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.4.4 - Documentation of biodiversity analytical models and workflows

Title	Documentation of biodiversity analytical models and workflows
Creator	Didier Leibovici
Creation date	17/11/2010
Date of last revision	05/04/2011
Subject	Protocols, models and ontologies, WP4 biodiversity models
Status	□ Draft ☒ Final
Publisher	EuroGEOSS
Type	Text
Description	Describes the building blocks (protocols, models and ontologies) used to be able to derive and implements the biodiversity workflows in a service orientated architecture.
Contributor	Didier Leibovici (UNOTT), Amir Pourabdollah (UNOTT), Grégoire Dubois (EC-JRC), Jon Skoien (EC-JRC), Michael Schulz (EC-JRC), Eamonn O'Tuama (GBIF)
Format	Docx and pdf
Source	
Rights	□ Restricted ☒ Public
Identifier	EuroGEOSS_D4_4. pdf
Language	En
Relation	WP2 and other thematic WP
Coverage	

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

Contents

Figures

Tables

ACRONYMS AND ABBREVIATIONS

ABCD	Access to Biological Collections Data
AEGOS	African-European Georesources Observation System
AJAX	Asynchronous JavaScript And XML
AML	Arc Macro Language
APAAT	African Protected Areas Assessment Tool
BDP	Biological Data Profile
BioCASe	Biological Collections Access Service protocol
BPEL	Business Process Execution Language
BIPM	Bureau International des Poids et Mesures
BPMN	Business Process Model Notation
CIESIN	Center for International Earth Science Information Network
CMS	Convention on Migratory Species
CNR	Consiglio Nazionale delle Ricerche - Italy
CSW	Catalogue Service for the Web
DiGIR	Distributed Generic Information Retrieval
DOPA	Digital Observatory for Protected Areas
EBA	Endemic Bird Area
EC-JRC	European Commission-Joint Research Centre
EML	Ecological Metadata Language
ENM	Ecological Niche Model
FGDC	Federal Geographic Data Committee
GBIF	Global Biodiversity Information Facility
GDAL	Geospatial Data Abstraction Library
GEOSS	Global Earth Observation System of Systems
GIS	Geographic Information System
GML	Geography Markup Language
GNA	Global Names Architecture
GROMS	Global Register of Migratory Species
GUM	Guide to the expression of Uncertainty in Measurement
HRI	Habitat Irreplaceability Index
HSI	Habitat Similarity Index
HTML	HyperText Markup Language
IBA	Important Bird Area
IOC	Initial Operating Capacity
IUCN	International Union for the Conservation of Nature
ISO	International Organization for Standardization
KML	Keyhole Markup Language
LGTG	LSID GUID Task Group
MDG	Millennium Development Goals
MIFTG	Metadata Implementation Framework Task Group
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index

NDWI	Normalised Difference Water Index
OGC	Open Geospatial Consortium
PA	Protected Area
PoHS	Probability of Habitat Similarity
RSPB	Royal Society for the Protection of Birds
SOS	Sensor Observation Service
SRI	Species Irreplaceability Index
TAPIR	TDWG Access Protocol for Information Retrieval
TDWG	Taxonomic Database Working Group
UNEP-WCMC	United Nations Environment Programme - World Conservation Monitoring Centre
UN-MDG	United Nations - Millennium Development Goals
USGS	U.S. Geological Survey
VMAP	Vector Map
WBDB	World Biodiversity Database
WCS	Web Coverage Service
WCS-T	Web Coverage Service - Transaction
WDPA	World Database of Protected Areas
WfMC	Workflow Management Coalition
WFS	Web Feature Service
WFS-T	Web Feature Service - Transaction
WMS	Web Map Service
WPS	Web Processing Service
WRI	World Resources Institute
WWF	World Wildlife Fund
XML	eXtensible Stylesheet Language
XPDL	XML Process Definition Language

PROJECTS RELEVANT TO THIS REPORT

Term	Definition
EUROGEOSS http://www.eurogeoss.eu/	EuroGEOSS is a large scale integrated project in the Seventh Framework programme of the European Commission. It is part of the thematic area: "ENV.2008.4.1.1.1: European environment Earth observation system supporting INSPIRE and compatible with GEOSS" .(DG RTD) EuroGEOSS demonstrates the added value to the scientific community and society of making existing systems and applications interoperable and used within the GEOSS and INSPIRE frameworks.
UncertWEB http://www.uncertweb.org/	UncertWeb is a small or medium scale focused research project (STREP) in the Seventh Framework programme of the European Commission. The project is contributing to the Objective ICT-2009.6.4 ICT for Environmental Services and Climate Change Adaptation (DG INFSO) UncertWeb is about how to compose workflows of resources within a GEOSS or web service context accounting for uncertainty.
GENESIS http://www.genesis-fp7.eu/	Genesis stands for Generic European Sustainable Information Space for the Environment. The objective is to provide those involved in environment management and health services in Europe with an efficient, web-based solution for monitoring air quality, fresh and coastal water quality and their impacts on health. The advanced, ICT-based solution that will result from this research and development will combine open, collaborative information networks while integrating systems that already exist in Europe

1 SCOPE

This document complements the deliverable D4.1 (2009) about the use scenarios described for this thematic. It aims at identifying and describing the protocols, models and ontologies of the exemplar biodiversity models (research questions) and defines the workflows to be implemented in service oriented architecture for the Advanced Operating capacity.

2 INTRODUCTION

The African Protected Areas Assessment Tool (APAAT), developed within the Joint Research Centre by MONDE, is an online information system based on a Geographic Information System (GIS) and satellite-derived data developed to aid decision makers assess the state and pressure of 741 protected areas (PAs) in Africa. It can be considered as the first consistent, continent-wide assessment of the state of protected areas in Africa. More information can be found at http://bioval.jrc.ec.europa.eu/APAAT/ and in Hartley et al. (2007).

In order to evolve towards an operational environmental monitoring and forecasting service, matching the needs from a broader community of end-users, the system needed to be able to capture more regularly the spatio-temporal environmental changes, implying that new means have to be put in place to ensure the automatic update of the information presented. It is the purpose of WP4 in EuroGEOSS to develop a biodiversity operating capacity conceptually similar to the APAAT as an interoperable web service capable of exchanging a large variety of data coming from different sources (Dubois et al., 2009, 2010). Coined the Digital Observatory for Protected Areas, or DOPA (http://dopa.jrc.ec.europa.eu/), this operating capacity based on SOA is expected to easily bring together the large variety of sensors, databases and systems. It is also conceived as a set of distributed databases and open, interoperable web services.

While the APAAT shows pre-processed information for each of the protected area it describes, DOPA will contain a minimum of such information, relying mainly on a number of key Web Processing Services (WPSs) to generate information on the fly and provide so end-users with greater flexibility. One example is the use of the Habitat irReplaceability Indicator (HRI) computed for each protected area. If the APAAT shows the HRI in the form of a map that is associated to each of the 741 African protected areas considered so far, the DOPA will compute such an indicator on the fly using a dedicated WPS designed to compute the likelihood of finding similar ecosystems in a given area. This WPS, called eHabitat, easily allows for the computation of the HRI at any location, including new areas to be protected and this anywhere on the globe. Another immediate benefice in relying on a set of fundamental web services is that these can be chained with other data/model providers and provide added functionalities at a very low cost. In the case of eHabitat, linking this WPS with a service providing climate change data allows for ecological forecasting in protected areas and for the assessment of temporal changes in the surfaces occupied by the ecosystems. In the same way, linked with a database of species data, the likelihood to find ecosystems suitable for the observed species allows for Ecological Niche Modelling (ENM).

Within EuroGEOSS, eHabitat is expected to encourage the multidisciplinary interoperability with the other thematic WPs by facilitating the derivation of new product datasets from existing data and scientific models (Wheeler 2006, GEO-AIP3 2011)). Facilities under developments, to be able to assess the quality of the workflow, *i.e.* the new product datasets, in term of uncertainties of the outputs are part of EuroGEOSS (Leibovici et al. 2009) and UncertWeb (Pebesma et al. 2010).

Figure 1 shows the Service Oriented Architecture of the DOPA with the eHabitat at the heart of its modelling capability: a set of distributed interoperable databases containing essential datasets (species range maps, species occurrences, boundaries of protected areas, ecological data) can be accessed through web clients. End-users, typically decision makers and researchers, can interact with the available data and compute a set of biodiversity indicators using modelling services like eHabitat. This architecture and set of services are expected to be used to assess and monitor the state and pressure of protected areas at the global scale (see Figure 1).

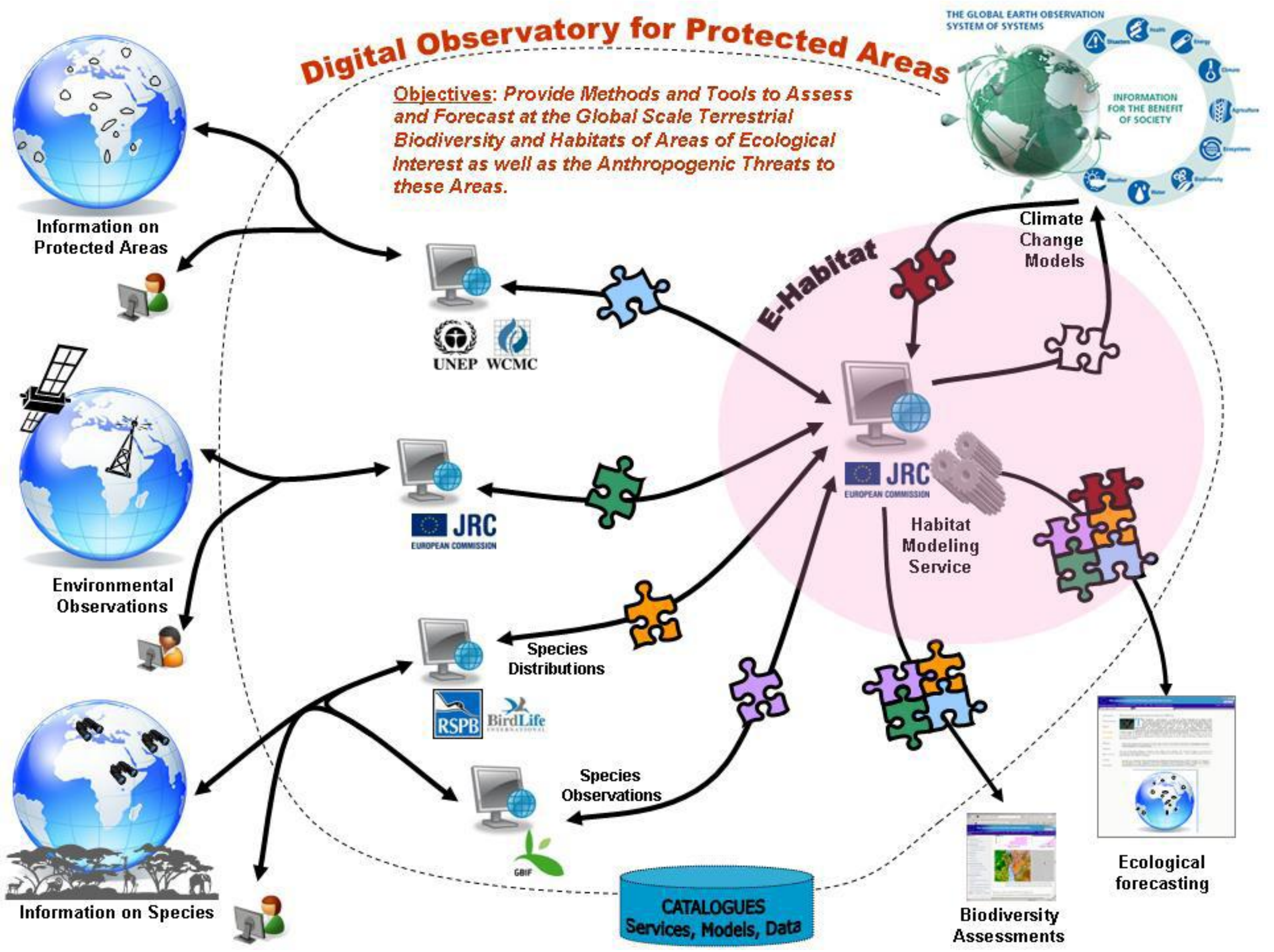
Digital Observatory for Protected Areas
Objectives: Provide Methods and Tools to Assess and Forecast at the Global Scale Terrestrial Biodiversity and Habitats of Areas of Ecological Interest as well as the Anthropogenic Threats to these Areas.
THE GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS
INFORMATION FOR THE BENEFIT OF SOCIETY
Information on Protected Areas
UNEP WCMC
E-Habitat
Climate Change Models
Environmental Observations
JRC EUROPEAN COMMISSION
JRC EUROPEAN COMMISSION
Habitat Modeling Service
Species Distributions
RSPB BirdLife
Species Observations
GBIF
Information on Species
CATALOGUES Services, Models, Data
Biodiversity Assessments
Ecological forecasting

Figure 1: The Digital Observatory for Protected Areas (DOPA)

DOPA is an interoperable web service version of the APAAT which will also allow interactions with a climate change model web service. (source D4.1)

2.1 Research scenarios for the biodiversity operating capacity

In order to implement the biodiversity operating capacity, the description of exemplar models within each thematic and across them needs to be formalised using as much as possible existing or reusable frameworks, protocols, and ontologies based on what has been already achieved.

Two main research scenarios for Work Package 4 have been adopted:

ParkSim: The APAAT is providing assessments for protected areas in the light of other protected areas of the continent and of the other protected areas of the country. The DOPA is aiming to further extend the use of protected areas to as well as to provide the tools and means to obtain an assessment based on biodiversity and threats to any area end-users might be interested in. The assessment of potential new protected areas by computing a number of key indicators for a simulated protected area. Figure 2 shows a graphical representation of indicators based on species and habitat irreplaceability (biodiversity indicators) and on anthropogenic pressures due to agriculture and population (threats indicators). Because these indicators are highlighting the relative value of a protected area against other protected areas from the same country and from the same ecoregion, all data are interlinked. Adding or removing a protected area will automatically affect the values of the indicators of all other related protected areas: removing a protected area will automatically increase the values of other protected areas from the same region while the addition of a new area would reduce the relative values of the existing protected areas. It is the purpose of the ParkSim scenario to allow end-users to simulate the changes to the indicators by adding or removing protected areas to existing sets and assess so the impact of such changes. Similarly, the species irreplaceability indicators are affected by the ranking of the species on the IUCN red list of threatened species that is regularly updated. This list is used to attribute an additional weight to threatened species, a weight that needs to be changed interactively by end-users willing to use a different ranking scheme.

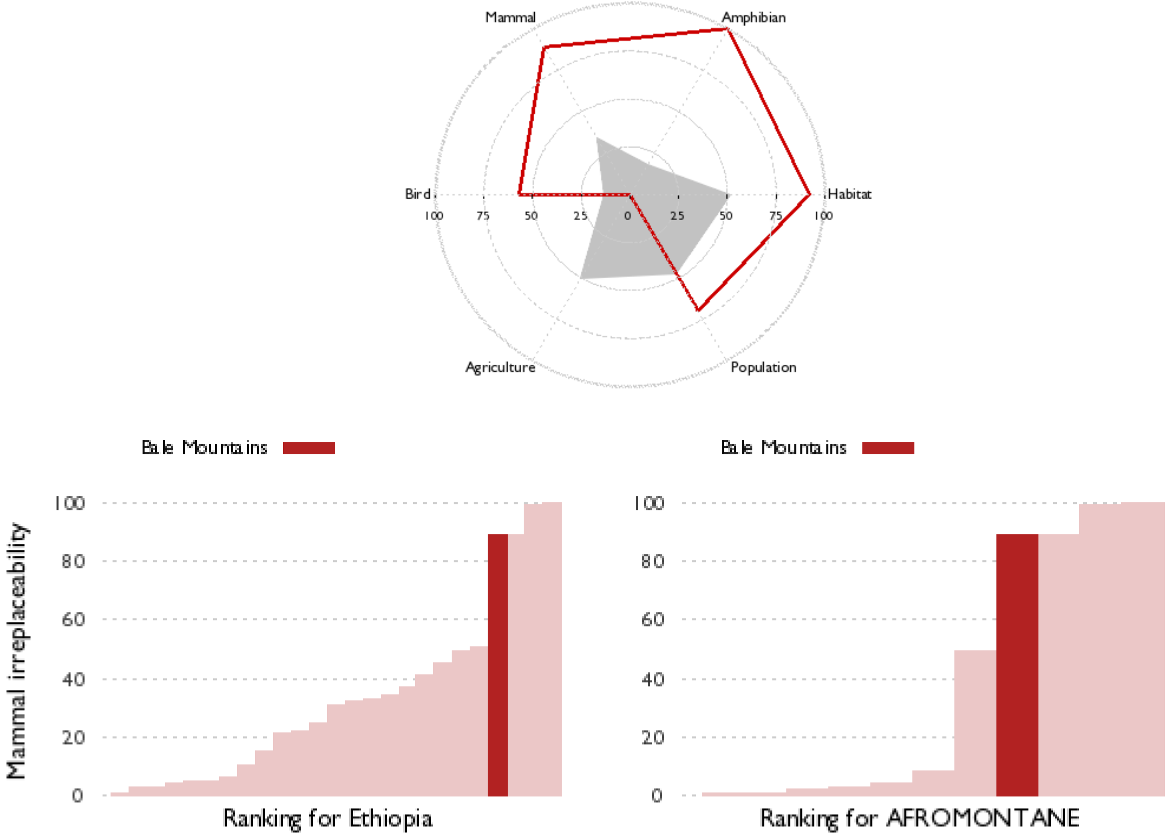

Figure 2: Irreplaceability indicators for protected areas.

The radar plot shows all six of the irreplaceability indicators for a protected area (here the Bale Mountain National park, Ethiopia) in red along with the country averages which are grey. Each indicator has been scaled from 0 (lowest) to 100 (highest) to allow easy comparison. The histograms show the ranking of the protected area for a single indicator (here mammal irreplaceability) against the same indicator of protected areas found in the same country (left) and in the same ecoregion (right).

eHabitat: This scenario demonstrates how the chaining of a basic WPS for computing likelihood of finding similar ecosystems with other interoperable services can dramatically improve the functionalities of an information system. By linking eHabitat WPS with services providing climate change data and/or species occurrences (i.e. GBIF), one can build a GEOSS-based infrastructure for decision makers to identify the ecosystems that are suitable for a given species (Ecological Niche Model) and/or assess the impact of climate change on these habitats, all essential information for assessing criticality of pressures on biodiversity (Figure 2).

The assessment of the impact of climate change on protected areas is a scenario that is part of a demonstration of the 3rd Phase of the GEOSS Architecture Implementation Pilot (AIP-3) as a contribution from FP7 EuroGEOSS project. It will also contribute to different approaches towards uncertainty propagation in workflow models: within EuroGEOSS (see T2.4) and within the FP7 project UncertWeb (Pebesma et al. 2010).

Based on the previous AIP-Phase2 experience, the EuroGEOSS and GENESIS projects will enhance the interoperability infrastructure with:

a) a discovery broker service which underpins semantics enabled queries: the EuroGEOSS/GENESIS Discovery Augmentation Component (DAC);
b) environmental modelling components (i.e. OGC WPS instances) implementing algorithms to predict evolution of PAs ecosystems;

If time allows for it, a workflow engine could be developed to:

i) browse semantic repositories;
ii) retrieve concepts of interest;
iii) search for resources (i.e. datasets and models) related to such concepts;
iv) execute WPS instances.

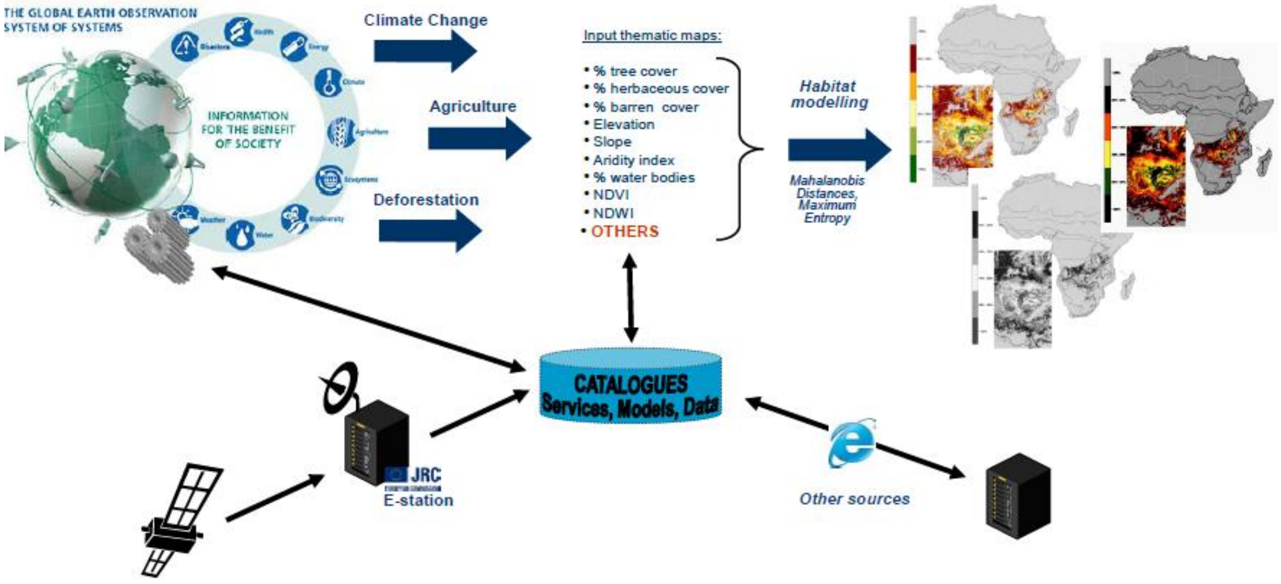

Figure 3: eHabitat model

eHabitat model principle to design a service where end-users can select the ingredients of the ecosystems they are interested in as well as the other modelling services capable of generating input data for modelling. Among the ingredients, climate change data and species occurrences can be used for climate change impact assessment and ecological niche modelling, respectively.

3 DOPA SPECIFIC MODELS

Here we call specific DOPA models the series of functionalities derived directly from the data infrastructure. eHabitat, as a more complex modelling entity, is thought as a DOPA component and is described in the next section.

Among the various services DOPA is expected to provide as a container of services for biodiversity management and assessments, we will, in the framework of EuroGEOSS focus on the gradual setting up of the following web services:

- Computation of species indicators (*e.g.*, variety, composition and irreplaceability)
- Computation of Indicators of the uniqueness of habitats (using eHabitat)

Outputs from the UncertWEB project (funded by DG INFSO) involving the use of eHabitat for assessing climate change impact on ecosystems will be further used to assess the vulnerability of protected areas to a variety of anthropogenic factors.

3.1 Interoperability components for the ParkSim scenario

The ParkSim scenario is described on Figure 3 as a workflow: succession of tasks. This Figure uses the BPMN (Business Process Model Notation, a standard for workflow drawing, maintained by the OMG, adopted by the WfMC) to represent workflows. For workflow notation and format see D2.3.1 and D2.4.1. In the particular workflow shown on Figure 3, the datasets involved are not represented but only the tasks of selecting them. This is due to the flexibility characteristic of this model (also present for the eHabitat model).

For interoperability purposes and implementation within an SOA framework, the workflows need to be fully described task by task. This can be done within the workflow editing tool (see D2.3.1); these fine descriptions linked to computational aspects of the workflow are then stored in the exchanging format XPDL or BPEL exported from the BPMN editor in an XML file.

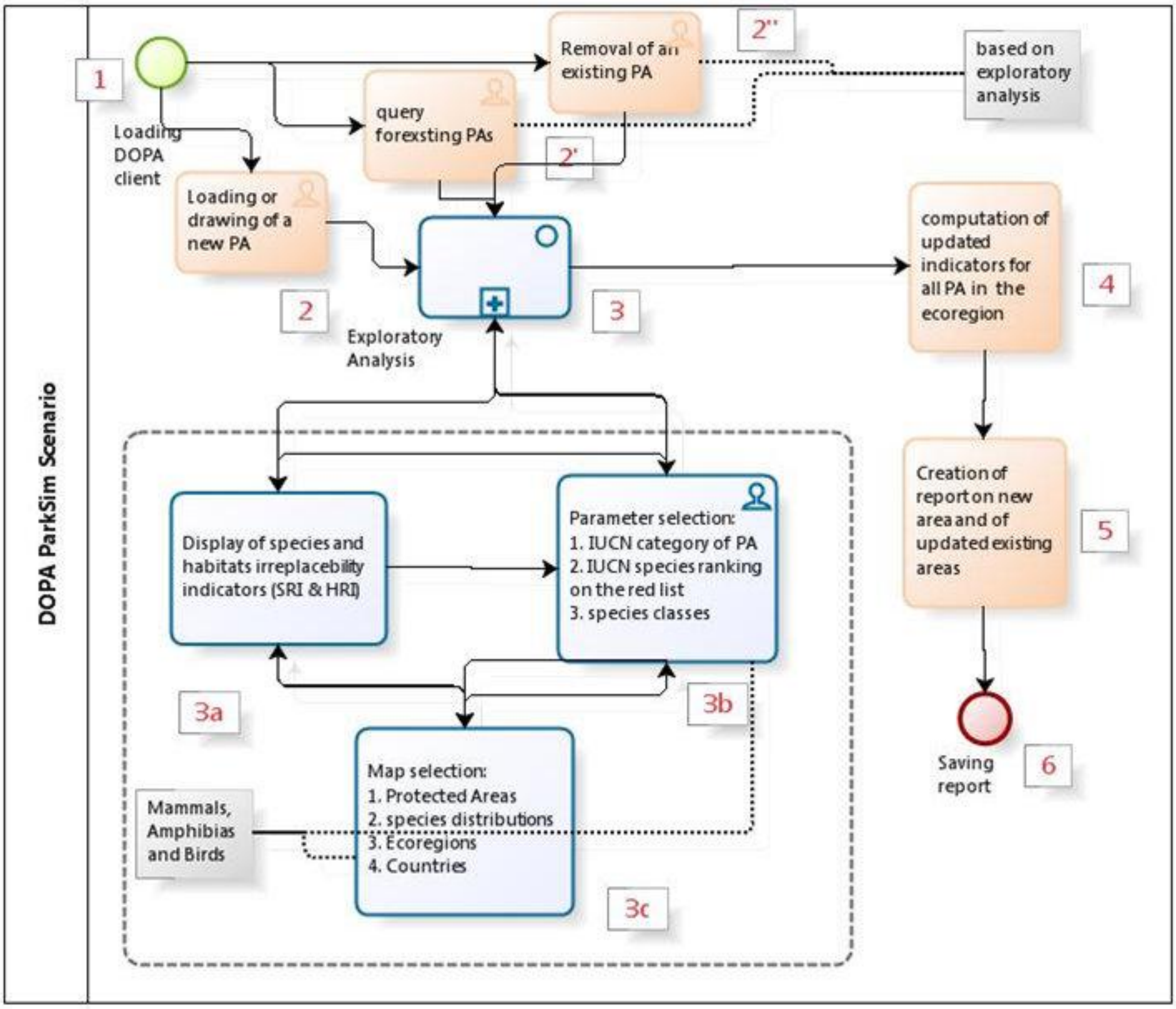
DOPA ParkSim Scenario
1
Loading DOPA client
Loading or drawing of a new PA
2
query forexsting PAs
2'
Removal of an existing PA
2''
based on exploratory analysis
Exploratory Analysis
3
computation of updated indicators for all PA in the ecoregion
4
Creation of report on new area and of updated existing areas
5
Saving report
6
Display of species and habitats irreplacebility indicators (SRI & HRI)
3a
Parameter selection:
1. IUCN category of PA
2. IUCN species ranking on the red list
3. species classes
3b
Map selection:
1. Protected Areas
2. species distributions
3. Ecoregions
4. Countries
3c
Mammals, Amphibias and Birds

Figure 4: The ParkSim scenario

The ParkSim scenario as seen by the DOPA system (the sub-process "Exploratory Analysis" as been embedded here to obtain only one diagram; numbered steps are further discussed in the text)

How to achieve these tasks within DOPA and EuroGEOSS SOA platform is described here:

Step/Task	**Description**	**[Protocol] involved and details**
1 start	- user locates the DOPA application/service and IOC, advanced features potentially require authentication	**[WWW/WMS]** EuroGEOSS/DOPA website
2 query	- user provides/loads the existing set of Protected Areas for its IOC	**[WFS]** server hosted within the DOPA system and discoverable using EuroGEOSS broker
2' query	- user provides/loads a new set of Protected Areas or modifies (add/change) the existing set of Protected Areas.	**[WFS]** server hosted within the DOPA system and discoverable using EuroGEOSS broker. The same as in step 2.
2'' query	- user provides/loads the existing set of Protected Areas and modifies it (remove)	**[WFS]** server hosted within the DOPA system and discoverable

		using EuroGEOSS broker. The same as in step 2.
3 exploring & setting		
3a	user activate the visualisation of the indicators. Indicators are automatically updated with changes of map selection and changes of parameters	**[WMS/WFS]** DOPA Web-Client
3b	selection of parameters affecting the indicators	**[WMS/WFS]** DOPA Web-Client
3c	selection of areas for which the indicators need to be computed for [Figure 5]	**[WMS/WFS]** DOPA Web-Client
4 geoprocessing	- with input from 2 processing service is triggered and outputs updated parameters and maps **[WPS/WCS]**	**[WPS]** server hosted within the DOPA system and discoverable by EuroGEOSS broker
5 processing	-a processing is activated to deliver a report **[interface linked to 2' or WPS]** - processing of finished results from 4 to deliver a report	**[WPS]** the WPS is under development, this task is at the present time part of the DOPA web interface calling a servlet ().
6 saving/exit	-storage of the report **[catalogue service]**	report should include changed set of Protected Areas and parameter settings to be re-runable. **[WFS-T/CSW]**

The parts in square brakets are the interoperable components needed to be existing in order to run the model: **[interoperable component]**. For most of them, they are related to web services which are already following specific standards.

This workflow, stored as an XPDL file could be running within a WPS or WWS (see D2.4.1) which would use the XPDL file as input of the service request. The WPS needs to run in the back end either a translation of the XPDL file into the executive language (such as BPEL) or run directly an XPDL workflow engine (e.g., Together Enhydra Shark (TM)). The choice or selection of parameters are either predetermined or provided as input of the service. This can be useful when testing different areas (delineated in 3b) in a more automatic way.

3.2 DOPA Web Client

DOPA's data exploration is facilitated through the development of a web client (Figure 5). A basic interface has been developed using ExtJS, a javascript development framework. The mapping functions have been set using OpenLayers. The API supports around 30 layer formats, that can be fetched, projected, processed using the API. It can also generate spatial vectorial structures on-the-fly like polygons, points markers etc. Openlayers also includes controls, triggers/events, popups, AJAX calls, DIV tag creation on the fly functionalities.

One important functionality is the interconnection between a WMS and WFS service that point to the same data and share the same name (OpenLayers.Protocol.WFS.fromWMS.Layer), allowing the loading of the 130 000 polygons of the WDPA instead of pushing a WFS-XML document of 700MB containing all the polygons. Using this architecture, it is possible to fetch the same information from a WMS and request specific polygon information using WFS filters.

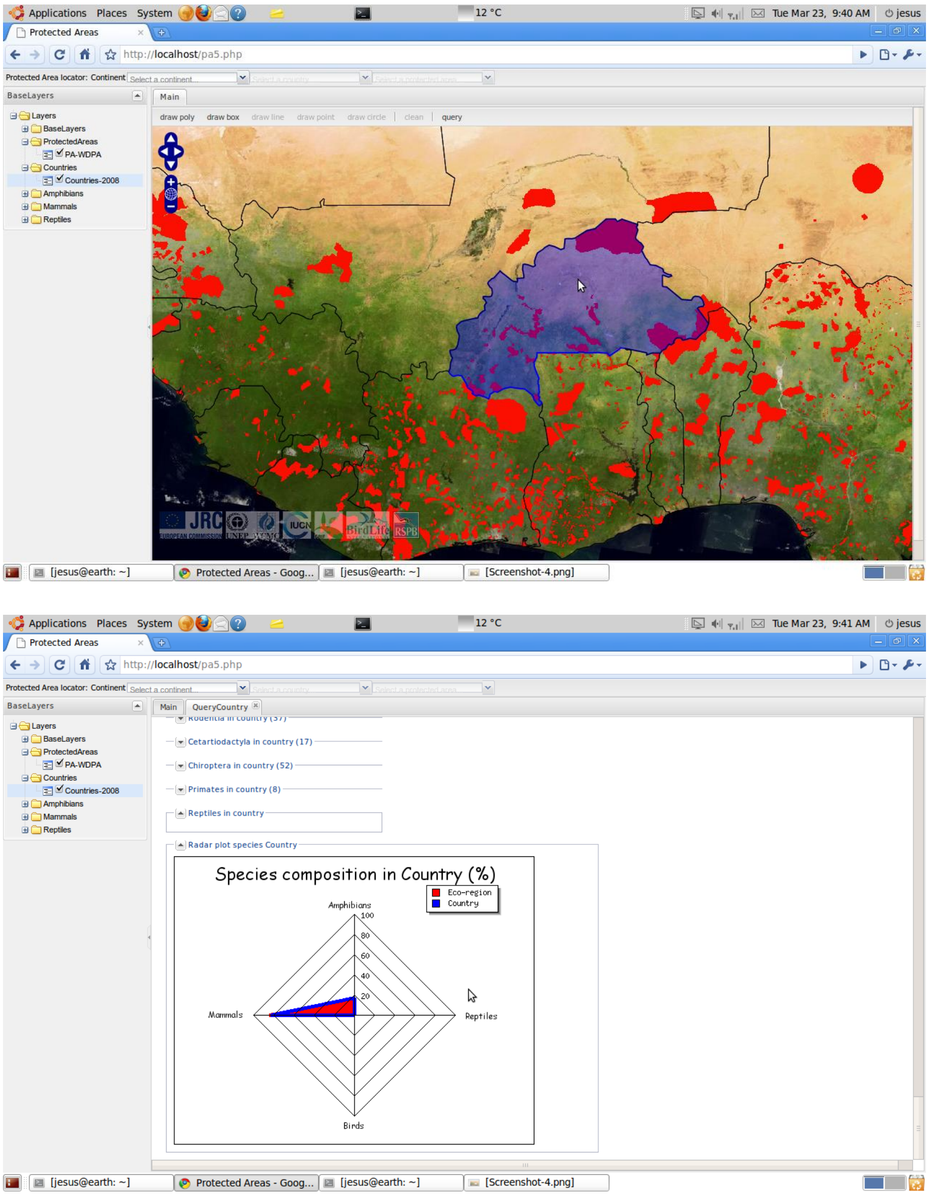

Figure 5: Example of a possible WebGUI for DOPA

Example of a possible WebGUI for DOPA: an area is selected for analysis triggering requests to the PostGIS database (top), of which results are summarised with diagrams in a new window (bottom).

3.3 Interoperability components for the eHabitat scenario

The eHabitat component of DOPA focuses on habitat as predictor for biodiversity. Within DOPA eHabitat provides protected areas (PA) management functionalities used as models:

(I) PA comparison model: assessing similarities in thematic maps

(II) Forecasting model: assessing changes in similarities in thematic maps when altering input data (typically using different forecasted climate data for climate change impact assessment)

The eHabitat models (I) and (II) use various thematic (environmental) map layers which are selected for modelling biomes, ecosystems or habitats using a multivariate dissimilarity measure based on the Mahalanobis distance (see e.g. Farber and Kadmon, 2003). Polygons are used to define a region of interest and the likelihood to find habitats that are similar to the one found in the polygon is computed outside of the polygon using the multivariate dissimilarity approach. A variant to minimum Mahalanobis distance criterion is the maximum entropy which is also implemented in eHabitat (see Figure 6). The generic workflow expressed in the BPMN form is given on Figure 5. For interoperability purposes and implementation within an SOA framework the workflows need to be fully described task by task.

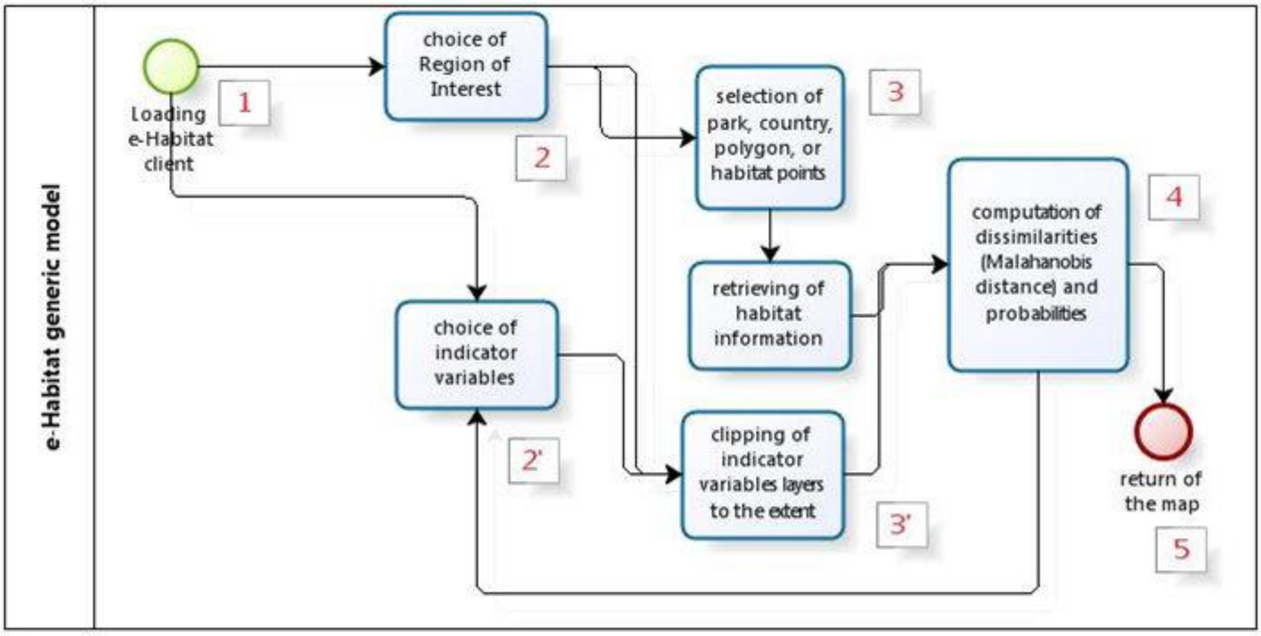
e-Habitat generic model
Loading e-Habitat client
1
choice of Region of Interest
2
selection of park, country, polygon, or habitat points
3
retrieving of habitat information
computation of dissimilarities (Malahanobis distance) and probabilities
4
choice of indicator variables
2'
clipping of indicator variables layers to the extent
3'
return of the map
5

Figure 6: eHabitat workflow

eHabitat general user-machine interaction scenario workflow (steps have been numbered for identification in the text

How to achieve these tasks within DOPA and EuroGEOSS SOA platform is described here:

Step/Task	Description	[Protocol] involved and details
1 start:	- user to locate the eHabitat applications via DOPA -potentially to log into it-	**[WWW]**
2 region filter:	- user delineates/selects a region of interest	**[WCS] [WFS-T]** using EuroGEOSS broker (WCSs), the DOPA server (a WFS)
2' data selection:	- query and/or selection of available indicators	**[WFS] [WFS-T]** host server within the DOPA system
3 data sub-filtering:	1.user selects a comparison 'profile'	**[WFS]** via the EuroGEOSS broker
3' geoprocessing:	- extent clipping of 2' onto 2	**[WPS]**
4 (geo)processing:	- core computation of dissimilarity criterion, and probabilities of similarity	**[WPS]** there are two tasks in 4 one retrieving information one computational, there can be existing in the one web service (a sub-workflow) or not
5 saving/exit:	- user visualises the returned map and make a decision to save and/or exit	**[WFS-T] [WCS-T]**

The parts in square brakets are the interoperable components needed to be existing in order to run the model: **[interoperable component]**. As described in the previous section this workflow can be stored as an XPDL file and running within a WPS or WWS (see D2.4.1) It can be useful when testing different areas (delineated in 2) in a more automatic way.

3.4 eHabitat for ecological forecasting and ecological niche modelling

When using different climate change scenario, the eHabitat component can be seen as a habitat predictor for biodiversity. This could either be done by combining a climate change model with the eHabitat model (previous section) in order to use data forecasted before the core similarity/probability of similarity being computed, or directly using the eHabitat model on pre-forecasted data. In contrast to Figure 5 that is showing a single run for computing habitat similarities, the assessment of ecological changes requires that a first run is computed to generate a reference instance and additional runs are performed to compare results with the reference point.

Similarly, instead of using a protected area as the reference area against which similarities of ecosystems will be computed, the selection of a single location corresponding to the observation of a species as the reference point will allow the end-users to identify areas that present similarities to this exact location and so a the outcome of the model of an ecological niche for this given species. Multiple realisations of such an exercise allows for the computation of an ensemble approach to ecological niche modelling (see e.g. O'Haney, 2009).

A first proof of concept on the use of eHabitat for forecasting climate change in habitats of protected areas has been set up for the GEO-AIP3 (2011).

4 VOCABULARIES AND ONTOLOGIES FOR THE BIODIVERSITY DOMAIN

For recent overviews of the status of ontologies and other Knowledge Organisation Systems (Hodge, 2000) in the Biodiversity domain, the reader is directed to GBIF (2010) and GEOBON (2010a; 2010b). In general, this is still the preserve of specialists and awaits transfer, adoption and application by the wider biodiversity community of practice. To advance the use of ontologies in Biodiversity informatics, challenges around their development, maintenance and governance need to be addressed and the community must to be educated and provided with easy-to-use tools. This is beginning to be addressed, e.g., by the OBO Foundry[1], a collaboration of developers who have adopted best practices in biomedical ontology development and provide several biological ontologies covering various domains (e.g., anatomy, biochemistry, environment, phenotype, and taxonomy). Of special note are the Environment Ontology (EnvO)[2] for environmental attributes of an organism or biological sample and the Gene Ontology (GO)[3] for gene and gene product attributes. The GBIF vocabularies site[4] is a prototype service providing support for a range of community developed vocabularies and Darwin Core extensions and includes support for term labels in multiple languages. A user can create a vocabulary, add a concept to a vocabulary, or create an extension to Darwin Core (i.e., add a set of additional terms that extend the base set).

Within species level biodiversity, most effort to date has gone into development of common terms/vocabularies (controlled vocabularies) but it is recognised that progressing to more advanced KOS constructs such as thesauri and ontologies is necessary to enable more powerful and automated ways of data integration. These levels of data integration are going much beyond the plans envisaged in the biodiversity workflows proposed here, at least for the early

[1] http://www.obofoundry.org/

[2] http://www.environmentontology.org/

[3] www.geneontology.org

[4] http://vocabularies.gbif.org

developments planned for the DOPA in the frame of the EuroGEOSS project. Still, it is probably worthwhile to recall the existing efforts made on vocabularies, thesauri, ontologies and metadata specifications that are currently used in biodiversity informatics.

The Taxonomic Databases Working Group (TDWG[5]) is the focal organisation for standards development and ratification within the biodiversity community. It provides several vocabularies and ontologies. These are listed in Table 1 together with a non-exhaustive selection of some other vocabularies relevant to the Biodiversity domain. Metadata specifications pertaining to biodiversity are also included.

Table 1.Vocabularies, thesauri, ontologies and metadata specifications in use in biodiversity informatics

ABCD	"ABCD Schema is a common data specification for biological collection units, including living and preserved specimens, along with field observations that did not produce voucher specimens. It is intended to support the exchange and integration of detailed primary collection and observation data." (http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdIntroduction). In contrast to DwC, ABCD aims to be comprehensive and is therefore complex. It contains some 1000+ terms, of which a subset have been mapped to DwC for data exchange in the GBIF network. The ABCD version 2.06 XSD schema is available from http://rs.tdwg.org/abcd/2.06/rddl-2007-10-18.html.
Biological Data Profile of CSDGM	Biological Data Profile of the Content Standard for Digital Geospatial Metadata "broadens the application of the CSDGM so that it is more easily applied to data that are not explicitly geographic (laboratory results, field notes, specimen collections, research reports) but can be associated with a geographic location. The profile changes the conditionality and domains of CSDGM elements, requires the use of a specified taxonomical vocabulary, and adds elements." http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf
CSDGM	The Content Standard for Digital Geospatial Metadata (CSDGM), (FGDC-STD-001-1998), the US Federal Metadata standard, "provides a common set of terminology and definitions for the documentation of digital geospatial data." http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html
Darwin Core	Darwin Core (DwC) (http://rs.tdwg.org/dwc /index.htm) is a "...body of standards. It includes a glossary of terms ... intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries. The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information". The specification includes some 170+ currently accepted terms organised into eight categories: Record-level Terms; Occurrence; Event; dcterms:Location; GeologicalContext; Identification; Taxon; ResourceRelationship; MeasurementOrFact. DwC is the exchange format for the GBIF network. *Simple Darwin Core* (http://rs.tdwg.org/dwc/terms/simple/index.htm) is a specification of a commonly used subset of DwC terms and defined in an XSD schema (http://rs.tdwg.org/dwc/xsd/tdwg_dwc_simple.xsd).

[5] www.tdwg.org

EML	"Ecological Metadata Language (EML) is a metadata specification developed by the ecology discipline and for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts... EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset." http://knb.ecoinformatics.org/software/eml/
EUNIS Habitat types	The European Environment Agency EUNIS Habitat types provides a pan-European habitat classification system covering all habitat types – terrestrial, freshwater, marine, and both natural and artificial . http://eunis.eea.europa.eu/habitats.jsp
IUCN Habitats	The IUCN publishes an authority file containing listings of habitats (http://intranet.iucn.org/webfiles/doc/SSC/RedList/AuthorityF/habitats.rtf). The GBIF vocabularies site provides experimental access to the IUCN habitat terms in a web accessible manner (http://vocabularies.gbif.org/vocabularies/habitat_iucn). http://www.iucn.org/about/work/programmes/species/red_list/resources/technical_documents/authority_files/
Metadata Profile of CSDGM for Shoreline Data	Metadata Profile of CSDGM for Shoreline Data provides “the format and content for describing data sets related to shoreline and other coastal data sets.” http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/shoreline-metadata
Natural Collections Description	Natural Collections Descriptions (NCD) is “a proposed data standard for describing collections of natural history materials at the collection level “. http://www.tdwg.org/standards/312/
NBII Biocomplexity Thesaurus (BCT)	The US National Biological Information Infrastructure (NBII) Biocomplexity Thesaurus merges several individual thesauri covering aquatic sciences, life sciences, pollution, sociology, ecotourism and fire ecology. http://thesaurus.nbii.gov/
OBO Foundry ontologies	The OBO Foundry provides access to several biomedical ontologies covering such domains as environment, anatomy, phenotype , biological processes, and biochemistry. http://www.obofoundry.org/
Taxon Concept Transfer Schema	The Taxon Concept Transfer Schema (TCS) provides a standard for the exchange of taxonomic information (taxon names and taxon concepts) relating to biodiversity and natural history data. It is modelled as an XSD schema. http://www.tdwg.org/standards/117/.
TDWG Structured Descriptive Data (SDD)	Structured Descriptive Data is a standard for the capture, transport, caching and archiving of various types of descriptive taxonomic data encompassing the semi-structured, semi-formalised descriptions of a taxon or individual specimen, descriptions in dichotomous keys, and the underlying raw data descriptions of parts of individual specimens. It is modelled as an XSD schema. http://www.tdwg.org/standards/116/
TDWG Species Profile Model (SPM)	“The species profile model is intended to be a specification of data concepts and structure intended to support the retrieval and integration of data that documents species, e.g., facts about biology, ecology, evolution, behaviour, etc.”. The vocabularies are expressed in RDF. http://wiki.tdwg.org/twiki/bin/view/SPM/WebHome
TDWG Vocabularies	The TDWG vocabularies include some 30 ontologies covering such

	categories as Taxon, Occurrence, Institution, Person, etc. These are expressed in OWL but have limited semantics. GBIF(2011) provides a brief summary. http://rs.tdwg.org/ontology/voc/

5 DOPA & BROKERING SERVICES

Both scenarios, ParkSim and eHabitat, would benefit from the SOA brokering approach developed in the frame of EuroGEOSS by Nativi et al. (2011) for the GEO-AIP3 which is illustrated in Figure 7.

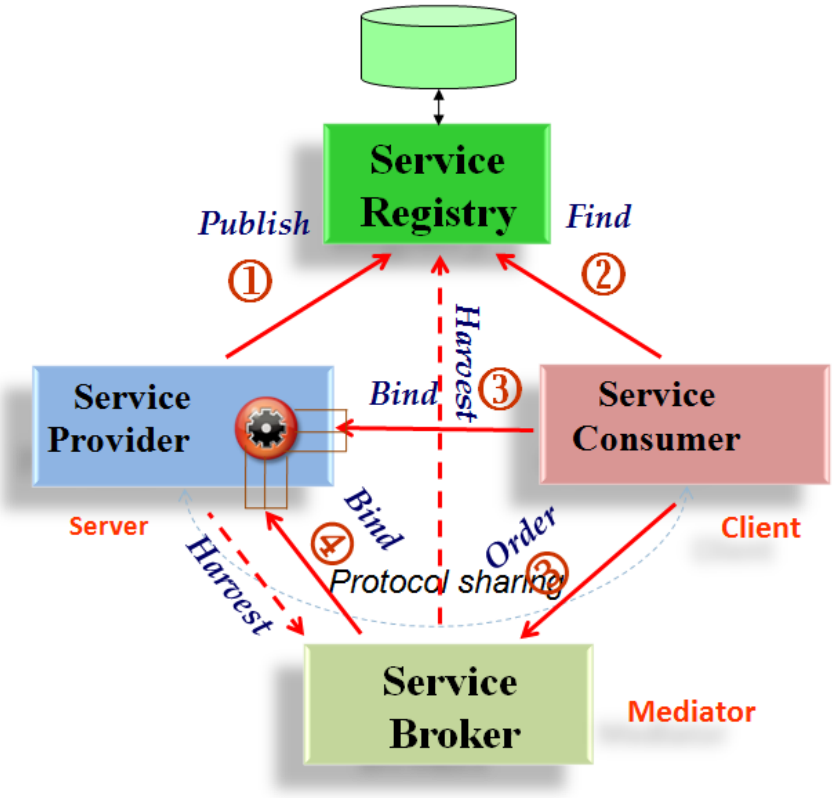

Figure 7: SOA architecture of the EuroGEOSS brokering approach

Important additions to traditional service brokers is the presence in the current design of a **Discovery Augmentation Component** (DAC) – a broker service which adds **semantic intelligence to queries**. This option would allow for improved means of finding potential input data as in Step 2' of Figure 5. The Service broker would also ideally possess **a workflow engine** in order to:

i) browse semantic repositories;
ii) retrieve concepts of interest;
iii) search for resources (i.e. datasets and models) related to such concepts;
iv) execute WPS instances.

Within the GEO-AIP3 (2011), a Graphical User Interface (GUI) has been developed for the broker to allow testing and interacting with the DAC.

6 REFERENCES

D4.1. (2009) Report of user requirements and the existing data infrastructures at partner institutions relevant to biodiversity. http://www.eurogeoss.eu/Documents/EuroGEOSS_D4_1.pdf

Dubois, G. Hartley, A., Nelson, A., Mayaux, P. and J.M. Grégoire (2009). *Towards an interoperable web service for the monitoring of African protected areas.* In: "Proceedings of the 33rd International Symposium on Remote Sensing of Environment (ISRSE)", May 4-8, 2009 Stresa, Italy

Dubois, G., A. Hartley, S. Peedell, J. de Jesus, É. Ó Tuama, A. Cottam, I. May, I. Fisher, S. Nativi and F. Bertrand (2010). DOPA, a Digital Observatory for Protected Areas including Monitoring and Forecasting Services. *European Geosciences Union (EGU) 2010*, Vienna, Austria, 2-7 May 2010.

Dubois, G., Clerici, M., Peedell, S., Mayaux, P., Grégoire and J.-M., Bartholomé, E. (2010). *A Digital Observatory for Protected Areas - DOPA, a GEO-BON contribution to the monitoring of African biodiversity*. Map Africa 2010, Cape Town, South Africa, 23-25 November 2010

Farber, O. and R. Kadmon (2003). Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160**:115-130.

GBIF (2011). Recommendations for the Use of Knowledge Organisation Systems by GBIF. Released on 04 Feb 2011. Authors: Terry Catapano, Donald Hobern, Hilmar Lapp, Robert A. Morris, Norman Morrison, Natasha Noy, Mark Schildhauer, David Thau. Copenhagen: Global Biodiversity Information Facility, 49 pp., accessible online at http://links.gbif.org/gbif_kos_whitepaper_v1.pdf.

GEO-AIP3 (2011). "eHabitat" Climate Change and Biodiversity WG Use Scenario. Engineering Report GEO Architecture Implementation Pilot, Phase 3 GEOSS Architecture Implementation Pilot. http://www.ogcnetwork.net/system/files/CCBio-eHabitat-ER-v2.0-FINAL.pdf

GEOBON (2010a). Group on Earth Observations Biodiversity Observation Network (GEO BON) Detailed Implementation Plan Version 1.0 – 22 May 2010; accessible online at http://www.earthobservations.org/documents/cop/bi_geobon/geobon_detailed_imp_plan.pdf.

GEOBON (2010b). Group on Earth Observations Biodiversity Observation Network (GEO BON) - Principles of the GEO BON Information Architecture, Version 1.0 – 14 June 2010; accessible online at http://www.earthobservations.org/documents/cop/bi_geobon/geobon_information_architecture_principles.pdf.

Hartley, A., Nelson, A., Mayaux, P. and Grégoire, J-M (2007). *The Assessment of African Protected Areas*. Joint Research Centre of the European Commission, JRC Scientific and Technical Research series, EUR 22780 EN, 2007

Hodge, G (2000). Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. 2000; accessible online at: http://www.clir.org/pubs/reports/pub91/contents.html.

Leibovici, DG Hobona, G Stock, K and Jackson, M (2009) Qualifying geospatial workfow models for adaptive controlled validity and accuracy. In: IEEE proceedings 17th International conference on GeoInformatics, August 2009, USA, pp. 1-5. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5293485

Nelson, A., A. Hartley, G. Dubois, M. Punga (2009) *Geoinformatics for the Environmental Surveillance of Protected Areas in Africa*. In: “Proceedings of the StatGIS09: GeoInformatics for Environmental Surveillance”, June 17-19, 2009 Milos, Greece

O’Haney, J.R. (2009) Neural ensembles: a neural network based ensemble forecasting program for habitat and bioclimatic suitability analysis. *Ecography*, **32**, 89–93.

Pebesma, E Cornford, D Nativi, S and Stasch, C (2010) The uncertainty enabled model web (UncertWeb). Environmental Information Systems and Services Infrastructures and Platforms, workshop at EnviroInfo2010, Bonn/Cologne, Germany 7th of October 2010 http://www.uncertweb.org/documents/presentations/the-uncertainty-enabled-model-web-uncertweb/download

Wheeler, T (2006), “Collaborative Multidiscipline/Multiscale Analysis, Modeling, Simulation and Integration in Complex Systems: System Biology,” in *Computational Science and Its Applications - ICCSA 2006*,pp. 654-664. Retrieved May 13, 2008, from http://dx.doi.org/10.1007/11751540_69

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