

Importing

Introduction

No database is ever an island. Data needs to be exported for use in other programs. Sumac accommodates this need by having an Export button in every list.

Data also needs to be imported. Sumac's Utilities menu has an Import command that lets you import any kind of data into Sumac.

Usually data coming from an external source needs to be cleaned up and reformatted. The Utilities menu also has an Import Workspace command that provides powerful tools for cleaning up and formatting data.

This lesson discusses how to clean data from an external source, then import it into your Sumac database.

Before viewing this video, you should view:

Lesson 1 *Sumac Basics for All Users*

Lesson 3 *Single Contact Record*

Lesson 4 *Contacts List*

Cleaning up Data

Here is the scenario: You recently held a fundraising event and received donations by cheque from attendees. You captured information from registration sheets at the door, along with the amounts that they donated, and put it all into a spreadsheet. You wish to import all this information into Sumac.

use a file of sample contact and donation data

The data is in a spreadsheet. Use your spreadsheet program to **save the spreadsheet** as a tab-delimited text file. The exact commands used for this vary depending on the spreadsheet program you use.

Now use Sumac. Choose the Import Workspace command from the Utilities drop-down menu. Import Workspace asks you to open a file. Choose the tab-delimited text file that you saved a moment ago.

Two Things to Do On Opening A File

The first thing to do when you open a file is to **specify how many header rows** are present. Sumac shows the header rows in grey, and assumes the rest of the rows are data. Operations performed on the data will not affect the header rows. For example, if you

sort by a particular column, the header rows remain unaffected.

The second thing to do is to Merge Split Records. The spreadsheet that you converted may have had single cells that contained multiple lines of data – there were paragraph breaks inside a single cell. When you convert the spreadsheet to a text file, these paragraph breaks split the single record in the middle of the field that held the paragraph break character. This makes the single record look like two or more records. So you need to put the two records together again.

Scroll to find a broken record. Control-click (or command-click on a Macintosh) to select two consecutive lines, then click Merge Split Records to put the broken record back together again.

Scroll to confirm that there are no other split records. Depending on your data, you may need to merge additional split records.

Window Overview

You have opened the file and done these two important tasks, so now stand back for a moment and look at the Import Workspace.

The Import Workspace window is similar to a spreadsheet. Click to edit any cell. When you are editing a cell, you can use Enter to move down, shift-enter moves up. Use Tab to move to the right, shift-tab to move to the left.

If you often use spreadsheets, you know that one of the risks of editing data is that you might accidentally move cells up or down, breaking up a record which occupies a single row. Sumac always keeps the contents of each row together, so you don't need to worry about this problem when you use the Import Workspace.

If you need to select a row, use control-click.

The scrolling list of data is splittable. Often people's names are in the left hand columns and you want to be able to show these names while scrolling horizontally to see other columns of data. **remove splitter**

Sumac enables buttons that apply and disables those that do not, usually based on what is currently selected. So if you click to select a column, buttons that operate on a single column are enabled. If you select two adjacent columns, then buttons that operate on two adjacent columns are enabled.

The buttons are roughly grouped by function. Some apply to managing files, some process columns of data, and some process rows of data.

Let's examine these buttons while cleaning up your data.

File Management

When you click Open to get another file, you have the option to replace or append to the data that is already showing. If you have several files with the same column layout, you can append them together into one file, and perform data cleaning operations on one larger set of data rather than repeating the operations for several smaller files.

As you edit data, you should save it regularly. The Save As button helps by adding a version number to the end of the file name, and increasing it each time you save the data. This lets you keep several versions as you perform your editing work. Usually, at the end when all the data has been imported and checked, you can throw out the intermediate files. But occasionally you discover that you made an error and want to go back one or two versions to repeat or change some editing operations.

Now, let's clean up the data.

Split Column

The Split Column button lets you split one column into multiple columns. Your data file has people's first and last names entered in one column, separated by a comma. You must break this into separate first and last names. Select the column to be split, click the Split Column button, and specify the character that marks the separation between the columns. Put appropriate titles on the columns. **first and last name in one column**

Delete Empty

Delete Empty is used to delete rows that have no value in a particular column. For example, if you are importing contacts, and some rows have no last name, you cannot import them. Click to select the last name column, then click Delete Empty to delete these rows. **some rows have no contact identification**

Delete Rows

While getting rid of unneeded rows, you may notice some at the end of the file that are commentary, not real data. Select them then click Delete Rows to delete them. **remove rows of notes at the end of the file**

Apartment Numbers

The Apartment Numbers button analyzes a column of street addresses and separates apartment numbers out of them into a separate column. **expand column to show there are apartment numbers** The advantage of this is that Sumac can then format postal mailing addresses using country-specific settings. Click to select the street address column, then click the Apartment Numbers button. If Sumac finds apartment numbers, it creates a new column to hold them. Give the apartment number column a title. **street**

addresses with apartment numbers

Change Case

This data has city names in upper case letters. This is hard to read, so you should switch it to mixed case. The Change Case button lets you convert such data into more readable mixed case. Click to select the column of city names, then click to convert them to Title case. **cities in upper case to mixed case**

Validate

If you know that a column should contain a certain type of data, click to choose a column, then use the Validate menu to specify what is supposed to be in the column. Sumac analyzes each row and selects those that are incorrect.

Let's check postal codes. Sumac understands the formats of five and nine digit U.S. zip codes, and also the format of Canadian postal codes. Note that if you have a postal code outside of North America, Sumac may consider it to be incorrect but it does not prevent you from importing it.

Click to select the column of postal codes, then choose from the Validate menu to validate them. Click to show the incorrect rows.

After you have validated and found some incorrect postal codes, Sumac can help you fix them. Often when U.S. zip codes go through a spreadsheet, they are treated as numbers and lose leading zero digits. Common errors in Canadian postal codes are substitution of a similar looking letter for a digit and vice versa. Click the column holding the postal codes, then click Fix Postal Codes, and respond to each change that Sumac proposes.

bad U.S. and Canadian postal codes

Click to show all the rows of data.

Multi Fill Down

Sometimes a spreadsheet holds data that is presented for convenient reading. Repeating values have been omitted. Your list has a column for state or province, but the state was entered only sometimes. You can copy and paste repeatedly to fill in the cells. However Multi Fill Down does this on a whole column with the click of a button.

Choose the city column and click to sort by city. Now make sure that MI (Michigan) is on the first Detroit entry, and that ON (Ontario) is on the first Windsor entry. Then click Multi Fill Down. Sumac fills in all the states. **several people in same state**

Another use for Multi-Fill Down is the add data that was not in the file at all. Let's add a column to hold countries. Then enter U.S. and Canada once in the column. Use Multi-Fill Down to fill the column. **enter countries**

Phone Numbers

The Phone Numbers button operates on a selected column to standardize the format of phone numbers, insert a three-digit area code in front of phone numbers that are seven digits long, and move extensions to a separate column. Choose the column holding the phone numbers, then click the Phone Numbers button. It looks like there are proper area codes for the Detroit phone numbers but not for the Windsor ones, so specify 519 as the default area code. Note that some extensions were put into a separate column. Enter the title for the extensions column so you don't forget what it holds. **phone numbers with extension numbers and no area code**

New Empty Column

In the data, there is an email address column. Click to select it, then validate it. Most of the email addresses are correct, but one has two email addresses in it. Sumac contact records have room for multiple extra email addresses, but they are stored in two different fields so the values need to be separated into two columns.

Click to select the email address column, then click New Empty Column. Put a title on the new column **email 2**. Then manually split the extra email address into the second column. Now re-validate both email address columns to make sure that the split was done correctly.

multiple email addresses in one column

Duplicate Column

Validate the currency amounts. Sumac removes extraneous punctuation and ensures that there is a valid number in each cell. **amounts with currency figures**

You are importing donations, and the incoming data has only one currency amount which is for both total donation and receipt value of the donations. Sumac requires two separate values, even though they may be the same. Click to select the column containing currency values, then click Duplicate Column. Sumac adds an additional column to the file and copies the selected column values into it. Put appropriate titles on the columns. **single currency column becomes two**

Delete Column

Use Delete Column to delete a selected column. Often imported data contains columns of information that are only useful to the system that originated the data. For tidiness, and to simplify the data you are looking at, you can remove unnecessary columns as you clean up the file.

The data file being imported has a column that indicates registration status for the event.

This column is not needed for importing donation and contact information. So click to select it, then click Delete Column. **remove registration status column**

Move Columns, Append Columns

Append Columns is often used when you need to append multiple columns of notes and put them into a single notes field.

Your data file has cheque numbers and notes for the donations. But Sumac donation records do not have a separate field to hold cheque numbers, so you should put both the notes and the cheque numbers together into the notes field. Let's put this information into the Notes field in the donations records.

First, use the column moving buttons to move the columns into place, the notes on the right and the cheque numbers on the left.

Click and control-click to select the two columns of interest. Then click Append Columns.

Sumac asks you what characters should be inserted when there is data in both columns. Often a semi-colon and space are appropriate. In this situation, you can put semi-colon and a space between the columns, separating the cheque numbers and notes. **cheque number and notes**

Find and Select

If you are curious about the records from Windsor, click to select the city column. Then click Find and Select. Enter the text to be found **enter Windsor** then click OK. Sumac selects any rows that contain the specified text in the chosen column.

Then you can use Hide Selected to hide the Windsor records or **use Show Only Selected** to get a list of only the matching rows. When you are done examining or editing the matching rows, click Show All to see all records again. **some contacts are in Windsor**

Assign IDs

The data file is being used to create contact records and also donation records. Once the data is cleaned up, you will import the new contacts, then import the donations.

The easiest way to connect the donations with their corresponding contact is by assigning ID numbers. If a contact is added to the database, Sumac automatically assigns a unique ID number. With this data, it will be convenient to assign IDs in advance, so when you import the donations you will be able to link them to known contact IDs.

You want to ensure that you don't assign a contact ID that is already in use. So look in the contacts list, and sort to find the highest ID number. It is xxx, so use xxx + 1 as the

first new contact ID number.

Click Assign IDs, and specify the first ID to create. Sumac adds an ID column on the left side of the data. Give it an appropriate column title. **need IDs for contacts**

Save Cleaned Data

Now **save** the clean contact and donation information. Sumac appends a version number to the end of the file. It is a good idea to save frequently so that if you make a mistake you can revert to a previously-saved version of the file.

Other Buttons

There are a few buttons that have not been used in cleaning up this spreadsheet.

Get Contact IDs

The Get Contact IDs button looks up contact names in the database, and shows you their IDs. If you know the contacts are already in the database, it is sometimes convenient to work with contact IDs instead of names.

Script

The Script button lets you tell Sumac to apply a file of editing commands to a particular column. Scripts can do lookup and conversion operations that change one set of values to another set of values. For details, see the *Sumac Users Manual*.

Group Duplicates

If your data file may contain duplicate records, click to select the columns which need to be the same for two rows to be considered identical. Then click Group Duplicates. Sumac moves duplicate records to the start of the data file, and highlights them. You can decide if you need to resolve the duplicates, or if they are okay just as they are.

Remove Duplicates

The Remove Duplicates button is similar to Group Duplicates, except that instead of just *showing* you duplicate records, it actually *merges* each group of duplicates to create a single record. One use of Remove Duplicates is to get all the unique values for a lookup list. For example, if you have a file of donations, you can remove duplicate records based on the donation type column, thereby producing a list of all the distinct donation types.

Contact Names

If you want to check to see which contacts, identified by first and last name in a spreadsheet, are already in the database, click and control-click to select the columns holding first and last names. Then click the Contact Names button. Sumac looks up all

the contacts, and selects those who are not in the database.

Importing

Now that you have cleaned up and organized the contact and donation data, you can import it into Sumac.

Only certain users are allowed to import data. If you are going to import data, your Sumac administrator must give you the Bulk Import capability in your user profile.

You should import the contacts first, so the donations can then be imported and connected to their corresponding contacts.

Choose the Import command from the Utility menu. Indicate that you are importing contacts, and that they are new. Choose the data file.

Import Contacts

The data file appears on the right hand side. You must tell Sumac which contact record fields should hold which data. The left hand side of the window holds a list of all the fields in contact records.

You match field names to data columns by dragging field names and dropping them onto the data columns. **match all the columns**

Now that all the contact columns have been matched, you can validate the data. Until you validate your data, you cannot import it. Click Validate to tell Sumac to check all the data.

Then click Import. All the contact records are imported into Sumac.

Import Donations

Now that you know all the contacts are in the database, import the donations. Choose the Import command again, indicate you are importing donations, and choose the same file.

Line up the fields in the donations.

There are some fields that are not actually in the file but which you know. So you can manually enter them. **enter event, date**

Click Validate, to ensure everything is right, then click Import to import the data into Sumac.

Import Communications

Since you know that all the contacts in our file attended an event, you can create communication records that indicate they did so.

Choose the Import command, indicate you are importing communications, and choose the same file.

Specify values for each of the fields being imported.

Click Validate, to ensure everything is right, then click Import to import the data into Sumac.

Import Done

If you now look at the contacts list, you see that the newly imported contacts have been added to the end of the list. Each has a donation and a communication record.

Conclusion

In this lesson you learned how to:

- ◆ clean up data to get it ready for importing
- ◆ import new contacts into the database
- ◆ import donations
- ◆ import communications indicating attendance at an event.

You should now proceed to other Sumac lessons to learn more about how Sumac can help you every day.