



Milestone M4.23

Scratchpad to DwC-A Mapping Module

Leading partner: BGBM

Compiled by: Lorna Morris

Date: September 2012

Introduction

For milestone M4.17 we implemented a Scratchpads module to export data from the Scratchpads2 database for DwC-A extension data linked to core data for a Taxon [1]. This module allowed the user to specify the Scratchpads data they wanted in their DwC-A export file via a web form. Each of the 5 DwC-A extensions in the default export module could be mapped to a Scratchpads view defined by the user. A static meta.xml file was provided in the export, which contained the default field mapping between fields in the Scratchpads view and terms in each extension. The mapping has been described in detail in ViBRANT-D4.1 document [2].

The mapping relied on the user specifying an appropriate view for the extension, with the correct field types in the correct order. Moreover some data consumers wishing to export data from Scratchpads (e.g. Encyclopedia of Life -EOL) are interested in different sub-sets of the data and simply providing a default set of extensions and fields to export is not sufficient. Therefore a flexible mapping solution is required.

Implementation

Initially we started by looking a specific use case. Together with Edward Baker and Simon Rycroft (NHM) we looked at the requirements for DwC-A export of Scratchpads data for EOL. In our previous implementation we created Scratchpads views to map to 5 DwC-A extensions ('reference', 'image', 'types and specimen', 'distribution', 'description'). EOL have an additional extension for 'agents' and require different fields in the export for 'reference' and 'image'. We created a new Scratchpads view for agent data and modified our existing views to include this data. The EOL to DwC-A mapping is described in this document [3]. The meta.xml file was modified to include the additional DwC-A terms required by EOL.

In the previous implementation the dwca.zip file (containing the text files for the Taxon and extensions, and the static meta.xml) was generated on page request from the appropriate Scratchpads URL. Together with Edward Baker and Simon Rycroft the

module was modified so that DwC-A export now runs as a cron job and the dwca.zip file is only updated if the data has changed.

To add further flexibility to our DwC-A export module we decided to modify it to dynamically create the meta.xml so that users were not restricted to using the extensions and terms that we have specified in our default and EOL export meta.xml.

The DwC-A mapping module is available from the Scratchpads git repository:

<https://git.scratchpads.eu/v/scratchpads-2.0.git>

(in the directory: sites/all/modules/custom/dwca_export_views)

User interface

Once the module is enabled in Scratchpads the user interface can be accessed by requesting this URL:

`http://<scratchpads-base-url>/admin/config/system/dwca_export_views`

The module provides an interface for the user to associate a view with a DwC-A rowType for an extension and specify the field to DwC-A term mapping within that view. It assumes that the user has sufficient knowledge of the Scratchpads user interface to create the views for the extensions and data fields they wish to export. The first form (Figure 1) in the mapping interface only presents views that are named with the prefix 'dwca_'. Thus any new views the user creates and then wants to associate with an extension must follow this naming convention. Figures 1, 2 and 3 show the user interface for the mapping.

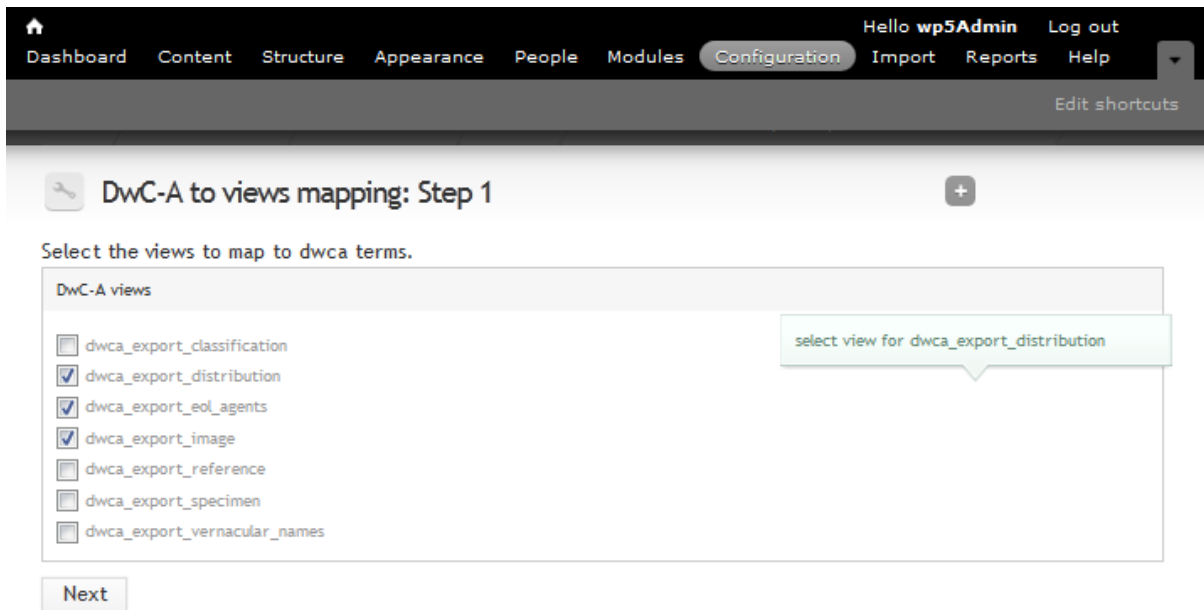


Figure 1: Screenshot of the dwca_export_views mapping module step 1. This first form shows the available DwC-A views. The user selects the view(s) to map and clicks next to progress to the next step of the mapping process.

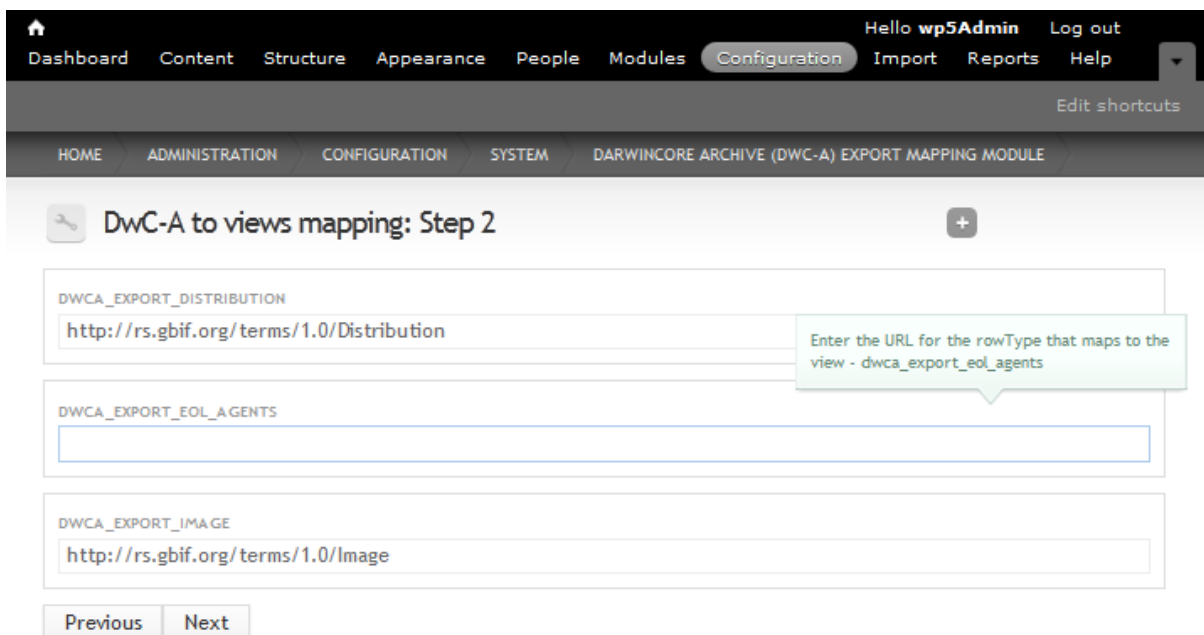


Figure 2: Screenshot of the dwca_export_views mapping module step 2. This shows the three view names selected in the previous step and the URL (rowType in the meta.xml) currently associated to the view. In the example the dwca_export_eol_agents view does not have an associated rowType. Adding or modifying a rowType to the form here associates a new rowType to the DwC-A view in the Scraphads database. The user then clicks next to progress to the final step of the mapping process.

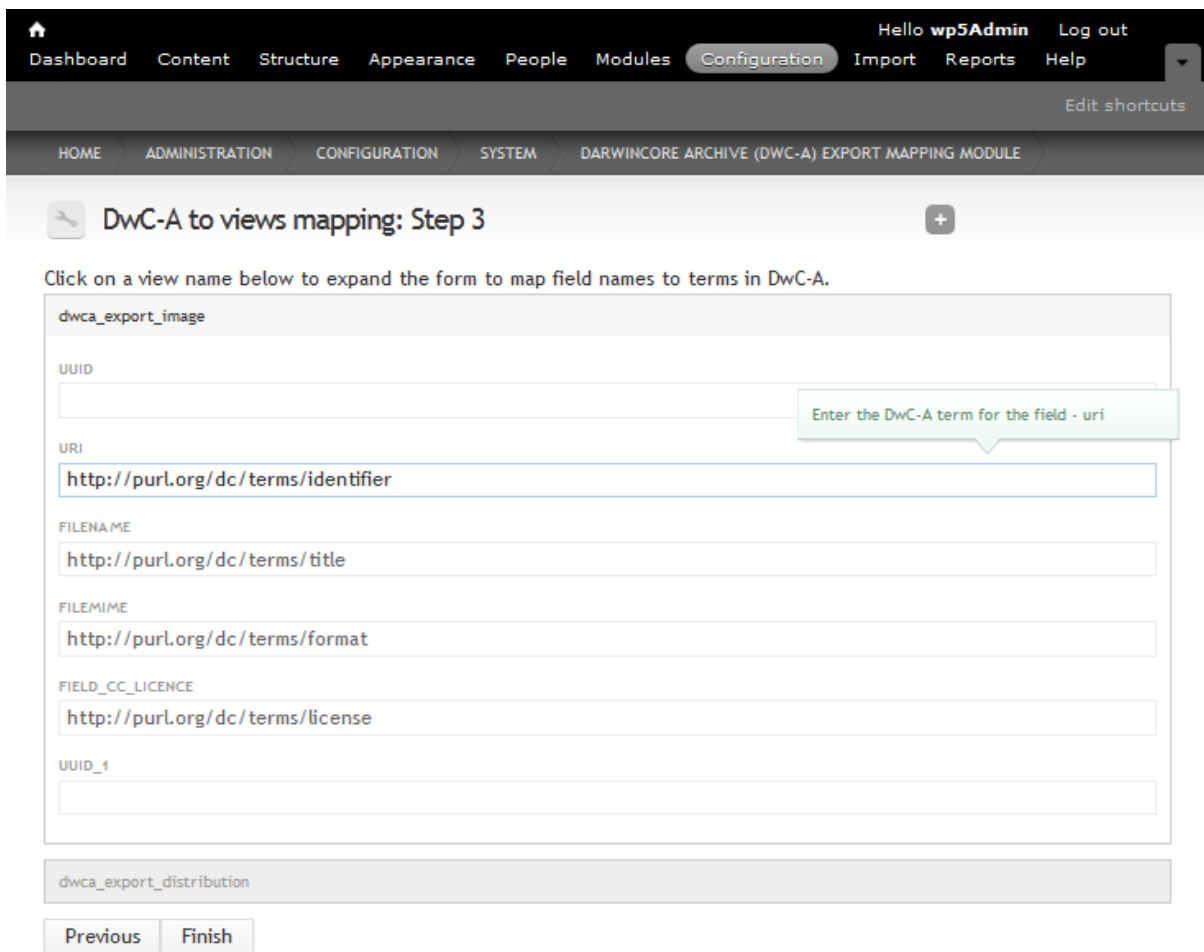


Figure 3: Screenshot of the `dwca_export_views` mapping module step 3. In this final step the user can enter terms to associate with fields in the selected view. Clicking on the view name opens the expandable form to show the individual fields. The form is pre-populated with terms previously stored in the Scratchpads database or with terms from the default `meta.xml` if this data has not been previously stored in the database. Clicking 'Finish' writes a `meta.xml` containing the specified `rowTypes` and terms to the file system and the database, so that next time the `dwca.zip` file is created (via cron) it will contain the `meta.xml` with the appropriate mappings.

Testing

We used the Scratchpads views and `meta.xml` created for our default DwC-A export module to check that the default values were set in the forms and to check the generation of the new `meta.xml` with `rowTypes` and terms entered via the forms. We also checked the Scratchpads database to ensure that the data was saved at each step. Further testing needs to be carried out to check that the generated `meta.xml` is added to

the dwca.zip file when the file is created (via cron). See also [4] for further discussion on using the generated DwC-A in the EDIT Common Data Model and the ViBRANT index database and new extensions and DwC-A terms that are required for import of the data into the CDM.

References

1. http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4%2017%20%E2%80%94%20Extensions%20for%20DwC-A%20export%20functionality.pdf
2. http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.18_%E2%80%94_Human_Interface_for_CDM-ViBRANT_index.pdf
3. http://vbrant.eu/sites/vbrant.eu/files/Darwin_Core_Archive_Export_Module_EoL.doc
4. http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.24.pdf