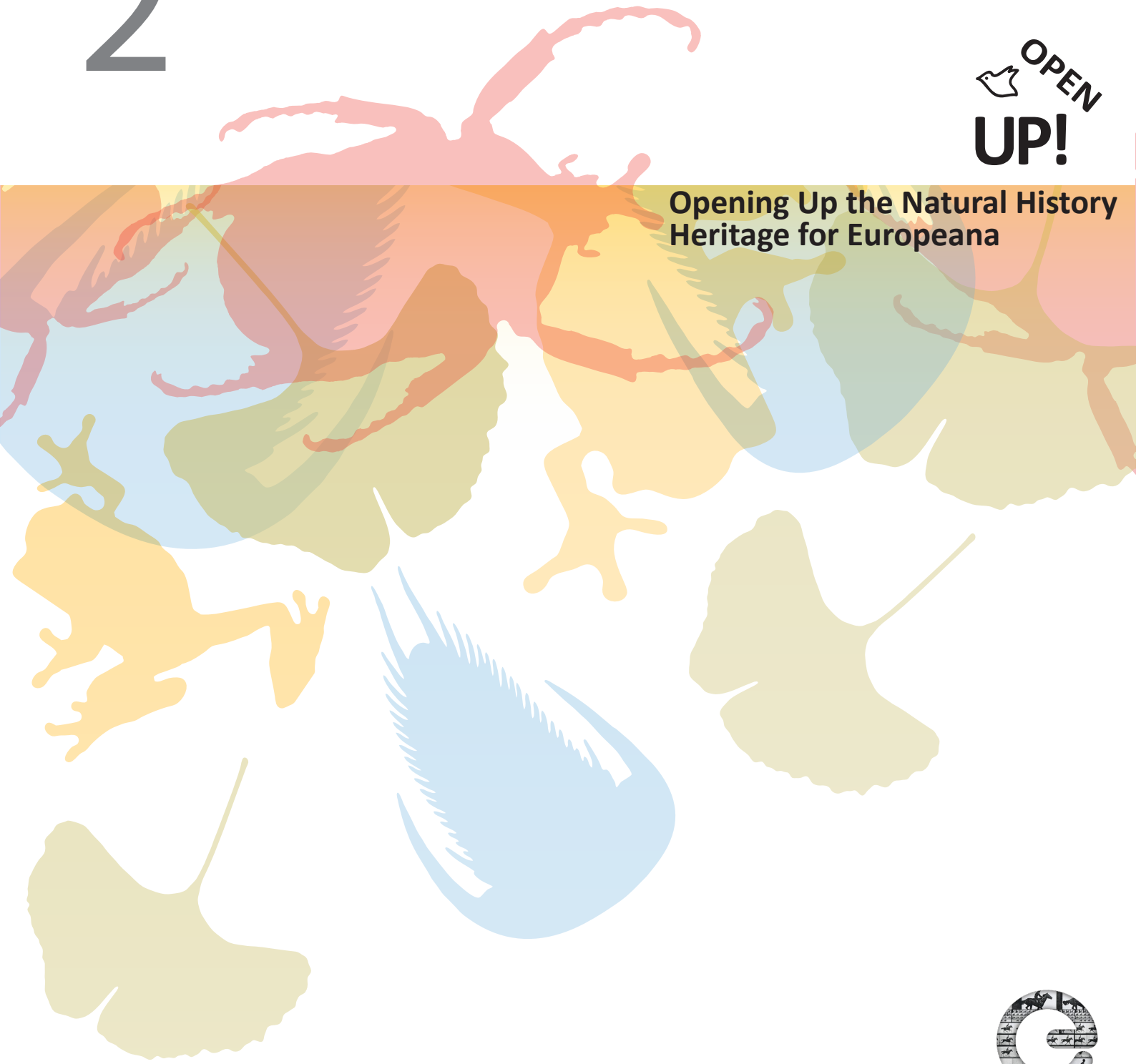
Newsletter

2nd issue

OPEN UP!

Opening Up the Natural History Heritage for Europeana



europeana contributor

co-funded by the CIP-ICT Policy Support Programme

http://ec.europa.eu/ict_psp

Content

OPEN UP!

www.open-up.eu

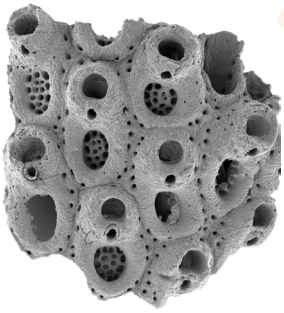
Editorial

Wow! The first 223,953 multimedia objects were delivered to Europeana through the OpenUp! project. Great! A lot of hard work involving all project partners is behind this event. The most challenging tasks are described in this Newsletter.

The delivered multimedia confirm that Europeana is a useful access point for free and verified natural science objects, ready for use in diverse applications, including educational purposes. We are looking forward to the next harvesting cycle, when still more partners and more diversified content will be online. We also expect to attract new content providers, who will then extend the network and transmit more multimedia specimens to Europeana.

Currently you can browse the content by selecting OpenUp! as "content provider" on the Europeana website. Enjoy!

Dr. Kamil Zágoršek, WP8



Knipowitschia montenegrina, holotype of recently described goby from watershed of river Morača in Montenegro; National Museum, Prague. One of the images which are being made accessable through OpenUp!

First OpenUp! content on Europeana!

J. Frank (NM, WP8)

The OpenUp! project delivered the first 223,953 images to Europeana at the beginning of March. The first package included content from 5 OpenUp! content providers. To see OpenUp! content on the Europeana portal, you can use this link http://bit.ly/y4HZiU or type *:* in the Europeana search field and filter your search by OpenUp! provider.

If you are looking for insects, the Finnish Museum of Natural History provided 3,768 images to the Europeana portal.

Each picture/specimen includes a thumbnail and metadata view (fig. 3) with hits of similar content. Clicking on the thumbnail lets you see or download the original image in high resolution (fig. 4).

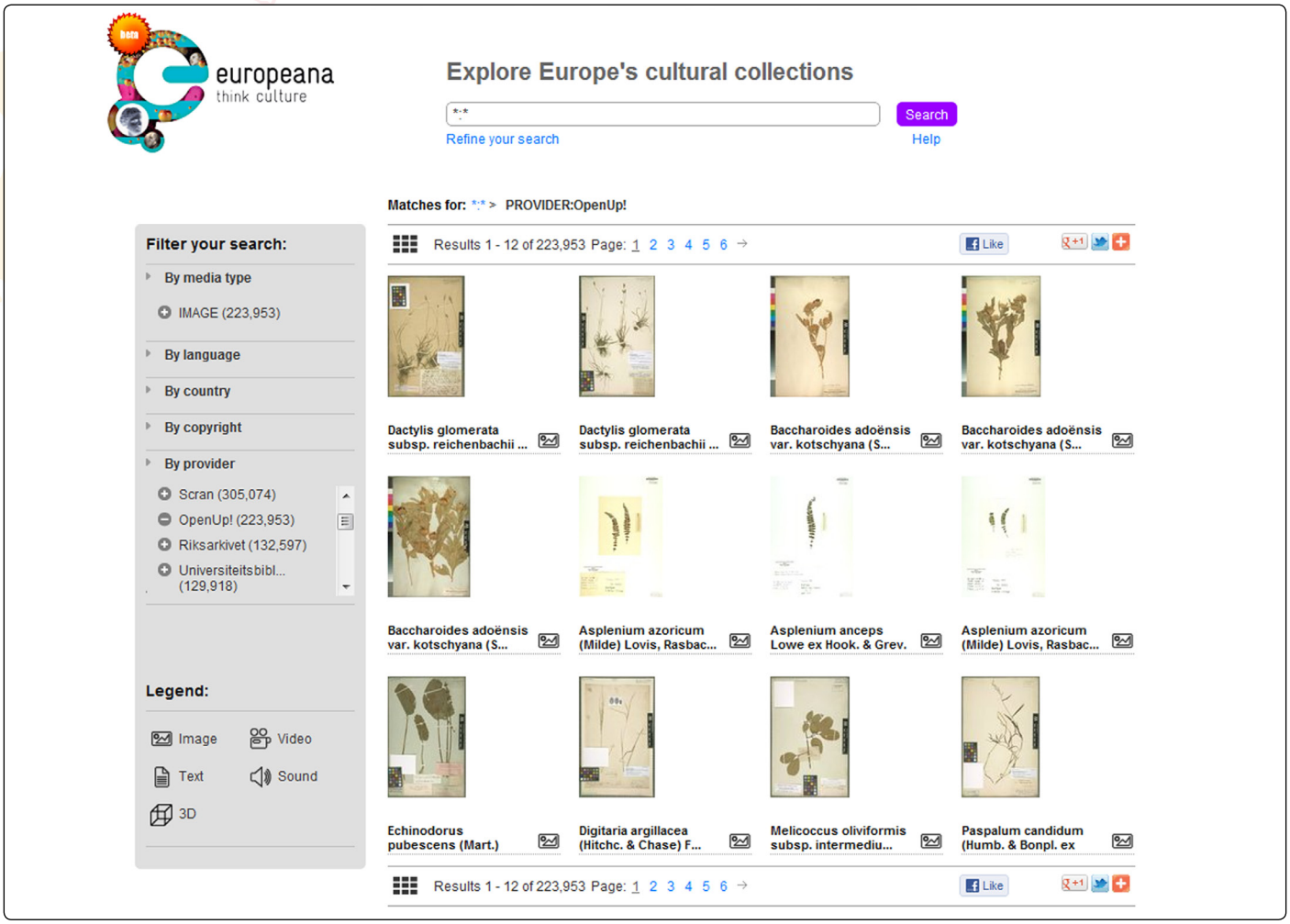


figure 1. Actual OpenUp! content on Europeana portal after first upload.

Within the content you can find 100,766 images from the herbarium collection of the Botanic Garden and Botanical Museum Berlin-Dahlem.

A great collection of mosses and other plants is in the 73,253 images from the Biologiezentrum der Oberoesterreichischen Landesmuseen. Additional 26,488 plants are from the herbarium of the Natural History Museum, Vienna and 19,678 from the herbarium of the University of Vienna, Institute for Botany – Herbarium WU.

Overview of OpenUp! content on the Europeana portal can be seen under this link:

http://www.europeana.eu/portal/europeana-providers.html.

We will inform everyone about Europeana functionality and next content harvest in the news section on our project website, through our social media and in our newsletters.

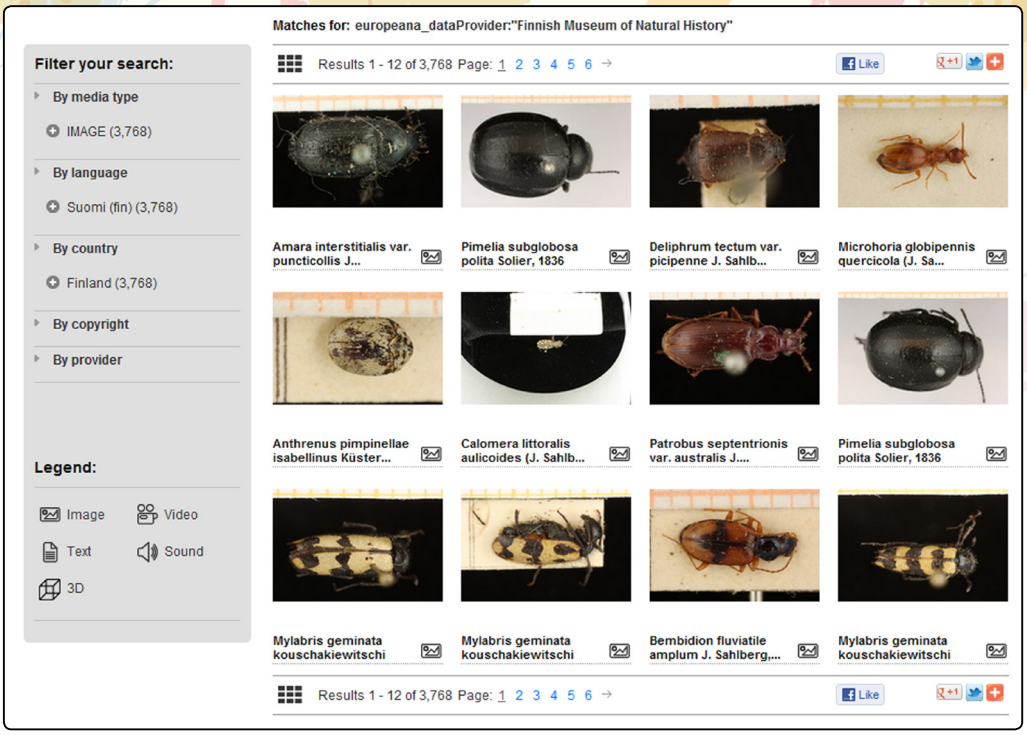


figure 2. Insect collection from the Finnish Museum of Natural History.

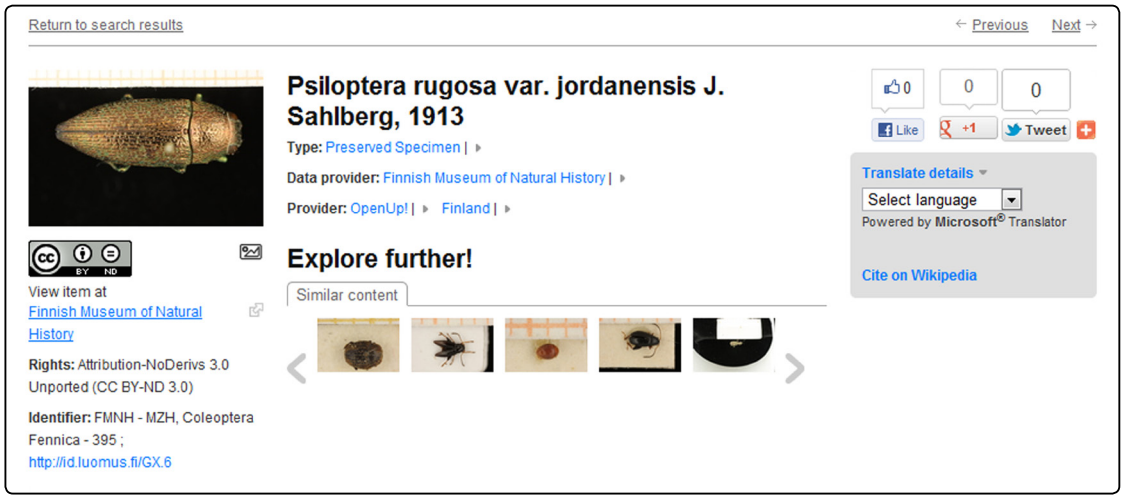


figure 3. image metadata view.

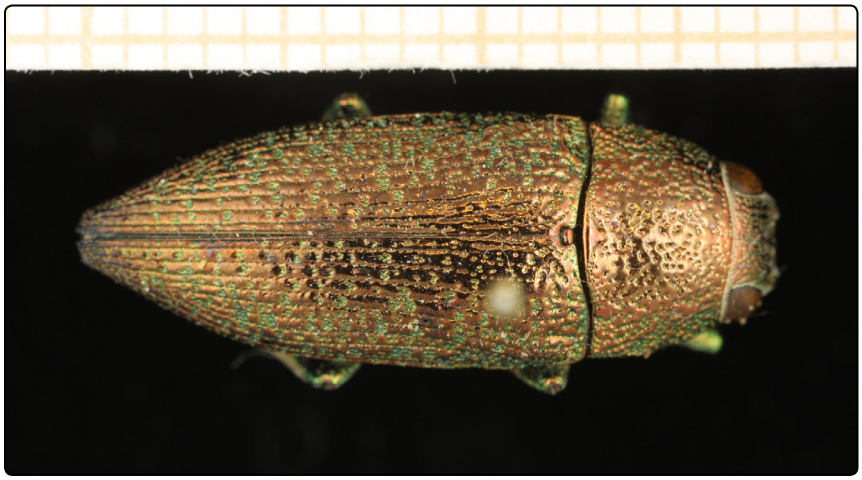


figure 4. Original picture of Psiloptera rugosa var. jordanensis J. Sahlberg, 1913 from Europeana portal

Data Quality Enhancement

E. Zippel (BGBM)

workflow example at the FU-BGBM

Providing botanical multimedia data of high quality is one of the tasks of work package 5 ("Botanical Community Content Provision, Metadata Catalogue and Quality Control") of the OpenUp! project. Along with the multimedia objects whose majority consists of digitised herbarium specimens, EUROPEANA will provide the data belonging to the herbarium voucher, e.g. scientific and vernacular names of the species, the author of the scientific name, the country of origin, the name of the person who collected the herbarium specimen as well as data about date, location, and habitat. Most of these data are usually recorded on a label on the herbarium sheet.

The digital herbarium of the Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM, Index Herbariorum: B) contains more than 120,000 digitised specimens of spermatophytes, ferns, and mosses. The main focus is put on the herbarium Willdenow (B-W, 57,450 digitalised specimens) and type specimens and other original material (more than 34,000 digitised specimens including 2582 Willdenow-types). All digitised specimens are referenced in the herbarium database.

The majority of the specimens documented in the herbarium database are historical specimens which are of particular importance for nomenclatural studies. More recent specimen labels usually provide detailed information on collectors, distribution of duplicates, date of collecting and often comprehensive geographical and ecological data on the collection locality. In contrast, the data on labels of historic specimens are mostly very sparse and sometimes only contain the name of the taxon. More detailed labels list the collector's name (sometimes as an abbreviation), sometimes with collecting number, and/or very short location information, mostly in Latinised form and abbreviated. Usually the labels are handwritten, sometimes in old German letters like Sütterlin or German Kurrent and/or indistinctly and illegibly written, so it is difficult and sometimes nearly impossible to read the labels.

These special characteristics of historic herbarium labels result in an imperfect and incomplete documentation of the specimens: on more than 30,000 specimens in the BGBM database, the country is marked as unknown, over 23,000 lack the entry for the collector and for over 25,000, the author of the taxon's name is missing. As mentioned, sometimes it is not possible to obtain this information, but often it is, although this may be a rather time consuming undertaking.

Another problem in the BGBM's specimen database is the handling of the geo-referencing of specimens collected in the botanic garden itself (marked as, e.g., "cult. in hort. Berol."). Some captured data list "Germany" as country of origin for these specimens, which of course leads to erroneous biogeographic data and must be corrected. The native origin of the specimens cultivated in the garden is checked, and if possible, the country or the continent of the original provenance is added. Like all interpreted data, these should be marked as such.

During the check of a missing country of a specimen, the complete dataset with all fields is verified for correctness, plausibility, and completeness.

Workflow for country check

For the OpenUp! project, I first started with the revision of all digitised specimens without any country information and corrected the origin entry of cultivated species where necessary. The reasons for the missing entries vary and so there are also various ways to solve the problem:

a) ***The specimen is clearly a type or other original material:*** The protologue may provide information of the origin of the species and can help to decipher illegible hand writing or to supplement the origin country, collector's name and/or number. An increasing number of historic titles with systematic publications is available online (BHL, www.biodiversitylibrary.org; BOTANICUS, www.botanicus.org). If the collector is known, the result of the search should be checked for plausibility with different directories of botanists, e.g. http://kiki.huh.harvard.edu/databases/botanist_index.html.

An example where it was necessary to find out the country of the specimen's origin is the type of *Isachne stricta* Elmer (image available online at http://ww2.bgbm.org/herbarium/viewLarge.cfm?Barcode=B%2010%200272956).

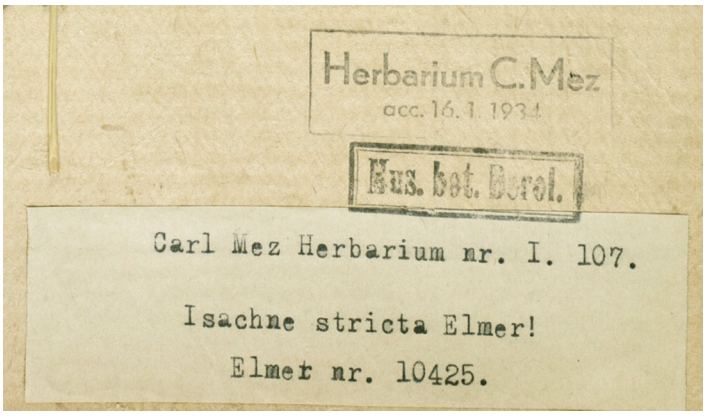


figure 1. Type specimen of *Isachne sticta*; general view of the herbarium specimen (above), detailed view of the label (below).

The label informs us just about the collector and the collector's number, and is marked, probably by C. Mez, as a type. The protologue in "Leaflets on Philippine Botany Vol.2", 1908 , ed. by A. D. A. Elmer (available online at http://www.biodiversitylibrary.org/page/767726#page/92/mode/1up) provides information about collector, collectors number, date, and location of origin:
"Type specimen 10425, A.D.E. Elmer, Dumaguete, Cuernos Mountains, Province of Negros Oriental, Negros, June, 1908". Therefore the origin and type status of the specimen are without any doubt. The City of Dumaguete is the capital of the Philippine Province of Negros Oriental of the Visayas Islands.

b) ***The location name is an old name which is unknown today:*** Research in directories which explain e.g. Latin names such as http://www.columbia.edu/acis/ets/Graesse/orblata.html will help to identify the name. Often a simple Internet query with a search engine search leads to success, since a lot of specimens are cited in various publications and the name of the location can be verified step by step, sometimes with the documentation of other specimen data of the same collector.

c) ***The location name is (nearly) illegible due to its old handwriting and/or ink:*** Asking people who can read the old handwritings will be the best way. If the entry is partly legible, a detailed internet search with changing letters may lead to success. To narrow the search down, it is generally helpful to learn about the distribution of the taxon (see below).

d) ***The data of the origin of the specimen is lacking on the label:*** If the species is endemic to one country or a stenoendemic species, it is possible to add the country or the region to the dataset. Sometimes it is only possible to enter the continent or to leave the origin blank. To avoid mistakes, it is crucial to check for possible homonyms and for plausibility e.g. with the collector's data. To learn about the distribution of the taxon, the use of floras, monographs, or directories like http://apps.kew.org/wcsp/home.do is necessary. Search in the EDIT specimen and observation explorer for taxonomists (http://search.biocase.org/edit/) will give an overview about the distribution as well as other herbarium specimens of the taxon.

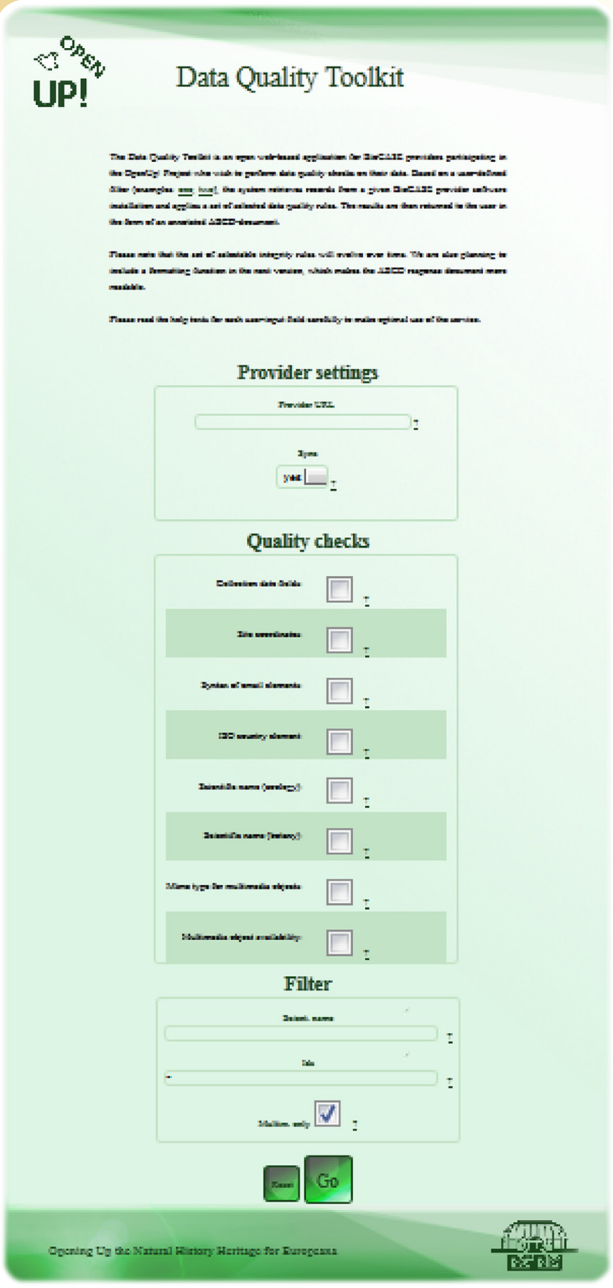


figure 2. OpenUp! Data Quality Toolkit window.

Further common gaps or mistakes in BGBM's database are:

e) ***Lack of the taxon's author:*** Check the author's name in TROPICOS, IPNI, floras, and monographs. Homonyms, which are frequently used in historical collections and sometimes unknown in databases like IPNI or TROPICOS, must be excluded. Sometimes it will not be possible to clear the author's name without the correct identification of the specimen and/or knowledge of the correct origin of the specimen.

f) ***Lack of the collector's name:*** For types, check the protologue, for illegible names, see above.

g) ***Misspelling of the taxon's name:*** Check the name in scientific databases like TROPICOS, IPNI or Euro+Med checklist.

h) ***Wrong entries or entries in wrong fields:*** Sometimes quite funny – like "Hort. Bonn" as collector's name instead of origin of the specimen. "Hort. Bonn" means that the specimen was cultivated in the Botanical Garden (hortus) Bonn!

The described workflow is very time consuming and cannot be substituted by technical tools. However, the OpenUp! Data Quality Toolkit (http://services.bgbm.org/dataqualitytoolkit) assists BioCASE collection data providers with an integrated interface for a useful set of automatic quality checks. These include simple syntax rules for selected ABCD elements (e.g. country, date, latitude, longitude) as well as the comparison of scientific names with international checklists. The set of quality rules is constantly evolving, based on the availability of services and requirements from OpenUp! partners.

The ingestion process for OpenUp! biodiversity data into the Europeana portal

G. Koch (AIT)

One of the main tasks of the OpenUp! project is to "provide a single access point for Europeana[1] to distributed multimedia content in the natural history domain". The envisaged amount of OpenUp!

multimedia objects that should be provided through a OAI-PMH[2] harvesting interface to Europeana are 200,000 records in the first project year, 600,000 in the second project year, summing up to 1.1 million records in the final project year. That is a huge task for the consortium, and the present article provides a

1 Europeana portal. http://www.europeana.eu, 28 Feb. 2012.

2 Open Archives Initiative Protocol for Metadata Harvesting OAI-PMH. http://www.openarchives.org/pmh/ 28 Feb. 2012.

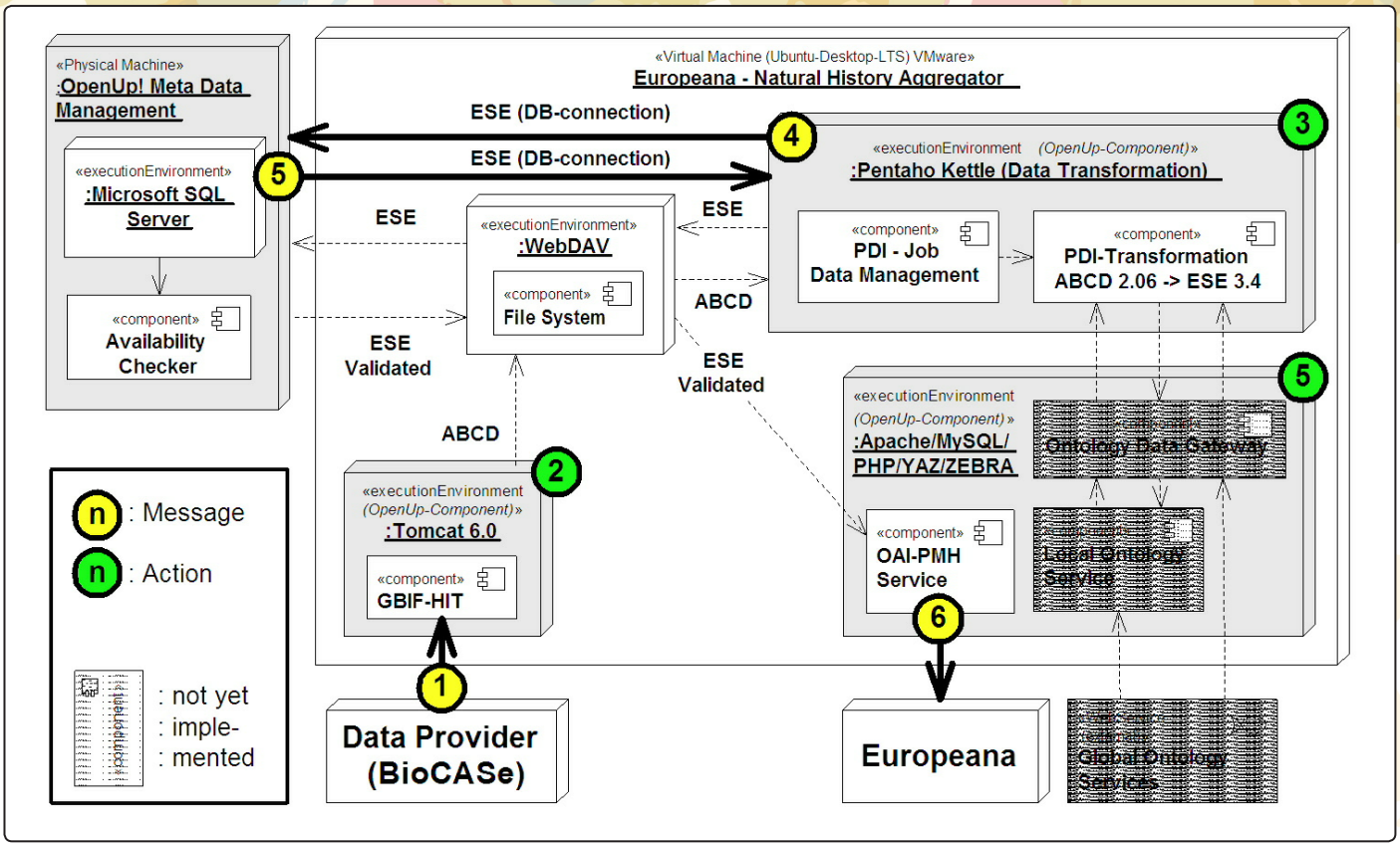


figure 1. Ingesting records into Europeana (overall technical workflow; status February 2012)
Description of the workflow: In step 1 (message) it is acknowledged that data is ready for harvest from the content provider's BioCase Data provider. The data is harvested (step 2 – action) and subsequently transformed (step 3 – action) and copied to the OpenUp! Metadata Data Management Server (step 4 – message). There the data is archived and link checked (step 5 – message). After that the data is provided in the OAI-PMH repository for Europeana harvesting (step 6 – message).

short overview on how the ingestion process to Europeana is being coordinated and carried out.

Where we start from, the content providers' side

The whole ingestion process starts at the OpenUp! content partners' side: The content partners select the content they want to publish in Europeana. They "clean" their data and install the required BioCase provider software[3] in order to test their data. The native databases of the content providers are mapped to the ABCD2.06[4] standard. During this process the content providers are supported by the OpenUp! helpdesk[5]. When the installations of the BioCase software and the mappings have been successfully finished, the content providers report the access point URL and their relevant data sources to the OpenUp! coordination and work package leader teams. Now the technical integration process into the Europeana portal starts.

What we build on, the technical side

The ingest of OpenUp! data to the Europeana portal is supported by the Natural History Aggregator (see overview in figure 1). The aggregator consists of four main components: the GBIF HIT Harvester, the Data Transformation tool, the Availability Checker and the OAI-PMH Service platform.

With the GBIF HIT Harvesting and Indexing Toolkit[6], the ABCD metadata is harvested from the content providers' repositories. The Data Transformation tool transforms this metadata into the Europeana metadata format. The data is then archived and links are checked with the Availability Checker. Finally the transformed data is exposed at the OAI-PMH Service platform for Europeana harvest and ingest into the Europeana portal.

3 BioCASE Provider Software. http://www.biocase.org/products/provider_software/ 28 Feb. 2012.

4 ABCD - Access to Biological Collection Data. http://wiki.tdwg.org/ABCD 28 Feb. 2012.

5 OpenUp! Helpdesk. http://open-up.cybertaxonomy.africamuseum.be/ 28 Feb. 2012.

6 GBIF Harvesting and Indexing Toolkit. http://www.gbif.org/informatics/standards-and-tools/integrating-data/harvesting-and-indexing-toolkit/ 28 Feb. 2012.

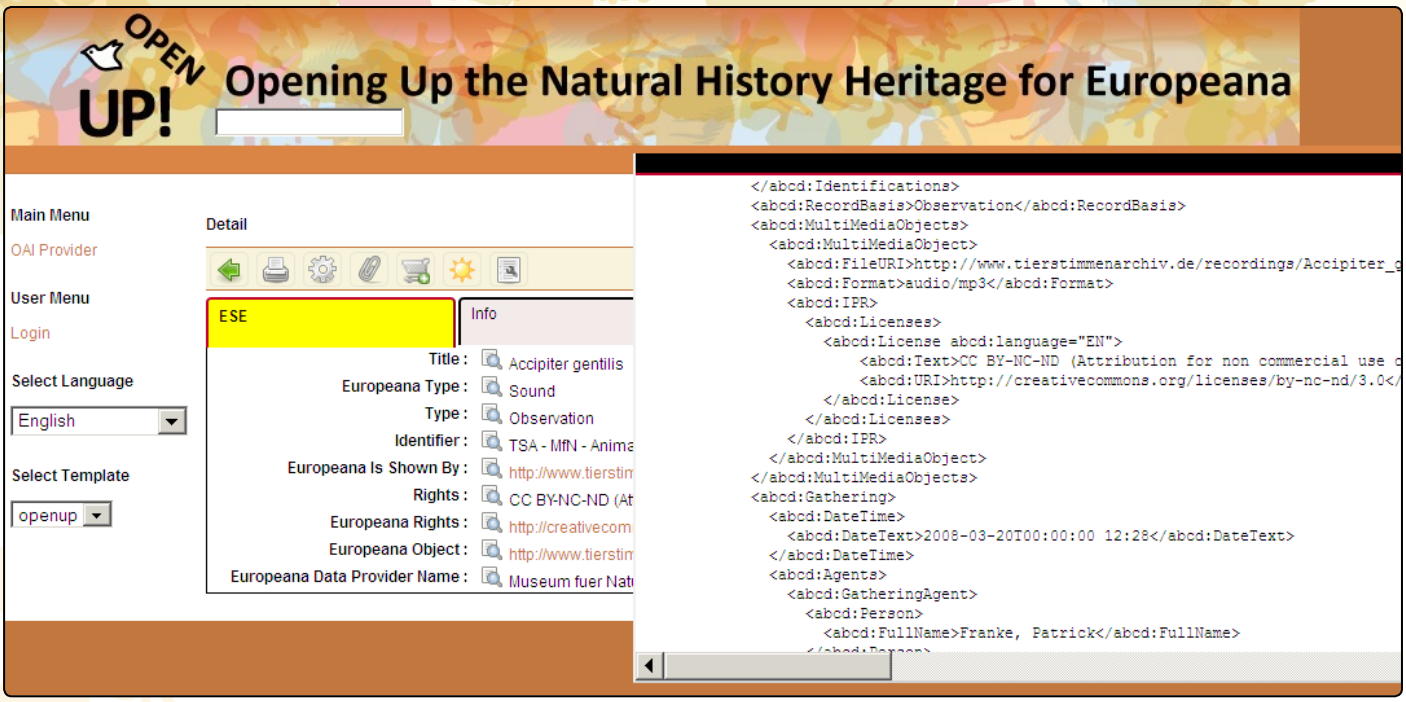


figure 2. Data check with the OpenUp! OAI-PMH platform

Finally getting ready for Europeana, the metadata transformation

In a first mapping workshop in June 2011 in Berlin, the initial mapping of the ABCD2.06 metadata to the Europeana metadata standard (ESE v3.4 – Europeana Semantic Elements) was developed. This mapping built a first basis for the transformation process of the data and is being subsequently refined during the project's lifetime.

As OpenUp! is a best practice project, the state-of-the art open source Pentaho Data Integration system[7] is used to support the data transformation process. Pentaho is an open source business intelligence (BI) [8]solution that features a core data integration engine based on ETL[9]. Extract, transform and load (ETL) is a standard process for integrating distributed data sets into a common target database.

The OpenUp! common database can be accessed via the OAI-PMH Service platform where it is checked to determine whether the transformed data complies with the current Europeana ESE standard. Hence, the platform displays both: the result of the ESE transformation (data in ESE format) and the view on the original ABCD2.06 Unit metadata (see figure 2).

The content providers receive feedback on the data check and get the chance to preview their data in a Europeana look-and-feel environment (see figure 3). The blue arrows point to the browser windows that open up when users click on a thumbnail (direct link to digital object) or on the "View item at …" link (the link to the digital object on the institution's website).

When the final checks are successful and the content provider agrees to the display, the data is released for subsequent publication in the Europeana portal.

A brief discourse on metadata

Metadata is the descriptive information about a data-set, an object (digital or real) or any other information resource (eg. website). Work package 3 of OpenUp! is dedicated to metadata modelling. The following paragraphs name some metadata standards that we have to bear in mind when ingesting OpenUp! data into the European digital library Europeana.

"**ABCD** – Access to Biological Collections Data – Schema is a common data specification for biological collection units, including living and preserved specimens, along with field observations that did not produce voucher specimens. It is intended to support the exchange and integration of detailed primary collection and observation data. […] Development of the ABCD content definition started after the 2000 meeting of the Taxonomic Databases Working Group (TDWG) in Frankfurt/Main, where the decision was

7 Pentaho. http://www.pentaho.com/ 28 Feb. 2012.

8 Business intelligence. http://en.wikipedia.org/wiki/Business_intelligence 28 Feb. 2012.

9 Extract, transform, load. http://en.wikipedia.org/wiki/Extract,_transform,_load 28 Feb. 2012.

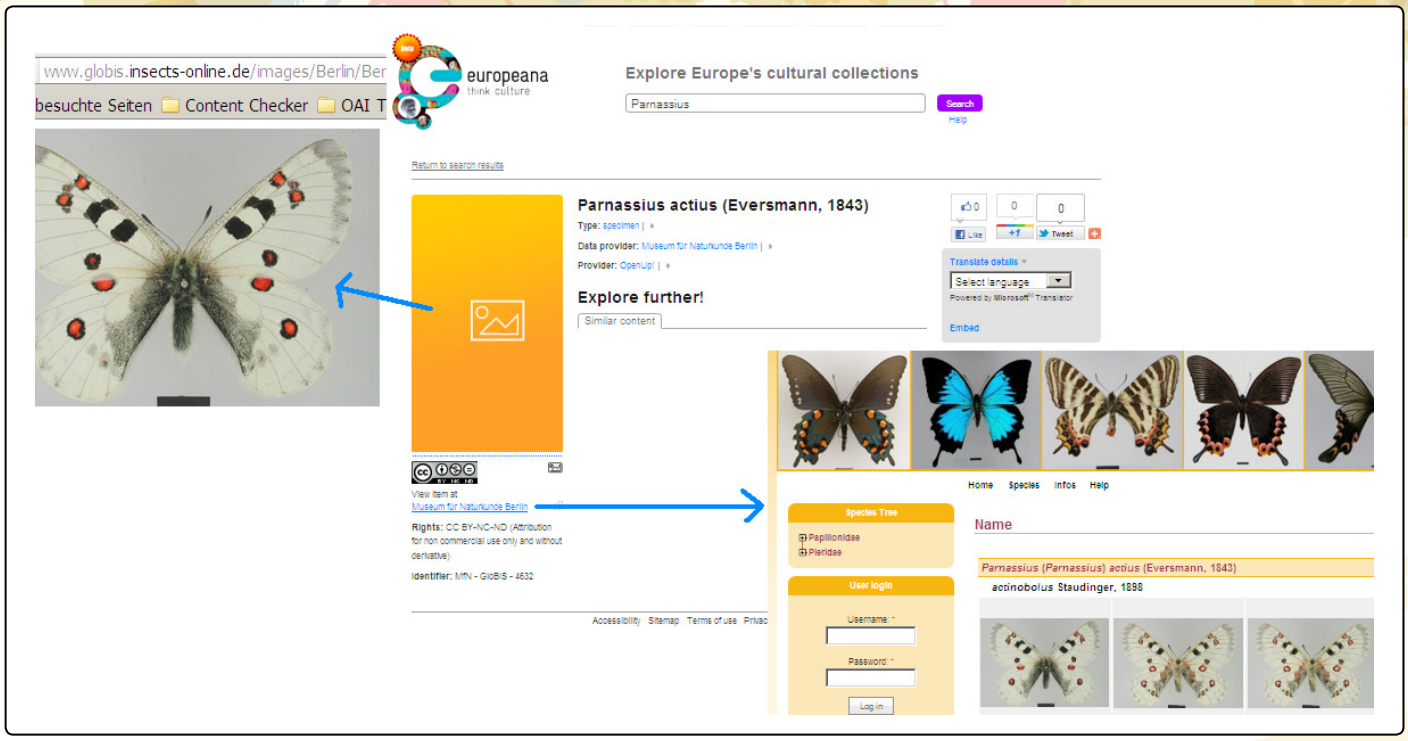


figure 3. Data check with the Europeana Content Checker Tool

made to specify both a protocol and a data structure to enable interoperability of the numerous heterogeneous biological collection databases then available."[10]

The current version of the ABCD standard is version 2.06 (July 2006). This common metadata standard builds one baseline for the transformation process of OpenUp! metadata to Europeana standards.

ESE (**E**uropeana **S**emantic **E**lements)[11] is the metadata standard of the Europeana portal. ESE has been in use in the Europeana portal since 2008.

The ESE standard is based on a subset of Dublin Core standard[12] metadata fields and is enhanced with several "Europeana data fields". The Europeana data fields have been added in order to allow a common display of the various data in the Europeana portal, and to enable certain portal browsing features (e. g. Browse by Copyright, Browse by Provider, etc.). The current version of the ESE standard is version 3.4 (March 2011).

The **E**uropeana **D**ata **M**odel (short: **EDM**)[13] is a far more comprehensive metadata model than ESE and will be the favoured data standard for Europeana in the future. EDM documentation was first published in 2010. One benefit of the data presentation via EDM will be the display of linked data in a wide semantic context. EDM will also support the presentation of the original data of each domain without reducing it to a minimum common denominator (like ESE does now). "EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate particular community standards such as LIDO, EAD or METS." (EDM Primer) The current version of EDM is version 5.2.3 (February 2012).

EDM presentation of data will be implemented in the Europeana portal (http://www.europeana.eu) step by step in 2012. The first OpenUp! ingest of data into Europeana in February 2012 was carried out with data still in ESE format. Future OpenUp! ingests will use the Europeana Data Model, taking eventually advantage of its potential to better display connected biodiversity data sets, records and digital objects.

10 An Introduction to the ABCD Schema v2.0. http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdIntroduction 28 Feb. 2012.

11 Europeana Technical Requirements. http://group.europeana.eu/web/guest/technical-requirements/ 28 Feb. 2012.

12 Dublin Core Metadata Initiative. http://dublincore.org/ 28 Feb. 2012.

13 Europeana documents. http://group.europeana.eu/web/europeana-project/technicaldocuments/ 28 Feb. 2012.

BHL-Europe launches the Biodiversity Library Exhibition!

J. Frank (NM)

The brand new Biodiversity Library Exhibition (BLE) of the Biodiversity Heritage Library for Europe provides a virtual tour of literature illustrations that anyone can enjoy.

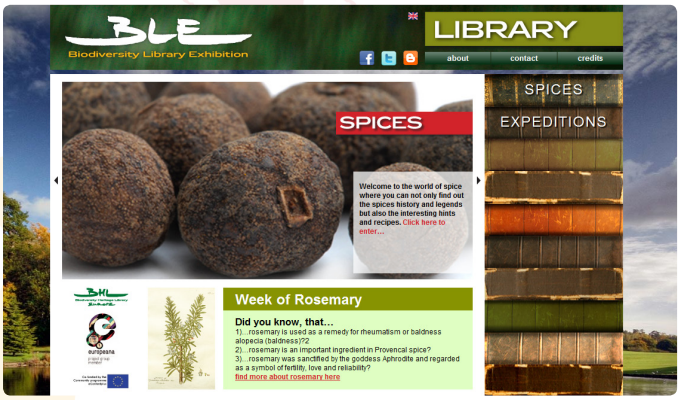


figure 1. Biodiversity Library Exhibition front page

Each virtual exhibition focuses on a specific topic and its representation in natural history literature. The first exhibition, for instance, focuses on spices and their corresponding literature in the BHL and BHL-Europe. Over 52 cards of spices are featured, each with their own illustrations, story and fun trivia. And for all you chefs, the BLE even contains useful recipes. Some of the purposes of this spices exhibition are to show the visitors where the spice came from, introduce the spice plants via beautiful illustrations and detailed pictures of different parts of the spice, such as roots, flowers, seeds etc.

If you are more interested in the extraordinary stories behind some of the most spectacular expeditions ever organized, then the second exhibition, which focuses on historical expeditions, is something for you! By means of trivia, summaries, diaries, and of course illustrations, the BLE makes these splendid stories accessible to the greater public. The voyage of the Beagle, the exploration of North and South America, or even the conquest of Mexico… they're just a few examples of what the expedition BLE has to offer. Currently, the exhibition features 18 expeditions from various periods of time. The illustrations or photos document the stories or interesting facts from different parts of the world.

A central feature of the informative card is the "Find me in books"-function, which groups the literature related to the specific topic (spice, spice plant, expedition or geographical position). A simple click on the illustrations or books will direct you to the BHL or BHL-Europe portals, and the literature featuring the spice, expedition or illustration of your interest. The literature on the exhibitions is also sorted in several categories:

figure 2. Spices – Did you know that Bay Laurel in Greek mythology guarded against lightning strikes and sprigs under the head were reputed to bring good dreams?

country, language and year of publishing. Last category is focusing on useful subjects as illustrations, taxonomy, medicine etc.

As the illustrations and books are linked to the BHL and BHL-Europe portal, the pictures on the cards could be linked for example with Europeana or EOL to show another type of open source content related to this topic. This website application doesn't just show the broad audience how the natural history content could be used, but it also has great potential in the educational sector. The interface is also very user friendly for creating new topics/exhibitions. We will see what the future will bring and maybe it will be possible to see the OpenUp! content at this exhibition through the Europeana portal.

If you are interested you can find a new spice of the week every Friday with additional information on http://bhleurope.blogspot.com/. Visitors can also contribute to the expeditions exhibition as well. If you have a favourite expedition that is not in there, you can prepare a text (according to the structure used on the website) and send it to us via the contact form at http://bit.ly/y25Lfz. You can find more information about expedition contributions at http://bit.ly/wJhceg.

You can visit the Biodiversity Library Exhibition at http://www.biodiversityexhibition.com/.

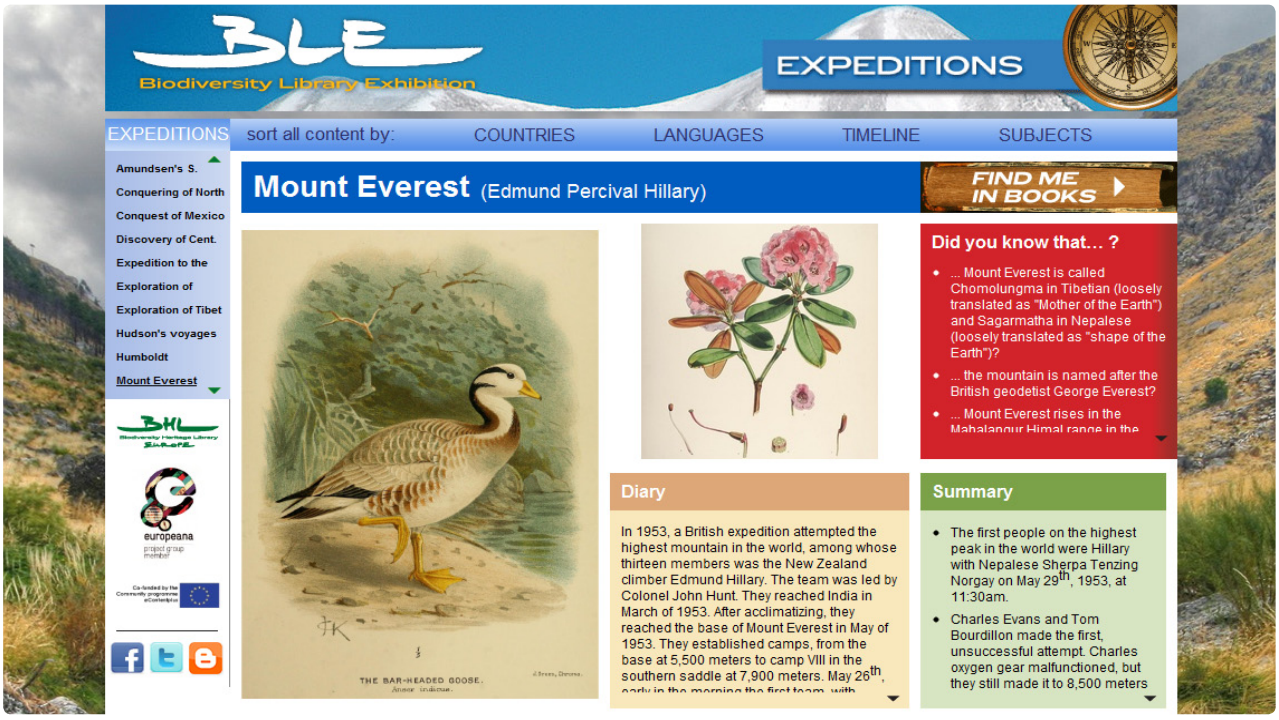


figure 3. Expeditions – Did you know that Anser Indicus the Bar-headed Goose on the illustration was spotted from Month Everest? or Rhododendron campanulatum ssp. aeruginosum grows at the food of Mount Everest?

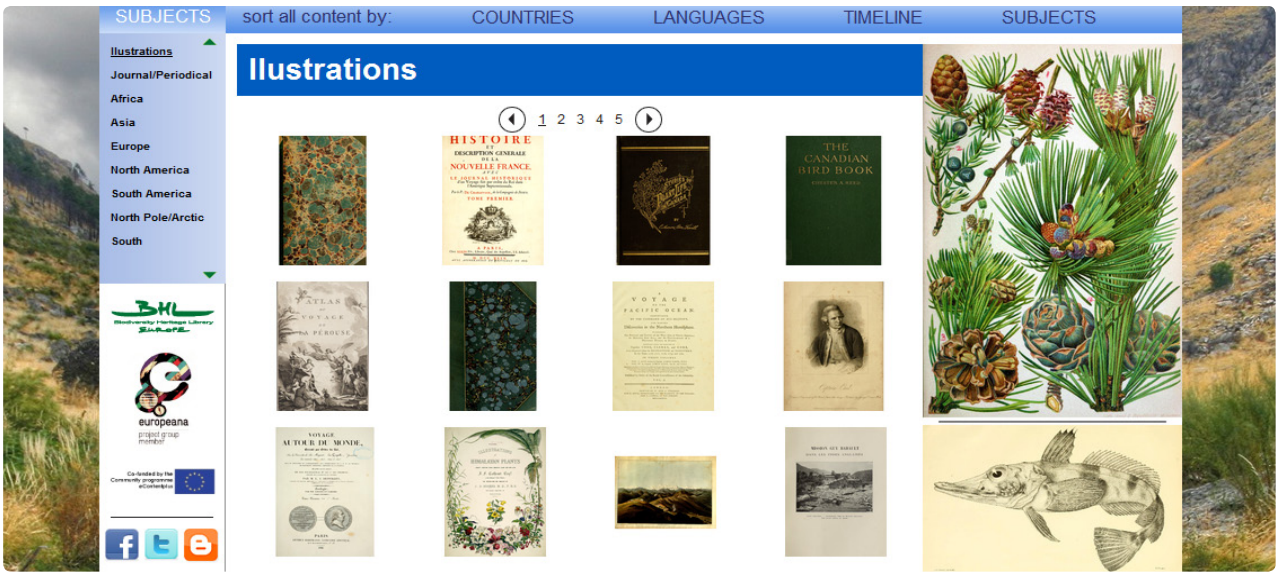


figure 4. category subjects showing books with illustrations

OpenUp! Newsletter

Editorial team: Kamil Zágoršek (National Museum, Prague), Jiří Frank (National Museum, Prague), Jan Sklenář (National Museum, Prague).

This communication tool is issued by OpenUp! at the National Museum, Václavské náměstí 68, 115 79 Praha 2, Czech Republic.

All natural objects figured in the newsletter come from collections of the OpenUp! content providers and will be displayed on the Europeana portal.

www.open-up.eu