BEFdata User Manual

V1.1

Date: 2012-02-20

Abstract: This document is the user manual of the BEFdata portal. It is a guide to the website structure including how to perform common actions. It details how to fill out a template spreadsheet with research data and upload it to the portal. The document concludes with an overview of naming conventions in the portal and on-going and future developments.

Author: Sophia Ratcliffe (ULEI), Karin Nadrowski (ULEI)

Version: V1.0

Table of Contents

1 The BEFdata Portal

The BEFdata portal is an online data management tool that allows cooperative research projects to upload, store and validate their data, in a secure environment. There are currently two instances online:

- **http://fundiv.befdata.biow.uni-leipzig.de/**
- **http://fundiv.befdata.biow.uni-leipzig.de/**

The portal was developed within the Biodiversity-Ecosystem Functioning (BEF-China http://www.bef-china.de/) research unit of the German Science Foundation (FOR 891). The version used by FunDivEUROPE is an instance of the data portal used in the BEF-China project.

Import of data into the portal is via a formatted MS Excel 2003 Workbook, hereafter called the workbook; the only way to upload data into the portal is by completing the workbook. A template of the workbook is available for download from the portal. Data export is organized using paper proposals, in which members can collect datasets into a cart and then issue a request to receive download rights for these datasets.

For help with the BEFdata portal, support details can be found under Imprint in the footer of the website.

This document is a guide to the website structure and shows how to perform common actions. These include filling out the template workbook with research data and upload it to the portal. The document concludes with an overview of naming conventions in the portal and on-going and future development.

In the following we use the FunDivEUROPE instance to describe the use of the BEFdata portal. To aid with understanding, Table 1 lists some of the common terms used in the portal, and this user manual, and their definitions.

Table 1: Description of terms used in the BEFdata portal

Name	Description
Workbook	MS Excel 2003 Workbook formatted to work with the BEFdata portal.
Worksheet	A sheet in the Workbook. There are five sheets that need to be completed to upload data to the portal from the workbook.
Dataset	The complete set of data, including the metadata and raw data.
Datacolumn	A column of data either in the Workbook or in the Dataset.

Datagroup	Sematic classification of the type of data in the Datacolumn.
Datatype	The type of data in the datacolumn: text, number, year, date or category.
Category	Data values of Category datatypes that can be reused between Datasets.
Paper proposal	Data requests accompanied by formulated research ideas
EML	Ecological Metadata Language. It is implemented as a series of XML files.
Cart	Web page on the BEFdata portal that stores the datasets a member has selected for inclusion in a paper proposal.

2 FunDivEUROPE BEFdata Portal website

2.1 The home page and logging in

Figure 1 illustrates the FunDivEUROPE Befdata portal home page. Members can log in through the boxes highlighted in the illustration. Unauthenticated users (those people who have not logged on) can navigate the portal but are not able to download any data, unless the dataset owner has allowed this level of access. They also cannot contribute data to the portal or request data via paper proposals.

Figure 1: Home page of the data portal

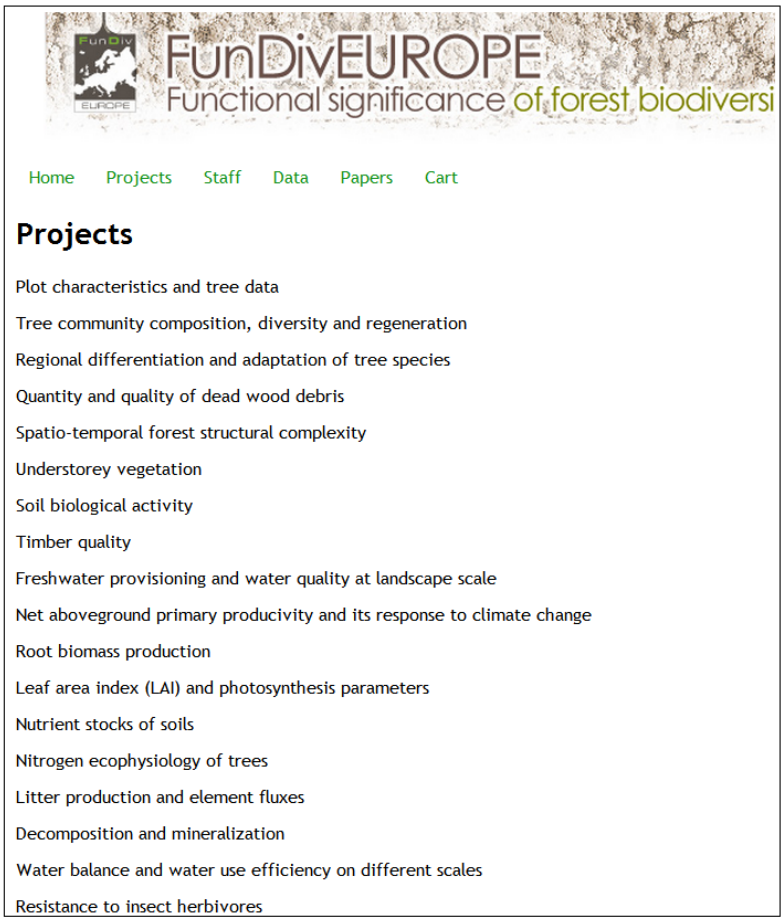
2.2 The navigation menu

The data portal has a very simple naviation, which is available on all pages. This section details the pages available from the main navigation.

Projects

The projects page lists all the sub-projects in the portal. Each FunDivEUROPE task has it's own project listed here, where members of that particular task can upload their data to. See Figure 2 for an illustration.

Figure 2: Projects page of the data portal



Clicking on the project name takes you to a page that lists all the datasets for the that project and any paper proposals that are linked to datasets within the project, see Figure 3. The box in the right-hand column lists all the members of the project.

Figure 3: A project page in the data portal

Staff

Each member of the portal has a page. The staff index page lists all members, including a portrait picture if one has been added. Clicking on the staff name takes you to their home page, which lists those datasets and paper proposals that they are involved in, and in the right-hand column, those projects that they belong to.

Data

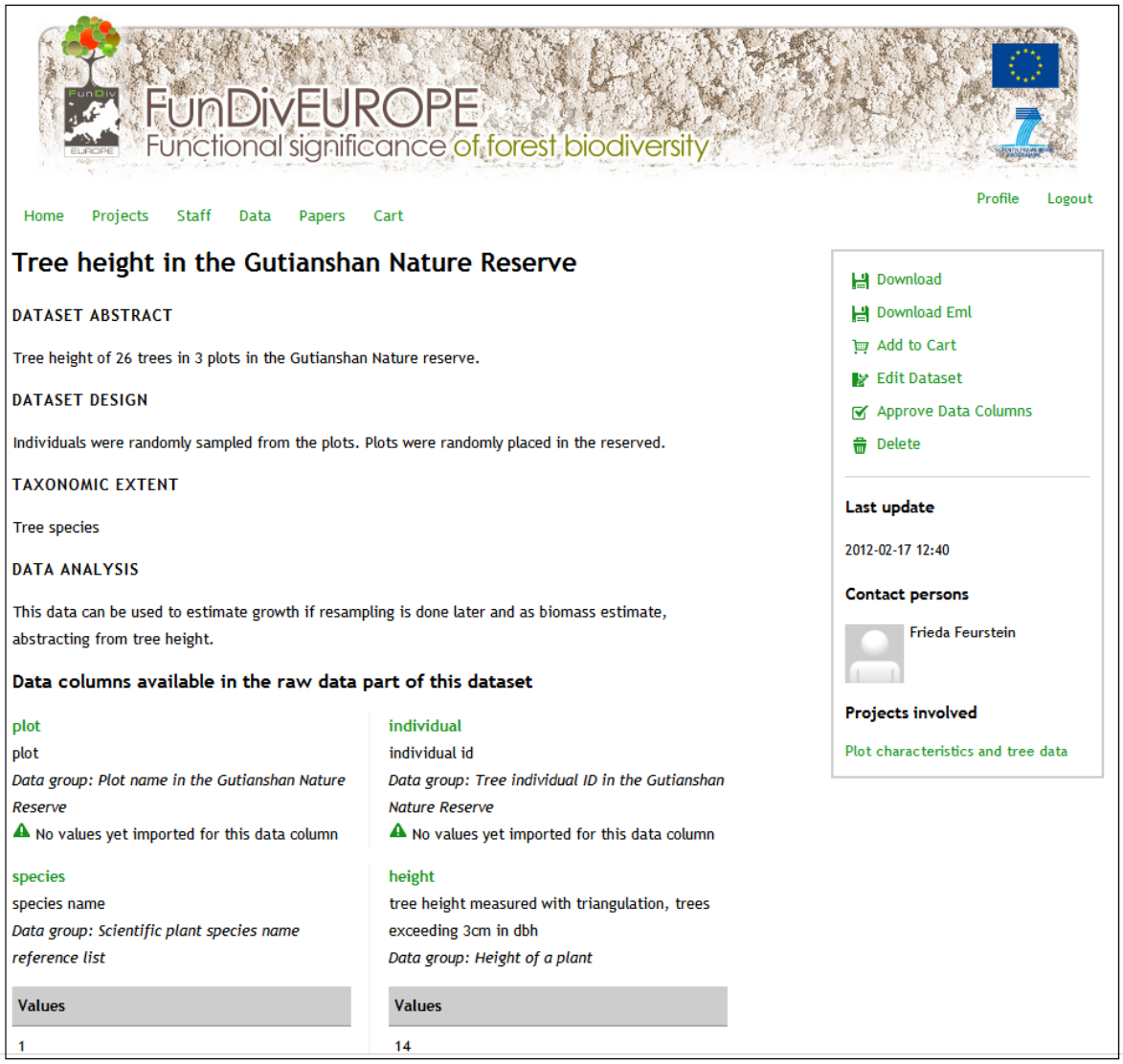
The data index page lists all the datasets that have been uploaded to the portal. Keywords from already uploaded datasets are displayed as a tag cloud in the right-hand column. Clicking on a key word displays links to datasets tagged by the keyword. Logged on users are also able to upload dataset from this page, by clicking on the 'Create new dataset' link. The upload process is described in Section 4.

Clicking on a dataset name opens the overview page of the dataset. This page displays the metadata, datacolumn details and actions that the current user is able to perform on the dataset. All of this information comes from the original uploaded workbook which is described in Section 3.

Metadata is the information required for people, not familiar with the data, to be able to understand and utilise the dataset in their research. Underneath the metadata on the page is a box of information for each datacolumn in the dataset. The title, description and datagroup of the datacolumn are presented, including the first five unique values from the column.

Figure 4 illustrates the dataset overview page as an owner of the dataset.

Figure 4: A dataset overview page in the data portal



The available options in the right-hand panel are different depending on the users' access rights.

Unauthenticated (not logged in users) are only able to see the date the dataset was updated and the list of members to contact about the dataset.

Members of the site who do not have download rights to the dataset and are not owners are able to:

1. Download the dataset metadata in EML format
2. Add the dataset to their Paper Proposal cart

Members of the site who have download rights to the dataset but are not the dataset owner are additionally able to:

3. Download the dataset, which downloads the full dataset to a complete workbook.

In addition to the items above, dataset owners are able to:

4. View the comment field.
5. Edit the dataset, including complete re-upload. This is discussed in more detail in Section

The comment field is only available for owners and those with 'project board' view. The comment also will appear in any short view of datasets in lists for the owners and project board members. The delete option is only available to administrators.

Paper proposals

Paper proposals are data requests accompanied by formulated research ideas that specify what data is needed and whose expertise should be consulted. Members who would like to use datasets for joint publication must submit a paper proposal that will be reviewed by the project board to make sure that it is novel and doesn't compete with other activities. In a second step, it is then reviewed by all dataset owners listed in a paper proposal. Datasets can be added to a logged-in user's cart and this collection of datasets is then used as part of a paper proposal. The proponents can only gain access to the datasets once all owners have approved the proposal. At the same time, a preliminary author list is collected based on the proponents and dataset owners. Dataset owners gain a 'co-author' role in the paper proposal (not necessarily in the resulting paper), whilst members mentioned in the Acknowledgement sheet (worksheet 2) of the data workbook will be added to the acknowledgements section of the paper proposal.

3 BEFdata portal workbook

Data, and it's associated metadata, are uploaded to the portal from a formatted MS Excel 2003 spreadsheet. As long as the raw data is in a table format (where the columns are the different measurements and the rows are each observation) then any type of data can be uploaded. There is no size restriction, apart from those of Excel (64,000 rows), but larger datasets (those over 10,000 rows) can take some time to upload. An empty, templated, workbook is available for download from the site and it is recommended to download the latest workbook so that you are always using the latest version. The title of the template workbook dataset is 'How to fill out a BEFdata Workbook'.

The workbook has five sheets. The first four sheets describe the data and the fifth contains the raw data. Table 2 describes each sheet in the workbook.

Table 2: Description of each sheet in the MS Excel workbook template

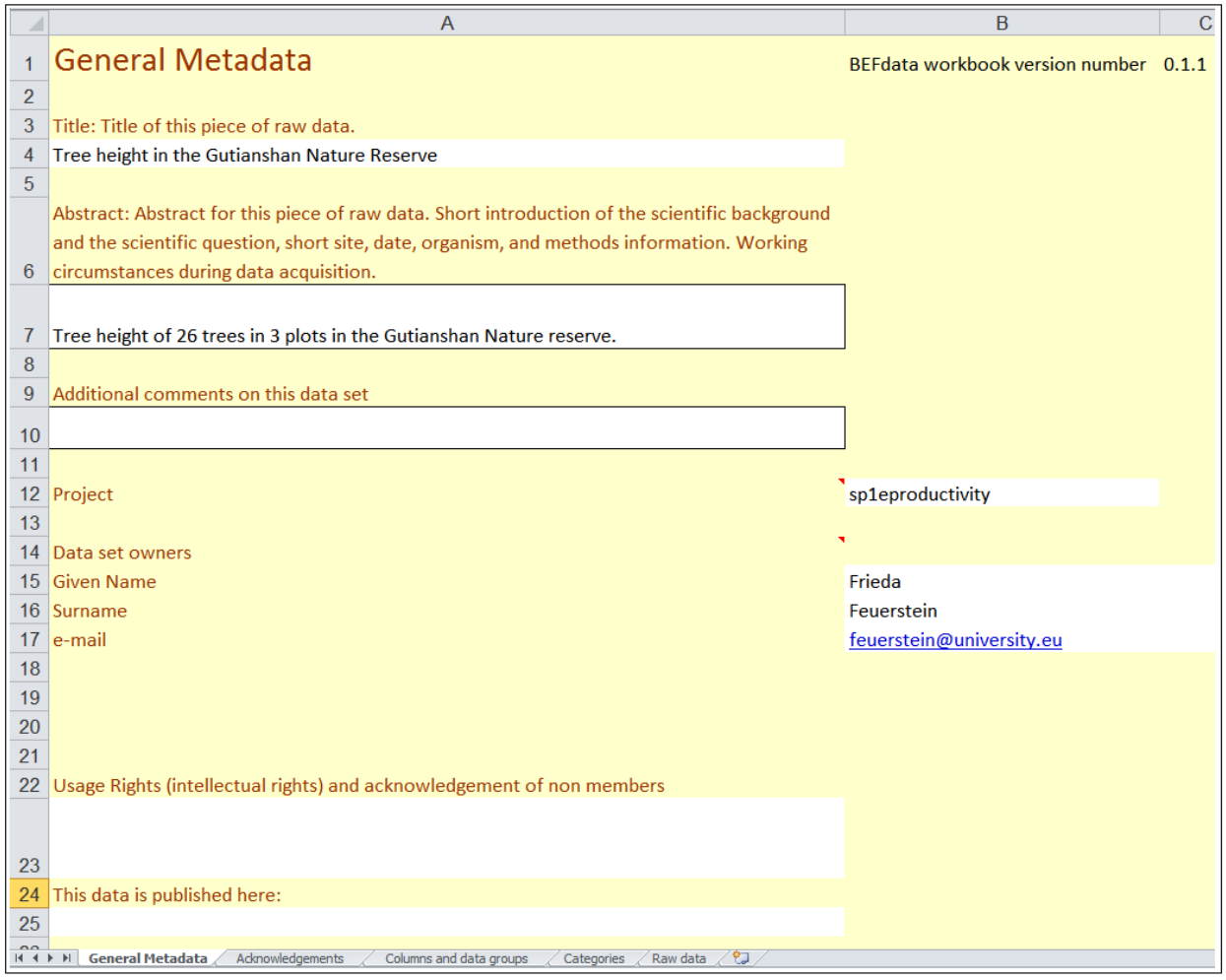
Name	Description

General metadata	Contains fields for dataset title, abstract, location, project and data ownership.
Acknowledgements	Lists any users who have contributed to data capture or collection for a particular column, other than those listed in the 'General metadata' sheet.
Columns and datagroups	Lists all the columns in the 'Raw data' sheet. Including definitions, unit of measure and data type.
Categories	Lists the individual options available for category datatypes used in the 'Raw data' sheet.
Raw data	Contains the actual data, each row as a single observation.

3.1 General metadata – sheet one

The general metadata sheet captures the title, abstract, project and any additional information required to understand and interpret the data correctly. Data ownership is also held in this sheet and during upload the current user is added as the data owner. Only the dataset title is required; all other fields are optional. The title must be unique within the portal otherwise an error will be raised.

Figure 5: An example of the General Metadata worksheet



	A	B	C
1	General Metadata	BEFdata workbook version number	0.1.1
2			
3	Title: Title of this piece of raw data.		
4	Tree height in the Gutianshan Nature Reserve		
5			
6	Abstract: Abstract for this piece of raw data. Short introduction of the scientific background and the scientific question, short site, date, organism, and methods information. Working circumstances during data acquisition.		
7	Tree height of 26 trees in 3 plots in the Gutianshan Nature reserve.		
8			
9	Additional comments on this data set		
10			
11			
12	Project	sp1eproductivity	
13			
14	Data set owners		
15	Given Name	Frieda	
16	Surname	Feuerstein	
17	e-mail	feuerstein@university.eu	
18			
19			
20			
21			
22	Usage Rights (intellectual rights) and acknowledgement of non members		
23			
24	This data is published here:		
25			

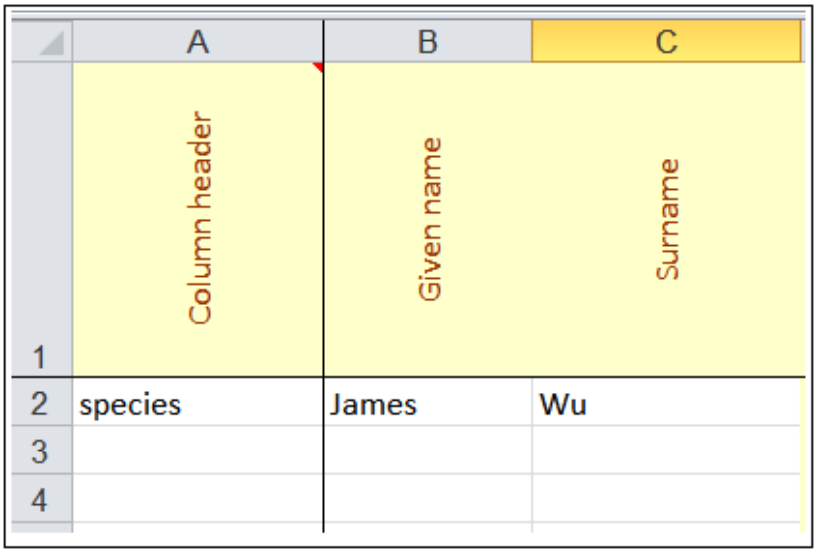
General Metadata | Acknowledgements | Columns and data groups | Categories | Raw data

The more information you are able to add to the metadata sheet the easier it will be for project members to understand and be enable them to use it.

3.2 Acknowledgments – sheet two

The acknowledgements sheet allows project members, who were responsible for the collection of data in particular datacolumn to be referenced, so that data ownership is not lost. Figure 6 illustrates the acknowledgement sheet where the data in two columns (Latitude and Longitiude) have their ownership set.

Figure 6: An example of the Acknowledgements worksheet



	A	B	C
1	Column header	Given name	Surname
2	species	James	Wu
3			
4			

To reference people simply add the column title and their first and last name in a single row.

3.3 Datacolumns and datagroups – sheet three

The datacolumns and datagroups sheet describes the columns in the data. Each datacolumn in the raw data sheet is represented by a row in the sheet. An example of the third sheet is shown in Figure 7. The sheet is separated into two parts: one that is specific to the dataset (from 'Column header' to 'Keyword'); and a part that can be reused within and between datasets ('Data group' ,'Data group description'). For example (see the highlight in Figure 7), the DBH of a tree can be measured in centimetres or in millimetres. In this case 'Tree size' is the data group and can be reused across datasets, whilst the dataset specific part is that tree height was measured in centimeters (stored in the 'unit' column).

Figure 7: An example of the Columns and Datagroups worksheet

	A	B	C	D	E	F	G	H	I
1	Column header	Column definition	Data type	Unit of measurement within dataset	Instrumentation for datagroup	Reference for data group	Keywords (comma separated)	Data group title	Data group description
2	plot	plot	category				Gutianshan stu	Plot name in th	Research plot n
3	individual	individual id	category				tree	Tree individual	Trees in the Gut
4	species	species name	category				species	Scientific plant	Scientific plant r
5	height	tree height mea	number	meter	Vertex		plant height, bi	Height of a plar	Estimating plant
6	date	date	date(2009-07-14)				date	Measuring date	Date of measur
7									

There are 5 different datatypes that can be used in the portal and the choice of datatype is important because the data will be validated against specific rules for each datatype. If no datatype is provided, or an invalid datatype is used, then the user will be promoted to select one from one of the 5 built-in datatypes during the upload process (see Section 4).

Table 4: Description of the five datatypes

Datatype	Validation rules
Text	No validation.
Date	Validated as a date. Once Excel recognizes the value as a date it will be handed over to the portal in the correct format. In this case you can choose 'date(2009-07-14)' as datatype.
Number	Validated as a numeric. Negative numbers and decimal places are allowed. If there are numerical categories predefined for the datagroup, they will be handled as categories (e.g. -9999 for missing altitude).
Year	Valitated as year (4 digit number)
Category	Validated against any categories already present in the portal for the given datagroup, and then against the list of categories included in the Categories sheet of the Workbook.

The datagroup and datagroup description are useful for describing or classifiying columns of similar data that are used across several datasets. Consistent use of categories and datagroups within and between datasets is essential when searching for, and merging, datasets. The use of naming conventions is described in more detail in Section 5. To help select a suitable datagroup the list of datagroups already used in the portal can be found here; http://fundiv.befdata.biow.uni-leipzig.de/datagroups.

Keywords included for a column are added to a 'tag cloud' of keywords on the Data page. Users can then use the tag cloud to find datasets.

3.4 Categories – sheet four

The categories sheet details any naming conventions used in the data, for example species codes instead of their full scientific name. Categories are the individual values of any category datatypes. If the category has already been uploaded in a previous workbook it is not required to list them again, assuming that they are identical. On validation the categories are matched to existing categories from the same data group and if one is not found the user will be prompted to create a new category during the upload process. We recommend that categories are added to the sheet as it speeds up the upload and validation process, however it is not necessary. See section 4 (Datatype approval) for more details.

Figure 8 shows an example of a completed category sheet.

Figure 8: An example of the Categories worksheet

	A	B	C	D
1	Column header	Category short	Category long	Category description
2	species	1	Ca.he	Castanea henryi
3	species	2	Pi.ma	Pinus massoniana
4	species	3	Sc.su	Schima superba
5	species	4	Ch.ax	Choerospondias axillaris
6				

Each category should be added as a row in the spreadsheet with the column name in which the category value appears in the first column. The category value goes in the second column (Category short) and the longer category name and description are in the third and fourth columns, respectively. A long category name is required when the purpose of the category is unclear from the short name. If no category long or category description are provided then the short name is used.

3.5 Raw data – sheet five

The Raw data sheet contains the data, one row for each observation. There is no limit to the number of rows, other than MS Excel's own limit, however, over 20,000 rows can take an hour to upload.

Figure 9 gives an example of a Raw data sheet.

Figure 9: An example of the Raw Data worksheet

	A	B	C	D	E	F
1	plot	individual	species	height	date	
2	1	1	1	5	05-Jan	
3	1	2	2	15	05-Feb	
4	1	3	1	4	07-Jan	
5	1	4	1	15	07-Jan	
6	1	5	2	14	05-Jan	
7	1	6	2	20	05-Feb	
8	1	7	3	5	07-Jan	
9	1	8	4	15	07-Jan	
10	1	9	4	4	05-Jan	
11	2	10	2	15	05-Feb	
12	2	11	2	14	07-Jan	
13	2	12	3	20	07-Jan	
14	2	13	3	5	05-Jan	
15	2	14	3	15	05-Feb	
16	2	15	3	4	07-Jan	

It is essential that the column headers in the raw data sheet match those used in the acknowledgements, columns and datagroups, and categories datasheets otherwise the import processs can't match the column descriptions to the data. The column headers must also be unqiue. The upload process will return an error if this is not the case.

4 Data upload and approval

Once the data has been compiled within the MS Excel 2003 Workbook and described correctly, the workbook can be uploaded to the portal.

The process of uploading the data has two steps: the first step is to save the metadata (from the first sheet in the workbook) and the initial upload of data; and the second is the process of going through each column of the uploaded data and verifying the datatype and data group, and adding any additional column descriptors. If the workbook has been described correctly then it is possible to approve all the columns in one go.

The following is needed for the data to be adequately described so that the upload processs runs with no problems:

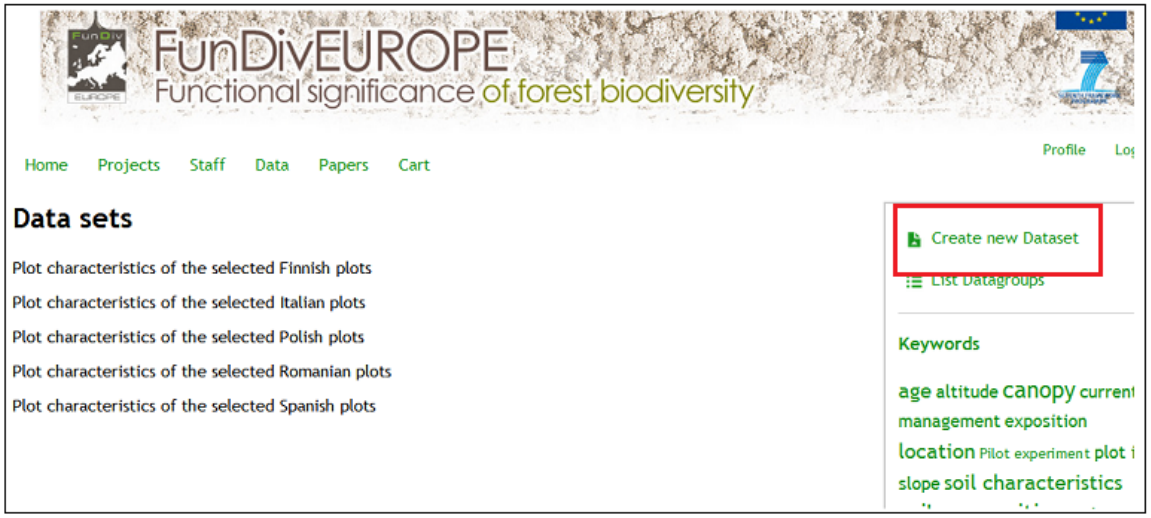
- Use the most current version of the workbook by downloading the dataset 'How to fill out the BEFdata Workbook'
- Unique column headers in the raw data sheet;
- Column headers from the raw data sheet match those in the other sheets of the workbook;

- Correct datatype used in the column and datagroup sheet (text, year, date(2009-07-14), number or category);
- A datagroup and datagroup description are included for each column in the column and datagroup sheet. These do not have to exist in the database, however, when applying naming conventions a known datagroup has to be chosen.
- When category datatypes are used, their individual values are stored in the category sheet, unless you know that the categories already exist in the portal for the datagroup.

4.1 Upload – selecting a dataset

To upload data, click on the 'Create new dataset' link on the **Data** page, as shown in Figure 10.

Figure 10: Location of the 'Create new dataset' link



On the next page use the file upload button to select the workbook from the file system, and click on the '**Create new dataset**' button.

4.2 Upload – saving the general metadata

Figure 11 illustrates the top section of the general metadata page which is the next page in the workflow. The information is taken from the General metadata sheet of the workbook and can be amended on this page. The additional metadata fields and the save button are not visible in Figure 11. There is an additional section that the user can complete that is not available in the General metadata worksheet, for user rights.

There are four checkboxes that determine the level of visibility of the dataset to users of the portal. Table 4 describes each option.

Table 4: Dataset acces level options

Access level	Description
Visible for public	This is checked by default. The dataset is visible to unauthenticated users. They are not able to download the data.
Free for public	Unauthenticated users are able to download the data.
Free for members	Members of the BEFdata portal (logged in users) are able to download the data.
Free within projects	Members who belong to the project in which the dataset resides are able to download the data.

Figure 11: The top section of a metadata page

To save the metadata click on the '**Save metadata**' button at the bottom of the page.

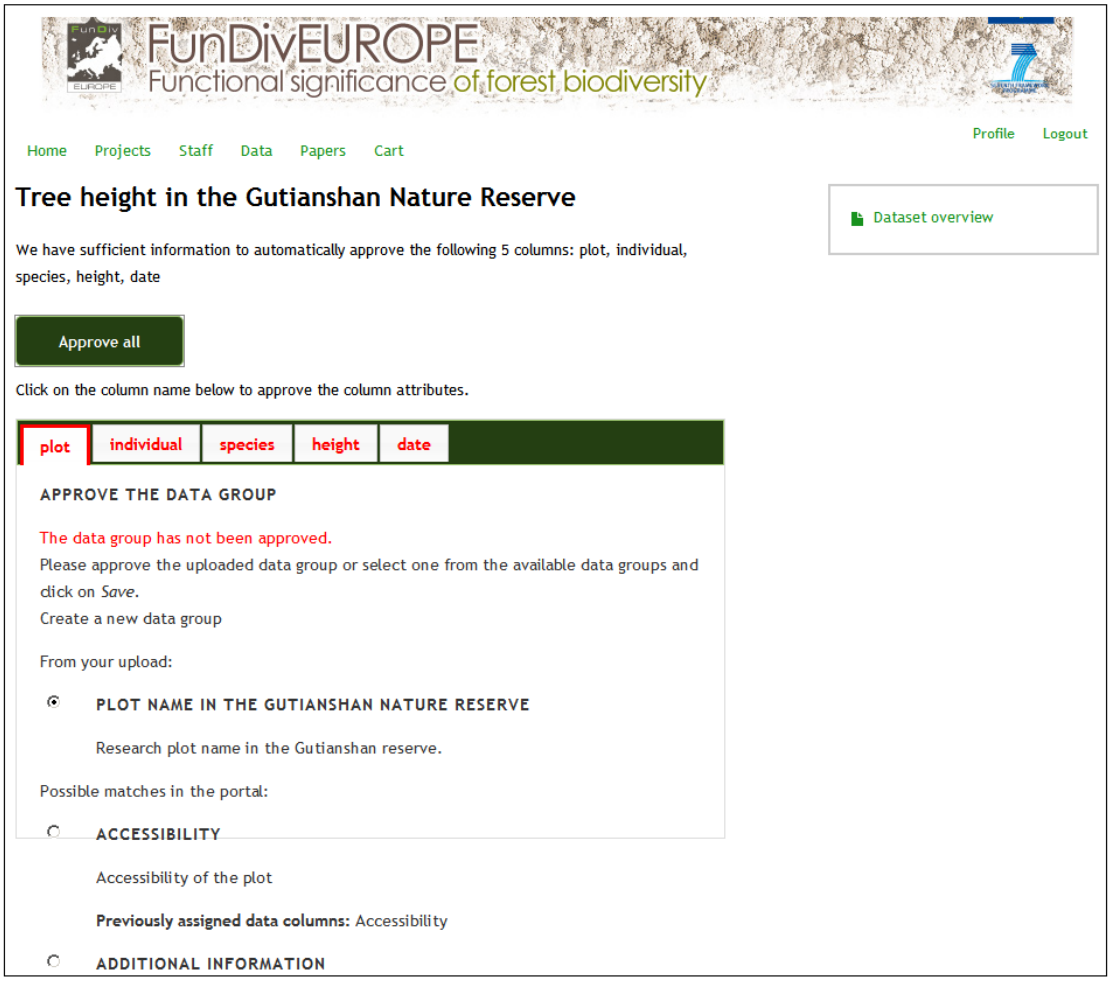
At this point all the data is uploaded to the database in an unvalidated form. For large datasets (>20,000) this can take upto an hour. A future development will be to make this upload run asynchronously so that the user can navigate away from the page and use the portal, whilst the upload continues.

Once the metadata has been saved the dataset is available on the portal.

4.3 Validation and approval

Once the data has been uploaded in an unvalidated form the dataset overview page will be displayed, as shown in Figure 12.

Figure 12: Dataset overview page after initial upload



The red text on the button tabs indicates that the column is unapproved. Orange text indicates partially approved (the datagroup has been approved but not the datatype), and green indicates that both the datagroup and the datatype have been approved.

In the case shown (Figure 12), all the data columns were correctly described in the workbook and so it is possible to automatically approve them in one go, rather than clicking through each of the columns. To automatically approve all of the columns click on the '**Approve all**' button.

Columns where both the datagroup and the datatype are approved, and that have no invalid values, will be green, those where the portal was able to approve the datagroup by not the datatype will be orange, and those where the portal not able to approve the datagroup will remain red. It is necessary to go through each of the red and orange columns and manually approve the datagroup and datatype.

Datagroup approval

Clicking on a 'red' column will display the form to approve a datagroup. This is also the default view, as illustrated in Figure 12.

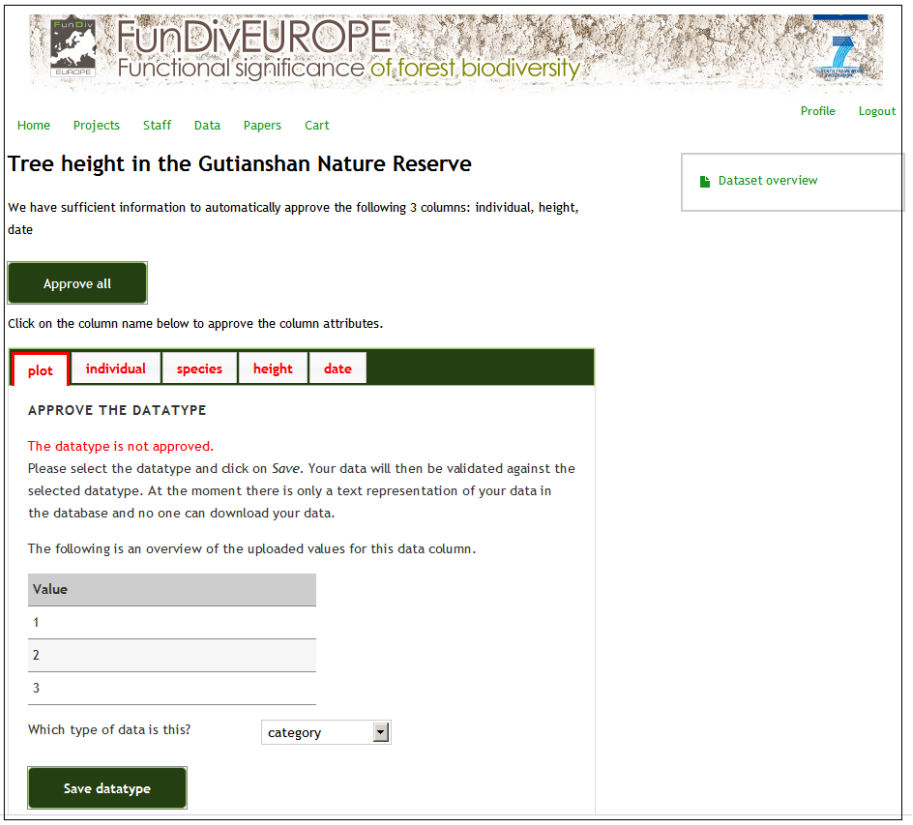
The approval page lists all of the available datagroups in the portal, plus the datagroup entered for the current column in the workbook. The datagroup from the workbook is the first in the list and is selected by default. It is possible to select another datagroup if a mistake was made in the workbook. Clicking on the '**Save**' button (not shown in Figure 12) will approve the datagroup.

It is also possible to add a new datagroup: clicking on the '**Create a new data group**' link will open up a dialogue box with a form in which the datagroup details can be entered and saved.

Datatype approval

Once the datagroup has been saved the datatype approval form will be displayed, as illustrated in Figure 13.

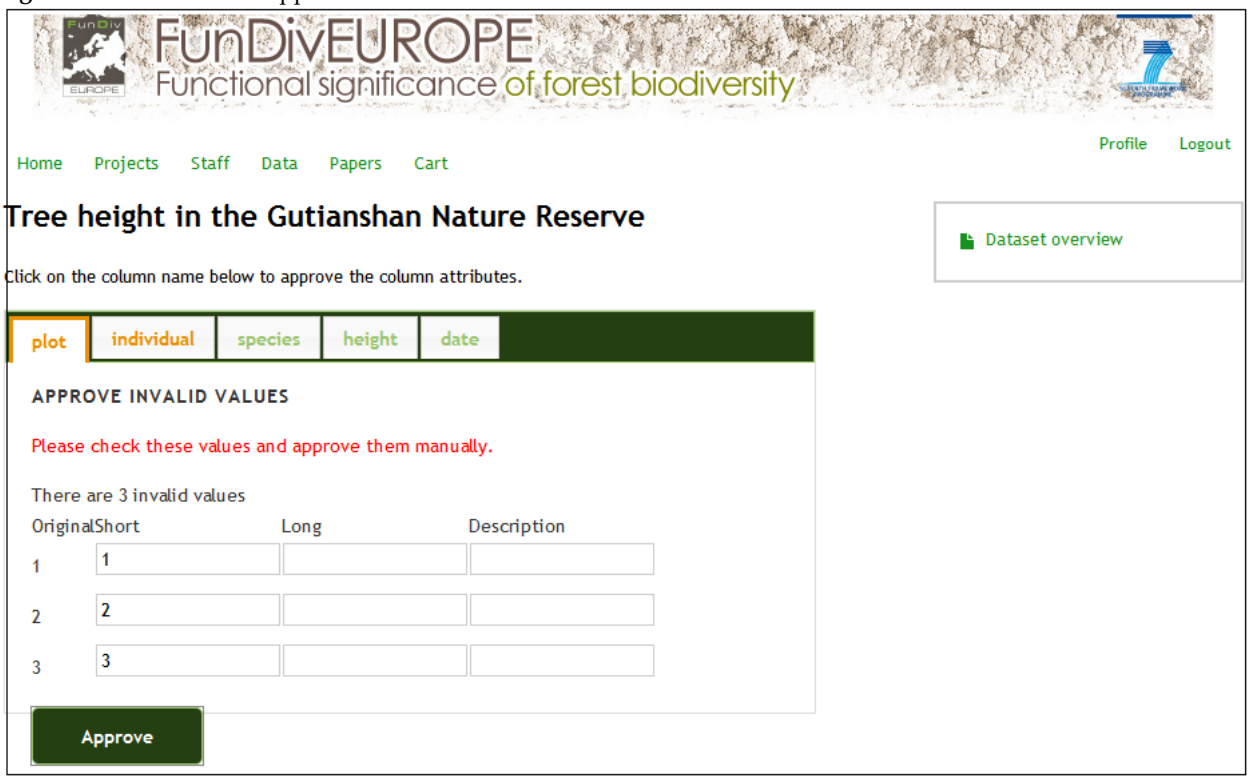
Figure 13: Datatype approval



The page displays the first 10 unqiue data values and a dropdown list of the five datatypes: text, date, year, number and category. Select the correct datatype and click on the '**Save datatype**' button.

The uploaded data for the column will be validated against the selected datatype. Any invalid values will be returned to the screen for the user to validate. An example of this is shown in Figure 14 using the category datatype where the data values did not exist in the database and they weren't included in the categories sheet of the workbook. Another example is for the number datatype where <100 was entered. This is not a valid number but may be correct in the context of the data.

Figure 14: Invalid values approval

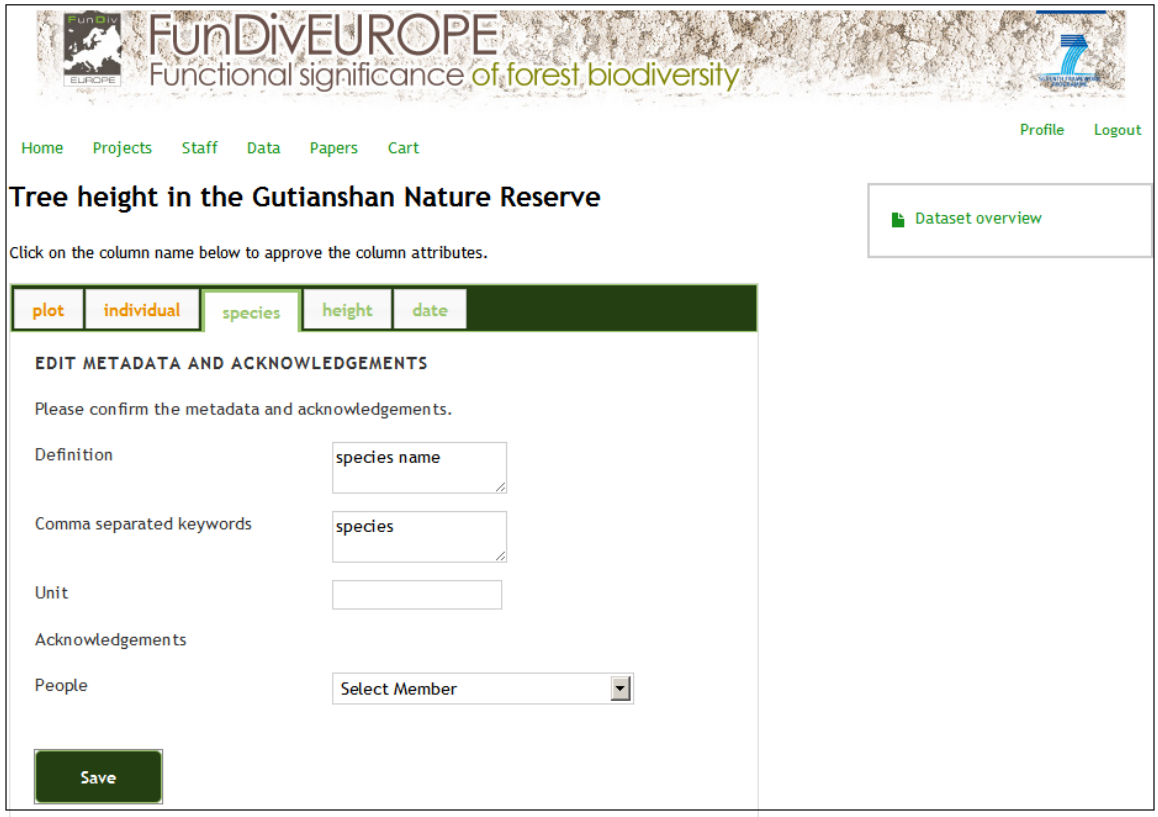


The user must then approve any invalid values and click on the '**Approve**' button. If the user doesn't provide a long name or a description for the category the short name will be used.

Column metadata and acknowledgements approval

The final page in the approval process is to confirm the metadata and acknowledgements (data provenance) for the column. Figure 15 illustrates an example of this page.

Figure 15: Saving column metadata and acknowledgements



It is possible to change the datagroup, datatype, metadata and acknowledgements of an approved column at any time.

Owners are able to edit their datasets. They can delete all the datacolumns and re-upload new values, however they are not able to completely delete the dataset. On re-uploading new data the user will have to go through the same approval process as described above.

5 Naming conventions

A central theme of the portal is naming conventions, the use of consistent names and codes for identical objects or bit of data. Consistent naming standards are essential when merging and sharing data.

Naming conventions are managed in the BEFdata portal by datagroups and categories.

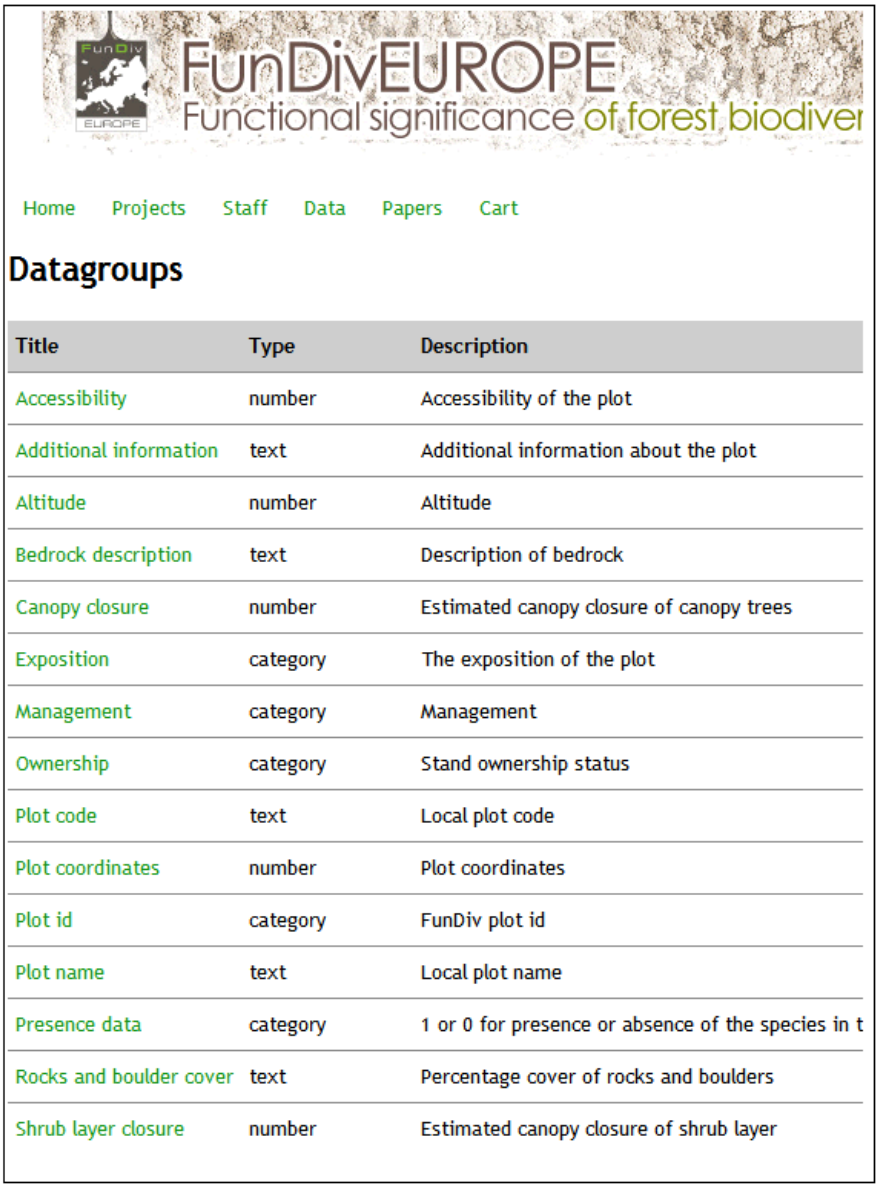
5.1 Datagroups

Datagroups are used for the semantic classification of the data. This enables similar data concepts to be grouped together within and between datasets and this improves the efficiency of data retrieval. The list of all datagroups in the portal can be found here:

http://fundiv.befdata.biow.uni-leipzig.de/datagroups, however, you must be logged in to view the page. Figure 16 illustrates the datagroups currently available on the portal. The listing includes the default datatype of the datagroup and a description. It is possible for a datagroup to have multiple datatypes (text, year, date, number, category).

As the project continues and more datasets are uploaded, the number of datagroups available will grow. Before completing the workbook it is recommended to check the datagroup list page to see if there is an existing datagroup that should be used.

Figure 16: Datagroup listing



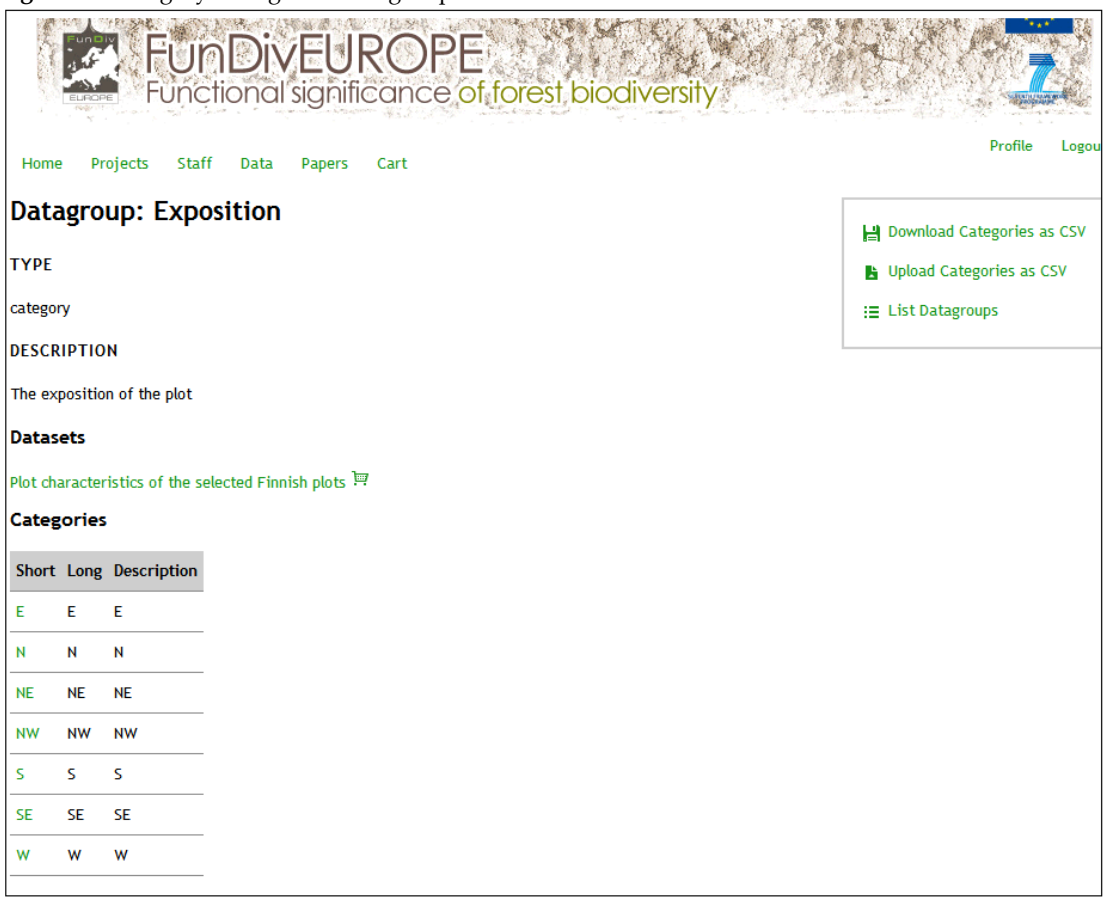
Title	Type	Description
Accessibility	number	Accessibility of the plot
Additional information	text	Additional information about the plot
Altitude	number	Altitude
Bedrock description	text	Description of bedrock
Canopy closure	number	Estimated canopy closure of canopy trees
Exposition	category	The exposition of the plot
Management	category	Management
Ownership	category	Stand ownership status
Plot code	text	Local plot code
Plot coordinates	number	Plot coordinates
Plot id	category	FunDiv plot id
Plot name	text	Local plot name
Presence data	category	1 or 0 for presence or absence of the species in t
Rocks and boulder cover	text	Percentage cover of rocks and boulders
Shrub layer closure	number	Estimated canopy closure of shrub layer

5.2 Categories

Categories enable identical objects to be referenced with the same name throughout all datasets. A good example of this is the use of codes for species names. It is essential that species have the same code throughout all the datasets in the project.

Clicking on the datagroup name displays the categories available under the datagroup. The portal ensures that category names are unique within each datagroup. Figure 17 shows an example of this page for the datagroup 'Exposition' which has 7 different category values, one for each of the points of a compass. From this page it is possible to download all the categories in the datagroup and upload new caetgories from a .csv file. The upload will error if there are duplicate categories in the spreadsheet you try to add categories that already exist in the datagroup.

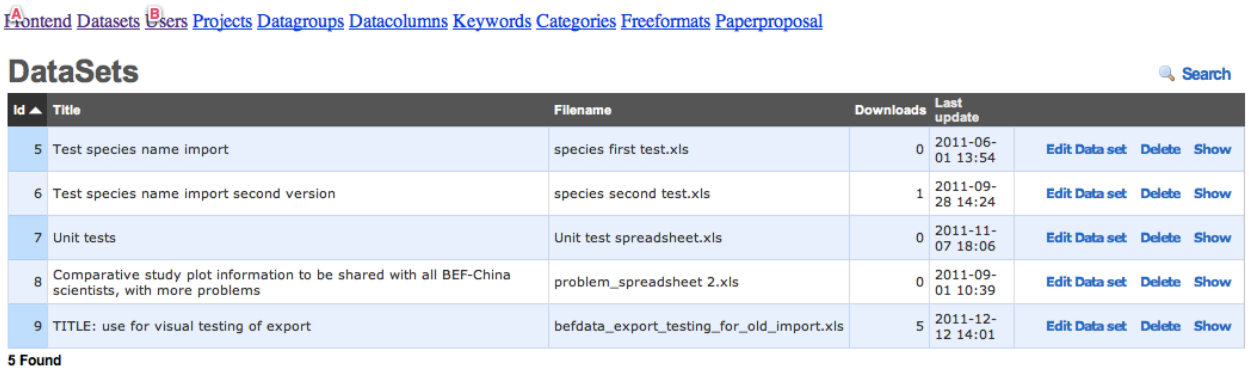
Figure 17: Category listing for a datagroup



6 The admin interface

Admins in the BEFdata portal can add new users and projects. Here we show how to add users and projects and link users to projects. To reach the admin interface, choose the "Admin" tab when logged in appropriately. Currently, on each fresh install of BEFdata, there will be an admin with the login "nadrowski" and the password "test".

The admin interface shows a table view on selected Ruby on Rails classes defined by BEFdata and corresponding to the database tables. The admin interface always starts by showing all datasets of the portal. With the leftmost link (A, Figure XX), you get back to the frontend. To add new users, you choose the "User" link (B, Figure XX).

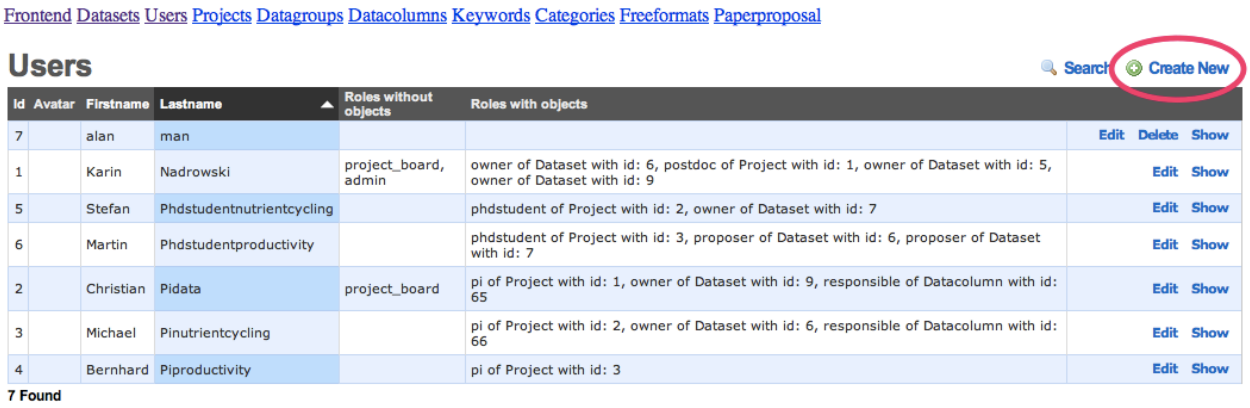


Id	Title	Filename	Downloads	Last update			
5	Test species name import	species first test.xls	0	2011-06-01 13:54	Edit Data set	Delete	Show
6	Test species name import second version	species second test.xls	1	2011-09-28 14:24	Edit Data set	Delete	Show
7	Unit tests	Unit test spreadsheet.xls	0	2011-11-07 18:06	Edit Data set	Delete	Show
8	Comparative study plot information to be shared with all BEF-China scientists, with more problems	problem_spreadsheet 2.xls	0	2011-09-01 10:39	Edit Data set	Delete	Show
9	TITLE: use for visual testing of export	befdata_export_testing_for_old_import.xls	5	2011-12-12 14:01	Edit Data set	Delete	Show

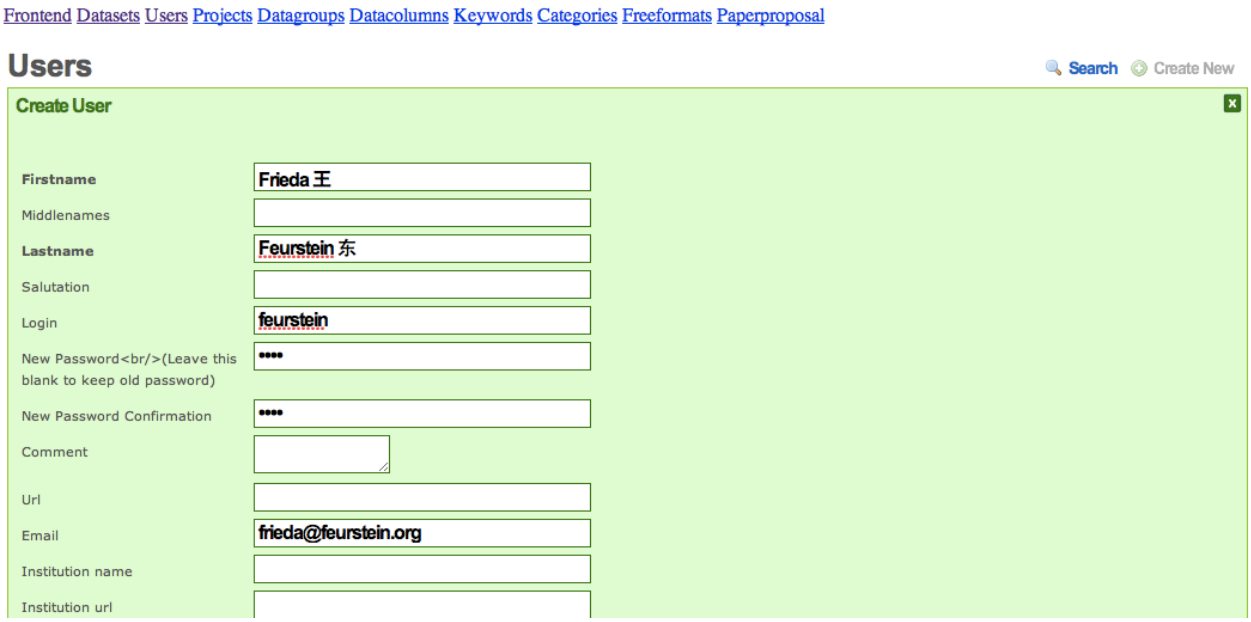
5 Found

6.1 Adding users

From the user list in the admin interface, users can be added by following the link in the far right corner of the screen.

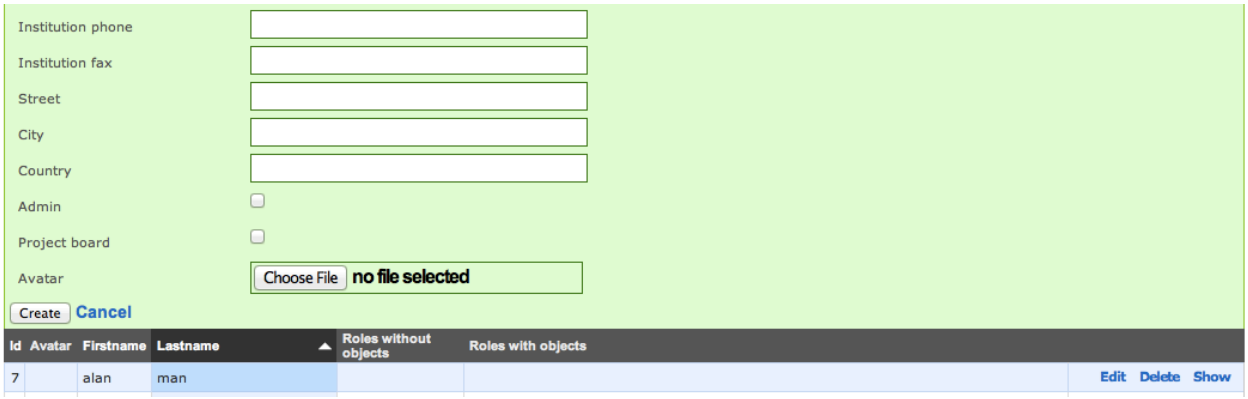


To create a new user, firstname, lastname, as well as e-mail address have to be provided. To enable the user to log on, he or she has to be given a login and a password. The password can later be changed by the user using the profile tab in the frontend.



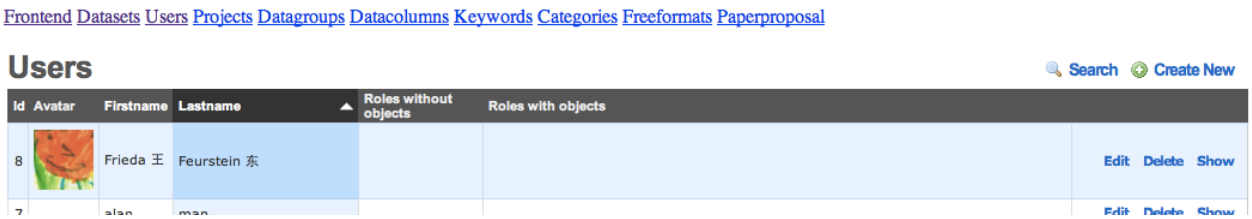
The user can be given an admin role to reach the admin interface of the portal, and / or the role of a project board member. Members of the project boards have responsibilities for the overall contents of the data and how it is used. They will be asked in the first round of every paper proposal, if the paper proposal is novel, if it competes with other paper proposals, or if further members of the research cooperation should be asked to take part in this analysis. They also can view the admin comments of a data-set, which can hold information on the metadata quality information on the dataset.

Although the user can herself provide a picture, the admin can also do this in the last line of the New-User dialogue.

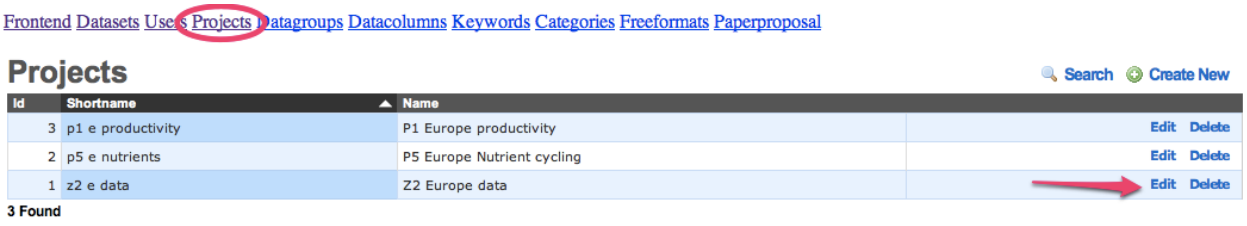


After choosing the "Create" button, the new user is online and the table in the admin backend also shows the avatar picture. The user now appears at the staff list, but is not linked to a project yet, which we will do next. For this we have to open the project, this user belongs to.

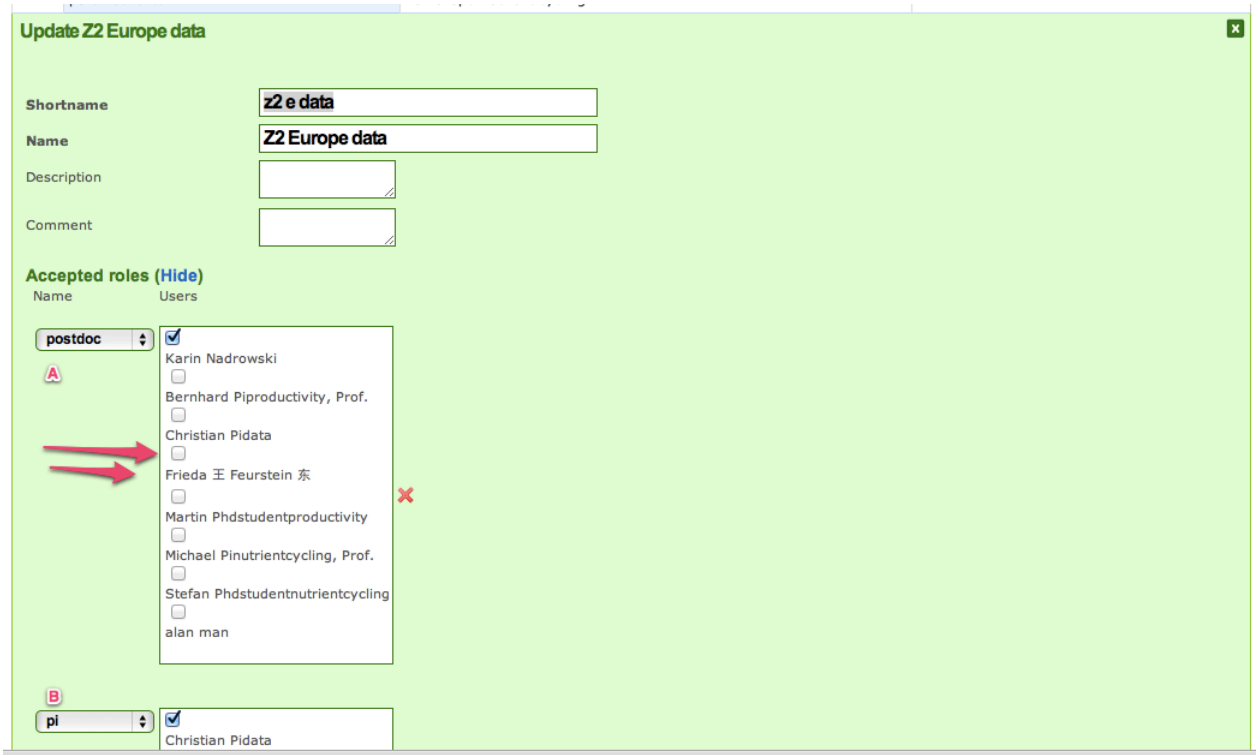
6.2 Adding a user to an existing project



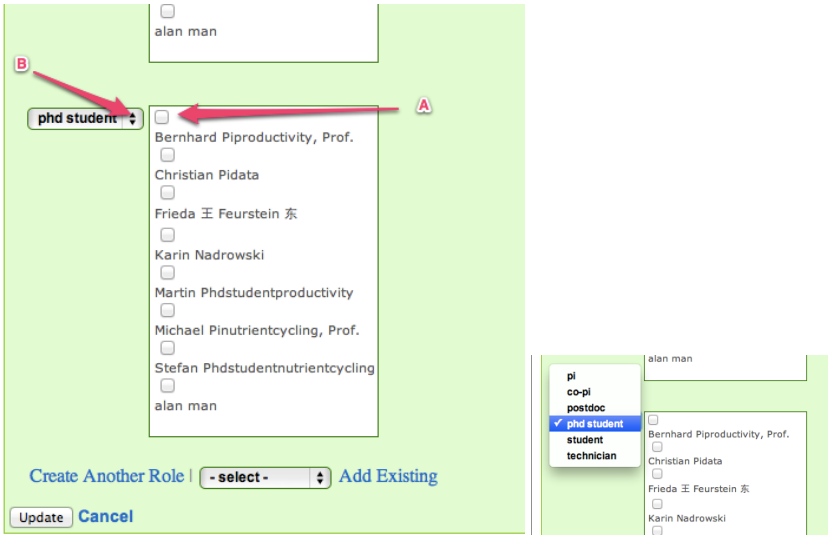
To add a user to an existing project, go to the "Projects" link on the top of the admin interface and choose the project. We will add Frieda to the "z2 e data" project.



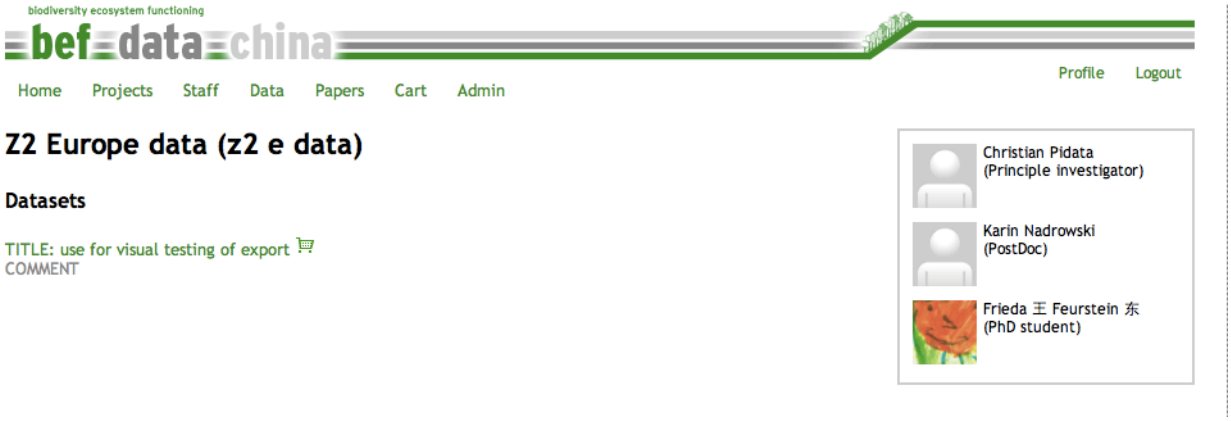
The dialogue for editing projects includes several lists of all portal users (A, B, Figure XXX following).



You see that Frieda is already choosable. For each role a user can have in a project, all users are listed. To add Frieda as postdoc to the "z2 e data" project, you have to choose the tick box above her name. If you would like to add her as a "phd student", a role that would be new to this project, you can choose the role in the dropdown list next to the last box of users on the project – edit page. You see that this is an unused list by the empty tick box at the top of that list (A). You can choose the appropriate role in the dropdown list (B), and then check the tick box above the user name.

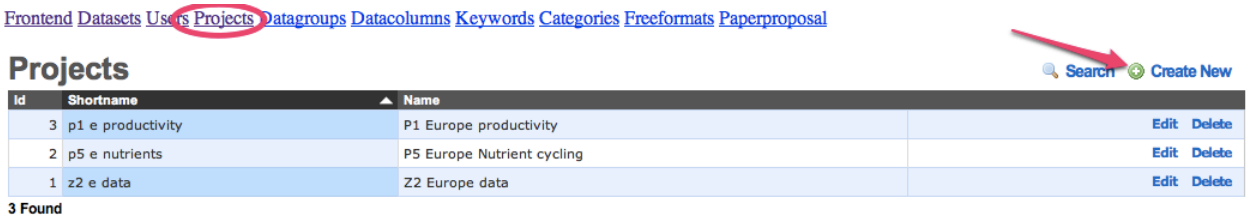


The project page for "z2 e data" now lists the new user as phd student in the frontend.



6.3 Adding a new project

Projects can be added through the admin interface after choosing the "Projects" link on the top of the admin interface. Instead of choosing "edit" for a given project as in the section above, you choose the create new button.



Projects only have few fields to fill out, which we do not show here. Mind that in the project list in the frontend projects are sorted by their title. Add users to projects just as described in the previous section. Mind that the users already have to be added to the portal to add them to a project.

7 Changing the general appearance: welcome page, header and footer

Although the majority of functions can be reused in every instance of the BEFdata portal, there are always differences from portal to portal. For example, the database has to be different for each portal instance. The database is configured in the /config folder of the application. Here the database.yaml holds the instructions for the database and the configuration.yaml holds the the name of the layout to be used as well as currently the default geographic boundary coordinates appended to an EML export of a dataset.

The layout itself is found under /app/views/layouts. There you find the templates not only for the BEF-China instance (befchina.html.haml), but also for the FunDivEUROPE instance (fundiv.html.haml). What is rendered in the end is the contents of the "application.html.haml", so that this has to be changed to give the portal a different appearance. The "application.html.haml" refers to partials in the /app/views/pages folder.

The configuration.yaml in the config folder holds the name of the of the custom layout, which will then be used to find the correct layout files. With this name (say "new_befdata_instance"), a folder must be created under views/pages.

There are presently two folders, one called "application", the other "fundiv". The application folder contains all the instructions for the default appearance. The "fundiv" folder examplifies, how this can be changed. You can create a new folder (in our case "new_befdata_instance") and copy the files from the existent ones to your new folder. You then can edit the files to your preferred appearance.

You can also change the "application" folder. But make sure to prevent this file from being overwritten when pulling code changes from the code repository.

```
~/befchina$ cd app/views/pages

~/befchina/app/views/pages$ ls

application     fundiv          home.html.haml     _menu.html.haml

data.html.haml  help.html.haml  imprint.html.haml  _search.html.haml

~/befchina/app/views/pages$ cd application/

~/befchina/app/views/pages/application$ ls

_footer.html.haml  _head.html.haml  _home.html.haml  _imprint.html.haml

~/befchina/app/views/pages/application$ less _footer.html.haml

~/befchina/app/views/pages/application$
```

Images are found in the /public/images folder.

```
~/befchina$ cd public/images
```

8 On-going and future development

This section details the on-going and next development projects for the BEFData portal development team. As these changes are implemented this document will be updated to reflect them.

8.1 Improve dataset upload and download performance

For larger datasets (>10,000) data upload and download can take some time. We are investigating the possibility of running both actions asynchronously so that after making the request to upload, or download data, the user can continue to use the site and they are notified when the action has completed.

8.2 Continued UI changes to improve usability

We are continually improving the usability of the portal as we receive feedback from members.

8.3 Site search

Currently the only way to search for datasets is to use the keywords or use the browser search on the data index page. It is planned to add a site-wide, full-text, search.

9 Acknowledgements

The BEFdata portal was developed within the Biodiversity-Ecosystem Functioning (BEF-China http://www.bef-china.de/) research unit of the German Science Foundation (FOR 891).

The BEFdata portal development team is:

Biology Institute I - Special Botany and Functional Biodiversity, University of Leipzig
Karin Nadrowski
Sophia Ratcliffe

Topic Maps Lab, Natural Language Processing Group, University of Leipzig
Michael Prilop
Karl Welter