Genesys Anno User Manual

Genesys Helpdesk

Version 1.4 95812911, 3 May 2019
# Table of Contents

1. Introduction .......................................................... 1
2. Installing Anno ......................................................... 2
3. Troubleshooting .......................................................... 3
4. Workspaces and projects ................................................ 4
   4.1. Creating a new workspace .................................. 4
   4.2. Loading an existing workspace ............................. 4
   4.3. Using the workspace ......................................... 5
   4.4. Under the hood ................................................. 5
   4.5. Projects ......................................................... 5
5. Window layout .......................................................... 6
   5.1. Toolbar .......................................................... 6
6. Connecting to a Genesys server ....................................... 8
   6.1. Authenticating with Genesys ............................... 9
7. Data sources .......................................................... 11
   7.1. Excel and CSV files ........................................ 11
   7.2. CSV file configuration ...................................... 12
   7.3. CSV and Excel header row .................................. 12
   7.4. Databases ...................................................... 12
8. Mapping to MCPD .................................................... 16
   8.1. Basic column mapping ...................................... 16
   8.2. Mapping your data to MCPD ............................... 17
   8.3. Handling multiple values ................................... 17
   8.4. Using regular expressions .................................. 18
9. Pushing data to the server ............................................. 19
   9.1. Parse all ...................................................... 19
   9.2. Upload ........................................................ 19
   9.3. Remove ........................................................ 20
   9.4. Log levels ...................................................... 20
10. Troubleshooting .......................................................... 21
1. Introduction

The Genesys upload tool Anno is a standalone Java application for managing accession-level data on Genesys PGR - a global database on plant genetic resources in the world’s genebanks. The Genesys database is accessible at https://www.genesys-pgr.org.

The Anno application allows users to map their Excel XLSX files, CSV files or database SQL queries to Multi-Crop Passport Descriptor (MCPD) format and push the mapped data to Genesys for publication.

Genesys Sandbox, a playground instance of the Genesys database, is available for developers and integrators to validate and test their configuration before pushing data to live Genesys servers. It can be found at https://sandbox.genesys-pgr.org.
2. Installing Anno

Anno is an open-source project, licensed under the Apache License v2. Anno requires the Java 8 Run-Time Environment (JRE) to run.

Make sure you don’t allow the Java JRE installer to install browser toolbars like Ask.com or make any other changes to your default browser configuration.

Disable Java in all Internet browsers on your computer. Java applet technology has been marked for deprecation and is seldom used today.

If downloading pre-compiled binaries, make sure to download the latest version of Anno for your platform. Most users will have a 64-bit CPU and JRE, and should use the package labeled `x86_64`. The package `x86` is for the 32-bit JRE.

Download the package from the downloads section. Extract, if necessary, and run the executable for your platform.

<table>
<thead>
<tr>
<th>Table 1. Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project page</td>
</tr>
<tr>
<td>Pre-compiled binaries</td>
</tr>
<tr>
<td>git repository</td>
</tr>
<tr>
<td>Issue tracker</td>
</tr>
</tbody>
</table>
3. Troubleshooting

The *WinRun4J* wrapper used for Windows OS may fail to automatically detect the Java run-time environment and fail with "Fail to find Java VM" error message.

In this case you will need to update the *Genesys Magic.ini* file and configure the `vm.location` property:

```
vm.location=C:\Program Files\Java\jre-9.0.4\bin\server\jvm.dll
```

Source: [Stackoverflow](https://stackoverflow.com)
4. Workspaces and projects

Upon starting the application, you will be presented with a **Workspace Launcher** that allows you to create a new workspace or load an existing workspace from disk.

A **workspace** allows you to save your Anno configuration files and any JDBC drivers needed to access your data in one location on your computer.

4.1. Creating a new workspace

When you first start the application, you will need to create a new workspace to store your configuration files.

![Workspace Launcher](image)

*Figure 1. Workspace Launcher*

Click the "Browse" button and navigate to the directory where you wish to create a new workspace. You may need to create a new folder for the workspace.

Confirm your selection by pressing "OK".

- The application will check if the selected folder is empty.

4.2. Loading an existing workspace

Click the "Browse" button and navigate to the directory with your existing workspace data. Confirm your selection by pressing "OK".

- The application will check if the selected folder is a valid Anno workspace folder.
4.3. Using the workspace

The workspace acts as a base directory for your configuration and data files. It is good practice to copy the source files (CSV, XLSX) to the workspace folder. This will help you better maintain your settings and data files you publish on Genesys.

4.4. Under the hood

Anno creates a sub-folder named "jdbc" in your workspace. This folder is used as a source location for any JDBC drivers you may need to access your databases.

4.5. Projects

Anno allows you to manage the settings, data sources and data mapping in project files. A project file contains:

1. Server settings, including the Genesys server URL, application keys and secrets
2. Data sources: CSV, Excel and database queries
3. Column configuration and mapping to MCPD

It is good practice to maintain one project file with the configuration used to test the data and push it to the Genesys Sandbox environment, and a separate project file to publish data to the Genesys production servers.
5. Window layout

After workspace selection, the main application window is loaded. The window has four sections:

- Toolbar (top)
- List of data sources (left)
- Data source view (center)
- MCPD descriptor list (right)

![Figure 3. Application window](image)

5.1. Toolbar

The toolbar provides access to top-level functions.

![Figure 4. Anno toolbar](image)

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load</td>
<td>Load an existing project file.</td>
</tr>
<tr>
<td>Save</td>
<td>Save the current project to a file.</td>
</tr>
<tr>
<td>Settings</td>
<td>Open the Settings dialog.</td>
</tr>
<tr>
<td>Add file</td>
<td>Add a new data source file to the project.</td>
</tr>
<tr>
<td>Automap</td>
<td>Automatically map columns of the currently open data to MCPD descriptors.</td>
</tr>
</tbody>
</table>

Table 2. Toolbar buttons
<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reset mapping</td>
<td>Remove assigned mapping for currently open sheet</td>
</tr>
<tr>
<td>Push</td>
<td>Open a dialog to send data to Genesys.</td>
</tr>
<tr>
<td>Add database</td>
<td>Add a new database-backed data source to the project.</td>
</tr>
</tbody>
</table>
6. Connecting to a Genesys server

The Settings dialog allows you to configure the current project and specify which Genesys server will receive your data.

![Figure 5. Settings dialog](image)

The settings are stored in the project file and will be saved and loaded with the rest of the project configuration.

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genesys server URL</td>
<td>The base URL of the Genesys server instance to use for this project. For testing purposes, the Genesys server URL should point to <a href="https://sandbox.genesys-pgr.org">https://sandbox.genesys-pgr.org</a>. For production use, it should point to the live Genesys URL, <a href="https://www.genesys-pgr.org">https://www.genesys-pgr.org</a>.</td>
</tr>
<tr>
<td>Endpoints</td>
<td>Authorization and token endpoints will automatically update when the server URL is changed. Do not modify these.</td>
</tr>
<tr>
<td>Client API key and secret</td>
<td>Contact the helpdesk at <a href="mailto:helpdesk@genesys-pgr.org">helpdesk@genesys-pgr.org</a> to obtain the valid client key and secret. Different values are used for the sandbox and production environments.</td>
</tr>
<tr>
<td>Access and refresh tokens</td>
<td>Authentication tokens are used to identify you to the Genesys server. These are obtained by clicking &quot;Authenticate&quot; or are loaded from the project file.</td>
</tr>
<tr>
<td>Scope</td>
<td>Scopes are granted to Anno to manage data on your behalf on Genesys. The scope must be: write.</td>
</tr>
<tr>
<td>Clear tokens</td>
<td>Clear access and refresh token values. You will have to re-authenticate with the server.</td>
</tr>
<tr>
<td>Authenticate</td>
<td>Validate the current configuration against the server or authenticate with the server.</td>
</tr>
</tbody>
</table>
6.1. Authenticating with Genesys

A valid user account on Genesys is required. You may use your Google+ account or create the account manually by providing a valid email address and an account password. Make sure you have valid user accounts for the sandbox environment at https://sandbox.genesys-pgr.org/login and the production servers at https://www.genesys-pgr.org/login.

You can use Google+ to create your user account on Genesys. You will not need to remember a separate password!

After you have obtained a valid client key and secret from helpdesk@genesys-pgr.org and created your Genesys account, you can authenticate against the selected server (sandbox or production).

Click the “Authenticate” button. When the access and refresh tokens are missing or have expired, the application will prompt you to authorize the application’s request to access Genesys on your behalf.

If tokens are still valid, their values in the dialog will be updated with the message: “Tokens are up to date.”

"Open link in browser" opens your default web browser (Chrome, IE, Firefox…) and prompts you to allow Genesys access to your resources. If you are not yet logged in to Genesys, you will be prompted to log in before the confirmation dialog is displayed.

Figure 6. Authentication dialog

Figure 7. Allow access
Select "Yes, allow access" and Genesys will generate a short-lived **verifier code** that you must copy and paste to the **Verifier code** field in the Authentication dialog. The verifier code is a 6-character string (e.g. t.d1S83). Unless an error occurs or the verifier code times out, your access and refresh tokens will be updated.

After obtaining the tokens, **save the project** by clicking the "Save" button in the application toolbar. Give the project file a name that tells you which Genesys server (production or sandbox) you have selected.
7. Data sources

Anno is able to load data from Excel XLSX and CSV files and through database SQL queries. Every data source must contain at least the following three columns that uniquely identify an accession on Genesys:

1. **INSTCODE**: FAO WIEWS Institute Code of the holding genebank.
2. **ACCENUMB**: Full identifier of the accession in the genebank.
3. **GENUS**: Genus of the accession.

7.1. Excel and CSV files

To add an Excel XLSX or CSV file to the project, click the "Add file" button in the toolbar. You will be prompted with an "Open file" dialog to select the source file to add to the project.

> Older versions of Excel files (with the .xls extension) are not supported.

Excel files may contain multiple sheets, which will be listed as individual data source sheets. CSV files contain only one sheet. The data sheets from source files are listed as sub-entries of the source file. To open a data sheet and load the first 300 rows, double click the sheet name.

*Figure 8. Project with XLSX and CSV data sources*

Loading data from Excel files is straightforward and requires no further configuration. This is not the case for CSV files. The file format of CSV files is much more flexible and may require additional configuration before it loads correctly.
7.2. CSV file configuration

CSV files are plain text files and do not provide any information about the character encoding, separator or quote character used to separate text strings from numbers.

Use Excel XLSX files instead of CSV files when possible. If you have a CSV file, we recommend opening it in Excel, making sure the data is well formatted, and saving it in XLSX format.

When opening a CSV data source, Anno may not be able to load the file until you provide information on the formatting of the file in the CSV tab of the data sheet.

Table 4. Configuration for CSV files: Formatting

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character set</td>
<td>CSV files generated from databases usually default to the character set of the operating system. Try &quot;windows-1250&quot; for files generated on Windows, or &quot;x-MacCentralEurope&quot; for Mac OSX.</td>
</tr>
<tr>
<td>Separator</td>
<td>Pick comma (,) or tab (blank).</td>
</tr>
<tr>
<td>Quote character</td>
<td>Pick single (’) or double quote (“).</td>
</tr>
</tbody>
</table>

You may have to experiment with different options. After making any changes to formatting, click "Reload" to load the CSV file with the new settings.

Even by providing the best settings for the CSV file, however, you cannot ensure that Anno will be able to read all data correctly. For best results, convert your CSV file to Excel XLSX format.

7.3. CSV and Excel header row

Some data files contain header rows at the top of columns, or within sheets, that should be ignored by Anno.

Table 5. Configuration for data files: Headers

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contains headers</td>
<td>Indicate that your CSV file contains a header row.</td>
</tr>
<tr>
<td>Header row index</td>
<td>Provide the index of the header row, whether it is at the top of a sheet or further down in the file.</td>
</tr>
</tbody>
</table>

Click "Reload" to read the data with new settings. Make sure that headers are read correctly.

7.4. Databases

Databases have many advantages over CSV and Excel files, and relational databases are often used to manage accession data. Anno allows you to directly query any database system using a valid
JDBC driver that permits the application to connect to the relational database management system.

Click "Add database" in the toolbar. This will add a JDBC connection to your database as a top-level source element. You will be able to add individual, tailored SQL queries as Anno data sources.

Figure 9. Adding a database as data source

Table 6. Add Database dialog

<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datasource type</td>
<td>Select the database type from the list of supported drivers: MySQL, MS SQL Server, PostgreSQL or ODBC.</td>
</tr>
<tr>
<td>Datasource name</td>
<td>Provide a name for the database connection to be used as the top-level label of the data source in the project.</td>
</tr>
<tr>
<td>Connection URL</td>
<td>Edit the JDBC connection string template. You will have to provide the database host name, port and database instance name.</td>
</tr>
<tr>
<td>User and password</td>
<td>Enter a valid username and password to access the database.</td>
</tr>
<tr>
<td>Connect</td>
<td>Attempt to connect to the database with provided settings.</td>
</tr>
<tr>
<td>Download driver</td>
<td>Attempt to download the JDBC driver for the selected database type.</td>
</tr>
</tbody>
</table>

Anno comes with an embedded MySQL driver; all other drivers need to be downloaded separately. The database type determines which JDBC driver should be loaded.

If your database type is not supported, contact helpdesk@genesys-pgr.org for assistance.

Click "Connect" to try to connect to the database. If all went well, you will be presented with a prompt to add the database link to the project. Otherwise, check the username, password and the JDBC connect string (search engines are a good resource to find a valid JDBC connect string for your database).
After the connection to the database is successfully established, the database connection is added as a top-level data source. You are now able to add SQL queries as individual data sources to the project. Right-click on the database data source and select “Add SQL query”.

This will create a data sheet entry under the database label, titled “Unnamed query”. Double-click the entry in the project data source tree and update the query label and the SQL query itself.

Press the “Reload” button to load data from the database. This will refresh the contents of the data sheet.

Save your project file regularly.
You should start with a simple SQL query to the database and then create additional data sheets in the project as you query for additional accession data.

All SQL queries must include **INSTCODE**, **ACCENUMB** and **GENUS** columns. Where necessary, try using SQL JOIN clauses to write an SQL query that includes this data.

Figure 13. A dummy SQL query with core columns: **INSTCODE**, **ACCENUMB** and **GENUS**
8. Mapping to MCPD

Once you have successfully loaded a data sheet with data you wish to publish on Genesys, you need to map the columns of your data sheet to Multi-Crop Passport Descriptors (MCPDs) listed on the right side of the application window.

1. Open the data sheet.
2. Click on the column heading label to load the current column configuration.
3. Drag the descriptor from the MCPD listing to the column configuration pane.

After dragging and dropping the descriptor into the column configuration pane, the RDF term field will be populated with the descriptor URL (e.g. http://purl.org/germplasm/germplasmTerm#germplasmID).

Your existing data must be compliant with MCPD for straightforward mapping!

8.1. Basic column mapping

Genesys requires the INSTCODE, ACCENUMB and GENUS for every accession in the data sheet. Load your data sheet and make sure it contains these three columns.

1. Click on the label of the column containing the FAO WIEWS Institute Code of your genebank.
2. Drag the INSTCODE descriptor from the MCPD list to the column configuration pane.
3. Select the column containing your genebank accession numbers.
4. Drag the ACCENUMB descriptor from the MCPD list to the column configuration pane.
5. Select the **genus** column of your accessions.

6. Drag the **GENUS** descriptor from the MCPD list to the column configuration pane.

![Figure 15. Mapping the ACCENUMB column](image)

Once these three columns are mapped, double-clicking on a row in the data sheet row listing will display an alert dialog with the accession data in mapped Genesys JSON format.

![Figure 16. Preview of data in JSON format](image)

As you continue mapping other columns, more information will be included in the JSON preview dialog.

⚠️ An application error dialog will be displayed when mapping is incomplete!

### 8.2. Mapping your data to MCPD

**TBD**

### 8.3. Handling multiple values

The MCPD standard specifies that multiple values may be provided for specific descriptors. One example of such data is the MCPD **REMARKS** field. You may manage this data in one Excel column or use multiple columns for each individual comment.

Anno allows you to specify whether a single column contains multiple values, as well as how the data
should be split. Alternatively, it allows you to map multiple columns of the data sheet to the same descriptor. In both cases, the individual pieces of the data will be converted to an array of values.

8.4. Using regular expressions

TBD
9. Pushing data to the server

After you have mapped your data to MCPD and confirmed that the JSON looks correct, you are ready to push the data to the Genesys server. See the Configuration section for details!

The Push dialog offers four functions:

1. Parse all
2. Upload
3. Remove
4. Change the log level

![Figure 17. The Push dialog](image)

9.1. Parse all

The "Parse all" action triggers a read-convert-parse operation of all records in the selected data sheet. This is a useful operation that will check whether all of your data will correctly load, parse and convert using the mapping definitions you have provided. Keep an eye on the log report and fix your data as needed before you attempt a push operation.

9.2. Upload

"Upload" will send the data from the selected data sheet to the Genesys server specified in the
Settings dialog. You will be prompted with the Genesys server URL before you begin pushing updates to the server.

### 9.3. Remove

"Remove" is an operation of last resort. It will move accession records from the active Genesys database to the archive. Only the three core columns must be mapped to successfully perform the operation: `INSTCODE`, `ACCENUMB` and `GENUS`.

⚠️ If applicable, use the `HISTORIC` flag instead of removing records from Genesys.

Note that Genesys never actually deletes accession data; it merely moves it to an archive that remains accessible if the record is referenced by its `PURL`, the Permanent URL.

### 9.4. Log levels

The dialog toolbar allows you to toggle the log level between `DEBUG`, `INFO` and `WARN`. This determines the level of detail rendered in the central log pane of the dialog. It defaults to `INFO`. 
10. Troubleshooting

This tool is perfect! No way you have a problem! :-)

However, if you do run into trouble using this tool, contact helpdesk@genesys-pgr.org for assistance. Based on your feedback, we will update the tool or update this section of the documentation with resolutions to commonly encountered problems.