Training course to enhance collection of fisheries and aquaculture statistics

Module 6 – Tools to support data collection, compilation and analysis
Introduction

• The success of data collection for SSF and aquaculture depends to a great extent on the quality of the work of data collectors.

• To guarantee the quality of the data collected, it is necessary to regularly train data collectors. Simplifying the tasks of data collectors is essential to ensure the good quality of the data.

• However, staff turnover may hinder the work, because of the absence of trained data collectors and the discontinuity of data collection.

• The use of technological tools facilitates the work of data collectors and supports routine data collections frameworks. It is more advantageous than paper-based data collection in terms of cost, speed of obtaining data and the quality of data.
Module 6 – Tools to support data collection, compilation and analysis

Outline

1. Role of data collectors of SSF and aquaculture statistics
2. Technological tools for data collection
3. Software for SSF data compilation and analysis
Role of data collectors of SSF and aquaculture statistics
Role of staff – OFFICE

• Primary data collected by field staff are of little or no utility unless there is an appropriate statistical office infrastructure

• Responsibilities and functions of statistical office staff:
  – Designing and planning of fishery surveys (implementation scheduling, training, equipment and logistics support)
  – Coordinating and monitoring field and office activities
  – Data processing (compiling and reviewing primary data obtained from the field, data checking, editing, correcting and analysing data)
  – Reporting results of data analysed to stakeholders
Role of Staff – FIELD

• The field staff are the data collectors and their supervisors. They are the primary interface between fishers (and aquaculture-practicing households) and fisheries management

• They collect and submit data to the fishery statistical office for further processing

• The important role of the field staff involved in data collection can be summarized as follows:

✔ Winning and maintaining cooperation of respondents

✔ Ensuring quality of data by correctly carrying out instructions on data collection received in regular training

✔ Providing feedback to office for evaluation
Technological tools for data collection
Technology for data collection and analysis

• Lower-cost, faster and higher-quality survey data collection is now achievable through Computer-Assisted Personal Interview (CAPI).

• CAPI uses tablet devices and cellular or Internet networks to assign interview cases to data collectors in the field, collect interview data and transmit finished interviews to survey headquarters, for real-time quality checking and analysis through automatic survey progress reports.

• Information is ready for statistical analysis as soon as an interview is completed and transmitted to the headquarters.

• This eliminates several steps and the need for printed paper arising in Paper-Assisted Personal Interview (PAPI) and allows data validation at the time of data collection via algorithms that define skip patterns and response quality checks.
CAPI versus PAPI

Studies have proved CAPI-based surveys to be cheaper, faster and of higher quality than PAPI-based surveys. Tablet device prices have continued to fall and the availability of cloud servers has reduced costs and eliminated the need to purchase and maintain own servers, thus lowering significantly the cost of moving from PAPI to CAPI.

<table>
<thead>
<tr>
<th>QUALITY</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple question types</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Error and response quality checks</td>
<td>automatic, high precision</td>
<td>manual, low precision</td>
</tr>
<tr>
<td>Survey progress monitoring</td>
<td>real time over cellular/internet</td>
<td>manual, at end of survey round</td>
</tr>
<tr>
<td>Capturing of sound, video, GPS position, etc.</td>
<td>yes</td>
<td>no, needs external devices</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TIMELINESS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Question skip</td>
<td>automatic, high speed</td>
<td>manual, slow</td>
</tr>
<tr>
<td>Interview case assignment to data collectors in field</td>
<td>electronic, real time over Internet/cellular network</td>
<td>Paper-based, slow by hand</td>
</tr>
<tr>
<td>On field location of interview case</td>
<td>fast, assisted by automatic inbuilt map in tablet</td>
<td>slow, paper-based</td>
</tr>
<tr>
<td>Finished interview transmission to headquarters</td>
<td>electronic, real-time over Internet/cellular network</td>
<td>Paper-based, slow by hand</td>
</tr>
<tr>
<td>Off network operation capability</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COST</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Printing</td>
<td>very minimum</td>
<td>very large</td>
</tr>
<tr>
<td>Paper storage and data entry</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Maintenance of server and tablet devices</td>
<td>free to low-cost cloud servers available, tablets last long if handled with care</td>
<td>no</td>
</tr>
<tr>
<td>Field staff training</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Lower-cost, faster and higher-quality data through real-time tablet-based interview data collection, response quality check, survey progress reporting and data export for analysis...

SERVERS
- Internet-connected
- Must have installed SUSO software (available for free)
- Can be hosted in paid-for cloud server or own local server
- Can be hosted free in survey solutions server at World Bank

NETWORK
- 3G or 4G cellular network
- WiFi internet connection
- Transmission of data from tablet: real-time or once network is available

PROGRAMS
- Other programs and applications
  - Allows data exchanges
  - Facilitates data analysis

PEOPLE
- People with survey experience
  - Supervisors, interviewers with survey experience
  - Basic IT skills
  - Skilled server administrator at HQ

TABLETS
- Android devices for data collection
  - Known Android devices
  - 7- or 8-inch screen
  - GPS-enabled
  - 3G or 4G connectivity
  - Long battery life
  - Charging sources: grid electricity, solar or electricity generator, portable charger, power bank, etc.
Why CAPI for SSF and aquaculture statistics?

• The collection of SSF and aquaculture statistics warrants deploying high-technology equipment to improve the quality, timeliness and cost of obtaining survey data.

• Pictures pre-saved in, or taken by, tablets can improve precision when identifying fish species, gears and vessels during fieldwork, or later validation and analysis of photographs taken during field work.
Why CAPI for SSF and aquaculture statistics?

• Reaching sampled homeports and remote landing sites can be done quickly, using tablets’ inbuilt GPS function.

• The instant transmission of finished interview data while still in the field allows for instant quality checks by the headquarters or office.

• The absence of large-volume paper printing and subsequent data entry, and cloud servers available free of charge, are great advantages that reduce the technological costs incurred by statistics producers.

Benefits of CAPI:

- Improve timeliness of data collection
- Ensure data quality and comparability
- Allow collection of new types of information/data
- Cost-effective, sustainable solution for NSOs
CAPI applications

• Two CAPI applications are worth mentioning:

  Survey Solutions (SuSo) and Open Data Kit (ODK)

• These two applications provide functions for questionnaire design and field data collection based on smartphones or tablet devices. They support all types of questions and capturing of data such as pictures and GPS positions.

• They allow for the automatic transmission of collected data to survey headquarters through mobile Internet network; otherwise, they may collect data off the network and transmit it soon as a network is available. The requirement of transmitting data over an Internet server is also met if the office has its own server or a server subscription to a cloud server, such as the free SuSo server hosted by the World Bank.

• These free applications are maintained and provided to the end users with support and training material.
CAPI applications – SuSo

The Global Strategy to improve Agricultural and Rural Statistics supported the development of the SuSo general online training material primarily regarding agricultural surveys. However, such material could be also tailored to SSF and aquaculture data collection.

It is available at [http://www.gsars.org/e-learning/index.html](http://www.gsars.org/e-learning/index.html).

SuSo possesses straightforward advantages and is therefore recommended when collecting household survey data for SSF and aquaculture statistics.
SuSo operates as follows:

1. Tailor-made questionnaires for smartphones or tablets are designed in the SuSo questionnaire designer and stored on a country-specific server or in the cloud.

2. The questionnaire is downloaded on the smartphone or tablet.
SuSo operates as follows (cont’d):

3. Data is collected at the home ports, landing sites OR households and recorded on the mobile phone’s memory;

4. During sample data collection, a connection to the Internet is made and the newly collected data are sent to the supervisor, and then to the survey headquarters, for data quality check, approval and analysis.
Researchers design questionnaires using visual tools and upload them to the central server.

Questionnaires with no errors are uploaded to HQ.

HQ distributes the teams of enumerators.

Supervisors monitor the submissions.

Supervisors assign households to individual interviewers.

Interviewers visit households and collect data.

Interviewers repeat interviews if errors are detected.

Researchers design questionnaires using visual tools and upload them to the central server.

Questionnaires with no errors are uploaded to the central server.

Supervisors assign households to individual interviewers.

Interviewers visit households and collect data.

Interviewers repeat interviews if errors are detected.
FAO supported the development of a toolkit for SSF data collection, composed of the Open ArtFish software and an ODK mobile phone application.

The mobile phone application (for Android only) operates as follows:

1. Forms are designed in ODK and tailored to the needs of each data collection system (fishing gears, vessels, strata, etc.);
2. Data is collected at the landing site and stored in the mobile phone’s memory;
4. As soon as an Internet connection is available, the data is sent to the OPEN ARTFISH database (at HQ-level) through Google Cloud;
5. At the central administration (HQ level), a fisheries statistician officer connects to the server and downloads the collected data into the OPEN ARTFISH database.
CAPI applications – OPEN ARTFISH (ODK)

Data submitted by enumerator

Use Open ArtFish installed in HQ

Data downloaded from cloud to HQ database

Design form

Data input by enumerator
Data compilation and analysis software for SSF (OPEN ARTFISH)
Open ArtFish

Open ArtFish stands for Open Approaches, Rules and Techniques for Fisheries statistical monitoring. It consists of the Open ArtFish software application and its “backbone”, the Open ArtFish database. The software application is built to estimate total catch and value by species for sampling schemes of SSF.

FAO made Open ArtFish available at:


• The Open ArtFish generic database estimates total catch and value by species for sampling schemes of SSF.

• Open ArtFish uses appropriate statistical procedures and sampling design as described in more detail by de Graaf et al. (2014) and Stamatopoulos (2002).
Open ArtFish

• It is “open” because it can be easily adapted to local requirements and more specific outputs or reports can be easily added, while maintaining appropriate statistical procedures

• It provides an indication of the reliability of the estimates through the relative error of the CPUE and of the total catch
Open ArtFish

• It is developed in MS Access because:
  ✓ MS Access is widely distributed and easy to use
  ✓ all estimation procedures can be based on queries, which facilitate understanding of the estimation process
  ✓ People with MS Access programmer skills can easily learn and operate the software

• The document entitled *Open ArtFish and the FAO ODK mobile phone application: a toolkit for small scale fisheries routine data collection* introduces the operation of Open ArtFish and the ODK mobile app for SSF

• To set up and use Open ArtFish, expertise in statistics and MS Access is required. When funds are available, FAO provides support to the development of national data collection systems with tailored-made Open ArtFish solutions.
Standardizing simple databases

The estimation of total catch in sample-based surveys is relatively simple.

Three data tables are needed:
1. Frame survey data
2. Vessel activity data
3. Catch data
Open ArtFish

In the main screen, the parameters are defined (left column) and the catch data (right column) may be entered or the mobile phone data imported.
Setting up Open ArtFish

Define:

• Major strata, minor strata and landing sites
• Name of the enumerators
• Major fish species with the FAO 3 alpha code
• Fishing units (boat gear)
• Maximum number of monthly fishing days
• Updated frame survey data
Setting up the parameters

Make your mobile phone application form
Set database parameters

SELECT YOUR DATA COLLECTION
MAKE YOUR OWN LANGUAGE SYSTEM
Enter catch data

THE FISHING UNIT DATA

Date (dd/mm/yyyy): 08-Feb-17
Major strata: CENTRAL REGION
Minor strata: ABURA-ASEBU-KWAMANKES DISTRICT
Landing site: ABOKUM ANO
Name Enumerator: JOSEPH ESHUN
Fishing unit: APW
Trip duration (days): 1

THE SPECIES DATA

Species: GARFISH
Catch (Kg): 1.00
Value (Cedis): 1
Validated: }
## Analyse data

**Data analyses and Estimation with BAC**

<table>
<thead>
<tr>
<th>Year</th>
<th>Major strata</th>
<th>Species</th>
<th>AGE</th>
<th>Minor strata</th>
<th>Fishing unit</th>
<th>Landing site</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SELECT OPTIONS TO BE ANALYSED**

- L 1.1 Year Major strate Catch Value
- L 1.2 Year Minor strate Catch Value
- L 1.3 Year Major strate Fishing unit Catch Value
- L 1.4 Year Minor strate Fishing unit Catch Value
- L 1.5 Year Site Fishing unit Catch Value
- L 1.6 Year Fishing unit Catch Value
- L 2.1 Year Month Major strate Catch Value
- L 2.2 Year Month Minor strate Catch Value
- L 2.3 Year Month Major strate Fishing Unit Catch Value
- L 2.4 Year Month Minor strate Fishing unit Catch Value
- L 2.5 Year Month Site Fishing unit Catch Value

- L 3.1 Year Major strate Species Catch Value
- L 3.2 Year Minor strate Species Catch Value
- L 3.3 Year Major strate Fishing unit Species Catch Value
- L 3.4 Year Minor strate Fishing unit Species Catch Value
- L 3.5 Year Site Fishing unit Species Catch Value
- L 3.6 Fishing unit Species Fishing Effort and CPUE
- L 4.1 Year Month Major strate Species Catch Value
- L 4.2 Year Month Minor strate Species Catch Value
- L 4.3 Year Month Major strate Fishing unit Species Catch Value
- L 4.4 Year Month Minor strate Fishing Unit Species Catch Value
- L 4.5 Year Month Site Fishing unit Species Catch Value

**CHECK YOUR SAMPLE SIZE**

- Check sample numbers CPUE
- Check sample numbers BAC
- Check mobile phone samples taken by enumerator
Analysed data: an example

<table>
<thead>
<tr>
<th>Year</th>
<th>Major strata</th>
<th>Fishing unit</th>
<th>Total catch (t)</th>
<th>Total value ('000)</th>
<th>Avg Rel Error CPUE 90%</th>
<th>Avg Rel Error Catch 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>CENTRAL REGION</td>
<td>APW</td>
<td>690</td>
<td>2,492</td>
<td>48%</td>
<td>85%</td>
</tr>
<tr>
<td>2017</td>
<td>CENTRAL REGION</td>
<td>HOOK AND LINE</td>
<td>25</td>
<td>124</td>
<td>18%</td>
<td>30%</td>
</tr>
<tr>
<td>2017</td>
<td>GREATER ACCRA REGION</td>
<td>APW</td>
<td>3,260</td>
<td>5,934</td>
<td>30%</td>
<td>56%</td>
</tr>
<tr>
<td>2017</td>
<td>GREATER ACCRA REGION</td>
<td>DRIFT GILLNET</td>
<td>130</td>
<td>554</td>
<td>13%</td>
<td>53%</td>
</tr>
<tr>
<td>2017</td>
<td>GREATER ACCRA REGION</td>
<td>HOOK AND LINE</td>
<td>80</td>
<td>529</td>
<td>16%</td>
<td>31%</td>
</tr>
<tr>
<td>2017</td>
<td>VOLTA REGION</td>
<td>APW</td>
<td>1,369</td>
<td>7,463</td>
<td>69%</td>
<td>79%</td>
</tr>
<tr>
<td>2017</td>
<td>VOLTA REGION</td>
<td>BEACHE SEINE</td>
<td>769</td>
<td>2,387</td>
<td>71%</td>
<td>309%</td>
</tr>
<tr>
<td>2017</td>
<td>VOLTA REGION</td>
<td>DRIFT GILLNET</td>
<td>4</td>
<td>20</td>
<td>186%</td>
<td>391%</td>
</tr>
<tr>
<td>2017</td>
<td>WESTERN REGION</td>
<td>APW</td>
<td>0</td>
<td>4</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>WESTERN REGION</td>
<td>DRIFT GILLNET</td>
<td>177</td>
<td>972</td>
<td>15%</td>
<td>35%</td>
</tr>
<tr>
<td>2017</td>
<td>WESTERN REGION</td>
<td>SET NET</td>
<td>61</td>
<td>894</td>
<td>35%</td>
<td>68%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Total</th>
<th>Avg Rel Error CPUE 90%</th>
<th>Avg Rel Error Catch 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,564</td>
<td>21,370</td>
<td>48%</td>
<td>114%</td>
</tr>
</tbody>
</table>
The estimation process

- All based on simple queries
- Follows exactly the procedures explained in the international training course
- Calculate the average and standard deviation, count the number of samples, calculate the relative error at a probability level of 90 percent
- Statistically sound
- All data stored in a final table
- The final table can be queried for additional analyses
Access queries
Conclusions

Technological tools for data collection improve the quality, timeliness and cost of obtaining survey data

CAPI offers the advantages of requiring lower costs and enabling faster and higher-quality survey data collection, compared to PAPI

Two free software applications (SuSo and Open ArtFish) are available, with support and training material software to assist in data collection, compilation and analysis

To set up and use Open ArtFish, expertise in statistics and MS Access is required. When funds are available, FAO provides support to the development of national data collection systems with tailor-made Open ArtFish.
References


Thank you