Global Record of Stocks and Fisheries

***A task of* Blue Assessment work package**

FIRMS - BlueBRIDGE TWG meetings
1-2 March 2016
Rome, Italy

BlueBRIDGE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675680

Outline

- Overall vision of the Global Record of Stocks and Fisheries (GRSF)
- Objectives GRSF (BlueBRIDGE Task 5.2)

Identifying the nature of GRSF

- List of key concepts/data fields of the sources of information FIRMS, RAM, FishSource and relationships *(implication with minimum data requirement for stocks and fisheries)*
- Comparison of list of stocks *(pilot cases including ICES and other data providers to cover different data coverages, implication with minimum data requirement)*
- Unique identifiers *(naming conventions for titles, best practices for human and machine readable identifiers, examples of DOI or other similar services, etc.)*
- Code lists and ontological representation *(species, area, etc. mapping among the 3 systems)*

- CMS (scope definitions), other supporting tools, workflow
- Extended knowledge base *(concept, competency questions, examples: FIRMS-WioFish, Aquaculture publications, results accuracy and implications at content provider level)*
- Governance and sustainability *(Content governance, system governance, data sharing policies, sustainability business model)*

Global Record of Stocks and Fisheries

- Vision

  an innovative environment supporting the collaborative production and maintenance of a **comprehensive and transparent global reference set of stocks and fisheries records** that will boost regional and global stocks and fisheries status and trend monitoring as well as responsible consumer practices.

GRSF Overall objectives

A global platform:

- **For compiling and sharing stock assessment results and management data for all of the world's fisheries**
- **To facilitate access to information on status & trends of stocks and fisheries**
- **To offer services to compute regional/global state of stocks indicators**
- **To offer services to public and private actors involved in ecolabelling, traceability and sustainable fisheries**
- **To foster improvement of data collection, assessment and governance in data poor contexts**
- **To improve visibility and quality of the data provided by the sources**
- **….**

Potential audiences

- **RFBs and their member states**
- **Seafood industry (from suppliers to retailers)**
- **General public**
- **National agencies** of governments dealing with fisheries reporting
- **Researchers and** Officers working on global **analyses on state of fishery resources**
- **NGOs** promoting sustainable fisheries

Main sources of Information

Comprehensive and shared global reference set of stocks and fisheries records

- **Fisheries and Resources Monitoring System (FIRMS)**
- **RAM Legacy Stock Assessment Database**
- **FishSource**
- **FishBase**

Other Sources

- **Chimaera (Portal to South West Indian Ocean Fisheries data)**
- **Fisheries Linked Open Data (FLOD)**
- **iMarine Top Level Ontology (TLO) Warehouse**
- **Geospatial data** (EEZ, FAO major fishing areas, etc.)
- **iMarine species data from providers** (FAO-ASFIS, GBIF, ITIS, OBIS, WoRMS,etc.)

Supporting applications

- **COMET (mapping services)**
- **COTRIX (Code lists management)**
- **GRADE (FLOD maintenance)**
- **iMarine data mining and annotations capacities**
- **MatWare**

Objectives for the 1st year of the GRSF

Core repository

- Combine records from the 3 sources to provide the critical mass to GRSF
- Unique Identifiers logic and mechanism for single stock/fishery
- Matching strategy for single stock/fishery across data providers
- Content Management System within a dedicated VRE for GRSF data
- Minimum data requirements for stocks & fisheries

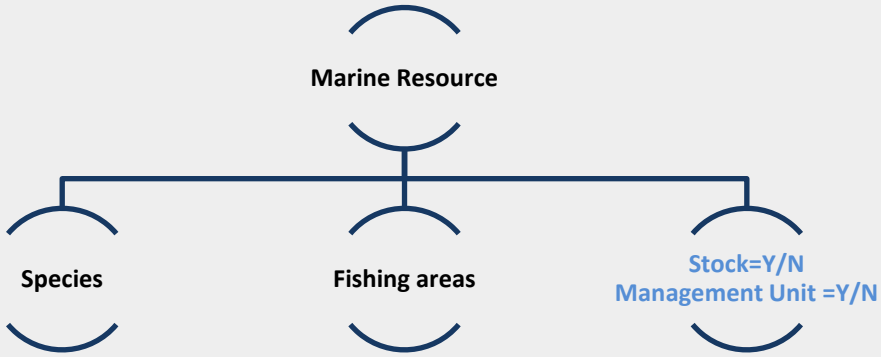
Extended Knowledge Base

- Collect additional requirements to enrich GRSF records with information through application of semantic harvesting and mining technologies (extended knowledge base)
- Formulate competency questions (users oriented searches)

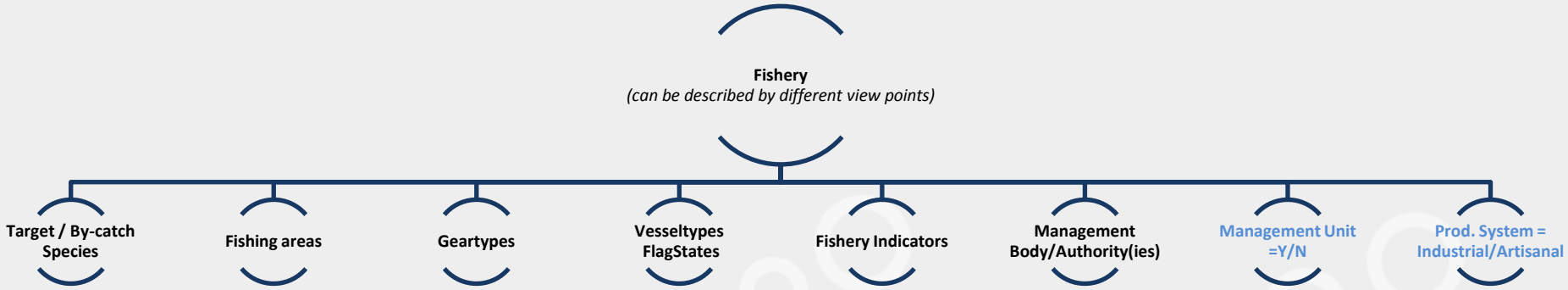
Foster participation of additional sources

- Identify additional information sources focusing on data poor stocks and fisheries (also in collaboration with regional projects)

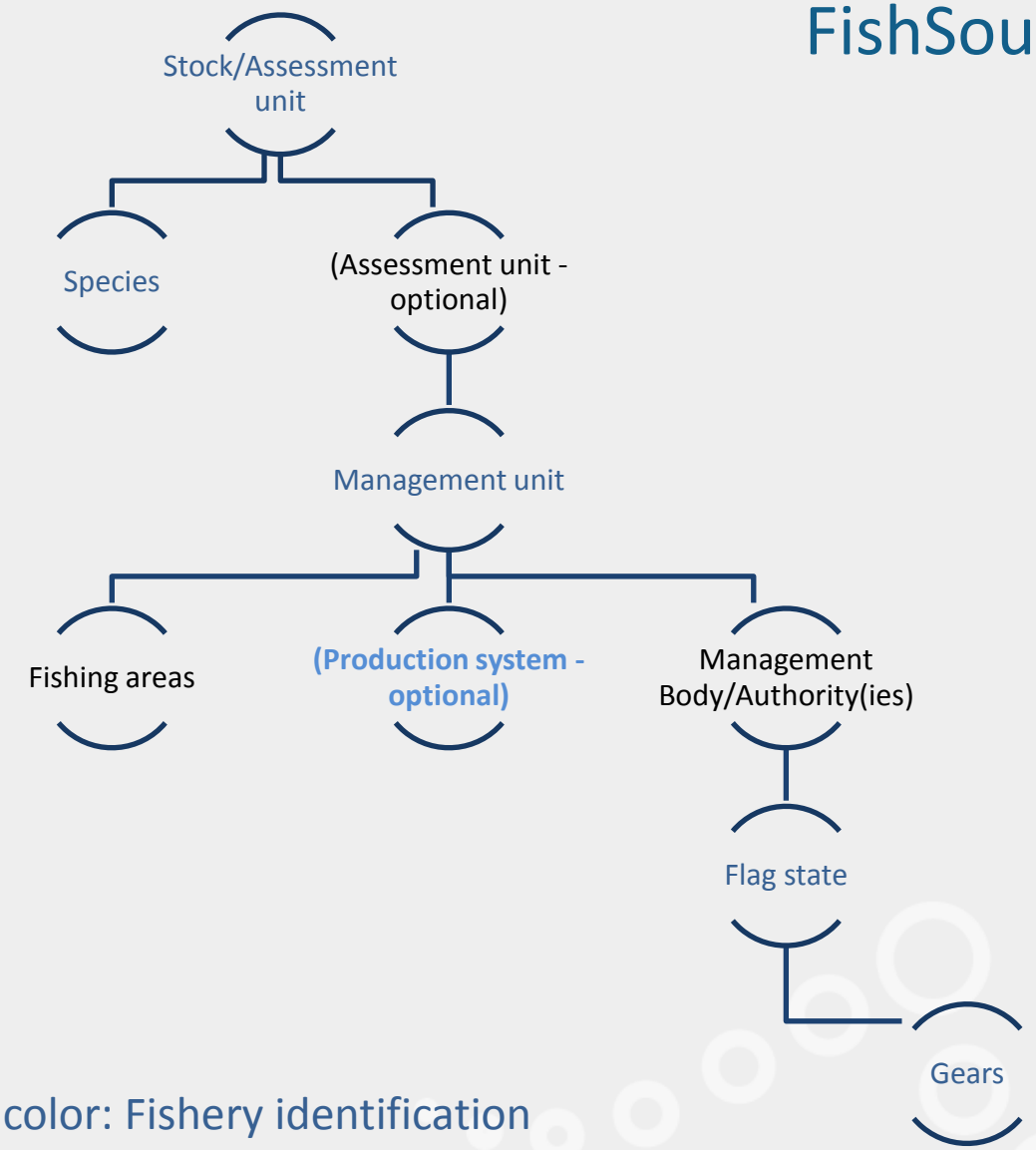
List of key concepts/data fields of the sources of information FIRMS, RAM, FishSource and relationships



FIRMS



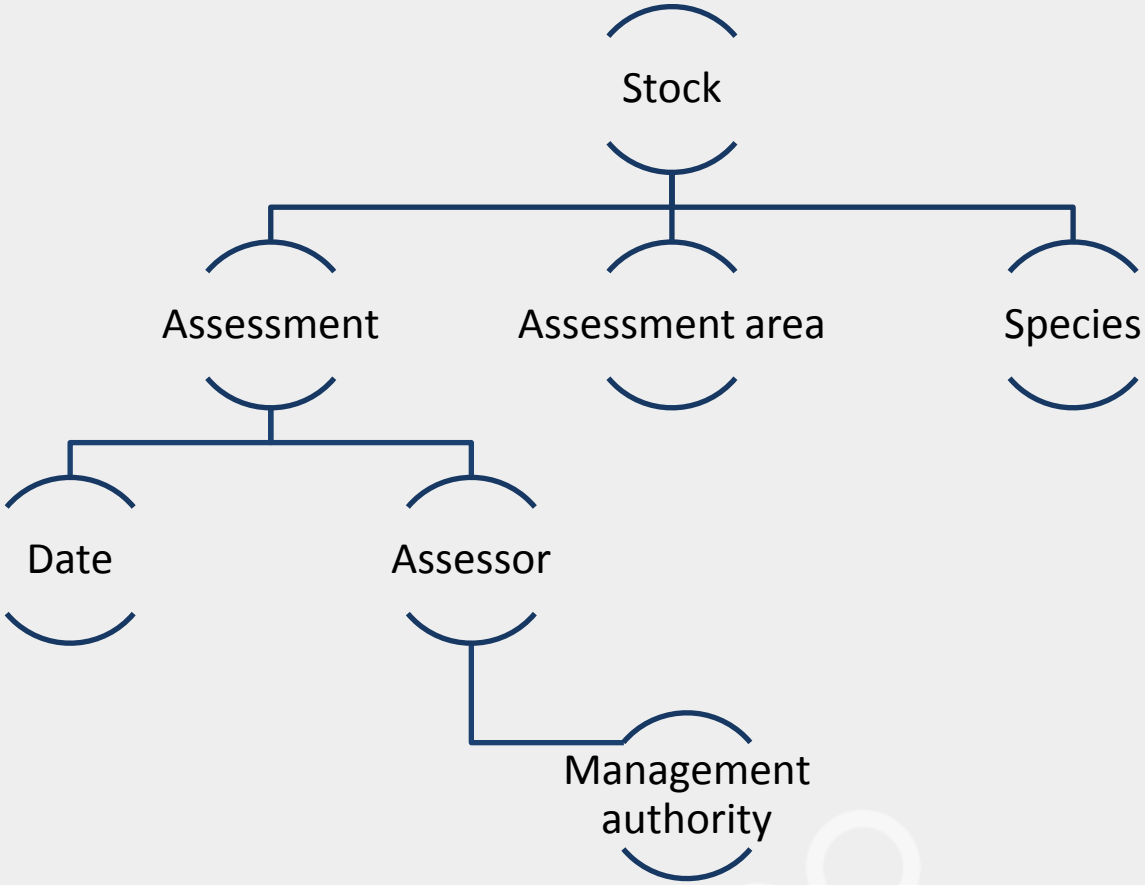
List of key concepts/data fields of the 3 sources of information FIRMS, RAM, FishSource and relationships



FISHSOURCE

Blue color: Fishery identification

List of key concepts/data fields of the 3 sources of information FIRMS, RAM, FishSource and relationships



RAM Legacy

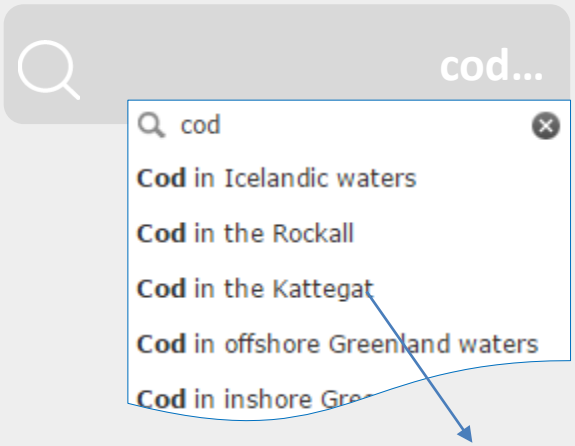
List of key concepts/data fields of the sources of information FIRMS, RAM, FishSource and relationships

	FIRMS		FISHSOURCE		RAM LEGACY	
	Concept	Description	Concept	Description	Concept	Description
STOCKS	Stock/Assessment unit ID	The field "*inventory identifier*" could be understood here also as a stock or as an assessment unit since sometimes makes reference to a portion of the stock managed under an authority (i.e. national level). The interpretation is made through the fields "Management unit" (yes/no) and "*Considered a stock*" (yes/no).	Stock/Assessment unit ID	Stock/Assessment unit/Unascertained	Stock ID	Stock ID
	Name of the stock	FIRMS standard name and local name of the stock	Name of the resource	Common name of the resource		
	Species Taxonomy Habitat	Taxonomy and marine habitat Scientific names, common names ClimaticZone, DepthZone, DepthBehav, BottomType, HorizontalDist, VerticalDist, GeoForm	Species Taxonomy Description	Taxonomy and description Scientific names, common names, scientific synonyms, FAO ID IUCN category,Image, ISSCAAP group	Species Taxonomy ---	Taxonomy Scientific names, common names
	Area of distribution	Georefencing standard zones (EEZ, RFB,FAO area...)	Area of distribution	FAO Fishing areas, occasionally use of local specific codes (i.e. Alaska coding system)	Area of distribution	Area where the stock was assessed
	Management	Description	Management	Management unit (management authority(ies) plus area, plus production system (artisanal/industrial)) and description (i.e. management reference points and narrative)	Management	Management authority
	Assessment Assessment methods	Type of model, name of model, uncertainty level Narrative, if available	Assessment Assessment methods	Narrative		

	FIRMS		FISHSOURCE		RAM LEGACY
	Concept	Description	Concept	Description	Concept
	Fisheries ID	Field "*inventory identifier*" and related fisheries (fisheries switching activity seasonaly, same fishery described from another national perspective or fisheries under the same management unit)	Fisheries ID	Name of fishery and description	---
	Fisheries title	FIRMS standard name and local name of the fishery	Fisheries name	There are not a specific field for the fisheries title, but a fishery common name can be defined by a combination of fields	---
	Species Taxonomy Habitat Bycatch	Taxonomy and marine habitat Scientific names, common names Environmental conditions of the fishing area: ClimaticZone, DepthZone, DepthBehav, BottomType, HorizontalDist, VerticalDist, GeoForm Target species, associated species, Discard species and protected species	Species Taxonomy Description	Taxonomy and description Scientific names, common names, scientific sinonyms, FAO ID IUCN category,Image, ISSCAAP group	--- --- --- ---
	Area of distribution	Georefencing standard zones (EEZ, RFB,FAO area...)	Area of distribution	Fishing areas	---
	Spatial scale	controleld terms (global, regional, national...)			
	Management Management Body/Authority(ies)	The field "Management Body/Authority(ies)" could be understood here also as a management unit. The	Management management unit - organizations	Mana	

Download at http://goo.gl/DizvHa

Minimum core data requirement - Stocks



The GRSF Database will contain:

- Mandatory elements
- Optional elements

And will be complemented by the extended knowledge base

Cod in the Kattegat

Stock Name & ID	Species Scientific name	Area	Exploiting Fisheries	Management Units	Assessment Method	Scientific Advice	State of Marine Resource (Exploitation rate, Abundance level)	DataBase Source	Source of Information	Data Owner
Unique Identifier	ASFIS	FAO, EEZ…	List of fisheries	Management Authorities	Type of model/name of the model	Narrative	F Mortality and B biomass (numeric values or descriptions) and set of fields for target limits or indicators	FIRMS/RAM/FishSource	Citation	RFB & other institutions
	X	X		X				X	X	X

X = Elements of the database functional for the identification of a stock among multiple sources

Minimum data requirement - Fisheries

Northwest Atlantic NAFO Flemish Cap groundfish fisheries

Fishery Name & ID	Species Scientific name	Area	Exploited stocks	Management Units	Production System Type	Flag State	Fishing Gear	Annual catch	DataBase Source	Source of Information	Data Owner
Unique Identifier	ASFIS	FAO, EEZ…	List of stocks	Management Authorities	Types: Industrial, Artesanal..	ISO3 country code	ISSCFG	Tonnes	FIRMS/RAM/FishSource	Citation	RFB & other institutions
	X	X		X	X	X	X		X	X	X

X = Elements of the database functional for the identification of a fishery among multiple sources

Comparison of list of stocks

	FIRMS			FISHSOURCE			RAM LEGACY		
	Title	Owner	URL	Title	Owner	URL	Title	Owner	URL
STOCKS	Patagonian Toothfish - Kerguelen Islands	CCAMLR	http://firms.fao.org/firms/resource/10524/en	---			---		
	Patagonian Toothfish - South Georgia	CCAMLR	http://firms.fao.org/firms/resource/10509/en	Patagonian toothfish - South Georgia	CCAMLR	http://www.fishsource.com/fishery/identification?fishery=Southern+bluefin+tuna+	---		
	---			---			Patagonian toothfish Macquarie Island	CSIRO	PTOOTHFISHMI
	Southern Bluefin tuna - Global	CCSBT	http://firms.fao.org/firms/resource/6010/en	Southern bluefin tuna (Country: AU; Gear: LL) FAO81. Australian eastern tuna and billfish fishery	CCSBT		Southern bluefin tuna Southern Oceans	CCSBT	CCSBT-SC-SBT-1931-2009-Parma
	Horse mackerel - Northwest Africa	CECAF	http://firms.fao.org/firms/resource/10095/en	---			---		
	Sardinella - Northwest Africa	CECAF	http://firms.fao.org/firms/resource/10090/en	---			---		
	Anchovy - Northern Spain	GFCM	http://firms.fao.org/firms/resource/13348/en	European anchovy - northern Spain (Mediterranean)	GFCM	http://www.fishsource.com/fishery/identification?fishery=Eur	---		
	Yellowfin tuna - Eastern Pacific	IATTC	http://firms.fao.org/firms/resource/21/en	Yellowfin tuna - Eastern Pacific (Gear: purse seines, longlines, pole-lines hand operated)	IATTC	http://www.fishsource.com/fishery/identification?fishery=Yell	Yellowfin tuna Eastern Pacific	multinational-IATTC-NEPAC	YFINEPAC
	Northern Bluefin tuna - East Atlantic and Mediterranean Sea	ICCAT	http://firms.fao.org/firms/resource/10014/en	Atlantic bluefin tuna - Eastern Atlantic and Mediterranean	ICCAT	http://www.fishsource.com/fishery/identification?fishery=Atl	Bluefin tuna Eastern Atlantic	multinational-ICCAT-EATL	ATBTUNAEATL
	Beaked Redfish - East Greenland and Icelandic Slope	ICES	http://firms.fao.org/firms/resource/10351/en	Beaked redfish - Icelandic slope	ICES	http://www.fishsource.com/fishery/identification?fishery=Bea	(only for NAFO)		
	Cod - Faeroe Bank	ICES	[illegible]	Atlantic cod - Faroe Bank	ICES	[illegible]	(Only for Faroe plateau, ICESVb1)		CODFAPL
	Atlantic Cod - Flemish Cap	NAFO	http://firms.fao.org/firms/resource/10319/en	Atlantic Cod - Flemish Cap	NAFO	http://www.fishsource.com/fishery/identification?fishery=Atl	Atlantic cod NAFO 2J		
	Patagonian toothfish - South East Atlantic	SEAFO	http://firms.fao.org/firms/resource/13375/en	---					
	Queen conch - Antigua	WECAFC	http://firms.fao.org/firms/resource/13107/en	(Only for Nicaragua and					

Download at http://goo.gl/wKSeoU

Unique identifiers

- There is a need for unique identification of concepts
  - To speak a common language across systems
  - To be able to combine, compare data / information from different sources
  - To enable automatic processing of information: indexing / comparison / aggregation
  - The most obvious needed one in this context is for stocks and fisheries.

  - There are two approaches for unique identifiers: one human readable and understandable, one machine oriented.
  - The unique ID can be used for the two activities:
    - Identification of unique records for populating the GRSF
    - Dissemination (coupled with the label)

Unique identifiers

- Which level of unicity do we want to match for unique stock
  - Any single report on status of stocks or fisheries
  - Any single stock assessment unit as defined for an assessment
    - Variants exist among assessments:
    - sometimes species within same family assessed together, then distinctly
    - sometimes the area assessed varies in scope
  - The ‘family’ of stocks assessment units monitored along a time series, at single species and ‘core’ area level

- Which level of unicity do we want to match unique fisheries

Unique identifiers

- Stocks and fisheries unique identifier: a human readable unique identifier for immediate understanding
  - Human readable: when data must be compared and reviewed with human action
  - The question to be asked is then: what makes the concept unique? For a species, the combination of scientific name + author defines its unicity. FAO has proposed a unique ASFIS code.
  - Same question must be asked for stock and fisheries

Unique identifiers

- The other approach is to have a machine readable unique identifier
  - A unique identifier for processes only operated by machine: data exchange (to ensure unicity of exchange information) and processing (to ensure traceability of source data after process).
  - This unique identifier can be human understandable: need to be carefully designed to ensure unicity (Cf. example: ASFIS code. 13824 max records with a 3 alphacode, and ASFIS has already 12600 records):
  - A unique ID generated from the object model serialization or randomly (the 128bits Universal Unique Identifier, UUID) : these unique identifiers are not human readable and can completely be disconnected from the concept itself (UUID). Is interesting when large amount of information exchange is involved and unicity must be guaranteed (uniquely identifying all VMS messages for the EU fleet)

Unique identifiers

- Recommendation: a human readable identifier
  - Unique identifiers are used here in the context of mapping (concepts, code lists). It seems relevant to the GRSF to have explicit and deterministic unique identifiers that can be used in the process of validating stock mapping in the CMS. The choice of these unique identifiers definition will be key.

Unique identifiers

Option 1 *(differentiating stocks from fisheries)*

- Stocks unique identifier for human understanding
  - What makes a stock unique? ➔ Species + area
- <FAO3AlphaCode> + <FAO Area Code / EEZ / RFB Area code / LME / Local Code> + <Management Unit Acronym>

- Fishery unique identifier for human understanding *(multiple entries per stock code)*
  - What makes a fisheries unique? ➔ Species + Area + Geartype + Management Unit +…
- <FAO3AlphaCode> + <FAO Area Code / EEZ / RFB Area code / LME / GRSF Code> + <Management Unit Acronym> + <ISSCFG> + <Prod System Type>

The outcome would be at least two labels one for stocks and one for the underlying fishery

Unique identifiers

Option 2 *(minimal elements approach)*

- A unique identifier applied
  - at single species level
  - in fishing area(s)
  - under a single management unit
  - including one geartype and one flagstate
  - specifying the type of production system
- <FAO3AlphaCode> + <FAO Area Code / EEZ / RFB Area code / LME / Local Code> + <Management Unit Acronym> + <ISSCFG> + <ISO3CountryCode> + <Prod. System type>

Example, Caribbean spiny lobster: SLC+FAO:31_EEZ:BHS+DMR-BHS+20.0.0+BHS+Ind_Art

The outcome would be one single label identifying the specific fishery elements

"Caribbean spiny lobster caught with miscellaneous gear in the national water of Bahamas by local industrial and artisanal fleets and managed by Department of Marine Resources Bahamas"

Code lists and ontological representation

- GRSF aims to provide access to information organized in a different way in 3 systems.
- In the same line of unique identifiers for stock and fisheries, there is a need to compare in a unique way the different pieces of information available in the 3 source systems.
- First step is to understand how information is structured in each system (Existing concepts and their relationships) to see if and how concepts can be compared: the first approach is the identification of the key concepts as presented before.

Code lists and ontological representation

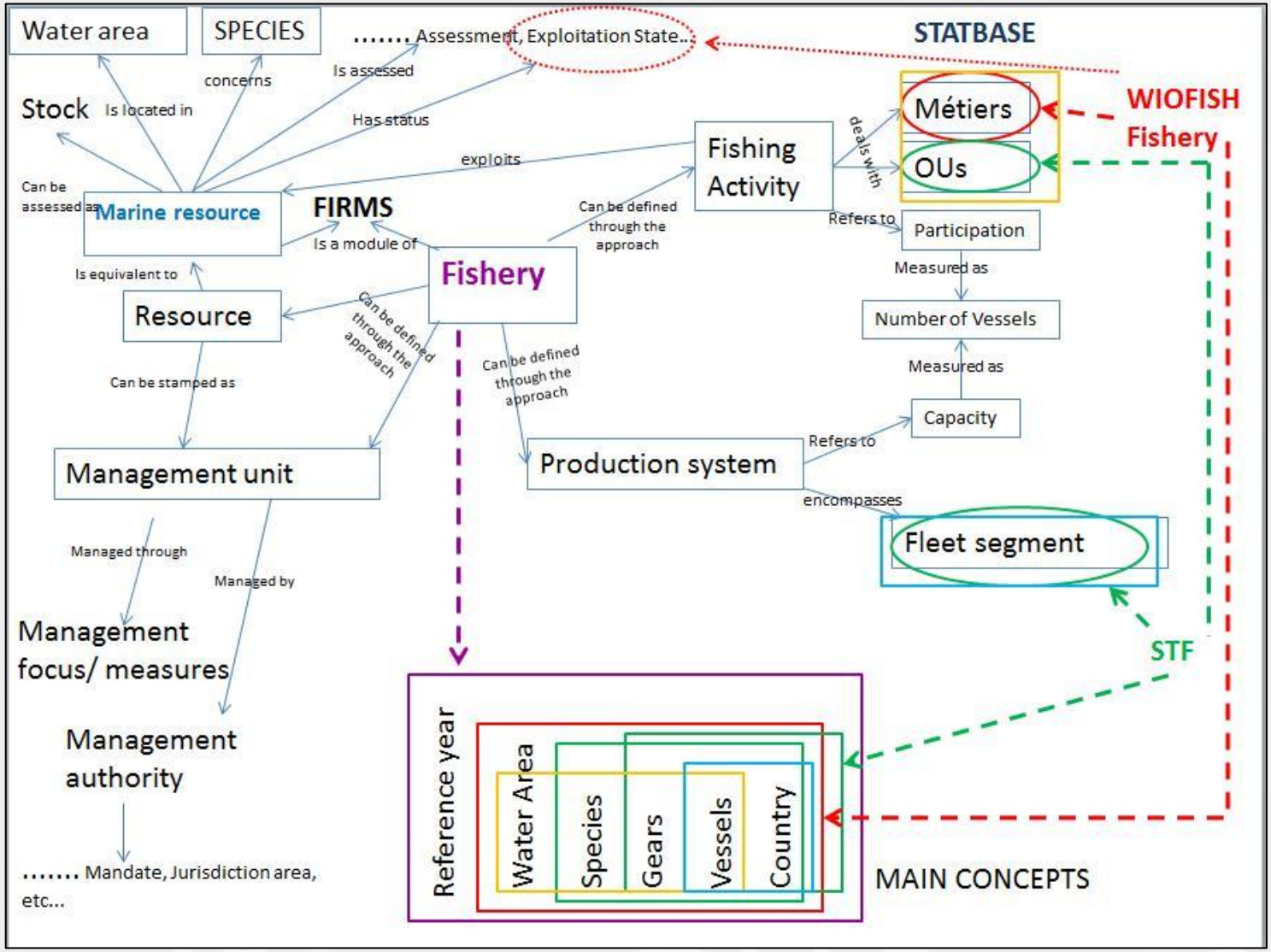
- Example: species seems to be a simple and shared concept across all fisheries systems: but it can be targeted species, by-catch species or discarded species
- Concepts and their relationship are presented in the system ontology, and ontologies can be compared (mapped) between systems to build a unique GRSF ontology
- Second step is to collect code lists used for each ontology's concept and map code lists to build gateways between systems definition content (Example: with gear types: line in system 1 = (pole line + long line) in system 2)

Code lists and ontological representation

- The final goal is to build tools that will exploit these mappings between concepts and between code lists to offer the end user tools to query from an unique entry point data and information coming from 3 different sources systems. It comes in complement to a “à la Google” generic search, providing thematic scalable research tools that can be expended on the short-medium term to include the extended knowledge database.

Code lists and ontological representation

- Example of the Chimaera ontology



Content Management System (CMS)

- The commonly accepted definition of a CMS is a computerized application to manage content in a centralized way: create, modify, delete, enrich, validate for publication a piece of content (can be news, blogs or shopping).

➔Publication workflow management

- 2 separated (but dependant) tools for (at least) 2 different users: an administration interface for the content management (*content manager*) and a publication interface for content browsing (*content user*)

Content Management System (CMS)

- CMS Administration = a tool to manage content: what is GRSF content ?
- Remember the initial GRSF definition "**Combine records from the 3 sources to provide the critical mass to GRSF**"
- CMS will propose administration tools to define the source of information and how these information are combined.
- CMS will allow to edit new sources from scratch

Content Management System (CMS)

- CMS Administration = example of tools
  - The CMS should provide tools to define how concepts are mapped between the 3 sources (management of the ontology), validate unicity of concepts (Example: manually validate the list of automatically mapped stocks and fisheries from the 3 sources systems based on the unique identifier previously defined). It could be a COMET like tools.
  - The content of GRSF is the result of automated process: the CMS must provide a tool to manage how the process is run: how and where the initial data sources are accessed. Is the process manual or semi-automated? How is this triggered when new data are available from authorized sources?

Content Management System (CMS)

- CMS Administration tool: who are the administrators / who has the responsibility in the validation of the definition
- CMS Administration = need to clearly define the publication / validation workflow to define the roles of each stakeholder (who creates mapping ? Who validates mapping? Etc…)

Content Management System (CMS)

- CMS dissemination = example of tools
- The machine-assisted indexing can allow automatic building of pages (aggregation of data coming from different sources in one block of information): the CMS dissemination management tool could define how these aggregations are organized when disseminating the information. Be aware of 3 source policies!
- List of thematic searches per topic could be managed through this CMS dissemination management interface: for a given topic, define the list of concepts allowed in the search refinement.

Extended knowledge base

- The extended knowledge base encompasses, beyond the core data requirements, all other data fields and information accessible from the stock or fishery records, including flat text files
- This extended information can be partially structured in order to accurately respond to most commonly identified competency questions
- Depending on his profile, the user will want/need to have access to certain types of information answering some his needs: "I'm a fisheries expert, I need to know all fisheries targeting a given species in a given area"
- These competency questions will lead to the identification of key concepts to be organized in the GRSF ontology together with relationships among them

Extended knowledge base

- The extended knowledge base is scalable;
- Basic need is to provide users with a generic tool to query any type of data and information
- More advanced needs require to set-up the extended KB for thematic searches: it's a very interesting tool to guide the user to find a certain type of information.
- building an extended knowledge base for a comprehensive set of thematic searches is a long process depending on needs expressed by final users and source systems data managers.

Extended knowledge base

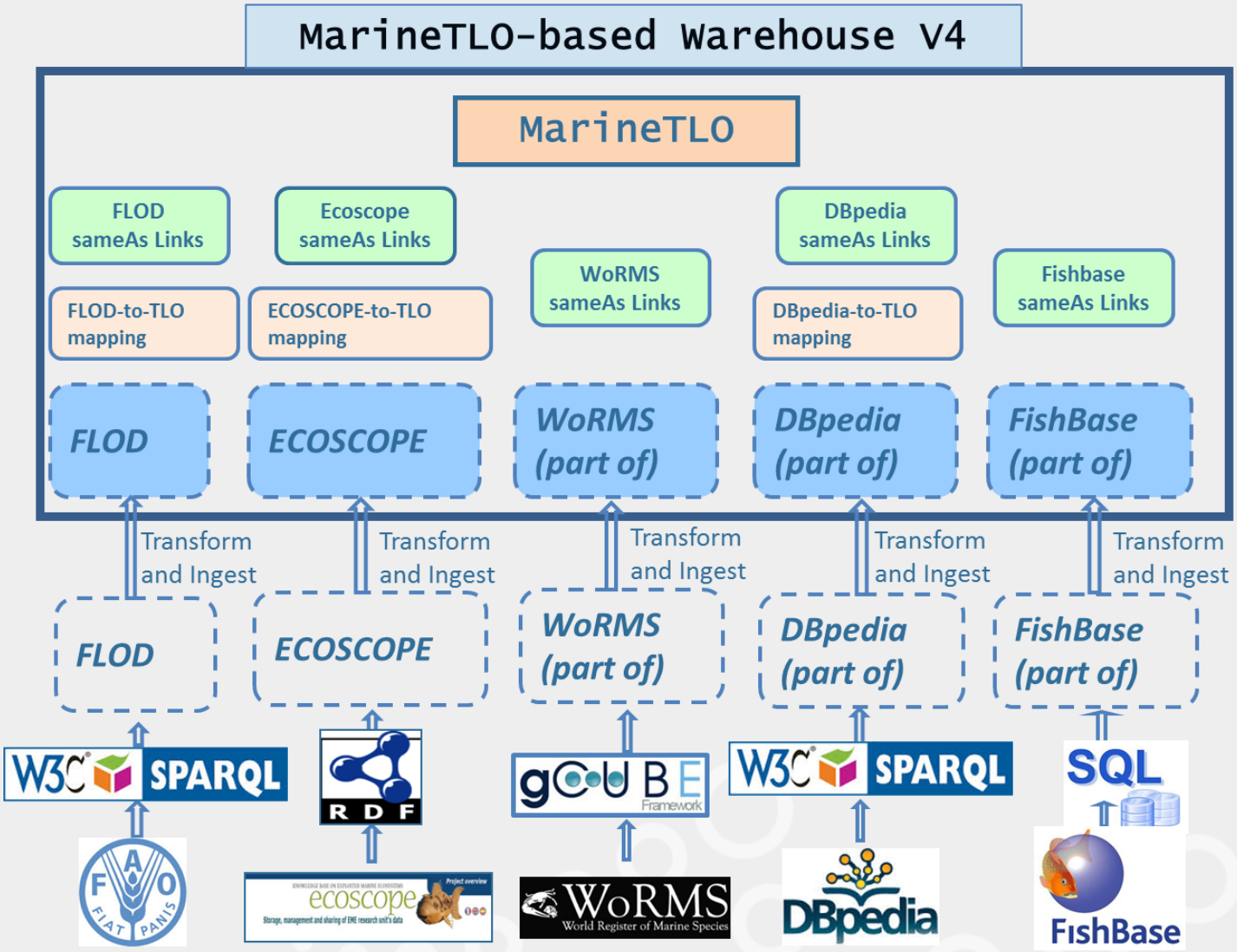
- This work has been already carried out during the Chimaera project (http://chimaera.d4science.org/)
- Data from FIRMS, WioFISH (a SWIO fisheries knowledge database) and StatBase (a SWIO fisheries statistical database) were analysed through the previously described ontological methodologies. Identified key Chimaera bricks for its ontology were species/gear type/vessel type/area.
- The extended knowledge base was built on the need for fisheries management information (stock status, fisheries management plans, measures etc…)

Extended knowledge base

- Lessons learnt from the project were:
- The more structured the information is available from the source, the easier the indexing is:
  - XML tag for FIRMS Vs PDF standard template for publication of WIOFish fisheries

    Simple control file for FIRMS Vs Complex indexing tool reading WIOFish PDF
- Data sharing and access policies: dealing with restricting policies makes indexing more complex (Cf. WIOFish: having access only to PDF files)
- Keep each system identity visible: the user is sent back to the initial source system to access the data / information (Require compliant policies and technicalities)

Extended knowledge base



Governance and sustainability

1. Content governance

- **data policy for access, managing and dissemination, citation, terms of use, ownership and infringements**

2. Data sharing policies

- **formats, interoperability, protocols, open data**

3. System governance

- **software policy, storage, data preservation, VRE management**

4. Sustainability business model

- **GRSF sustainability**
- **iMarine White Paper**

1. GRSF Content Governance

Content governance

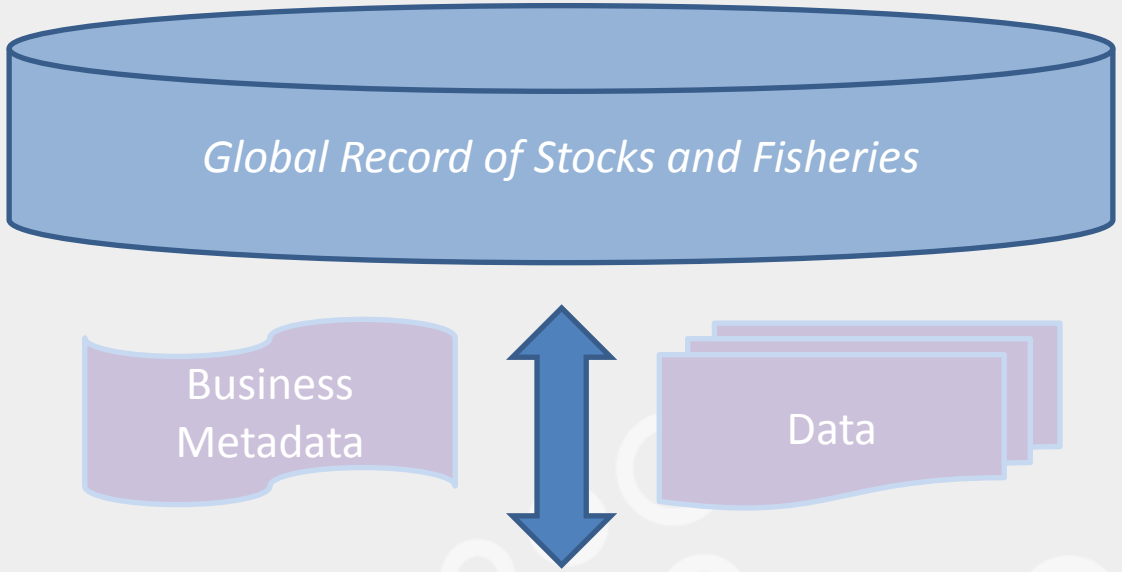
- data policy for access, managing and dissemination,
- citation,
- terms of use,
- ownership and infringements

The iMarine **Data Access and Sharing Policies** support the data sharing among contributors

1. GRSF Content Governance

iMarine **Data Access and Sharing Policies**

Business Metadata accompany the in-and-out flow of data to acknowledge and to preserve provenance and to provide instructions for derivative works



FIRMS Fisheries and Resources Monitoring System
RAM legacy
FishSource Status and environmental performance of fisheries worldwide
FishBase

2. Data Sharing Governance

GRSF data sharing policies

- are defined by the GRSF Partners
- extend iMarine policies
- may feedback with new requirements for iMarine policies

Policies encompass

- formats,
- interoperability,
- protocols for data sharing,
- open data, access, confidentiality requirements

The iMarine Board can ensure long term sustainability

- provides organizational structure
- defines the iMarine policies
- manages resources (data, computing)

2. Protocols for data sharing

Content management

- **Harmonized core repository**
- **Generic extended knowledge base**
  - With services which can be tailored to partners expectations
- **Principles driving optimization of partners' investments**
  - avoid duplication (single source of truth)
  - geographic complementarity
  - Topics complementarity (according to each system objectives)

Dissemination

- **Priority - Common GRSF services**
- **Potential - Specific services geared to serve each partner**

2. Protocols for data access

- Who can access
- Who get free access
- Who should register
- Etc…

3. System Governance

CMS System governance

Overall System governance requirements

- **software policy:** open source as per iMarine policies
- **Storage:** to be discussed
- **data preservation:** to be discussed
- **VRE management:** GRSF Administrator – who, what
  Content manager(s) – who, what aspects

Infrastructure governance

- **the iMarine White Paper proposes a roadmap**

4. GRSF Sustainability

Content

- **Sources** - maintained by contributing partners
- **Data sharing artifacts (master data, mappings, ontologies)** – managed by the 'GRSF Secretariat'

System

- **Core system -** under iMarine general Governance
- **Interoperable bridges –** under source systems responsibility
- **GRSF specific tools (Master Data Management, Ontology, LOD) –** under GRSF specific provisions

Resources to sustain

- **the iMarine White Paper proposes a roadmap**

4. iMarine sustainable Business Model

The iMarine White Paper provides a core model

- a public-led Partnership
- MoU's to operationalize collaborations
- a core legal entity
- a vison for sustainability, including:
  - **in-kind inputs,**
  - **sponsors,**
  - **membership fees,**
  - **subscription fees,**
  - **pay-per-use**

iMarine e-Infrastructure provides operational model

- joint responsibility for a complex system
- a governing structure => the iMarine Board
- a technical structure => D4Science core services
- a e-Infrastructure => the iMarine services

شكر

謝謝

Merci

Thank You

Благодарю

¡Muchas Gracias!