A Bottom-up Approach for the Identification of Requirements and ICT solutions for Environmental Information Sharing

Monica De Martino[1], Sergio Farruggia[3], Marina Monti[1], Nicole Ostlaender[2], Emanuele Roccatagliata[3], Robin S. Smith[2]

[1] CNR-IMATI, Via De Marini, 6 – Torre di Francia - 16149 Genova, Italy
[2] European Commission-JRC, Via E. Fermi,2749 - I-21027 Ispra (VA)
[3] GISIG, Via Piacenza 54 - 16149 Genova, Italy
{demartino, monti}@ge.imati.cnr.it,{nicole.ostlaender, robin smith}@ jrc.ec.europa.eu,
e.roccatagliata@gisig.it, sergio.farruggia@fastwebnet.it

INTRODUCTION

European guidance and policy for the development of information infrastructures recommends that new data and information handling resources should be based on existing examples, as proposed by the European Interoperability Strategy (EIS[1]) and the Interoperability Solutions for European Public Administrations (ISA[2]). In addition, the recent Communication on a European Shared Environmental Information System (SEIS[3]) outlines the need to modernize and simplify the collection, exchange and use of data and information required for the design, implementation and monitoring of environmental policy. These policy drivers present challenges in the actual identification and comparison of often heterogeneous systems within the environmental information sharing domain; including the processes and resources which support the capture/discovery, processing, validation, analysis and dissemination of data/information about the environment. To address this challenge, we present a method to gather and analyze the components of environmental information systems that can contribute to the development of information infrastructures such as SEIS. Our approach is described as (principally) 'bottom-up': a community of practitioners propose candidate systems for analysis, illustrating what approaches are currently adopted to create, manage, use environmental data/information that aim to meet the goals/principles (or 'top-down' setting) of the SEIS policy arena, while also informing this policy's development itself.

Specifically, while the top-down approach identifies the high level strategic drivers, defining axes and focus areas, the bottom-up approach helps to understand and describe the current situation by supporting the selection and assessment of actions, scenarios and replicable solutions. Gathering such information is a challenge as, from the outset: Which questions should be asked? How should answers be processed? And how can we compare systems with entirely different scope but that support the same policy vision?

Addressing these challenges and questions, the paper proposes a structured stepwise approach combining the top-down and bottom-up aspects for the analysis of existing heterogeneous information systems and good practices within environmental data sharing. In a nutshell, the stepwise approach starts by grounding itself in both the political context and priorities that SEIS aims to support. Next, in a second step, a generic request is issued for good practices in the form of existing systems and procedures supporting these principles. These submissions are then analyzed to identify a range of discrete high level processes that act as 'lowest common denominators' and that describe the procedures the systems support in the form of business processes. The resulting high level model provides a framework for describing and comparing in more detail these heterogeneous systems that

---

[1] http://ec.europa.eu/idabc/en/document/7854

[2] http://www.epractice.eu/files/Interoperability%20Solutions%20for%20European%20Public%20Administrations%20(ISA).pdf

[3] http://ec.europa.eu/environment/seis

should support the same policy principles from a business process perspective, while, at the same time, providing a bottom-up vision of the policy itself.

The approach has been developed and successfully applied in the European project "a Network to enhance a European Environmental Shared and Interoperable System" (NESIS[4]). NESIS aims to provide input to SEIS by focusing on the ICT issues of SEIS implementation, providing an ICT roadmap and guidelines for its technical implementation. To achieve such a goal, NESIS has been conceived as a network to share and exchange experience (i.e. good practices) about ICT solutions to manage environmental information.

Following this brief introduction, section 2 describes the general stepwise approach, while the third section applies it to the NESIS case. Conclusions and future work end the paper.

THE STEPWISE APPROACH

The aim of the stepwise approach is to collect, analyze and compare different information systems and procedures that support the same policy, and to feedback resulting information on identified functional requirements and existing good practices to policy makers and to developers for technical specifications. The diversity of applications that are reflected through such an approach bears the risk to, metaphorically-speaking, compare apples and oranges. For example: an application that mainly deals with monitoring of the environment will have different characteristics and components than a system whose main purpose is to support information dissemination.

To address this issue we apply the concept of Business Process Modeling. In system engineering, this activity is typically performed by business analysts and managers to represent an enterprise's processes, so that they may be analyzed and improved. As a well-established discipline, it is also applied in the context of software development where it represents the first step of requirement analysis by modeling the environment which a software system could support. By applying business process modeling we acknowledge the potential diversity of applications, and turn this into an opportunity: we create a Business Process Model (BPM) that represents, on a high level, the various processes that are supported by the sum of the applications. This approach has two advantages: (i) it provides a first bottom-up vision of the business model towards the policy, (ii) it allows systems to be analyzed for each high level business process they support. Thus, we no longer analyze entire systems, but subsets of systems and how they support specific high level business process, which immediately become much more comparable.

The steps

The description of the steps covers the entire process from the initial request for examples, to the creation of the high-level business process, and, finally, to the use of this description as a means to analyze existing good practices.

As the steps mark a sequence of activities (represented as boxes) and events (represented as circles), they are represented in a workflow diagram (Figure 1). For the graphical representation of the High-level BPM the standardized Business Process Modeling Notation (BPMN)[5] is used.

We identify three main actors and roles, represented in figure 1, as horizontal lanes or pools:

- Policy maker that supports the development of a particular policy through the bottom-up approach.
- Stakeholder community (thematic network) that has the 'good practice' systems which are believed to support the principles and main drivers identified in the policy.

[4] NESIS, ICT-PSP grant agreement no. 225062, http://www.nesis.eu

[5] Version 1.2 of the BPMN (see http://www.omg.org/spec/BPMN/1.2) is currently undergoing a major revision towards BPMN 2.0.

- Network coordination body that acts as a switch between the policy maker and the stakeholder community. Its main task is to mobilize and interact with the thematic community and to report back to the policy maker.

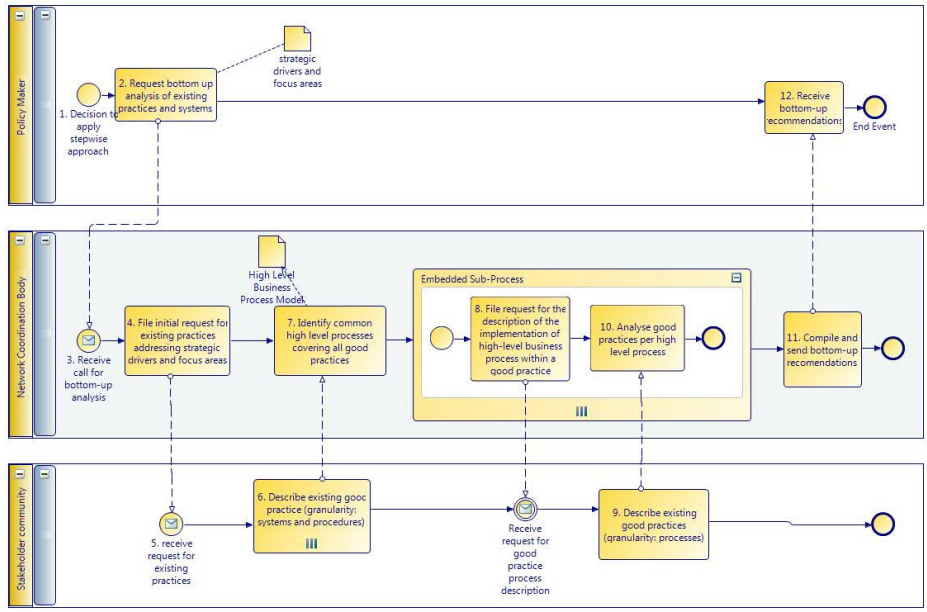

Figure 1: Approach overview.

The activities and events are described in the following paragraphs, using the numbered steps used in figure 1:

(1) The entire process starts with the decision to apply the top-down approach for a particular policy. (2) Following this trigger, the policy maker makes a request to the Network Coordination Body to perform an analysis of existing practices and systems addressing the policy's strategic drivers and focus areas. (3) Receiving this request, (4) the Network Coordination Body has the responsibility to identify and mobilize the community involved, filing an initial request to the stakeholder community that specifies the strategic drivers and focus areas to be supported. (5) Activated by this request, (6) the stakeholder community describes existing systems they consider to be good practices and sends these candidates back to the Network Coordination Body. Ideally, systems and practices should be described in an identical way, e.g. through a questionnaire or structured interviews. (7) The Network Coordination Body collects the good practices and performs an initial analysis to create a high-level BPM. This is a complex activity that may contain several sub-steps and the amount of work for the analyst depends on how the information in step 6 has been compiled. This step is described below in more detail involving a free text description from the questionnaire used in the NESIS case. This step aims to create a high-level BPM establishing a list of common high level processes from the existing material. This model represents all business processes (as well as abstract actors and the objects that are created or exchanged through the processes) supported by the systems and serving the policy drivers. (8) Given this frame, we match these systems with their high level processes, (9) by asking the system designers exactly how the system supports a particular process. Thus, we can focus on certain business processes of a system that system designers consider to be a particularly good practice. (10) We then examine, in detail, how two or more systems support the same business process, e.g. which parts of the processes are covered, the amount of automation the system provides, and the number of users that are served. This analysis allows us to point out which technological solutions (encodings, architectures, workflows etc) can be used for particular purposes

(i.e. processes, objects and roles). This information is compiled (11) and fed back to the policy maker (12).

THE APPLICATION OF THE STEPWISE APPROACH TO NESIS

In this sections we describe how the method has been applied to the NESIS project to create a bottom up vision of what SEIS should offer, based on the functionalities of existing systems. Please note that this discussion only covers steps 1-10, as these address the main methodological issues. Numbers identifying the various steps in the example follow the steps outlined above: (1) The analysis was initiated by the policy maker, (2) who requested the bottom-up analysis from the NESIS project. The policy maker referred the Network Coordination Body to the SEIS communication, as the source of strategic drivers and focus areas that the reference community should use when providing examples. (3) By receiving this request, NESIS took the role of the Network Coordination Body.

(4) NESIS initiated a call for existing practices addressing the strategic drivers and focus areas of SEIS. The stakeholder community that NESIS addressed includes the European Environment Agency's EIONET Community[6], including National Focal Points involving all the EU27 and EFTA Countries and the involvement of their National Reference Centres. (5) Receiving this request, the community selected good practices that: i) were proposed by EIONET National Focal Points partners and/or by the Members of the NESIS network, ii) are in compliance with the top down SEIS principles and iii) offer transferable experience for a wider audience of stakeholders. (6) As noted above, stakeholders used a questionnaire to describe entire systems and procedures. The results were collected in the NESIS Good Practice Catalogue (available on the project website). (7) The content of this catalogue allowed high-level business processes to be modeled. As indicated in the previous section, this step is complicated and is described in more detail, below.

High-level BPM identification (steps 8 and 9)

The first step towards the creation of the High-level BPM was to establish the list of candidate processes supported by the existing good practices. The relevant section of the questionnaire for the identification of business use cases was free text describing goals and procedures that the system carries out and supports, as well as the system architecture. Therefore, we used text analysis to identify process descriptions for the business processes involved, such as 'create environmental information', 'publish reports', and 'monitor air quality' that the systems support or fully automate. For each process, the roles involved were identified and named, as were the objects created or used/consumed (e.g. report, raw data etc). The freedom to describe the system (or procedure) as free text led to varying levels of detail. To deal with this inconsistency, we applied a simple classification with three levels of granularity: high, medium and low. High-level processes represented complex activities that might be applied in a large number of systems. Processes at the medium level involved more specific activities and low level processes might only be supported by a single system. Table 1 illustrates examples of processes and their categorization after a first screening.

After this first screening, the resulting list of processes and the linked roles and objects were analyzed further to establish the High-level BPM. Processes classified as "High-level" were immediately considered for the high-level BPM. "Medium" and "Low" level processes, on the other hand, were examined in more detail. If they could be considered as a part (or specialization) of an existing high-level process, then they were discarded as already included. If not, a corresponding high-level process had to be formulated. If this was not possible, or if the resulting process was not within the scope of the strategic drivers, the process was disregarded altogether.

[6] http://www.eionet.europa.eu

ID	Process Name	LEVEL (High | Medium | Low)
1	Adapt system to reflect changes in required information	Low
2	Analyze data	--
3	Raise Awareness	--
4	Carry out applications for land use planning *(NOTE In the sense of public participation)*	Low
5	Collect best practices	Medium
6	Collect (harmonized) reports	Medium
7	Collect monitored environmental information	Medium
8	Create definitions to be used in reporting obligations	High
9	Create information products	High
10	Create metadata	High
11	Deliver report	Low
12	Discover environmental information	Medium
13	Discover geographic information	Medium
14	Discover information	High
15	Discover monitoring stations	Low
16	Discover reporting obligation	Medium
17	Download data	Low
18	Enrich monitored information with red-list information *to create information products.*	Low
19	Harvest metadata	Medium
20	Manage databases	
21	Create metadata (through online form)	Low
22	Monitor content (of reportnet)	Low
23	Monitor environment (ozonelevel) through monitoring stations	Low
24	Monitor environmental information	High
25	Monitor soil quality	Low
26	Perform data quality assurance	High
27	Perform reporting	High
28	Prepare information product based on reporting	High
29	Process data to create information	High
30	Produce scientific publication	--
31	Provide information to the public	--
32	Perform public participation in environmental planning	--
33	Publish data	High
34	Publish data from monitoring stations	Medium
35	Publish environmental information	Medium
36	Publish geographic information	Medium
37	Publish information	High
38	Publish metadata	High
39	Publish report	High
40	Raise awareness for biodiversity issues	--
41	Register reporting obligation	Low
42	Report to public authorities	Medium
43	Search environmental information	High
44	Search for ozone information	Low
45	View (distributed) information	High
46	Search metadata	High
47	View data	High
48	View metadata	High
49	View data of monitoring stations	Medium
50	Visualize information in Form of Maps	High
51	Visualize ozone information	Low

Table 1: Identified processes and categorization after first screening of good practices.

Process categories	Process category refinement	Candidate Processes IDs (from Table 1)
Creation	Guideline creation	8
	Data creation	24
	Information Creation	9, 27, 28, 29, 42
	Metadata Creation	10,19
Dissemination		11, 33, 34, 35, 36,37,38 39 42
Discovery		12, 13, 14, 16, 43, 46
Consumption		45, 46, 47, 48, 49, 50
Feedback		26

Table 2: Mapping of processes into the High-level BPM represented as categories.

As a last step, to validate the high-level processes resulting from this procedure, all processes in Table 1 were arranged into categories representing the high level processes. This was done to ensure that all business processes addressed by any of the systems would appear in the resulting High-level BPM. The identified categories and the relationships with the processes listed in Table 1 are summarized in Table 2.

As the resulting high-level processes are abstract, the descriptions of roles and objects for these processes also had to be modeled on a higher level. In the following, we start describing the roles and objects which are necessary to understand the processes.

We identified eight actors and their roles. Each role is the main actor of at least one process dealing with certain objects:

- The *Creator* of an item. This actor is further specified by the type of item it produces: e.g. *Metadata Creator*, *Data Creator* etc.
- The *Supporter* of the efficiency of other processes through the design of specifications and guidelines, often undertaken by public authorities or standardization bodies.
- The *Data Provider*: often undertaken by Member States' public authorities but also by NGOs and citizens (as volunteers).
- The *Analyst* of data and information. This actor creates information products for the *End User/Consumer*.
- The *Metadata Provider* is typically the same individual that provides the data or information product which the metadata describes.
- The *Feedback Provider* checks items and provides feedback (e.g., on an item's quality) and is normally a third party, i.e. not the *Data Provider* themselves.
- The *Disseminator* registers and publishes items for the *End User/Consumer*.
- The *End User/Consumer* consumes items published by a *Disseminator*.

The objects we distinguish in the High-level BPM are:

- *Service* any entity that allows operations to be performed on the related item(s).
- *Item* an abstract concept, describing an object linked to information exchange. It has the sub-concepts: *data, information, metadata* and *guideline*.

This latter *Item*, *guideline*, is addressed in the example below. It supports the creation, publication or consumption of items. Examples for guidelines are *data specifications*, *templates* and *schemas to support harmonized reports*, and *service specifications* that support interoperability in a service-oriented architecture.

Analysis of a subset of Good Practice (Step 10)

Structured interviews were performed for a deeper analysis in order to identify how the high-level business processes have been implemented on a subset of good practices from the catalogue: Reportnet, PortalU, WISE, Artsdatabanken, NatureSDIplus Thesaurus Framework. A complete description of the analysis results, as well as the indications captured for the SEIS ICT guidelines, are included in the NESIS deliverables. Below, we provide just one example from the Thesaurus Framework developed within the NatureSDIplus project[7] and discuss some of the main aspects, organized by *item* and process.

In NatureSDIplus, a Thesaurus Framework aims to assemble well-known Knowledge Organizations Systems (KOS; e.g. thesauri, classification, and taxonomies) to address multilingual/multicultural issues in data sharing for nature conservation. The thesaurus framework is exploited at the 'geoportals level' to support two processes identified in the High-level BPM: metadata creation and data discovery.

As noted above, the framework and its sub-thesauri are *Guidelines*. The three processes covered are:

Create/Maintain Guideline. The framework is used to enrich the metadata profile with semantics. Semantic Web solutions have been adopted to exploit as much as possible the knowledge organized in the different thesauri. In particular, Simple Knowledge Organization Systems (SKOS) has been adopted as a solution to encode each KOS: it provides a standard way to represent KOS using the Resource Description Framework (RDF). RDF encoding allows it to be passed among computer applications in an interoperable way. Through the SKOS representation and RDF encoding, the

[7] See http://www.nature-sdi.eu

framework allows links to be created between the concepts in the different thesauri that exist within the framework (SKOS, 2009).

Publish guideline. The framework[8] and the sub-thesauri are published as Linked Data (Bizer *et al*. 2009). Linked Data is the optimal solution as it allows the framework requirements to be met: due to the modular structure, the framework content can always be extended with new thesauri and for each thesaurus to be interlinked with the existing thesauri in the collection.

Applying guideline. The framework has been employed for the metadata compilation and the data discovery at the geoportals level. Metadata compilation is exploited within the MDWeb tool[9] by using the thesaurus concept as a code list within the metadata profile. Data discovery is employed at the geoportals level by employing dedicated web services. The thesaurus supports the user in data search by suggesting keywords for query formulation.

This example of good practice shows that many components are already in place within the overarching concept of SEIS but efforts are needed to enable the identification and analysis of candidates for the overall architecture. By looking into other good practices supporting the same processes, and by comparing their workflows, components and technological solutions, we believe this is a first step towards identifying candidates for best practices and, thus, developing SEIS from the bottom-up.

CONCLUSIONS

This paper presents a method for comparing heterogeneous information systems in order to provide input to the development of ICT solutions and policies for environmental data sharing. The method, based on Business Process Modeling, has been successfully applied in the European project NESIS, thus demonstrating the validity of the approach. A set of business processes that currently exist in the environmental data and information communities of the EU Member States has been analyzed according to the proposed methodology. The analysis has been used to derive some requirements for the ICT services that SEIS should support, providing examples of processes needed for such an infrastructure to work beyond the information-sharing principles it sets in place. Arguably, many of these activities have tended to focus on the national-to-European level, with stronger links to the e-government related domain. This level of activity and related systems are, however, less likely to engage with environmental monitoring and more bureaucratic data-capture activities often found at sub-national levels that may also contribute to the development of SEIS.

It is believed that further research should take place to verify if our general approach can be readily applied to information infrastructures in other contexts, helping to create greater integration of other data and information in the wider European governmental sphere, from local to global levels and across domains.

Acknowledgements

This work has been supported by the EC under the ICT-PSP Programme within the project NESIS No. 225062. Thanks to the NESIS and NatureSDIplus Project Partners for their contributions.

REFERENCES

Bizer C., Heath T., Berners-Lee T. (2009) Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.* 5(3): 1-22 (2009)

Kruchten , M. (2000) *The Rational Unified Process. An Introduction*. Addison -Wesley, Reading MA

Jacobson I., Booch G., Rumbaugh J. (1999) *The Unified Software Development Process.* Addison-Wesley Professional, ISBN 0-201-57169-2

SKOS (2009) *SKOS: Simple Knowledge Organization System*, http://www.w3.org/TR/skos-reference/

[8] Available at http://linkeddata.ge.imati.cnr.it:2020/

[9] See http://mdweb.codehaus.org

SKOS (2009) *Simple Knowledge Organization System Primer,* W3C Working Group Note 18 August 2009 http://www.w3.org/TR/skos-primer