منظمة الأغذية والزراعة للأمم المتحدة | 联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Organización de las Naciones Unidas para la Agricultura y la Alimentación

COORDINATING WORKING PARTY ON FISHERY STATISTICS

Twenty-fifth Session

Rome, Italy, 23-26 February 2016

Handbook on GIS matter: a concept note

Author: FAO

GEMS - GIS ENFORCING METADATA & SEMANTICS (FROM FIGIS WIKI)

The need to disseminate and describe properly GIS data across the web is growing, and becomes part of institutional objectives. Best practices are reinforced by the implementation of internationally-recognized standards such as ISO and OGC, and regulations such as the EC INSPIRE directive. Standards target various key elements such as *data*, *metadata* and *services*. To reach these objectives and guarantee that standards are implemented and best practices followed, tools are needed with the data managers as main target audience.

GEMS is one of these tools, and intends to facilitate the co-management of GIS web data (layers) and their metadata description, in an operational, iterative & cost-efficient approach. GEMS is born from an initial brainstorming in the FIGIS team that highlighted the need to improve the public dissemination and discovery of GIS products and guarantee key business metadata such as data access & use constraints and data provenance (lineage). Its implementation started within the i-Marine Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources. It builds its strength by interacting with:

- Fishery & Marine institutions: French IRD Institute, VLIZ Flanders Marine Institute
- GIS development communities: through its contribution (user feedback, software development) to well-established open-source projects such as Apache SIS, GeoToolKit and geoserver-manager
- Other technology-oriented communities: connecting the GIS ‘’world’’ with others such as Linked Open Data, e.g. the Grade initiative

More info at https://github.com/openfigis/gems/wiki

SUBJECT: SPATIAL GRIDDED SYSTEMS FOR STATISTICAL COLLECTIONS

Purposes: The need, for harmonization and robustness purpose, to enforce a set of guidelines when reporting spatially gridded statistical collections to FAO is highlighted. The authors may refer to specific use cases like the tuna statistics provided by Tuna Regional Fisheries Management Organizations (tRFMOs) as examples of data collated by FAO and assembled into a global harmonized dataset.

Current status

Grid reporting systems: Fisheries statistics reported by grid of cells are generally collected on board of vessels using exact coordinates or a geo-referenced grid system. These data are then aggregated at a lower resolutions and reported to FAO. It is therefore paramount that the grid reporting system is well documented. For what concerns tRFMOs, the most common aggregation level is the 1x1 or 5x5 degrees resolution but occasionally rectangular or irregular grids/polygons are used.

Grid coding systems: The CWP Handbook, in its section G, suggests an "Areal grid system" as "a graticule-based system [that] could be used throughout the globe for coding purpose". (more info at http://www.fao.org/fishery/cwp/handbook/G/en). FAO has implemented the above "areal grid system" (with some minor modifications) for the statistical collection supporting the Atlas of Tunas and Billfishes (http://www.fao.org/fishery/statistics/tuna-catches/en) for the purpose of providing a global harmonized data set on tuna catches. It is noted here that, at present, only IOTC has endorsed an areal grid coding system following the above CWP Handbook suggestions for their public domain datasets. All others tRFMOs make their datasets available using latitude and longitude coordinates of each cell using a variety of different methods especially when defining the quadrant and the location of catches within each grid cell. The same "Areal grid system" is used for disseminating tuna statistics through web-based data services, such as iMarine. There are other grid coding systems, like the "c-square system" (more info at (http://www.cmar.csiro.au/csquares/spec1-1.htm), that might be considered when reporting statistical data to FAO. The above mentioned grid coding systems (Areal grid and c-square) refer to vector data structure types (polygonal or grid of points).

Data exchange formats:

Fisheries statistics reported, for example by tRFMO, are generally provided in a variety of tabular formats, like CSV, Excel, MSAccess tables, etc.

OGC or other standard data exchange formats exists and can be considered for exchanging statistics using the above grid coding systems. We can mention:

- GML (OGC official geospatial vector data XML format), for which providing an explicit geometry is not mandatory.
- SDMX (XML or JSON), which compared to GML has the benefit not to be bound to syntactic interoperability and to the geographic dimension. Indeed, SDMX datasets are explicitly defined with a Data Structure Definition (DSD), which gives all the meta information required to define/describe/produce a dataset, such as dimensions (which may include a geographic one), attributes, concepts, and reference data (codelists / vocabularies).

It is worth mentioning that other grid spatial formats exist, where data is not area-coded but explicitly geo-referenced with coordinates, like, as an example, the raster CF-NetCDF format (OGC standard).

Recommendations:

According to the authors, the main points to consider for recommending an appropriate grid system(s) for collecting and reporting catch and effort statistics to FAO, are the following:

1. *Define the characteristics of the grid reporting system*

   Characteristics of the grid reporting system may be mandatory or optional according to the case, and may be the following:
   a. shape of the cells (squares, rectangles, others)
   b. minimal resolution for a given set of species/gear/fisheries combination

2. *Define one or more grid coding systems*:

   Various grid coding systems may be considered, as long as they comply with the above characteristics and are compatible with the data aggregation made by FAO. Grid coded data reduce the geographic dimension to a simple codelist thus facilitating data reporting to FAO. Mappings and conversion tools between "Areal grid system", and "c-square" coding systems are already available in FAO hence these two coding systems are recommended for consideration.

3. *Define one or more data exchange formats*
   Various exchange formats may also be considered for the data exchange / reporting to FAO. This includes tabular (CSV, XSLX, MSAccess, SDMX) or geospatial data formats (coordinates, geometry, GML, NetCDF).
   It is recommended to investigate the use of standard formats that are not limited to syntactic interoperability but also guarantee a proper data structure definition. This includes the geographic dimension but also other required vocabularies used in the dataset (e.g., species, gears, etc.), such as in SDMX. Other geospatial grid formats (such as CF-NetCDF), although they seem more complex for the implementation by data providers, might be appropriate for complex data reporting when applying a grid coding system may not be sufficient.

The definition of grid coding systems and formats relevant for geo-referencing purpose should be discussed as part of the broader objective to define data and metadata standards for fishery data reporting (*e.g.* SDMX fishery profile / Data Structure Definition, NetCDF fishery-oriented CF conventions, etc.). This could be the object of a technical working group.

For reference purpose, the following tables provide:

- a comparison of the main characteristics of the CWP "Areal grid system" and the C-Square coding system to support data provider when choosing a grid coding system.
- a comparison of data exchange/reporting formats

Comparative table between two well known grid coding systems

	Areal Grid System	**C-square**
Support of shapes others than squares	Yes	No
Highest resolution	Up to 10 minutes of degree resolution (for squares)	Up to 1/100th of degree
Lowest resolution	Down to 30 degrees (for squares)	
Allowed resolutions	specific list (see http://www.fao.org/fishery/cwp/handbook/G/en)	resolution multiple of 5 (10°, 5°, 1°, 0.5°, 0.1°, 0.05°, etc)
Availability of coding/decoding from/to coordinates	Yes, in iMarine but not in desktop, e.g. Excel, R.	Yes, in iMarine plus other externals, in different languages
Need to define the reference point of the grid cell (center or corner)	Yes	No
Nesting/child-parental relationship	Yes, with limitations (it depends on the size and shape of the grid)	Yes
Allow non-squared grid	Yes	No

Comparative table between data exchange/reporting formats

Format	GML	SDMX-ML SDMX-JSON	CF-NetCDF
Standard	OGC	SDMX	OGC
Type of spatial format	Vector	-	Raster
Support for area-encoded data	Yes	Yes	No

Support for explicit geo-referenced data	Yes	Yes	Yes (regular grid only for the time-being)
Syntactic interoperability	Yes	Yes	Yes
Vocabulary control (with metadata)	No	Yes	Yes, using CF conventions
Existence of fishery profile / metadata conventions	No	No, to be defined	No, to be defined