PLANNINGSPACE ANONYMIZER

TOOL DOCUMENTATION
16.5

FEBRUARY 2022

DOCUMENT HISTORY

Version	Description of Changes	Author	Date
1.0	Document Created	Michael Seth	September 14, 2018
2.0	Aucerna branding update	Robert Will	February 18, 2020
2.1	Updated user interface screenshots	Robert Will	October 20, 2020
2.2	Update for functional changes in 16.5 Update 11 (February 2021): new options to enable/disable document names and user information from being anonymized	Phillip Kent	April 12, 2021
2.3	Update for functional changes in 16.5 Update 13: new command line interface. General tidying of document format.	Phillip Kent	July 09, 2021
2.4	Update for change in 16.5 Update 17: The setting 'Anonymize version names and history' has been split into two separate settings for 'version names' and 'document history'.	Phillip Kent	February 15, 2022

Copyright

Copyright © 2022 Aucerna

This document may be freely distributed and printed but only in its entirety. No individual part of it may be reproduced or quoted without the written consent of Aucerna.

Disclaimer

The information in this document is subject to change without notice. The software described in this document is furnished under a license agreement. This software may be used or copied only in accordance with the terms of such agreement. It is against the law to copy the software on any medium except as specifically allowed in the license agreement.

Table of Contents

1. Overview

Tool Name	Anonymizer
Application	PlanningSpace Dataflow, PlanningSpace Reserves
Compatible Releases	16.5
Tool Category	Database tool
Recommended User Level	Consultant, requires access to PlanningSpace tenant database in SQL Server

The Anonymizer is used to anonymize sensitive client data so databases can be shared with Aucerna staff, such as for investigating support incidents and UAT testing.

The following data content will be affected:

- Dataflow
    - Document history, including usernames and comments.
    - Audit Logs.
    - Document names will be anonymized, if enabled by the user.
    - Document data will be anonymized within a specific date range which the user can specify.
    - Document dates will be shifted by a specific range which is specified by the user.
    - Price decks.
    - Historic revision data.
    - Version names will be anonymized, if enabled by the user.
    - Document history will be anonymized, if enabled by the user.
    - Lists will be anonymized, if enabled by the user.
    - User information will be anonymized, if enabled by the user.
    - Removal of old revisions, if enabled by the user.
- Reserves
    - Reconciliation Names, Reconciliation Document Names.
    - Reconciliation Values, String Values, Change Reason Values, Economic Values.
    - Reconciliation Change Reason Values, Users.
    - Reconciliation Approval History.
- Decline
    - Well names, well group names.
    - Well properties.
    - Well data (including historic data).

2. Functionality

2.1 Prerequisites

This tool makes changes to a PlanningSpace tenant database in SQL Server. You must have a SQL Server login with read/write access to the database (i.e., the same permissions as required by the PlanningSpace IPS Server: roles 'db_datareader', 'db_datawriter', and 'pes_datawriter')
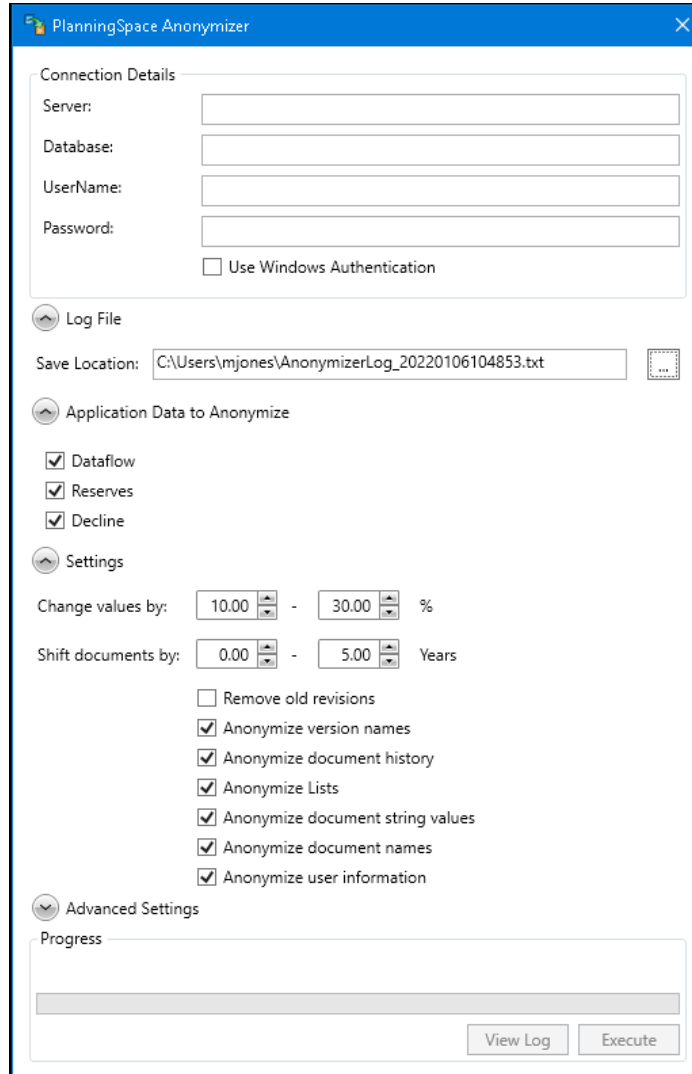
Before running the Anonymizer, the IPS Server must be shut down to ensure that all cached data for the PlanningSpace tenant is written to the database.

To improve the speed of the Anonymizer, ensure that the **Recovery Mode** of the database being anonymized is in **Simple** mode.

A command line interface is included in version 16.5.13 of the tool, and later versions.

The Anonymizer process cannot be undone, so you must ensure that you work with a copy of each tenant database and that the original database is not touched.

2.2 Using the tool

The tool has one set-up screen, as shown in the above screenshot.

The following table summarizes the **Connection Details** settings.

Server	Enter the network name of the server which contains the database you want to anonymize.
Database	Enter the name of the database that you want to anonymize.
UserName	Enter the username for login to the SQL Server database (*not needed for Windows Authentication*). *The user must have Owner privileges for the database*.
Password	Enter the password for login to the SQL Server database (*not needed for Windows Authentication*).
Use Windows Authentication	If this is checked then authentication will use the currently logged-in Windows user account (the 'DOMAIN/account' name will appear in the UserName field).

The **Log File – Save Location** setting sets the location of the log file.

The **Application Data to Anonymize** settings contain check boxes for **Dataflow**, **Reserves** and **Decline**. All are selected by default.

Fill in the **Settings** for the Anonymizer, and the **Advanced Settings** if required, (see the details below) and click **Execute** to run the tool.

The Anonymizer changes cannot be undone, so you must ensure that you are working with a copy of each tenant database.

Progress is indicated by the progress bar. Click **View Log** to see the log file.

2.3   Settings

The following settings are checked by default, and can be unchecked if desired:

- **Anonymize version names**: anonymize hierarchy version name, description, and the user that created the version.
- **Anonymize document history**: anonymize user information for document revisions.
- **Anonymize Lists**: anonymize all lists in Dataflow.
- **Anonymize document string values**: anonymize all string values in documents, including those found in the document revision history.
- **Anonymize document names**: anonymize Document names in the hierarchy; they will be renamed to use generic names (for example, Country 1, Country 2, etc.).
- **Anonymize user information**: anonymize user account information (i.e., Name, UserID, and Description).

Change Values By: The tool will change documents' numeric values based on this setting.

The original values will be changed by an amount within the specified percentage range. For example, if the setting is '0-20%' and the original value is 100, the anonymized value will be 100 plus or minus a random value between 0 and 20. The same random number will be applied to every numeric value throughout one document, meaning that the shape of the anonymized data will be similar to the original data's shape. Each document will use a different random value in the specified range.

Shift Documents By: Document dates will be shifted (plus or minus) by a random number of years specified by the user. Each document will be shifted by a different random amount to anonymize the data. For example if the user specifies a range of '0-5 years' and a document has a start date of 2020, the anonymized start date can be any year between 2015 and 2025.

Document scalar date time variable values will also have date values changed based on this setting.

Remove Old Revisions: Selecting this option will delete old document revisions from the database. The latest document revision will be kept, leaving only one revision per document.

2.4   Advanced settings

Reconciliation values batch size: (default: 2000) If you experience a timeout error during the Anonymizing Reconciliation Values phase when anonymizing a database, decrease this value and run the Anonymizer again. Decreasing the batch size has a negative impact on performance, so make any changes in small steps to minimize the impact.

Reconciliation changes batch size: (default: 1000) If you experience a timeout error during the Anonymizing Reconciliation Changes phase when anonymizing a database, decrease this value and run the Anonymizer again. Generally, this value should be around half of the value of the *Reconciliation values batch size*. Decreasing the batch size has a negative impact on performance, so make any changes in small steps to minimize the impact.

Maximum SQL connections: (default: 500) This setting determines how many connections the Anonymizer can establish with the SQL Server while running. You may need to decrease this value for highly-loaded SQL Servers, but generally it can be left at the default. For high-performance SQL Servers, this value may be increased.

2.5   Command line interface (version 16.5.13 and later)

The Anonymizer tool can be used from the PowerShell command line in version 16.5.13 (16.5 Update 13) and later. In this way the tool can be setup as a scheduled task, or it can be included in automation scripts, etc.

The executable file is 'Palantir.Dataflow.Tools.Anonymizer.exe'.

The tool has two modes: 'testconnect' and 'run'. For help information, enter either:

```
.\Palantir.Dataflow.Tools.Anonymizer testconnect --help
```

or:

```
.\Palantir.Dataflow.Tools.Anonymizer run --help
```

In 'testconnect' mode, only the database connection is tested. For example:

```
.\Palantir.Dataflow.Tools.Anonymizer testconnect -s SQL.company.com
 -db TENANT_ATLANTIS
```

In this case, Windows authentication will be used, based on the Windows account under which the PowerShell is running. If a username and password are needed then the `-u` and `-p` parameters must be included.

If the database can be accessed then the response looks like:

> *Starting...*
> *Connecting to SQL.company.com TENANT_ATLANTIS*
> *Finished. Duration: 00:00:00.7633415*

If the database details or authentication are not correct this will be shown in the (red) error messages.

In 'run' mode the input parameters for the SQL connection/authentication are again required. The Anonymizer settings that are specified in the GUI dialog box are instead written into a JSON file, which is specified using the parameter `-as` or `--appSettings`. There is a default file included in the tool folder 'appsettings.json', with contents as follows:

```json
{
  "applicationsToAnonymize": {
    "dataflow": true,
    "reserves": true,
    "decline": true
  },
  "dataflowOptions": {
    "deleteOldRevisions": false,
    "changeDateMin": 0,
    "changeDateMax": 5,
    "anonymizeVersions": true,
    "anonymizeDocumentHistory": true,
    "anonymizeLists": true,
    "anonymizeStrings": true,
    "anonymizeDocumentNames": true
  },
  "reservesOptions": {
    "reconciliationValueBatchSize": 2000,
    "reconciliationChangesBatchSize": 1000
  },
```

```
    "logFilePath": "%UserProfile%\\Documents\\Anonymizer_Log.txt",
    "changeValueMin": 10,
    "changeValueMax": 30,
    "SQLMaxPoolSize": 500,
    "anonymizeUserInformation" : true
}
```

3. Technical Information

Value anonymization: values are modified using the Microsoft Random class, which is an industry standard pseudo-random number generator. A random value modifier is generated and applied to each PlanningSpace document based on the range specified by the user. A second randomly-generated number determines the direction of change (increase or decrease) in the numerical values.

The random values are not stored, and the values in the database are modified directly (i.e. no copy is made before randomization); *the operation cannot be undone, therefore ensure you are working with a copy of the database and not the original*.

String anonymization: Document Names, Reconciliation Names, comments, Reconciliation string values, user names, and any other strings that are modified are given generic names based on a pre-defined convention. For example, Dataflow documents are renamed based on the node type plus an incremental integer; Reconciliation Nodes are renamed based on the Reporting Standard plus an incremental integer. There is no way to recover the original strings after they have been anonymized.

Price Deck anonymization: Price decks (which are shared between Economics, Financials and Dataflow) will be anonymized if the Dataflow checkbox is checked. Gas prices will be set to 5 for all cells containing a value, and both Liquid and Generic prices will be set to 50 for all cells containing a value (regardless of the currency in the price deck). The Price Deck names will be made generic (e.g., Price_Deck_1) and so will the Price Scenarios (e.g., Scenario_1).

WI Variables: WI Tree variables will be anonymized if the Dataflow checkbox is checked. WI values will be randomly-changed for one or more partners. Note: if there is only one partner, the WI cannot be anonymized (because the WI Tree variable only allows the value of 100 when one partner is present).