SECOND I-MARINE BOARD MEETING REPORT

Chair: M. Taconet

1 MEETING SUMMARY RESULTS

1.1 PROJECT PROGRESS

Summary of progress of the project's Board related activities.

1. The iMarine Board Chair (M.Taconet) reported progress on activities described in the Board work plan. The second Board meeting is a key step to consolidate the vision for iMarine project developments. Overall, the project is on schedule, and several services can already be exploited.
2. The Board has overseen the consolidation of several activities initiated in the previous reporting period. The progress is reported in the Business Cases; BC1 for the support to EU Common Fisheries Policy, and BC2 for the support to FAO High Seas Deep Seas program; as well as for BC3 "support to tropical pelagic LME EA to fisheries implemented through co-funding with other projects.
3. In BC1, the iMarine project was instrumental to foster a dialogue on a unified approach to develop a unified reporting format for all fisheries operations data, spanning from vessel reported information, to scientific observer data, and statistical reporting. The BC can now exploit the statistical service, and the community will be engaged to validate the infrastructure to analyze their data. The importance of reliable reference data was stressed again, and the harmonization feature development facilitated by a Code List Manager will continue. Here, the availability of semantic features for the discovery and use of alignments was considered important.
4. In BC2, FAO continued to specify through co-funding the VME-DB requirements and the role of the iMarine infrastructure; on the biodiversity community side, the collaboration with the EA-CoP continued. The release of the taxonomic discovery functions, and the many features now available in the e-infrastructure to search and share biodiversity information as data-tables, maps and in reports was reported. FIN and OBIS were responsible for drafting realistic use-cases, and they remain committed to their implementation. For other use-cases, effective collaborations with data providers have been established. This enabled the specification of proposed reconciliation services that may find a fertile ground when they manage to offer alert and trigger mechanisms when updated taxon information one source is out of sync with other data sources. The modeling part of this BC requires closer collaboration between Board members and their EA-CoP to:
    a. Better advertise the existing services, and
    b. Elaborate the use cases that can be supported.
5. In parallel to BC2, FAO identified a co-financing opportunity and defined a Species Fact Sheets VRE that will be used to draft and publish (through the FAO web site) expert reviewed species profiles.
6. In BC3, FAO continued framing iMarine potential for the fisheries community through two projects: SmartFish where collaboration is already under way to utilize services offered by LOD and semantic services in which iMarine intervene in the value adding chain, and ABNJ Tuna which is at project preparation stage under the GEF framework, and which would potentially benefit from tools developed under BC1 and BC2.
7. The Board, through Work Package 3 (WP3), defined the high-level governance model, and how that will determine the planning for sustainability. The model is described in the iMarine wiki and the Board will be called to examine the elements of a Public Partnership sustainability model. Two topics were deemed of priority and action will focus on CoP Products and the Policies.
8. The same WP will exploit that model to define how VRE services are made available to the EA-CoP, including as products, data services, and collaborations, where also the iMarine wiki will contain detailed descriptions of requirements by Business Case.
9. Through its discussions, the Board identified a number of products or services that would contribute to a strategic positioning of iMarine on the market: building a registry of (UN led) global

geospatial data sources compliant with OGC standards, including through transformation services from EML/IPT to OGC, would contribute to feed GEOSS and demonstrate alignment with the EU/INSPIRE Directive; delivering to data providers transformation services to publish their data in RDF-LOD; harmonizing the SDMX produced through different DSDs; offering to scientists through the statistical manager reproducible data processing capacities; enabling scientists to access a global registry of data sources from their PC based "R" console; plugging web portals on top of the infrastructure to exploit views on the data made available in VREs, for the dissemination to policy makers (indicators) or to the general public (information type AppliFish).

10. Several partners stressed the importance of a data access and sharing policy framework, especially for Terms of Use or when re-publishing data. The importance of correct citations and references is important for data-providers to grant access for re-use of their data, for scientist to be able to re-produce tests, and for infrastructure users to enhance the visibility of their products. In the short term, MoUs might compensate for the lack of an established iMarine policy and help shape it.
11. In WP3, the identification and description of Data Access and Sharing Policies continued. A second draft has been made available. This will have to be elaborated further with inputs from iMarine Board and EA-CoP representatives.
12. The governance instruments that implement the policies should also include Software Policies for components that are developed in the Communities of Practice. The need here is bring this into the e-infrastructure as an integrated component, and there are technology challenges to overcome.
13. The Project Technical Director (P. Pagano) informed on the existing data policies that are already in place, and which can be capitalized on for new and fine-grained metadata driven policies. All datasets in the e-Infrastructure already have access level security (private / VRE / Public). However, future exploitation, or even extending into data-point security, of these services requires that a careful balance be struck between the effort needed to define and implement such policies, and the expected benefits the EA-CoP expects to draw from these policies. Also, the feasibility of establishing such advanced services within the limited time-frame of this project in competition with other features has to be further considered.
14. The Technical Director also presented the technology progress of the iMarine project, and the plans for developments. These were highlighted in detail in several joint sessions, open for both Board Members as well as representatives from the Technology Providers. The developments in Period 1 have resulted in services that improve the discovery and access of data in the Biodiversity and Geospatial clusters, and activities for Code List management and Work Flow support in the statistical cluster have commenced. The coming Months and Period 2 will see the improvement of data management services that could only be elaborated once solid data access services were provided in Period 1.
15. These achievements imply that support for formats such as FishFrame, FLUX and SDMX can now be expected to be developed, and that services exploiting Darwin Core data set will be released. The iMarine Board will have to develop cohesive use-cases and requirements for the technology providers based on interactive engagements with the EA-CoP.
16. Detailed progress reporting of interest to Board Members through the project technical clusters was provided by the respective representatives:
    a. For the statistical data management cluster, Anton Ellenbroek, FAO
    b. For the biodiversity cluster, Nicolas Bailly, FIN
    c. For the spatial data cluster, Hervé Caumont
    d. For the semantic cluster, Julien Barde, IRD

1.2 BOARD PLANNING

1. Thirteen of sixteen Board members attended this second Board meeting, which is a very good indicator of continued interest.

Board Membership

2. A few changes to the Board membership were reported: The change of Deputy Chair, held by E. Vanden Berghe, OBIS, and now replaced by N.Bailly, FIN, was confirmed. Ward Appeltans replaces Edward Vanden Berghe as OBIS representative. Tjess Hernandez (for VLIZ) attends for the first time the iMarine Board, following Ward Appeltans appointment at IOC/OBIS. The replacement of ESTAT member Matt Elliott with Friderike Oehler was accepted, and Hervé Caumont was accepted as alternate to F. Brito. The Observer Philipe Schoeneman was replaced by Ingrid Bergeret of MEDDE. Karl Morteo of FAO considers to resign. The status of the Board can be seen in the excel file.

Policies

1. The development of a Data Access and Data Sharing policy will continue with iMarine resources allocated to FAO. The second draft of the policy was discussed at the iMarine Board meeting, and the coming intersession is intended to elaborate a final draft for the next Board meeting (3).
2. All documents and data that are supposed to be under the control of a user of the system are in the scope of this policy. The policy document does not touch on the technicalities required to implement such a policy. For that, the Board has to elaborate on a Software policy.
3. The coverage of the policy not only addresses the policies to apply, but extends to governance issues, such as how to define a shared view within and across infrastructures on data ownership and custody, and how to properly guarantee confidentiality, attribution and integrity.
4. The instrument also has to cover the re-use and re-publication of information from many data sources. An iMarine policy must maintain and publish proper attributions for datasets that combine multiple datasets. The role and function of iMarine as a custodian needs to be discussed in detail for each product and service.
5. The discussion that took place under this item will enable the Secretariat to produce and distribute to Board members a final draft during the intersession for review at the next Board meeting.
6. The Chair concluded that during the coming reporting period a final draft of the Data and Access Sharing policy will be produced, and that all information for the iMarine software policy has to be collected. All partners were invited to share relevant documentation.

Harmonization

7. In BC1, the Board noticed an increased interest in the EA-CoP on studying the feasibility of a set of formats that cover the entire data flow in the BC1 domain. This would need extensive and reliable reference data mechanisms and the iMarine project has assumed a role in fostering such dialogue on these issues. The project is willing to offer support in the further development of such standards.
8. This activity already results in collaboration across Board members' institutes to establish a Community of Practice driving the development of interoperable data and infrastructures in support of fisheries operations data collection and exchange.
9. In BC2, the iMarine services that manage biodiversity data were released and provide an example of how expertise domain experts can now rely on infrastructure support.
10. The technical clusters were instrumental in offering a discussion platform to expertise domain representatives, even when not always in direct support to either BC. These expertise domains are

not defined on BC or specific knowledge domain, and often overlap. Their role in the project is to foster discussion on functionality, collect requirements, and to some extent design and plan implementation. One example are the environmental services required for the enrichment of biodiversity data, where geospatial experts engage in a discussion with biodiversity experts on the services required.

11. Immediately following the iMarine Board meeting, Cluster coordinators were invited to summarize progress and planning made with respect to the Board Work-plan for their EA-CoP. This will provide an overall progress report and an updated Board work plan. These updates are included in this iMarine Board report for review by the Board.
12. The Chair observed that iMarine has already had an impact by widening the scope of data that can be brought under a generalized set of standards in the fisheries operations domain. A shared set of standards and reference data for both Monitoring/Control/Surveillance/reporting, and scientific needs are now being discussed (the FLUX initiative). This can be considered already a success for the project, but it also requires timely and effective collaboration in the EA-CoP to consolidate the requirements. This will be a main Board activity for Period 2, closely related to the Data policies and governance inputs.

1.3 BOARD AND CLUSTER ACTIVITIES

13. The joint sessions outlined how the technology partners have taken up the clusters' related work plans for the past reporting period, and the planning for the next intersession. The presentations were largely prepared by CNR, T2, IRD, FIN, and FAO. In these presentations, summary activity reports were presented to provide the Board with facts and figures.
14. Details including short term activities can be found in the Meeting notes section, and will be progressively moved into the iMarine Board Channel and technical wiki.
15. **Summary statistical data cluster including code list management** (Anton Ellenbroek)
    1. The activities towards the vision for this reporting period as outlined by the Board Work Plan, are progressing well. For the reporting period, the increased dialogue on fisheries operations data reporting standards and the release of the statistical service deserve to be mentioned.
    2. Specific activities on **short to medium term objectives include:**
        a. Supported objective: **an ICIS (and VTI) VRE usable by fisheries statisticians CoP;** improve and persist curation (and other) settings; include curation in a workflow; add read and publish SDMX services, and complement these with other formats such as FishFrame, (to be considered as a beta version of FLUX). During the reporting period, ICIS was enriched with geospatial referencing components for the SPREAD use case. The SDMX activities of FAO and CNR are on-going.
        b. Supported objective: **a Code List manager/mapper for use by fishery and biodiversity CoP;** This will require the development of several functionalities; i) a code list manager that covers codes and reference data, time and space dimensions, hierarchies, translations; ii) a Code list mapper that will establish relations between codes in code lists; manage evolving relations, manage interpretations, errors (misidentifications); iii) a Rules-based distance calculation and mapping engine will support users' mapping process. The released code-list manager in the infrastructure will be enriched with a semantic technologies driven customized code list manager, published as Linked Open Data. This activity, led by FAO, will be bolstered in Period 2.

c. Supported objective: **ICIS (and VTI) harmonized statistics repository scientifically exploited.** This exploitation model implies enhance existing capacities in ways that attract existing communities (such as COST users) to work on the e-infrastructure. This includes i) analytical services in e-Infrastructure (Statistical service, Integrated R, Parallel R); ii) data exchange / Read-Write capacities between SDMX and R (SDMixeR); iii) Reallocation support (SPREAD). The EA-CoP has engaged in a dialogue with large data providers to define the contribution of the iMarine project to large initiatives such as FLUX, and aligning the outputs with EU directives such as INSPIRE. In Period 2, the requirements for COST, FishFrame and FLUX support will be analyzed, and the Board will have to produce a concrete use-case.

16. **Summary biodiversity cluster** (N.Bailly)
    1. Two gCubeApps; BiodiversityResearchEnvironment and EcologicalModeling were made available to the communities.
    2. In addition, progress was made in the AquaMaps VRE, and the validation started.
    3. Towards the realization of the vision of the Board Work Plan, the cluster will work on the following **Topics/focus areas:**
       - **Accessing main taxonomic and biodiversity repositories** through Darwin Core services, either as external infrastructures (GBIF, CoL, WORMS), or by hosting the database in iMarine infrastructure (OBIS)
       - **Enhanced data availability and enrichment/integration of data** through association of oceanographic data with bio-geographic data
       - **Enhanced quality control** through i) visualization, duplicates detection, outliers detection in 'geographic' space (i.e. based on the location) for bio-geographic data; ii) reconciliation of lists of taxonomic names, detection of duplicates and near-duplicates, identification of synonyms; iii) detection of outliers in environmental space (i.e. base on location and physical oceanographic measurements associated with the biological observation)
       - **Enhanced inputs to species predictive modeling** by increased quality of the data, and the availability of data enriched through association with physical and chemical oceanography
       - **Diversified - modularized species predictive modeling** through openModeller, R and other statistical processes.
    4. The plans for the coming period include establishing MoU's with data providers to ensure correct citation and referencing when data are re-used in the e-infrastructure.
17. **Summary geospatial cluster** (H.Caumont)
    1. The activities in this cluster are spread between CNR, Terradue, IRD, and FAO. The activities were presented in a technology session with project partners and Board Members.
    2. IRD outlined their approach to contribute through the Geospatial Cluster, and provide a CS-W endpoint to iMarine (in line with the INSPIRE Directive), with entries providing URLs to data access services.

       IRD Catalog:
       i. Uses GeoNetwork, includes ISO 19119 Service Metadata documenting data access endpoints
       ii. Using Terradue's software: possible to transform such ISO/OGC metadata into RDF triples (as a bridge towards CS-W 3.0 OpenSearch) - Scale: Global or Regional
       iii. Types: Coastal, Deep Seas
    3. CNR commented on IRD proposal in the session. There may be a risk that iMarine becomes catalog of geospatial resources. iMarine can deliver more and the Board should advocate

developing processing capacity and specific subscription processes and services. CNR commented that having 2 approaches requires balancing effort.

4. Terradue shall undertake this discussion point further with IRD and CNR in order to formalize the complementarities and compatibilities of both Spatial Data Infrastructure (SDI) and gCube perspectives.
5. There is also a need for the Geospatial Cluster to provide a reconciling vision for these INSPIRE (SDI) and iMarine (gCube) approaches, and the way they jointly answer the project's requirements. Considering the two main iMarine business cases (fisheries, Deep seas), they lead to discussions on the service sustainability in a distributed system or infrastructure.
6. Key sustainability factors such as OpenSource developments and added-value approaches on top of the databases of the providers were discussed. From there, the particular 'service' of Occurrences Point Enrichment was considered, for which IRD code can be adapted, building on FIN requirements that have been published on the topic.
7. One required step in Period 2 will be to review the Essential Observation Variables (EOV; environmental parameters like SST) that could be considered, including taking possible insights from the work conducted within OpenModeler.

18. **Summary semantic cluster** (J.Barde)
    1. A first set of applications has been set up to demonstrate the interest of this approach from a theoretical point of view. However, the use cases scenarios must now be driven by the needs of the communities of users.
    2. The different activities related to the Semantic cluster are held by FAO (**FLOD**), FORTH (**xSearch**) and IRD (**Ecoscope**). The first ten months have been dedicated to dedicated to compare approaches and identify similarities among needs expressed by communities of involved partners, and to manage technical issues in order to enable these partners to use the existing tools and knowledge bases (KBs).
    3. With FORTH exists the possibility to increase rdf generation, such as with Agrotagger. This turns biographic references into uri's and makes these available in the search. Similar to highlighting literals, it is entity mining. This activity brings repositories closer to clients.
    4. The development of a TLO is important. The design has to stay close to needs of users, not of researchers. IRD recommended a simple ontology and a use case with a fact sheet generator.
    5. Summary of activities: So far, the focus of the various activities can be summarized in three key areas (see Figure 1) whose aims are:
        i. ***Sustainable generation of RDF*** (or Linked Open Data, LOD) from sources of heterogeneous data formats and access protocols.
        ii. ***Scalable storage and retrieval of RDF*** from a single or distributed knowledge bases through dedicated endpoints. Work on Top Level Ontology,
        iii. ***Publishing and consuming RDF*** with client applications: search engines / generation of landing pages populated with data from remote LOD by using ***(2)***; enhanced search by exploiting network of connections in LOD (query expansion: multiple languages, synonyms); web content enrichment with LOD (e.g. for fact sheet generation).

2 MEETING NOTES - PRESENTATIONS, DISCUSSIONS, AND DECISIONS

Wednesday, October 3, 2012; iMarine Board - TCom Joint meetings

2.1 GEOSPATIAL CLUSTER

Presenters: F. Brito (Terradue) *Slides*: .pptx file and G. Coro (CNR) *Slides*: .pptx file

Participants: F. Brito (Terradue, Chair), A. Antoniadis (NKUA), W. Appeltans (UNESCO), N. Bailly (FIN), C. Baldassarre (FAO), J. Barde (IRD), C. Bekiari (FORTH), E. Blondel (FAO), L. Candela (CNR), V. Canhos (CRIA), H. Caumont (Terradue), P. Cauquil (IRD), G. Coro (CNR), A. Ellenbroek (FAO), P. Fabriani (E-IIS), N. Holdsworth (ICES), G. Kakaletris (NKUA), A. Manzi (CERN), Y. Marketakis (FORTH), P. Pagano (CNR), M. Taconet (FAO), E. Travaglino (E-IIS);

Progress summary from the Geospatial cluster chair:

Progress is made in integrating components in the e-infrastructure. The community should now define a concrete list of data sources connecting to Business Cases to be made available in the infrastructure.

The GeoSpatial cluster goals/objectives are:

1. **Data discovery** in main repositories of environmental, fisheries and biodiversity geospatial sources like netCDF-CF repositories; My Ocean, SeaDataNet, IRD World Ocean Database, Ocean data portal and other iMarine partners geospatial sources (defined by OBIS, IRD, FAO, ICES);
2. **Data access** in main taxonomic and biodiversity repositories. Increased granularity in data access with OPeNDAP, WCS, WFS (including GML2-3). Defining geospatial needs (including product subscription approach) and their availability;
3. **Data processing**; R geospatial, WPS/HADOOP, FAO Intersection Engine potential, biodiversity data enrichment, geospatial analyses and modeling. Define enrichment needs of bio-ecological or activity occurrences with environmental data: OBIS Ocean Physics, VTI, VME Design and planning for enrichment capacity. Advanced geospatial analytical and modeling features - e.g. using R geospatial for reallocation and aggregation. Defining advanced geospatial processes required in reallocation, aggregation, interpolation. Designing and planning implementation for geospatial processes capacity.
4. **Data visualization** Interactive map viewer enabling visualization of geospatial biodiversity and fisheries data. Defining visualization functionalities for SPREAD, OBIS, VME-DB, VTI. Design and plan implementation for visualization features.

Discussion:

- Terradue proposes to focus on the variables (e.g. SST, Salinity) to be made available, then to prioritize variables and finally to identify the potential providers/data sources;
    - Identify variables (SST, PAR, Ice concentration) => prioritization
    - Identify datasources (repositories) => prioritization
    - Identify tools (including data policies) => prioritization
- Terradue addresses similar challenges in other infrastructure in Europe. The key is how the infrastructure addresses downstream services. The next question is how the infrastructure assists scientists. For that, scientists need to see what data is available (e.g. environmental data, time-series) in terms of quality, time/spatial coverage, etc. iMarine should strive to identify a unique position, e.g. by providing a dashboard to seek and customize access to environmental data from other infrastructures.

- The scale issue of global variables or local variables can be added. Variable prioritization is also driven by data coverage (temporal and spatial) and quality.
- Yet another approach could be to build a catalogue of products and services by relying on OGC standards, e.g. register a CSW URI; E.g. SeaDataNet is a catalogue of CSW services;
- The project technical partners pointed out that this might have limitations on effective data exploitation as we become limited to the protocols supported. A catalogue of OGC compliant sources is effective for resources discovery and direct access, but we can offer more capacities for processing.
- IRD data providers are requested to describe their data and products by ISO19115 and ISO19119 metadata formats (this is part of INSPIRE);
    - they work with a company developing code to transform NetCDF files to ISO standards;
    - it is expected that the infrastructure links data sources compliant with such formats;
    - clients are willing to use / are expecting to use these approaches;
- In addition to the services described, existing resources of iMarine partners can be re-used, or further extended. An example is the FAO Intersection Engine that provides services and data to allow the reallocation of catch statistics within i-Marine.
- Several partners raised the issue of sustainability of the federating infrastructure that depends from the sustainability of the resources providers. These providers concerns should be listened to and taken into account in the envisaged policy;
- The Chair summarized the discussion. The main point being that the Board should start identifying the needed variables while connecting to Business Cases, then the resources and define their scope.

2.2 SEMANTIC CLUSTER

Presenters: J. Barde (IRD) *Slides*: .pdf file and C. Baldassarre (FAO) *Slides*: .pptx files

See Document iMB2/2012/4a and Document iMB2/2012/4b

Participants: J. Barde (IRD - Chair), A. Antoniadis (NKUA), W. Appeltans (UNESCO), N. Bailly (FIN), C. Baldassarre (FAO), C. Bekiari (FORTH), E. Blondel (FAO), F. Brito (Terradue), L. Candela (CNR), V. Canhos (CRIA), H. Caumont (Terradue), P. Cauquil (IRD), G. Coro (CNR), M. Doerr (FORTH), A. Ellenbroek (FAO), P. Fabriani (E-IIS), S. Garavelli (TRUST-IT), N. Holdsworth (ICES), J. Keizer (FAO), G. Kakaletris (NKUA), A. Manzi (CERN), Y. Marketakis (FORTH), P. Pagano (CNR), M. Taconet (FAO), E. Travaglino (E-IIS);

Progress summary from Semantic cluster chair:

The different activities related to the Semantic cluster are held by FAO (FLOD), FORTH (xSearch) and IRD (Ecoscope). The first ten months have been dedicated to compare approaches and identify similarities among needs expressed by communities of involved partners, and to manage technical issues in order to enable these partners to use the existing tools and knowledge bases (KBs).

A first set of applications has been set up to demonstrate the interest of this approach from a theoretical point of view. However the Board reaffirmed that use cases scenarios must now be driven by the needs of the communities of users.

Summary of activities

So far, the focus of the various activities can be summarized in three key needs (see Figure 1) whose aims are:

- ***(1) Sustainable generation of RDF*** (or Linked Open Data, LOD) from sources of heterogeneous data formats and access protocols.
- ***(2) Scalable storage and retrieval of RDF*** from a single or distributed knowledge bases through

dedicated endpoints. Work on Top Level Ontology,

- ***(3) Publishing and consuming RDF*** with client applications: search engines / generation of landing pages populated with data from remote LOD by using ***(2)***; enhanced search by exploiting network of connections in LOD (queries expansions: multiple languages, synonyms..); web content enrichment with data pooled from LOD (eg fact sheets generation).

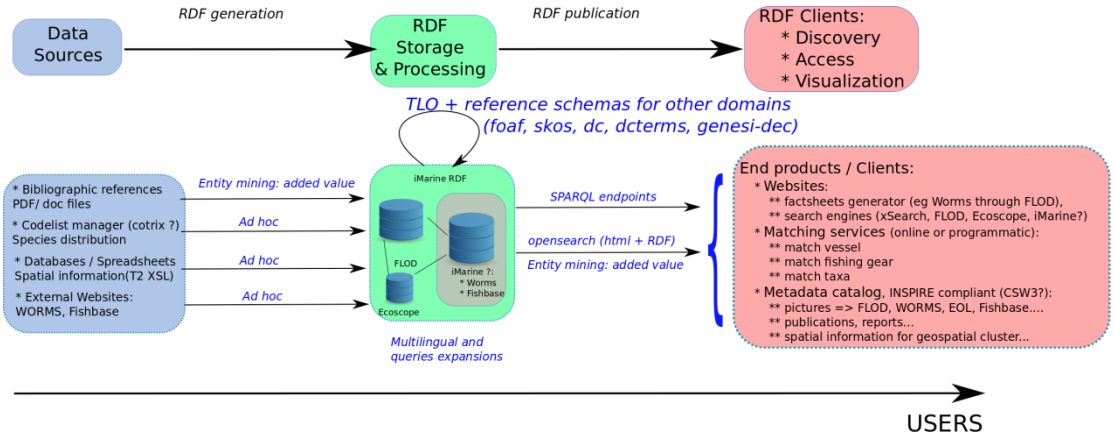

Figure 1: main groups of activities of the semantic cluster

Involvement of partners (first 10 months)

The first period of iMarine enabled to assess the data asset, the tools and the approaches brought from the partners into the project to achieve these three needs:

- ***(1) Sustainable generation and maintenance of RDF****:* FLOD approach includes the usage of java components that adapt in reading the data source (mainly code lists), and produce RDF with writers ad-hoc for the data type found in the source. The Ecoscope approach is similar but depends on data types: drivers can be either specific (for relational databases with ad hoc SQL queries, or for image annotations that cannot be automated so far) or generic when described by other standards (OGC metadata for spatial information or publications from Zotero RDF export).

- ***(2) Scalable storage and retrieval of RDF:*** both FLOD and Ecoscope have a triple store with Apache Jena[1] technology to store the RDF generated in ***(1)***. The first application layer on top of the triple store is a SPARQL endpoint where semantic clients can send SPARQL queries to retrieve a subset of data focusing on the relationships holding between them. FLOD and Ecoscope implemented their endpoints sharing the same technology named Joseki[2], further incubated by Apache with the name of Fuseki[3].

- ***(3) Publishing and consuming RDF****:* on top of the SPARQL endpoint, more consumer applications have been built to hide the complexity of RDF and SPARQL in ***(2)***: search engines to retrieve and browse the content of KBs (FORTH xSearch), an OpenSearch[4] interface for most common query capability, and graphical user interfaces to present the entities and relationships in the KBs in a friendly way. The need to refer homogeneously to the entity types contained in both FLOD and Ecoscope generated the need

1 http://jena.apache.org/

2 http://www.joseki.org/

3 http://jena.apache.org/documentation/serving_data/index.html

4 http://www.opensearch.org/Home

for a common model: the Top Level Ontology (TLO) is a transversal activity (led by FORTH) dealing with all aspects of knowledge management, in the domain of Ecosystem Approach to marine resources.

So far, FAO and IRD deal with activities (1), (2) and (3) and FORTH mainly with (3). However FORTH work on TLO is expected to impact all activities and entity mining application can help to generate additional RDF.

Next steps with "relevant" use cases for the coming Months:

- **Methods for RDF publishing and querying:** The Top Level Ontology (TLO): an initial design will be provided by FORTH driven by user requirements gathered for the dynamic fact sheet generator (expected in November 2012).

- **Data sources of users to be published in RDF.** The generation of new RDF triples (TLO compliant) to enrich existing KBs has to focus on relevant data sources from the users point of view. The suggested sources are:
  1. Bibliographic references made available from users (Zotero use case, cf above). The goal is to populate KBs with existing references that are going to be related with URIs of entities (species, fishing gears..) and agents (researchers, projects..) of various KBs by using FORTH entity mining application.
  2. Documents annotation: annotation of text based documents (words and pdf files, cf above) will enable relationships with agInfra project and old documents to be made available online.
  3. Relational databases: to enable the mapping of databases content with the TLO in order to facilitate the generation of RDF from SQL queries.
  4. iMarine data sources dealing with data sources from partners of the project (e.g. Fishbase, WoRMS..) such as: code lists or species manager, spatial information metadata (GeoNetwork of geospatial cluster with OpenSearch of Terradue, etc.). This activity finds a synergy with the work in WP11 on disseminating data in RDF format. WP11 creates, in fact, a good starting point to re-factor RDF disseminated data into LOD format and conventions.

- **Storage and retrieval of iMarine RDF.** Resulting RDF triples will be either managed in existing KBs or in new ones set up in the framework of iMarine with dedicated SPARQL endpoints on top of them.

- **Client applications using iMarine RDF**
  1. Search engine: setting up an iMarine search engine to browse existing data and information (resources) from different iMarine sources, by reusing principles of current partners search engines and by extending the kinds of information resources that are dynamically gathered (images, maps, statistics, reports, illustrations, people and other documental resources) from distributed RDF repositories.
  2. Dynamic factsheet generation: iMarine fact sheet generator will aim to set up factsheets by querying the various SPARQL endpoints made available (with the TLO schema),
  3. Document annotation: iMarine annotation service (cf Agrotagger) for (old) pdf, doc files and zotero RDF export.
  4. Entities Matching: iMarine matching services: "match taxa", "match vessels"...(similar to Worms "match taxa" service).

Discussion:

CNR aims to offer an incremental approach. The TLO should evolve. Request is entity mining on textual documents. The current infrastructure offers little support to these applications. It is important to note that these were not requested or included in the project DoW, and they have evolved in the project.

IRD proposes to start with OpenSearch and Sparql to use the TLO over Ecoscope and FAO-resources. The next question then is how to generate more relevant RDF to include. The Code list manager, Geoserver and other repositories could be RDF sources. A proposal will be distributed.

The Technical Director noted that more is needed to enable semantic technologies to exploit infrastructure resources. For instance, an AquaMaps must be an entity with a description, but these are not yet published. This effort will require that we change the use of technologies. Our current metadata are not rich enough. But indeed this would grant iMarine with a service which would allow Partners to generate RDF data sources. Propose to discuss the proposal first in PEB.

C. Baldassarre: FLOD has been developed in FAO, but one of our expectations is that we can outsource the management of the facilities developed to the infrastructure;

- this aspect should be discussed; a concrete plan and a deep knowledge of the technology is a prerequisite for this;
- the content, coverage and maintainability of the KB must be better documented;
- the scalability and stability of the components and their ensemble must be understood;
- the use cases and sustainable exploitation models need further elaboration.

The Top Level Ontology (TLO) is a fundamental approach for a number of use cases:

- it 'bridges' between knowledge domains and concepts;
- it is a prerequisite for concept driven data exchange between infrastructures;
- it describes the knowledge boundaries and concept information structures;
- it sets the lower boundary of required information with a concept.

This TLO was discussed in various parallel sessions beyond the Board meeting. Reports can be provided to Board members on request.

The Board, (through J. Keizer) asked which RDF data sets are already (publicly) available, how they are maintained, and if they are exploitable by outside 'clients'. It was agreed that AgInfra and iMarine engage in a discussion to plan a collaboration experiment. This could start with exposing content to each other.

iMarine (C. Baldassarre) provided a future vision for FLOD and TLO activities that should strive to open data that can be exposed through a SPARQL end-point for the entire infrastructure. A single SPARQL end-point would offer several benefits at project level:

- Reduce development and operational costs;
- Offer economies of scale to user groups wishing to exploit semantic services;
- Enrich the e-infrastructure if components are embedded in gCube.

Thursday, October 4, 2012; iMarine Board - TCom Joint meetings

2.3 BIODIVERSITY CLUSTER

Session chair N.Bailly FIN (see Document iMB2/2012/5), *Slides*

Participants: N.Bailly (FIN - Chair), A. Antoniadis (NKUA), W. Appeltans (UNESCO), C. Baldassarre (FAO), J. Barde (IRD), E. Blondel (FAO), L. Candela (CNR), V. Canhos (CRIA), D. Castelli (CNR), H. Caumont (Terradue), P. Cauquil (IRD), G. Coro (CNR), A. Ellenbroek (FAO), G. Farantatos (NKUA), S. Garavelli (TRUST-IT), F. Hernandez (VLIZ), N. Holdsworth (ICES), G. Kakaletris (NKUA), V. Marioli (CNR), J. Neves (NEAFC), F. Oehler (Eurostat), P. Pagano (CNR), F. Simeoni (FAO), F. Sinibaldi (CNR), M. Taconet (FAO), R. Tsantoulli (NKUA), E. Van Ingen (FAO);

Progress summary from Biodiversity cluster chair:

The Session Chair introduced expected features and use cases through slides;

Use cases for which the project can develop services are:

- Species name reconciliation
  - Darwin Core related tools
  - Develop new tools and integrate existing tools such as Taxamatch;
  - Integrated data and data services in the e-infrastructure
  - Alert and update mechanism; 1 species a day of fish
  - The goal is name reconciliation VRE. Asks Board for opinion on i4Life cross checklist.
- Species occurrence data tools
  - AquaMaps existing VRE exploitation plan
- Statistical features VRE
  - for e.g. outlier detecting in DwC datasets
  - Map comparison; starting with a simple visual comparison
  - Duplicate detection
  - Quality control
  - Clustering in environmental space, not geographic space
  - Proposes a closer collaboration with geospatial cluster
- Engage Communities

Demonstration of **features** (CNR) Slides

Discussion:

- The delivered features are in line with the Board expectations,
  - The data access features are well-advanced;
  - The possibilities for data reconciliation are promising;
  - Care must be taken to properly heed data provider's data policies.
- Modifications to meet specific users' requirements need to be documented
  - By existing cluster participants; initial comments included:
    - Search priorities (e.g. search through WS, add a repository)
    - Save a search result as a DwC in WS,
    - Compare two DwC instances (e.g. at family level)
    - Trigger feed-back to data providers
    - Ensure copyright and attribution in all products
  - Through a validation exercise driven by WP3
    - This will be based on a set of expected work-flows.
    - These scenarios will be defined in the community,
    - Unesco/IOC explained the OBIS scenario, with data updates every 2-3 months that needs to be compared to other data providers.

    - Unesco/IOC want to study the establishment of a node in VLIZ
    - FIN explained the need for 1:1 comparison of species information, where on average 1 species per day is added to "science for fish" only
- Data access and sharing issues
    - Attribution and copyright to be managed at dataset level
    - Modification tracking and local persistence of 'working sets', e.g. a taxon to be reviewed would be stored in the user work-space.
    - There are no plans to negotiate data update agreements with large data providers such as GBIF. The plan is to develop tools that assist users of the infrastructure to modify data-sets. How these are reconciled with the data providers, and how these are updated in the data sources is another issue.
    - A triggering and notification mechanism can be a useful addition to the infrastructure, and it was proposed to define a project position.
- Data modeling and statistical analysis
    - The cluster leader repeated his list of services for modeling and analysis. A first step would be access to a host of visualization services to generate maps on the same data. The example of lampreys was given, where now 4 maps can be compared.
    - For bycatch analysis, the access to larger volumes of data will improve the predictive and analytical model quality. However, the tools are very specific, and will require that users can load and extract data into their specific models (Example, a gear specific R model loaded in the e-infrastructure R-environment)
    - The EA-CoP will have to further define these opportunities. IRD offered to provide feedback and drive some propositions and run such analysis on top of the RDB. Also ICES which has a working group on by-catch is willing to consider a contribution, and assist in the identification of suitable data sources and modeling expertise.
    - Chair: Proposes to 1. run a few models, 2. Present at RFMO meetings (e.g. IOTC and ICCAT)
- Collaboration with geospatial cluster
    - The generation of maps uses algorithms that rely on environmental data. Access to these environmental data is currently cumbersome, error prone, incomplete, and manual. In collaboration with partners in the geospatial cluster, iMarine offers seamless access to a wide range of geospatial data products.
    - The biodiversity cluster will continue to collaborate with the geospatial cluster to define and precise the requirements.
    - Of particular interest is the potential to run algorithms that do not only take geographic gradients into account, but also environmental differences in e.g. salinity, slope, or temperature.
    - IOC/UNESCO referred to the EBSA process where iMarine could be a provider of geospatial explicit data in biodiversity assessment. In addition, the process to establish MPA's can benefit from these iMarine products.
- Relevance to BC1- biodiversity models for by-catch analysis (NEAFC, ICES and IRD)
    - What objectives in the EA-CoP require products based on species occurrences and distribution maps beyond the biodiversity and taxonomy scenarios? In fisheries scenarios, such as By-catch management, how could these tools assist?
    - The tools offer a means to improve e.g. the by-catch reporting, e.g. if we discover a great mismatch, or if it is reported by weight.
    - Depending on gear, a cross-check with by-catch reports could be useful.
    - By 2020 the UN wants to protect 20% of the oceans. The EBSA process could be a client to iMarine products. Also to identify MPA's.
    - FAO is involved in VME and encounters. The aim is to have a work flow of data from rough reports to curated data.

- RFMO often are not active in scientific analysis, which is e.g. outsourced to ICES. But there definitely is a need of data. Many areas are poorly sampled. The iMarine tools could be presented to RFMO scientist/managers.
- If the tools are to be used for by-catch assessment, the data access and sharing policies are important, and have to be ready.
- iMarine can have a role in QA and later serve as a data node.

2.4 STATISTICAL CLUSTER

Presenters: G. Coro (CNR) *Slides*: .pptx file and A. Ellenbroek (FAO) *Slides*: .pptx file

- (Document iMB2/2012/6)

Participants: A. Ellenbroek (FAO-Session Chair), A. Antoniadis (NKUA), W. Appeltans (UNESCO), C. Baldassarre (FAO), J. Barde (IRD), E. Blondel (FAO), L. Candela (CNR), V. Canhos (CRIA), D. Castelli (CNR), H. Caumont (Terradue), P. Cauquil (IRD), G. Coro (CNR), G. Farantatos (NKUA), S. Garavelli (TRUST-IT), F. Hernandez (VLIZ), N. Holdsworth (ICES), G. Kakaletris (NKUA), J. Keizer (FAO), V. Marioli (CNR), J. Neves (NEAFC), F. Oehler (Eurostat), P. Pagano (CNR), F. Simeoni (FAO), F. Sinibaldi (CNR), M. Taconet (FAO), R. Tsantoulli (NKUA), E. Van Ingen (FAO), I. Bergeret (French Minister for Fisheries);

Progress summary by Statistics cluster chair:

Progress in the statistical cluster was illustrated using the iMarine Board Work Plan. Progress was noted on the interaction with the wider EA-CoP, the requirements for data formats (FishFrame and FLUX), and in the development of technology components in support of SDMX and the data analysis with the Statistical Service, the custom built data analysis tool of iMarine.

Discussion*:*

- The Statistical Manager was first presented by GianPaolo Coro, who precised that it is connected to CSV, SDMX and the storage manager, and will be soon connected to Open Modeller. In the future it should be able to connect to databases. In summary, it is pluggable to datasets and algorithms, and 26 algorithms of them are already supported. Future work will match algorithms to biological models needs. A benefit of the Statistical manager is that one can reproduce the experiments.
- CNR mentioned progress regarding the implementation of the SDMX registry building on OpenSDMX. This integration will enable iMarine to play a role in harmonizing SDMX data corresponding to different DSDs. FAO indicated that more information and exchange should prevail on these developments considering that OpenSDMX is a FAO initiative. IOC/UNESCO indicated the interest in Time Series of the biodiversity partners.
- FAO indicated that the focus will now be on Cotrix, the Code List manager, which many partners indicated as a priority for iMarine. The Cotrix project charter will be disseminated and ICES, DG-MARE and MEDDE stated their interest in providing feedback.

During the reporting period, several scientists have expressed interest to experiment with the tool, or share experiences in their generation and analysis of species distribution models. These include:

- Experimenters; R-modelers at Universities and RFMO's; two were selected to conduct an experiment. iMarine will engage in an effort to use ICIS and R to load a dataset and perform an analysis.
- Advisors; R and geospatial data modelers willing to review the approach, services and algorithms hosted in the iMarine infrastructure to develop EA-CoP tools, without directly becoming involved in the tools.

- Testers; several persons have expressed an interest to experiment with R. However, the current implementation will have to be improved before it will be considered ready for full-testing.

R is an important tool to consider. Clearly, there are a number of arguments to support this tool.

- R is already integrated in the infrastructure, e.g. it is possible to process through an R IDE the data that are in the infrastructure;
- There are large numbers of R packages that should / might be added to the R IDE. A particularly strong point of R is that individual scientist can load packages with administrators intervention. This opens a host of possibilities where ICIS users can operate on curated data. ICIS thus becomes a data acquisition, curation and synthesis tool, with the embedded R ready to offer analytical functions on top;
- The iMarine statistical service is a natural companion to R.

The open nature of the infrastructure allows a "client" to interact with the iMarine facilities

- The entire infrastructure is based on the SOAP protocol; these can be exposed to external services to offer e.g. Biodiversity data; the taxonomy and occurrence point services in theory can also be accessed from outside the infrastructure, from an R client residing on a desktop. An organization such as IOC/UNESCO or VLIZ could use a specific data provisioning or processing service.
- There are also REST interfaces, e.g. to access statistical data repositories.

2.5 INTRODUCTION – PRESENTATION OF MEETING AGENDA AND ADOPTION

Participants: A. Ellenbroek (FAO), W. Appeltans (UNESCO), J. Barde (IRD), L. Candela (CNR), V. Canhos (CRIA), D. Castelli (CNR), H. Caumont (Terradue), S. Garavelli (TRUST-IT), F. Hernandez (VLIZ), N. Holdsworth (ICES), G. Kakaletris (NKUA), J. Keizer (FAO), J. Neves (NEAFC), F. Oehler (Eurostat), P. Pagano (CNR), M. Taconet (FAO), Ingrid Bergeret (French Minister for Fisheries); F. Callewaert (DG-MARE). N. Bailly (FIN)

Welcome and Agenda approval – M. Taconet

The chair welcomed the participants to the meeting, with particular attention to the newcomers, and chair introduced the agenda which was approved by the Board with few scheduling modifications. It was agreed that the report will be drafted after the meeting and approved by correspondence.

2.6 ACTIVITY REPORT OF THE IMARINE BOARD

Review of iMarine Board Membership – M. Taconet

The chair introduced the new participants while indicating a number of changes in the membership and in the participants present to the Board:

- Edward Van den Berg doesn't work anymore for IOC/OBIS and has resigned from the iMarine Board;
- Ward Appeltans replaces Edward Vanden Berg as OBIS representative;
- Tjess Hernandez (for VLIZ) attends for the first time the iMarine Board, following Ward Appeltans appointment at IOC/OBIS;
- Friderike Oehler (ESTAT) replaces Matt Elliot;
- Ingrid Bergeret (MEDDE) replaces Philippe Schoenemann;
- Joao Neves (NEAFC) attending the Board for the first time (after Stefan Asmundsen);
- K.Morteo expressed a possibility to resign because of other duties

Board's approval of SB's proposed iMarine deputy-chair – M. Taconet (see Document iMB2/2012/1)

The chair regretted the resignation of Edward Vanden Berg from the Board and from his position of Deputy chair, and congratulated him on behalf of the Board for his very active and leading action as Deputy chair. With reference to Document iMB2/2012/1 which recalls procedures and action required from the SB, and from the Board, the Chair asked the newly proposed Deputy-chair Nicolas Bailly to introduce himself.

Nicolas Bailly presented his CV, and activity plan. The Board approved his candidacy, and he was elected Deputy-Chair to the iMarine Board. The election became effective immediately.

Brief review of iMarine overall work-plan and progress achieved during the intersession (Secretariat and Board members - see overall work-plan and June plan)

The chair referred to the June Board work-plan, mentioning that a number of events occurred in the meantime, and invited the participants to describe progress and comment on past or envisaged activities.

Discussion:

Several Board Members presented iMarine to initiatives relevant to their institute (ICES to the FishFrame Regional Data Base, IOC and CNR to the OBIS Committee) , and reported interest and a willingness to engage in discussions with project representatives on management of data and data formats. The project is invited to forward plans on e.g. data formats for fisheries data, management of fisheries data, and potential to use the infrastructure for data exchange and analysis.

FAO and CNR have presented in August the iMarine opportunity to VLIZ, which in turn has introduced its project portfolio (in particular WORMS and EMODNET). VLIZ expressed interest but also need to better understand data sharing/consumption policies before reporting in December to the WORMS steering committee and proposing a MoU. Major concern is visibility of the contributing partner.

The IOC/OBIS representative indicated the opportunity to introduce iMarine and its possible role in support of the identification of EBSA's, and the forthcoming October CBD's CoP.

Also, partners are willing to host infrastructure nodes, but the project should evidence specific benefits to the host of allocating resources for this purpose.

The Board has to produce clear guidelines of how the infrastructure manages citation and acknowledgement. This policy has to be very transparent in order for data providers and consumers to fully understand the benefits of data through the iMarine e-infrastructure.

FAO had similar needs for clarification that emerged from the SmartFish workshop on Indian Ocean fisheries. Here FAO is aiming to use iMarine and FLOD to harmonize across systems, but the policies attached need to be very clear, yet easy to implement for data contributors.

Board liaison and mobilization with scientist:

The debate then focused on how to involve super users (e.g. data managers of big infrastructures) and scientists and receive their guidance in Business case scenarios.

- In BC1, IRD prepares scientific reports for IOTC and ICCAT, and has a pivotal position understanding the needs of both scientist and policy makers. It is thus in a good position to collect feedback on science and policies. The board should seize on this opportunity, and if interested, propose collaborations in the field of tuna data.
- In BC2, CRIA has good contact with the niche modeling community, and can invite representatives to make an assessment of the tools. However, there are no established contacts with marine biologists. One of the CRIA partners offered to collaborate with some of their PhD students that

could assist in a review of the services provided through the e-infrastructure. Such a review would also provide an entry point to other scientists.

- Other ecosystem modeler scientists at various universities were mentioned, and several of these face computational limits in their work. Here too, the project could offer a solution to computational or data needs.
- The Chair proposed to contact a few key players, and proposed to reproduce the approach taken with ICIS, where several 'ambassadors' were active in the requirements and validation phase.
- The Technical Director outlined that iMarine with the current technologies can already serve a diverse set of potential users, categorized as follows:
  - data managers (OBIS, FishBase, WoRMS, EoL), to simplify their services;
  - scientists involved in BCs related scenarios,
  - the educational community;
  - General users in need of biodiversity data, statistics, or environmental data.
- He invited the Board members to study the technology, and liase with their EA-CoP to identify the services they need, and how iMarine can support. Some examples that the Board can discuss with EA-CoP representatives are:
  - SDMX production work-flows, from field data to a hosted registry;
  - DwC taxonomy and occurrence comparison;
  - Trend analysis or model analysis,
  - Subscriptions to Monthly Environmental Data Map service with user specified resolution.
- For some of these categories of users (in particular the scientists), the opportunity of a training workshop was discussed.
- IOC has experience in rolling out user-facing services, and advised to focus first on serving general users in need of data (e.g. OBIS related data). This evidences the need for, and the advantages of exposing data through an e-infrastructure. Hosting a data node is a big step, and this must provide clear benefits.
- The Board asked for a few showcases of concrete results. IRD proposed to use FishFrame to provide a convincing argument to show the value of the infrastructure. When identifying users for this e-infrastructure, the data managers and scientists are often the same persons, and the showcase could target both at once.
- The iMarine AppliFish Android application was welcomed as a good example of an integrated use of e-infrastructure data. Indeed, the interest was such that several partners wanted to start distributing the App immediately, either as an App, or as a component to expose through their respective websites. The project technical partners informed that there was still considerable effort required to finalize the App, and that the Board will receive further notice when the App has reached a mature stage (expectation is Nov 2012).

Conclusion: the Chair summarized the discussion: the project will need to provide a few convincing use cases. After these have been identified, the Board needs to decide on a training approach where we at the same time train and explore collaboration opportunities. The scientific exploitation can only start after a training, and therefore we have to be selective: the trainees need to seamlessly fit with the exploitation community.

2.7 BUSINESS CASE 1; EU FISHERIES POLICIES - FISHERIES OPERATIONS DATA WORKFLOW FOR SCIENTIFIC USE (CFRAME)

M.Taconet, iMarine Board Chair introduces the topic

iMarine has many characteristics that places it in a position to support data frameworks for fisheries management. For instance, it offers management tools for data collection, harmonization and exchange using statistical data formats, semantic technologies and integrated code list management.

iMarine has started high-level discussions with primary concerned players in the governance and data management disciplines to identify the opportunities for collaborations and shared efforts. The Chair presented a set of slides containing a simplified overview of EU's eRS[5] and DCF[6] data workflow, focusing on future opportunities.

One of the outcomes is the CFrame discussion paper on data work-flows in support of the scientific communities and this CFrame initiative is open for iMarine partners' contributions.

The slides presented were then discussed:

- ICES and IRD: both eRS and DCF concern commercial fisheries, not scientific survey data. FishFrame has nothing to do with research survey. For science data and research surveys, there are separate flows. The Regional Databases (RDB) should be better positioned on this diagram. ICES need to be split apart from the RDB.
- ESTAT: The situation in Europe is quite complex, and several data flows were missing from presentation. E.g. the overview missed data reported directly by Member States to FAO and DGMare, or the data flow from ESTAT to ICES. The chair clarified that the single flow from MS through DG-MARE, through ESTAT, and finally FAO was the expected future situation.
- DG-MARE: In Europe, the direction is towards a central node managed by DG-MARE to realistically manage data streams with one line of communication. In addition, there will be a strong incentive to adopt one format: FLUX. FLUX will describe individual data elements (e.g. logbook, capture, vessels, landing, trade, enforcement, IUU, traceability, ...) in such as way that it will clearly split business layers from content, conformant to the UN-CEFACT standard..In short, it will cover everything from "net to fish", except large statistical datasets. For VMS data transfer between parties, there still is a discussion with MS regarding using a central node. The DCF format could also become part of the FLUX format. All code lists for eRS will be on Master data registries. iMarine can find a place in servicing parts of these data streams.

Brief introduction of CFrame - A.Ellenbroek (see Document iMB2/2012/2)

CFrame is an initiative in the iMarine project, currently at iMarine Board level, to offer support from the infrastructure to the management of fisheries operational data. The initiative focuses on an inventory of the capabilities to manage existing data formats such as FishFrame, SDMX and FLUX. CFrame aims to facilitate the development of technologies in support of formats yet established.

Discussion: the chair invited the group to debate on FLUX and the proposed technology experimentation.

- The EA-CoP community has several related initiatives under way (DG-MARE eRS and FLUX, ICES and IRD FishFrame Regional Database), which the Board was recommended to engage.
- DG-MARE described that requirements for the FLUX development have been initiated regarding e-logbook. A first version of eRS is expected to be released soon, with an MDM available by the end of the year. iMarine could support the description of business rules for FLUX, the provision of code list services, the transformation of FishFrame to FLUX format
- FAO said it may contribute to the development of the FLUX model for e.g. the vessel part.
- IRD can contribute to the development of RDB - FishFrame capabilities. The group agreed that it is complicated for some member states to comply with advanced XML. eLogbooks and FishFrame are complicated to manage. iMarine could assist to generate FishFrame formatted datasets.
- Indeed, the possible relations with FLUX must be ensured, including during a transition period, and this requires that the underlying data-model is flexible. An experiment based on FishFrame data

[5] Electronic Reporting System
[6] Data Collection Framework

model seen as beta version of FLUX would be workable. Also ensure that you can access other databases, for example through adapters.

- The recurrent statement on reference data enforced the idea that the project needs a code list manager/mapper. FAO has a task to develop in the iMarine project such a functionality. Through regional databases, it is hoped that iMarine can play a role with e.g. support to FLUX or FishFrame.
- Regarding the opportunity of a technological experiment, the example was given in the biodiversity domain with the GBIF / IPT approach, where a schema was defined to meet the reporting needs. The Technical Director confirmed that same approach could be applied for FishFrame and iMarine could be asked to develop a small demo on 2 or 3 sources with FishFrame adaptors.
- MEDDE (involved in the VALID and eRS projects) and NEAFC (which Commission is expected to endorse a policy for data sharing) support the FLUX initiative as well as the idea of an experiment and said they could provide data sources , requirements, and testing.
- But Flux is much more, beyond the technology. It requires policies on security and confidentiality, ownership (who owns the data – states or fishermen?), data validation rules, and reference to use. Another complexity is also in exchange procedures.
- One example of such a complexity is in that data are managed by e.g. a secretariat, but the questions related to confidentiality and security is with the member states. Ownership of data is very sensitive to members, and it may be difficult to even provide test data.

Conclusion: the Chair summarized that the value of iMarine was in the management of master reference data, data sharing and transformation, and support tools to data validation, Three action points were endorsed by the Board in the following priority order:

- 1. Code List Manager/mapper,
- 2. Regional database strategic contacts through DGM and ICES contributions,
- 3. An experiment with IRD / iMarine on FishFrame from a VMS and a tuna source.

The Board will have to further consider if there is a dependency or priority required in these.

2.8 GOVERNANCE AND SUSTAINABILITY ISSUES

General overview and report from General Assembly (Donatella Castelli or Philippe Rohou)

- Background information at Governing mechanism and D4Science business model proposal

Discussion:

- Board members commented about data sharing policy developments. GEOSS has produced a white paper including a statement on GEOSS data call, putting together a set of principles for data sharing. Should iMarine link to this kind of forum, and should Terradue (represented in this Board) present a statement on iMarine GIS data compliant with this call in the Upcoming GEOSS meeting? Likewise, FIN said that there is a white paper on biodiversity informatics and we need a list of datasources mentioning if they comply with e.g. the OpenData like GBIF.
- The Scientific Director answered that we can liaise with them on concrete aspects, but we need to keep in mind the difference between an infrastructure and a dataset.

Priority actions expected from Board members (Marc Taconet)

The chair then presented slides presented at GA on the sustainability plan, seeking Board's opinion on priority actions.

Conclusion: considering that we cannot have all expertise in particular on legal matters, the Board agreed on the importance of setting a time-frame and actions, and to consider options of hiring a consultant. Two topics were deemed of priority to the Board: the CoP identification, and the policies.

2.9 SUSTAINABILITY – PROSPECTS OF USER COMMUNITY AT PROJECT END

Types of resources in iMarine - Linking the EA-CoP to iMarine services, by Fabio Simeoni (slides)

Discussion:

- The presenter explained the iMarine vision to provide a Platform As A Service (PAAS). This is an approach similar to Google apps, but iMarine offers an extra integration not found in other systems, as we have full control over the entire set of hardware and software resource. This control over all interactions is iMarine strongest point. The PAAS however has to be further developed.
- A key consideration are the boundaries. It is important to be clear as to whether iMarine is an open or closed infrastructure connected to the web? Can iMarine consider to offer an entry point for a collaboration with a CoP.
- At present, the main interactions are on VRE user interfaces, rather than services. iMarine offers an approach similar to google apps, plus an extra integration that is not found in other systems. Direct interaction with users through the iMarine Gateway is currently iMarine's strongest point. The PAAS has to be further developed.
- Regarding services, our top-down approach may seem closed, but in reality iMarine is open and our efforts are to lower the barriers for technical collaboration.
- A particular selling point is the capacity to integrate across systems' administrative boundaries. But a policy is required for sharing data with proprietary systems.

Board's guidance: CoP targets for a sustainable infrastructure (Marc Taconet)

Based on a recommendation made by the Advisory Council, the WP3 has developed a Wiki page aimed at identifying prospect CoP and products within scope of project's BCs. The Board is requested to review and fill this Wiki page for presentation at the next Advisory Council.

Discussion:

- The rational underlying this proposal was questioned. How is this document positioned? How will it be used? As a guide? As a journal? Can we foresee a progress and action part?

Conclusion: the Board agreed that this EA-CoP description page would be instrumental for various goals:

- Publicizing and outreach: developing a products' catalog based on a matrix of Services by Communities. The main aspects should be featured on the iMarine website
- Planning and organizing the Board activities and the sustainability plan, based on a concrete knowledge of the users, their types, and numbers
- Monitoring as a journal of progress, which requires that the page should be regularly updated, and include Prospecting activities, Pipeline activities, Concrete current activities, dead activities.

The chair concluded that WP3 will rearrange the Wiki page according to the provided guidance, and that Board members will be requested to fill it in the next coming weeks.

2.10 IMARINE DATA ACCESS AND SHARING POLICY

Needs for MoUs regarding data usage (Donatella Castelli)

The scientific Director introduced the topic with reference to the experience gained through interactions with data providers, and the discussion was engaged.

Discussion:

- The iMarine e-infrastructure is faced with services from the provider infrastructure which lack the required capacity for large consumers. One example is the CoL: iMarine complies with the policy conditions as given by the web-site. Their WS was however not powerful enough for the large use by the infrastructure. CoL required a special agreement on usage and users, which necessitated a MoU which would for example stipulate that a copy of the CoL database would be required for performance considerations, and under which conditions such copy is synchronized with the source. The MoU may evolve if new users join the infra and consume CoL data. The MoU specifies the use of the CoL data for a time frame.
- With WoRMS a similar situation exists, iMarine limiting its access to WORMS data through on-line requests. In addition, the WoRMS management has raised issues on the re-use of their data, even though iMarine feels complying with their policies. Redistributing a copy of the database is not allowed, and credit should be given. iMarine complies with both aspects.
- Other, similar situations were given by meeting participants; Some aggregators become partners in EU projects, instead of the original data contributors. Other data providers, when interacting with large consumers, faced a costly license issue.
- CRIA stressed to define a sustainability model across the value chain. Currently, the data is open, and it will not be simple to find paying customers for the data use, if that even is possible under the data-license. Donors pay for structure and secretariat, not for the scientists and the data. Data become ever more dynamic, and it is important for data providers to know where and how their data are used, also for them to prove their right of existence. If we can prove that, it will benefit all.
- The project Director added that indeed, there is a discussion in the commission that publicly funded data should be open. We need to understand the direction and the scope. The commission allocates costs to sustain infrastructure and pay for access, and our model show cater for that.
- FIN has experienced issues with multiple donors, with multiple expectations on data re-use etc.. In addition, the partner organizations need also to consider their own sustainability beyond iMarine, making the case for a data policy even stronger.
- The Technical Director referred to the de-facto policy of citing the credit, (WoRMS, CoL, GBIF) and including credits to authors., iMarine does not store entire datasets (except for GBIF with which a specific agreement was made for storing entire dataset during one week), and most of these are read-only. The caching of a part of a dataset is a technical artifact which considerably speeds up search and retrieval, and if that is done with proper care, then conflicts with data owners can be avoided. It is also very important to note that iMarine does not plan to be the publisher, but to provide the means to data owners to improve their products by using iMarine services.
- At VRE level, usage quota which would restrict data access cannot be imposed. Providers such as WORMS stated that they would privilege their users to access through WORMS products elaborated in iMarine. The Technical Director agreed on this view that OBIS and WoRMS should become iMarine users, exploiting services provided by iMarine, not the other way around (AquaMaps is a good example). The policy should make this very clear so to encourage data providers to join the initiative.
- The responsibility of the data providers is to clearly state their license and terms of use, and this should be propagated in iMarine. The MoU then guarantees the behavior accordingly. iMarine should consider proceeding with a review of all licenses and understand how to behave with each type of license. The support of a lawyer would be required there.

- The Policy might differ depending if the data are static (historical data sets) or dynamic (data is changing everyday). The policy should cater for a guarantee to i) not modify the data in case of redistribution, ii) always refer to the master source, iii) ensure specified type of use which iMarine will make of the data, documented through usage statistics; The policy should promote the principle of freshness of the data made available and ensure proper documentation to enable user to judge it.

Conclusion: the Chair summarized the discussion: the iMarine data access and sharing policy document must make explicit how we refer to the proper attribution of data. In the absence of policy document, MoUs are necessary at this stage and contribute to the development of a comprehensive policy. In the future, endorsement of the iMarine Policy by the providers will in most case substitute the need for ad-hoc MoUs, This policy needs to also take composite products into consideration. The policy may need a review by a lawyer specializing in the subject matter.

Policy document draft 2 (Document iMB2/2012/3)

Aureliano Gentile introduced the draft version 2 of the Policy document elaborated during the intersession, building on the feedback obtained on draft version 1 during iMarine Board 1. The chair invited the Board to organize the discussion on the following key principles, towards their adoption.

Discussion:

Metadata:

- The terminology “submitted to” was deemed inappropriate, and should be substituted by “made available through”. However that would not apply to the different Use Case of a user directly uploading data.
- The elements of the citation need to be well understood. Owner (e.g. FAO) is often a different entity from “author” (an individual person). “Publisher” is the distributor. “How to Cite” should be defined at data provider level: there is the concept of “preferred citation” in Terms of use and providers should be encouraged to make use of it.

General principles:

- It was clarified that iMarine can make data public through the existing standards such as SDMX, or OGC e.g. produced by GeoServer services.
- The current draft suggests producing metadata at dataset generation time (a dataset is a larger corpora of data, and stresses the importance to maintain the Metadata record through the data-flow. The Metadata applies at dataset level, and this is inherited at record level. However with 1.000.000 TS records, it would be unrealistic to think managing Metadata at record level.
- AgInfra distinguishes Collections from Datasets (OBIS, GBIF). AgInfra has two deliverables, and it would be fruitful to exchange ideas.
- FAO goes towards CC-zero, i.e. there is no limitation in the use of data. There are still problems where FAO is custodian of data owned by others.
- A policy should be defined at collection level, while a dataset is managed according to the format that conveys it. This discussion is about the specific iMarine policy, and iMarine is the custodian of information provided by others, including those having confidential data. Providers should consider their license and policies.
- With metadata chaining and flows, GEOSS experiences the frequent problem of absence of the required Metadata at provider level. A question is therefore whether the platform can get a complete metadata. In iMarine the effort is for promoting CC, and since rdf elements are defined in CC framework (a specific CC even exists for scientific works), semantic technologies could be used to improve the provision of Metadata to manage waivers and licenses. iMarine should invite providers to map their Terms of use with CC, and could consider facilitating providers task in implementing CC in their websites. Another difficulty is that none of the current providers of

iMarine has yet adopted CC. IRD for example, as provider, is in the process of assessing its own licensing with the support of lawyers, and mentioned.

- iMarine should promote openness, and state that those no complying are not really part of this initiative. However in the EA-CoP there is a strong need for industry data and an emerging movement to promote the participation of the industry in contributing its knowledge; this will not be possible with a strict open data policy.
- In its work towards promoting adoption of CC, should iMarine involve lawyers. On one side, lawyers would advise how source data can be mapped to CC, on the other hand involving lawyers might become unworkable. The pros and cons of keep it as simple as possible should be assessed.
- The lack of clear statement on redistribution in the current draft is seen as a gap.

Structure of the iMarine policy document:

- With reference to the usual corpus Policy/Procedures/best practices, the wordings “best practices” in the title was felt difficult to understand.
- The higher Policy level would be structured with generic principles and terms and use and should enforce the general agreement. And at VRE level, the infrastructure technology would allow to have more specific terms of use. For example, the VMS VRE would require a more restrictive terms of use than related to other VREs. If a user registers, s/he will accept generic terms of use. When accessing a specific VRE, a more restricted use policy can be accepted. Although not yet implemented, the infrastructure could if required restrict data download and sharing for a given VRE.
- Proposes to list a set of iMarine standards, a selection of Community Standards.

Roadmap towards adoption of the Data access and sharing policy

- Need for MoUs: The question was raised as to how to proceed in the short term with the VREs and data sharing without a completed policy document. It was agreed that MoUs should be established with VLIZ-WORMS and CoL (with Peter Shark), in ways that echo the current version of the developing policy. Reference could be made to MoUs established between CoL and GBIF or EOL. VLIZ will provide an MoU example.
- iMarine policy should give stronger evidence of mandate and sharing goals with clearer statements, and evidence of alignment with EU goals. In this respect, the GEOSS policy is simple and clear (mandate includes science, benefit to science etc) and could be used as a model. If iMarine aligns with the GEOSS data core and is perceived as an instrument able to feed GEOSS with UN sourced data, it would gain stronger support from EU. Biodiversity is also a clear case of complication regarding use and re-use of biodiversity data; in this biodiversity field, GEOBON has objectives similar to GEOSS and it would be beneficial to align iMarine policy ideas with GEOBON ones.
- It was agreed that the iMarine policy should help clarifying its role and should consider aligning with such other policies. It was also agreed that in more advanced stages, a lawyer should be consulted in order to review this policy document.

Conclusion: the roadmap was agreed to be:

- Feedback on this draft version 2 by end of October
- Development of MoUs for WORMS and CoL
- Development of version 3 according to the document structure approved by this Board, taking into account version 2 feedback, need to align with the foreseen MoUs and other use cases of shared data Collections.
- Participation to the RDA initiative
- Version 3 will be presented to Board 3 as a pre-final version: one more year will be necessary to fine tune it with actual use cases and feedback from other initiatives (e.g. GEOSS, GEOBON), and to make it legally stamped.
- Guidelines providing practical extensions of the policy should be initiated now and will be focus of development in 2013
- A Final version should be signed-off by the Board5 by Q1 2014 for endorsement by Organizations members of iMarine.

2.11 IMARINE STANDARDS AND GUIDELINES

Presenter: Anton Ellenbroek

The iMarine policy will list among existing Metadata format those considered as iMarine standards. The WP3 leader introduced those and the debate was opened.

Discussion:

- Biodiversity: Both Darwin Core, and Darwin Core Archive are concerned by iMarine standards. iMarine pinpoints on needs to evolve DwC archive, and this will be the role of Nicolas Bailly participating in TDWG.
- Geospatial: Reference was also made to EML/IPT, but this duplicates OGC and in line with the INSPIRE directive, OGC should be the preferred geospatial standard. However iMarine could provide the service of transforming the EML/IPT into OGC and thus contribute to INSPIRE (see annex 3 of implementation rules)

Statistical:

- the SDMX proposed Metadata standard was accepted, but a question would be whether there are requirements for handling and/or publishing Annotations at record level.
- FLUX: fat present, there is no official policy documentation on FLUX, but instead two documents on FishSoft exploiting the current eRS format. The idea is to propose a central node for EU MS with a first version to test in October. A Fleet register management will be a first step.

Semantics:

- Some RDF models should be given priority: FOAF for describing projects and people; Genesi-DEC should be considered a candidate standard; the TLO (Top Level Ontology) is about a way to annotate information resources. FIMES was considered a candidate standard for species fact sheets.
- The Board deemed necessary to list the criteria to be considered for declaring a Metadata format as “iMarine candidate standard”. A preliminary list of criteria was proposed:
  - Internationally used
  - Pushed by the community with a valid use case
  - Politically compliant with EU orientations
  - Assessed as feasible, with the technical team

Conclusion: the chair summarized saying that there is already consensus on some of the listed Metadata format as iMarine standard: DwC, OGC, SDMX. A distinction will have to be made between a full iMarine standard and a candidate iMarine standard. A candidate standard might find an entry point (e.g. capacity to read and transform), while a full standard would be fully supported along the whole workflow and essential functionality deemed necessary by the Community.

2.12 SUMMARY OF BOARD RECOMMENDATIONS FOR THE NEXT 6 MONTHS

The Board reviewed and summarized the main recommendations for each WP3 main task, highlighting sequence of activities and priorities, and assigning responsibilities. The Board agreed to formalize this as the Board October 2012 work plan in Excel format after the meeting.

Governance

- The Board will describe the target community of users and their products as a journal which will enable to monitor progress, distinguishing concrete users, pipeline users, and new prospects.
- FAO, CNR and FIN will develop further proposals for the sustainability model under SB guidance

Policies

- A draft version 3 of the iMarine data access and sharing policy will be produced for Board3 according to the structure agreed upon by Board2 and following guidance received during this meeting. The Board Members will review the iMarine policy document version 2 by end of October

Sub-contracts and MOUs

- CNR (D. Castelli) will take the lead for the MoUs with WORMS and CoL
- FAO (M. Taconet, A. Ellenbroek) will develop the sub-contract with NEAFC after their commission meeting in November with the goal to sign it in January 2013

Towards a demonstrator for Fisheries operational data (CFrame)

- Within a first phase of 3 to 6 months, IRD will set-up on their server a Regional Database with a Postgress data model generated from FishFrame schema (this is funded by DCF for a web-portal by end 2013). In the following period, the RDB will be connected with iMarine infrastructure and the goal will be to plug on top of this data base transformation capacities from CSV files into SDMX, FLUX, DwC, OGC, and Code list mapping capacities. It was agreed that in phase 1, there is no need for strong technology interactions.
- FAO will make sure that DG-MARE's code lists be made available in the infrastructure.
- Regarding eRS, use cases on validation of data and curation will be taken from the VALID project, and IRD, ICES and MEDDE will consolidate the most common ones, and propose for implementation.

Geospatial

- The SPREAD use case made good progress with recent TCom and technical activities will be pursued
- Progress on data discovery will await inputs from OBIS and VME-DB
- Occurrence enrichment can proceed with an IRD use case, while IRD will have to gain more knowledge on WPS; additional use cases should be discussed with biodiversity members. The community should help identify hydrographic parameters of interest to the business cases, and Terradue will be able to help select suitable sources. The selection of processing methods will have to come from the Use cases. Progress should proceed in incremental ways.
- Time Series visualization will not proceed before the requirements work initiated by FAO will be validated by the Community

Biodiversity

- The taxa-matching and occurrence reconciliation objectives will receive priority and a consultant (Edward Vanden Berghe was proposed and agreed) will support the community in developing requirements and testing cases.
- The Board agreed that the OpenModeller library, contributed by CRIA and integrated on the D4Science platform under OpenBio/i4Life, will be made available to iMarine;
- The Deputy-chair will investigate with CRIA what their contribution to WP3 and WP6 can be; among the likely inputs were pre-identified i) the validation of OpenModeller in iMarine, ii) inputs to quality control procedures related to taxon or occurrence reconciliation.

Statistics

- FAO will proceed with Cotrix (Code list manager), later followed by the Code list mapper;
- The Statistical environment will be enhanced with the SDMX workflow with a delivery foreseen by end 2012/early 2013; other features requested by the Community in ICIS for the Tuna Atlas use case will be implemented; these format handling and workflow realizations will in turn provide the foundations for other XML data formats such as FishFrame or Flux
- Use cases for the statistical manager will be looked for among community members

Semantics

- The community will work on generating more RDF from the identified use cases (dynamic species fact sheets, SmartFish, ...)
- The Board privileges to stick to practical user needs and this will influence the approach chosen for the TLO development which should answer to practical requirements, even if not perfect
- Forth should work on entity mining algorithms and integration with search engine capacities for the infrastructure
- The Board agreed that a valuable direction for iMarine would be to acquire a capacity to generate RDF and LOD from iMarine published products, in ways defined by the policy

Finally, the iMarine Board decided:

- to set-up one mailing list per technical cluster;
- to hold its next meeting during the second or third week of March 2013.

2.13 END OF MEETING

The meeting was closed at 17h00 on Friday 5th October.