In the context of a sudden crisis, such as the COVID-19 crisis, there are often a lot of surveys and hence databases that are created and launched in a hurry to assess the populations’ needs and address them quickly in the field. This means that there will be many Excel files being sent around, with different versioning and privacy levels. A component at stake here becomes the size and performance of the database, and the potential analysis file, due to a rather common lack of or the bad quality of the internet connection in the field. The need for it to be sent, reviewed and finalized involves a lot of steps, for which the file should be at its best performance and minimal size.

This tutorial will help you undertake a few changes in your process, in order to improve the performance of your Excel databases and facilitate analysis when you need rapid insights. Implementing these reflexes within your organization may take some time at first since it will force you to think through your whole process but, in the end, you should gain a lot of time. Those changes are simple and easy enough to implement, and it will help your organization harmonize its practices and get more efficient.

Tips for setting up a database which facilitates your analysis

1. Have an idea of the desired results before starting the project

   Design the questions/fields according to the desired analysis

   You need to collect data for which you will perform analysis on! There is no need to collect additional data or to collect data “just in case” as it will increase for no good reason the size of your database. This will ensure that what you collect is minimal and proportionate to your needs (in accordance to data protection principles) but also that your database remains the most efficient possible.

   Limit respondent fatigue

   - Ask essential questions
   - Add interviewer observations
Pay attention to the data collected

- Limit sensitive data collection
- Provide for categorization directly at the time of collection, to facilitate any analysis afterwards

For instance, ask a choice between age ranges rather than an open numeric question, if you are sure that your categories will not change.

Choose the relevant analysis for the different variables

Make sure the data collected will actually be possible to analyze and it will make sense. The best way to do so would be to prepare an analysis plan ahead of the collection!

If you are using Mobile Data Collection (MDC) to import data, include calculations directly in the questionnaire

- Add automatic calculations such as scoring
- Insert calculation checks for errors

This will allow you to simplify your survey, have a quicker analysis following the data collection and reduce the need of formulas in your database.

Key words: simplicity and clarity

Clarify everyone’s needs from HQ to the field by asking yourself:

- Will we re-use this analysis file in other cases?
- Who is involved in the process and who will need to have access to the results of the analysis?
- How complex does the analysis tool need to be?
- Will I use a final dataset or will I need to feed the dataset later on with new information?

Having an analysis tool is useful for reporting to HQ but also for improving operations’ work in the field. It is very important to involve all relevant parties within your organization to develop a tool that can either fit or be adapted to everyone’s needs. In order to do so, you will have to communicate with various staff members on their specific needs for analysis, and come up with a collaborative solution.
Simplify the setup of the tool and provide clear guidance on how to use it

This will make sure that your team will be able to use it, and pick it up quickly. For instance:

- Use formulas that your team is able to replicate in case of bugs, or
- Protect fields with complex formulas from modification,
- Mark clearly places (with color coding for instance), where actions on the tool are needed, as well as the cells not to be modified.

Simplicity in the desired analysis

To stick as closely as possible to the needs identified in the field, there is often no need to create a big dashboard with loads of graphs that will never be used for the project.

3 Anticipating the need for analysis tools by building competences

Develop the technical skills of the teams over time to be able to deploy a rapid survey analysis tool when a crisis occurs:

- Have them follow webinars
- Allow training time

Good practices to adopt

1 When creating your database in Excel, use simple functions that will facilitate analysis

Use the table function

- To repeat formats and formulas
- To autocomplete cells when adding a new line
- To have absolute references in formulas

Where to find it?
Menu Insert > Tables > Table
Maximize the use of the name manager function

- To facilitate formulas
- To simplify referencing
- But be careful not to have useless connections or duplicates

Use functions to simplify data entry and gain time

Implement functionalities when necessary, among those:

- Dropdown lists
- Data entry constraints

Where to find it?
Menu Formulas > Defined names > Name manager
Menu Data > Data tools > Data validation
• Conditional formatting **ONLY WHEN NECESSARY** because it’s heavy when there are too many of them

Where to find it? Menu Home > Styles > Conditional formatting

Protect sheets / files when possible

Where to find it? Menu Review > Changes > Protect sheet / Protect workbook

Use manual calculation mode

It's not just where there is a large volume of data, but when a spreadsheet gets beyond a certain size it becomes slow. Depending on the complexity of your spreadsheet (number of formulas, quality of the internet connection, etc.), this may occur quicker than expected.
To control when Excel does the calculations
To save time by not having calculations done over and over
Especially if there are a lot of formulas and/or data

Where to find it?
Menu File > Options > Formulas > Calculation options

3 Minimize the size of the final file (for improved performance)

Delete formulas that are no longer relevant

If certain formulas are used once, convert them later to simple values, if there is no need for them to be "dynamic" or updated.

Do not use relative references in the formulas (cells) but rather use direct references set up in the name manager or use columns names set up in the table.

A relative cell reference is having a reference to a cell by simply clicking on it or by entering its "coordinates", for instance G2 (when you insert formulas and you refer to a certain cell with relative reference). From the time you will need to copy and paste the formula or drag it across rows or columns, it will move accordingly, by keeping the relative position of the G2 cell to the initial cell. The issue is when you have to repeat the formula over multiple columns or rows, having formulas referring to specific cells, then it is recommended to use absolute references or names.
Take a good look at the size of the file before and after any significant modification (adding a pivot table, copying a large number of formulas, etc.)

Use more efficient formulas:

- Use **IFERROR** instead of **IF** and **ISERROR** combo
- Use **MAX(A1,0)** instead of **IF(A1>0,A1,0)**
- Use the **INDEX/MATCH** combo, instead of **VLOOKUP**
- Use — **(double negatives)** to convert TRUE’s and FALSE’s to 1’s and 0’s **(instead of multiplying it by 1 or adding 0 to it)**.

Limit complex formulas

- Array formulas
- Chains of formulas between multiple sheets and/or elements

Limit the link between files

---

It’s important to have a filing system well organized and good reflexes for versioning the files you’re working with

Raw vs ongoing vs final dataset

- Make a copy of the raw dataset you have/received before you modify it in the “raw” folder
- When you are in the process of cleaning and / or building your database, keep your files (with various versions) in the “ongoing” folder
- Once you have your finalized file, you can put it in the “final” folder

This separation allows you and your team to differentiate the working documents, and make sure the tracking is up to date. You can also go back easily to a previous version if needed.
Do not modify / move / rename files and folders unless it is absolutely required

Else, connection of queries, tables, etc. should be updated when refreshing connections: it’s a complex and unnecessary process.

In the file name, make sure to include:

- The date of the last update, for instance 20210105
- The initials of the person who modified it, for instance CGA
- v# for the number of the version
- A significant change in the file implies a new version
- And separate the elements with an underscore

**Example: 20210105_food-security-db_CGA_v2.xlsx**

This name tells us that this Excel file is the second version of a database about food security that was last edited on the 5th of January, 2021 by C. Gaxxxx. Using a clear filing system will allow you to save time as you won’t need to open the document to check what is inside, which version it is, who modified it last, etc. and limit potential errors (using the wrong file, etc.).

Thank you for reading!