



<b>Integrated Capture Information System (ICIS)</b>
<b>User Requirements Gathering</b>
<b>FAO, Rome, Italy, 10 – 13 June 2008</b>
<b>Working document - ICIS Functional Requirements</b>
<b>Author: FIES/FAO</b>

# ICIS

## FUNCTIONAL REQUIREMENTS

### Data import

*The system shall be able to connect to or import from participating information systems:*

- All catch statistics available in RFB webs and FAO catch statistics.
- Connect to FAO GIS services.
- Connect to AquaMap services.
- Develop data sharing protocols with tier institutions, including data security agreements and accessibility to intermediate/final products.

### Data harmonization

*The system shall provide mechanisms to persistently map local to global schemas:*

- Develop/identify central standard format, flexible enough to accommodate all existing catch related statistics and potential expansion in near future.
- Develop translation module between individual datasets and central standard format.
- Local to global code/classification translation.

### Data model

*The system shall provide a common model for the storage of harmonized data from participating information systems:*

#### A. CATCH STATISTICS COLLECTIONS KEYS

- Area: FAO major areas (19), FAO subareas (92) and Divisions (154), EEZs (219), RFB defined areas (41), 5x5 deg., 1x1 deg., 0.5x0.5deg.
- Species: species, ISSCAAP Group, ISSCAAP division, FAOSTAT group, main group, order, family, tunas (*any other levels of aggregation?*)
- Time: year, quarter, month, day (*need to consider different year frames.*)
- Flag state: country, RFB memberships, and grouping levels (eg continents, EU).
- Stock.
- Gear (available for tuna catch statistics).

*note: Incorporation of efforts and size information will be considered in the next phase.*

#### B. RELIABILITY INDEXES

- CWP rules (e.g. Highest reliability for data from RFB with management responsibility):

*note: Need a table of areas and species of competence by RFBs.*

- Data of RFBs for area and species of their competence will be considered as the most reliable.
- When no priority information available, take a geometric mean.
- Statistical reporting:
  - Estimation through repetition from previous years, or from trend for missing data, will be considered as less reliable than reported data.

- Differing levels of spatial aggregation.
- Scientific revision status.

### C. SPECIES HABITAT INDEX

*note: Advanced species habitat function to be incorporated only when reasonably reliable maps developed by AquaMap. Until then, only simplistic models will be used:*

- Coastal .vs. offshore/oceanic: use the same criteria as those used for estimation of high seas catch of highly migrated species (need to define “coastal”.)
- For coastal species, substantial proportion (how many? e.g. 90%?) to be allocated to EEZ of flag country.

*note: Some sources may have adequate resolution.*

- For offshore/oceanic species, prorata based on FAO Species distribution maps (?).
- Eventual elements of species habitat function data model:
  - Polygon identifier.
  - Species 3 alpha code.
  - Time / seasonality component.
  - Source identifier.
  - a Method identifier
  - A normalized species habitat index value (value between 0 and 1.)

### D. GIS DATA MODEL

*To be defined but should meet needs of AquaMap and FIGIS-GeoNetwork/Geoserver standards.*

- Eventual elements of GIS Metadata data model:
  - Source.
  - Thematic domain.
  - File format (vector, raster).
  - Scale (resolution).
  - Temporal characteristics.
  - Reference system (projection).
  - With/without extended polygons.
- Eventual elements of GIS datamodel:
  - Vector (polygon/points).
  - Attributes conforming with RTMS standards.
  - Stored in Oracle Locator (but other alternatives are possible, i.e., PostGis).
  - Served using GeoServer through WMS and WFS.

## Catch data spatial reallocation

*For a given specie, the system shall be able to dynamically re-allocate catch statistics upon user queries using the following methods:*

- **Algorithm:**  $C(i,j)=C(aj)*P(i)*A(i)/\text{sum}[A(i|aj)*P(i|aj)]$ 
  - where **aj** is Area of source **j**
  - where cell **i** locates and
  - **P(i)** is *Spatial allocation function* reaching value between 0 and 1 and
  - **A(i)** is *area coverage* of cell **i**
- **Area function:** If  $P(i)=1$  for all **i**, the formula in fact allocates  $C(i,j)$  in proportion to an extent of area, without incorporating species habitat information.

- **Spatial allocation function:** species distribution filters: ad-hoc, simple distribution (coastal, oceanic, depth).
- **Aggregation/disaggregation functions:** average, best guess, absolute diff., percent diff., etc.

## Query Scenarios

*The system shall allow authorized users to query the database:*

- Single source statistics (for public use):
  - Let user select Source collection and single geographic system for output
  - Provide numeric extraction of data in accordance with user selected filter and classifications (like Fish Stat) with and without spatial reallocation.
  - Provide maps of data in accordance with user selected filter and classifications (like Fish Stat) with and without spatial reallocation.
  - Provide graphs of data in accordance with user selected filter and classifications.
- Dual source statistics (for RFB use, as part of working system): comparisons
  - Let user select two Source collections and single geographic system for output
  - Provide numeric extraction of difference between two statistics in accordance with user selected filter and classifications (like Fish Stat) with and without spatial reallocation.
  - Provide maps of difference between two statistics in accordance with user selected filter and classifications (like Fish Stat) with and without spatial reallocation.
  - Provide graphs of data in accordance with user selected filter and classifications.
- All source statistics combined (expert view for public use, pre-processing needed):
  - Let expert-internal user select Multiple Source collections and single geographic system for output.
  - Provide numeric extraction of consolidated statistics as expert guess in accordance with user selected filter and classifications (like Fish Stat) with and without spatial reallocation.
  - Generate dissemination products, and store them under restricted access with all query parameters, until eventual publishing:
    - Tables,
    - maps and
    - graphs.
  - Provide query mechanisms for general public in order to access pre-processed and published map-graphs products.

## Data analysis

No data analysis required at this stage

## Data annotation

*The system shall allow authorized users to annotate data and configurations:*

- Document discrepancies across dataset collections.
- Track changes to dataset collections across versions(?).
- Annotate rules.

## Data output

*The system shall be able to disseminate data to authorized users:*

### **E. ACCESS**

- Multiple aggregations depending on user type (higher resolution internally for data harmonization, lower resolution externally for dissemination).

### **F. FORMATTING**

- User-defined formatting.
- Need intermediate output that can then be prepared/cleaned and used as a basis for final public output.

### **G. METADATA**

*General principle is to provide along with data:*

- Data source

*Transparent access to sources used.*

- Data provenance

*Metadata regarding where data comes from and how collected.*

- Data reallocation

*Provide spatial reallocation rules and functions applied to harmonize and/or re-allocate data.*

## SUPPORTING TOOLS

*These are tools which while not directly part of the ICIS system will be eventually needed to support more precise reallocation scenarios*

### Reallocation tools

#### H. SPECIES DISTRIBUTION:

- A working system for the comparison of different GIS sources for species spatial distribution:
  - Calling two sources of GIS based species distribution information.
  - Overlaying geographically these two sources.
  - Based on outlier mechanisms (to be defined), producing a file of doubtful data (annotated with reason for doubts), for consideration by data owner.
- Should be able to work with FAO species distribution and WorldFish Center's AquaMap product.

#### I. SPECIES HABITAT:

- Tool supporting creation and editing of habitat profiles for the use of the Habitat function:
  - Will be defined for each species as multiple of distribution probability by different factors, such as coastal/offshore, depth, temperature, etc.
- Support creation of profiles by:
  - Manually uploaded profile.
  - Profile filled from simple GIS processes based on overlay of FI Species distribution maps with selected geographic resolution layer.
  - Profile filled using external processes such as AquaMap.

*note: in the case of multiple species, aggregation of above described processes.*

## NON-FUNCTIONAL REQUIREMENTS

### Data granularity

- Least-aggregated time resolution.
- Highest area resolution.

### Metadata

- Source identification.
- Last update.

### Extensibility

- Future ability to access national statistics.

### Availability

- Web.

### Performance

- Processing speed for results should not be slower than if it were stand-alone.
- Queries should be delivered within five seconds.

### Technologies

- Use GIS.
- Use ontologies.
- Use web-based technology.

### Existing services

- Make use of FISat working system services.
- Consider recycling FishStat+ forthcoming services.
- Make use of forthcoming FIGIS-Geonetwork/Geoserver GIS services.

### Standards

- OWL-compatible for classifications and reference terms.
- CWP classifications:
  - ISSCAAP.
  - ISSCFV.
  - ISSCFG.

- FAO Statistical areas.
- GIS:
  - ISO19115 for spatial Metadata
  - Open Geospatial Consortium (OGC) and their 2 leading specifications (WMS, WFS).
  - Geography Markup language (GML), which describes geographic features (lines, polygons, points) and associated DBF attributes (name, ...) for spatial data exchange.

*Note: more at [http://www.ird.fr/informatique-scientifique/methodo/standards/normes\\_iso\\_ogc/](http://www.ird.fr/informatique-scientifique/methodo/standards/normes_iso_ogc/)*

## **Usability**

- Should provide novice/expert interface.

## **PROJECT RISKS**

### **Technical**

- Way of handling different “year” definitions, especially when no monthly data are available.
- Difficulty in defining reliable habitat function.
- Handling of species aggregations.

### **Managerial**

- Data sharing agreements.
- Difficulty distributing work between trust funds.
- Lack of adequate involvement from partners.
- interferences with on-going developments of FAO GIS features, and FishStat



Annex: Conceptual application model

