



# Training on Socioeconomic Monitoring (SocMon) Methodology for Evaluation of Socioeconomics and Marine Resources Utilization at Selected Coastal Communities in Myanmar

Mawlamyine University, Mon State and  
Sakhanthit Village & Thayawthadangyi Island Group, Kyunsu Township  
9-19 January 2014

## Session 10: Coding and Data Management for Household Interview Variables (Coding/Encoding Data using Excel and SPSS)

**Prof. Marissa Pontillas**  
Palawan State University

<sup>1</sup>Prepared by Dr Michael Pido, Prof Marissa Pontillas and Ms Eva Marie Ponce de Leon Palawan State University Puerto Princesa City, Philippines / Dr Robert S Pomeroy, University of Connecticut, USA,

**Preparing the data for analysis is  
the first step in data management.**

# Learning Objectives

1. Transform the questionnaire responses into codes using the coding guide in the field manual;
2. Expand the coding guide by providing additional codes to the interviewees' responses that are not yet in the coding guide
3. Use and complete an MS Excel worksheet for recording coded data in preparation for data analysis;

# Learner's Activities

- ***Perform hands-on exercises on:***
  - ***coding questionnaire responses using coding guide prepared;***
  - ***Expanding coding guide based on interviewees' actual responses to both closed-ended and open-ended questions***
  - ***inputting coded responses in the MS Excel worksheet***

# Expected Outputs

The expected outputs from undertaking the exercises will include:

- ***Final Coding Guide***
- ***Coded Questionnaires***
- ***MS Excel Database for the Variables***

# Introduction

- Preparing the mass of data gathered from the household surveys and key informant interviews for statistical analysis is the next step in data management after the HH/KI interviews have been conducted and recorded.
- The quality of data analysis largely rests on the quality of data gathering, particularly on the accuracy of its recording.

# Introduction

- No sophisticated data analysis procedure can compensate for poor data quality. Hence, the necessity to ensure that all questions are asked completely and interviewees' responses are recorded accurately cannot be over-emphasized.

# Introduction

- The GIGO (“Garbage in, garbage out”) rule for data processing certainly applies.
- Thus, care must be taken that all gathered data are recorded accurately, coded correctly, and organized into a data base.
- A data base facilitates statistical analysis.
- Use a numerical coding guide for ease in recording and inputting data.



# Purpose of the Coding Guide

Makes the life of the data manager easier by:

- simplifying data management
- transforming responses to a database
- for consistency and/or homogeneity
- treatment of variables
- for analytical purposes

# Data Encoding

- Each question in the HHS and KII is identified by a variable name represented by a letter and a number specified in the coding guide.
- For ease in coding, numbers are assigned to the possible responses that may arise for every question.
- For variables that are measured at the nominal scale, the number codes only serve to differentiate one response from the others, and do not have any meaning beyond naming or labeling.

# Purpose of the Coding Guide

Makes the life of the data manager easier by:

- simplifying data management
- transforming responses to a database
- for consistency and/or homogeneity
- treatment of variables
- for analytical purposes

# Data Encoding

- Each question in the HHS and KII is identified by a variable name represented by a letter and a number specified in the coding guide.
- For ease in coding, numbers are assigned to the possible responses that may arise for every question.
- For variables that are measured at the nominal scale, the number codes only serve to differentiate one response from the others, and do not have any meaning beyond naming or labeling.

# Data Encoding

- Ordinal level variables would give rise to measures that can be ranked, as in Likert-type scales, where higher ranks are given higher scores and lower ranks are assigned lower scores. However, the scoring must be reversed for negatively worded statements.
- Always refer to the coding guide to ensure accuracy and consistency in data recording.

# Coding Responses to Open-Ended Questions

- Distill and record the essence of the responses given by the interviewees, so that it would be easy to determine the appropriate code to be assigned to each response.
- Be able to distinguish which answers are similar and should therefore have the same code.
- Ensure that each code is distinct and different from the other codes for the same variable.

# Coding Responses to Open-Ended Questions

- A list of possible responses are enumerated in the coding guide, yet it is still possible that a response would not fall into any one of the categories that were assigned codes. Should this arise, it is necessary to expand the list of answers and assign additional codes to the new categories of responses.
- The coding guide may be further developed to accommodate responses that were not provided in the list.

# Coding Responses to Open-Ended Questions

- Start with a limited number of possible answers assigned with specific codes, and to add other answers (and their corresponding codes) as they arise in the accomplished questionnaire. Then, after seeing the range of responses, the researchers may lump specific responses to general categories.
- There might be cases when an interviewee provides multiple answers. If there is only one expected answer, probe further to determine what is the primary answer, and note down other secondary responses.



# The MS Excel Database

# Recording data in MS Excel Database

For the HHI, the **demographics (H1 – H7 + CC1a)** section does not only gather data about the household head being interviewed but for each member of the household. Thus, a separate worksheet need to be prepared for this portion.

# Recording data in Excel Database

- ❖ Rule of Thumb for inputting the data into the MS Excel file: a column should contain data for only one variable and a row should contain data for only one unit of measure (such as a household head).
- ❖ For purposes of identification, the first row would usually contain the HHI/KII variable names (refer to coding guide) while the first column would list the identification number of the questionnaire.

# EXERCISE on CODING

1. Using a FILLED Household Survey Questionnaire, code responses for questions H1 to H9B using the initial coding guide prepared.
2. Add and assign new codes for responses not included in the initial coding sheet.

# Questions?