FP7 Project nr	226487	Project start date:	01 May 2009
Acronym:	EuroGEOSS		
Project title:	EuroGEOSS, a European approach to GEOSS		
Theme:	FP7-ENV-2008-1: Environment (including climate change)		
Theme title:	ENV.2008.4.1.1.1: European Environment Earth Observation system supporting INSPIRE and compatible with GEOSS (Global Earth Observation System of Systems)		
Web site:	www.eurogeoss.eu		

D.2.3.1. Interim Report on advanced workflow modelling

Title	D.2.3.1. Interim Report on advanced workflow modelling
Creator	Lorenzino Vaccari
Creation date	25/08/2010
Date of last revision	28/06/2011
Subject	Interoperability, User Requirements, Initial Operating Capacity, Workflow
Status	☐ Draft ☒ Final
Publisher	EuroGEOSS, WP2
Type	Text
Description	
Contributors	Max Craglia (JRC), Carlos Granell (JRC), Laura Diaz (UJI), Michele Chinosi (JRC), Nicole Outslander (JRC), Sven Shade (JRC), Didier Leibovici (UNOTT), Mattia Santoro (CNR)
Format	Doc
Source	
Rights	☐Restricted ☒Public
Identifier	EuroGEOSS_D_2_3_1.doc
Language	En
Relation	D2.4.1, D2.6.1, D3.4, D4.4, D5.4
Coverage	Not applicable

These are Dublin Core metadata elements. See for more details and examples http://www.dublincore.org/

TABLE OF CONTENTS

FIGURES

TABLES

ACRONYMS AND ABBREVIATIONS

Abbreviation	Name
AOC	Advanced Operating Capacity
APAAT	African Protected Areas Assessment Tool
APM	Area Production Model
ASRA	Adaptive Spectral Representativity Analysis
BPEL	Business Process Execution Language
BPEL4WS	BPEL for Web Services
BPMI	Business Process Modeling Initiative
BPMN	Business Process Modeling Language
CEN	European Committee for Standardization
CHE	Ebro River Basin Authority
CIFOR	Center for International Forestry Research
CNIG	Centro Nacional de Información Geográfica
CNR	Centro Nazionale delle Ricerche
CORINE	Co-ordination of Information on the Environment
CSW	Catalog Service for the Web
DCMSEE	Center for drought management in Southeastern Europe
DEM	Digital Elevation Model
DOPA	Digital Observatory of Protected Areas
EDISOFT	Empresa de Serviços e Desenvolvimento de Software, S.A. – Portugal
EDO	European Drought Observatory
EFDAC	European Forest Data Centre
EFFIS	European Forest Fire Information System
EFICP	European Forest Information and Communication Platform
ENM	Ecological Niche Model
EU	European Union
FAO	Food and Agriculture Organization

FGUA	Fundacion General de la Universidad de Alcala
FP7	Seventh Framework Programme
GBIF	Global Biodiversity Information Facility
GEM	Global Environment Monitoring
GEO	Group on Earth Observations
GeoRM	Geo Rights Management
GeoRSS	Geospatially-enabled RSS and Atom feeds
GEOSS	Global Earth Observation System of Systems
GISCO	Geographic Information System of the European Commission
GMES	Global Monitoring for Environment and Security
GSCB	Ground Segment Coordination Body
GSF	GEOSS Service Factory
IBA	Important Bird Area
IDEE	Spanish Spatial Data Infrastructure, Infraestructura de Datos Espaciales de España
IES	Institute for Environment and Sustainability
INSPIRE	Infrastructure for Spatial Information in Europe
IOC	Initial Operating Capacity
IR	Implementing Rules
IUCN	International Union for Conservation of Nature
ISO	International Organization for Standardization
JRC	Joint Research Centre
LISS	Linear Imaging Self-Scanning System
LMNH	Land Management and Natural Hazard
LUCC	Land Use and Cover Change
MARM	Ministerio de Medio ambiente Y Medio Rural Y Marino
MAS	Multi-Agent Systems
MCPFE	Ministerial Conference on the Protection of Forests in Europe
MS	Member State
NDVI	Normalize Difference Vegetative Index

NFI	National Forest Inventory
NGO	Non-Governmental Organization
NSDI	National Spatial Data Infrastructure
O&M	Observation and Measurement
OASIS	Organization for the Advancement of Structured Information Standards
OFAC	Observatory for the Forests of Central Africa
OGC	Open Geospatial Consortium
OSE	Observatorio de la Sostenibilidad en España
OWS	OGC Web Services
RDSI	Reference Data and Services Infrastructure
RSPB	Royal Society for the Protection of Birds
RWER	Rural, Water and Ecosystem Resources
SAIH	Automated Hydrology Information System
SAWSDL	Semantic Annotations for WSDL
SCAS	Service Chain Access Service
SDI	Spatial Data Infrastructure
SEIS	Shared Environmental Information System
SIA	Integrated Water Information System of Spain
SOA	Service Oriented Architecture
SOS	Sensors Observation Service
SPOT	Satellite Pour l'Observation de la Terre
SRTM	(NASA) Shuttle Radar Topographic Mission
SWE	Sensor Web Enablement
TREES	United Nations Environment Programme
ULBF	University of Ljubljana Biotechnical Faculty
UNEP	United Nations Environment Programme
USGS	U.S. Geological Survey
WBDB	World Biodiversity DataBase
WCMC	World Conservation Monitoring Centre

WCS	Web Coverage Service
WCS-T	Web Coverage Service, Transactional
WDPA	World Database on Protected Areas
WfMC	Workflow Management Coalition
WFS	Web Feature Service
WFS-T	Web Feature Service, Transactional
WMS	Web Map Service
WPS	Web Processing Service
WP	Work Package
WS-BPEL	Web Service – BPEL
WSDL	Web Service Definition Language
XPDL	XML Process Definition Language

1 INTRODUCTION

GIS desktop applications provide to the user many complex functions to perform GIS data acquisition, creation, analysis, process, and mapping. For years these functions were accessible only through the GIS desktop applications, but recently, GIS services have became published and available on the web. At the moment, the majority of these geo-services exist as single services. After discovering, services can be composed or coordinated to provide complex functionalities. In this case, a manual and static composition of a number of predefined geo-services has to be performed thus generating simple or complex workflows which have to be designed, developed, executed and distributed on the web. In this context one of the goals of the EuroGEOSS project, within the work package related to the multidisciplinary interoperability, is the investigation of advanced service chaining methods and mechanisms for modelling workflow of distributed geo-processing services.

This document investigates gaps, requirements and methodologies currently available when modelling workflows of distributed geo-processes published on the web. This analysis will provide the background knowledge to support the applicability of workflow modelling methodologies to the scenarios, use cases and processes available from different EuroGEOSS thematic areas.

Different issues have to be considered in order to apply workflow modelling within a multidisciplinary context. First of all, to design a process model and its workflow counterpart it is necessary to use a notation which is capable to fill the gap among all the different kinds of designers and users, and it is either sufficiently recognized as a standard or it is used as a common workflow modelling practice. Moreover, this notation has to provide the modeller the possibility to design a workflow, to exchange it with other modellers and users, to validate it and, if needed, to run it. Then, a mechanism to discover appropriate workflows, either to know their characteristics or to re-use them in different contexts, has to be implemented, adopted and provided to the users of different disciplines. Finally, mechanisms able to chain and run workflow executable instances have to be analysed and, eventually, adopted.

In the following sections we will first focus on workflow modelling issues, and then present a list of solutions available to chain and execute workflows. Specifically, Section 2 will briefly illustrate basic workflow modelling issues. Section 3 will extensively illustrate BPMN, of the most adopted solutions to design, develop, exchange and run workflows. Section 4 will present advanced methods and mechanisms to chain distributed geo-processing services. Section 5 will give some conclusions and will summarize future work.

2 WORKFLOW MODELLING

As defined by the Workflow Management Coalition (WfMC), "*a workflow is the automation of a business process, in whole or part during which documents, information or tasks are passed from one participant (a resource; human or machine) to another for action, according to a set of procedural rules"* (Hollingsworth, 1995). Currently, thematic experts provide use case descriptions (very often using simple natural text) to workflow modellers. The modellers analyse these descriptions and generate draft workflow descriptions. This approach faces three problems. First of all, use cases are frequently described by technical staff rather then real thematic experts. Such people have usually thematic software systems in mind when describing a use case. Consequently, the use case descriptions are similar to a specialistic software-designer view instead of illustrating a clear high-level view on user requirements. Second, thematic experts are usually not trained in use case development. Descriptions become ambiguous and incomplete. For example, actors in a workflow are not well separated, distinct steps are not represented at the right level of detail, and listed actor names are not consistenly used throughout the natural language description of the workflow. Third, the transcription of natural language use case definitions to workflow modelling languages is error prone and the workflow modeller may inpose own interpretations. Moreover, when the context involve different disciplines, there is the need to adopt a notation that allows scientistis and users of different disciplines to understand and exchange process models and workflows. The notation has also to be understood by different kinds of users such as decisors, modelers, scientists, IT people, etc.

To better analyze these issues, for each EuroGEOSS thematic area, an investigation of the gaps on the textual description of the use case workflows (sequence of steps) was done. Some of the issues, found while analyzing workflow descriptions, include:

Actors definitions

- Actors' names often change between use case header and description.
- Acronyms are sometimes used, making difficult to recognize the actors.
- Actors' names similarities generate ambiguity (e.g., User, EuroGEOSS user, System user, etc.).
- Definition of what an actor does is missing or incomplete.
- Declared actors don't appear inside the use case description.
- Lack of meaningful details about actors.
- Sometimes, actors' roles are not so clear.

Actions / Steps

- The description of the tasks is ambiguous.
- Tasks/steps have no actors associated.

Alternatives paths and flows

- Exception handlings are often missing.
- A possible failure case is introduced but the failure management is missing.
- Some paths, illustrated in the use cases, are missing.

Steps definition granularity

- Two steps definitions are declared inside the same step description.
- Some steps lack specified actions.

Parameters and attributes specification

- Parameters and attributed specifications are often missing

Use of 'formalized' language

- Some confusion in using "AND", "OR", "OTHERWISE" was observed.

Pre- and Post-Conditions

- Pre- and post conditions are often missing

To overcome these issues various design approaches can be adopted and a formal description of the workflows has to be used. In the following section we present some of the approaches used to model workflows.

2.1 Workflows design approaches and formats

Generally, users from various environmental and geosciences disciplines would adopt three different workflow design approaches to model workflows. Modellers may use them in their natural separated order from purely conceptual to programming viewpoints or in a mixed way:

- The first approach, referred to as **abstract workflow specification**, defines complex tasks in an abstract way. This workflow modelling is suitable to describe workflows from a high level of abstraction and makes the descriptions more suitable also for users without programming skills. This is suitable for a conceptual primer modelling.

- The second approach can be considered as an **intermediate approach** which leverages the expressive power of the complete notation to describe the activity flow precisely, including the exception paths significant to key performance indicators. In this case, the models are not still executable but they have a defined semantics and are subject to validation rules. Moreover, in this phase the modeller can, for example, define pre-conditions, post-conditions and select instances of the processes which participate to the workflow. These workflows are not still executable because some technical details (e.g. specification of data structures), that would be required to execute the model on a process engine, are omitted. Nevertheless, these workflows can be simulated to test their validity and completeness. This approach will correspond to an algorithmic description of a conception model.

- The third approach, called **executable modelling**, makes it possible to execute workflow specifications by using workflow execution engines. In this case, a number of technical details, enabling the execution of the workflow, have to be programmed. These details can be ignored when modelling using the first and the second approaches. However, it is possible to apply some methods and styles to the first approach and the second approach so that extending it to the third approach requires only refinement, not major structural changes.

Only recently, some convergences appeared around the adoption of different formats and tools which offers the domain users the possibility to use any of the three approaches previously discussed (Barker & Van Hemert, 2007), (Lu and Sadig, 1997), (Leibovici & Pourabdollah, 2011).

Alike UML modelling allows various levels of description of a system from purely conceptual to its implementation (e.g. database, object programming), the Business Process Model and Notation (BPMN) is one of the most adopted solutions for modelling workflows[1] (Chinosi and Trombetta, 2011).

3 WORKFLOW MODELLING THROUGH BPMN

The BPMN, originally developed by the Business Process Management Initiative (BPMI), is a graphical notation with a clear semantic. BPMN presents also a mechanism to obtain (semi-)automatically XML code to deploy in workflow engines or to be shared across different domains.

[1] http://www.bpmn.org/BPMN_Supporters.htm

The most used release of BPMN is the version *1.2*[2], even if a new major version (*BPMN 2.0*) has been published in January 2011. The primary goal of BPMN is to provide a notation that is readily understandable by all business users, from the business analysts, that create the initial drafts of the processes, to the technical developers, responsible for implementing the technology that will perform those processes, and finally, to the business people, who will manage and monitor those processes. Thus, BPMN creates a standardized bridge for the gap between the business process design and process implementation. Another goal, but no less important, is to ensure that XML languages designed for the execution of business processes, such as BPEL and XPDL can be visualized with a business-oriented notation.

Figure 1 shows a summarised overview of the whole set of BPMN elements.

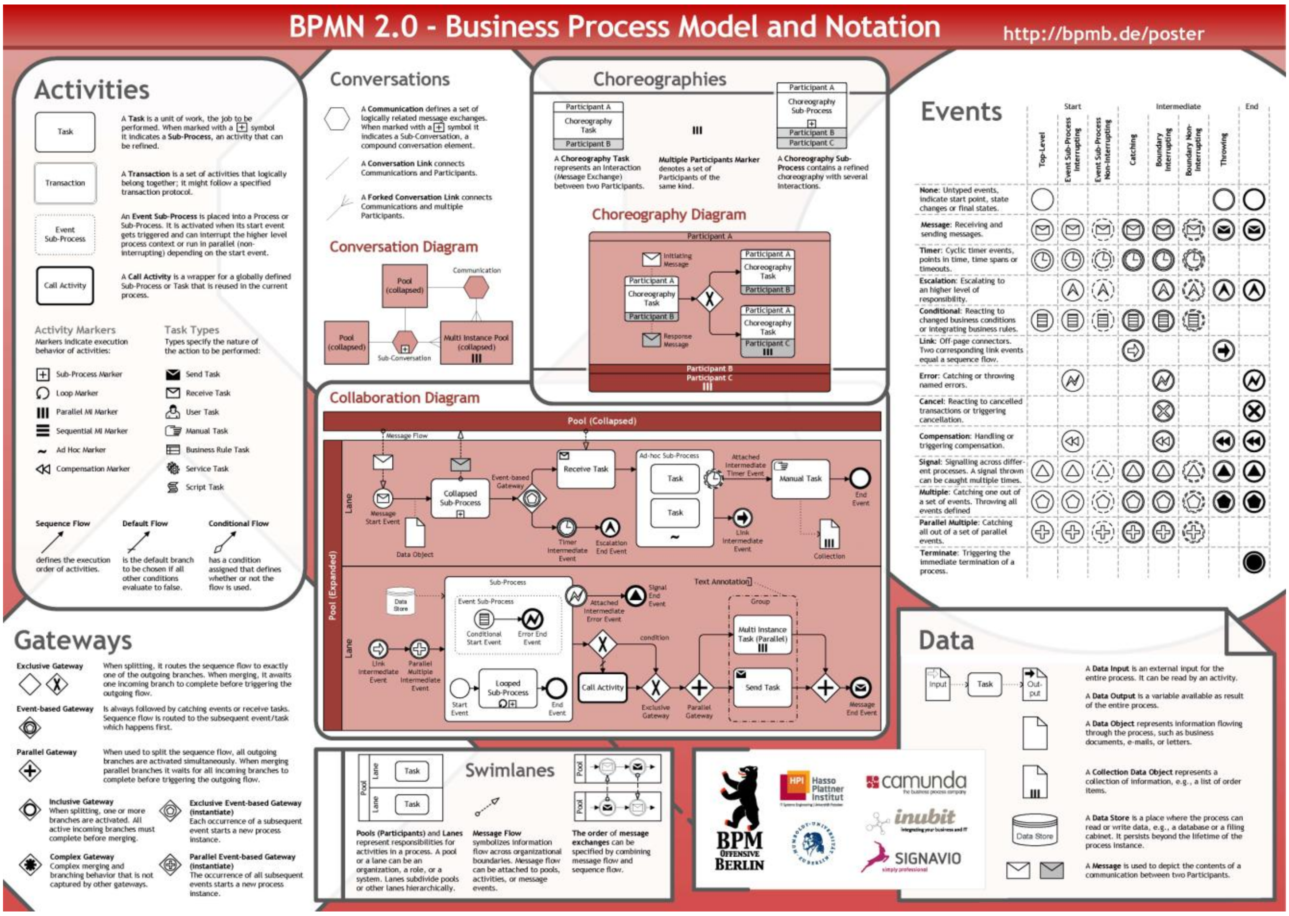

Figure 1 - BPMN Notation

BPMN provides a small set of notation categories so that the reader of a BPMN diagram can easily recognize the basic types of elements and understand the diagram . Within the basic categories of elements, additional variation and information can be added to support the requirements for complexity without dramatically changing the basic look and feel of the diagram. The basic BPMN categories of elements are:

1. *Flow Objects*, which are the main graphical elements to define the behaviour of a Business Process (they define events, activities, gateways).
2. *Data*, which represent generic data objects, inputs, outputs and data stored.

[2] http://www.omg.org/spec/BPMN/1.2/

3. *Connecting Objects*, which connects the flow objects to each other or to other information.
4. *Swimlanes*, which group the primary modelling elements.
5. *Pools,* which represent the Participants to the workflow.
6. *Artefacts*, which provide additional information (groups, text annotations).

3.1 BPMN at work: modelling thematic workflows

This section presents an example (see Figure 2) on how to design workflow activities by using BPMN and compare this representation to a flow-chart diagram of the same workflow.

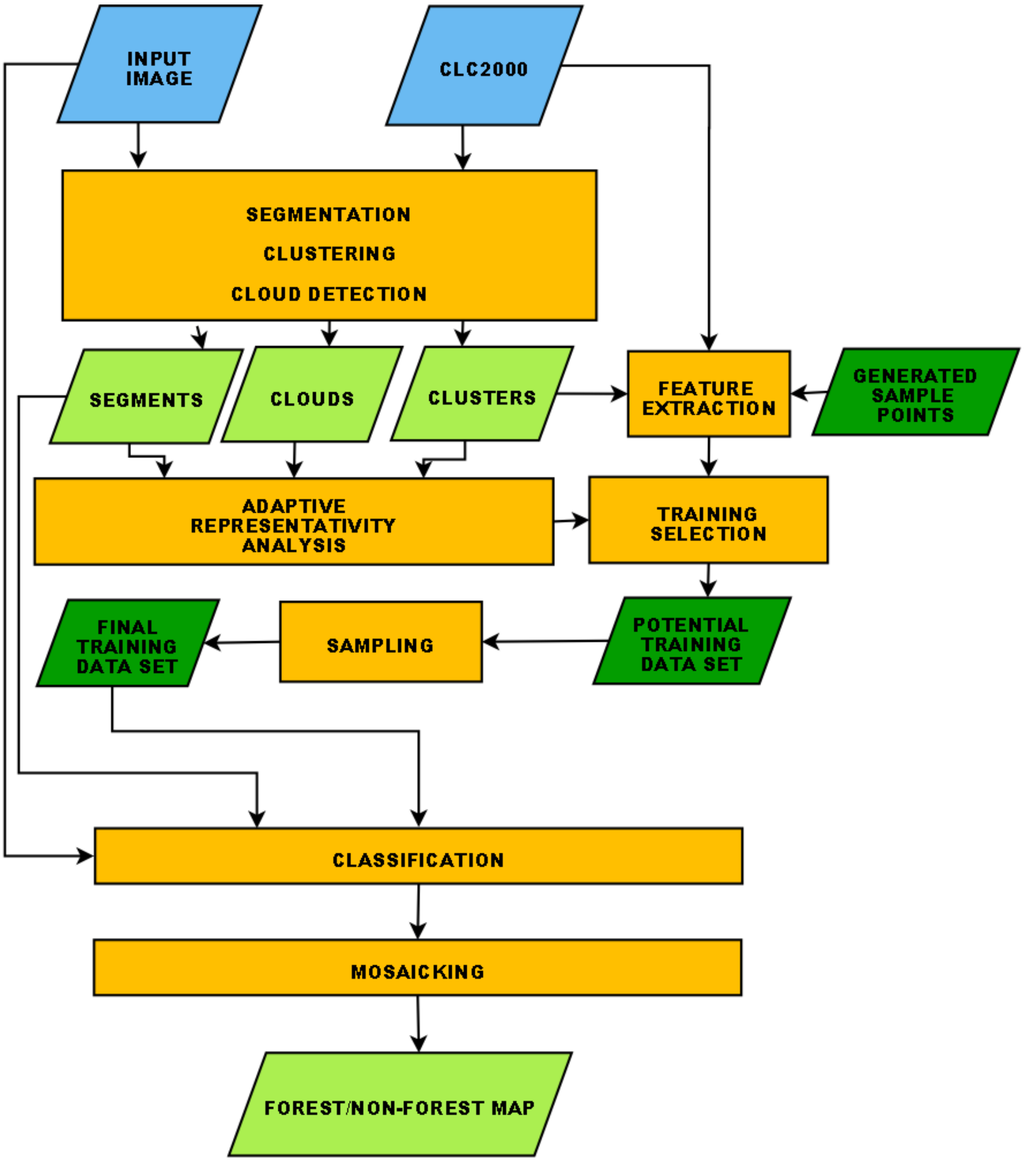

Figure 2 – Basic workflow of Forest/Non forest Map production

The example is taken from an internal JRC initiative which goal is, among the others, to improve understanding of the data flows inside and among different JRC institutes. The example illustrates the workflow to produce the pan-European forest/non-forest map[3] .
The primary objective of the map production, performed by the Forest action of LMNH (JRC), is to produce a Pan-European high resolution forest/non forest map (FMAP2000) and the forest type map of Europe[4] which would support pan-European reporting on sustainable forest management (Pekkarinen et al., 2009). To produce the map, an automated processing chain consisting of several phases that included image segmentation, clustering, adaptive spectral representativity analysis (ASRA), training data extraction, supervised nearest-neighbour classification, and the composition of final output mosaic, are developed. The purpose of the image segmentation is to reduce the number of entities to process. Working with segments instead of pixels increased the image processing efficiency in the clustering and classification phases. Clustering is used to derive spectral classes needed in the ASRA. The ASRA is applied to recognize representative combinations of spectral and informational, i.e. the CLC2000, classes. The training data for the nearest-neighbour classifier is derived using generated sample points and the information obtained in the ASRA. Finally, the output mosaic is computed using the classified scenes and uncertainty information derived during the classification process.

The resulting methodology is used to map European forests using more than 400 Landsat ETM+ images. The resulting forest/non-forest map is validated with two different point level datasets. Furthermore, the map is used to compute forest area estimates that are compared to reported country-level forest area statistics.

While BPMN's similarity to flowcharting notation makes it friendly and familiar to business users, it presents some additional features in respects to the traditional flowcharting (Silver, 2010).

Figure 3, for example, shows the same workflow represented by using the BPMN formalism[5] but, from the comparison to the flowchart, it is possible to identify a number of elements which are not represented in the basic flowchart represented in Figure 2:

- BPMN is a standard and has a precise semantic, because each shape, symbol and connection has a well defined meaning.
- It is possible to clearly identify the following participants to the workflow, represented by *Pools*:
    - the forest action of the IES LMNH unit, which performs the activities to generate FMAP2000,
    - the providers of the input and ancillary datasets:
        - Global Land Cover Facility[6], a centre for land cover science, which provides the Landsat ETM+ imagery data
        - IES SDI (I&CLC2000), the JRC IES which provides the Corine Land Cover 2000 dataset (seamless Vector layer re-sampled to 25m)[7]
        - USGS, the U.S. Geological Survey, which provides the SRTM Digital Elevation Model[8]
        - Netapp, a JRC intranet GIS repository which provides local forest data and which collects the final results of the workflow,

[3] http://forest.jrc.ec.europa.eu/forest-mapping/forest-cover-map/2000

[4] http://forest.jrc.ec.europa.eu/forest-mapping/forest-type-map

[5] http://forest.jrc.ec.europa.eu/forest-mapping/forest-cover-map/2006-forest-cover-map

[6] http://www.landcover.org/

[7] http://www.eea.europa.eu/data-and-maps/data#c5=all&c11=landuse&c17=CLC2000&c0=5&b_start=0

[8] http://srtm.csi.cgiar.org/

  - the Internet Web site[9] of the Forest action where the final datasets are published and available to download.
- It is possible to clearly identify input datasets (*Landsat*, *local forest data*, *SRTM-DEM, CLC2000)*, intermediate products (*segments*, *clusters, clouds, potential training datasets, final training datasets*) and final results *(MSPA, Forest Map-ZIP, Forest map-KMZ, 1Km forest percentage map)*.
- It is possible to define sub-tasks, such as the *data correction* task showed in Figure 3.

Moreover, BPMN supports the *event-triggered behaviour*, where an event is “something that happens” asynchronously while a process is happening.

Finally, notice that the diagram is just the surface layer of a complete XML language for process definition. Using the XML serialization, the model can be exported and run through a simulation engine, or even the corresponded process can be automated by using a process engine.

[9] http://forest.jrc.ec.europa.eu/

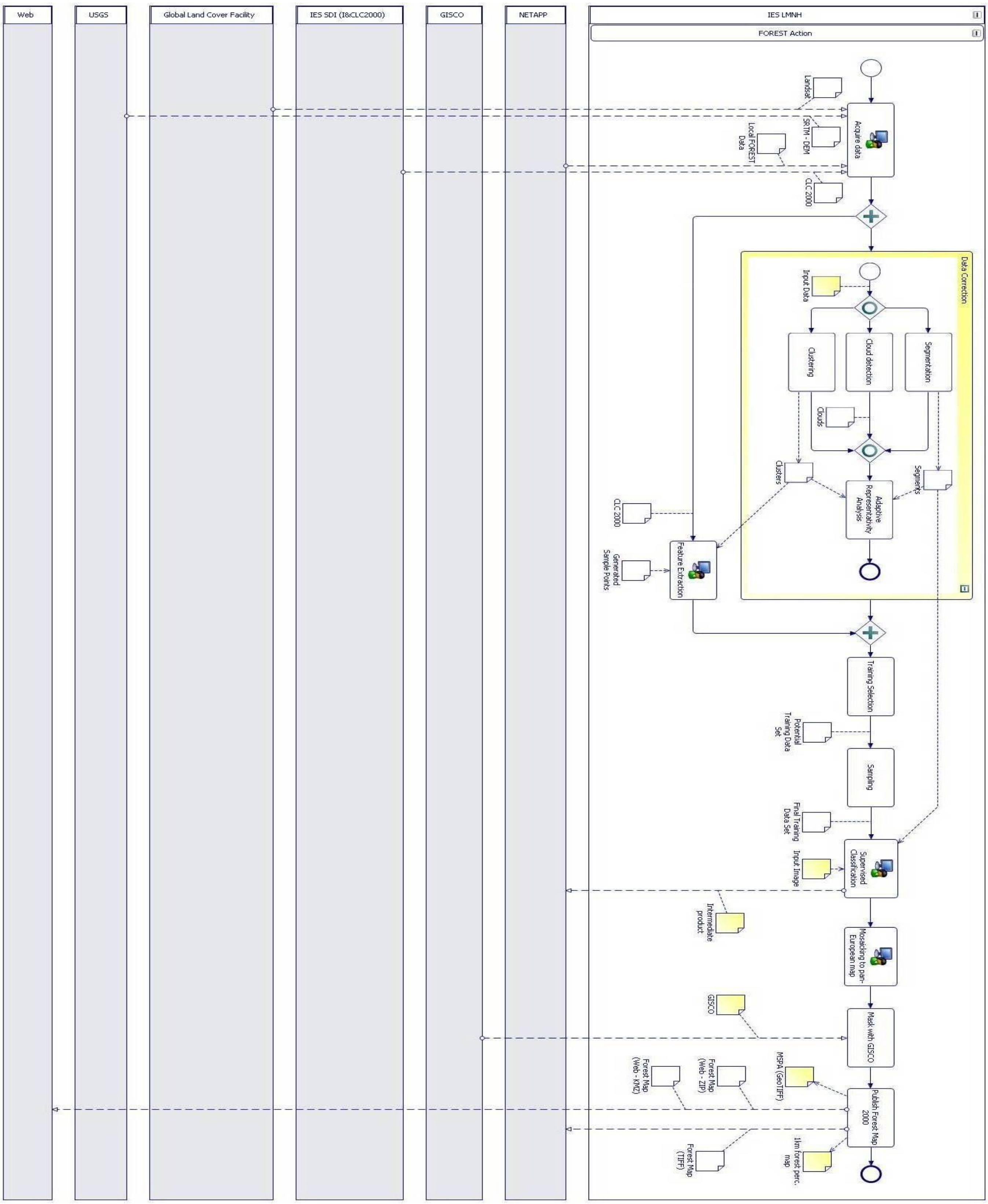

Figure 3 – BPMN representation of the pan-European Forest Map generation

3.2 Serializing BPMN

Up to now, they are two widely adopted solutions to serialize BPMN diagrams using XML[10]. Both the XML Process Definition Language (XPDL)[11] and the Business Process Execution Language (BPEL)[12] are largely adopted solutions for serializing BPMN even if they were not proposed as native BPMN serializations. A third solution, namely BPMN 2.0, was officially proposed in 2011 as native BPMN serialization.

It is worth noting how the XPDL definition was adapted to support BPMN, while the BPEL definition is completely independent from BPMN. BPEL came before BPMN (WS-BPEL 2.0 has been published in 2007, while BPMN 1.1 in 2008) while XPDL has been modified in 2008/2009 according to BPMN specifications in order to support this emerging standard. Most relevant characteristics of BPEL, XPDL and BPMN 2.0 follow:

- BPEL is an execution language. It is a programming language that has variables and operations. The operations can send and receive SOAP messages, and there is strong support for XML and XML transformation. It has constructs that make it easy to call multiple web services at the same time, and synchronize the results. It does not have any concepts to support the graphics of the diagram. Ultimately, the goal of BPEL is to provide a definition of web service orchestration, the underlying sequence of interactions, the flow of data from point to point. It does not however attempt to represent the drawing that specifies the orchestration.

- XPDL is mainly a process design format which represents the “drawing” of the process definition. It has X & Y coordinates and node sizes. It has a concept of lines, and points along the line that give it a particular path. The nodes and lines have attributes which can specify executable information such as roles, activity descriptions, timers, web service calls, etc. XPDL 2.1[13] contains extensions in order to be able to represent all aspects of BPMN, including executable properties: “*XPDL provides a file format that supports every aspect of the BPMN process definition notation including graphical descriptions of the diagram, as well as executable properties used at run time. With XPDL, a product can write out a process definition with full fidelity, and another product can read it in and reproduce the same diagram that was sent*”[14].

- In order to propose a notation that could include both design and executable aspects, in 2009 a OMG team started to work on the new version of BPMN specifications which was published on January 2011. The new version of the specifications come together with a complete and native XML Schema definition, namely *BPMN20.xsd*, filling the gap still not filled by neither XPDL nor BPEL. This new native and official XML Schema allows separating the graphical

---

[10] http://social-biz.org/2006/05/26/bpmn-xpdl-and-bpel/, http://www.brsilver.com/2007/03/21/the-real-issues-with-xpdl-bpel-and-bpmn/

[11] http://www.wfmc.org/xpdl.html

[12] The acronym BPEL is commonly used if not discussing a specific version. IBM and Microsoft firstly decided to combine their workflow definition languages into a new language, BPEL4WS. In April 2003, BEA Systems, IBM, Microsoft, SAP and Siebel Systems submitted BPEL4WS 1.1 to OASIS for standardization via the Web Services BPEL Technical Committee. The Committee voted on 14 September 2004 to name their specification "WS-BPEL 2.0".

[13] http://www.wfmc.org/index.php?option=com_docman&task=doc_download&Itemid=72&gid=132

[14] http://www.wfmc.org/

information from the semantics of a process using different namespaces. Moreover, the BPMN 2.0 specification extends the scope and capabilities of the BPMN 1.2 in several areas:

- Formalizes the execution semantics for all BPMN elements.
- Defines an extensibility mechanism for both Process model extensions and graphical extensions.
- Refines event composition and correlation.
- Extends the definition of human interactions.
- Defines a Choreography model.

There was a great debate in the past few years on the difference between BPEL and XPDL and the difference in meaning between process diagram and process model. As it can be seen now, BPEL and XPDL are mainly used differently for different purposes. The BPEL representation of a diagram conveys the process semantics but it does not capture, as XPDL does, the precise shapes of the BPMN activities, gateways, and events, the bends in the arrows, etc. In other words, even if XPDL 2.1 has executable properties, it mainly captures the diagram. Instead, a BPEL description mainly captures the process semantics and it can be automatically executed because all of the execution details are part of the BPEL description process. But which of these two notations best represents the process model? It depends on which level of abstraction the designer is working on. XPDL contains some executable details attached to high level concepts (BPNM terms) while BPEL only contains low level concepts (e.g., port-types, schemas, XML transformations, etc). The argument over whether BPEL or XPDL is more “portable” is based on different interpretations of what “portable” means. If you mean the same process semantics can be executed on two different engines, then BPEL is more portable. If you mean that the same diagram can be created in two different tools, then XPDL. Which aspect of portability is more valuable? It depends on the final goal. If it is to glue together tool A and tool B, then XPDL has more flexibility. Otherwise, if the goal is the simulation or the execution a business process (even using different tools) and the graphical aspect of the process are not functional, then BPEL could be the right choice.

3.3 Workflow Orchestration and Choreography with BPMN 2.0

Modelling workflows requires defining sequences of tasks that can either be used to perform a single workflow or combined in more complex workflow structures. These structures are mainly indicated as Workflow *Orchestration* of Workflow *Choreography,* the exact definitions of which are still matter of intense debate[15]. Roughly speaking, we can consider the *Orchestration* as a composition of workflows from the point of view of participants and the *Choreography* as a composition of workflows from an external point of view, considering mainly the exchange of messages between different participants. This section will highlight the point of view of the BPMN specifications, which precisely defines both the structures. In fact, a recent BPMN capability, which is very important for the definition of complex and chained workflows, is the possibility to define choreography models. More precisely, BPMN 2.0 specifications define three basic types of sub-models within a BPMN model:

1. **Private (Internal) Business Processes** (O*rchestration*), which are the ones internal to a specific organization. These Processes have been generally called *workflow* or *BPM Processes* and are contained within a single *Pool* or workflow participant. The process flow is therefore contained within the *Pool* and cannot cross the boundaries of the *Pool*. The

[15] http://www.infoq.com/news/2008/09/Orchestration

flow of *Messages* can cross the *Pool* boundary to show the interactions that exist between separate *Orchestration*. Figure 4 shows an example of *Orchestration* sub-model which illustrates an internal procedure of the *Pool* "Forest action".

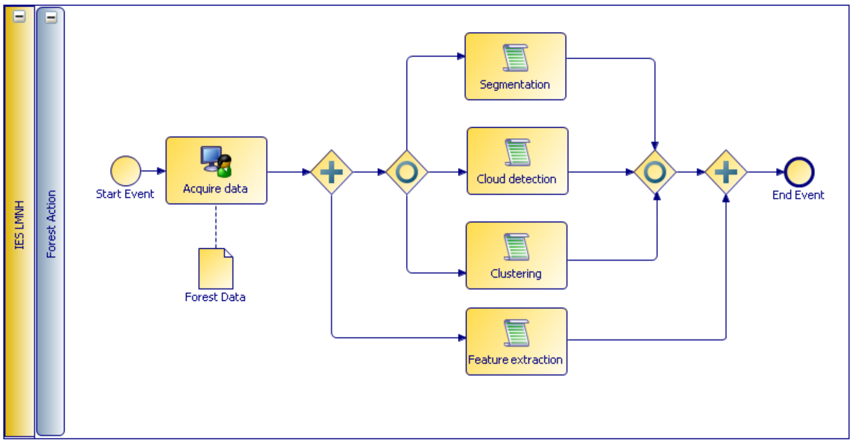

Figure 4 – Example of Orchestration diagram

2. **Abstract (public) processes** (*Choreographies*), which represent the interactions between a private Business Process and another process or *Pool* or *Participant*. Each step in the *Choreography* involves two or more *Participants* (these steps are called *Choreography Activities*). This means that the *Choreography*, in BPMN terms, is defined outside of any particular *Pool*. *Choreographies* are used to focus on *Messages* exchange rather than on processes internal behaviour. In this way, a *Choreography* diagram can be used to better address the goal of modelling advanced service chaining in the frame of distributed geo-processing services. Each service can be modelled as a separate *Orchestration* diagram, representing their interactions by mean of a *Choreography*. Figure 5, for example, defines the *Choreography* of the Forest map workflow depicted in Figure 2.

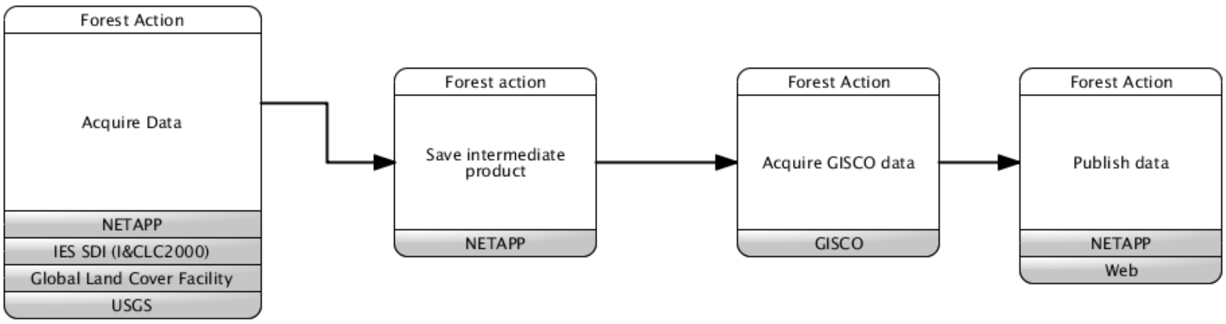

Figure 5 – Example of Choreography diagram

3. BPMN 2.0 defines a third type of representation diagram, namely **Collaboration**, which can include both *Orchestrations* and *Choreographies*. In Figure 2, for example, a *Collaboration* diagram to represent the workflow about the generation of the pan-European Forest Map was used. Moreover, *Collaboration*s can include also the view *Conversations* to sketch relationships among participants. For example, Figure 6 represents a *Conversation* between participants of the Forest Map workflow illustrated in Figure 2. Once the general picture representing relationships among different services is depicted using a

Conversation diagram, it can be further investigated and detailed by using a *Choreography* diagram.

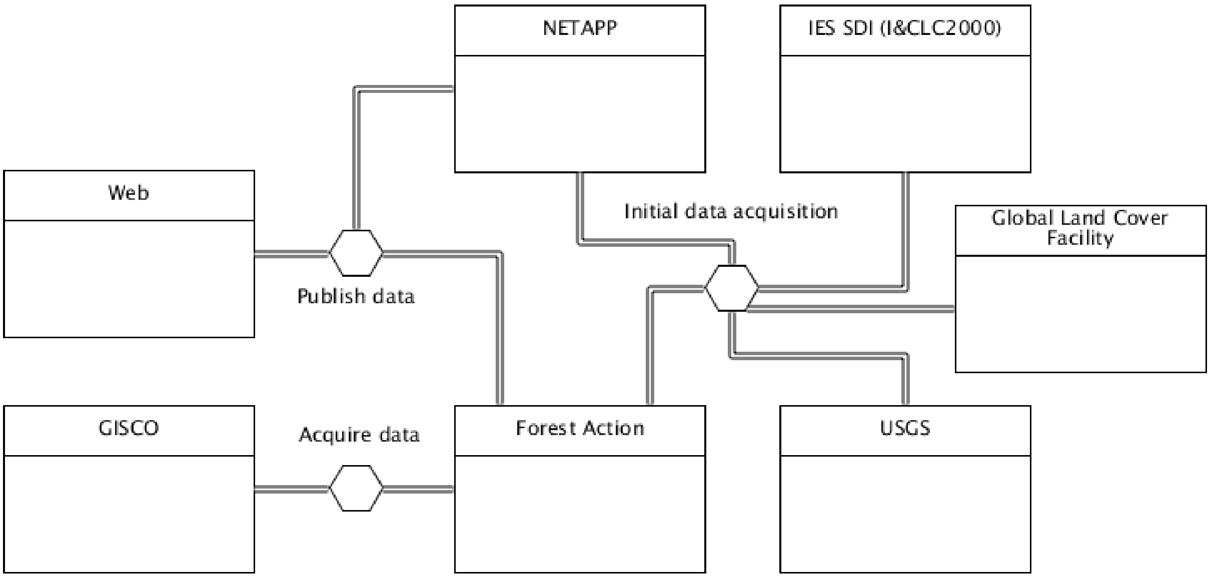

Figure 6 – Example of Conversational diagram

3.4 BPMN editing tools

The aim of this section is to give the EuroGEOSS partners an overview on the editing tools currently available and their characteristics, so that every user can choose among the solutions which best satisfy his/her requirements. The survey of the most widely used BPMN editors, preferably with simulation/validation support, is part of a recent internal activity performed at the SDI unit of JRC. Among them, the selection chose to evaluate the editors available free of charge (FLOSS – Free Libre Open Source Software – projects or at least software released under a free license) or with an evaluation version to try, even if with limited capabilities. Where only the BPMN editor, part of a suite, was provided for free, the characteristics have basically been evaluated on the free version available along with the textual descriptions of the features (white papers, fact sheets and so on) provided on web sites. Table 1 shows a summary of 17 different solutions. Some metrics have been then applied to the results in order to drive the choice of a preferred application to better fulfil the needs better. For a complete list of implementations and software supporting BPMN, visit http://www.bpmn.org (a list of more than 60 supporting implementation is provided) or http://www.businessprocessincubator.com (where it is possible to find a list of 60 different software). Apart from the criteria expressed above, also the possibility to have an integrated solution which connects the workflows definition modelled with BPMN to the UML definition of data structures was considered. Moreover, it is worth underlining some of the metrics used to choose the editor to be adopted.

- LICENCE: this would not be a real problem, even if there is a predefined budget; but whereas the features are the same, the cheaper the better.
- PLATFORMS: platform-independent solutions have an added value with respect to software platform-oriented.
- BPMN: a full compliance with BPMN 1.1/1.2 is requested, even if a full compliance to BPM 2.0 was considered, too.

- BPEL/WSDL: many active or past projects produced (or aim to produce in the future) code in BPEL, or WSDL,-compliant languages and formats; anyway, this condition is not the most constraining one at the moment.
- XPDL: XPDL could be used to share BPMN diagrams across different tools and platforms and thus this is a very important point.
- SIMULATION and VALIDATION: these features help users to check the validity and the behaviour of processes.
- REPOSITORY and COLLABORATION: these two features could help setting up a single and complete infrastructure without the need to use different applications during modelling phase.
- EASE TO USE and GUI: the user interface is fundamental, as well as to have a quick learning curve.

Table 2 represents a subset of the editors list provided in Table 1 obtained applying the considerations mentioned above.

Table 1: BPMN Editors Survey

Editors (link)	Version /Date	License			Platforms					Main Characteristics								
		Free/Mixed/ Commercial	MS Windows	Unix/Linux	Mac OS	Independent	Web-based	BPMN 1.1/1.2	BPMN 2.0	BPEL	XPDL	WSDL	Simulation	Validation	Repository (Not public)	Collaborative	Ease to use	GUI
ActiveVOS http://www.activevos.com/	7.1	C				X		P[M]	P[M]	Y[A]	Y	Y	Y	Y	Y	Y	2	3
ADONIS:CE http://www.boc-group.com/	3.90.1.98	M[1,2]	X					P[C]	N	Y	Y[H]	P[N]	Y	P	Y	N	2	2
BizAgi http://bizagi.com/	1.5.1.5	M[2]	X					Y	N	N	Y	N	Y[H]	Y	Y[I]	Y[J, H]	4	4
BONITA Studio http://www.bonitasoft.com/	5.1.1	F				X[a]		P[C]	N	N	Y	N	Y	Y	P[G]	Y	4	4
Cuecent BPMN Modeller http://www.cuecent.com/cuecent_bpmn.html	1.0	M[1,2]	X[a]					Y	N	N	Y	N	Y	Y	Y	Y	2	3
eKuar http://www.ekuar.com/	2.0	M[2]					X	Y	N	N	Y	N	Y	Y	Y	Y	4	4
GLOBAL 360 AnalystView http://www.global360.com/	3.0	C	X[b]					Y	N	N	Y	P	Y	Y	Y	Y	4	3
IDS SCHEER – ARIS Express [vi] http://www.ariscommunity.com/	2.0/7.1.0	M[2]				X[d]		N	Y	N	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	3	3
Intalio BPMN Designer CE + Server	6.0.3.015	F[1]				X[a]		P[F]	N	Y	N	Y	Y	Y	Y	Y	3	3
ITP Process Modeller http://www.itp-commerce.com/	5.2237	M[1,2]	X[b]					Y	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	Y[H]	4	3
ORACLE BPM Studio / Suite [iii] http://www.oracle.com/technology/products/bp	10.3.1.0.0	M[2]	X	X	X			P[C]	N	Y[H]	Y	Y	Y	Y	Y	P	3	3
SAVVION Process Modeller http://www.savvion.com/model_xl	7.5 SP1	M[1,2]	X[a]					P[C]	N	N	Y	Y	Y	Y	Y	Y	4	4
Signavio Academic / Oryx [i] http://oryx-project.org/	May 2010	M[2]					X	Y	Y	Y	Y	N	Y	Y	Y	Y	4	4
Sketchpad [v] http://sketchpadbpmn.sourceforge.net	SVN rev 91	F				X[d]		Y	Y	N	Y	N	N	P[D]	N	N	2	3
Soyatec eBPMN designer http://www.soyatec.com/ebpmn/	1.0.2	M[2]				X[a]		P[E]	N	N	N	N	N	P[L]	N	N	2	3
SPARX Enterprise Architect [ii] http://www.sparxsystems.com.au/	7.5.844	C	X	X[c]	X[c]			P[C]	N	Y	N	Y	N	Y	Y	Y	2	3
StarPound [vii] http://www.starpound.net/	1.2.0GA 30.0.011	F	X[a]			X[d]		P	N	P	N	Y	N	P	N	P	2	3
TIBCO Business Studio [ii] http://developer.tibco.com/business_studio/	3.2.0	C	X	X				Y	N	P[K]	Y	Y	Y	Y	Y	Y	3	4
Trisotech Visio Stencil [iv] http://www.businessprocessincubator.com/	4.2.4/1.2.4	F[1]	X[b]					Y	Y	N	P[B]	N	N	P[B]	N	N	4	3
Visual Paradigm BP-VA [ii] http://www.visual-paradigm.com/product/bpva	3.2	C				X		Y	Y	Y	Y	Y	Y	Y	Y	Y	4	4
Together Workflow Editor (JaWE)[O] http://www.together.at/prod/workflo	4.0.1	F				x		Y	Y		Y 2.		Y	Y	Y		4	4

Notes:	
- Numerical values are between 1 (worst) and 5 (best) - Some cells have values like Y(es), N(o), P(artially) 1) Free registration required 2) Only the BPMN editor is free, the rest is part of a commercial solution a) Based on Eclipse framework b) MS Visio is required c) Crossover compatibility layer by CodeWeavers is required d) It is a JAVA product i) Oryx is a free product, while the Signavio is the commercial solution with a free access reserved for Universities (Academic version) ii) It supports also UML 2 iii) The BPM Studio is part of the whole Suite iv) It supports also UML 2 as a basic stencil of the tool v) Originally developed by Global 360, now a Sourceforge hosted project vi) It supports also UML2 in its commercial version vii) The CORE is a pure JAVA product, while the Studio is available both for Windows and as an Eclipse plugin	A) With BPEL4People support B) Validation is available online after a free-of-charge registration C) Supports only a core subset of BPMN 1.1/1.2 D) Validation available online on the XPDL Validator webpage E) Full support of BPMN 1.0 and only partial support of BPMN 1.2 F) Some constraints on elements positioning and use G) Many connectors are present (JDBC, SAP, DB, …) H) Available in the full version I) Using Microsoft Sharepoint J) Using a Wiki K) Using a plug-in L) Just some hints and constraints during design activity M) Almost complete BPMN 2.0 support but unclear mapping to BPEL 2.0. From the Supported Standard Technical Note: "BPMN is supported with an internal model that allows executable BPEL". N) Not automatically created O) A Together Workflow Server (Enhydra Shark (TM)) also exists to be used as a workflow engine associated with JaWE http://www.together.at/prod/workflow/tws

Table 2: Selected tools

Editors (link)	License			Platforms					Main Characteristics										
	Free	Commercial	Mixed	MS Windows	Unix/Linux	Mac OS	Independent	Web-based	BPMN 1.1/1.2	BPMN 2.0	BPEL	XPDL	WSDL	Simulation	Validation	Repository	Collaborative	Ease to use	GUI
ORACLE BPM Studio / Suite [iii] http://www.oracle.com/technology/products/bpm			X[2]	X	X	X			P[C]	N	Y[H]	Y	?	Y	Y	?	?	3	3
SPARX Enterprise Architect [ii] http://www.sparxsystems.com.au/		X		X	X[c]	X[c]			P[C]	N	Y	N	Y	N	Y	Y	Y	2	3
TIBCO Business Studio [ii] http://developer.tibco.com/business_studio/	X			X	X				Y	N	P[K]	Y	Y	Y	Y	Y	Y	3	4
Trisotech Visio Stencil [iv] http://www.businessprocessincubator.com/	X[1]			X[b]					Y	Y	N	N	N	N	P[B]	N	N	4	3

In the end, it is possible to select two software from the survey. Currently, the best free software that satisfy the internal JRC requirements, is the TIBCO Business Studio, while the most suitable commercial solution is the SPARX Enterprise Architect. Considering the few differences in features between them, the final choice could be oriented toward the TIBCO Business Studio. Other partners in EuroGEOSS may come to different conclusions based on their own requirements.

4 PUBLISHING AND CHAINING GEOGRAPHIC SERVICES

This section illustrates specific solutions to publish and chaining distributed geographic Web Services. We first illustrate a methodology which assists users to deploy and publish environmental data and products. Then, we list some proposed solutions to chain geographic services published in the web and, finally, we illustrate the composition framework proposed by the FP7 UncertWeb project.

4.1 Assisted content web deployment and publication

One of the issues in environmental integrated modelling (flood predictions, forest fires simulation, etc.) based on existing resources (data, services, etc.) available on geographic information infrastructures is the lack of mechanisms to assist thematic users and experts in making environmental data and products persistent in a distributed and standard manner rather than storing such content locally and often isolated from other users. Diaz et al. (2011) proposed to extend current geospatial information infrastructures with publication services that involve actively users in content provision. This component provides EuroGEOSS/GEOSS users an integrated service with functionality to perform publication actions over the EuroGEOSS services, components, and data sources, and then increasing the availability of content to exploiting the operating capacity of EuroGEOSS. The realization of the component is called GEOSS Service Factory (GSF). The GSF allows users to convert smoothly new content in existing standard service instances. It prepares a channel that hides complexity and facilitates content sharing, while remaining loyal to the geospatial initiatives agreements and standards to reach the required level of interoperability.

Technically, GSF is internally aligned to the Abstract Factory design pattern (Gamma et al, 1995) that makes it flexible and scalable. Externally, the GSF exposes the OGC WPS interface to promote interoperability with existing WPS-aware clients. For instance, temporal process results may be chained with the GSF process to persist them as standard geospatial services. The GSF is designed to be implemented as a service component with a standard interface to be re-used in different scenarios. Since WPS is used to reach processing interoperability, it was chosen as the standard to implement the GSF as a service. Offering a unique interface, GSF implements the functionality to publish different content types in the EuroGEOSS system. The content can differ in nature, like data content (raw data, processed data, discovered data), processing content, etc. Each content type will be published according the publishing policies of the concrete system. The content is published in the system to be available for different purposes, like visualization, download or discovery, GSF delegates the final publication to each concrete factory according to these purposes. These factories offer the functionality to publish data content into standards-based data services, according to INSPIRE Implementing Rules, such as WMS, WFS, and WCS.

4.2 Chaining and executing geo-processes

Once they are published on the web, the full potential of web services as an integration framework will be achieved only when applications and business processes are able to integrate their complex interactions by using a standard process integration approach. This section will investigate advanced geo-service chaining methods, i.e. different techniques and approaches to combine of geo-web services. ISO (ISO119, 2005), distinguishes between three main chaining approaches:

- *User defined (transparent) chaining***.** The human user defines and controls the order of execution of the individual services. The user discovers and evaluates the available services, determines their fitness to the need, determines a valid sequence of services, and controls the chaining. As the details of the services are not hidden from the user this pattern is also called *transparent chaining*. In this scenario, no explicit description of the service chain exists.

- *Workflow-managed (translucent) chaining*. The human user invokes a service that controls the chain (called *Workflow Management Service* in ISO 19119). The execution of the chain is managed by a workflow service (or multiple workflow services). The user's involvement in the steps of the chain is mostly one of watching the chain execute the individual services that are apparent to the user, hence the alias of translucent chaining. A key distinction for this pattern is the existence of a defined chain prior to the user executing the pattern. He may also need to provide parameters particular to the specific instance, but relies on the workflow service to carry out the chain. This predefined chain is assumed to have a degree of semantic validity. It is not specified whether this means an explicit *description of the chain* in some workflow language (like BPEL) that can be executed by a (loosely coupled) workflow service or a *deployed instance of the chain* that is tightly coupled to a specific instance of a workflow service (and hence does not require an explicit description).

- *Aggregate (opaque) service.* In this pattern the services appear as a single service which handles all coordination of the individual services that are part of the chain. The user may or may not be aware that there is a set of services behind the aggregate service but has no possibility to watch the execution of the individual services (hence the alias of *opaque chaining*). As in the previous pattern, the existence of a pre-defined service chain, in the sense of a deployed instance of the chain, is assumed. In contrast to translucent service chaining, however, the user cannot observe the execution of the chain or access intermediate results.

The three main chaining approaches are used by different OGC initiatives.

Distributed geo-processes are mainly published on the web through the OGC Web Processing Service (WPS) specification[16]. The WPS specification defines the communication protocol to execute processes managing geographical resources. In particular, the WPS specification provides rules for standardizing inputs and outputs (requests and responses) for geospatial processing services, such as polygon overlay, features merging, map analysis and generation, etc. Technically, WPS has three interface operations: *GetCapabilities, DescribeProcess* and *Execute*. These interfaces define how a client can: *(i)* retrieve the list of processes available from a specific WPS server (*GetCapabilities* operation), (*ii*) retrieve detailed information about each supported process, such as structure, data types of input and output parameters (*DescribeProcess* Operation), and (*iii*) request the execution of a process (*Execute* operation). Notice that the WPS can represent a specific task in a workflow but also can be used to chain tasks (see below).

Implementation of geo-processing chaining services includes the activities developed by OGC within the "OGC Web Services Networks" initiatives. In particular, within the OWS-5 initiative, major achievements of OGC follow:

- The development of SOAP and WSDL interfaces for four foundation services: WMS, WFS-T, WCS-T, and WPS, allowing these services to be integrated into industry standard service chaining tools.

[16] http://www.opengeospatial.org/standards/wps

- The deployment of Service Implementations for WFS-T, WCS-T, WMS and WPS to demonstrate SOAP and WSDL binding patterns.
- The development of a BPEL script for SWE[17] Geo-Referenceable workflow.
- The demonstration of a workflow supporting EO Wildfire scenario within a SWE framework.
- A design of a Conflation workflow process and BPEL script to demonstrate service chaining and workflow, web processing services, and service interoperability using a variety of OGC service standards.

Also, within the OWS-6 initiatives, the Geo-Processing Workflow (GPW) thread aimed to develop and demonstrate interoperability among geo-processes through service chaining, workflow and web services, with emphasis on implementing security capabilities for OGC web services, including SWE services. Work in this thread was built on the results from previous test-beds, which includes authentication/authorization from OWS-4 and SOAP/WSDL recommendations from OWS-5. The results in the GPW thread were realized through a workflow scenario to demonstrate interoperability in a service-oriented architecture and RESTful architecture. The main topics for investigation and experimentation in the OWS-6 GPW thread were about:

- Security for OGC Web Services.
- Asynchronous Workflows & Security.
- Grid-Enabled Web Processing Service (WPS) profiles.
- GML Application Schema Development, Validation & ShapeChange Enhancements.

In parallel to the OGC initiatives, a number of solutions have been adopted to chain and execute geo-processing distributed services. In the following we will give an overview of some of the most used approaches.

- WPS-based solutions: The easiest method is when a WPS is seen as a black box performing a single process even if it implements a conceptual model which could be disaggregated in a series of combined/chained processes. In this case, three different possibilities are available to chain WPS geo-services:
    - The simplest way is when cascading HTTP GET operations is proposed, in which an input in one WPS execute request is the output from another WPS execute response. However, this method will make complex syntax if more than two WPSs are chained.
    - A WPS service can internally call some other WPSs or other kinds of Web Services[18]. Thus the workflow definition can be implemented on the server side codes. It can be imagined to be done in two different ways (Leibovici, 2010):
        - hard-coded workflow, in which the workflow definition is permanently programmed,
        - a workflow engine is used in the back-end.

- Workflow Web Service: A variant of the last solution could be when the workflow definition (a BPEL file, an XPDL file or any other format) is an input parameter of the WPS execute query. In this case, the WPS acts as a web workflow engine service, thus potentially accepting any workflow. This solution, called Workflow Web Service (WWS) and illustrated

[17] http://www.opengeospatial.org/projects/groups/sensorweb

[18] http://www.vector1media.com/component/content/article/75-feature/12431-geoprocessing-in-spatial-data-infrastructures

in (Leibovici, 2010), provides a solution for workflows as web services using the WPS framework.

- ORCHESTRA: the EU ORCHESTRA project (ORCHESTRA, 2007) uses several services distributed over the network that are orchestrated by an additional service. This service directly interacts with the client and executes the workflow defined by using *activeBPEL*, an Open Source BPEL engine[19]. In order to deploy the defined workflow, the Service Chain Access Service (SCAS), an ORCHESTRA service type which provides a means to create and delete service chain instances, was used. The creation of a new service chain instance, that can be invoked as a 'single' service accordingly with the opaque chaining defined in ISO 19119, is based on an explicit description of the workflow.
- Deployable WPS: a similar solution is proposed by (Shaffer, 2008). It extends the WPS specification by adding two new operations to the WPS interface (deploy/undeploy). The new proposal, namely WPS-T, is not limited to workflows only but presents a generic way of dynamically deploying and undeploying processes. Therefore, different kinds of WPS-T profiles can be developed to foster interoperability, including a specific BPEL profile. Although the ORCHESTRA project developed a similar service, this approach goes beyond the ORCHESTRA service specification by allowing both services and service chains to be deployed.
- GENESIS: the EU GENESIS project[20] also investigated the feasibility of using a high level notation for workflow definition (e.g. BPMN) and the automatic generation of the corresponding workflows in executable format (e.g. BPEL). One of the aims of the project in this field was also the dynamic generation of workflows by enabling on-the-fly service chaining by users (Mikulicz, 2010). This task was based on web service descriptions with semantic annotations and on an editing environment that proposes and checks consistency of service outputs and service inputs. Achievements delivered by the project included:
    - A proof of concept prototype of workflow generation tool based on high-level notation.
    - A proof of concept prototype of the workflows optimisation on-the-fly.
    - A generic uploading user interface (UI) and workflow monitoring.
    - A Survey on semantic annotation of processes by using the SA-WSDL notation.
- TAVERNA: a representative of many scientific workflow systems is also Taverna[21], an open source and domain-independent workflow management system, which has been created by the myGrid[22] team. The Taverna suite is written in Java and includes the Taverna Engine (used for enacting workflows) that powers both the Taverna Workbench (the desktop client application) and the Taverna Server (which allows remote execution of workflows). Although many examples of using Taverna lie in the bioinformatics domain, Taverna is actually domain independent because it can invoke generic WSDL-style Web services.
- KEPLER: The Kepler scientific workflow system (Altintas et al., 2004) is another system developed to execute scientific workflows. It provides domain scientists with an easy-to-use yet powerful system for capturing scientific workflows. Scientific workflows are a formalization of the ad-hoc process that a scientist may go through to get from raw data to publishable results. Kepler attempts to streamline the workflow creation and execution process so that scientists can design, execute, monitor, re-run, and communicate analytical procedures repeatedly with minimal effort. Kepler is unique in that it seamlessly combines

[19] http://ww.**activebpel**.org

[20] http://www.genesis-fp7.eu/index.php/dynamic-workflow

[21] http://www.taverna.org.uk/

[22] http://www.mygrid.org.uk/

high-level workflow design with execution and runtime interaction, access to local and remote data, and local and remote service invocation.

- GRID COMPUTING: Some recent solutions have been proposed by new technologies such as Grid Computing and Web 2.0 mashups[23] (Yu et al. 2008). The Grid computing community proposes an architecture (Hobona et al., 2010) that bridges web services based on the abstract geospatial architecture and the Open Grid Services Architecture (OGSA)[24]. A workflow management system, called SAW-GEO, supports orchestration of Grid-enabled geospatial web services. The implementation of SAW-GEO is based on both the Simple Conceptual Unified Flow Language (SCUFL)[25] and the BPEL.
- MASHUPS: Web services composition via Mashups have also been recently proposed by the Microsoft Composable Virtual Earth project[26]. Within the project, the MapCruncher tool was developed (Elson et al., 2007). It allows users to add custom raster overlays onto the existing road and aerial imagery provided by Virtual Earth or Google Maps. Overlays are typically detailed maps, such as a bicycle route map, building floor plan, or campus map. The resulting web site is an interactive web map that features both the user's maps and the standard imagery.
- BPMN 2.0, BPEL & XPDL Implementations: as illustrated in section 3.2, a BPMN diagram can be serialized into XML by using three main formats: BPEL, XPDL and BPMN 2.0. In the case XPDL, an update list of current implementations are available at the following web page, published by the WfMc: http://www.wfmc.org/xpdl-implementations.html. Available BPEL execution engine are listed at http://en.wikipedia.org/wiki/Comparison_of_BPEL_engines. BPMN 2.0 execution engines are starting to be proposed and published (see http://www.omg.org/bpmn/BPMN_Supporters.htm).

In the following section, we will present the "Composition as a Service" (CaaS) solution proposed by the FP7 EU UncertWeb project. The CaaS solution, due to its characteristics, which include the execution of BPMN workflows and the possibility to chain geographic service types, will be considered as a possible workflow chaining and executing solution to be tested for the EuroGEOSS project.

4.3 The UncertWeb CaaS

Recently, the FP7 UncertWeb project has been proposed a system architecture for the Uncertainty-enabled Model Web. In (Mazzetti et al., 2011) after the analysis of the requirements a system architecture is proposed for environmental model discovery and chaining. The document describe the design and the implementation of a distributed information system for creating, validating, editing, storing, publishing, and executing scientific workflows, referred to as CaaS system. The UncertWeb CaaS aims to provide added-value services on top of remote processing and access services. The CaaS is designed to address the following use cases:

UC1. Create a workflow
the user composes a workflow by means of a suitable editor.

---

[23] http://www.scribd.com/doc/4839186/Web-20-mashups-vs-Gridworkflows

[24] http://www.globus.org/ogsa/

[25] http://www.taverna.org.uk/developers/taverna-1-7-x/architecture/scufl/

[26] http://research.microsoft.com/jump/75291

UC2. Discover a workflow
the user browses and examines a (possibly distributed) library of available workflows.

UC3. Discover a process
the user browses and examines a (possibly distributed) library of available processes.

UC4. Execute a workflow
the user launches the execution of a selected workflow, e.g. a model run.

UC5. Store a workflow
the user saves the workflow to a local library, possibly publishing the workflow to the other users.

The CaaS publishes a WPS interface to execute the workflows that are described using the BPMN 2.0 formalism. A BPMN 2.0 file can be uploaded to the CaaS that will publish it as a single WPS process.

In order to enable the execution of BPMN files, the CaaS makes use of a set of conventions to bind a BPMN Task to a particular service type. Among others, the CaaS will support the following service types:

- OGC WPS
- OGC WCS
- OGC WFS
- Uncertainty Transformation Service by University of Munster

The complete list of supported service types will be released in September 2011 together with the conventions needed to execute a BPMN using the CaaS. A BPMN respecting such conventions is runnable by the CaaS, that provides the necessary mediation modules in order to execute the single tasks composing the workflow. As workflow engine, the CaaS will make use of the jBPM engine.

5 CONCLUSIONS AND NEXT STEPS

One of the main aims of the EuroGEOSS project is to investigate advanced service chaining methods and mechanisms for modelling workflows of distributed geo-processing services. This interim report investigated on some preliminary issues on how to model and implement workflows based on distributed geo-processes. In the first part of the document some of weaknesses and gaps of existing methodologies to illustrate workflows descriptions were identified. The analysis was based on the description of the use cases proposed by the different EuroGEOSS partners. After the identification of the gaps, three different workflow design approaches were discussed. In the central part of the document, the Business Process Modelling Notation (BPMN), one of the most adopted workflow modelling solution, was presented and discussed. In particular, the following issues were presented. :

- A brief description of the BPMN main elements
- A practical example on how to model a pan-European forest by using BPMN.

- Approaches and mechanisms for publishing workflows in catalogues and then reusing them in new model workflows through the description of two of the most used serialization solution for BPMN, namely XPDL and BPEL.
- Complex workflow orchestration and choreography methods, recently adopted by the BPMN 2.0 specifications .
- A recent survey on workflows editing and publishing tool, which let the users to model, describe and sometimes, depending on the tool, execute workflows.

In the last part of the document the focus was on how geo-processes have been represented and published on the web. Specifically, after an introduction on the OGC WPS specification, a discussion on a recent methodology developed within the EuroGEOSS project, which proposes a mechanism to assist users in the publication of geo-processes, was given. Moreover, a selection of workflow applications, implemented for the management of distributed geo-processes, was discussed.

Next steps of the project will include the following activities:

- The collection and the analysis of the workflow models and geo-processes proposed by each EuroGEOSS thematic area (Forest, Biodiversity and Drought).
- The investigation on how geo-processing and service chaining technologies can be improved regarding performance and reliability.
- The assessment of the CaaS solution proposed by the UncertWeb Project.

Acknowledgments

We thank all the EuroGEOSS partners and external contributors which helped us with the preparation of this report, and in particular:

- The GENESIS partners and Michele Chinosi (JRC) for the discussion and the preparation of the section related to BPMN.
- Michele Chinosi and Nicole Outslander (JRC) for the preparation of the section related to the survey of the BPMN editing tools.
- Carlos Granell (JRC) and Laura Diaz (UJI) for the discussion and preparation of the section related to the "assisted content web deployment and publication".
- Paolo Mazzetti of the UncertWeb project for the discussion on the CaaS.

REFERENCES

Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., Mock, S., 2004. *Kepler: an extensible system for design and execution of scientific workflows*. Proceedings of the 16th International Conference on Scientific and Statistical Database Management. Pages: 423--424.

Barker, A., Van Hemert, J. I., 2007. *Scientific Workflow: A Survey and Research Directions*. In R. Wyrzykowski, J. Dongarra, K. Karczewski, & J. Wasniewski (Eds.), *PPAM* (Vol. 4967, pp. 746-753). Springer.

Chinosi, M., Trombetta, A., 2011. *BPMN: An introduction to the standard*. Computer Standards & Interfaces, In Press, Accepted Manuscript, Available online 13 June 2011: http://www.sciencedirect.com/science/article/pii/S0920548911000766

Díaz, L., Granell C., Fonts, O., Gil J., 2010. *D.2.6.1. Report on Web 2.0 and alternative search*

mechanism. EuroGEOSS.
Available at: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_2_6_1.pdf

Elson, J., Howell, J., Douceur, J.R., 2007. *MapCruncher: integrating the world's geographic information*. Operating Systems Review.

Figueiredo, C., Gaigalas, G., Achard, F., Rubio Iglesias, J., M., 2009. *D.3.1: report on user requirements for the EuroGEOSS Forestry operating capacity*. EuroGEOSS.
Available at: http://www.eurogeoss.eu/Documents/EuroGEOSS_D3_1.pdf

Figueiredo, C., Revez, G., Gaigalas, G., Achard, F., Rubio Iglesias, J., M., Whitemore, C., Giovando, C., 2009. *D.3.2: Design Specifications for EuroGEOSS Forestry Components and Interfaces*. EuroGEOSS.
Available at: http://www.eurogeoss.eu/Documents/EuroGEOSS_D3_2.pdf

Gamma, E., Helm, R., Johson, R., Vlissides, J., 1995. *Design Patterns: Elements of Reusable Object- Oriented Software*. Addison Wesley, P.87.

Hobona, G.; Fairbairn, D.; Hiden, H.; James, P. , 2010. *Orchestration of Grid-Enabled Geospatial Web Services in Geoscientific Workflows*. Automation Science and Engineering, IEEE Transactions on , vol.7, no.2, pp.407-411.

Hofer, B., Niemeyer, S., 2009. *D.5.1: report on user requirements for the EuroGEOSS Drought operating capacity*. EuroGEOSS.
Available at: http://www.eurogeoss.eu/Documents/EuroGEOSS_D_5_1.pdf

Hollingsworth, D., 1995. *The Workflow Reference Model*. Workflow Management Coalition. Document Number tc00-1003.

Jones, R, Bastin, L, Cornford, D and Williams, M., 2010. *Handling and communicating uncertainty in chained geospatial Web Services*. Spatial Accuracy, Leicester, UK.

Leibovici, D., Pourabdollah, A., 2011. *D2.4.1: Interim Report on Integrated Multiscale and Multisource Modelling*. EuroGEOSS. To be published.

Leinenweber, L., 2009. *OGC OWS-6 GeoProcessing Workflow (GPW) Thread Summary Engineering Report*. OGC.

Lu R., Sadiq S., 2007. *A Survey of Comparative Business Process Modeling Approaches*. In Proceedings 10th International Conference on Business Information Systems (BIS), Lecture Notes in Computer Science, Volume 4439/2007, 82-94.

Mazzetti, P., Nativi S., Cornford, D., Pebesma, E., 2011. *UncertWeb, The Uncertainty Enabled Model Web, Deliverable 2.1-User Analysis*. To be published.

Mikulicz, H., 2010. *GENESIS D5700.1_3.0 - Report on dynamic workflow assessment activity*. GENESIS project.
Available at: http://www.genesis-fp7.eu/images/publications/newtechnologies/genesis_d5700.1_orcl_report_on_dynamic_workflow_assessment_activity_2.00_public.pdf

ORCHESTRA, 2007. *Service Chain Access Service Specification 0.4*. Available at: http://www.eu-orchestra.org/docs/OA-Specs/Service_Chain_Access_Service_Specification_v0.4-JRC-IES.pdf

Pekkarinen, A., Reithmaier, L., Strobl, P., 2009. *Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data*. ISPRS Journal of Photogrammetry and Remote Sensing 64: 171-183.

Schaffer, B., (2008). *Towards a transactional Web Processing Service (WPS-T).*In: Pebesma, E., Bishr, M. & T. Bartoschek (Eds.): GI-Days 2008 - Proceedings of the 6th Geographic Information Days. IfGIprints Nr. 32, Institut für Geoinformatik, Münster.

Silver, B., 2010. *BPMN Method and Style*. Cody-Cassidy press.

Taylor, I.J., Deelman, E., Gannon, D.B., 2007. *Workflows for e-Science: Scientific Workflows for Grids*. Springer-Verlag.

Tuama, É.Ó., Dubois, G., Cottam, A., May, I., Fisher, I., 2009. *D.4.1: report on user requirements for the EuroGEOSS biodiversity operating capacity*, EuroGEOSS.

Available at: http://www.eurogeoss.eu/Documents/EuroGEOSS_D4_1.pdf

XPDL WfMC, 2008. *Workflow Process Defiition interface - XML Process Definition Language (XPDL)*. Workflow Management Coalition, Document WfMC-TC-1025. Available at: http://www.wfmc.org/index.php?option=com_docman&task=doc_download&itemid=72&gid=132.

Yu, J., Casati F., Benatallah, B., Daniel, F., 2008. *Understanding Mashup Development*, Internet Computing September/October 2008 (vol. 12 no. 5) pp. 44-52.