D4Science Infrastructure - Task #7899

VRE # 7892 (Released): Creation of new VRE - EcoEvo

Create a catalogue profile for EcoEvo VRE

Mar 30, 2017 10:30 AM - Leonardo Candela

Status:	Closed		**Start date:**	Mar 30, 2017
Priority:	Normal		**Due date:**	
Assignee:	Miriam Baglioni		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	Data Publishing			
Infrastructure:	Production			

Description

The profile should be created by using the fields specified at the following Google Spreadsheet

https://docs.google.com/spreadsheets/d/1pfR0omqTjmqZtVzX5UzSFNomJEnJXRHcWUGPo_8QLd4/edit?usp=sharing

---

History

#1 - Mar 30, 2017 05:04 PM - Miriam Baglioni

this is to let you know that the spreadsheet was slightly modified by adding the list of possible values for the "subject" property

#2 - Mar 31, 2017 11:42 AM - Miriam Baglioni

The spreadsheet has been modified again. We have split the properties in two different subsets, and removed one property.

#3 - Apr 05, 2017 07:06 PM - Leonardo Candela

- % Done changed from 0 to 90

A possible schema is the following

```
<Body>
 <metadataformat>

  <metadatafield>
     <fieldName>TypeOfSource</fieldName>
     <dataType>String</dataType>
     <mandatory>true</mandatory>
     <defaultValue />
     <note>The type of the resource</note>
     <vocabulary isMultiSelection="false">
        <vocabularyField>Aggregator of data repositories</vocabularyField>
        <vocabularyField>Data repository</vocabularyField>
        <vocabularyField>Data source registry</vocabularyField>
        <vocabularyField>VRE</vocabularyField>
        <vocabularyField>Literature data source with links to datasets</vocabularyField>
     </vocabulary>
     <validator />
  </metadatafield>

  <metadatafield>
     <fieldName>Identifier</fieldName>
     <dataType>String</dataType>
     <mandatory>false</mandatory>
     <defaultValue />
     <note>The unique identifiers for the data source. Expected format is <identifier value>[namespace], e.g.
10.1002/...[DOI] You can specify the namespace the identifier belongs to between square brackets. For example:
 10.1002/... [doi] if the identifier code is a DOI. You can use DOI, re3data.org, OpenDOAR, URL,</note>
     <vocabulary />
     <validator />
  </metadatafield>

  <metadatafield>
     <fieldName>FirstAppearance</fieldName>
     <dataType>Time</dataType>
     <mandatory>false</mandatory>
```

```xml
      <defaultValue />
      <note>When the datasource first appeared</note>
      <vocabulary />
      <validator />
  </metadatafield>

  <metadatafield>
      <fieldName>Organization</fieldName>
      <dataType>String</dataType>
      <mandatory>false</mandatory>
      <defaultValue />
      <note>Organizations supporting and maintaining the resource. Expected format is <organization name> [name
space:code], e.g. IFG - CNR [ISNI:000000041756390X] You can specify the code of the organization with respect
to a given namespace. For example the institute of clinical phisiology of the cnr is associated to the 0000000
41756390X ISNI number, so it could be shown as IFG - CNR [ISNI:000000041756390X]. You can choose among ISNI, M
ARC,</note>
      <vocabulary />
      <validator />
  </metadatafield>

  <metadatafield>
      <fieldName>DegreeOfDomainPertinence</fieldName>
      <mandatory>true</mandatory>
      <dataType>String</dataType>
      <defaultValue />
      <note>A measure of 'pertinence to  the domain' for the resource</note>
      <vocabulary isMultiSelection="false">
         <vocabularyField>1</vocabularyField>
      <vocabularyField>2</vocabularyField>
      <vocabularyField>3</vocabularyField>
      <vocabularyField>4</vocabularyField>
      <vocabularyField>5</vocabularyField>
      <vocabularyField>6</vocabularyField>
      <vocabularyField>7</vocabularyField>
      <vocabularyField>8</vocabularyField>
      <vocabularyField>9</vocabularyField>
      <vocabularyField>10</vocabularyField>
      </vocabulary>
      <validator />
  </metadatafield>

  <metadatafield>
      <fieldName>EcoEvo Dataset Description</fieldName>
      <mandatory>false</mandatory>
      <dataType>String</dataType>
      <defaultValue />
      <note>A description of how EcoEvo relevant datasets can be identified within the data source. E.g. for Eu
ropePMC it could be 'search articles by MeSH terms and categories in order to identify relevant datasets'</not
e>
      <vocabulary />
      <validator />
  </metadatafield>

  <metadatafield>
        <fieldName>Subject</fieldName>
        <mandatory>true</mandatory>
        <dataType>String</dataType>
        <defaultValue />
        <note>Scientific subjects relative to the scholarly objects hosted by the data source
        </note>
        <vocabulary isMultiSelection="true">
            <vocabularyField>Animal ecology</vocabularyField>
            <vocabularyField>Microbial ecology</vocabularyField>
            <vocabularyField>Aquatic ecology</vocabularyField>
<vocabularyField>Terrestrial ecology</vocabularyField>
<vocabularyField>Evolutionary ecology</vocabularyField>
<vocabularyField>Conservation biology</vocabularyField>
<vocabularyField>Behavioral ecology</vocabularyField>
<vocabularyField>Climate studies</vocabularyField>
<vocabularyField>Macroecology</vocabularyField>
<vocabularyField>Urban ecology</vocabularyField>
<vocabularyField>Social evolution</vocabularyField>
<vocabularyField>Disease ecology</vocabularyField>
<vocabularyField>Plant ecology</vocabularyField>
<vocabularyField>Food webs</vocabularyField>
```

```
<vocabularyField>Evolutionary adaptation</vocabularyField>
<vocabularyField>Population genetics</vocabularyField>
<vocabularyField>Ecosystem functions</vocabularyField>
<vocabularyField>Spatial and landscape ecology</vocabularyField>
<vocabularyField>Community ecology</vocabularyField>
<vocabularyField>Phylogenetics</vocabularyField>
<vocabularyField>Evolutionary genetics</vocabularyField>
<vocabularyField>Evolutionary ecology</vocabularyField>
<vocabularyField>Biodiversity</vocabularyField>
        </vocabulary>
        <validator />
    </metadatafield>


<metadatafield>
        <fieldName>Links</fieldName>
        <mandatory>true</mandatory>
        <dataType>Boolean</dataType>
        <defaultValue>true</defaultValue>
        <note>Availability of links to other scholarly objects
        </note>
        <vocabulary />
        <validator />
    </metadatafield>


<metadatafield>
        <fieldName>MetadataFormat</fieldName>
        <mandatory>false</mandatory>
        <dataType>String</dataType>
        <defaultValue />
        <note>Name of supported metadata formats/schema
        </note>
        <vocabulary />
        <validator />
    </metadatafield>


<metadatafield>
        <fieldName>DatasetFormat</fieldName>
        <mandatory>false</mandatory>
        <dataType>String</dataType>
        <defaultValue />
        <note>Name of supported format of the dataset
        </note>
        <vocabulary />
        <validator />
    </metadatafield>


<metadatafield>
        <fieldName>MetadataReuseLicence</fieldName>
        <mandatory>true</mandatory>
        <dataType>String</dataType>
        <defaultValue />
        <note>Metadata reuse policy, e.g. CC-BY, CC-0
        </note>
        <vocabulary />
        <validator />
    </metadatafield>


<metadatafield>
        <fieldName>DatasetReuseLicence</fieldName>
        <mandatory>true</mandatory>
        <dataType>String</dataType>
        <defaultValue />
        <note>Content reuse policy, e.g. CC-BY, CC-0
        </note>
        <vocabulary />
        <validator />
    </metadatafield>

        </metadataformat>

      </Body>
```

#4 - Apr 06, 2017 09:37 AM - Leonardo Candela

- Status changed from New to In Progress

- % Done changed from 90 to 100

The profile has been registered in the IS, as soon as the Catalogue instance is created items compliant with this profile can be registered.

#5 - Apr 06, 2017 09:38 AM - Leonardo Candela

- Status changed from In Progress to Feedback

#6 - Apr 06, 2017 10:31 AM - Miriam Baglioni

Hi,

there are three missing properties in the schema:
property with code 2: Name of the resource (mandatory) multiplicity 1, value free text
property with code 6: Data source description (mandatory) multiplicity 1, value free text
property with code 7: Language (mandatory) multiplicity 1..N, value ISO639 3-letter codes

#7 - Apr 06, 2017 02:12 PM - Leonardo Candela

The profile describes only the "additional fields" to be collected, by default every catalogue entry has a name/title and a description (plus other fields).

Re Language it has to be added.

BTW, I'll wait for having the catalogue working to check if the rest is working as expected thus to edit the profile one time only.

#8 - Apr 08, 2017 08:44 AM - Leonardo Candela

- Assignee changed from Leonardo Candela to Miriam Baglioni

@miriam.baglioni@isti.cnr.it could you please check if everything but language is OK? To do so please enter the VRE / Dataset and make a try to publish a dataset.

@costantino.perciante@isti.cnr.it is there any simple way to add a field having values the ISO 639-3 languages (the code and the name) other than specifying all the values?

#9 - Apr 08, 2017 09:03 AM - Costantino Perciante

Unfortunately there is not if you want to show the values in a listbox. The other way around would be to let the user insert the iso code which can be then validated by a possible regular expression that fits this purpose.

#10 - Apr 10, 2017 09:06 AM - Miriam Baglioni

- File Screen Shot 2017-04-10 at 09.04.42.png added

I cannot insert. It stops at the first page. I followed the "publish item" link. It seems the "Publish in" field is required, but no value is present in the combo box. Attached a screen shot

#11 - Apr 10, 2017 09:28 AM - Costantino Perciante

@miriam.baglioni@isti.cnr.it have you tried to close the browser and reopen it? Some information is stored in session. Please close the browser, reopen it and retry. Let me know if it works this way

#12 - Apr 10, 2017 10:14 AM - Miriam Baglioni

@costantino.perciante@isti.cnr.it Nothing has changed.

#13 - Apr 10, 2017 10:56 AM - Costantino Perciante

Ok, I guess we have to wait the problem related to this ticket #8042 to be fixed

#14 - Apr 11, 2017 06:26 PM - Costantino Perciante

@miriam.baglioni@isti.cnr.it now you should be able to publish. Please let us know

#15 - Apr 12, 2017 08:59 AM - Miriam Baglioni

@costantino.perciante@isti.cnr.it It is not possible to publish. Now the error is related to the "Tag" field. I inserted a tag and when I press Continue it says I should add at least one meaningful tag for the item. Attached a screenshot

A remark: Is it possible to change Dataset to Datasource both in the EcoEvo starting page and in the "Types" field in the insertion of the elements?

#16 - Apr 12, 2017 09:03 AM - Miriam Baglioni

- File Screen Shot 2017-04-12 at 08.55.47.png added

sorry I forgot the attachment

#17 - Apr 12, 2017 09:20 AM - Costantino Perciante

- File attach_tag.png added

- File attached_tag.png added

Once you have written a tag there, you need to push enter to attach it to the item (the info note on the right side tells you this).

See my screenshots

#18 - Apr 12, 2017 09:28 AM - Miriam Baglioni

Sorry, I didn't push enter. Do not consider my precious posts

#19 - Apr 12, 2017 10:26 AM - Miriam Baglioni

I have a couple of questions:
In the insertion page the Licence field is associated to the information I am inserting in the catalogue?
In the summary, after the insertion of the item in the catalogue in the interface that opens when you push Manage, Licence, visibility and searchable are for the item I am inserting in the catalog, aren't they?
What are Acquire URL and Source for?

#20 - Apr 12, 2017 12:49 PM - Leonardo Candela

Miriam Baglioni wrote:

> I have a couple of questions:
> In the insertion page the Licence field is associated to the information I am inserting in the catalogue?

This is more a semantic issue that a technology issue. In the catalogue you are inserting "metadata" about a certain "item", the licence is one of the metadata associated with the item. Every catalogue item must have one and only one licence.

> In the summary, after the insertion of the item in the catalogue in the interface that opens when you push Manage, Licence, visibility and searchable are for the item I am inserting in the catalog, aren't they?

Manage is for catalogue managers only. Licence, visibility and searchable refer to the specific item.

> What are Acquire URL and Source for?

Please, ignore them ... (we have to override the Manage facility thus to make it compliant with the "insert" phase).

#21 - Apr 12, 2017 12:54 PM - Miriam Baglioni

Some issues:

Is it possible to remove all the fields that should have a fixed value in the first page?
In the first page: "Licence" (always academic free licence), "Publish in" (always EcoEvo), "Types" (always datasource)

In the second page: is it possible to have a bigger text field for the EcoEvo dataset description element?
Is it possible to add in the information for the Subject field "Hold down the CTRL or CMD button to select multiple options."?
MetadataFormat, dataset format, MetadataReuseLicence and DatasetReuseLicence may have multiple values.
The filed language is missing.

After the insertion of the item, in the additional info table, is it possible to remove the "version" and "system:type" fields?
In the interface associated to the manage button, is it possible not to visualize the fixed fields (Licence) and the fields Acquire URL and Source?
Why is the Item URL a changeable field? Shouldn't it be fixed?

In the page of the data catalogue when you watch the item in the catalogue. Is it possible not to visualize the Organizations, Type and Licence information on the left? We always have EcoEvo, and Datasource so this information is not useful.
Furthermore, the group should be associated with the typology of datasource element in the catalogue that is the field "Type of source", and we should consider the values in the subject field as tags. Is it possible to have "Groups" named as "Datasource typology", and "Tags" as "Subjects"?

#22 - Apr 12, 2017 02:46 PM - Leonardo Candela

@miriam.baglioni@isti.cnr.it thanks for the valuable feedback, yet tacking into account the application context (I understood it is just for writing a paper) we should carefully evaluate whether the effort needed to accommodate them and the exploitation timeline you have make them really feasible.

Please, find below detailed replies.

Some issues:

Is it possible to remove all the fields that should have a fixed value in the first page?

No, it isn't without changing to source code (this is not a matter of configuring the schema).

In the first page: "Licence" (always academic free licence), "Publish in" (always EcoEvo), "Types" (always datasource)

Ditto, to support this we have to modify the code. Moreover, I'm wondering why you are selecting the academic free licence.

In the second page: is it possible to have a bigger text field for the EcoEvo dataset description element?

This is not configurable yet, maybe in future yes.

Is it possible to add in the information for the Subject field "Hold down the CTRL or CMD button to select multiple options."?

Added to the notes associated with the field.

MetadataFormat, dataset format, MetadataReuseLicence and DatasetReuseLicence may have multiple values.

With the current version of the specification language this is not supported (i.e. it is only possible to specify that a field has multiple values). However, it is possible to add any field yet these are "free fields". I hope that this is included in the next version (@francesco.mangiacrapa@isti.cnr.it @costantino.perciante@isti.cnr.it) ... these fields are "free text fields";

The filed language is missing.

In order to make this a drop-down field we have to copy all the values of the selected ISO schema ... and this is problematic. Another option is to enable the user to enter the three letters .. yet this is not user friendly. I'm wondering whether you really need the entire set of languages is needed.

After the insertion of the item, in the additional info table, is it possible to remove the "version" and "system:type" fields?

This requires changes in the code, it is not achievable in a short time frame.

In the interface associated to the manage button, is it possible not to visualize the fixed fields (Licence) and the fields Acquire URL and Source?

Not in a short time.

Why is the Item URL a changeable field? Shouldn't it be fixed?

Manage is for catalogue managers, they might have the need to fully customize the item including the URI.

In the page of the data catalogue when you watch the item in the catalogue. Is it possible not to visualize the Organizations, Type and Licence information on the left? We always have EcoEvo, and Datasource so this information is not useful.

Not immediately.

Furthermore, the group should be associated with the typology of datasource element in the catalogue that is the field "Type of source", and we should consider the values in the subject field as tags. Is it possible to have "Groups" named as "Datasource typology", and "Tags" as "Subjects"?

Not immediately.

All in all, these request require some effort and time. We have to carefully evaluate them and prioritize them with the rest of tasks we have. Dunno what is blocking, what is a wish. Moreover, dunno the real timeline of the VRE ... it was urgent and the urgency is about a possible paper.

#23 - Apr 13, 2017 02:26 PM - Miriam Baglioni

```
Moreover, dunno the real timeline of the VRE ... it was urgent and the urgency is about a possible paper.
```

The urgency was due to the fact that we wished to cite it in a paper, but the catalogue will be maintained and, hopefully, used by researchers from the Ecology and Evolution community.
Apart from the insertion issues that I pointed out, I think it really important to be able to group by typology of datasource element in the catalogue, ant to use the value in the subjects field as tags. It is also important to have the possibility to select the language from a predefined vocabulary.

#24 - Apr 19, 2017 06:35 PM - Pasquale Pagano

> Apart from the insertion issues that I pointed out, I think it really important to be able to group by typology of datasource element in the catalogue,

This is fine and it will be solved with the release of the new version (version 3).

> and to use the value in the subjects field as tags.

This is also fine and it will be solved with the release of the new version (version 3).

> It is also important to have the possibility to select the language from a predefined vocabulary.

This has still to be solved. I will discuss a solution with the team.

#25 - Apr 21, 2017 09:16 AM - Miriam Baglioni

@leonardo.candela@isti.cnr.it Could you please add another field in the schema? They would like to have a "Search interface webpage" field as free text with multiplicity 1.

Thanks

#26 - Apr 21, 2017 10:37 AM - Francesco Mangiacrapa

@miriam.baglioni@isti.cnr.it

-1 Is "Search interface webpage" field needed for each "Types" (DataRepository, AggregatorOfDataRepositories, VRE, etc.)?

-2 Do you want locate it in a specific position? Es. At the start, at the end, between two fields already existing..

-3 It seems to be a URL, Would you like to add a suggestion to show for this field also?

Moreover, We can add a regular expression to validate the input as a URL. Is it required?

#27 - Apr 21, 2017 10:43 AM - Leonardo Candela

@francesco.mangiacrapa@isti.cnr.it

> -1 Is "Search interface webpage" field needed for each "Types" (DataRepository, AggregatorOfDataRepositories, VRE, etc.)?

Yes, all the types are equal i.e. represented by the same metadata;

> -2 Do you want locate it in a specific position? Es. At the start, at the end, between two fields already existing..

I forgot about that! @miriam.baglioni@isti.cnr.it it is possible to define the ordering of the fields appearing in the second step of the data entry process;

> -3 It seems to be a URL, Would you like to add a suggestion to show for this field also?

Yes.

> Moreover, We can add a regular expression to validate the input as a URL. Is it required?

Yes, please add it ... there are other fields expected to be URL ... add the regexp to all of them.

#28 - Apr 21, 2017 02:49 PM - Miriam Baglioni

- Status changed from Feedback to Closed

thank you very much

Files

Screen Shot 2017-04-10 at 09.04.42.png	111 KB	Apr 10, 2017	Miriam Baglioni
Screen Shot 2017-04-12 at 08.55.47.png	232 KB	Apr 12, 2017	Miriam Baglioni

attach_tag.png	40.5 KB	Apr 12, 2017	Costantino Perciante
attached_tag.png	40.5 KB	Apr 12, 2017	Costantino Perciante