

BEXIS 2.8.0

Data Collection Module

User Guide

Authors

Nafiseh Navabpour, Roman Gerlach, David Blaa

Contact

Website: <http://fusion.cs.uni-jena.de/bexis>

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, Germany (Dept. of Computer Science, Dept. for Geography, Dept. of Ecology) and the Max-Planck-Institute for Biogeochemistry Jena, Germany.

Contents

- 1. Overview..... 3
- 2. Create a new dataset..... 3
 - 2.1 Content..... 4
 - 2.2 Messages 5
- 3. Upload Data..... 5
 - 3.1. Structured Data 5
 - 3.1.1. Select File..... 5
 - 3.1.2. Get File Information 6
 - 3.2. Specify Dataset 7
 - 3.3. Define Primary Key 7
 - 3.4. Validation 8
 - 3.5. Summary..... 8
 - 3.6. Unstructured Data..... 9
- 4. Push big files to server 9
- 5. Import metadata structure 10
 - 5.1. Select File..... 11
 - 5.2. Read Source..... 11
 - 5.3. Set Parameters 12
 - 5.4. Summary..... 13

1. 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 three workflows available under the **Collect** tab and one in **Admin** tab:

- Create a dataset (Collect)
- Upload data to a dataset (Collect)
- Push big files to server (Collect)
- Import metadata structure (Admin)

2. Create a new dataset

This wizard will assist you in creating a new dataset in BEXIS. 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 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.

Dashboard Search Plan Collect Administrator Help

BEXIS 2.8 - Create Dataset

* TaxonCoverage Select, please

Family

* ScientificName

Subspecies

LifeScienceIdentifier

VoucherSpecimen (1)

Remarks

SequenceLevelMetadata

* NumberOfSamples 0

TypeOfMoleculesExtracted

PermitNumber

Samples

SequenceProtocols

SequenceDataProcessingInfo

Remarks

Validate Submit

2.1 Content

The content area in the center 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.

VoucherSpecimen (1) ✖



* TaxonCoverage	<input type="text" value="Select, please"/>	* NumberOfSamples	<input type="text" value="0"/>
Family	<input type="text"/>	TypeOfMoleculesExtracted	<input type="text" value="Type 1"/>
* ScientificName	<input type="text"/>	TypeOfMoleculesExtracted 2	<input type="text" value="Type 2"/>
Subspecies	<input type="text"/>	PermitNumber	<input type="text"/>
LifeScienceIdentifier	<input type="text"/>		

- * Required attributes
- + Add an attribute.
- Remove an attribute.
- v ^ Change order of the attribute.
- v ^ activate / deactivate
- v ^ 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.

Validate

Submit

2.2 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.

+ - ∨ ∧

* SourceInstitutionID

* SourceID

* UnitID

Validate Submit

3. Upload Data

To upload your data, please go to the **Update data to a dataset** tab under **Collect** tab. This wizard will assist you in uploading data into the BEXIS repository. A dataset can be structured or unstructured.

The screenshot shows the BEXIS 2.8 - Upload Data wizard. The navigation bar includes Dashboard, Search, Plan, and Collect. The breadcrumb trail is BEXIS 2.8 - Upload Data. The main heading is "Add Data or Update Dataset". Below this, a message states: "In order to add to or update data of an existing dataset in the system, please select whether the dataset is structured or unstructured." Two buttons are visible: "Structured Data" and "Unstructured Data".

3.1. Structured Data

The term “structured data” is used for all datasets where their 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.

3.1.1. Select File

In the first step an existing file containing your data 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) or ASCII (*.txt, *.csv). Microsoft Excel files are required to use a template created with the Data Planning Module (Plan) 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.

- Select File
- Get File Information
- Specify Dataset
- Define Primary Key
- Validation
- Summary

Select File

Please select a data file to be uploaded to the system.

Note: Large data files may take several minutes or hours to transfer, depending on your network speed. The system is frozen until the file has been transferred completely. Please be patient!

For multiple large files you may use the "Push Big Data to Server" option before entering this Upload Wizard.

Supported file formats: (*.xslm) (*.txt) (*.csv)

Maximum file size: 1024 MB

Select a data file from your local computer.

Select a data file previously uploaded to the server.

3.1.2. 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 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).

Column – Wise

date	character	integerNumber	character
10.10.2010	Tree 1	125	wet
11.10.2010	Tree 2	123	wet
12.10.2010	Tree 3	105	dry
13.10.2010	Tree 4	95	dry

Row – Wise

date	sampleName	Tree 1	Tree 2	Tree 3	Tree 4
10.10.2010	Tree 1	125	123	105	95
11.10.2010	Tree 2	wet	wet	dry	dry

Data sets 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).

Offset 1 – one empty row

Variables start in row 3

	B	C	D	E	F	G		A	B	C	D	E	F	G
1								1						
2		date	character	integerNumber	character			2						
3		date	sampleName	gram	dry/wet			3						
4		10.10.2010	Tree 1	125	wet			4						
5		11.10.2010	Tree 2	123	wet			5						
6		12.10.2010	Tree 3	105	dry			6						
7		13.10.2010	Tree 4	95	dry			7						
8								8						
9								9						
10								10						

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

3.2. Specify Dataset

In BEXIS 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 of the existing dataset from the dropdown list.

Dashboard
Search
Plan ▾
Collect ▾

BEXIS 2.8 - Upload Data ▾

Select File

Get File Information

Specify Dataset <

Define Primary Key

Validation

Summary

Specify Dataset

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. in structured datasets the number of variables and their properties must be identical in each file.

Select ▾

Select an existing dataset to attach your file with.

3.3. Define Primary Key

While adding data to an existing dataset you need to specify a unique identifier (e.g. primary key) for 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.

The screenshot shows the BEXIS 2.8 - Upload Data interface. At the top, there is a navigation bar with 'Dashboard', 'Search', 'Plan', and 'Collect'. Below this is a header 'BEXIS 2.8 - Upload Data'. The main content area is divided into two columns. The left column contains a vertical list of steps: 'Select File', 'Get File Information', 'Specify Dataset', 'Define Primary Key', 'Validation', and 'Summary'. The right column is titled 'Define Primary Key' and contains three checkboxes: 'Species', 'Count', and 'Plot'. Below these checkboxes is an orange 'Check' button with an ellipsis to its right.

3.4. 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.

The screenshot shows the BEXIS 2.8 - Upload Data interface. At the top, there is a navigation bar with 'Dashboard', 'Search', 'Plan', and 'Collect'. Below this is a header 'BEXIS 2.8 - Upload Data'. The main content area is divided into two columns. The left column contains a vertical list of steps: 'Select File', 'Get File Information', 'Specify Dataset', 'Define Primary Key', 'Validation', and 'Summary'. The right column is titled 'Validation' and contains a paragraph of text: '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.' Below this text is an orange 'Validate' button with the text 'Validate selected data file against a given data structure.' to its right.

3.5. Summary

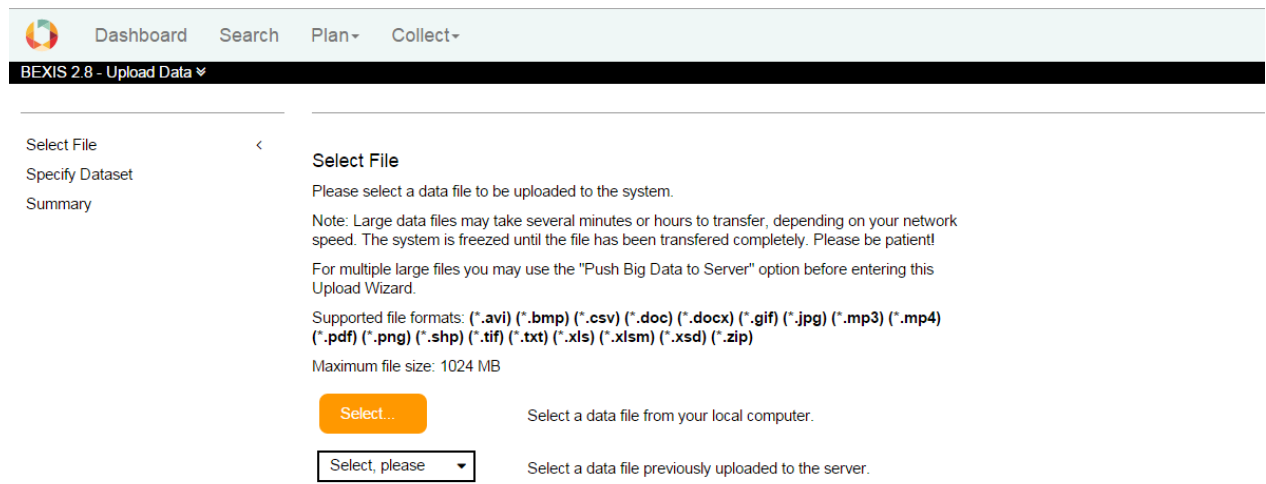
With this final step a summary of your uploaded data file is provided. Please check the information and click the **Finish** button to confirm and finalize the upload.

3.6. Unstructured 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) (*.csv) (*.doc) (*.docx) (*.gif) (*.jpg) (*.mp3) (*.mp4) (*.pdf) (*.png) (*.shp) (*.tif) (*.txt) (*.xls) (*.xls) (*.xsm) (*.zip).

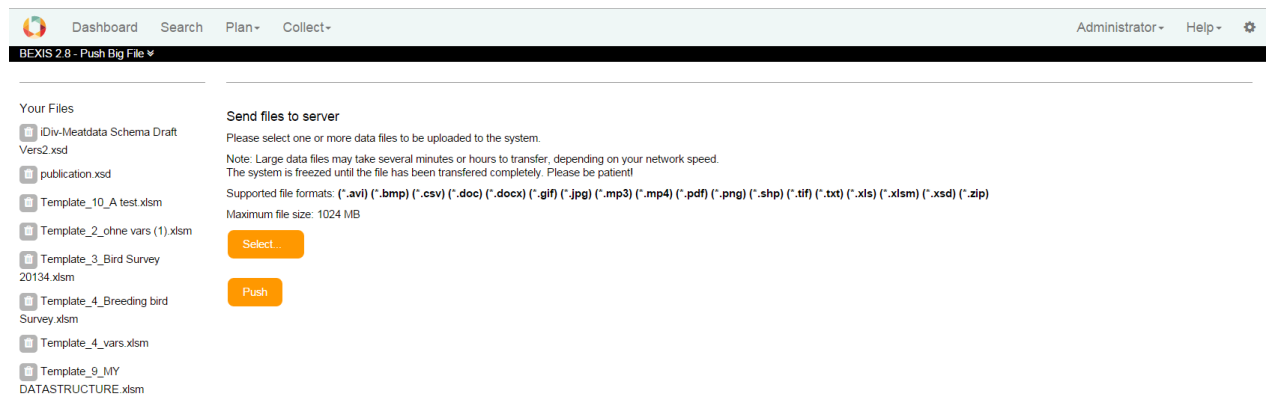
The Maximum acceptable file size up to now is: 1024 MB.



After selecting an unstructured data, you need to specify a related dataset.

4. 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.



5. Import 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 under the **Administration** tab in **Data Collection**. The wizard will assist you in importing your metadata structure into the BEXIS system. A metadata structure must be defined in a XSD schema file.

When importing a metadata schema into BEXIS, 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.

```
<xs:import namespace="http://www.w3.org/XML/1998/namespace"
           schemaLocation="http://www.w3.org/2001/03/xml.xsd">
</xs:import>
```

The current BEXIS system requires all referenced files to be locally available on the server (no URL to external resource). So 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 files to server).

```
<!-- ===== -->
<xsd:include schemaLocation="./baseTypes.xsd"/>
<xsd:include schemaLocation="./researchObjects.xsd"/>
<xsd:include schemaLocation="./dataStructure.xsd"/>
<!-- ===== -->

<xs:import namespace="http://www.w3.org/XML/1998/namespace"
           schemaLocation="./xml.xsd">
</xs:import>
```

5.1. 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.

The screenshot shows the 'Select File' step of the 'BEXIS 2.8 - Import Metadata Structure' wizard. The interface includes a navigation menu on the left with options: Select File, Read Source, Set Parameters, and Summary. The main content area is titled 'Select File' and contains the following text: 'To import a new metadata structure into the system, please select a XSD schema either from your local computer or from the previously uploaded files.' Below this text are two options: an orange 'Select' button for 'Select an XSD file from your local computer.' and a dropdown menu labeled 'Select' for 'Select an XSD file previously uploaded to the server.' At the bottom of the wizard, there is a '(1 of 4)' indicator and 'Cancel' and 'Next' buttons.

5.2. 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).

The screenshot shows the 'Read Source' step of the 'BEXIS 2.8 - Import Metadata Structure' wizard. The navigation menu on the left includes: Select File, Read Source, Set Parameters, and Summary. The main content area is titled 'Read Source' and contains the following text: 'Please specify a 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). Generating the new metadata structure in the system may take a few minutes, depending on the size and complexity of the XSD schema. Please wait for the confirmation message before heading to the next step.' Below this text are two input fields: 'Schema name' with the value 'ABCD Test' and 'Root node' with the value 'dataset'. An orange 'Generate' button is positioned below the input fields. At the bottom of the wizard, there is a 'Prev' button, a '(2 of 4)' indicator, and 'Cancel' and 'Next' buttons.

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.

```

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:sea="http://www.
  <xs:element name="DataSets">
    <xs:annotation>
      <xs:documentation>TDWG root element: container for one or more Da
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" name="DataSet">
        </xs:sequence>
      </xs:complexType>
    </xs:element>
    <xs:complexType name="Gathering">
    <xs:complexType name="Identification">
    <xs:complexType name="NamedArea">
    <xs:complexType name="Unit">
    <xs:complexType name="Sequence">

```

If no root node is selected then the wizard will automatically select the first element which is a complex type. But 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.

5.3. 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.

5.4. Summary

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

Dashboard Search Plan Collect Administrator Help

BEXIS 2.8 - Import Metadata Structure

Select File
Read Source
Set Parameters
Summary

Summary

Metadata structure name: ABCD Test

Root node

You successfully created metadata structure:
ABCD Test
Please use the **Create a Dataset** to create a dataset with this metadata structure.

Prev (4 of 4) Cancel