

STAT:5400
Lab session 4
Introduction to LibreOffice Base

Oct. 16, 2015

1 LibreOffice

LibreOffice is a free and open-source office software suite, comprised of modules for word processing, spreadsheet, database management, presentations, drawing, and equation editing. It reproduces much of the functionality of Microsoft Office and can read and write many MS Office file types, including .doc and .docx, .ppt and .pptx, and .xls and .xlsx. It runs under recent versions of Linux and Mac operating systems, as well as Windows.

Documentation is available at <http://www.libreoffice.org/get-help/documentation/>.

LibreOffice was developed by The Document Foundation as an offshoot of OpenOffice.org in 2010. See <http://en.wikipedia.org/wiki/LibreOffice> for a history of OpenOffice.org and LibreOffice.

When used to describe computer software, the word “libre” means that the software may be used, modified, copied and shared without permission from the copyright holder.

Note for Ubuntu Linux users: OpenOffice.org was part of the standard Ubuntu distributions through Ubuntu 10.10 (Maverick). However, the Base component didn't come with it and had to be installed separately (for example, by using the Synaptic package manager). Beginning with Ubuntu 11.04 (Natty), the LibreOffice suite has been included instead of Open Office. Several other Linux distributions (including Fedora, which is on the DIVMS network) have made the same decision. Now that Apache has taken over Open Office, it is possible that the Linux distributions may switch back.

2 Creating a database in LibreOffice

Before diving into the software to create a database, we must plan the structure of the tables and relationships. We wish to create a database to store some information about patients in a clinical trial. Some information about each patient is stored only once: patient id, birthdate, date of randomization, treatment assignment. Other information (CD4 counts) is measured and recorded many times for each patient: date of blood draw, CD4 count. We will create two tables within our database and establish a one-to-many relationship between them.

2.1 Creating the database

Create the subdirectory in which you wish to place the files for this lab. To start Base on our Linux network, click the Application Menu tab and select “Office/LibreOffice Base.” Alternatively, you can go into the subdirectory and type `oobase` at the Linux prompt. (Yes, it's still `oobase` even for LibreOffice.) Either way, the database wizard will come up.

Make sure that “Create a new database” is checked, and click “Next.”

On the next screen – “Deciding How to proceed after saving the database” – make sure that “Yes, register the database” and “Open the database for editing” are checked. Registering the database makes it accessible to the other components of OpenOffice.org for exchange of data. Click “Finish.”

You will be prompted for the filename of the database and the folder in which to save it. The only downside of registering a database is that it makes it tricky to move or rename it later, so specify the name and path with care. Then click “Save.”

If you get an error message about a problem with Java at any time during the above process, use the pull-down menus to select Tools/Options/Java. Then make sure that the button for “Free Software Foundation” is ticked. Click OK. Then exit out of LibreOffice, and start over.

2.2 Designing tables

On the next screen, click “Tables” in the left hand panel. Notice that a different list appears under “Tasks.” Click “Create table in design view.” After a pause, a spreadsheet-like screen comes up that enables you to define fields for the table.

We will begin with the Patients table. We want a unique character patient ID number as the primary key for this table. Enter “Patid” under “Field name.” Tab to the “Field type” column. Use the pull-down menu to select “Textfix [CHAR].” A new section opens at the bottom of the window to allow you to enter “Field properties.” Set “Entry required” to “Yes” and length to 5. Next create a field called “Birthdate.” Set the field type to “Date.” In the “Field properties” section, click in the small second box after “Format example,” and select “12/31/1999.” Create another field called “Randdate” with the same characteristics. Finally, create a field called “TrtGroup.” Set it to type “Text(fix)-CHAR” and to length 1. To make the “Patid” field the primary key for this table, right click in the box to the left of its field name, and check “Primary key.” If you forget to do this before saving the table, Base will give you a message that there is no primary key.

To save this table structure, click “File” and “Save.” Specify the table name “Patients” and click “OK.” When you get back to the main database window, you must click “File” and “Save” again to save the new table in the larger database structure. This two-step save process must be carried out for each new table, form, relation or whatever.

Now we will create the CD4 count detail file. It will have its own primary key, a “CPatid” field as a foreign key for linking back to the Patients table, fields for the

date of blood draw and the measured value. Setting the CID field to AutoValue will cause Base to automatically enter unique numbers.

| | | |
|--------|---------------|----------------------|
| CID | integer | AutoValue = Yes |
| CPatid | Text fix | Entry required = Yes |
| CDate | Date | 12/31/1999 |
| CD4 | small integer | |

Designate CID as the primary key if it isn't already. Save and close the Design window, and save the new table.

3 Defining the relationship between tables

Next we need to create the link between the Patid fields in the two tables. Click "Tools" and "Relationships." Add each of the tables by clicking its name and then clicking "Add." Then close the Add tool. You will get a window with all the fields of each table listed. Click the "New Relation" icon. This opens the Relations window. Click on the CD4 table name in the Tables Involved section, then on the CPatid field name. Then select Patid from the dropdown menu for the Patients file. Tick "Update cascade" and "Delete cascade," and then click OK. You should see a little zigzag line connecting the CPatid and Patid fields in the two table listings. Save and then close the Relation Design window, and save in the main database window.

4 Data entry

4.1 Creating a data entry form

In the main database window, click "Forms" in the left panel, then "Use the wizard" in the Tasks window. Select the Patients table. Then select all the fields from the Patients table by clicking >>. Click "Set up a subform." Check "Add Subform" and "Subform based on existing relation." Click "CD4" as the relation, and click "Next." Add all the fields, and click "Next."

The next step is "Arrange controls." It allows you to rough out the style in which the fields will be presented for data entry in the form and subform. Choose what you like, and click "Next."

Leave the default data entry mode "This form is to display all data."

Choose the style and color that you like.

Name the form "Complete patient." Choose "Modify the form." Click "Finish," close the form, and then save your work in the main database window.

We also need to create a form for entering CD4 data for existing patients.

4.2 Entering data

Left click on the Complete Patient data entry form that you created. Enter data in the fields at the top of the form. When you click in "CPatid" in the CD4 subform area, Base will automatically fill in the value from the top of the screen. When you have entered all data, find the "Save record" icon at the bottom of the screen and click it. Find the "New record" arrow. To add more CD4 data for the same patient, click it with your cursor in the subform. To start entering a brand new patient, put your cursor in the Patid field before clicking the "New Record" icon.

5 Creating a query

We will create a query that pulls data from both tables, using the relationship. It can then be used to view the data, or as the basis for a formatted report. Click "Queries" in the left panel, then "Create Query in Design View." Add both the Patients and the CD4 tables to the query. Close the Add window. You should see a window with a box for each table, and the relationship between them. Below is a spreadsheet in which you will specify the contents of the query. Click in the Field row of the first column, and select Patid from the Patients file. Move across the columns, selecting the desired field from the desired table for each one. Be sure that "Visible" is checked for all the fields. Use the Sort row to set the Patid and Cdate fields to Ascending sort order.

Save and name your query, then Save when you get back to the database window.

To display the entire contents of your linked tables, double-click on the name of your query. Close it when you're finished looking at it.

To display only a subset of your records, right click on the query name, and choose "Edit." Go to the "Criterion" row and enter a value to match in one of the fields. Save your query, and double-click its name in the database window.

6 Creating a report

Base can create reports based on linked tables. Click Reports in the left panel, then "Use Wizard to Create Report."

7 Exporting data from Base

The following instructions are taken from http://help.libreoffice.org/Common/Importing_and_Exporting_Data_in_Base

"You copy a table from Base to a new Calc sheet, then you can save or export the data to any file format that Calc supports.

1. Open the database file that contains the database table to be exported. Click Tables to view the tables, or click Queries to view the queries.

2. Choose File - New - Spreadsheet.
3. In the Base window, right-click the name of the table to export. Choose Copy from the context menu.
4. Click cell A1 in the new Calc window, then choose Edit - Paste.

Now you can save or export the data to many file types.”

Try exporting the Patient table to a Calc spreadsheet. Save the spreadsheet in .csv format. It could then easily be imported into R.

8 Importing data into Base from other software packages

The following instructions are taken from http://help.libreoffice.org/Common/Importing_and_Exporting_Data_in_Base

You can import text files, spreadsheet files, and your system address book in read-only mode only.

When you import from a text or spreadsheet file, the file must have a first row of header information. The second row of the file is the first valid data row. The format of every field in the second row determines the format for the entire column. Any format information from a spreadsheet file gets lost when importing to Base.

For example, to ensure the first column has a text format, you must make sure that the first field of the first valid data row contains text. If a field in the first valid data row contains a number, the whole column is set to number format, and only numbers, no text, will be shown in that column.

1. Open a Base file of the database type that you want.
Either create a new Base file using the Database Wizard, or open any existing Base file that is not read-only.
2. Open the Calc file that contains the data to be imported to Base. You can open a *.dbf dBASE file or many other file types. Preferably, the field names should appear in the first row of the Calc spreadsheet.
3. Select the data to be copied to Base.
You can enter a range reference like A1:X500 in the Name Box if you don't want to scroll.
If you copy a dBASE sheet, include the top row that contains the header data.
4. Choose Edit - Copy.
5. In the Base window, click Tables to view the tables.

6. In the Base window, choose Edit - Paste.
7. You see the Copy Table dialog.
Most databases need a primary key, so, unless one of the existing fields is to be the primary key, you may want to check the “Create primary key” box.
If you have copied column names as well as data values, tick “Definitions and data” and “Use first line as column names.”
Click “Next.”
8. On the “Apply columns” screen, copy the names of all columns that you want to include in the table into the right-hand window. Click “Next.”
9. In the “Type formatting” window, specify the formats for each field. If you are using an existing field as the primary key, left click on its name. Then right click the name and check “Primary key.” Click “Create” when you are finished with formats for all fields.

Download the .csv file called “software.csv” from the Datasets section of the course web page. Start LibreOffice Calc (from the Applications menu or by typing `oocalc` at the Linux prompt). Read the csv file into a spreadsheet. Then use the instructions above to import into a Base table (either in the patients database or in a new database).