Data entering, data management and QC
Laurent Saint-André, Matieu Henry

To cite this version:
Laurent Saint-André, Matieu Henry. Data entering, data management and QC. Training Workshop on Tree Allometric Equations, May 2014, Colombo, Sri Lanka. pp.7 slides. hal-02792113

HAL Id: hal-02792113
https://hal.inrae.fr/hal-02792113
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Data entering, data management and QC

- L. Saint-André (CIRAD - INRA), and M. Henry (FAO)

Training workshop on tree allometric equations, Forest Department Sri Lanka, May 20 – 24th 2014
How data resulting from inventory and monitoring efforts will be acquired, managed, processed, checked, analyzed, distributed, and archived.

Objective:

To transfer field data into ONE electronic database

Different types of software:
- Microsoft Office
- Access
- MySQL

http://science.nature.nps.gov/im/datamgmt/index.cfm
In order to avoid errors:

To check all the data several times: double entry

Comparison of the two files to detect the errors

To well define the variables to be measured (text of numbers)

Exemple: Geographic coordinates

<table>
<thead>
<tr>
<th></th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lat</td>
<td>25.456</td>
<td>7°28’55.1</td>
</tr>
<tr>
<td>Long</td>
<td>12.567</td>
<td>13°41’25.9</td>
</tr>
</tbody>
</table>
METAINFORMATION

When the measurements were performed

Details of the code and acronyms

Definition of the variables

Height of the measured diameter or circumference etc.

Units of the measured parameters

Any additional information

Crucial to allow the use of the data at different time period and different users
If you have different level of information, you need to have different tables.
This structure of information minimize the repetition of the information.
Objective: Delete the artifacts/ impossible data
The difficulty depends on identifying what is normal or not normal
Very often, the errors come from a change of units
QUALITY CONTROL

Graphic analysis

Descriptive statistics

DATA STRUCTURE

One line for one individual of the sampled population

One column -> one descriptive variable