

Antares_Data_Organizer 2.0.0

REFERENCE GUIDE

<i>Simulation package</i>	
<i>Script Editor package</i>	
<i>Graph Editor package</i>	
<i>Data Organizer package</i>	<i>X</i>

DATA ORGANIZER

REFERENCE GUIDE

Table of contents

1 Introduction	3
2 Data manager	4
2.1 Choice of origin and destination workspaces.....	4
2.2 Choice of actions to perform.....	5
2.3 Choice of the moment when the actions should be performed	6
3 Catalog manager	6
3.1 Choice of source and sink workspaces	6
3.2 Creation and suppression of catalogs	6
3.3 Duplication of catalogs	6
3.4 Concatenation of catalogs.....	7
4 Data browser	7
4.1 Study: identification and metrics	7
4.2 Archive: Identification and metrics	7
4.3 Study and Archive: Antares notepad.....	8
4.4 Chests: Identification, metrics and notepad	8
4.5 Packs: Identification, metrics and notepad	8
5 Chest Editor.....	8
6 Assessment of hash signatures	9
7 Performances	10

1 Introduction

The purpose of this document is to detail the operations that can be carried out with the **Antares Data Organizer** application, which may be installed as a companion tool to the **Antares_Simulator** application. This document does not address issues related to the use of **Antares_Simulator**. The reader is assumed to be familiar with the concepts developed in **Antares_Simulator**, to which an introduction can be found in the document named “Antares_Simulator General Reference Guide”.

Antares studies often involve a great number of files and may require a lot of space for storage and a lot of time for copy/transfer operations. Besides, complex study process may require many simulations on multiple datasets variants. The need can then arise to categorize these datasets in different manners (e.g. simulations related to such time horizon or to such technical strategy, etc.).

In this context, the purpose of the **data organizer** is, by providing a flexible environment dedicated to the organization of the application data, to help users to classify and keep track of the work completed on an everyday basis, while saving time and disk resources on the long run.

The **Antares Data Organizer** is built around the following concepts:

- Antares study: folder containing data used and/or produced by **Antares_Simulator**
- Antares archive: file containing a compressed view of an Antares study
- Antares chest: any folder which is not a study and which contains a root file named “info.ado”
- Antares pack: file containing a compressed view of an Antares chest
- Antares data object: **either** a study **or** an archive **or** a chest **or** a pack
- Antares portfolio: **either** a single Antares data object **or** a folder containing any kind of Antares data objects (studies, chests, archives, packs)
- Antares study catalog: file containing references (storage paths) to Antares studies or Antares chests
- Antares archive catalog: file containing references (storage paths) to Antares archives or Antares packs
- Antares directory: a folder containing Antares catalogs

The application provides different services listed hereafter:

- Conversion of Antares studies into Antares archives, back and forth
- Conversion of Antares chests into Antares packs, back and forth
- Trimming of Antares studies (removal of all redundant output and log files)
- Edition of Antares study digest reports
- Registration/unregistration of Antares studies and chests in/from Antares studies catalogs
- Registration/unregistration of Antares archives and packs in/from Antares archives catalogs
- Copy of Antares studies, archives, chests and packs into Antares portfolios
- Copy of Antares studies catalogs and Antares archives catalogs into Antares directories
- Concatenation of Antares catalogs within an Antares directory
- Creation of Antares chests

These services are provided by three independent modules/panels whose content is further detailed in the next sections:

- Data manager (Management of Antares Studies, Archives, Chests and Packs within Antares portfolios)
- Catalog manager (Management of catalogs within Antares directories)
- Data browser (Outlook of a single specific Antares study / archive/ chest / pack)
- Chest editor

Note: In this software, hovering the mouse over a function button or an input field displays help about the function or about the expected field content.

Note: Paths too long to be displayed within the application standard layout are shortened and terminate by “...”
Hovering the mouse over the pathname displays the complete path.

2 Data manager

A data manager session involves three steps carried out through the GUI and described hereafter.

2.1 Choice of origin and destination workspaces

This may be done either by using commands “origin” and “destination” of the main data organizer menu or within the data manager panel itself.

- The “origin” workspace to be used by the data manager for the current session is either a portfolio (a single Antares data object or a folder containing data objects) or a catalog (containing references to studies or to archives) on which the user wishes to perform a number of actions.

Note: The selection of an origin workspace is mandatory. After selection, the list of archives and/or studies that are stored (portfolio) or referenced (catalog) within it is displayed

- The “destination” workspace contains one mandatory element and two optional elements. The mandatory element is the location at which new data generated by actions listed in 2.2 should be stored. Setting up this location requires to choose between two possibilities:
 - Unstructured: selection of a portfolio within which the results of all data-generating actions will be stored in a “flat” way (i.e. at the root of the portfolio)
 - Structured: for each item to process, use of the parent folder containing the native item to store the results of the action(s) scheduled for this item.
- In addition to the destination location, the destination workspace may optionally include two catalogs (one study catalog and/or one archive catalog)

Once these choices have been made, the list of the items found in the origin workspace is displayed. Two options available top right of the data manager GUI may help speed up the operation:

Exploration depth: Items stored deeper than the value of this parameter (default value: 10) will not be examined and will not appear in the displayed list.

Display size: The “size” parameter will not be assessed for displayed items¹.

The next stage of the session consists in selecting actions to perform on some or all of the displayed items. Three options available top right of the data manager GUI may help speed up some of these actions:

Compression algorithm: choice is given between saving time or saving space during compression operations.

Include hash: when compressing an object or “showing” its content, a digital signature of the original may be added or not to the result. This parameter applies to the “archive” and “show” functions described here below. Unticking this option may save a lot of time when dealing with large objects.

Check integrity: Compression of studies or chests into archives or packs is done twice and results are compared so as to guarantee data integrity. Unticking this option may save a lot of time (and HDD space) when dealing with large objects.

¹ After turning “on” this parameter, the origin portfolio definition needs to be refreshed for the change to take place

2.2 Choice of actions to perform

For each item displayed in line (archive or study), three information fields and eight checkboxes are shown in columns. The three Information fields stand as follows:

- Item size (space on disk used for the storage of the study or of the archive))
- Date of last item modification (for a study : last modification through the simulator GUI; for an archive or a pack: date of creation of the archive or pack file; for a chest: last modification of the folder)
- Status ("archive", "pack", "study", "chest" if the origin workspace is a portfolio ; "OK", "not found" or "corrupted" if the origin workspace is a catalog)

The eight checkboxes stand as follows:

- Show (S): produces in a general report (stored in the "log" folder) a section presenting the general metrics of the current item The content of this section is the same as that shown item-wise by the data browser (see section 4)
- Register (R): adds a new reference to the current item in the ad hoc destination catalog
- Unregister (U): removes the reference to the current item from the ad hoc origin catalog
- Trim (T): removes all output files and log files from the current study (not available for archives)
- Copy (C): copies the current study folder or archive file to the destination portfolio (if different from the origin workspace)
- Archive (A): compresses a study or a chest in files "study_name.antar" or "chest_name.antpack"
- Expand (E): Decompress the current archive or pack file in a study or chest folder bearing its original name
- Delete (D): deletes the current study, chest, archive or pack from the origin portfolio.

Note: Depending on the nature of the item shown in line and on the nature of the destination workspace(s), each checkbox can either be ticked or must remain blank. For instance, studies can be compressed as archives, but archives cannot be compressed again. Likewise, it is not possible to register a study in a catalog if no catalog has been selected, etc.

Note: The name given to an archive is that of the Antares study folder it comes from (also called "external name"). This name is often identical to that of the proper Antares study name (also called "internal name"), but it may differ from it.

Note: The **Delete** command actually removes the dataset (archive or study) selected in the origin workspace (portfolio or catalog). It should not be confused with the Unregister (U) command, which only removes from an origin catalog the reference to an external dataset, without alteration of the dataset itself.

Note: The new items produced by the **Archive**, **Expand** and **Copy** commands are stored in the destination portfolio.

Note: The new items produced by the **Archive**, **Expand** and **Copy** commands are automatically registered in the archives and/or studies catalogs that may have been selected on the destination side. These destination catalogs should not be selected if registration is undesirable.

2.3 Choice of the moment when the actions should be performed

Once the desired set of actions to perform on the items of the origin workspace has been defined by ticking the appropriate checkboxes, these can be either performed on the spot or planned for a later execution, by defining an ad hoc delay in the bottom left field of the data manager panel.

When several checkboxes are ticked on the same line, they are performed from left to right.

Note: a planned operation may be canceled at any moment by using the ad hoc button

Note: the countdown is not affected by the Windows session entering locked state

Note: the countdown is stopped by the Windows session entering sleeping mode

3 Catalog manager

A catalog manager session involves a sequence of steps carried out through the GUI and described hereafter. To the exception of the very first of them, they are all optional.

3.1 Choice of source and sink workspaces

This may be done either by using commands “source” and “sink” of the main data organizer menu or within the catalog manager panel itself:

- The “source” workspace to be used by the catalog manager for the current session is a directory (containing studies catalogs and/or archives catalogs).
- The “sink” workspace to be used by the catalog manager for the current session is a directory (containing studies catalogs and/or archives catalogs).

Once the source and/or the sink have been chosen, the lists of catalogs found in both source workspace (left side) and sink workspace (right side) are displayed in the catalog manager panel and any² of the steps described further in 3.2, 3.3 or 3.4 may be performed.

3.2 Creation and suppression of catalogs

New catalogs may be created in either source or sink workspaces by carrying out the following sequence:

- Typing of a character string forming a valid file name³ (without the specific catalog extension)
- Using either the “New Archives Catalog” button or the “New Studies Catalog” button of the appropriate side of the panel

Existing catalogs may be deleted in either source or sink workspaces by carrying out the following sequence:

- Ticking the checkbox(es) associated with the catalog(s) to delete
- Using the “Delete” button of the appropriate side of the panel

3.3 Duplication of catalogs

Catalogs may be copied either way (from source to sink or from sink to source) by carrying out successively the following operations:

- Ticking the checkbox(es) associated with the catalog(s) to copy

² If the source or the sink is not defined, only the (one-side) 3.2 command can be performed

³ Symbols for valid names are: a-z, A-Z, 0-9, -, _, (,), &, space, comma, @

- Using either of the “Left-to-Right” or the “Right-to-Left” arrow-shaped buttons located in the middle of the panel

3.4 Concatenation of catalogs

The content of a set of catalogs selected in the source workspace can be appended to catalogs selected in the sink workspace by carrying out successively the following operations:

- On the source side, ticking the checkbox(es) associated with the catalog(s) whose content should be transferred
- On the sink side, ticking the checkboxes associated with the catalog(s) to which the transferred content should be appended
- Using the cross-shaped button located in the middle of the panel

Note: The content of a catalog cannot be appended to a catalog of a different type (i.e. archives catalogs cannot be appended to studies catalogs and studies catalogs cannot be appended to archives catalogs).

Note: When the content of catalog A is appended to catalog B, references found in A that may already exist in B are not duplicated. Only missing references are copied.

4 Data browser

The Data Browser panel has no interactive GUI element. Its sole purpose is to display in a way easy to read a number of characteristics related to a specific Antares study, archive, chest or pack selected beforehand in the Data manager panel. This selection is carried out by clicking on the relevant item in the Data Manager list (see section 2). As a result, the following information is displayed in the Data Browser panel.

4.1 Study: identification and metrics

- Study External name (i.e. name of the folder containing the Antares study)
- Study Internal name (i.e. Antares study name)
- Study storage location
- Original Study Hash (information printed in the “show” file but not displayed in the browser window)⁴
- Version (Antares format of the Antares study)
- Last saved (first archive creation date; subsequent copies of the item will still refer to this date)
- Size and file number of the “Input” data folder
- Size and file number of the “Output” data folder
- Size and file number of the “User” data folder
- Total size and file number of the Antares study

4.2 Archive: Identification and metrics

- Archive file name
- Original Study location
- Original Study catalog (if any : catalog used to find the study when the archive was created)
- Study External name (i.e. name of the folder containing the Antares study)
- Study Internal name (i.e. Antares study name)
- Archive storage location
- Original Study Hash (digital signature of the study)
- Version (Antares format of the Antares study)
- Last saved (first archive creation date; subsequent copies of the item will still refer to this date)

⁴ Computing the study digital signature may require a long time, hence the information is not delivered when using the interactive data browser. It can, however, be accessed by using the “show” command.

- Size of the archive file
- Size and number of files of the uncompressed original study

4.3 Study and Archive: Antares notepad

This field displays the all-purpose information stored in the “user’s notes” notepad that users may have filled out when building the dataset with the Antares simulator

Note: The content of this study-related notepad cannot be modified by the Data browser. Since it is a proper part of the study dataset, it can be edited only with ***Antares_Simulator***

4.4 Chests: Identification, metrics and notepad

- Chest name (i.e. name of the folder containing the info.ado file)
- Chest path
- Last saved : folder’s last modification date
- Size and number of files contained in the folder
- In the user’s notes section : content of the readme.ado file

4.5 Packs: Identification, metrics and notepad

- Archive file name : name of the pack file
- Original location: location of the original chest folder compressed into a pack
- Original Study catalog (if any : catalog used to find the chest when the pack was created)
- External name (i.e. name of the folder containing the Antares chest)
- Location: path to the pack file
- Last saved (first pack creation date; subsequent copies of the item will still refer to this date)
- Size of the pack file
- Size and number of files of the uncompressed original chest
- In the user’s notes section : content of the info.ado file located at the root of the original chest

5 Chest Editor

The Chest Editor panel is meant to provide a simple way for the user to declare that a folder (whose content, size, structure are strictly arbitrary) should be further considered as an Antares Chest.

An Antares Chest is a folder which:

- Is not an Antares Study
- Contains a root file named “info.ado”.

Folders that cannot be defined as Antares chests are:

- Antares studies.

The functionalities offered by the Chest Editor panel are:

- Choice of the folder to declare as a new chest (button “make”)

- Edition of the content descriptor (info.ado file) of an already existing chest (button “update”)
- Restoration of the folder’s status to “normal” by deletion of the info.ado file (button “unmake”)

6 Assessment of hash signatures

When compressing data objects (studies or chests), the data manager may include or not a numeric signature of the object, under the form of a “SHA-1” string of characters.

Many ready-made open source software exist for the assessment of such signatures but there is no established standard regarding how signatures of complex objects (folders containing other folders and files) should be assessed in a universal and safe cross-platform way (the final signature depends on the order in which individual items are aggregated, with or without inclusion of their names).

The Antares_Data_Organizer SHA-1 assessment code is able to compute hash for folders and files, along with their names, in the same fashion as Git, with no limit on file size, and gives strictly identical results in windows and Linux environments (note: not tested on Unix-style systems).

7 Performances

Compressing, expanding or copying large objects may require a lot of time and storage space. To help manage these issues:

- The operations programmed within the data manager panel may be deferred to a time of comparatively low charge (countdown button)
- Chests and studies may include Antares data objects in compressed forms (i.e. archives “.antar” or packs “.antpack”). When compressing such chests or studies, compression is not re-run on these items, which are simply added to the archive.
- The Antares_Data_Organizer compression engine is the well-known 7-zip GPL archiver (version distributed along with the ADO in windows packages: 7-zip 18.05) Limits in terms of size and time are those of this library.

In practical terms, hereafter are some indicative results obtained for a compression-decompression cycle as well as for a copy operation, on the following Windows configuration:

Windows 10, CPU: Intel I7-4900 MQ @2.8GHz, RAM: 8 GB

To assess the upper bound of time requirements, all operations were carried out **From** an external USB 3 drive **To** another USB 3 drive

DATA USED FOR THE TEST:

Chest **C** to compress as a pack **P**:

Size: 1 127 706 138 788 bytes (# 1 TB)

Number of files: 10 221 546

Size of largest files: chest includes two files of 61 GB each (.antpack, excluded from compression)

Pack **P** to expand as a chest **C**:

Size: 335 472 349 570 (# 313 GB)

Number of files: 1

Pack **P** to copy as a pack **P**

Size: 335 472 349 570 (# 313 GB)

Number of files: 1

RESULTS:

Compression C > P	with hash assessment and integrity check:	79 hours
Compression C > P	without hash nor integrity check:	15 hours
Expansion P > C		11 hours 20 minutes
Copy P :: P		2hours 30 min.

Besides, the time for opening the compressed pack without expanding it (so as to browse through the pack's content and/or extract some data from it), using the 7-zip console with the password indicated in the troubleshooting notes, is about: **1 minute**

Antares_Data_Organizer 2.0.0

TROUBLESHOOTING NOTES

Problem: The data manager does not find all of the studies/archives located within the “origin portfolio”.

Solution: The exploration carried out by the data manager within the origin portfolio is limited to a number of levels (field “research depth” of the data manager panel, default value 10) adjustable between 0 and 50. Items located deeper than the maximum level cannot be found. If level 0 is chosen, only items located at the root of the portfolio will be found.

Problem: The data manager takes a long time to find all of the studies and archives included in the “origin portfolio”.

Solution: The in-depth exploration of the portfolio begins to be performed as soon as the origin portfolio is chosen, and it goes as far down as allowed by the selected exploration depth. Full examination of the content of C:\ may require much more time than that of C:\users\username\myAntares_space.

Problem: A study (or archive) that was correctly registered within a catalog in the course of a Data Organizer session cannot be found later, in another session in which the relevant catalog is used as origin workspace for the data manager.

Solution: In the interval between the two Data Organizer sessions, the study (or the archive) has been deleted or moved elsewhere by direct Windows Explorer commands (cut/paste, etc.) or other applications. The Data Organizer is not aware of operations performed by other applications on data it is supposed to manage. To keep registered keys consistent, copy or deletion of a registered item should be performed by the Data Organizer, in a session including appropriate register/unregister commands.

Problem: The data organizer signals that an error occurred during the copy of a study (or of an archive), but the copied data seems to be correct.

Solution: An error message is proof that the copied item is not the exact replica of the original item that it is required to be. The copied item is corrupted somehow and should not be considered as being equivalent to the original.

Problem: The data organizer signals that an error occurred during the deletion of a study (or of an archive)

Solution: Two cases are possible: a) the user does not have the proper rights to perform the action b) deletion was not performed because previous errors made it unsafe (for instance: in a “Copy and Delete” sequence, “Delete” is not performed if “Copy” fails).

Problem: The data organizer GUI main window gets partially mixed up with other application main windows (Outlook, Word, Excel, etc.). Minimizing and restoring the application window seems to fix the behaviour but the problem reappears after some time.

Solution: The graphic framework used to design the Data Organizer is not fully compatible with “old-fashioned” Windows desktop themes, that is to say themes that do not manage transparency (Aero) effects. To permanently fix this issue, please change the Windows Desktop Theme to one of the “Aero” family (for instance, “Windows 7” instead of “Windows 7 basic”). If “Aero” themes are not available, minimize all opened application windows before starting a Data Organizer session and restore them to normal size at the end of the session.

Problem: Crash without message.

Solution: Examine the content of the file located in:

C:\D:\....\Users\username\AppData\Local\rtel\data_organizer\logs\data_organizer.log

Problem: Need to restore the GUI default aspect (missing tab, empty GUI, etc):

Solution: Delete the folder: *C:\D:\....\Users\username\AppData\Local\rtel* (it may be invisible)

Problem: Need to extract some information from within a large archive or chest, without expanding it.

Solution: Browse through the archive or chest with the 7-zip application, with the password:

Antares_Data_Organizer_Password_Format_001

Antares_Data_Organizer 2.0.0

ATTRIBUTION NOTICE

Antares_Data_Organizer

Copyright 2007-2019 RTE - Author: Sylvain Marandon

Antares_Data_Organizer is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

Antares_Data_Organizer is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Antares_Data_Organizer. If not, see <<http://www.gnu.org/licenses/>>.

Antares_Data_Organizer 2.0.0 uses external libraries. Source and binary forms of these programs are distributed along with Antares_Simulator with NO WARRANTY. The list of these libraries can be found in the document ATTRIBUTION-NOTICE.txt located in the root folder of a distribution of Antares Data Organizer 2.0.0 in source form.