iMarine

Supporting data driven scientific advice workflows

Donatella Castelli
(CNR-ISTI)
Data Infrastructure supporting the implementation of the principles of the Ecosystem Approach to Fisheries Management and Conservation of Marine Living Resources

Interdisciplinary & multifacets collaboration at the local, national, regional and international levels

Knowledge generation
- Access to large amount of heterogeneous, distributed, across-domain data
- Rich data mining, analysis and processing capabilities
Data infrastructure for research
Data infrastructure offering services for the research communities

Data infrastructure
“e-Infrastructure” offering services for collection, deposition, storage, preservation, access, retrieval, analysis/mining/processing, publication, etc.

e-Infrastructure
Electronic platform operated by a responsible entity offering an open set of basic enabling services (including access to resources) to a distributed Community of Practice.
The Community of Practice (the community)

The e-infrastructure (the operational platform)

The system (the enabling sw system)
• **Distributed and dynamically created** environment
• where **subset of** data, services, computational, and storage **resources**
• regulated by **tailored policies**
• are **assigned to a subset of users** via interfaces
• for a **limited timeframe**
• at **very little cost** for the providers of the participatory data e-infrastructures

VRE Definition

1. Specify VRE metadata (including policies)
2. Select applications
3. Configure applications
4. Select data collections

Software deployment and hardware setup completely hidden
Evolving needs of its users completely supported
iMarine VREs (13 May 2015)

iMarine Gateway
https://i-marine.d4science.org/

- **Public VREs** (used to offer an application environment to a subset of users of a community)
- **Private VREs** (used for experiments, data access and preparation, and data analytics)
A single point to

- Get **status and updates** from applications and other users they are interested in;
- Get **notifications** about messages, jobs completion, new generated products, etc.

**Collaborative Environment**

- **Workspace**
- **Messages**
- **Search in your Workspace**
- **Notifications Page**

**Home Social**

**Share updates**

**User news feed**

**VREs user is a member of**

**gCube: just an overview**
Collaborative Environment

Feed from Users

Feeds from Applications
A single point to

- Manage data, store and preserve them
- Share data
• The Social Portal extended with

- Discussion topic
- Notification to a user
- Members page
- Member profile
- Member contacts

Share an update or a link, use “@” to mention and “#” to add a topic.
Management and interpretation of biological and ecological data in the environment

Complete full life-cycle data framework, from observational data to aggregated data repositories enriched with validation and analytical tools

Storage and interpretation of geospatial explicit information, including WPS processing

Flexible sharing, storage, reporting, search and retrieval, aggregation and projection facilities

A BUNDLE is a set of services and technologies grouped according to a family of related tasks for achieving a common objective.
Data-centered workflows

- Collaboration
- Sharing
- Visualization
- Publication
- Storage
- Processing
- Analysis
- Modeling
- Integration
- Enrichment
- Transformation
- Validation
- Access
- Discovery
- Collection

**BIOL CUBE**
- Occurrence and Taxonomic Data Discovery
- Occurrence Data Processing
- Species Distribution Modeling
- Species Distribution Maps Discovery
- Taxonomic Data Comparison
- Taxonomic Data Matching

**STATS CUBE**
- Code List Discovery
- Code List Management
- Statistical Engine
- Tabular Data Discovery
- Tabular Data Enrichment
- Tabular Data Management
- Tabular Data Processing

**GEOS CUBE**
- Geospatial Data Discovery
- Geospatial Data Processing

**CONNECT CUBE**
- Enhanced Documents Management
- Fact-sheets Management
- Information Object Discovery
- Messaging
- Shared Workspace
- Social Networking Facilities

21 April 2015, CNR-Pisa
Data collection: SmartForm VRE

On Board Bycatch recording

- Dutch Elasmobranch society
- SEAFO RFB, ..... 

Define forms, controlled vocabularies, ....

Select a Fishery survey

Validate & Enrich

Deposit

Analyse

Workshop on e-infrastructures supporting Food Safety Risk Assessment, 13 May 2015, Parma
Data curation & integration: Tuna Atlas VRE

Data collected by RFBs maintained in FAO managed DBs
- The Atlas of Billfish Catches
- The Global Tuna Catches by Stocks

Curation
Normalise data, delete columns and aggregate catch, delete duplicated records, aggregate valid duplicates, aggregate time dimensions, apply reference data codes, discovery error in codes, fix errors in codes, simple preprocessing, ....
Analysis: Scalable Data Mining VRE

- 100+ statistical models
- Transparent use of cloud computing
- Automatically generated interfaces
- Integration with R

Data preparation & access

Monitoring the status of the computation
BioDiversityLab: uniform access to data

Workshop on e-infrastructures supporting Food Safety Risk Assessment, 13 May 2015, Parma
The International Guidelines for the Management of Deep-Sea Fisheries on the High Seas. VME database to assist in informed decision making and the development of further measures to increase sustainability and reduce impacts.

VME record

Description (Habitat & Biology)

Specific measures

Historical information on fishing areas and closed areas

Meetings & other Sources of Information

General Measures

RFMO

Web site
http://www.nafo.int/

Regional Fishery Body fact sheet

Management Body/Authority:
Northwest Atlantic Fisheries Organization (NAFO)

The Northwest Atlantic Fisheries Organization's (NAFO) overall objective is to contribute through coordination and cooperation to the optimum utilization, rational management and conservation of the fishery resources of its area of competence, and to ensure the long-term conservation and sustainable use of the fishery resources and, in doing so, to safeguard the marine ecosystems in which these resources are found.
The Tuna Atlas factsheet

Fact sheet: describes an entity (species, vessel..) and its characteristics (including relationships with other entities) as well as linked data (images, publications, indicators...) coming from different sources.
**BlueBRIDGE main objectives**

1. **Blue Assessment**
   Assessing the status of fish stocks and producing a global record of stocks and fisheries

2. **Blue Economy**
   Analysing socio-economic performance in aquaculture

3. **Blue Environment**
   Fisheries & habitat degradation monitoring

4. **Blue Skills**
   Boosting education and knowledge bridging between research and innovation in the area of protection and management of marine resources

5. **Blue Commons**
   Publishing & re-use of scientific outcomes
AppliFish

www.i-marine.eu