Using IDimager to catalog photographs of Natural History Collections with the Darwin Core XMP metadata standard.

by F. Bungartz, Charles Darwin Foundation, Galapagos Islands

Background

IDimager is currently the only software product for Digital Asset Management (DAM) that fully supports Adobe's e*X*tensible *M*etadata *P*latform (XMP). Unlike EXIF (http://www.exif.org/), the standard used by the industry to embed specific metadata in digital images (e.g., camera model, f-stop, lens type, etc.) and unlike the standard established by the International Press and Telecommunications Council (IPTC, http://www.iptc.org/cms/site/index.html;jsessionid=aWoP2PsPMip_?channel=CH0089), the XMP (http://www.adobe.com/products/xmp/) platform has been designed to allow flexibility and the option to extend the often limited metadata fields. This means that XMP can be used to design your own custom metadata fields that will embedded inside your photographs.

XMP was based on and has similar characteristics as the e*x*tensible *M*arkup *L*anguage (XML, http://www.w3.org/XML/), which was introduced as a descriptive language for digital content. A markup language is basically set of annotations to text that describe how this text is to be structured, laid out, or formatted. The basic idea is to establish a standardized way how to describe digital content that can then be interpreted intelligently by any piece of software.
In theory this option to implement a standard that allow the design of custom metadata offers unlimited flexibility. Nevertheless, the very concept of establishing a "standard", which fully supports "extensibility", is unfortunately a bit of an oxymoron. Custom fields by their very nature are "customized", that means they are not standardized and therefore they are not necessarily compatible across different software and platforms.
Therefore, although IDimager includes tools to design and develop custom XMP, it is also currently the only software that actually reads this custom XMP.

Today most DAM software fully support EXIF, IPTC (http://iptc.org/std/Iptc4xmpCore/), and most are also capable to read a core set of XMP standards. For, example, the Dublin Core (http://dublincore.org/) is a commonly used metadata standard, Adobe has established another set of fields that became widely accepted across different image editing and image cataloging software, and across different operating systems. Some software products like ExpressionMedia/iView (http://www.iview-multimedia.com/) or Adobe LightRoom (http://www.iview-multimedia.com/) also introduced their own XMP schemas. IDimager uses its own XMP schema (ics) to store tags, which allow re-construction of an IDimager database from scratch if that schema is embeded inside the digital images. Most recently The Picture Licensing Universal System (PLUS) released an XMP standard to simplify and facilitate image licensing (http://ns.useplus.org/go.ashx).

For most professional photographers these standards will be more than sufficient to address all their needs to store metadata inside their photographs. Press, wedding, and landscape photographers typically do not need additional metadata fields.

Even wildlife photographers probably do not necessarily need additional fields. They can use fields like Headline (photoshop:headline) or Title (dc:title) for catalog their photograph of a "*Siberian Tiger*" and use a field like Description (photoshop:description) to describe that this photos shows, e.g., "*A Siberian Tiger in the Russian Taiga near Novosibirsk*".

The Darwin Core (dwc) - *Metadata for biological diversity*

For scientists who take images of organisms (animals, plants, fungi, etc.), however, no adequate standard currently exists that permits to store scientific information as XMP within their photos. For example, no XMP standard supports fields that describe the scientific classification of an organism, like *species name*, *family*, *order*, *class*, *kingdom*, *domain*, etc. Scientist who work on biological diversity rely to a large extent on collecting specimens, but again, no common standard exists to store metadata data like *collection location, collector, collector's number, herbarium acronym, catalog number, etc.* inside a scientific photograph.

Biodiversity Information Standards (TDWG) is an international not-for-profit group who developed the first internationally established standard to facilitate sharing of biological diversity information – the Darwin Core (http://rs.tdwg.org/dwc/index.htm and http://rs.tdwg.org/dwc/terms/index.htm). This standard conforms with XML requirements, but until now an XMP implementation of the Darwin Core did not exist.

At the Charles Darwin Research Station (http://www.darwinfoundation.org/) in the Galapagos Islands, we have used IDimager to developed an XMP implementation for the Darwin Core (dwc) to manage photographs of our Natural History Collections. Our IDI implementation of the XMP Darwin Core has not yet been officially adapted by TDWG, but we are in contact with TDWG to suggest that they may want to officially support this implementation of the Darwin Core as XMP before a variety of different, incompatible XMP custom implementations emerge.

The Encyclopedia of Life (http://www.eol.org/) website already has developed a set of Flickr (http://www.flickr.com/groups/encyclopedia_of_life/) compatible tags which currently do not conform with XMP specifications. They are implemented as proprietary Flickr tags embedded inside the photo metadata as keywords. The tags are limited to taxonomic classification of organisms and are therefore insufficient for storing additional specimen information typically gathered for scientific specimens deposited in Natural History Collections.

The Darwin Core XMP schema that we have developed at the Charles Darwin Foundation (Fig. 1) includes all Darwin Core 149 standard fields, plus an additional 16 fields of auxiliary terms. Terms listed in the Darwin Core that are shared with the Dublin Core have not been included; these are the following fields: dcterms:type | dcterms:modified | dcterms:language | dcterms:rights | dcterms:rightsHolder | dcterms:accessRights | dcterms:bibliographicCitation

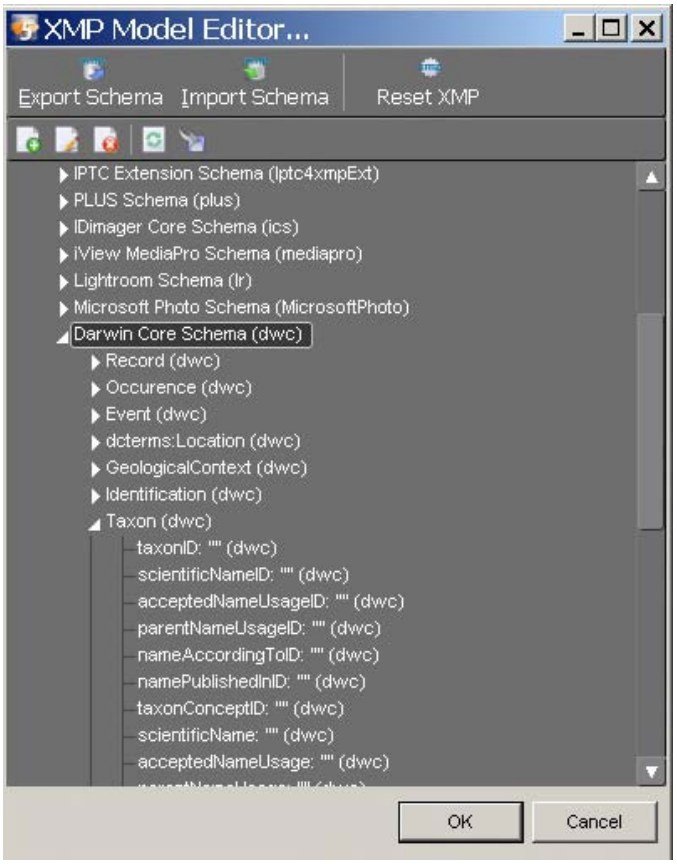

Figure 1: The Darwin Core (dwc) XMP Schema

For the implementation of this schema we followed the hierarchy of terms suggested in the Darwin Core Quick reference guide.

The schema provided here includes one deviation from the Darwin Core standard as established by TDWG: the way how dates are stored.

The recommendation by TDWG is to follow ISO 8601:2004(E). The example given at the Darwin Core Quick Reference Guide suggests that date fields should include a notation to allow both start and end date of an event to be entered (see eventDate):

Examples: "1963-03-08T14:07-0600" is 8 Mar 1963 2:07pm in the time zone six hours earlier than UTC, "2009-02-20T08:40Z" is 20 Feb 2009 8:40am UTC, "1809-02-12" is

12 Feb 1809, "1906-06" is Jun 1906, "1971" is just that year, "2007-03-01T13:00:00Z/2008-05-11T15:30:00Z" is the interval between 1 Mar 2007 1pm UTC and 11 May 2008 3:30pm UTC, "2007-11-13/15" is the interval between 13 Nov 2007 and 15 Nov 2007.

This is impractical. In the XMP implementation it would require that dates are not entered as data type “date” but as “string”. It is much safer for any user of a photo database like IDimager to enter dates into date fields than into text fields. If an event has to be entered as “2007-03-01T13:00:00Z/2008-05-11T15:30:00Z” the likelihood that errors will occur is very high. Also, a database can easily interpret and sort, query, filter date fields according to dates.

Therefore the XMP schema that we include here in this case deviates from the current standard and instead adopts the previous draft standard, the Darwin Core (see http://rs.tdwg.org/dwc/terms/history/versions/index.htm: EarliestDateCollected and LatestDateCollected).

You can download the full IDimager Darwin Core XMP schema here.

It is unlikely that all 165 XMP fields in the Darwin Core Schema are needed by the majority of users. There is, for example, some overlap with existing Adobe Photoshop XMP fields and it is seems best practice to use these fields on this already widely distributed XMP implementation, for example “photoshop:Country” instead of Darwin Core “dwc:country”.

IDimager allows highly flexible customization of Image Detail Panels that not only display custom metadata, but also permit data entry. At the Charles Darwin Foundation we are using a set of custom panel to add Darwin Core XMP metadata to catalog our photos of Natural History specimens (Fig. 2).

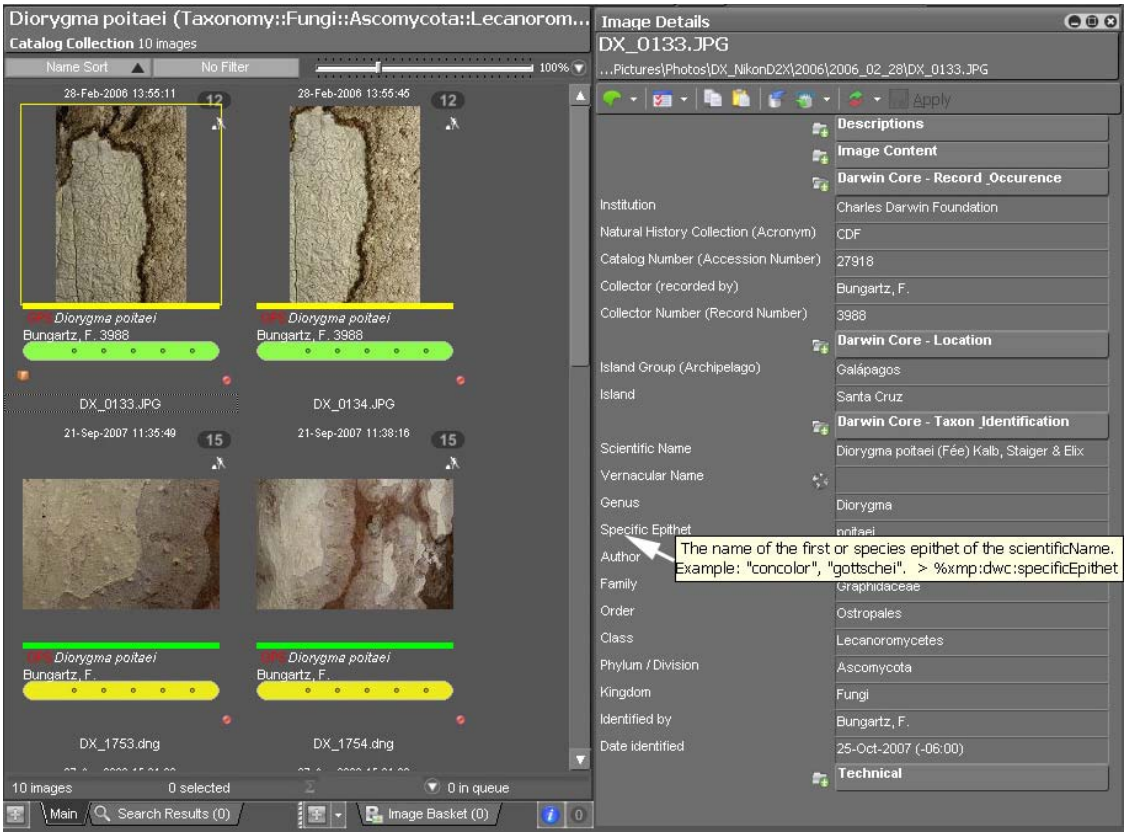

Fig. 2: The Custom Image Details Panel for the Darwin Core XMP used at the Charles Darwin Foundation.

Here you can download our custom IDimager Image Detail Panel that contains a subset of fields that are useful for the management of natural History Collections.

Data entry can be facilitated significantly using IDimager's sophisticated way to map its catalog of hierarchical keywords (labels) to any XMP field (Fig. 3). Additionally, a CSV import script facilitates synchronization of XMP metadata with any external database (you can download this script here). A CSV file exported for example from an MSAccess database that has the same field names as the Darwin Core XMP schema can be easily imported into the IDimager catalog. This way you can export the collection data for photographs of specimens that you host in your Natural History Collection and then upload this important information to your catalog and embed it as metadata into your images.

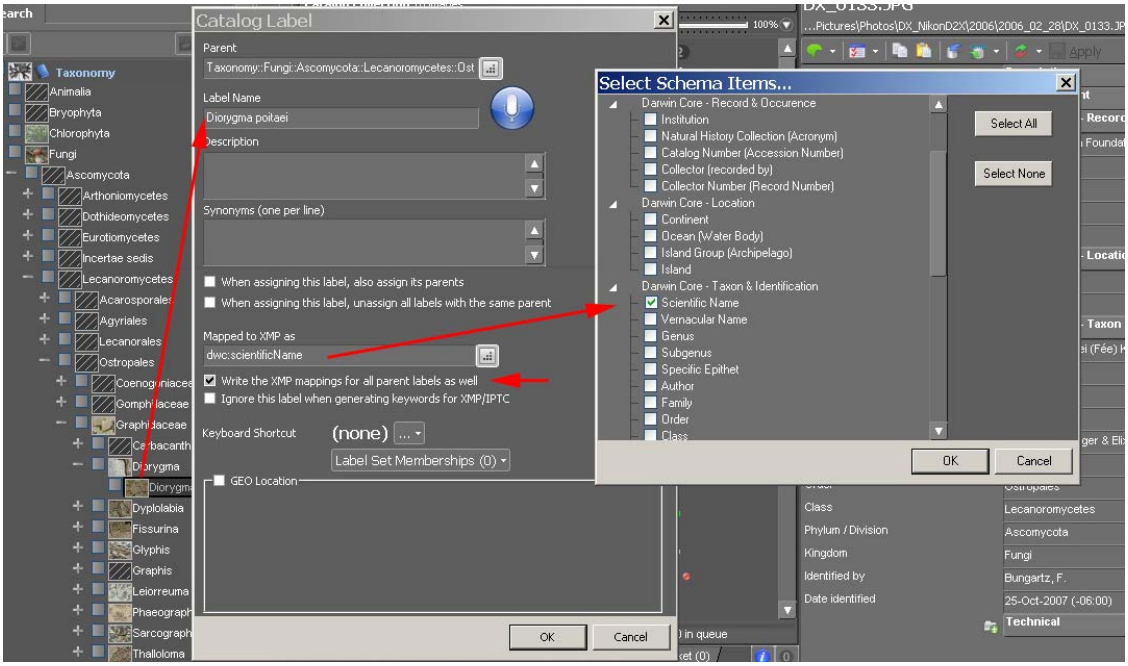

Fig. 3: IDimager allows you to map IDI hierarchical labels to the Darwin Core XMP fields.

Darwin Core XMP

The *XMP Darwin Core Schema* (dwc) schema can be downloaded from the IDI Resource Repository.

There are two sets of files: (1) the *XMP Darwin Core Schema* (dwc), and (2) several *Darwin Core Custom Image Detail Panels* for data entry in IDimager. To better organize the metadata, these custom panels are grouped according to their parent categories according to the Darwin Core Quick Reference Guide.

You can download the following Panels from the IDI Resource Repository:
Darwin Core – Record & Occurrence (with fields like Collector and Collector Number, Institution, Natural History Collection)
Darwin Core – Location (with fields like Continent, Island Group, etc.)
Darwin Core – Taxon & Identification (with fields like Kingdom, Division, Class, Order, Family, Scientific Name, Identified by, etc.)

To be available in IDimager, at least the *XMP Darwin Core Schema* (dwc) must be imported.
It is not necessary to import all Image Detail Panels. For some users who would simply like to add only the taxonomy, loading the *Darwin Core – Taxon & Identification* Detail Panel would suffice. Once you have imported the *XMP Darwin Core Schema* (dwc) you can also set up your own Image Detail Panels or customize the ones that you download.

You can add any field from the Darwin Core to these panels (for detailed instructions how to edit IDimager Custom panels visit this popular post at the IDI forum).

BEFORE designing Image Details Panels it is required to import the *XMP Darwin Core Schema* (dwc), then you can add and modify the *Darwin Core Custom Image Detail Panel*(s).

(1) Import the *XMP Darwin Core Schema* (Fig. 4a & 4b):
Highlight any thumbnail, ALT+Enter opens Image Details, go to XMP Editor – Import Schema.

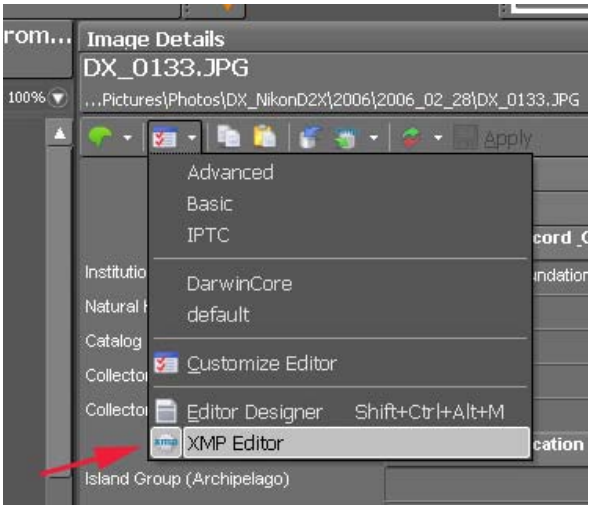

Fig. 4a: Open the XMP Editor to load the Darwin Core XMP Schema.

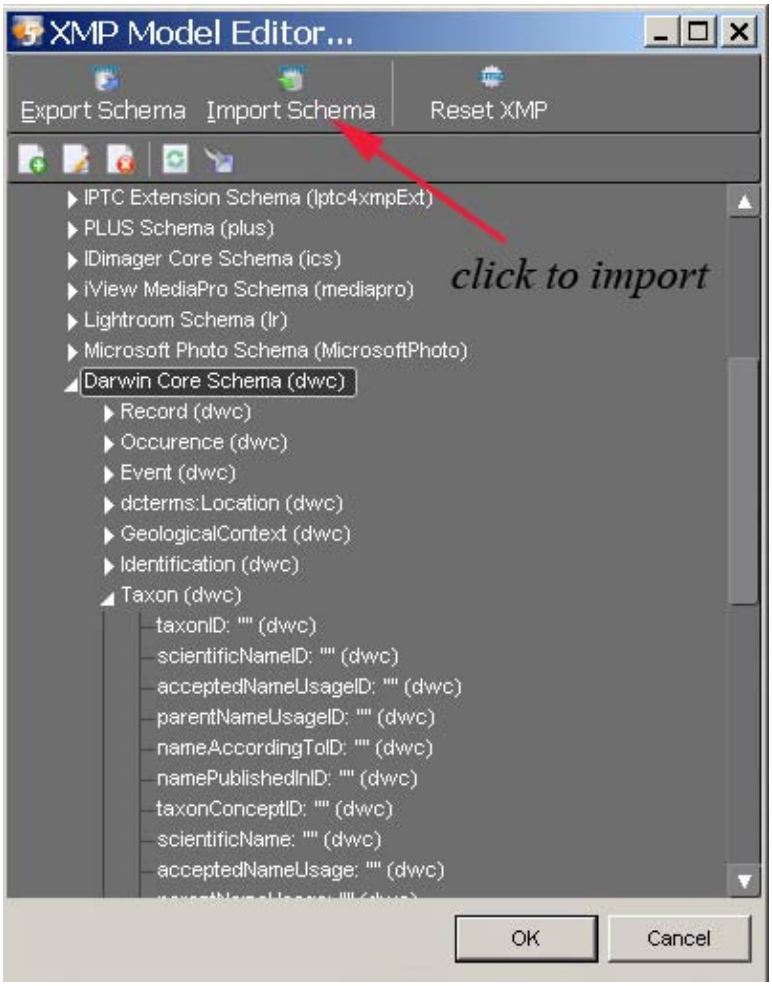

Fig. 4: Import the Darwin Core XMP Schema.

Restart IDImager

(2) Add the *Darwin Core Image Detail Panel(s)* to IDimager (Fig. 5a & b):
Highlight any thumbnail, ALT+Enter opens Image Details, go to Editor Schema – click the small arrow in the far upper right of the window and select: “manage – load” to import the *Darwin Core Image Detail Panel(s)*.

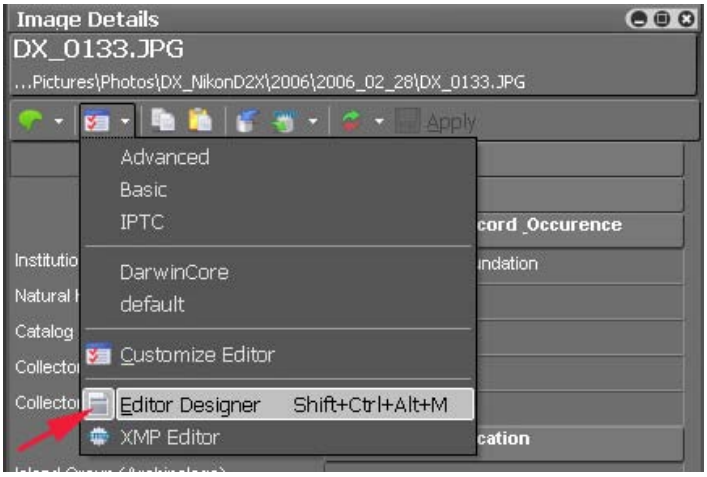

Fig. 5a: Access the Editor Schema Designer to load the Image Detail Panel.

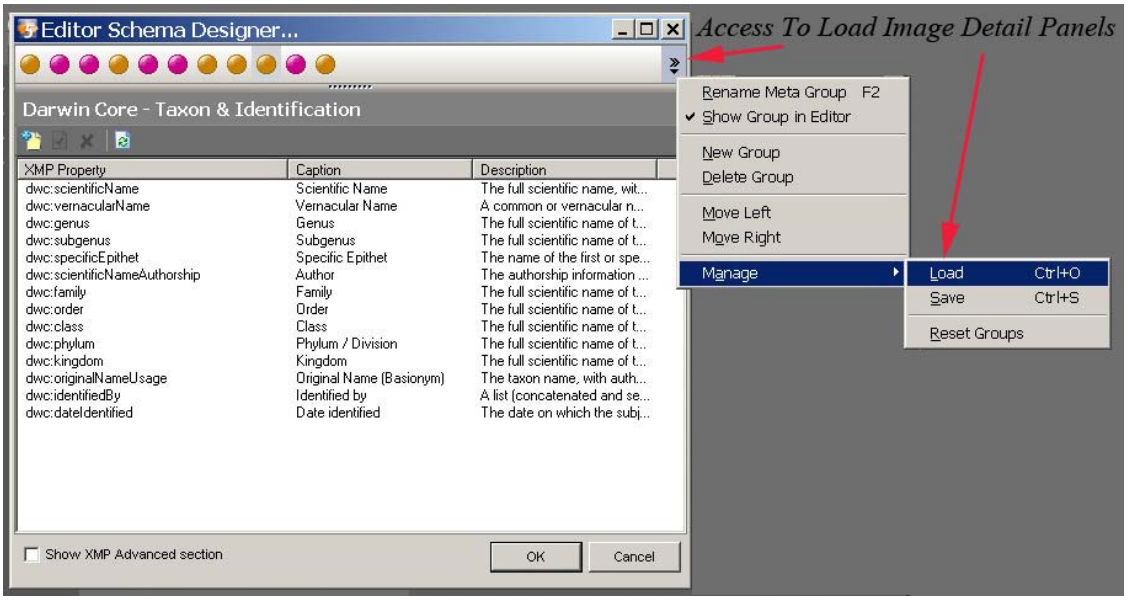

Fig. 5b: Use the Editor Schema Designer to load the Charles Darwin Foundation's Image Detail Panel into IDimager.

Restart IDimager.

Data Entry to Darwin Core XMP using IDimager

Now Image Details contain the ***Darwin Core Image Detail Panel(s)*** and you can enter content in three ways:

(1) Manually fill the Darwin Core Metadata fields using the Image Details that you just imported (watch a video here how you can easily customize the IDI Image Detail Panel).

(2) Using IDI's keyword mapping: this is a very elegant way how to make sure that your taxonomic classification is always up-to-date and reflected in IDI's catalog tree. Create a new top level category Taxonomy and create a taxonomic hierarchy of IDI Catalog Labels (Fig. 6). These keywords can then be mapped to the corresponding XMP Darwin Core fields (right click any label in your tree – edit: Fig. 2). If you check "write XMP mappings for all parents as well" assigning a lower rank (species, genus, etc.) will automatically also assign all corresponding taxonomic hierarchy to the XMP Darwin Core Fields.

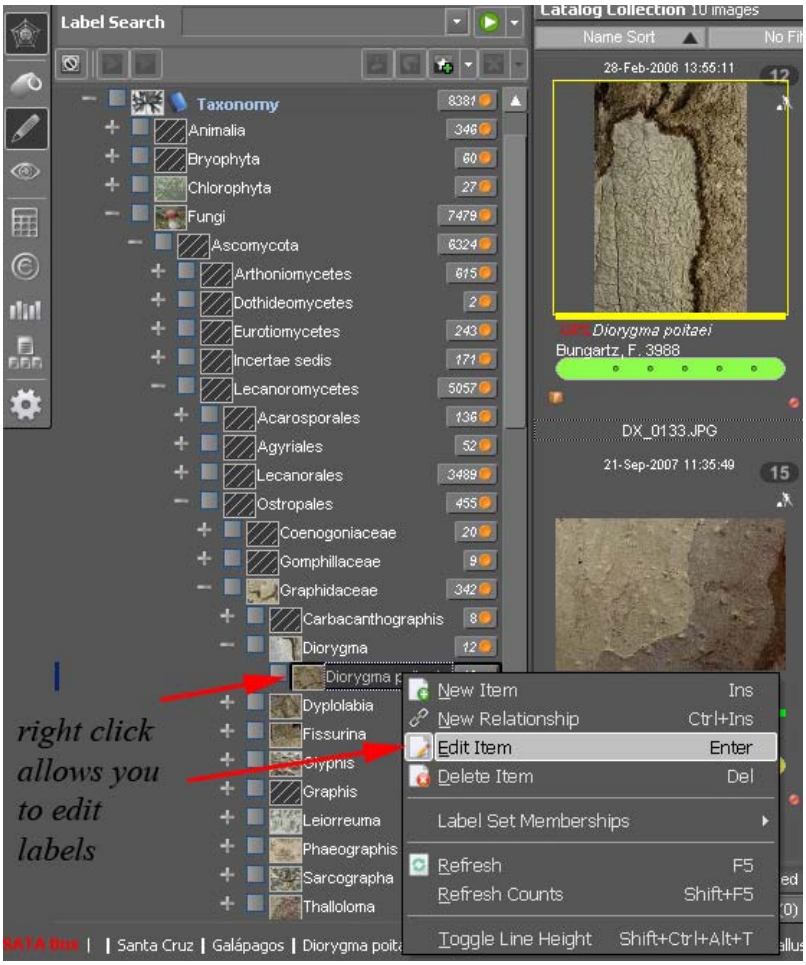

Fig. 6: In IDimager keywords can be organized hierarchically labels; right-click a label to map it to a metadata field (see next figure).

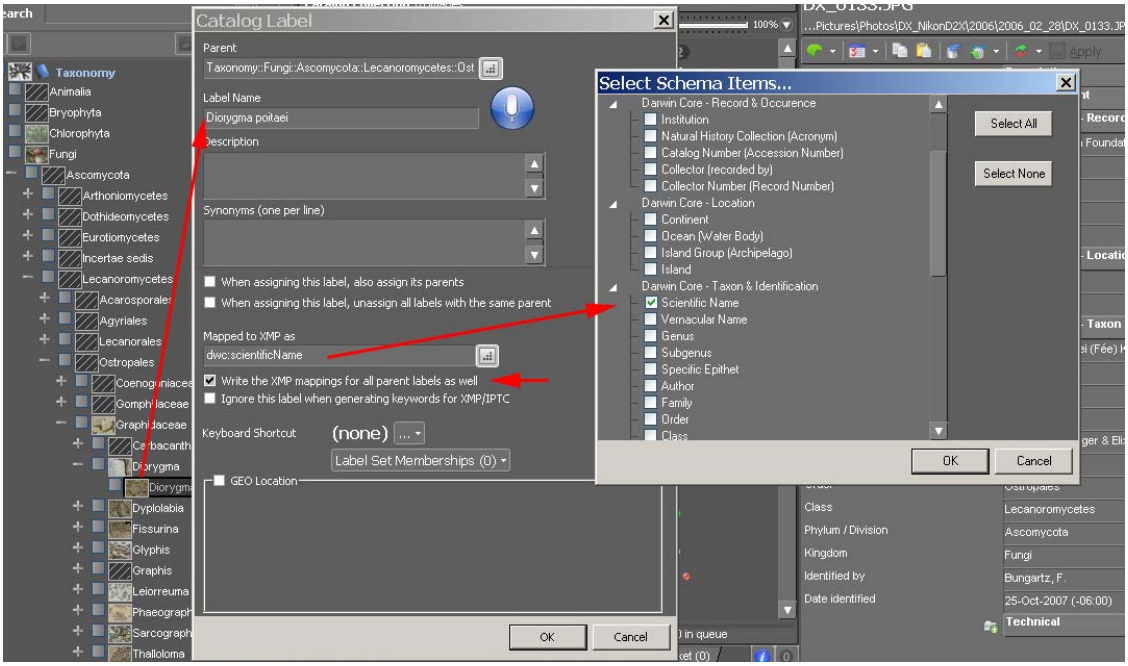

Fig 7: IDimager labels can be mapped to XMP fields. Assigning a label to an image will automatically fill the content of this XMP field.

(3) Finally, if you are managing your specimen data outside IDI in a separate database you can download the script "Import form CSV" from IDI's Resource Repository. This script matches XMP field(s) with records in a CSV table and if the fields match then imports the data into IDI.
For example: For all photos that correspond to a particular specimen, you enter a specimen accession number, or combination collector and collector's number, or any other unique ID that identifies your specimen into the Darin Core XMP fields. Then you export the data that you have in your own Natural History Collection database in CSV format. Make sure that the headers of our fields exactly correspond to the field names in the Darwin Core XMP Schema. In IDI you can now select the images for which you want to add this collection data and run the script.
This way it is very easy to batch import a large amount of collection data that will from now on now be stored inside your files, residing inside the XMP Darwin Core schema. This means each image now has the same complete set of collection and identification data as your specimen.

Detailed instructions and screenshots still follow...

IMPORTANT NOTICE: The Charles Darwin Foundation provides this XMP Schema and the Image Panels free of charge, "*as is*". We have tested the schema and panels. They are being used regularly by us. Nevertheless, we are not responsible in any way for problems you may encounter in using them.

WARNING!!!

Please be aware, that changing the ***XMP Darwin Core Schema*** will make our effort to provide a *standardized* version to manage biodiversity data obsolete. Worse: Although you will be still able to use a modified schema with your copy of IDimager, exchange of the data with other photographers will be difficult or even impossible. Modifying the schema and using different, but highly similar versions in different installations of IDI may also cause malfunctioning.
This warning does *not* apply to the ***Darwin Core Image Detail Panel(s)***. Customizing these panels does not modify the underlying XMP schema. Feel free to add new panels, select new fields according to your needs.

We hope that you find the information and resources provided helpful to improve your workflow as a scientific photographer working with Natural History Collections.