Artisanal Tuna Data Workshop

11-14 November 2013, Noumea, New Caledonia

SESSION 8

Data Management

(in the context of Artisanal tuna fisheries data…)

Oceanic Fisheries Programme (OFP)
Secretariat of the Pacific Community (SPC)

Presentation Outline

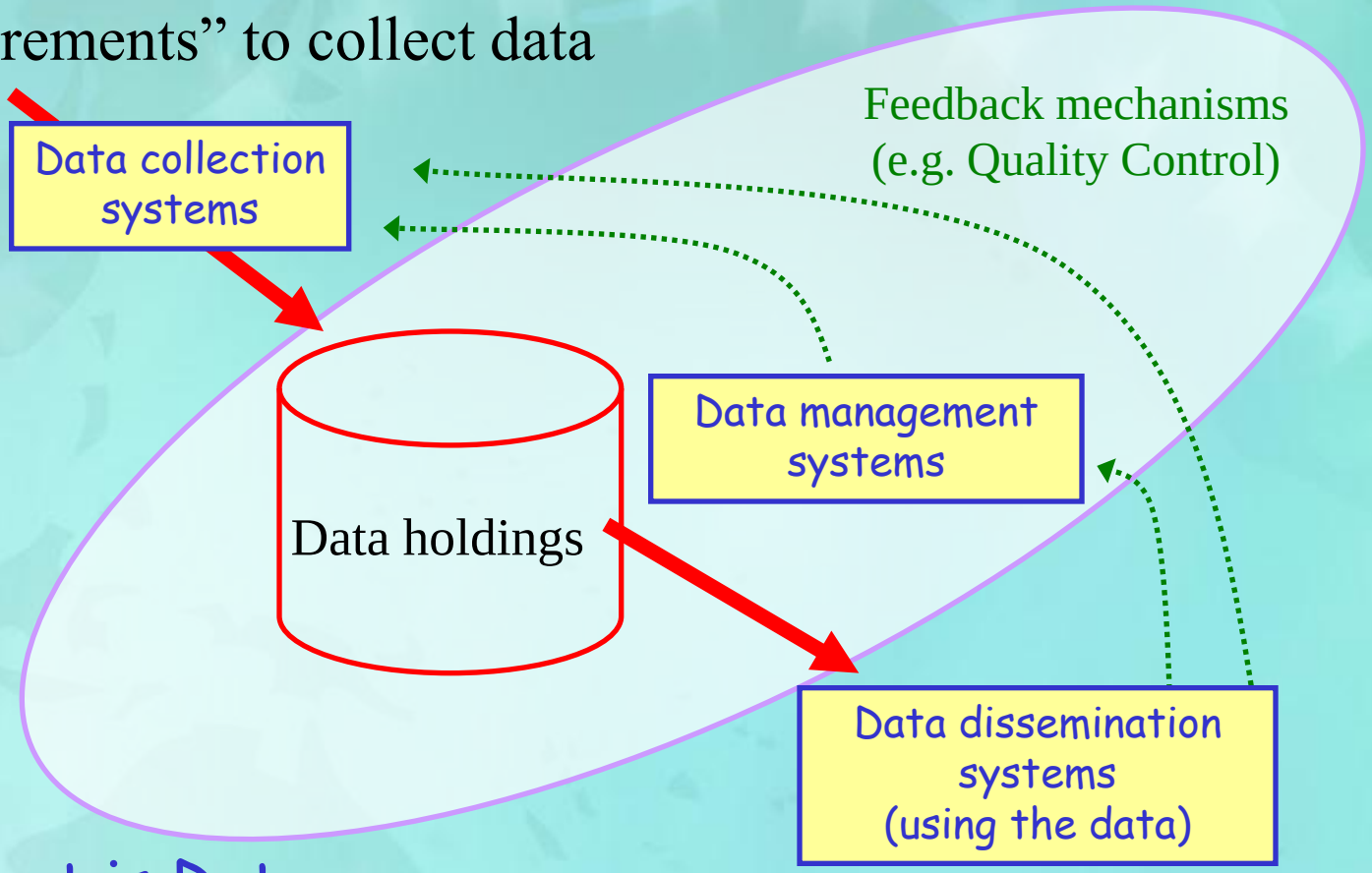
- What is Data Management ?
- Why Data Management is important …
- "Best Practice" in Data Management
- Artisanal Database systems
- Data Quality and Coverage
- Data Archiving

What is Data Management …

"Data management" is used to ensure …

- Data are stored in an "efficient" form (e.g. in an integrated manner)
- Data are of the highest "Quality" (e.g. retain their accuracy)
- Data are complete (e.g. represent the desired coverage)
- Data are readily accessible (i.e. facilitates dissemination)
- Data are secure

"Requirements" to collect data



What is Data Management ?

Output to users of data

Why Data Management is important…

… because …

- Data collection is an investment and the "management of data" protects and enhances that investment
- Users of the data want …
  - Data that are readily accessible
  - Data that are representative (both high quality and coverage)
- Owners of the data want …
  - Systems to store data in an efficient form
  - Systems to ensure data quality
  - Data security

Why Data Management is important…

… from Artisanal tuna fisheries data perspective …

- Significant effort to collect the data, so you want to ensure it is available in a form that can be easily used !
- Issues in the data collection are identified in data management and feedback used to enhance the data collection system and the quality of data …

Best Practice in Data Management

"The term '*best practice*' generally refers to the best possible way of doing something."

Principles of Best Practice

- Accuracy
- Effectiveness
- Efficiency
- Reliability
- Accessibility
- Transparency
- Timeliness
- Relevance

Best Practice in Data Management

Accuracy

– a measure of how well the data represent true values.

For example, "Catch in weight" may be recorded as a visual estimate, or may be weighed which provides a more 'accurate' value.

Best Practice in Data Management

Effectiveness

- the likelihood that desired objectives are achieved.

For example, Artisanal vessel activity data improves the coverage of artisanal data which serves the purpose of achieving "effectiveness"

…that is, these data more "representative" of the actual 'activity' and when applied to artisanal catch/effort, they improve the coverage, an so the data are therefore more "effective" when they are used …

Best Practice in Data Management

Efficiency - the ratio of output to input.

This is the amount of effort that is needed to produce an acceptable output.

It also refers to the amount of input data the user has to obtain to produce an acceptable result.

Database systems are used to "efficiently" store the data collected and allow "efficient" access of data through reporting systems.

Best Practice in Data Management

Reliability – related to accuracy, and refers to the consistency with which results are produced.

Some types of data are inherently more "reliable" than others, for example, GPS positional data are considered very reliable.

Data collected by an experienced data collection officer is usually considered more "reliable" than data collected by a first-time data collection officer.

The relative degrees of "reliability" in the artisanal data should be considered in a Data Management System and by the users of these data.

Best Practice in Data Management

Accessibility

- how accessible are the information to the users.

This relates to the ease with which users can access the information in detailed or summary form

-- database systems facilitate "accessibility" through REPORTING SYSTEMS.

Best Practice in Data Management

Timeliness

- relates to the frequency of data collection, its reporting and updates.

Refers to …

How frequent data are collected…
How frequent databases are updated…
How frequent databases are quality-control checked …
How frequent are reports produced and made available to senior managers/researchers/etc. …

Best Practice in Data Management

Relevance – the data collected should meet the needs of the user – i.e., should fulfill the principle of "fit for purpose".

Answers the following questions :

- Do the users of the data contribute to the data collection design and process ?
- Are users aware of the characteristics of the data they extract (i.e. is it "relevant") ?
- Is there documentation (metadata) available to the user on the parameters selected and sources of data used to generate a report ?

Consequences of Best Practice in Data Management

If "Best Practice" in Data Management is adhered to, then you should expect …

- Data collection/management staff will be confident in knowing exactly what they are dealing with…
- Users (senior management/researchers) will be confident in the data…
- Presentation of relevant and accurate information will ensure data collection/management staff will earn respect from the FISHERS …

Artisanal Database System (TUF-ART)

Definition of an Integrated Database System

" *…an integrated database system manages and organises the storage of all types of relevant data in electronic form through data entry, data quality control and other database management system modules.*

An integrated database system also allows for the efficient retrieval of information through a REPORTING module …"

Artisanal Database System (TUF-ART)

Benefits of an integrated Database System

- Maintains consistency and ensures standardisation
- Facilitates the determination of the coverage (completeness) of data
- Ensures standardisation (e.g. consistent naming of vessels, skippers, species, etc.)
- Facilitates the REPORTING of summarised data

Artisanal Database System (TUF-ART)

Data Processing Module

- Provides a means of transferring hard-copy data to electronic form
- Allows for efficient storage of data in a database
- Includes Data Quality Control, for example,
  - Online error checking
  - "Field range checks"
  - "Reference Table Edit checks"

Artisanal Database System (TUF-ART)

Why a database is better than an EXCEL file …

- Databases provide on-line quality control checking → better quality of data
- EXCEL file formats can be easily modified and therefore lose "standardisation" (i.e. the ability to combine data from EXCEL files)
- Large EXCEL files become very difficult to manage
- Database reporting systems make it easy to combine different types of data (e.g. vessel activity log and the catch/effort data)
- It is difficult to produce reports from multiple EXCEL files

Why Data Quality is important …

"Data Quality is important in ensuring the data collected and managed are accurate / representative".

"*You can have the best database system in the world, but if the data collection is not appropriate, then summary reports are not representative*".

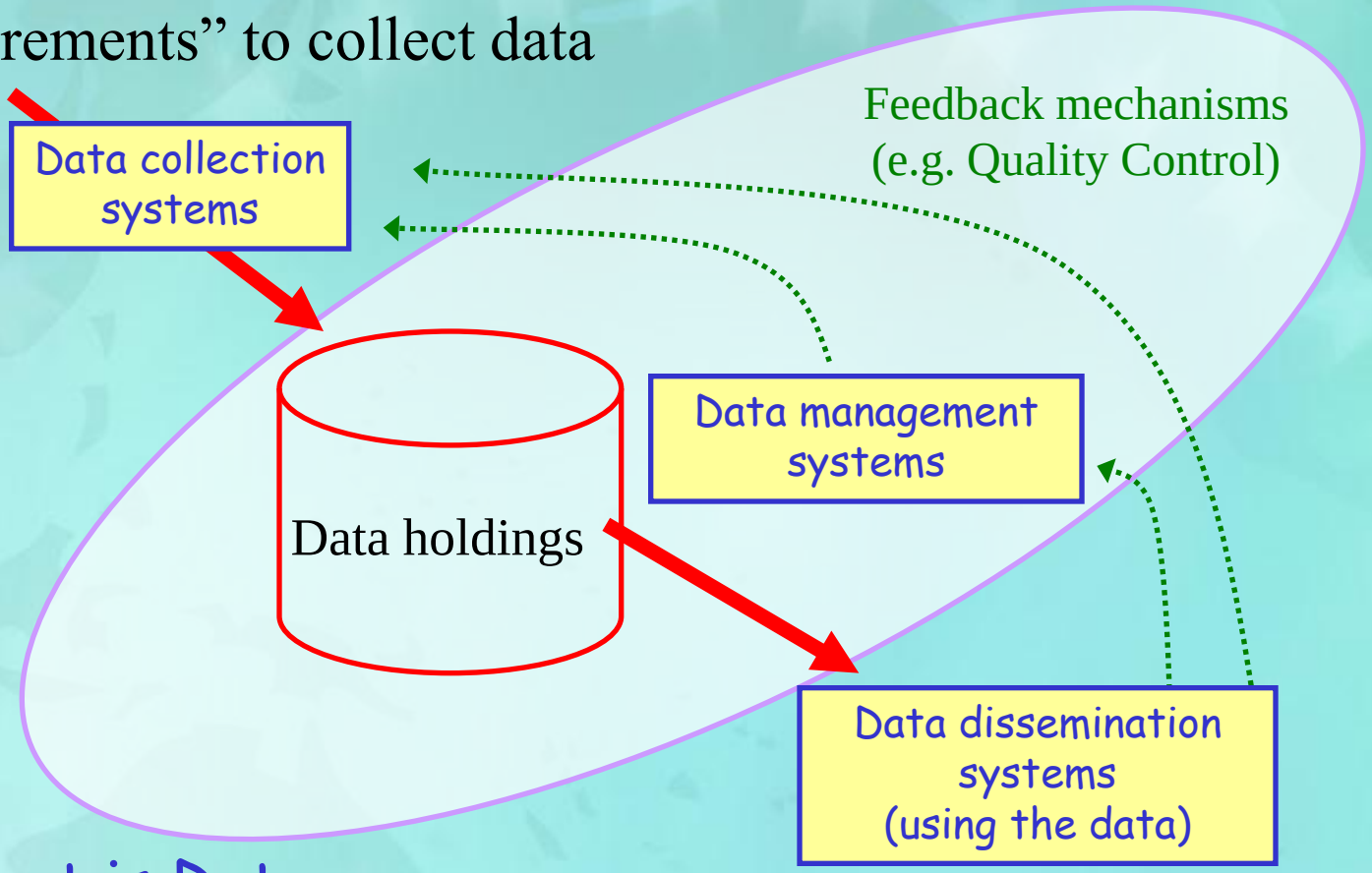
…put another way, "Garbage in" means "Garbage out" …

"Data coverage is one component in Data Quality Control"

Data Quality Control

- The "Quality Control" concept covers all systems …
  - Quality Control in data collection systems
  - Quality Control in data management systems
  - … linked through feedback mechanisms …
- Main purpose of Quality Control is to ensure accurate / representative data are available
- Quality Control in Data management systems
  - Pre-data processing
  - Data Processing
  - Post-data processing
  - Data quality reporting

"Requirements" to collect data



What is Data Management ?

Output to users of data

Data Archiving System

Data Backup/Recovery Strategy – Answers the following questions …

- What media is used to backup electronic data…
- What frequency of backup is used…
  - Daily, Weekly, Monthly, Annually
- Who is responsible for backup and recovery of electronic data…
- Are there instructions/procedures for the Data Backup and Recovery
- Consider whether all historic electronic data should remain on-line
- Consider whether historic hard-copy data should be archived off-site

Data Management

SESSION 8 -- DISCUSSION

Discussion : Consider any area where you could improve the MANAGEMENT of your Artisanal data, for example, …

a) Are there any "Best Practice" areas mentioned here that you could consider adopting ?

b) Could you improve your procedures for filing/backup/recovery of your data ?

c) Do you use the TUF-ART database system ?

d) Do you need the latest version of the TUF-ART database system ?

e) Could you consider any other areas for improvement ?