Martus Data Summarizer Software User Guide

Version 1.1

The Martus Data Summarizer software has been developed and is being distributed under a free/open source license.

TABLE OF CONTENTS

1. Introduction to the Martus Data Summarizer Software

The Martus Data Summarizer shows descriptive summaries of your Martus bulletin data, showing how often each field is filled in, and what the most common and least common values are for each field. These field level descriptions of data entered in Martus bulletin fields is helpful to Martus users for assessing data quality, and to get a big picture understanding of the data that has been entered in Martus. For example, you can see how many bulletins were entered by a certain user, or mention individuals of a given ethnicity (if your project is collecting that information).

Important Notes:

- Interpretation of Martus Data Summarizer output should be limited to descriptive summaries of the characteristics of the data you have collected, and should not be used to draw conclusions about patterns of behavior more broadly, beyond your collected data.

    1. Martus Data Summarizer output is intended to give you information about the status/progress of data that has been gathered and entered in Martus, but not to analyze "what happened" from the data that has been entered. So for example we can count the number of bulletins documenting killings from a particular timeframe or region (if your project is collecting that information), but not how many killings actually occurred in a particular timeframe or region.
    2. Martus Data Summarizer output does not correct for the following:
        a. Sampling bias (e.g. more bulletins entered from one particular region than others, due to data collection circumstances, even if the project is covering incidents that happened across all regions).
        b. Duplicate reporting (i.e. more than one bulletin may exist about the same incident, or the same person may be mentioned using different names, or there may be multiple people with the same name).
        c. Data quality issues (e.g. typos that might lead to multiple values in a field that were meant to be the same, fields being left blank where values were expected). Some of these issues can be addressed in a Martus project customization, by requiring certain fields to be entered before a bulletin or grid row can be saved, or using dropdown fields to avoid typos and guarantee consistency in data entry. For more information about this, see section 10b ("Customizing bulletin fields") in the Martus User Guide (available at https://www.martus.org/downloads/, or in the Martus\Docs directory of your Martus installation).
        d. Possible inconsistency in how data is entered in Martus by different people. For more on this topic, see the "Inter-rater Reliability" and "Controlled Vocabulary" sections on http://www.hrdag.org/resources/core_concepts.shtml.

3. Martus Data Summarizer output especially should not be used to attempt to count the number of violations, victims or perpetrators, beyond counting the number of bulletins that match certain field values. Before counting violations or victims/perpetrators, important data processing needs to take place before such counting can be considered scientifically defensible. This includes addressing the possibility of duplicate reporting (more than one bulletin may exist about the same incident); representing the complexity of human rights violations (an individual who was the victim of a given act may later become the perpetrator of a different act, or vice versa); and maintaining consistency in meaning and counting (ensuring, for example, that all the data entered in bulletins used consistent definitions for violations.)

   The Martus Data Summarizer can be used for descriptive or qualitative analysis of the information entered into Martus, but please keep in mind that quantitative or inferential statistical analysis cannot be performed until additional steps have been taken in processing the data, and without enough volume and different sources of data, statistical analysis may not give meaningful or reliable results.

- Please also note that, unlike in the Martus software, where all data is encrypted, the information in the Martus Data Summarizer output files is not encrypted (so that anyone who gets a copy of this file or a printed version of the output will be able to read all the data). Please only store Martus Data Summarizer output files in secure locations on your computer or network, and do not leave printed versions of them in an insecure location.

2. Getting Started

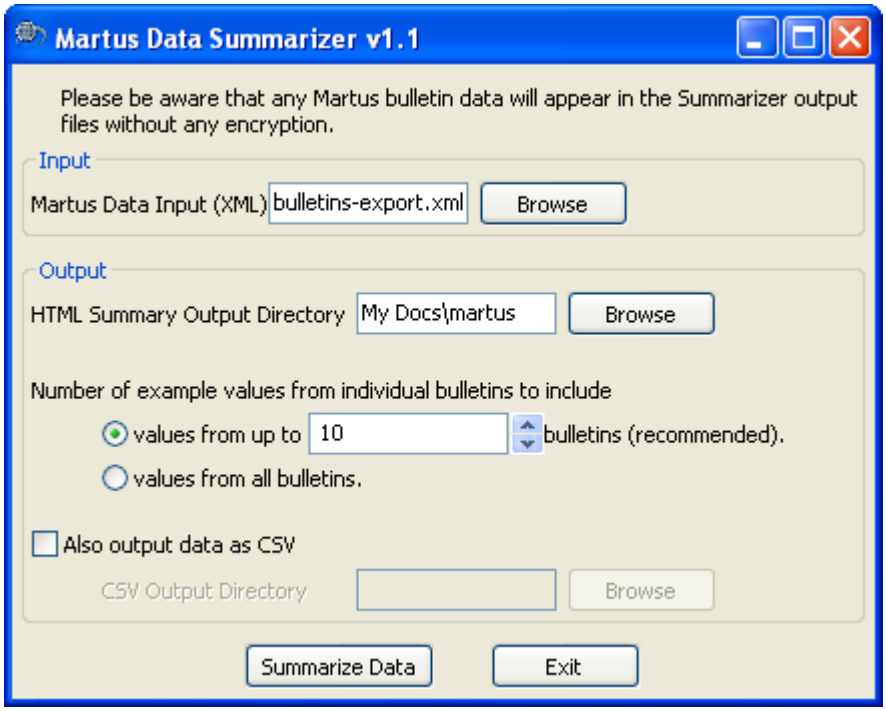
You can run the Martus Data Summarizer on any computer running Windows or Mac OS that meets the same requirements as the Martus software. For more information about the most current system recommendations, see the software download page on https://www.martus.org, or section 2 ("Getting Started") in the Martus User Guide (available at https://www.martus.org/downloads/, or in the Martus\Docs directory of your Martus installation).

The Martus Data Summarizer accepts a Martus exported bulletin file (.xml), which contains the final version of a bulletin, as an input file. The Martus Data Summarizer provides an output file (or files) in .html format. You can drill down into each field entry in the main .html output files to see more details about the values entered in the field. There is also an additional option to save the summarized data as .csv format

3. *How to run the Martus Data Summarizer*

The Martus Data Summarizer can only be used after Martus bulletin data has been entered and exported.

1. Once the bulletins are created in Martus, export the bulletins as an XML file. For instructions on this step, please refer to section 6p. ("Exporting bulletin data") in the Martus User Guide (available at https://www.martus.org/downloads/, or in the Martus\Docs directory of your Martus installation). Please make sure that your Martus .xml file only contains the final version of your bulletins.
2. Get the Martus Data Summarizer program for Mac or Windows from Benetech, by emailing info@martus.org.
3. Start the Martus Data Summarizer program by doing the following:
   a) Windows: unzip the Martus Data Summarizer zip file, and double-click the "Martus Data Summarizer.exe" file
   b) Mac: unzip the Martus Data Summarizer zip file. This should extract a Martus Data Summarizer .app file that you can put in your applications folder or on your Desktop, or wherever you prefer on your computer. Double-click on the "Martus Data Summarizer" .app file.
4. The Martus Data Summarizer window is displayed.



5. In the "Martus Data Input (XML)" field, browse to the location on your computer where your exported Martus bulletin (.xml) file is, and select the file.

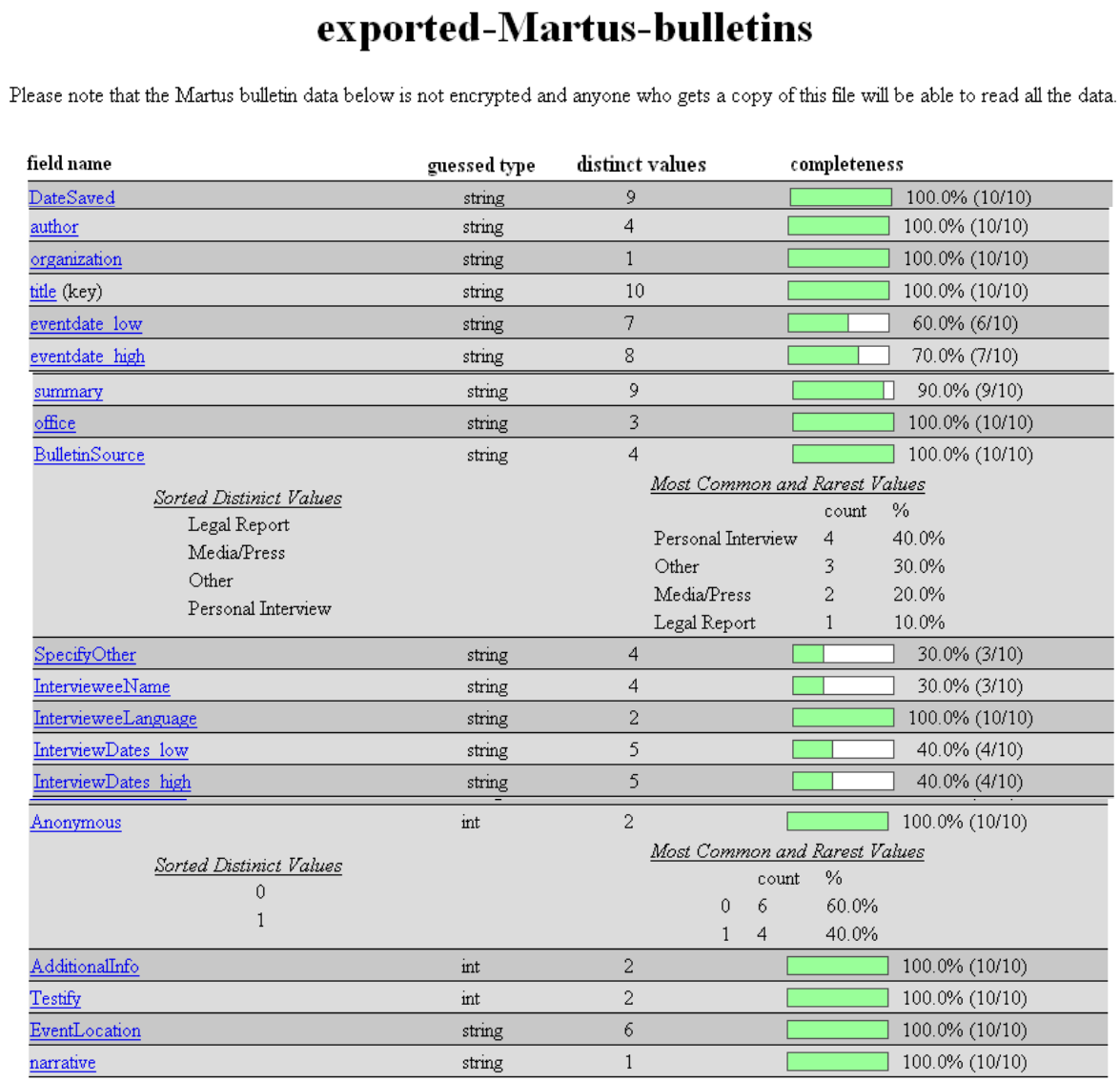
6. In the “HTML Summary Output Directory” field, browse to the location on your computer where you would like the Martus Data Summarizer output (.html file(s)) to be created.
7. In the “Number of example values from individual bulletins to show” field, choose the number of values you want displayed for each field in the Martus Data Summarizer .html output. For example if you have 100 Martus bulletins, but only want to show the top and bottom 5 values for each Martus bulletin field, you would enter 10 in this field. If desired, you can choose the “values from all bulletins” option to display the field values from all the bulletins, however we recommend against using this option as it will show all your Martus bulletin data in unencrypted format which is insecure because anyone can read the content (and will make the output files very large as well).
8. If desired, you can also check the box “Also output data as CSV” to save data as a “flattened” CSV file and browse to the location where you would like the CSV file (or files) to be created. These files can be loaded into spreadsheets or other tools for further analysis. Please see FAQs on how to open output .csv file(s).

9. Click on the “Summarize Data” button to create the output file(s). Each Martus Data Summarizer run will create a main output file that shows all the fields that exist in any of your Martus exported bulletins. If your Martus exported bulletin data has grids (tables) in it, the Martus Data Summarizer will create additional output files for each grid. For more information about grids, see section 10b (“Customizing bulletin fields”) in the Martus User Guide (available at https://www.martus.org/downloads/, or in the Martus\Docs directory of your Martus installation).

If you receive an error screen when running the Martus Data Summarizer, please take a screenshot of it and send it, along with an explanation of any details of the steps you took, to help@martus.org.

4. *Sample Martus Data Summarizer output*

The following screenshot shows an example of a Martus Data Summarizer html output file with the number of example values to be shown set at 10.



exported-Martus-bulletins

Please note that the Martus bulletin data below is not encrypted and anyone who gets a copy of this file will be able to read all the data.

field name	guessed type	distinct values	completeness
DateSaved	string	9	100.0% (10/10)
author	string	4	100.0% (10/10)
organization	string	1	100.0% (10/10)
title (key)	string	10	100.0% (10/10)
eventdate_low	string	7	60.0% (6/10)
eventdate_high	string	8	70.0% (7/10)
summary	string	9	90.0% (9/10)
office	string	3	100.0% (10/10)
BulletinSource	string	4	100.0% (10/10)
Sorted Distinict Values: Legal Report, Media/Press, Other, Personal Interview			*Most Common and Rarest Values*: Personal Interview 4 40.0%; Other 3 30.0%; Media/Press 2 20.0%; Legal Report 1 10.0%
SpecifyOther	string	4	30.0% (3/10)
IntervieweeName	string	4	30.0% (3/10)
IntervieweeLanguage	string	2	100.0% (10/10)
InterviewDates_low	string	5	40.0% (4/10)
InterviewDates_high	string	5	40.0% (4/10)
Anonymous	int	2	100.0% (10/10)
Sorted Distinict Values: 0, 1			*Most Common and Rarest Values*: 0 6 60.0%; 1 4 40.0%
AdditionalInfo	int	2	100.0% (10/10)
Testify	int	2	100.0% (10/10)
EventLocation	string	6	100.0% (10/10)
narrative	string	1	100.0% (10/10)

For each field in the Martus bulletin input file, the Martus Data Summarizer shows:

1. The name of the field from the Martus bulletin.
2. The type of field (e.g. “string” for text fields, “int” for integer number fields, etc). Please note that checkboxes in Martus will be displayed in the Martus Data Summarizer as “int” type, since the Martus bulletin export file shows a “checked” checkbox as “1” and an “unchecked” checkbox as “0”.

3. The number of distinct values entered in the field. For example if you have 10 bulletins, some fields might have a different value for each bulletin (i.e. 10 distinct values), others might only have 1 value (i.e. all bulletins had the same value in this field, others might have 2 values (i.e. 0/1 for a checkbox), and still others may have 3 – 9 distinct values depending on what data was entered for that field.
4. A "Completeness" field that shows how often the field had data entered in it. This is very useful when doing a data quality assessment, as you may want certain fields to be entered a high percent of the time. If you are seeing that fields that in which you want to have data entered have a less than 100% completeness value, you might consider setting those fields to be "Required" in Martus. For more information about setting fields to be Required (and other data validation settings), see section 10b ("Customizing bulletin fields") in the Martus User Guide (available at https://www.martus.org/downloads/, or in the Martus\Docs directory of your Martus installation). Please note that a high level of completeness does not state anything about the quality of the data entered, just that there was SOMEthing entered, so we recommend you drill down to see examples of the data entered, as described below.

For each field, you can click on the field name link, and see more details about the values entered in each field. For each field, you can see:

5. The "Sorted Distinct Values", which shows the values entered in any of your bulletins for this field, sorted alphabetically. Note that the number of values displayed will not exceed the number you entered in the "How many example values to show" field on the initial Martus Data Summarizer screen. In the above screenshot, you can see that the BulletinSource field had 4 different values entered across the 10 Martus bulletins that were exported. Please note that if the number of values entered in a field in your bulletins is larger than the "How many example values to show" field on the initial Martus Data Summarizer screen, only the first and last values (sorted alphabetically) will be shown, instead of all the values that were entered.
6. The "Most Common and Rarest Values", which shows the values entered in any of your bulletins for this field, sorted by how often they were entered. So the most frequently entered values will be at the top of the list, and the least frequently entered values will be at the bottom of the list. The Martus Data Summarizer shows both the count of and percent of time (across all your bulletins) for how many times a value was entered. Note that the number of values displayed will not exceed the number you entered in the "How many example values to show" field on the initial Martus Data Summarizer screen. In the above screenshot, you can see that the BulletinSource field had 4 different values entered across the 10 Martus bulletins that were exported, and "Personal Interview" was the most commonly entered value (at 40% of the time) while "Legal Report" was the rarest value (at 10% of the time). Please note that if the number of values entered in a

field in your bulletins is larger than the “How many example values to show” field on the initial Martus Data Summarizer screen, only the most frequent and least frequent values (sorted by number of times the value was entered) will be shown, instead of all the values that were entered.

5. Frequently Asked Questions

1. Can I get Martus Data Summarizer output in a different format than html?

Yes, in this version of Martus Data Summarizer you have the option to create a flattened .csv data output file in addition to html format file. Multiple .csv files will be created if the exported bulletin data contained grids. These files can be loaded into spreadsheets or other tools for further data analysis. Please see the FAQ below on how to manage these files.

2. Is the Martus Data Summarizer available in languages other than English?

This version is only available in English, but we have plans to translate the Martus Data Summarizer into other languages in future versions.

3. How do I open any .csv output files for further analysis?

Please note that the .csv output files will be delimited NOT by commas (which can cause issues if there are commas in bulletin text), but instead by "pipes" which look like this: |

If you want to load the .csv into a spreadsheet, we recommend that you use OpenOffice instead of Excel to open the csv file and specify the delimiter when you open the file (uncheck "comma" and pick "other", and then enter | in the box).

If you want to open it in Excel, you need to open a blank spreadsheet and then go to the Data tab, pick the option to get data "From Text", select the CSV file you wish to open, and select to use | as the delimiter (instead of Tab or comma).

If you want to load the .csv into other tools, you should be able to specify the delimiter to do so.