Fisheries Data Interoperability WG Introduction

A.Ellenbroek, A.Charef, E.Blondel - FAO of the UN

J.Barde - IRD

CWP-IS

21 June 2017

Copenhagen

Fisheries Data Interoperability WG:

Put five scientists in a room and ask them about copyright and you will get seven wrong answers.

They could have used their time better using an open license and then going about their science …

The RDA; RESEARCH DATA ALLIANCE SHARING WITHOUT BARRIERS

Build the social and technical bridges that enable open sharing of data.

- **enable data shared across barriers through focused Working Groups and Interest Groups**
- Participation in RDA is open to anyone
- Launched in 2013 by the European Commission, the United States National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation
- over 5,700 members from 128 countries (June 2017)

The FDI-WG

The Fisheries Data Interoperability Working Group (FDI-WG)

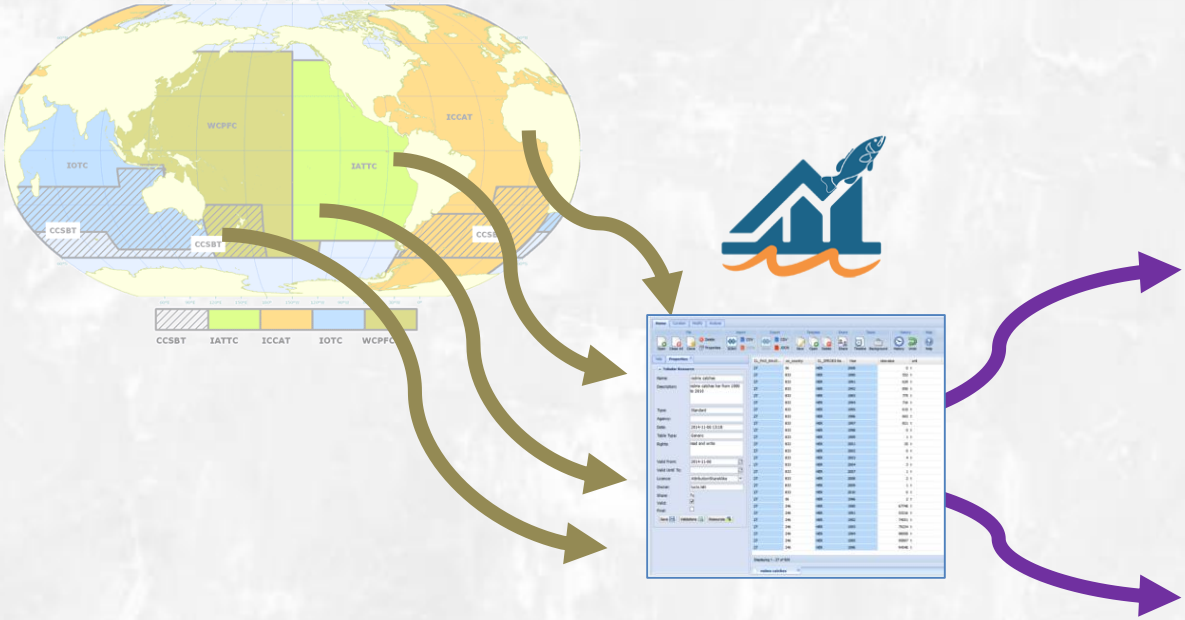
- Devise a global data exchange and integration framework on fisheries data
- Facilitate the use of de-facto, and preferably open, standards for the identification, description, mapping and publication of fisheries data
- Connect existing data networking initiatives

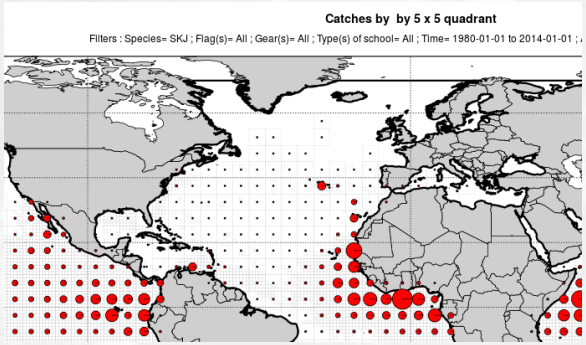
Formatting; A big picture
Global Tuna Atlas; from fishnet to internet

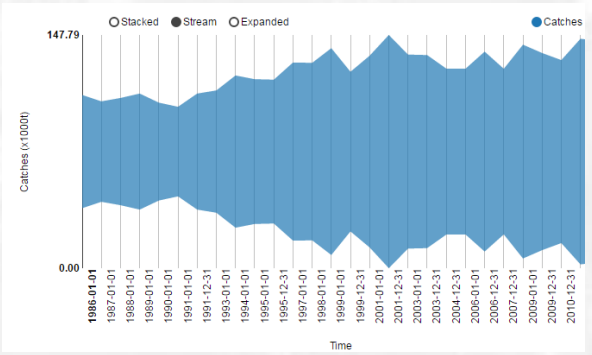
Collate global public tRFMO data

Integrate and harmonize

Publish locally and in 'your' website







FDI-WG has two informal Task Groups (similar to CWP)

- Task Group on Reference data harmonization for capture fisheries and aquaculture statistics with main objectives: (A. Charef Co-Chair)
  - Discuss a proposal for a global data structure (e.g. a Catch DSD)
  - Discuss harmonization/mappings
  - Discuss guidelines for structure extensions.

- In the context of reference Geographic data, a GIS group is tasked to recommend grid reporting systems, their codification, and format (Co-Chair Julien Barde of IRD, Emmanuel Blondel represents FAO)

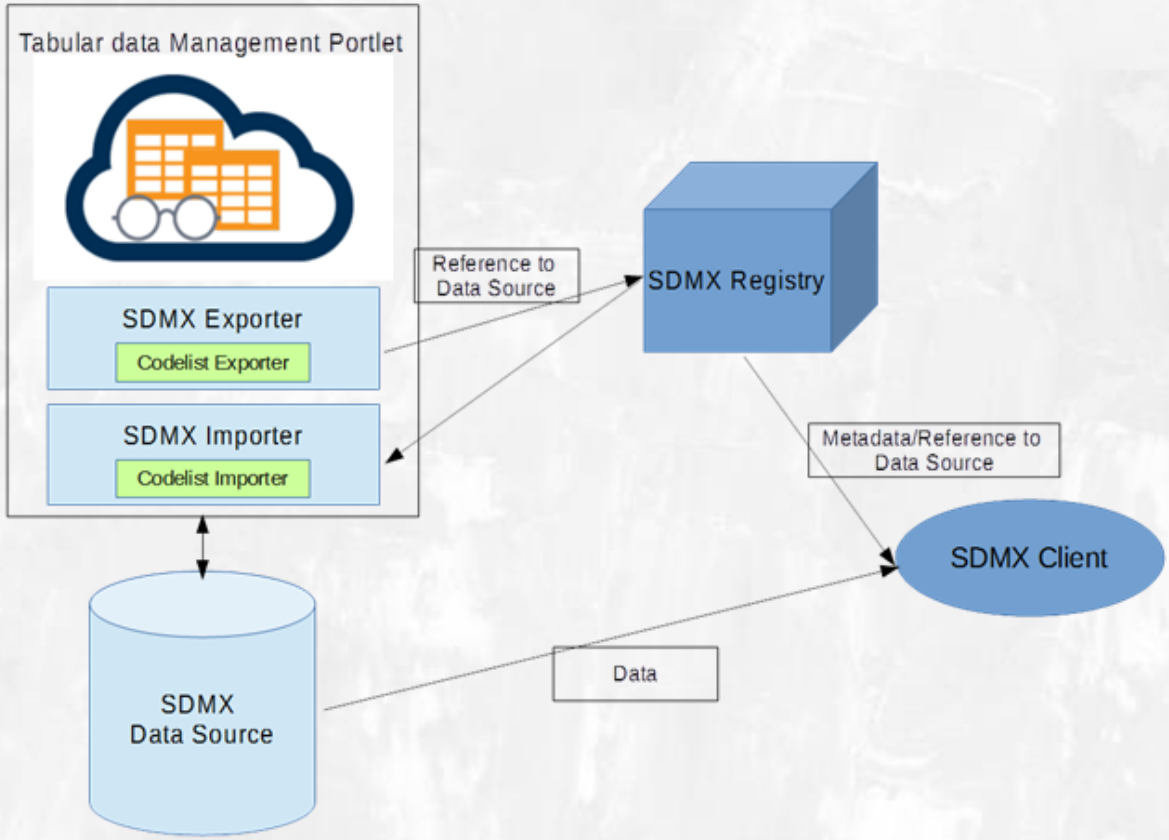
Reference data - Data Structure Definition (DSD)

Suggested DSD for Global Capture Production

Order	Concept_id	Role/Type	Codelist_id	Codelist_Code_id	Description/reference
A	COUNTRY	Dimension	CL_FI_COUNTRY_GROUPS	UN_CODE	Country code
B	FISHING_AREA	Dimension	CL_FI_AREA_GROUPS	CODE	FAO major catch areas
C	SPECIES	Dimension	CL_FI_SPECIES_GROUPS	3ALPHA_CODE	ASFIS species
D	YEAR	Time Dimension			Reference year
E	QUANTITY	Primary measure			Quantity of production
F	SYMBOL	Attribute	CL_FI_SYMBOL	CODE	FAO standard symbols
G	UNIT	Attribute	CL_FI_UNIT	CODE	Quantity unit

The DSD's and the related Code Lists will be publicly available

Example Toolset based on iMarine Tabular Data Management Service and SDMX architecture



Final deployment

RDA Fisheries datasets
Is interoperability an issue?

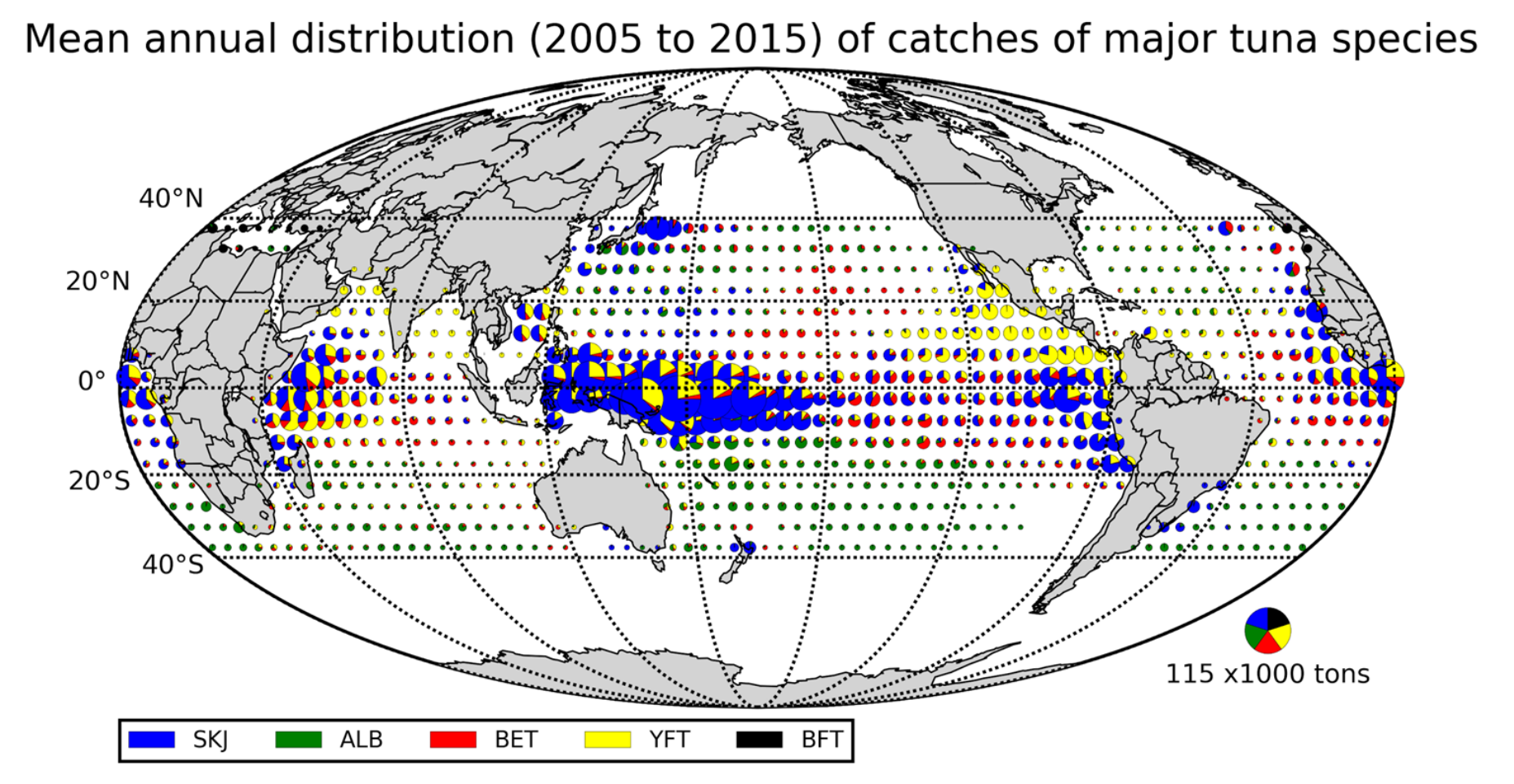
Some types of fisheries datasets:

- **observations**:
  - **high resolution datasets** (individuals: fishing vessels & fishermen, scientific cruises)
    - occurrences / sighting & fishing operations, related samples
    - trajectories: eg VMS, FADs, animal tracking
  - **gridded** / aggregated datasets: usual type of datasets delivered by RFMOS (various spatial and temporal resolution, sometimes delivered by areas).
- **forecasts**: eg stock assessment datasets (models inputs and outputs)

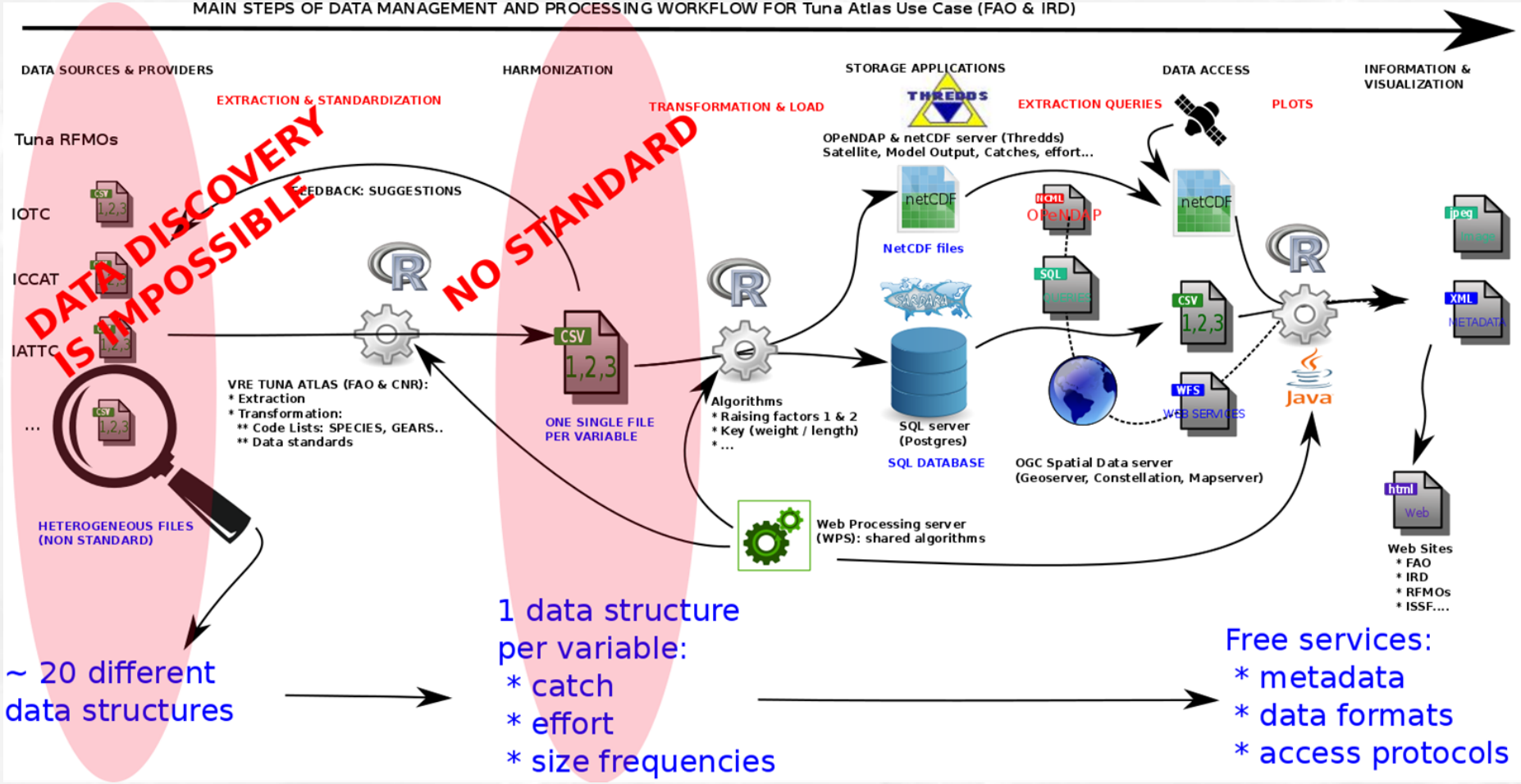
Fisheries data: summary of interoperable fields

- Fisheries domain objects = {*Fleet, Fishing Vessels, Fishing Gear, Fishermen,species,population, stocks, schools, individuals, Fishing Areas..*} => *Fishing Gear (longline, FADS…), Locations / Areas: EEZ, MPAs,*
- Fisheries domain activities = {*Fishing operations, landings, sampling*}
- Fisheries domain observations / parameters = {*effort,catch, nominal catch, bycatch, discards, biometry(size class,length-weight relationships..), reproduction / maturity stage, mortality, recruitment*}
- **RDA: How to make these interoperable**

Getting global.. eg tuna data



Getting global...complicated



Fisheries datasets interoperability: main goals

Adopt standards to implement guidelines:

- **data discovery** and proper use of datasets,
- **data access** (whatever data format),
- **data transformation** (multiple data formats),
- work first on concepts more than formats
  - ISO/TC 211 discussions at conceptual levels
  - OGC UML diagrams with XML schemas

Can RDA facilitate Fisheries data reuse with existing standards for (meta)data formats?

RDA Fisheries datasets interoperability: main goals

- **Comply** with conventions:
  - with generic format:
    - CSV,
    - NetCDF and other OGC standards (metadata, WMS, WFS..),
    - SDMX
  - with domain specific formats: EML/Darwin Core,
- RDA: Tools to assist Data **providers** to comply with standards: eg OGC (softwares & services), URI's, DOI, Metadata, ...

Reuse existing conventions in related domains

- OPEN DATA / Science (**generic**) metadata:
  - Datacite (OpenAire Connect, eg DCMI)
  - CEFAS, SEANOE IFREMER, NOAA/IOOS (CKAN can harvest other catalogs: eg Geonetwork).
- OGC standards for **spatial** (meta)data formats and access protocols are widely used:
  - Europe (INSPIRE: eg Ifremer, IRD CSW , Thredds & NetCDF, WFS)
  - FAO, other countries (eg IMOS/CSIRO)
- **Statistics**: SDMX

Reuse existing conventions in RDA related domains

- CF Conventions ocean (meta)data:
  - **dimensions**: time, latitude, longitude, depth,
  - **variables /** parameters: physical, chemical, biological (to be extended),
- TDWG standards: for biological (meta)data (species, observations and measures)
  - GBIF / OBIS: EML for ecological and biodiversity data (occurences, samples, biometry..)
  - from scientists: NOAA, CSIRO,..
- RDA assists reuse / extend standards for fisheries.

Geo-referencing concepts for fisheries data interoperability

- **Synergy between RDA Fisheries Data interoperability WG and Coordinating Working Party on Fishery Statistics**
- **Scope and definitions**
  - Scope of “Fisheries data”?
  - Geo-referencing / geographic dimension
- **Geo-referencing for fisheries data**
  - Levels of geo-referencing
  - Data and Metadata standards

Geo-referencing levels

Classification and Coding systems

- **Classification systems**
  - **Area systems** (polygon)
    - Aggregated
    - Irregular
    - Possible hierarchies / breakdown

      Statistical areas, Reporting areas, Jurisdictional areas, Competence areas Management areas (protected, restrictions, closures) / units, etc.
  - **Locations / Sites** (point)

    Ports, landing sites, sampling sites, etc.
  - **Transects** (line)

- **Coding systems**
  - **Geo-codelists**
  - **Registries**

Geo-referencing levels

RDA: Area system classification for fisheries data?

ISO 19115 Topic categories

value	description
farming	rearing of animals and/or cultivation of plants. Examples: agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock
biota	flora and/or fauna in natural environment. Examples: wildlife, vegetation, biological sciences, ecology, wilderness, sealife, wetlands, habitat
boundaries	legal land descriptions. Examples: political and administrative boundaries
climatologyMeteorologyAtmosphere	processes and phenomena of the atmosphere. Examples: cloud cover, weather, climate, atmospheric conditions, climate change, precipitation
economy	economic activities, conditions and employment. Examples: production, labour, revenue, commerce, industry, tourism and ecotourism, forestry, fisheries, commercial or subsistence hunting, exploration and exploitation of resources such as minerals, oil and gas
elevation	height above or below sea level. Examples: altitude, bathymetry, digital elevation models, slope, derived products
environment	storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape
geoscientificInformation	information pertaining to earth ?nces. Examples: geophysical features and processes, geology, minerals, sciences dealing with th composition, structure and origin of the earth s rocks, risks of earthquakes, volcanic activity, landslides, gravity information, soils, permafrost, hydrogeology,
health	health, health services, human ecology, and safety. Examples: disease and illness, factors affecting health, hygiene, substance abuse, mental and physical health, health services
imageryBaseMapsEarthCover	base maps. Examples: land cover, topo
intelligenceMilitary	military bases, structures, activities. Ex information collection
inlandWaters	inland water features, drainage systen lakes, water utilization plans, dams, cu
location	positional information and services. Ex zones and services, place names
oceans	features and characteristics of salt wat waves, coastal information, reefs
planningCadastre	information used for appropriate actio maps, cadastral surveys, land ownersh
society	characteristics of society and cultures. traditional beliefs, manners and custor impact assessments, crime and justice,
structure	man-made construction. Examples: bu
transportation	means and aids for conveying persons routes, tunnels, nautical charts, vehicl
utilitiesCommunication	energy, water and waste systems and hydroelectricity, geothermal, solar and sewage collection and disposal, electri

INSPIRE data themes

Annex I
1. Coordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

Annex III
1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities

?

11. Area management/ restriction/regulation zones & reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions

No Fisheries data area types common vocabulary

Towards building fisheries data vocabularies (definitions, code list)?

Asssessment / Distribution area

Fishery Statistical area

Fishing area

Competence area

Reporting area

Management unit

Jurisdiction area

Metadata enforcing

OGC GeoAPI

OGC/ISO 19115-19139

OGC / CSW

INSPIRE

Standard Vocabularies (species names & codes)

ASFIS

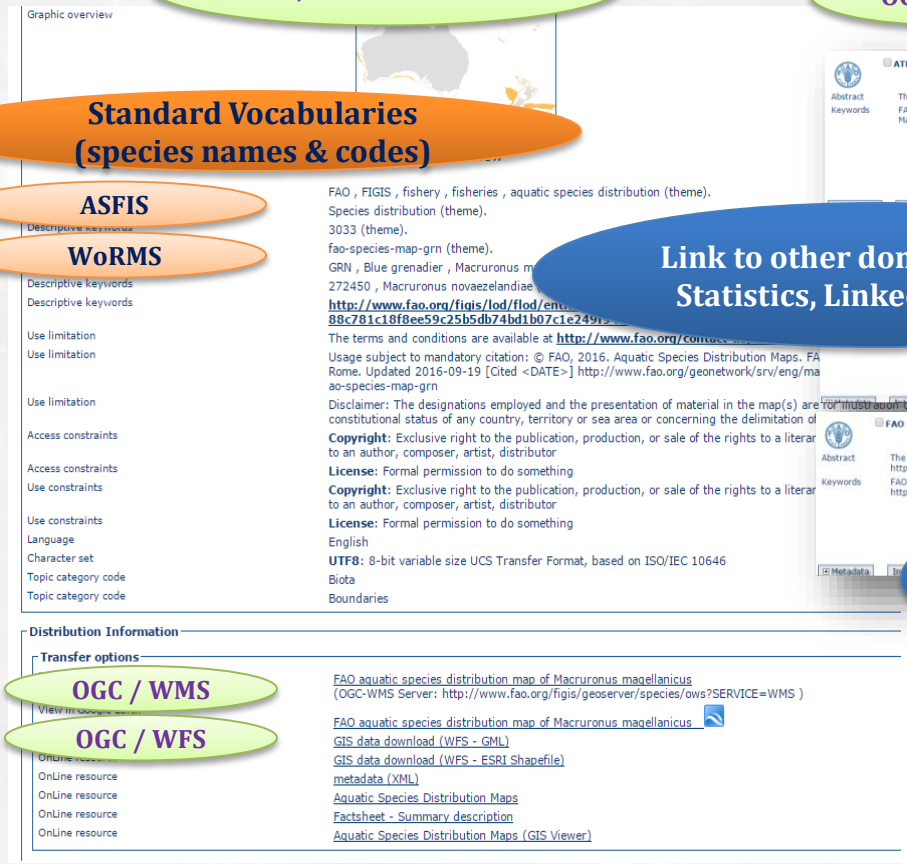
WoRMS

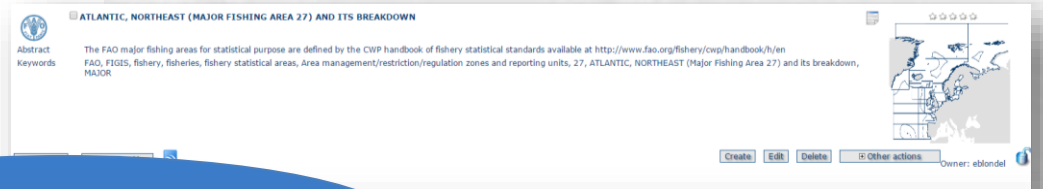
Link to other domain systems Statistics, Linked Open Data

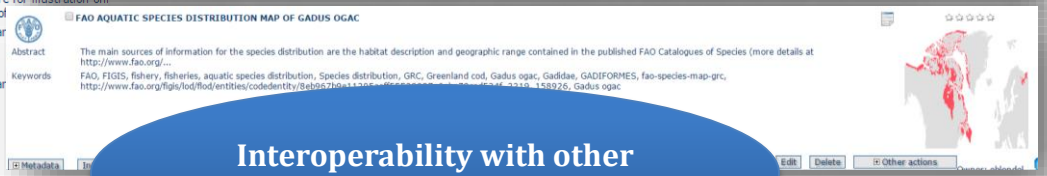
Interoperability with other Search Engines

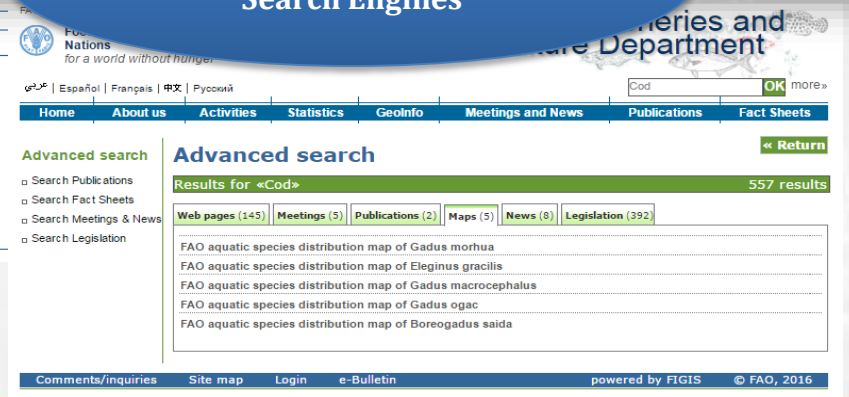
OGC / WMS

OGC / WFS









Metadata enforcing – **Emerging needs**

- **RDA Use Case – Tuna fisheries data Atlas**
  - **Mainly supported by i-Marine/BlueBridge, and building on shared FAO GIS knowledge & practices.**
  - **Work with BlueBridge partners (FAO, IRD, CNR, etc.)**
  - **Objectives:** Building on CWP and RDA recommandations:
    - Standardized, transparent and sustainable data ingestion flow from Tuna RFMOs into a common Tuna fisheries database.
    - Standard public access & discovery of Tuna regional and global fishery geo-referenced datasets, **annotated with harmonized fisheries data vocabularies,** through complementary Geospatial (OGC) and Statistical approaches
    - Enhanced Tuna fisheries data portal

Conclusion

RDA WG's related services

1. Fisheries data providers should focus on standardized files (CSV, NetCDF..) & metadata,
2. Services for data management should come for free: BlueBridge Tuna Atlas VRE as a playground for Fisheries gridded datasets:
   - Spatial data (OGC): Geoserver, Geonetwork / CKAN, Thredds, Postgres / Postgis),
   - Biodiversity => TDWG/GBIF and IPT
   - SDMX

Conclusion

Next activities in RDA FDI-WG:

- Agree on recommendations:
  - adopt some data structures & metadata (CSV)
  - validate & implement workflows.
- BlueBridge infrastructure to be used:
  - to publish (meta)data on FAO Geonetwork:
    - gridded datasets, acoustics
    - model outputs: stock assessment; trajectories simulations,
  - give access to datasets (CSV, NetCDF, SDMX..)
- Virtual meetings to discuss details

RDA WG to implement «Solutions» for Data management

- ***Innovation: Re-useable catch data and fisheries related data products***
  - *Recommendations for statistical and model data (RDA)*
  - *Recommendations for geospatial data and formats (RDA)*
  - *Demonstrate feasibility of data services (BlueBRIDGE)*
  - *Describe exploitation scenarios in Datathon(s) (Both)*