Building an IT infrastructure to support marine SDI: IPMA case study

Valeria Pacheco, Portuguese Institute of Sea and Atmosphere, Portugal

Abstract

The Portuguese Institute for the Sea and Atmosphere, IPMA, I.P.'s mission is to promote and coordinate scientific research, technological development, innovation and services on the sea and atmosphere. This Institute hosts several interdisciplinary projects: providing weather, oceanographic, sea biological and climate data, forecasts and warnings towards safeguarding people and property within the national territory. IPMA has responsibility to tackle issues of data availability to third parties, public and private entities and some of its main challenges are the handling of large quantity of produced data, format diversity and intra-departmental dispersion. A partnership agreement between IPMA , I.P. and the Portuguese Task Group for the Extension of the Continental Shelf (EMEPC), more specifically on the implementation of SNIMAR project, is presently serving as a trigger to address these issues.

The SNIMAR project *Preparation of integrated geographic information for marine and coastal water management* – predefined project of the EEA Grants, aims to develop a Marine Spatial Data Infrastructure (MSDI) to increase capacity of assessing and predicting environmental status in marine waters. SNIMAR will be developed in accordance with the Directive INSPIRE and based on open source technologies, thus ensuring interoperability and compatibility between geographic data and services.

This presentation outlines IPMA's contributions to the SNIMAR project and the procedures used to identify structure and create relevant data classification in compliance with the adopted metadata profile (SNIMar). To reach these goal methods to normalize and harmonize IPMA's marine spatial data are required, thus, the definition of its data policies. Besides wanting to adopt SNIMar profile, IPMA also intends to improve its levels of monitoring and control of the production process, reason why IPMA has the necessity of collecting and treat the information.The proposed methodology will serve as a case study that will be applied later on to the Institute's remaining core areas.

Introduction

The importance of a sustained management of earth's resources it has been recognized more and more by the majority of governments and private sector entities. Due to this, the requests for spatial information increased significantly. Spatial information is extremely valuable since it provides geographical context in management of many subject areas and it represents a critical asset in decision support. [1]

In order to manage this information it is necessary to create new infrastructures to assist this data integration into a common system. Therefore, the creation of a Spatial Data Infrastructure (SDI) becomes crucial for supporting governments and other private entities in their decision activities.

A SDI is a collection of data, technologies and institutional arrangements that has the purpose of creating an environment in which all stakeholders could search, discover and query spatial information in a collaborative manner. The advantages of developing a SDI are usually associated with a reduction of duplicate efforts in collecting and maintaining geographic data, availability of the data and interoperability between datasets.[1] Therefore, the entities owners of such data consider the development of an SDI a primary concern.

In response of the increasingly necessity for spatial data a number of initiatives have been taking place. An idea of a global infrastructure capable of enabling users to discover, search, view and understanding geospatial information started to become a reality. One of those initiatives was INSPIRE (Infrastructure for Spatial Information in Europe), under the competence of European Commission, intended to fulfil the need of setting a legal framework to support the availability of spatial information, for the formulation, implementation and evaluation of union policies. That initiative represented, a step towards building an European SDI.[2]

In Portugal there have been requests for spatial data, regarding the marine sphere, to be available free of charge and without legal restrictions, in response to those needs the National Information system of Sea (SNIMar) was created. This project has the main objective of providing geospatial data and an IT infrastructure for the Portuguese communities regarding marine data. This SDI is crucial since it will host important marine information and IPMA will have a key role in preparing all this geospatial information.

In this paper it will be explained how IPMA took the first steps towards the development of a marine SDI. It will be shown the importance of the creation of sub-area to support SDI's future activities, namely a controlled environment to create reliability in the delivery of spatial data.

1 - Global SDI - brief historical perspective for the reason underlying its creation

The first references in the recognition of the value of geospatial data, dates back to the early 90's, in which geospatial data was used to assist in crime reduction management by United States Department of Justice. It wasn't long before partnerships between local law enforcements agencies and the state started to happen and explored the potentiality of Geographic Information Systems (GIS) applications, in identification, visualization and analysis of crime trends and patterns [3]. These efforts led to development of the field, and enable authorities to spread these partnerships to regional and local levels.

In 1992, the agenda 21 resolution establishes measures with the purpose of reversing the impacts caused by environmental deterioration. The global realization that action needed to be managed more effectively and based in factual evidences was emphasized. In congruence with the fact that more global infrastructures are needed to support a global level management and a lot more commitment in solving those issues.

In this context, GSDIA Global Spatial Data Infrastructure Association (GSDIA) developed the GSDI cookbook in order to enable national and regional initiatives to create their own interoperable SDIs. The GSDI cookbook was developed with the purpose of globalizing and standardizing this idea. GSDI include conceptual disambiguation, best practices, standards of software open source solutions and supportive organisational strategies and policies. It intends to set a framework capable enough of guaranteeing that the tools resulting from this approach will be reliable and it will fulfil the primary need of the SDI, the interoperability. [3]

1.1 European SDI

In 2001 Europe started the implementation of policies and guidelines for the European countries to establish a common SDI that could help the integration between data and systems. The main focus was to ensure that spatial data infrastructures of the Member States were compatible and usable in the European Community context. In this context it was established the INSPIRE directive.

1.2 Inspire

The directive 2007/2/EC of 14 March 2007 establishes an Infrastructure for Spatial Information in the European Community [4]. This directive is an initiative that intends to create a legal framework to support the availability of spatial information within the European scope, with the purpose of supporting legal activities in the creation, implementation and assessment of union policies. The major idea behind it is to create a legal base to support future developments regarding the development of a European SDI. The Directive addresses 34 spatial data themes needed for environmental applications.

1.3 Portuguese SDI

The first SDI in Portugal was created in 1990 and was the first one to be available through the Internet in 1995 [5]. Portugal started the investment in this field due to the impacts that SDIs have in economic, social and environmental field. The most important economic benefit of SDIs is the promotion of economic growth as a result of an expanding market for geographic information products and services both locally and internationally [6].

As our first approach, IPMA decided to use 2 ,of the recently identified, products to study them and prepare them, collecting information under the form of metadata in accordance with INSPIRE directive and also collecting information about the IT environment as a way of characterizing the product's life cycle. The conducted study about these 2 products, intends to enable IPMA's IT managers to systematize the process to apply to a wide collection of data,

2 - SNIMar project - IPMA's contribution

The project SNIMar arises out of the necessity that existed in creating a centralized information node about the marine environment, with the purpose of understanding the environmental status of the marine and coastal waters in Portugal. This information is disperse in different public institutions that work in accordance with their own policies and specific needs not concerning with third parties entities needs or requests. In this sense, the necessity of creating a centralized node that could enable the inclusion of all the information about the Portuguese marine sphere and enable users to combine and reference that information, emerged. To respond to this necessity, project SNIMar started its activities with the purpose of preparing integrated geographic information for the management of marine and coastal waters with the ultimate objective of building a thematic SDI about the marine sphere. With this thematic SDI would be possible to discover, search and query geographic information and with this information it would be possible to support decision regarding the marine and coastal environment.

Project SNIMar is divided in 6 Work Packages (WP) - Table 1, in order to manage the development of the work and ensure the objectives set will be accomplished.

WP	WP Name	Description
0	Project Management	General project management
1	Data Policies	Definition of Data Sharing policies
2	Technical framework	Software and Hardware Management
3	Metadata and Geographic Information	Contents and harmonization of the data in accordance with INSPIRE directive
4	Geoportal and Services	Service's Catalogue and technology for making the data available
5	advertisement	Project promotion

Table 1 - Work Packages

2.1 IT Architecture

Within the scope of work package 2 of the project, IPMA is responsible for designing and implementing the IT architecture for the solution . IPMA's architecture will have the behaviour of a local node and will also be a central node. As a local node, IPMA will respond to the project through the characterization of its data in accordance with the regulations imposed by the project, creating its own metadata which will feed the global metadata catalogue. As central node, IPMA will create conditions to host the geoportal with the metadata catalogue and make available the accessibility services to internal and external users. The communications between central nodes and local nodes will be ensured by VPN communications. Any authenticated user will be able of searching the catalogue, in accordance with the accessibility policies defined individually for every dataset included in the geoportal catalogue.

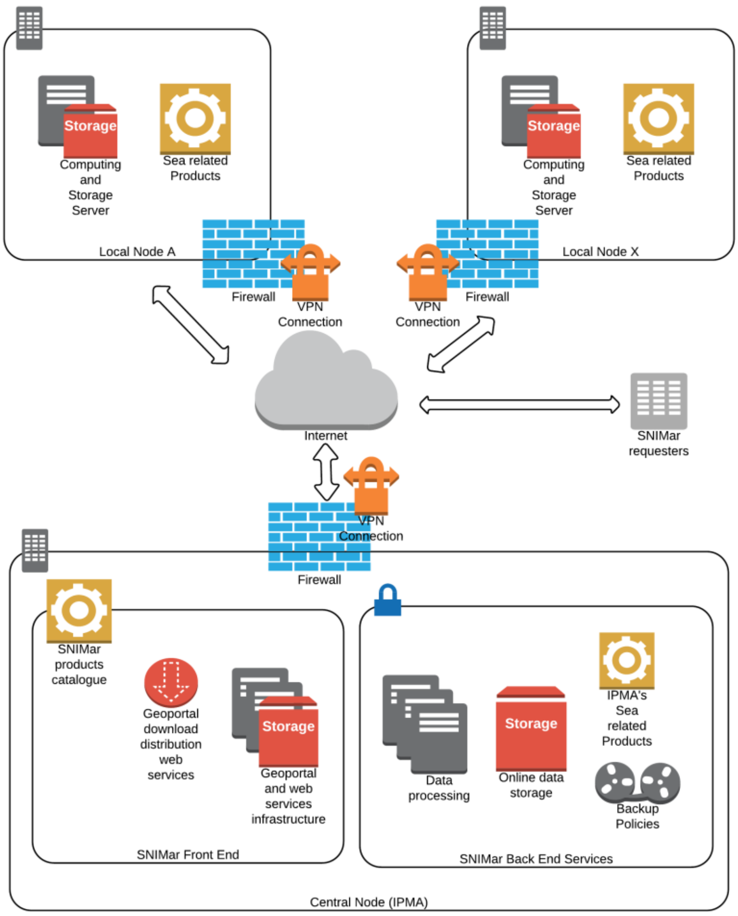

fig 1. IT SNIMar Architecture

2.2 Constraints and issues with IPMA's datasets.

IPMA as an important partner in project SNIMar and one of the most important data provider had to rearrange their internal processes in order to prepare the information for the project.

IPMA's situation regarding data was as it follows:

- There was a great amount and diversity of data disperse.
- The products and product's production process weren't defined - the necessity of defining products and processes in order to organize the data was a reality
- There wasn't a unique identification of products inside IPMA, no way of referring to the products easily without confusion or mistake.
- The documentation was inappropriate, didn't follow any standards.
- Data was in a number of formats and platforms.

While defining some metadata, according INSPIRE directive technical guidelines, it was possible to give some structure and context to some of IPMA's geographical datasets and from there define some products. Within an attempt to find a solution to the above mentioned issues, and as an intervenient in the local node, IPMA saw the opportunity to structure its data and its internal processes, as way of improving its levels of quality, availability of products and operational rates within the scope of its responsibilities.

2.3 - SNIMAR profile: to accomplish harmonisation

In order to fulfil requirements of harmonisation it was developed a profile for the metadata. SNIMar profile tried to harmonise and extent the utilization of the component of keywords as way of describing metadata (element of ISO 19139: **descriptiveKeywords**) Besides the types of keywords already existent in the codelists of ISO 19139.[8] there are contemplated others that group the keywords in a more specialized manner and oriented to the data and services considered within the project.

To harmonise the contents, it’s being compiled a thesaurus which include the participation of all entities part of project SNIMar. This thesaurus will serve as base for filling the SNIMar keywords.

As the SNIMar thesaurus only includes terms relative to the marine sphere, IPMA will extent it to gather terms relative to the remaining subject areas producers of data, such as the climate, meteorology and environment.

This utilization more harmonised of the component of keywords descriptive of ISO 19139 [6] for project SNIMar will be considered by IPMA in the production of its own metadata. Beyond fulfilling the requirement of the harmonisation of its contents, it will also serve as baseline to construct a hierarchical structure of the products, produced by IPMA.

In the context of the Geoportal, which is being developed, this collection of keywords structured in a hierarchical form will enable the implementation of more efficient metadata oriented queries, which will facilitate the search for metadata. Metadata is essential to support the discovery, evaluation, and application of geographic data [3]

3 - IPMA SDI

The implementation of a internal Marine SDI is the main purpose of IPMA, within the scope of project SNIMar. Building a Geospatial Data Infrastructure to hold the marine sphere is a complex task that requires petabytes of geospatial data processed within different software and hardware. To manage all this information it was created a model to exemplify IPMA's data lifecycle (figure 2). Sometimes with more, sometimes with less complexity of processing, this model characterize the cycle of processing of the products inside IPMA. Within the scope of project SNIMar, this characterization of the production process will enable the creation of metadata which will feed metadata catalogue. However, this processing flows are also supported by a strong IT component, reason why IPMA's IT managers also intend to take the moment of collecting the data to collect information also about the IT environment where the production process occur.

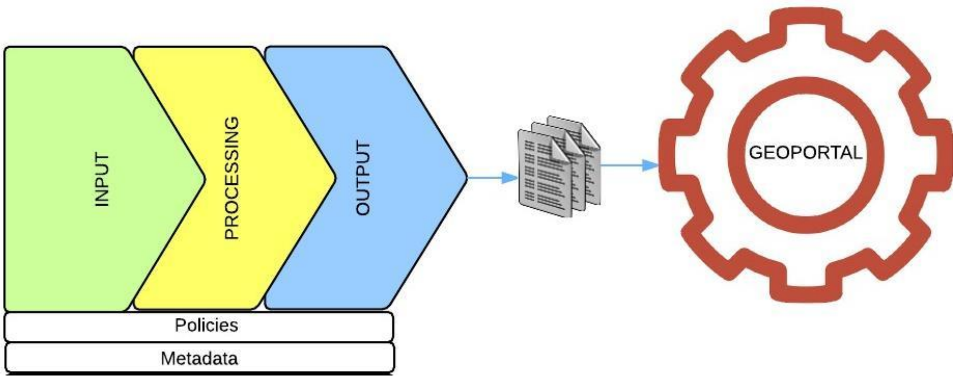

Figure 2 - IPMA Data Infrastructure

At IPMA one of the major issues identified was the great amount of data and the level of specificity with no correlation or interoperability between datasets. Concerning this issue, this methodology was implemented to create a model capable of defining commonalities between datasets regardless their field of knowledge. Thus, by implementing this model, IPMA's IT managers were able of transposing the production process to fit in this model and in this way, be able of segmenting the large quantities of data into a well organized process with all the information regarding the dataset and its production process.

4 – An IT Perspective towards the implementation of the marine SDI

The implementation of a controlled environment is crucial for tracking data production flows that will allow a creation of a new dimension of information. This new dimension will enable a complete monitoring and control of the production chain. Although this IT infrastructure has information regarding hardware and software the main scope is relied on the product and not in IT. The necessity of implementing this controlled environment is to translate a complex data production chain (figure 3) into a simplified and controlled data production chain (figure 4).

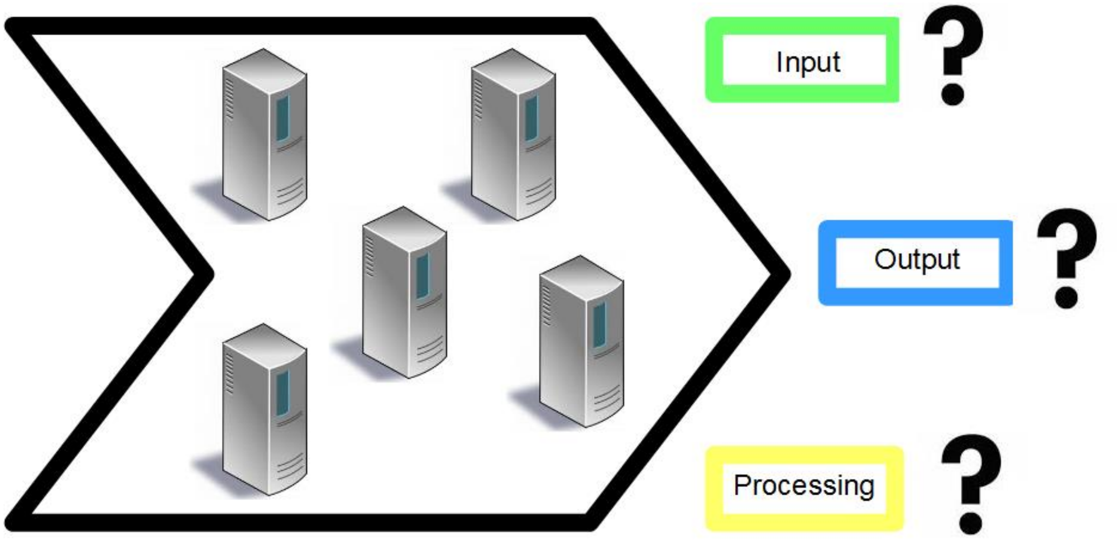

Figure 3 – Complex data production chain

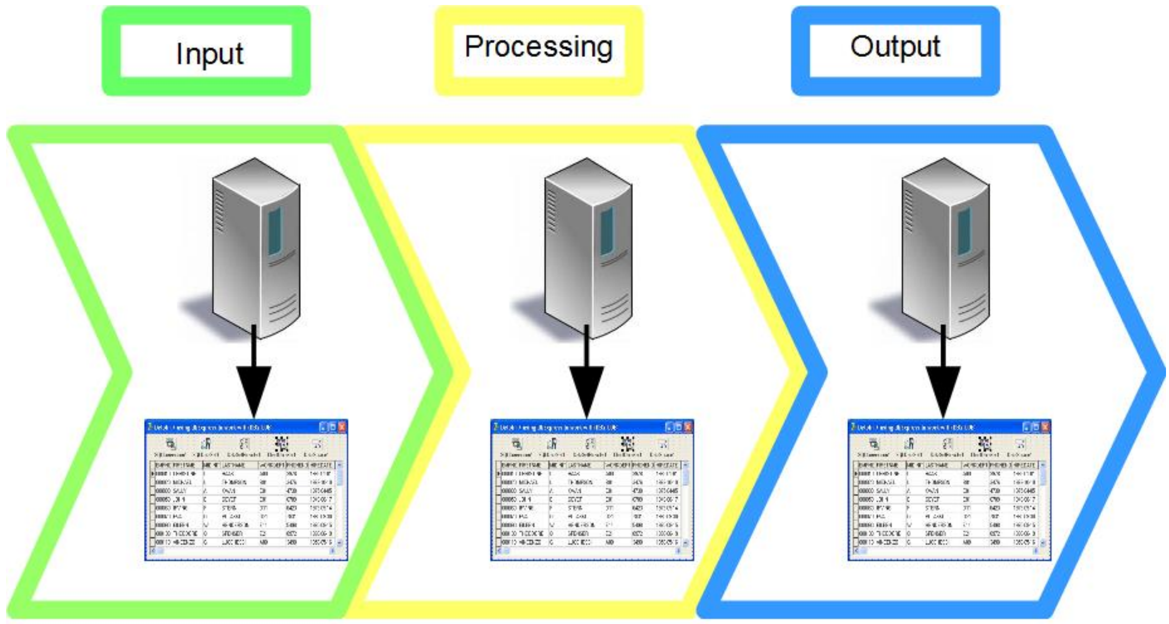

Figure 4 – Controlled data production chain

To ensure the availability of every product, in each different phase, the implementation of this methodology represents a turning point for IPMA. With this methodology is now possible to track where the processes for developing the products occur. Thus, it was decided that a new dimension, to cover the IT infrastructure of the products, was required, in order to guarantee the availability of data in every part of the process of production. This new dimension should include the description of equipment, processing elements and users access policies, it is intended with this new dimension to be able of defining all the dependencies and assets associated with the product in order to be responsive in case of failure. To the collection of information, regarding these dependencies, it will be called Information Technology Metadata (IT Metadata). IT Metadata intends to describe all the information regarding IT environment in which the product is involved and to provide reliability to the mechanisms that enable data delivery.

In this way, IPMA now has the possibility of be successful relatively to the monitoring and control of the quality of the processing cycle.Which will enable IPMA to improve their intervention in any part of the process, in case of any problem occurs , even the problems that could affect the correct behaviour of the geoportal. .

To manage an entire IT infrastructure and to respond to their daily needs, regarding the management of their activities, IPMA implemented a Configuration Management Database (CMDB) with the objective of centralize all this IT information into a single application. Following this strategy and methodology it was integrated the controlled data production chain into this internal CMDB (figure 5). The role of this CMDB is to provide a centralized information repository of core configurable IT components and relationships to the associated business service hierarchy. The main benefits of this integration are inherent to three main factors: Control; Integration; Decision Support [7].

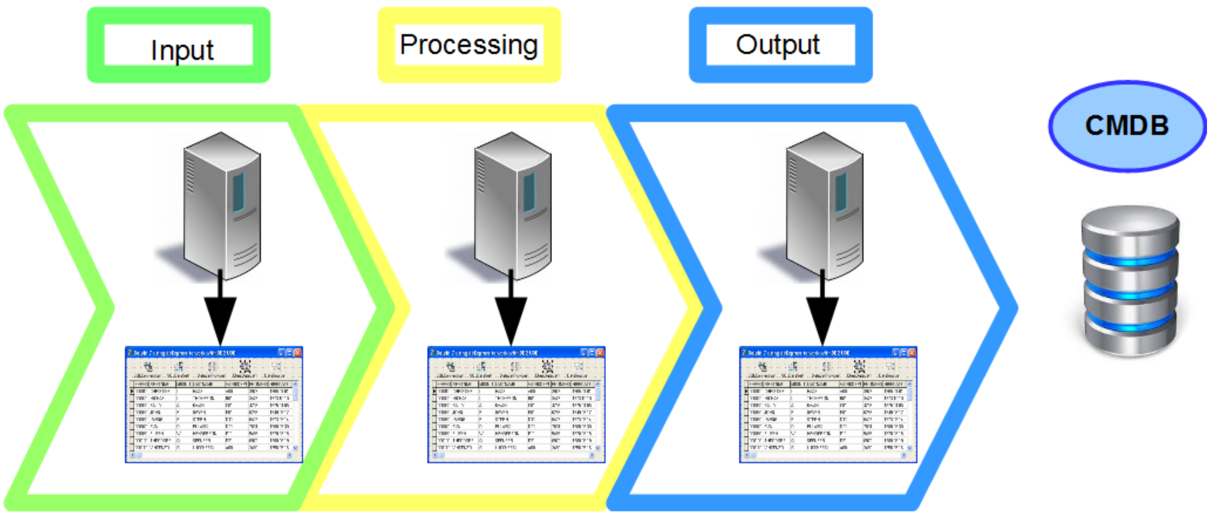

Figure 5 - Controlled Data Production Chain and CMDB

4.1 Extension of the metadata profile to the IT metadata

The figure 6 represents the information that should be associated with each product. Each layer of this model represents a layer in IT metadata, which are related with the number of dependencies that each product has.

It is IPMA intent to extend the metadata profile to the IT sphere and be able of including this information within the metadata available on the geoportal. To do that it will be necessary to develop a model to include these tags in the metadata structure. This kind of information will be enabling to users of the geoportal to query for server or procedure and the geoportal could include policies so this information only will be accessible to IT managers or IPMA internal users.

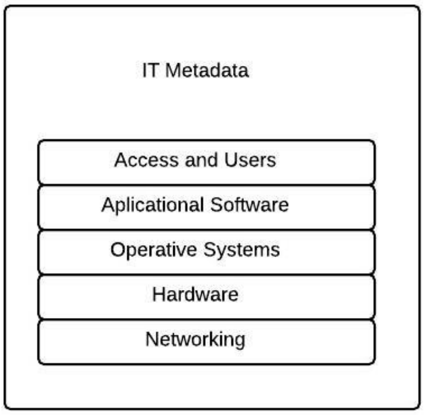

Figure 6 - Metadata IT

4.2 - IPMA Case Study

In this chapter it will be presented two case studies regarding the implementation of this methodology, the Bivalve analysis and the SWAN´s model. Both of these two case studies are practical examples of an initial adoption of this new methodology implementation at IPMA. The initial information was collected through direct interviews with scientist and IT managers responsible for data management.

4.2.1 - Bivalve capture zones

In this subsection it will be provided an example applied to the model mentioned above relative to Bivalve Analysis. Bivalve analysis is a common procedure at IPMA namely to identify and determine the zones where the capture of bivalve molluscs is permitted or prohibited. To do that, IPMA provides a map where the restricted and free-capture zones are flagged. These maps are an extremely important product, because the outcome of such information represents a matter of public health so this information should be always available and updated on IPMA's website.

In figure 7 it is represented a generic perspective of the process that it takes to produce the final map.

This conceptual separation it is necessary to address the issue of having to deal with too many processes and assets. In this way, it was possible to divide the process of production of this product and register all the dependencies inherent to the process chain that could cause conflict or troubles.

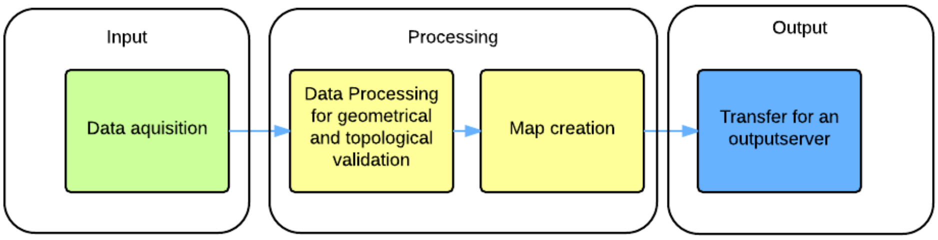

Figure 7 - Bivalve Processing Chain

4.2.2- SWAN's model

In this subsection it will be provided a case of a specific example of a product's production process using this methodology. SWAN is a third-generation wave model, developed at Delft University of Technology, which computes random, short-crested wind-generated waves in coastal regions and inland waters [10]. IPMA develops a product named *sea wave forecast according with SWAN's model (SWAN's product)*, this product is responsible to provide forecast about the sea state and has a complex processing chain.

Sea wave forecast is an operational product, with an hourly availability frequency and with a number of dependencies extremely high. Also this product is extremely relevant for civil protection and nautical tourism.

The process of implementing this model into the production chain followed the guidelines mentioned in the figure 8. In this figure 8 it is explained how it was integrated this SWAN´s product into the production chain following the input, processing and output methodology.

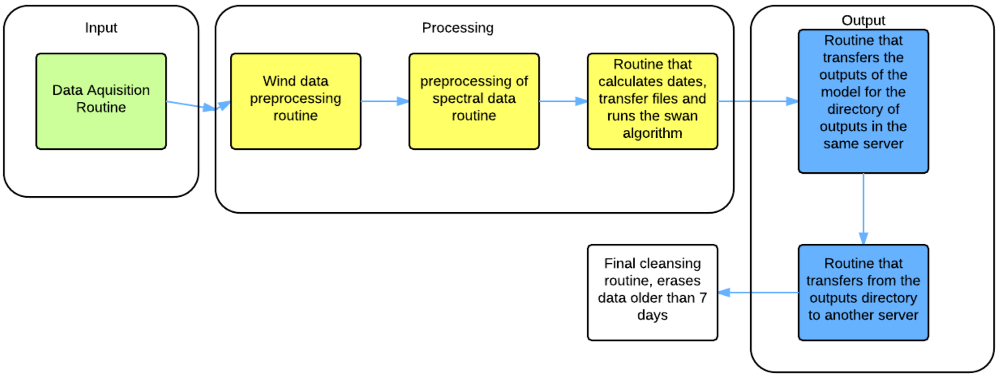

Figure 8 - SWAN´s Model Production Chain

By identifying the SWAN´s process activities it becomes clear which phase is needed to be controlled. SWAN´s product is dependent on to many routines, without them the product would make wrong forecasts. There are series of events that have to be processed through a certain hierarchy and at certain time in order to deliver the forecasts properly. After identifying the commonalities between these two processes it is necessary to create a model for the database capable of storing all the information required to track the product production process. To create this database model it was made an abstraction of the model mentioned in the figure 7.

4.3 - Systematization Process

In order to systemize the process of data collection and organization, in this chapter will be clarified the process in which the products go through to be available on project's SNIMar platform. if we want to put a product available, here's the process that has to be made:

- Identify the phases of the production process of a product

- Collect information in accordance with SNIMar profile and collect information related to the IT component of a product

- Creating the metadata SNIMar to feed the catalogue for the geoportal

- Creating the IT metadata to feed an internal platform (internal consumption)

5 - Current status, a IT management perspective

To fulfil international obligations, IPMA intends to have their processes certified in a near future (ISO 27000). In order to accomplish it, IPMA will have to implement three major logic separations regarding the development process: A production environment, a homologation environment and development environment. At this moment, there is no separation and in case of failure it's extremely difficult to track the phase in which the system fails.

This separation is important because there is a real necessity for a controlled production environment, to guarantee the availability of this type of product. To control the production environment IPMA needs to organize their processes and to gather as much information as possible about the assets where the activities of the processes occur.

After gathering this information, it was decided that this information should be integrated within a information system already existent in IPMA. This integration will allow a more centralized monitoring ,with the focus on the product, of the production chain. Thus, allowing to IPMA to respond more objectively in case of failure.

6 - Conclusions and Future Work

During the initial phase of implementation of this methodology it was foreseen that marine SDI was crucial to prepare geographic information existent at IPMA. In order to accomplish SNIMar objective was necessary to prepare geographic information existent in the institution and to guarantee these information is produced. In order to fulfil this mission there were made some structural changes with the objective of incorporate the recently created new dimension for IT metadata. Following this methodology, defined throughout this article, it was proved, in a first stage, that using this methodology it will be guaranteed products availability at different phases of production (input, processing and output).

Previously, the focus was given to the IT assets and not to the products, which did not describe IPMA's reality, regarding its core business. By changing this focus and giving more emphasis on the product and not in IT assets, it was establish the correlation between assets and products, giving a better perception to the users of the impact that a failure could have in product's delivery.

The two case studies, Zones of Bivalve Analysis and SWAN's model, described two different processes with different specificity. To accomplish the objectives of implementing these two study cases it was necessary to create a normalization regarding the identification and indexing of information to simplify this process.

Although the SDI is still far from being accomplished, all data infrastructure was identified and the relationships between all the assets and products were created, therefore in this way it is possible now to guarantee product’s availability, because with the CMDB it was possible to integrate within an information system all the information about the product’s production process.

Each time, a product delivery fail, a mechanism of tickets is available for reporting this issue. Therefore guaranteeing in this way, support to all the phases of production, and providing integrity to processes that run inside IPMA.

The process described throughout this article, was in its majority a data collection process which represented a step towards the identification of all assets, including algorithms and procedures required to compose the geospatial information that will be a constituent part of the SDI.

.

References

[1] Strain, L.; Rajabifard, A.; Williamson, I. Marine Administration and Spatial Infrastructure. Marine Policy 2006, 30, 431-441.

[2] Bernard, L.; Kanellopoulos, A.; Annoni, A.; Smits, P. The European geoportal – one step towards the establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems 2005, 29, 15-31.

[3] Global Spatial Data Infrastructure Association.Spatial Data Infrastructure Cookbook, 2009. Available at http://www.gsdidocs.org/GSDIWiki/index.php/Main_Page (accessed on 21 April 2015).

[4] European Commission. Regulation on Inspire Data and Service Sharing. Available at http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32010R0268 (accessed on 19 March 2015).

[5] Direção Geral do Território. Sistema Nacional de Informação Geográfica. Available at http://snig.dgterritorio.pt/portal/ (accessed on 10 March 2015).

[6] Yalcin, G. Initial organizational studies on national Spatial Data Infrastructure at government level. In Proceedings of the 7th International Conference Interdisciplinary in Engineering (INTER-ENG 2013), Tirgu Mures, Romania, 2013; Procedia Technology 12 (2014) 572-576.

[7] BMC Software. What Do You Need from a Configuration Management Database (CMDB). Best Pratices White Paper 2006.

[8] ISO 19139:2007: Geographic Information –Metadata—XML schema implementation. International Organization for Standardization, Geneva, Switzerland.

[9] European Commission. INSPIRE Directive. Available at http://inspire.ec.europa.eu/ (accessed on 19 March 2015).

[10] SWAN Model. Simulating WAves Nearshore. Available at http://swanmodel.sourceforge.net/ (accessed 28 April 2015).

Paper reference no 293
Name of the presenter: Valéria Pacheco e José Santos
Author's affiliation: Instituto Português do Mar e da Atmosfera
email address : valeria.pacheco@ipma.pt , jose.santos@ipma.pt
Telephone number: (351) 218 447 000

Valéria Pacheco is from Porto and currently she lives in Lisbon. She graduated in computer engineering in 2010 from ISCTE-IUL (Instituto Superior de Ciências do Trabalho e da Empresa). She started her professional career in banking and insurance industry where she worked as analyst, tester and developer. She took a post-graduation in Information Systems and Knowledge management. Valéria also have experience in giving classes to undergraduates. Currently, she is working on her thesis related with Information Systems and she is working has a Junior researcher at IPMA (Portuguese Sea and Atmosphere Institute) where she is providing IT support to one of IPMA's internal projects.

José Santos is a graduate student in Geography with specialization in Cartography and Geographic Information Systems (GIS). He obtained his master degree in 2012 in the field of GIS and Spatial Modelling from the Institute of Geography and Spatial Planning (IGOT). In the past he worked at the European Space Agency and until the begin of May he worked as a Junior research at the Portuguese Sea and Atmosphere Institute (IPMA). He is an enthusiastic user/developer of GIS Open Source and Linux Operating Systems. Currently, he is a charter member of the Open Source Geospatial Foundation (OSGeo).