length of 100 characters (of course we can cut it)

So, @marketak@ics.forth.gr @aureliano.gentile@fao.org any other ideas?

#5 - Mar 22, 2023 01:05 PM - Luca Frosini

- Status changed from New to In Progress

- % Done changed from 0 to 50

#6 - Mar 22, 2023 01:57 PM - Yannis Marketakis

Overall I do not think we have a problem with the solution you initially proposed, because:

- in GRSF Admin VRE

  - GRSF records do not include their database sources, they only include their corresponding legacy records (e.g. https://data.d4science.org/ctlg/GRSF_Admin/6119509e-dc3c-33d9-b3d5-26c61171a500) using the name of the source
  - legacy records only include their database sources (e.g. https://data.d4science.org/ctlg/GRSF_Admin/1aa384c5-7ae5-39e2-b284-a119c63bf67f) using the name of the source

- in GRSF public VRE

  - GRSF records include only their database sources (e.g. https://data.d4science.org/ctlg/GRSF_Admin/6119509e-dc3c-33d9-b3d5-26c61171a500)
  - legacy records are not published in GRSF public VRE

So I do not see that there is a possibility of a clash there

#7 - Mar 22, 2023 02:02 PM - Luca Frosini

Yannis Marketakis wrote in #note-6:

> Overall I do not think we have a problem with the solution you initially proposed, because:
>
> - in GRSF Admin VRE
>
>   - GRSF records do not include their database sources, they only include their corresponding legacy records (e.g. https://data.d4science.org/ctlg/GRSF_Admin/6119509e-dc3c-33d9-b3d5-26c61171a500) using the name of the source
>   - legacy records only include their database sources (e.g. https://data.d4science.org/ctlg/GRSF_Admin/1aa384c5-7ae5-39e2-b284-a119c63bf67f) using the name of the source
>
> - in GRSF public VRE
>
>   - GRSF records include only their database sources (e.g. https://data.d4science.org/ctlg/GRSF_Admin/6119509e-dc3c-33d9-b3d5-26c61171a500)
>   - legacy records are not published in GRSF public VRE

So I do not see that there is a possibility of a clash there

Ok now everything is clear. So the actual code was not wrong as I thought.

So I come back to my initial request, which avoid to lookup the target Record, i.e. having the name in the JSON instead the id.

Sorry, for my confusion

#8 - Mar 22, 2023 04:32 PM - Luca Frosini

I just had a call with @aureliano.gentile@fao.org and for the name of the Resource is fine having the source.

So, to summarise everything I kindly ask you if it is possible have the property refers_to as in the following example:

```
"refers_to" : [ {
    "name" : "RAM",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/2109694f-17d1-3206-815d-179bc52e5ec7"
  }, {
    "name" : "FishSource",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/c0d7d44f-71cb-3b28-9b65-ddf72c57d44f"
  }, {
    "name" : "FIRMS",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/1b3ce0de-1ed1-3ba8-8a80-f866b287617e"
  } ]
```

#9 - Mar 22, 2023 04:39 PM - Luca Frosini

- % Done changed from 50 to 80

#10 - Mar 23, 2023 11:54 AM - Yannis Marketakis

I confirm that we will change the objects for the refers_to records to show the name of the corresponding source, instead of the catalogue id, as suggested.

```
"refers_to" : [ {
    "name" : "RAM",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/2109694f-17d1-3206-815d-179bc52e5ec7"
  }, {
    "name" : "FishSource",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/c0d7d44f-71cb-3b28-9b65-ddf72c57d44f"
  }, {
    "name" : "FIRMS",
    "url" : "https://data.d4science.org/ctlg/GRSF_Admin/1b3ce0de-1ed1-3ba8-8a80-f866b287617e"
  } ]
```

#11 - Mar 23, 2023 12:17 PM - Luca Frosini

- Tracker changed from Support to Task

- Subject changed from About 'refers_to' to Change the content of 'refers_to'

- Status changed from In Progress to Paused

- Target version set to GRSF

As agreed with @marketak@ics.forth.gr I'm going to pause this ticket.
I'll ask to change the produced JSON to @marketak@ics.forth.gr as soon as ready with the new service.

#12 - Apr 14, 2023 11:11 AM - Luca Frosini

@marketak@ics.forth.gr @aureliano.gentile@fao.org Which String will you use for FAO SDG 14.4.1 Questionnaire referred record?
Even if there are not yet, maybe they can be in the future right?

@aureliano.gentile@fao.org do you want to align the name of the resource with the name used for Database Source see #24817 i.e.

- "FishSource",
- "Fisheries and Resources Monitoring System (FIRMS)",
- "RAM Legacy Stock Assessment Database",
- "FAO SDG 14.4.1 Questionnaire"

or do you want to keep the short name/acronym?

#13 - Apr 14, 2023 12:26 PM - Aureliano Gentile

Sorry, I do not understand where this will be displayed in the catalogue. Can you help me please?

#14 - Apr 14, 2023 02:46 PM - Luca Frosini

- File Screenshot 2023-04-14 at 14-45-33 Records Management - GRSF_Admin.png added

The name of the Resources pointing to the referred Records.

Data and Resources

RAM		Explore
FishSource		Explore
FIRMS		Explore

#15 - Apr 14, 2023 02:47 PM - Luca Frosini

See

https://data.d4science.org/ctlg/GRSF_Admin/cb468fb0-fed5-370f-9bc6-08fa99c54f3c

#16 - Apr 14, 2023 03:32 PM - Aureliano Gentile

Thanks Luca, we spoke and agreed to adopt all names in full both in "database_sources" and in "refers_to".

#17 - Apr 26, 2023 10:19 AM - Luca Frosini

- Parent task changed from #24611 to #25008

Files

cb468fb0-fed5-370f-9bc6-08fa99c54f3c.json	33.4 KB	Mar 21, 2023	Luca Frosini
Screenshot 2023-04-14 at 14-45-33 Records Management - GRSF_Admin.png	56 KB	Apr 14, 2023	Luca Frosini