BEXIS 2.12.1

Data Collection Module

User Guide

Authors

Nafiseh Navabpour, Roman Gerlach, David Schöne

Contact

Website: <http://bexis2.uni-jena.de>

Email: bexis-support@uni-jena.de

Phone: +49-(0)3641-948968

Acknowledgement

The development of the BEXIS 2 software would not be possible without the German Research Foundation (DFG) funding the BEXIS++ project. BEXIS++ is a collaboration of the Friedrich-Schiller-University Jena (Dept. of Computer Science, Dept. for Geography), the Max-Planck-Institute for Biogeochemistry Jena, the Technical University Munich (Dept. Ecology and Ecosystem Management), the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), and the Georg-August-University Göttingen.

Contents

[1. Overview 3](#_Toc526925008)

[2. Create Dataset 3](#_Toc526925009)

[2.1 Copy an existing Dataset 4](#_Toc526925010)

[2.2 Content 4](#_Toc526925011)

[2.3 Messages 4](#_Toc526925012)

[3. Upload Data 5](#_Toc526925013)

[3.1 Upload Tabular Data 5](#_Toc526925014)

[3.1.1 Select File 5](#_Toc526925015)

[3.1.2 Get File Information 6](#_Toc526925016)

[3.2 Specify Dataset 7](#_Toc526925017)

[3.3 Choose Update Method 8](#_Toc526925018)

[3.4 Validation 8](#_Toc526925019)

[3.5 Summary 9](#_Toc526925020)

[3.6 Upload Data 9](#_Toc526925021)

[4. Import Data 10](#_Toc526925022)

[4.1 Select File 10](#_Toc526925023)

[4.2 Metadata 11](#_Toc526925024)

[4.3 Select Areas 11](#_Toc526925025)

[4.4 Verification 12](#_Toc526925026)

[5. Push big files to server 13](#_Toc526925027)

[6. Manage Metadata Structure 13](#_Toc526925028)

[6.1 Select File 14](#_Toc526925029)

[6.2 Read Source 15](#_Toc526925030)

[6.3 Set Parameters 16](#_Toc526925031)

[6.4 Summary 17](#_Toc526925032)

# Overview

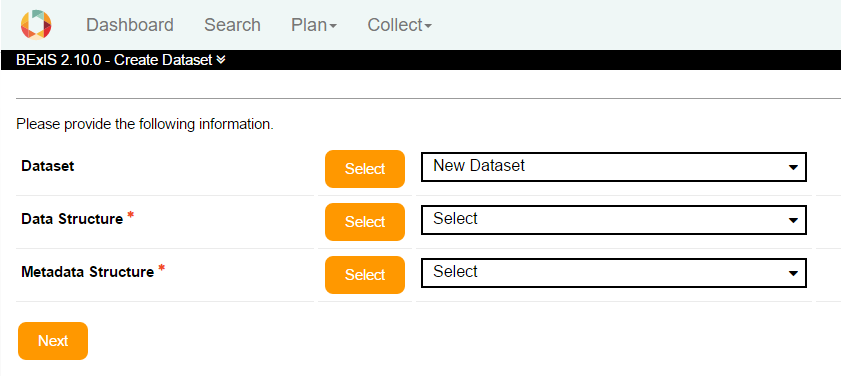
The Data Collection Module provides tools to create new datasets, enter metadata, upload data to the system, and import metadata structures (i.e. schemas). There are some workflows available under the **Collect** tab and in the **Setup**:

* Create Dataset
* Upload Data
* Push Big File
* Manage Metadata Structure
* Export Metadata

# Create Dataset

This wizard will assist you in creating a new dataset in BEXIS 2. The Wizard is very flexible and builds up differently depending on the selected Metadata structure. Therefore, we describe only the basic functions here.

The first step is to generate an *empty* or a *copy of an existing dataset* based on your selection of the two mandatory elements: data structure, and metadata structure.

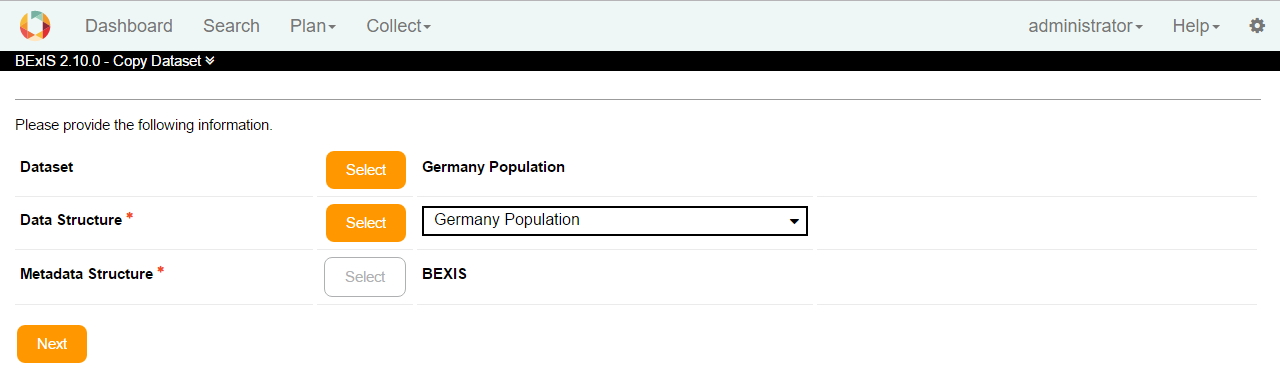


The next stage is determined by the selected metadata structure.

## Copy an existing Dataset

By using Select button, you can choose a Dataset to make a copy of it. Related to the Dataset, you can choose a Data Structure, but there is only one related Metadata Structure.

You are able to use predefined content or change fields as you want.



## Content

The content area is where you enter metadata describing your dataset. The forms provided here may look different and contain different attributes depending on the metadata schema (structure) you have chosen in the first step.

 Required attributes,

 Add an attribute,

 Remove an attribute,

 Change order of the attribute,

 Expand / Collapse

On the Bottom, there is a button titled Validate to examine whether required attributes have been filled and whether the information complies with the business logic. The validation may also be triggered by using the Submit.

You could edit a submitted dataset or make a copy of that by clicking on the Edit or Copy buttons.

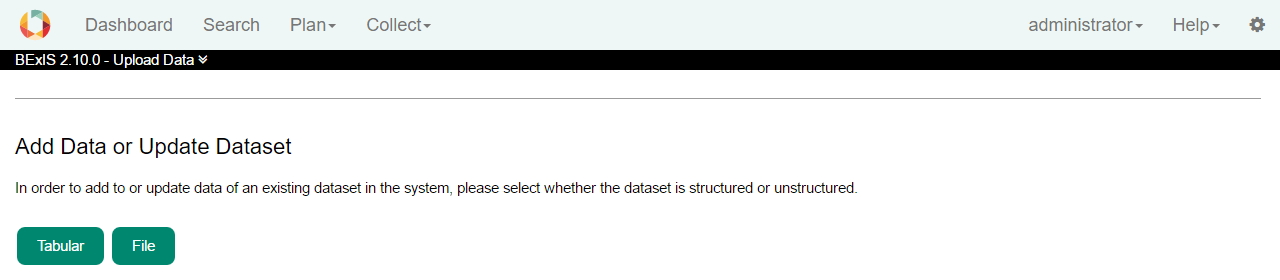
## Messages

When an input is faulty, the input field is highlighted in red. If you go with the mouse over the box, you get information about what is wrong.



# Upload Data

To upload your data, please go to the **Collect > Upload Data** via main menu. This wizard will assist you in uploading data into the BEXIS 2 repository. A dataset can be structured or unstructured (i.e tabular or file).

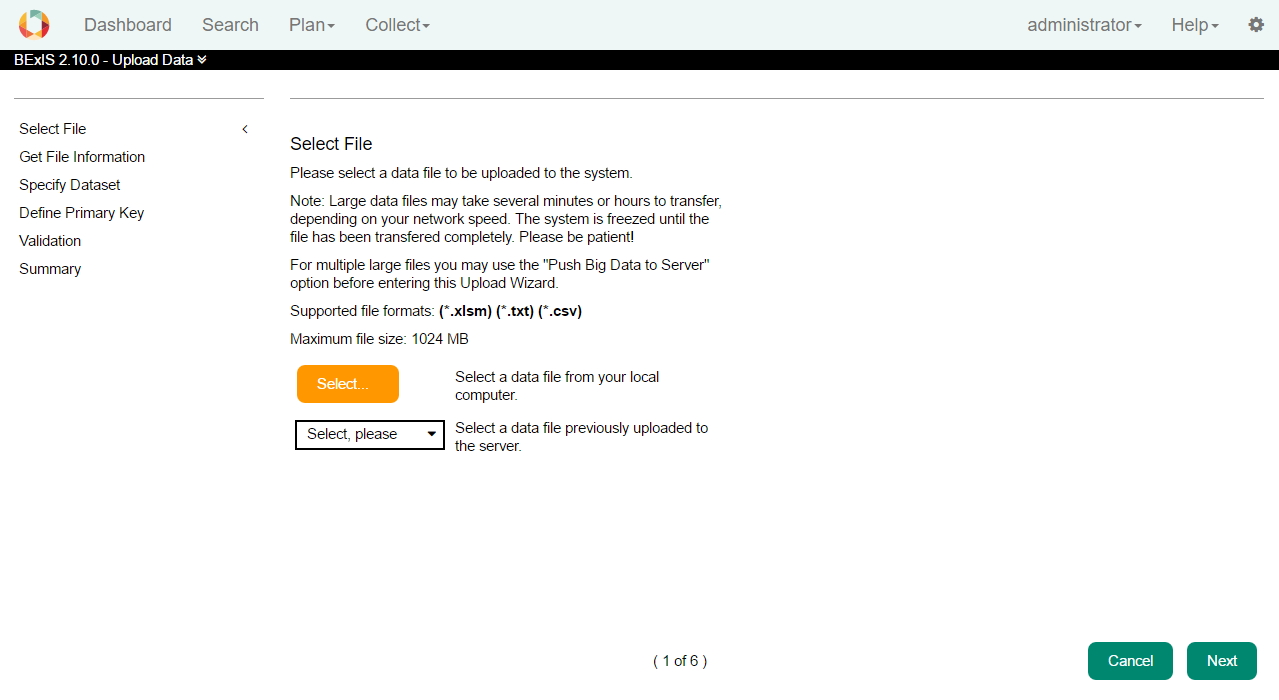


## Upload Tabular Data

The term “Tabular data” is used for all datasets where there internal structure of the data is “known” to the system. For example, in a data table the header, which defines the columns (i.e. variables), is the structure of the data. Before uploading/importing data to the system, the data structure needs to be created with the Data Structure Manager of the Data Planning module.

### Select File

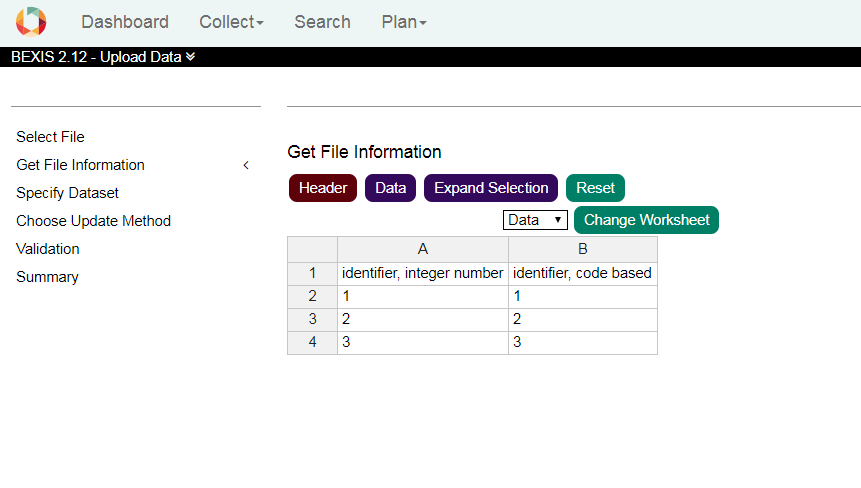
In the first step, an existing file containing your data needs to be selected. You can select either a file from your local computer or a file that has been uploaded to the server prior to starting the Upload Wizard. The second option is designed for files larger than 4 MB that may take several minutes to transfer. The wizard supports file formats of Microsoft Excel (\*.xlsm) or ASCII (\*.txt, \*.csv). Microsoft Excel files are required to use a template created with the Data Planning Module of BEXIS 2 (refer to Data Planning User Guide for more details). Once a file has been successfully selected, click the **Next** button and proceed to the next step.



### Get File Information

For all Microsoft Excel files using a BEXIS 2 template the file information and data, structure is automatically extracted and this step is omitted. Please refer to the Data Planning User Guide for more details on how to create such a template.

For all other **Excel files**, usersneed to provide the information by selection the data on the screen.



For all **ASCII files**, users need to provide information on the file structure and formatting.

First, please choose a **separator** that is being used to separate data values from each other in your ASCII file.

Depending on your language, different punctuation is used for decimal values. Please choose the one present in your ASCII file.

Next please specify whether the **orientation** of your data is column-wise or row-wise (see figure below).



Datasets may contain empty rows or columns on top or to the left before the header and the actual data values start. Please specify this offset in number of columns or rows.

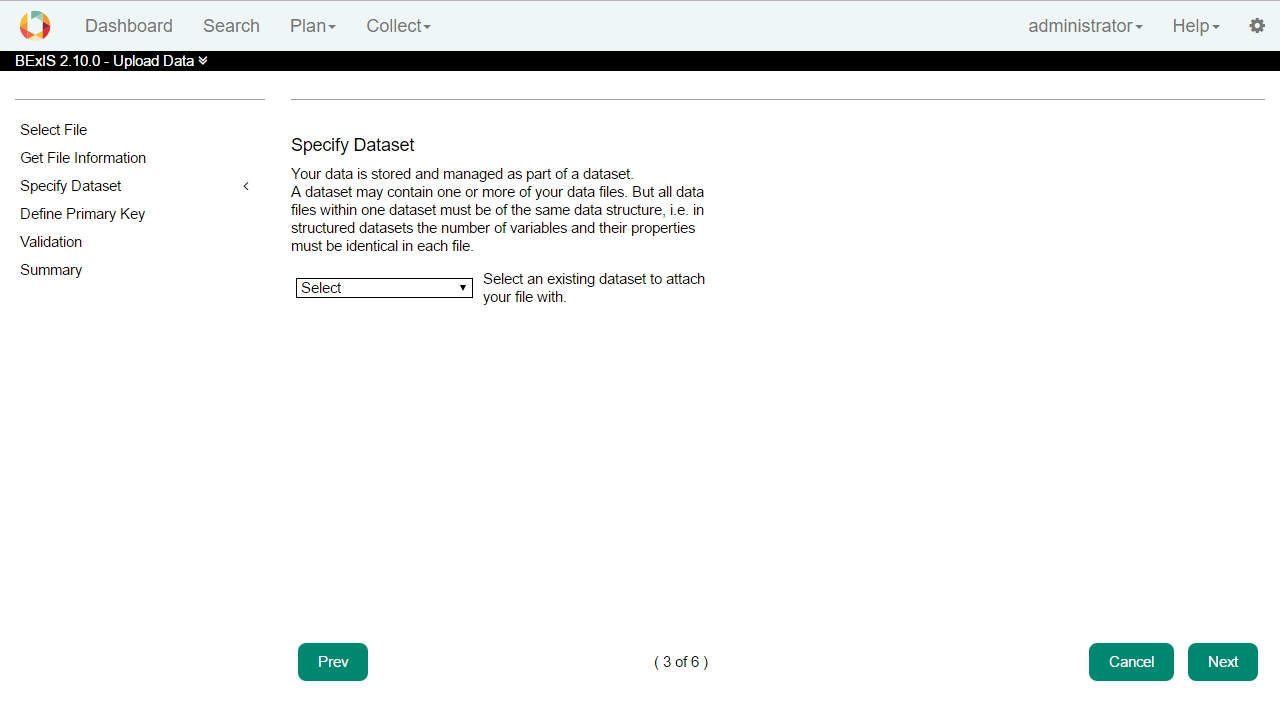
Further, your data file may contain a header defining variable names, types etc. The row/column where this header starts needs to be specified (see figure below).



Finally, the row/column where the actual data values start needs to be specified.

## Specify Dataset

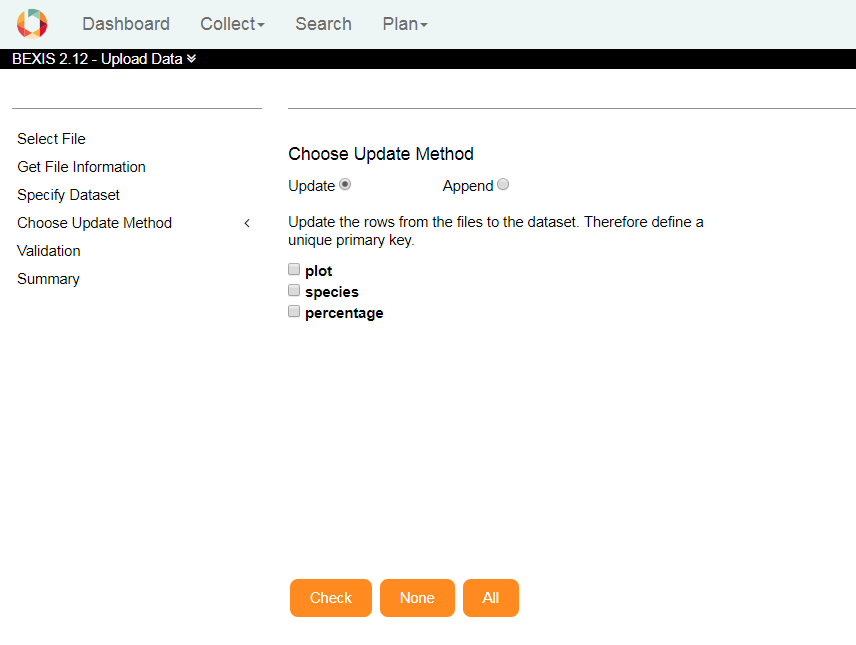
In BEXIS 2 your data is stored and managed as part of a dataset. A dataset may contain one or more of your data files. But all data files within one dataset must be of the same data structure, i.e. the number of variables and their properties must be identical in each file. To upload your data to the system, please select one existing dataset from the dropdown list.



## Choose Update Method

While adding data to an existing dataset you need to specify how you want to update.

By **Update** the user need to specify a unique identifier (e.g. primary key) for each tuple (i.e. row) in your dataset. If your dataset already contains a variable with such a key, please select it. Otherwise, a primary key can be created by combining available variables. Please click the **Check** button to verify whether the selected combination is unique. If you go back and change something in the process of uploading, you need to check the primary key again.



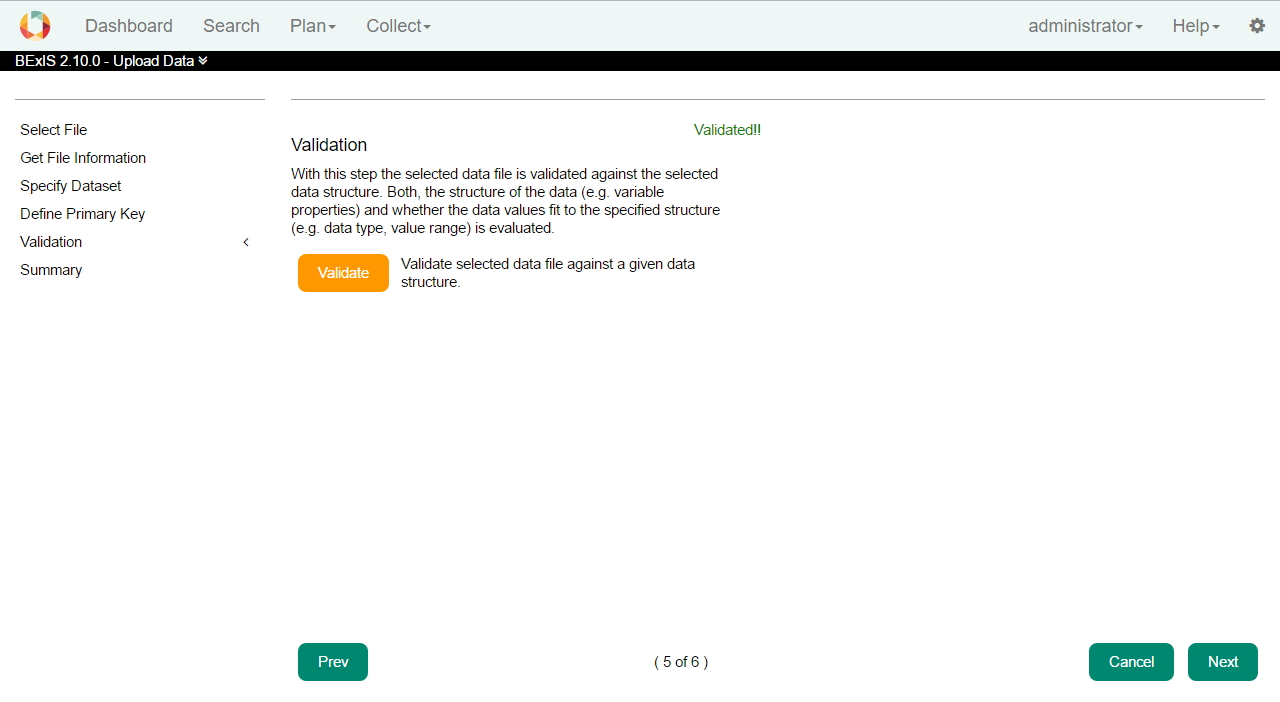
By **Append**, the lines are uploaded directly to the data without checking for duplication.

## Validation

With this step, the selected data file is validated against the selected data structure. Both, the structure of the data (e.g. variable properties) and whether the data values fit to the specified structure (e.g. data type, value range) is evaluated.

Click on **Validate** button to validate the data file.

If you go back and change something in the process of uploading, you need to validate the file again.



## Summary

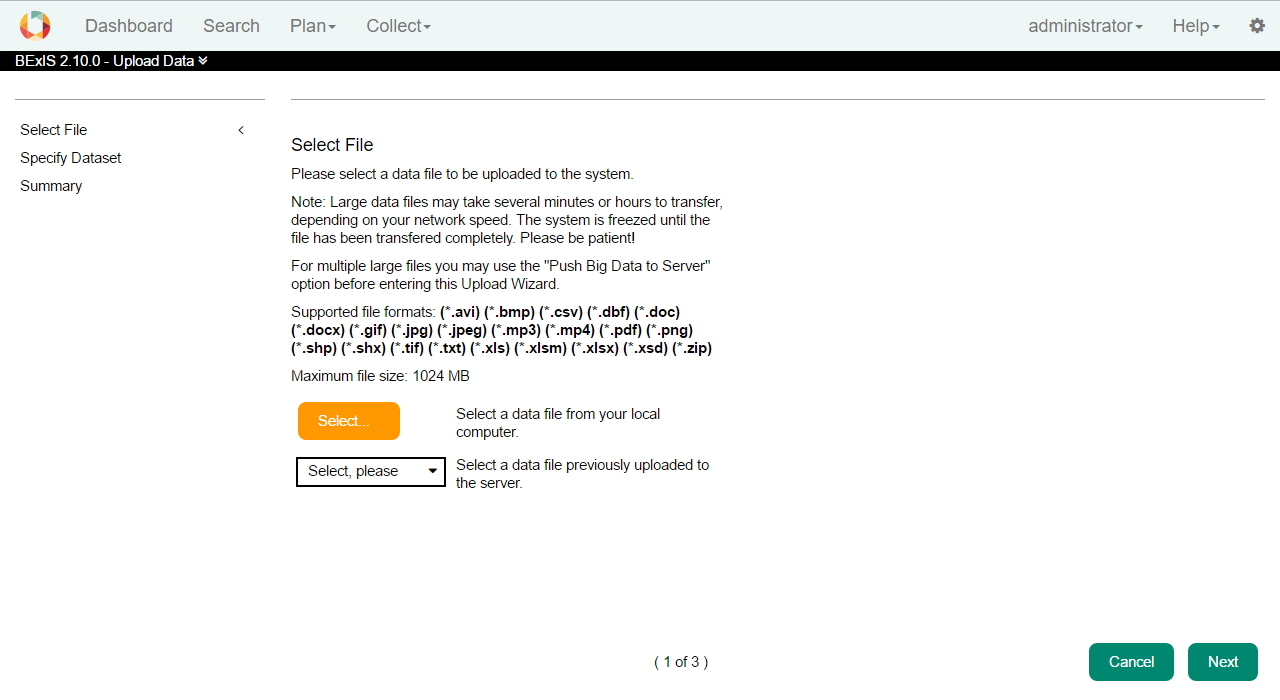
With this final step a summary of your uploaded data file is provided. Please check the information and click the **Finish** button to confirm and finalize the upload.

## Upload Data

An unstructured data could be either selected from your local computer or could be a file that has been uploaded to the server. In the case of unstructured data, we do not read the contents of the data. We copy the files to the server and place them in relation to the dataset.

BEXIS 2 application can support many file formats such as (\*.avi) (\*.bmp) (\*.csv) (\*.dbf) (\*.doc) (\*.docx) (\*.gif) (\*.jpg) (\*.jpeg) (\*.mp3) (\*.mp4) (\*.pdf) (\*.png) (\*.shp) (\*.shx) (\*.tif) (\*.txt) (\*.xls) (\*.xlsm) (\*.xlsx) (\*.xsd) (\*.zip).

The Maximum acceptable file size up to now is: 1024 MB.

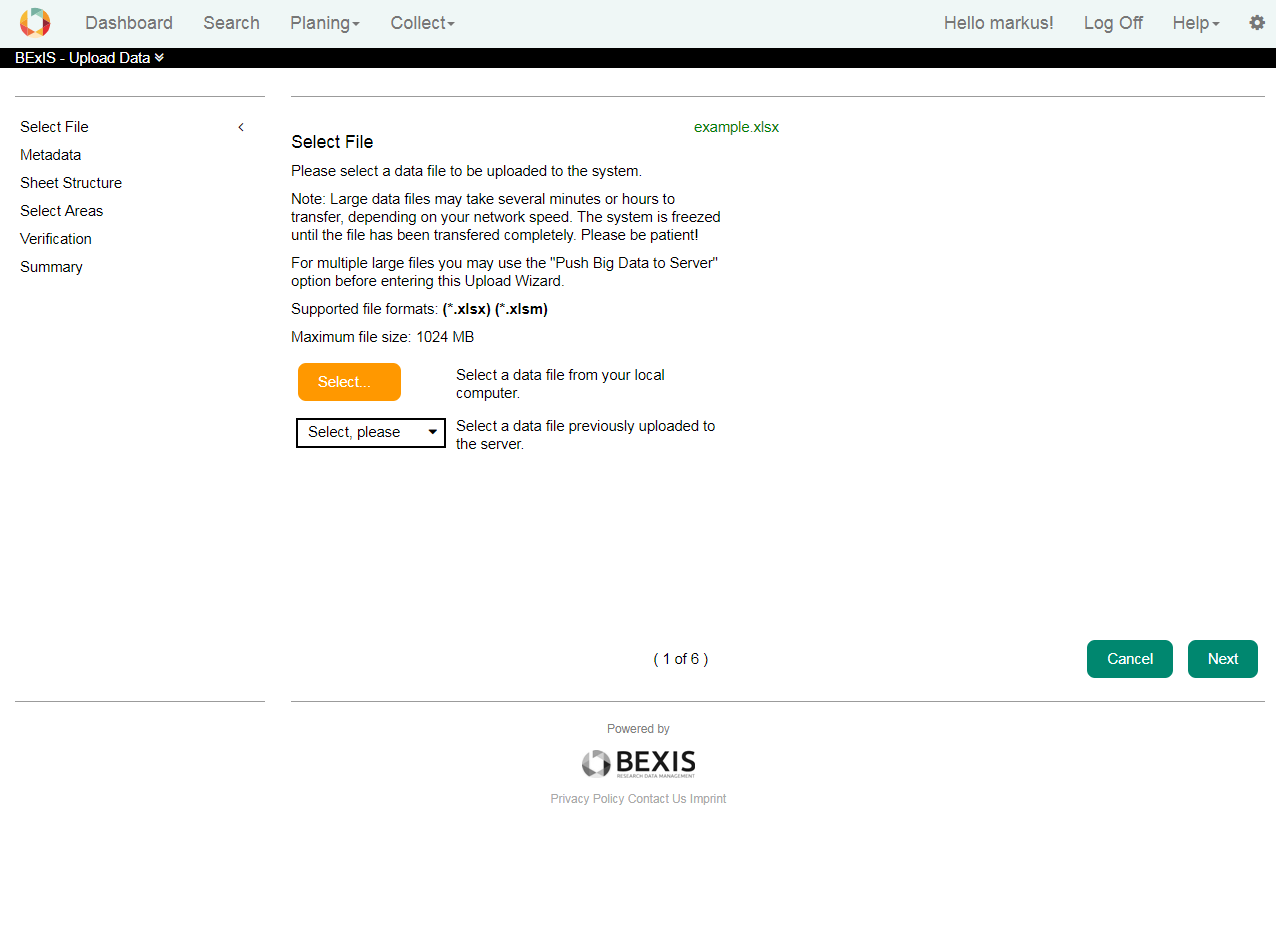


# Import Data

The Import Data wizard enables you to import both a tabular data structure and data in a single workflow. Import a file follows the following steps.

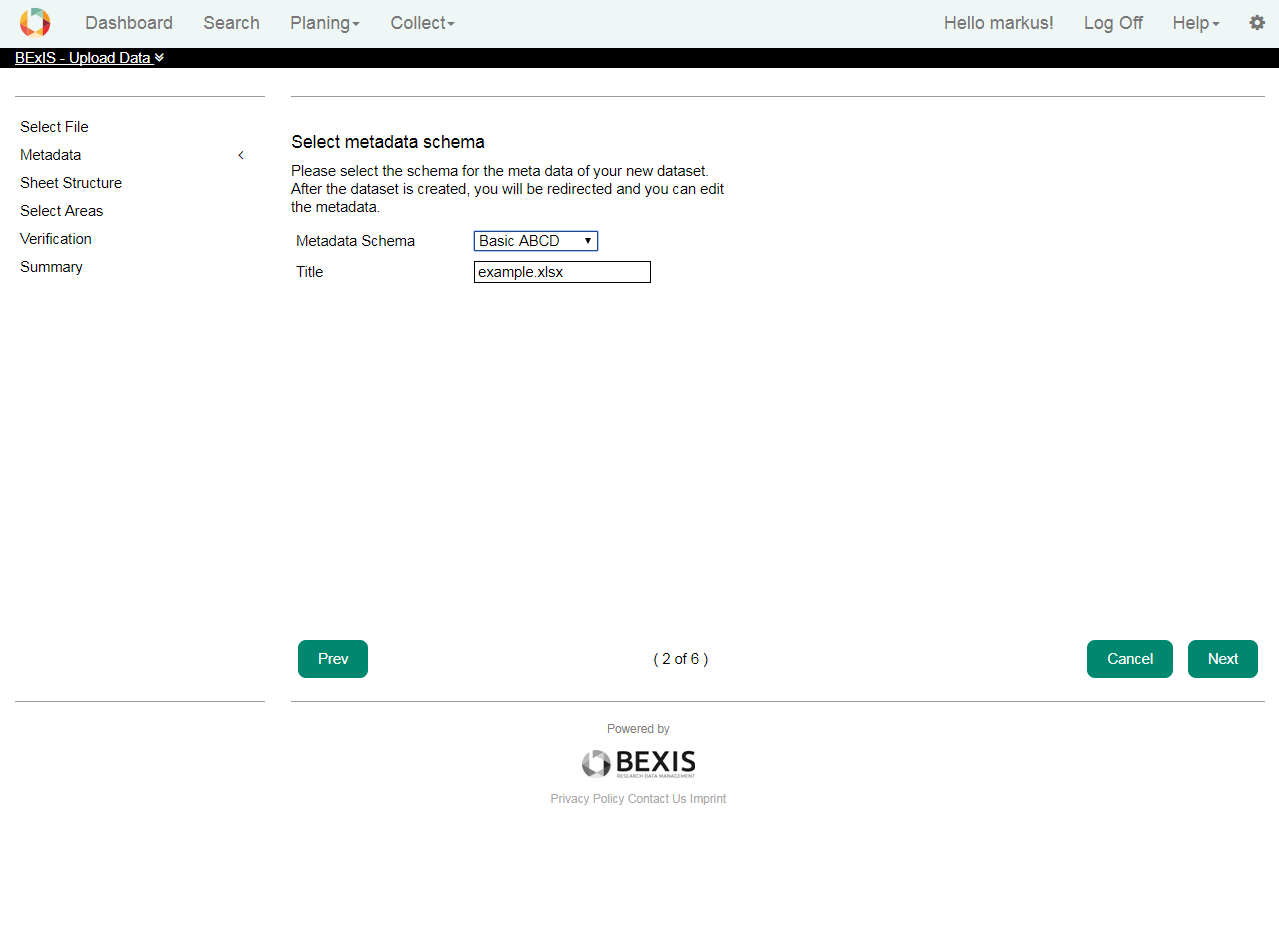
## Select File

In the first step an existing file containing both your data and the names of the variables needs to be selected. You can either select a file from your local computer or a file that has been uploaded to the server prior to starting the Upload Wizard. The second option is designed for files larger than 4 MB that may take several minutes to transfer. The wizard supports file formats of Microsoft Excel (\*.xlsm, \*.xlsx). Once a file has been successfully selected, click the Next button to proceed to the next step.



## Metadata

This step allows you to select the metadata schema that you would like to use for the dataset. There is no need to enter any metadata yet - after the last step you will be redirected to a page that allows you to enter the metadata. You can, however, change the title of the dataset. If you don't wish to change it, it will default to the name of the file you are using. Once you have selected a schema, click the Next button to proceed to the next step.



## Select Areas

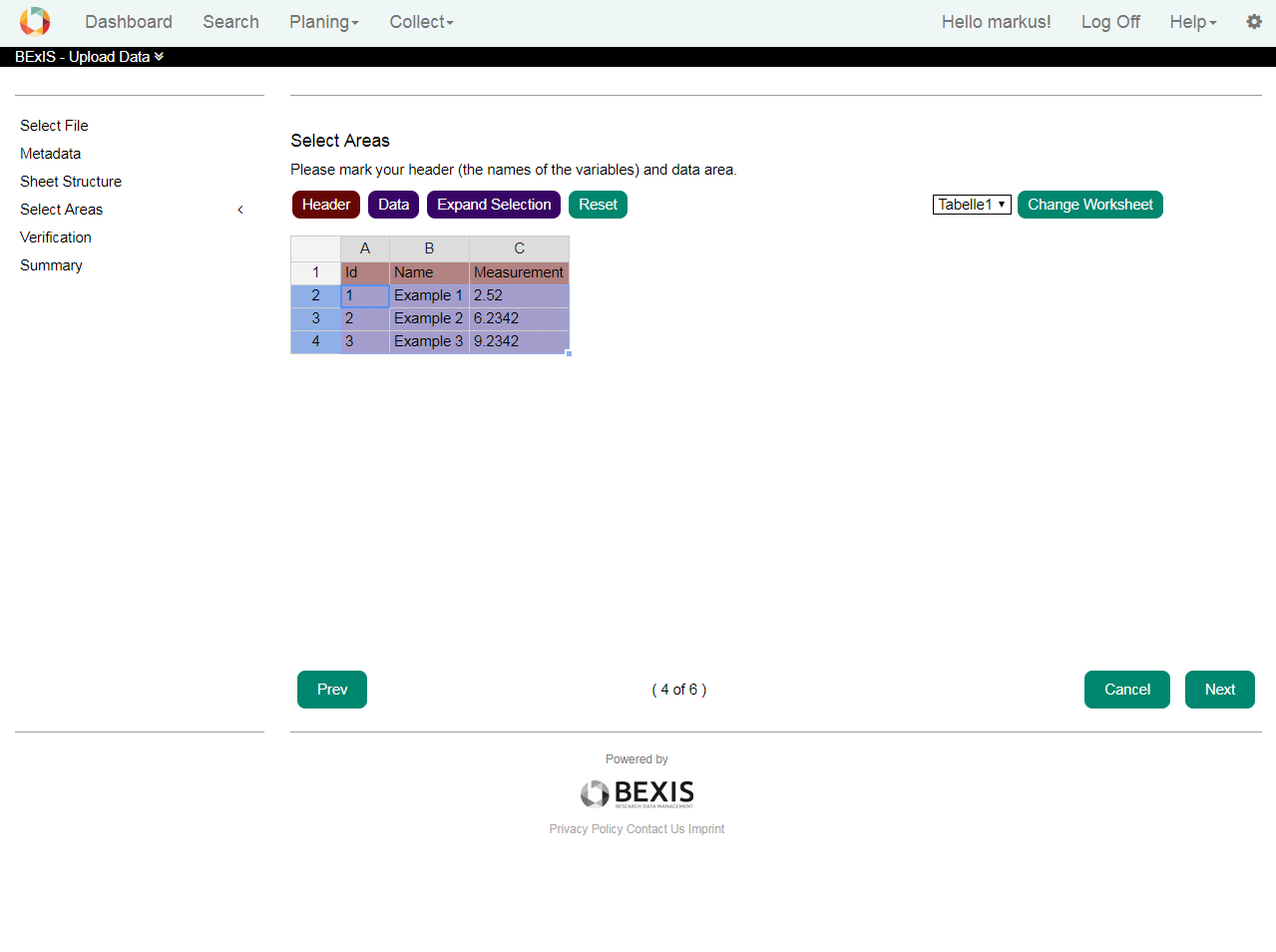
You will see a table that represents the file that you selected during the first step. Here you can select, which part of the file should be used as variables and which parts should be interpreted as data.

For the variables, select a row or one part of a row and click the "Header" button. The selected area should now be highlighted in red. If you are using a BEXIS2 template file there is no need to select the whole header with unit, datatype, etc. Just select the names of the variables.

For the data, you can select multiple rows and click the "Data" button. The selected area should now be highlighted in blue. Please make sure that the data area contains as many columns as the header area. You can skip one or more rows by selecting the area above them, mark it as data, and then select the area below them and mark it as well. You can use the "Expand Selection" button to expand the last data area you selected to the last row of the file. This can be very helpful when working with large datasets.

If you made any mistakes during the selection process just use the "Reset" button to remove all markings and start over.

If you wish to upload data from a different worksheet you can select it from the dropdown-menu and click the "Change Worksheet" button. Please be aware that you can only upload data from one sheet at a time. Once you have selected the header and at least one data area, click the Next Button to proceed to the next step.

****

The representation of the data you will see in the table might differ from what you see in Microsoft Excel or similar programs.

Don't worry, your data will be uploaded correctly if you choose the correct datatypes during the next step.

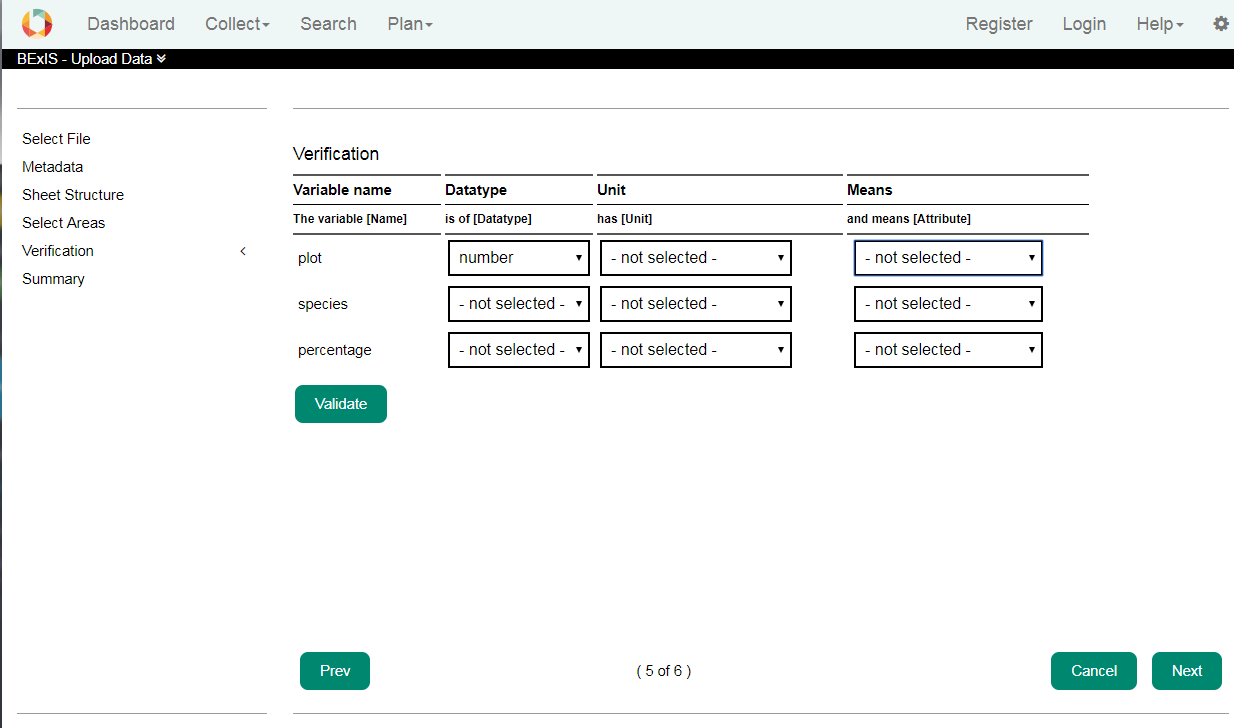
The following differences are known:

* Dates and times
  + Microsoft Excel users: Dates and times will be displayed as full timestamps, containing both a time and a date.
  + Libre Office users: Dates and times will be displayed as real numbers.
* Boolean values
  + Libre Office users: Boolean values might be displayed as "0" and "1".
* Real Numbers
  + Real numbers might be displayed with scientific notation.

## Verification

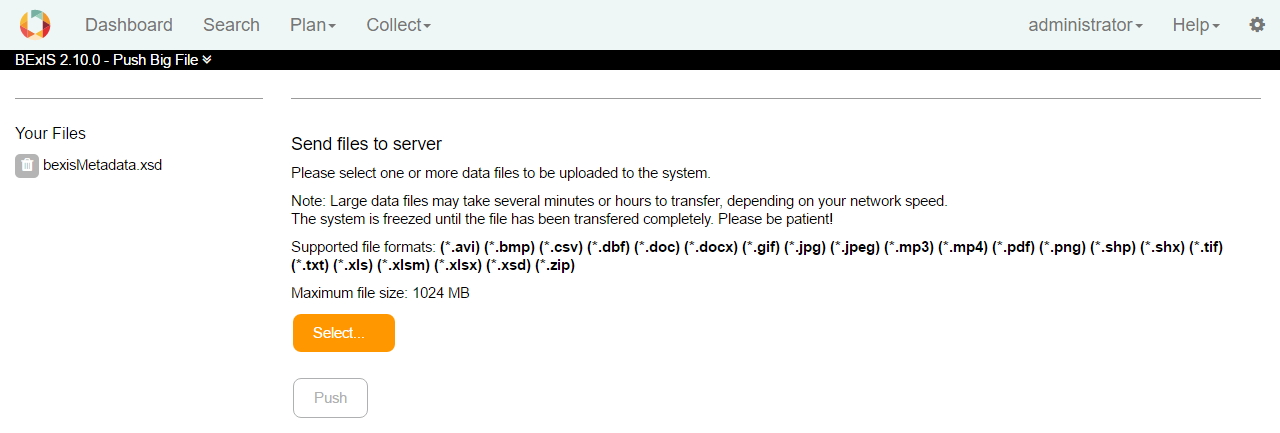
With this step you can define which units and datatypes your variables are using. The first column of dropdown-menus provides suggestions for attributes that are being used in other datasets and are similar to your variables. If you select one of the suggestions, unit and datatype are automatically adjusted. If you don't wish to use the suggestions, feel free to choose unit and datatype yourself.

The "Validate" button allows you to check if the datatypes you selected are suitable for the data you selected in the previous step. This allows you to recognize errors early and correct your selection. Once you've selected your units and datatypes, click the Next button to proceed.



# Push big files to server

Each user has a personal folder on the server where files are stored temporary. On this page, you can see the uploaded files. You can delete each file by clicking on the X or use these files later, when you want to upload data to a dataset.



# Manage Metadata Structure

Metadata structures (also called schemas or profiles) are typically created and imported by a data manager or administrator of the system. Thus this import function is available underthe **Setup > Manage metadata Structure**. The wizard will assist you in importing your metadata structure into the BEXIS 2. A metadata structure must be defined in a XSD schema file.

When importing a metadata schema into BEXIS 2, each element of the XSD file(s) is analyzed for its type, name, annotations, attributes, data types, constraints etc. Based on this information a form is automatically being created. For example, if an element is of data type Date, a date picker UI component will be used in the form. Also all names and descriptions are used exactly as they are in the XSD file(s).

NOTE: There are metadata standards available for almost any domain or type of data. It is good practice to follow one of them in order to ensure interoperability later on. However, although technical possible, most standards are very complex and should not be used as a whole. Users would just be overwhelmed and may need only a small selection of elements to describe their data. Standards are designed to cover a great range of use cases and data managers (in collaboration with their community) should make the effort in defining a set of feasible metadata elements in a profile (XSD file).

IMPORTANT: Please check, whether the XSD schema files have any dependencies to other files. You can find the dependencies in the **import** or **include** tags.



The current BEXIS 2 system requires all referenced files to be locally available on the server (no URL to external resource). Therefore, you may need to store all references first to a local folder, change the schema location path in every file (e.g. ./fileName.xsd) and then upload all files to the server. (**See 4**. Push big files to server).

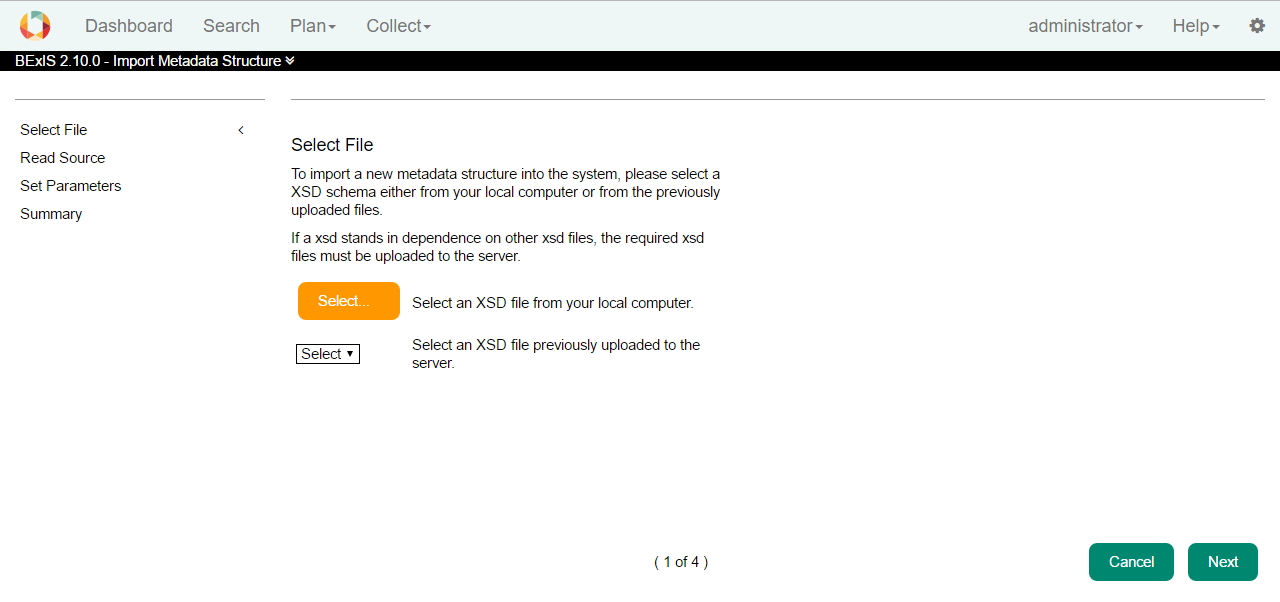




## Select File

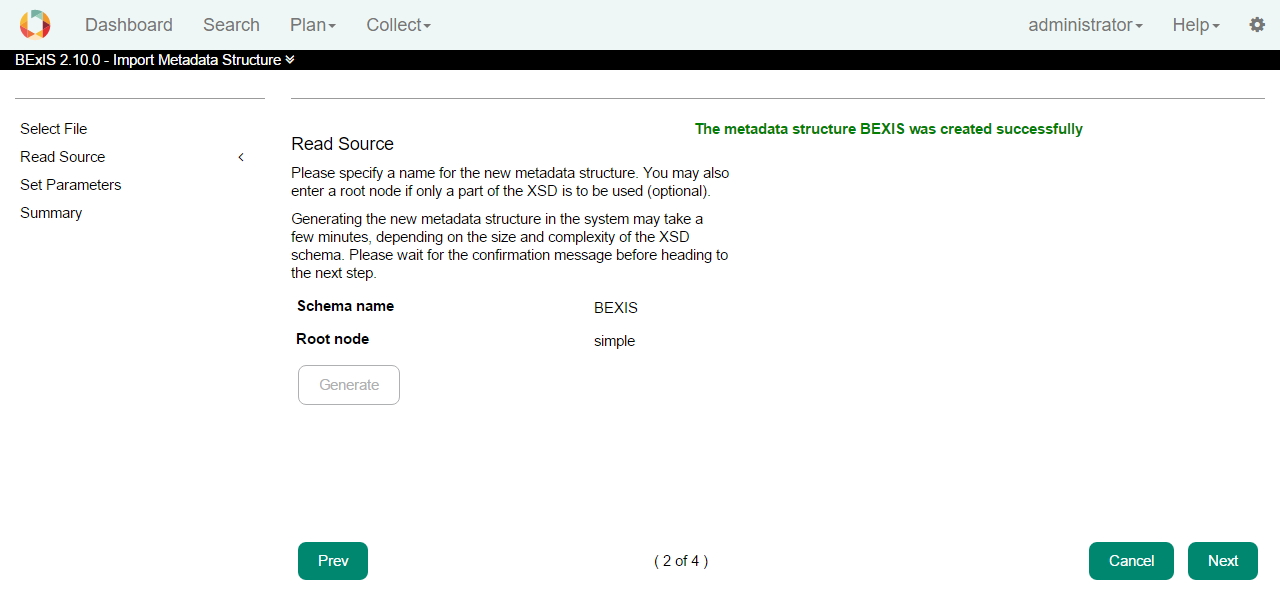
In the first step, an existing file containing your data needs to be selected. You can either select a XSD file from your local computer or a file that has been uploaded to the server prior to starting the Wizard. You may use the “Push big data to server” function in the Collect menu to upload multiple related XSD files.

Note: Please upload a valid XSD structure. BEXIS 2 does not check this kind of validation.



## Read Source

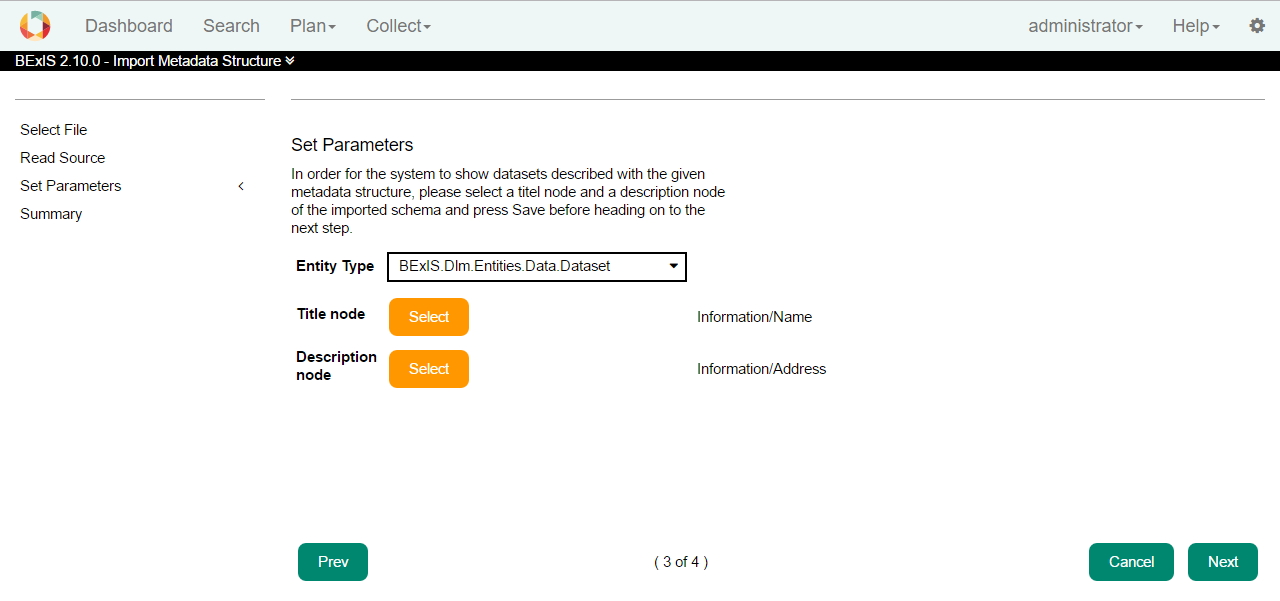
Please specify a name (i.e. display name) for the new metadata structure. You may also enter a root node if only a part of the XSD is to be used (optional).

  
  
To find the root node open the XSD Schema file and have a look on the element tags. In the example of ABCD it looks like this.

  
  
If no root node is selected then the wizard will automatically select the first element, which is a complex type. However, it is also possible to define the element “DataSet” as root node and the metadata structure starts from this element. The Name of a metadata structure must be unique and the root node must exist.

## Set Parameters

For the system to handle a dataset at least the title and a description is needed. In this step these two elements, which are typically available in all metadata structures, should be identified and made explicit to the system.



## Summary

The Summary page is an overview about the created metadata structure.  
  
