D4.8: Assessment and Validation Results of the Biodiversity Operating Capacity developed in EuroGEOSS

Title: D 4.8 Assessment and Validation Results of the Biodiversity Operating Capacity developed in EuroGEOSS

Creator: G. Dubois

Creation date: 26/04/2012

Date of last revision: 30/04/2012

Subject: Assessment and Validation Results of the Biodiversity Operating Capacity developed in EuroGEOSS

Status: ☑ Draft ☒ Final

Publisher: EuroGEOSS

Type: Text

Description: This document summarizes the developments made by in Workpackage 4 in setting up a Biodiversity Operating Capacity. The report presents the developments made in the frame of EuroGEOSS with a focus on the added modelling capacity provided by WP4 and a discussion on the benefits and limitations of the proposed web services.

Contributor: Gregoire Dubois, Michael Schulz, Jon Skoien (JRC) Éamonn Ó Tuama (GBIF), Ian Fisher (RSPB), Mark Balman and Ian May (Birdlife International), Stefano Nativi and Mattia Santoro (CNR), Craig Mills (UNEP-WCMC)

Format: Word document.


Rights: ☑ Restricted ☒ Public

Identifier: EuroGEOSS_D4_8.doc

Language: EN

Relation: WP4, WP2

Coverage: Not applicable

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

ACRONYMS AND ABBREVIATIONS .................................................................................. 3
Executive Summary ........................................................................................................ 5
1 Towards a Digital Observatory for Protected Areas (DOPA) ....................................... 6
   1.1 Assessing protected areas .................................................................................... 6
   1.2 On the potential benefits of a Service Oriented Architecture (SOA) .................... 6
   1.3 The biodiversity operating capacity in EuroGEOSS ............................................ 9
2 WP4 partners and status of their data and modelling infrastructure before EuroGEOSS ...... 10
   2.1 The Joint Research Centre (JRC) ....................................................................... 10
   2.2 The UNEP World Conservation Monitoring Centre (WCMC) ............................ 10
   2.3 The Global Biodiversity Information Facility Secretariat (GBIF) ......................... 11
   2.4 BirdLife International ......................................................................................... 12
   2.5 The Royal Society for the Protection of Birds (RSPB) ......................................... 13
3 Setting up critical data infrastructures for biodiversity ................................................. 15
   3.1 Coordination of the DOPA and development of web based modeling tools by JRC ... 15
      3.1.1 Coordinating the migration from the APAAT towards the DOPA .................... 15
      3.1.2 Setting up of modeling services ................................................................. 15
         3.1.2.1 eHabitat, a web processing service for ecological modeling ..................... 16
   3.2 Provision of park boundaries and of web based editing tools by WCMC ................. 18
      3.2.1 Official data collection of protected area boundaries .................................... 18
      3.2.2 Citizen data collection of protected area boundaries .................................... 18
      3.2.3 Data publishing of the World Database on Protected Areas ......................... 20
      3.2.4 Data standards ............................................................................................ 20
   3.3 Web services and catalogues for species occurrences by GBIF ............................... 21
      3.3.1 Provision of species occurrences by GBIF .................................................... 21
      3.3.2 Development of a metadata catalogue and supporting activities .................... 21
3.4 Provision of important bird areas and species range maps by Birdlife International .... 22
      3.4.1 Release of species range maps and fact sheets ................................................. 23
   3.5 Increased data collection by RSPB and validation of modeling scenarios by the RSPB. 27
4 Enhanced modelling capability developed by WP4 for EuroGEOSS ......................... 28
   4.1 A short introduction to ecological modeling with eHabitat .................................... 28
      4.1.1 Using Mahalanobis distances for computing ecological similarities ................ 28
      4.1.2 Ecological forecasting with eHabitat ............................................................. 29
      4.1.3 Ecological niche modeling with eHabitat ....................................................... 33
         4.1.3.1 Data and associated Web Services ......................................................... 33
         4.1.3.2 The case of the Black Harrier (Circus maurus) ....................................... 35
         4.1.3.3 Bioclimatic area of the Black Harrier (Circus maurus) ............................ 36
         4.1.3.4 Use of third party web services for decision making ................................. 38
      4.1.4 eHabitat WPS and the EuroGEOSS brokering approach ............................... 39
         4.1.4.1 Discovery Broker .................................................................................... 40
         4.1.4.2 Discovery Augmentation Broker (Semantic discovery augmentation) ........ 41
         4.1.4.3 Access Broker ....................................................................................... 42
         4.1.4.4 Scenario Steps ....................................................................................... 42
   4.2 Multi-disciplinary modeling in EuroGEOSS: conclusions and further considerations ... 45
5 Future Developments .................................................................................................. 47
REFERENCES .................................................................................................................. 50
ACRONYMS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Architecture and Data</td>
</tr>
<tr>
<td>AIP-2</td>
<td>Architecture Implementation Pilot, Phase 2</td>
</tr>
<tr>
<td>AOC</td>
<td>Advanced Operating Capacity</td>
</tr>
<tr>
<td>APAAT</td>
<td>African Protected Areas Assessment Tool</td>
</tr>
<tr>
<td>CAAS</td>
<td>Composition as a service</td>
</tr>
<tr>
<td>CSR</td>
<td>Components and Services Registry</td>
</tr>
<tr>
<td>CSW</td>
<td>Catalogue Service Web</td>
</tr>
<tr>
<td>DOPA</td>
<td>Digital Observatory for Protected Areas</td>
</tr>
<tr>
<td>GCI</td>
<td>GEOSS Common Infrastructure</td>
</tr>
<tr>
<td>GDAL</td>
<td>Geospatial Data Abstraction Library</td>
</tr>
<tr>
<td>GBIF</td>
<td>Global Biodiversity Information Facility</td>
</tr>
<tr>
<td>GEO</td>
<td>Group on Earth Observations</td>
</tr>
<tr>
<td>GEOSS</td>
<td>Global Earth Observation System of Systems</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information Systems</td>
</tr>
<tr>
<td>GRASS</td>
<td>Geographic Resources Analysis Support System</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>IOC</td>
<td>Initial Operating Capacity</td>
</tr>
<tr>
<td>JRC</td>
<td>Joint Research Centre</td>
</tr>
<tr>
<td>OGR</td>
<td>OGR Simple Feature Library</td>
</tr>
<tr>
<td>OWS</td>
<td>OGC Web Services</td>
</tr>
<tr>
<td>RSPB</td>
<td>The Royal Society for the Protection of Birds</td>
</tr>
<tr>
<td>SBA</td>
<td>Societal Benefit Area</td>
</tr>
<tr>
<td>SIR</td>
<td>Standards Registry System</td>
</tr>
<tr>
<td>SOA</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>SWG</td>
<td>Scenario Working Group</td>
</tr>
<tr>
<td>UIC</td>
<td>User Interface Committee</td>
</tr>
<tr>
<td>UNEP</td>
<td>United Nations Environment Programme.</td>
</tr>
<tr>
<td>WCMC</td>
<td>World Conservation Monitoring Centre</td>
</tr>
<tr>
<td>WCS</td>
<td>Web Coverage Service</td>
</tr>
<tr>
<td>WFS</td>
<td>Web Feature Service</td>
</tr>
<tr>
<td>WP</td>
<td>Work package</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>WMS</td>
<td>Web Mapping Service</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
<tr>
<td>WPS</td>
<td>Web Processing Service</td>
</tr>
</tbody>
</table>
EXECUTIVE SUMMARY

The number of interoperable research infrastructures has increased significantly with the growing awareness of the efforts made by the Global Earth Observation System of Systems (GEOSS). One of the Social Benefit Areas (SBA) that is benefiting most from GEOSS is biodiversity, given the costs of monitoring the environment and managing complex information, from space observations to species records including their genetic characteristics.

In the frame of the setting up of the Digital Observatory for Protected Areas (DOPA) through the EuroGEOSS project, a number of international organizations, namely the Joint Research Centre (JRC) of the European Commission, the Global Biodiversity Information Facility (GBIF), the UNEP-World Conservation Monitoring Centre (WCMC), Birdlife International and the Royal Society for the Protection of Birds (RSPB), significantly improved and increased access to their data or sometimes even started having a capacity on their own for delivering datasets that are essential for biodiversity assessments. The availability on the Internet, for the first time, of more than 10,000 reference bird species maps, is an example of this effort.

But GEOSS goes beyond simple data sharing, to encourage the publishing and combination of models, an approach which can ease the handling of complex multi-disciplinary questions. eHabitat for example, a basic Web Processing Service (WPS) developed in EuroGEOSS to compute the likelihood of finding ecosystems with equal properties, can provide essential information for conservation management. When chained with other services providing data on climate change, eHabitat can be used for ecological forecasting and becomes a useful tool for decision-makers assessing different strategies when selecting new areas to protect. Further enhanced by the EuroGEOSS discovery broker service, eHabitat can access virtually any kind of thematic data that can be considered as useful when defining ecosystems and their future persistence under different climatic or development scenarios.

This report will present the data and model services developed by these international organisations in the frame of the EuroGEOSS project and illustrate the benefits of these newly developed infrastructures through a few case studies which forecast the impact of climate change on protected areas.
1 TOWARDS A DIGITAL OBSERVATORY FOR PROTECTED AREAS (DOPA)

1.1 Assessing protected areas

Protected Areas (PAs) play an essential role in conservation programs and in the sustainable use of natural resources. The African Protected Area Assessment Tool (APAAT)\(^1\) developed at the Joint Research Centre (JRC) of the European Commission by Hartley et al. in 2007, is an online information system designed to assess the state and pressure on African PAs and to prioritize them accordingly, in order to support decision making and fund allocation processes. The APAAT contains information on 741 PAs, across 50 countries, and includes information on 280 mammals, 381 bird species and 930 amphibian species, and a wide range of climatic, environmental and socio-economic data. This information is further used to compute 6 indicators for each protected area, quantifying its value with regards to species diversity of amphibians, birds and mammals, the level of irreplaceability of the ecosystems it is hosting, and to the threats coming from population pressure and agricultural pressure at the park boundaries. These indicators can be further summarized to compare each PA with other protected areas from the same country or the same ecoregion. Figure 1 summarizes the process of generating the indicators of the APAAT as well as the type of information used as input.

The APAAT was set up using ArcGIS, a popular Geographic Information System (GIS), and its programming language, the Arc Macro Language (AML), to develop scripts that process the data and which call scripts written in other languages as well as system processes to perform tasks that cannot be achieved in AML. The system is based on a central repository, regularly mining information in the distributed databases: It can answer simple questions on the representativity and value of protected areas. Figure 2 indicates the main data sources used for generating the biodiversity indicators which require boundaries of PAs from the UNEP-World Conservation Monitoring Centre (WCMC), landcover and environmental data from the JRC and species range maps from the International Union for the Conservation of Nature (IUCN) for what concerns amphibians and mammals and from the Birdlife International for the bird data. The data collection process is suffering from the absence of interoperability between systems and of common formats, preventing so the system from being automated and further used for monitoring purposes. Updates like the deletion or addition of PAs are also an issue as such changes need to be made by hand. A really distributed system would therefore drastically improve the efficiency of this system and would allow an end user to submit complex queries and download datasets matching requirements.

1.2 On the potential benefits of a Service Oriented Architecture (SOA)

The next stage of the development of the APAAT would be to create an environmental monitoring system that is able to capture the dynamic of spatio-temporal changes in habitats and anthropogenic pressure on PAs (Nelson et al., 2009). This would be extremely valuable for screening the impact of policies and projects that dealing with the management of biodiversity and natural resources. Such a monitoring system would however require an automated access to regularly updated data, something that is still thorny mainly because of the lack of interoperability with most data sources.

Extending the capabilities of the APAAT, one of the few existing operational biodiversity information systems covering Africa, to a global distributed database of biodiversity information, requires that information from a wide array of data providers around the world is easily available.

\(^1\) http://bioval.jrc.ec.europa.eu/ APAAT/
and coupled with model servers that can provide vital biodiversity analysis tools to the GEOSS community of science.

Figure 1. Updated version of the same figure from Hartley et al. (2007) giving an overview of the different parts of the process and how they fit together to create the online reports. The information presented by the APAAT is derived from a large variety of data sets, including those used to characterize habitats (land cover, terrain, water bodies, ecoregions, climate, vegetation vigor) species distribution (location maps of amphibians, birds, mammals) or anthropogenic activities (agriculture, road network, urban areas).
This means that the large variety of data types and data providers need not only to be interoperable and efficiently networked, in particular for species related information, but also that the archiving of large datasets will need to be properly considered. Developing such a working environment in which large distributed and interoperable databases are intensively interrogated will require some significant coordination efforts as well as the development of the appropriate analysis framework such as the Model Web (Geller and Turner, 2007).

The Shared Environmental Information System (SEIS) is a recent collaborative initiative of the European Commission and the European Environment Agency (EEA) to establish together with the Member States an integrated and shared EU-wide environmental information system. Its purpose is to improve, modernize and streamline current information systems. In return, SEIS will provide member states and EU institutions with more coherent environmental information to facilitate the drafting, implementation and effectiveness of environmental protection policies. Although focusing on Europe, its principles are extremely relevant to the further development of monitoring tools for PAs.

In particular, as stated in a communication of the European Commission (COM/2008/0046 final), information in SEIS should be

1) managed as close as possible to its source;

2) collected once, and shared with others for many purposes;
3) readily available to public authorities and enable them to easily fulfill their legal reporting obligations;

4) readily accessible to end-users, primarily public authorities at all levels from local to European, to enable them to assess in a timely fashion the state of the environment and the effectiveness of their policies, and to design new policy;

5) accessible to enable end-users, both public authorities and citizens, to make comparisons at the appropriate geographical scale (e.g. countries, cities, catchment areas) and to participate meaningfully in the development and implementation of environmental policy;

6) fully available to the general public, after due consideration of the appropriate level of aggregation and subject to appropriate confidentiality constraints, and at national level in the relevant national language(s); and

7) shared and processed through common, free open-source software tools.

Even if point 4 focuses only on one of the key end-users (policy-makers, scientists and natural resource manager), the overall idea behind SEIS remains very relevant for the long term strategy of the African PA related work.

A service oriented architecture (SOA) in which interoperable web services, currently managed mainly by large environmental organizations, is not sufficient to monitor changes affecting biodiversity and the associated habitats. A web service for the monitoring of protected areas, would require that end-users, in particular those that can provide ground truth data, can upload, correct and edit data/metadata. Customized and reusable Graphical User Interfaces (GUI) can nowadays be easily designed to allow data providers to validate, correct and add to existing repositories of species data, protected area boundaries, threat and habitat information, basic GIS layers etc. In this point of view, point 7 of the SEIS principles is not only an absolute condition for an increased capacity building of end-users and for the empowerment of regional authorities but it is also essential for the dynamic monitoring of PAs through for ground validation of African data.

With the steady growth of environmental web services, distributed databases and common data formats, rises the risk of data redundancy. It should be ensured, for example, that borders of PAs are maintained and update data repositories, catalogues, registries and associated web services are at least as important as the growth in accessible and updatable information.

1.3 The biodiversity operating capacity in EuroGEOSS

Workpackage 4 (WP4) of EuroGEOSS is aiming at developing a biodiversity operating capacity derived from the APAAT that provides access not just to data but also to analytical models made understandable and useable by scientists from different disciplinary domains. To become an operational spatio-temporal web service for the monitoring of protected areas, efforts have focused on porting the core elements of the APAAT on the web with the objectives of facilitating the maintenance and portability of the system. Other boundary conditions for further development are the use of open standards for spatial data such as developed in the context of the open geospatial consortium (OGC) and the use of R as our environment for modeling. These efforts should ease the sharing of experience and knowledge between experts and allow the APAAT to evolve from a sophisticated web client to a fully operational web service within a Service Oriented Architecture (SOA). The use of open standards for spatial data and of open source programming languages for the development of the core functionalities of the system is also expected to encourage the participation of the scientific community beyond the current partnerships and ease the sharing of biodiversity related information.
This new infrastructure for biodiversity, called the Digital Observatory for Protected Areas (DOPA\(^2\)), was developed by the Joint Research Centre of the European Commission in collaboration with the partners of WP4, namely the GBIF, the UNEP-WCMC, Birdlife International and the Royal Society for the Protection of Birds (RSPB) (Dubois et al, 2010). In addition of the assessment strategy proposed by the APAAT, the DOPA would be global and further allow the monitoring and forecasting of the state and pressures in Protected Areas (PAs) in view to provide the scientific and technical support to decision makers but also to the scientific community.

In the following, after a brief overview of the partners of the WP4, we will discuss the original data and model infrastructures of the JRC partners before EuroGEOSS, the developments made in the frame of the project and finally illustrate with case studies the benefits provided to the community of the newly developed data and modeling infrastructures.

WP4 partners and status of their data and modelling infrastructure before EuroGEOSS

1.4 The Joint Research Centre (JRC)

JRC the European Commission, engaged in direct research. The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. The Institute for Environment and Sustainability (IES) is one of the JRC Institutes, the mission of which is to provide is to provide scientific and technical support to EU policies for the protection of the global environment contributing to sustainable development. The MONDE (Monitoring of Natural Resources for Development and cooperation) group has developed the APAAT and monitors terrestrial ecosystems in EU development-assistance priority areas. Among the main tasks of this group are to reduce uncertainty in global land cover statistics for climate and resource studies; contribute to an improved land management and to assist European Commission funded aid projects by mapping areas subject to land degradation, assessing disturbance by fire, monitoring the rainforests and improving the management of the biodiversity.

The APAAT being already discussed in the introduction of this report, it will not be further detailed here.

1.5 The UNEP World Conservation Monitoring Centre (WCMC)

The UNEP World Conservation Monitoring Centre is the biodiversity assessment and biodiversity policy support arm of the United Nations Environment Programme, the world’s foremost intergovernmental environmental organization. The Centre has been in operation for over 25 years, providing objective, scientifically rigorous products and services to help decision makers recognize the value of biodiversity and apply this knowledge to all that they do. The Centre’s core business is locating data about biodiversity and its conservation, interpreting and analyzing that data to provide assessments and policy analysis, and making the results available to both national and international decision makers and businesses. UNEP-WCMC has a long history of supporting the conservation and sustainable use of biodiversity through provision of information products and services that contribute directly to decision making processes. It has also played a significant role in supporting implementation of biodiversity-related agreements at all levels, not only through these assessment processes, but also through more direct support in data management, synthesis of reports, development of information strategies, and so on. Since 1981 UNEP-WCMC has been identifying and compiling information on the protected areas of the world to produce comprehensive global dataset and maps. The resulting dataset, known as the World Database on

Protected Areas (WDPA), is widely utilized by policy makers, park managers and by industry. UNEP-WCMC's work on protected areas is carried out in close collaboration with the World Commission on Protected Areas and the IUCN Programme on Protected Areas.

The development of the Protected Planet website\(^3\) began in early 2010 by UNEP-WCMC with a goal to improve decisions about the world’s protected areas by making the data in the World Database on Protected Areas as freely available as possible and by bringing together information about them with other information about species, habitats and threats as well as contextual information from other sources. This work would be achieved by working towards the following objectives:

1. Improve data held within the World Database on Protected Areas
2. Promote good conservation outcomes in protected area by facilitating development of ‘communities of interest’ around protected areas
3. Ensure key decision makers have easy to understand information available to them about the many values of protecting biodiversity

### 1.6 The Global Biodiversity Information Facility Secretariat (GBIF)

Established and funded by governments in 2001 through an OECD Global Science Forum recommendation, the Global Biodiversity Information Facility (GBIF) is the world’s largest multilateral initiative for enabling free access to biodiversity data via the Internet to underpin scientific research, conservation and sustainable development. GBIF serves science and society as a global ‘public good’ and is well placed as a long term, sustainable network to play a central role in GEOSS as exemplified through its participation in EuroGEOSS.

Halting biodiversity loss remains one of humanity’s most critical and urgent challenges and requires access to credible, authoritative information to support scientific analyses which can ultimately inform policy decision making. This is GBIF’s mandate as a global science infrastructure – to provide the means to enable ready access to such data based on open data sharing principles. It achieves this by building an informatics platform, by promoting the development of digital data capture and exchange standards, enabling online publishing and sharing of biodiversity data, catalyzing the development of analytical tools, promoting global participation, linking up a global network of participants and building capacity to use this infrastructure – all in support of scientific research and science-based policy. And though its network comprised of national, regional and thematic Biodiversity Information Facilities, the GBIF Data Portal (www.gbif.org) provides unified access to a continually expanding set of biodiversity data records – currently over 321 million records from some 8842 datasets from 379 data publishers (Figures 3, 4).

GBIF had two main roles in the EuroGEOSS project: provision of species occurrence data through standardized web services, and implementation of a distributed metadata catalogue system. An informatics infrastructure of this kind, relying on the use of common data standards and protocols to support interoperability, particularly via web services, and enabling the integration of primary biodiversity data with other data types (e.g. climate data, protected areas, as in the DOPA) offers many benefits in support of biodiversity assessments by allowing diverse systems (and individual scientists) to discover and access data in order to undertake scientific analyses and other investigations.

\(^3\) [http://www.protectedplanet.net/](http://www.protectedplanet.net/)
Figure 3. GBIF has built a global informatics platform through partnership with some 57 countries and 45 organizations.

1.7 BirdLife International

The BirdLife International Partnership is a world-wide network of autonomous non-governmental organizations (NGOs), working together to conserve biodiversity through shared priorities, programs and actions. A decentralized Secretariat, made up of a global office in the UK and six regional offices around the world, coordinates activities, disseminates information and provides services to the Partnership. BirdLife’s mission is to conserve wild birds, their habitats and global biodiversity, by working with people towards sustainability in the use of natural resources.

BirdLife’s work builds on a solid foundation of science. Birds are the best known group of organisms on earth, and so provide a sound basis for BirdLife’s vast network of observers to monitor the state of the planet’s biodiversity. BirdLife is the Listing Authority for birds in the IUCN Red List of plants and animals at risk of extinction, assessing the status and trends of individual species. The BirdLife Partnership is also working to identify, document and conserve a network of Important Bird Areas (IBAs) using standardized criteria. These IBAs are critical sites for threatened, restricted range, biome-restricted and migratory bird species. The BirdLife Partnership undertakes research, monitoring and conservation action, education and advocacy, across a broad agenda, focusing not only on birds but also on other biodiversity, as well as on social issues such as sustainable development. To support these programs of work, BirdLife has developed and maintains a number of data management applications. These applications enable staff within its Secretariat and Partnership to manage and query data on individual bird species global status, IBAs and individual species observations. These data are available to the public through the
BirdLife Website and through WorldBirds⁴ (a web tool developed in collaboration with RSPB (BirdLife in the UK) and Audubon (BirdLife in the USA)). BirdLife also maintains spatial data on the distributions of individual species and the boundaries of individual IBAs.

Before the EuroGEOSS project, BirdLife International had range maps for approximately three quarters of the world's birds (7,000). These maps were slowly being created as and when funding and resources allowed. The maps were primarily used internally (but with occasional requests from academia to use the maps within analyses). Only 2,000 threatened species maps were shown on the BirdLife International website. These existed as a static image format (GIF), no web services (WFS, WMS) were used at this point.

Figure 4. As of April 2012, the GBIF data portal provides access to more than 321 million occurrence records from 8,842 datasets from 379 publishers and spanning a wide range of geospatial, temporal and taxonomic coverages. The map shows counts (observations) per one degree cell ranging from 1 to 100,000+.

1.8 The Royal Society for the Protection of Birds (RSPB)

RSPB is Europe's largest conservation organization, with over one million public members and a permanent staff of over 1200. It seeks to achieve its UK objectives through the conservation of habitats and species, either by owning and managing land itself or by influencing land use practices and government policy, backed up by quality scientific research. It manages 181 nature reserves covering over 121,000 ha, which are home to 80% of the UK's most threatened species.

⁴ http://www.worldbirds.org
RSPB adopts an evidence-based approach to conservation, and has significant data holdings both from its reserves and site/species surveys, both in the UK and abroad. Internally, RSPB has several large home-grown databases, including Merlin, a distributed reporting and mapping tool, which currently holds almost 3 million bird observations. The RSPB is a member of the National Biodiversity Network (NBN), and is working to mobilize conservation datasets onto the network. As the UK Partner of BirdLife International, RSPB aims to maintain the numbers, diversity and geographic distribution of the world's most important sites, species and habitats for birds and helps to coordinate.

The original plans for the RSPB related mainly to the digitization of source data and the delivery of those data and others to the Global Biodiversity Information Facility (GBIF) for use within the various streams of EuroGEOSS work. There was no intention to provide any web services directly from the RSPB, but rather use those at BirdLife for delivery of Important Bird Area (sites) information and GBIF for bird observations.

RSPB initial objectives were to:

i. deliver UK datasets into the UK National Biodiversity Network and onwards into GBIF;

ii. complete implementation of WorldBirds across Africa, develop systematic Common Bird Monitoring functionality, and encourage involvement of BirdLife Partners in Common Bird Monitoring (WorldBirds is an on-line system designed to allow birders and professionals to submit their casual and systematic bird observations to local BirdLife Partners);

iii. facilitate digitization of African datasets through provision of small grants to BirdLife Partners;

iv. deliver global WorldBirds data to GBIF, through development of an intermediate data warehouse and functionality to pass subsets of data to the Avian Knowledge Network (AKN) for onwards transmission as part of the regular AKN to GBIF updates;

v. provide IBA data and boundaries to BirdLife international;

vi. conduct scientific analyses to show the potential uses of ad hoc vs systematic lists;

vii. train BirdLife Partners in data management and Citizen Science.

An additional role was to help validate and test services provided by other WP4 participants, such as the Digital Observatory for Protected Areas (DOPA). It was also important to have the link all the way from grass-roots recorders to providers of biodiversity services.

Benefits for biodiversity services were through the provision of data to enable a more accurate and rounded assessment, filling gaps in knowledge and increasing the number of observations available on which analyses could be performed.
2 SETTING UP CRITICAL DATA INFRASTRUCTURES FOR BIODIVERSITY

2.1 Coordination of the DOPA and development of web based modeling tools by JRC

2.1.1 Coordinating the migration from the APAAT towards the DOPA

JRC’s main task was to coordinate WP4 to ensure the overall migration of the APAAT towards a system constructed around web based services. This would require from the partners that their species range maps (Birdlife) and occurrences (GBIF) become available via web services and the flow of the input data (RSPB) to these critical data infrastructures would also be based on the Internet. Rather than being available only through an online database, boundaries of protected areas supplied by the WCMC would become also available via web services and new means would be put in place to allow remote users to improve via the internet the contents of these critical datasets. Figure 5 illustrates the new flow of information as proposed in the DOPA and should be contrasted with Figure 2 describing the same process for the APAAT.

![Diagram of data flows in the DOPA](image)

Figure 5. Data flows of the initial version of the Digital Observatory for Protected Areas (DOPA). In contrast to the APAAT, the DOPA is built around interoperable web services and its modeling capability is relying mainly on web processing services.

2.1.2 Setting up of modeling services

Better results from ecological modeling can be obtained either by improving existing models or by developing new ones. Chaining interoperable model components is a third alternative that is particularly interesting because such a chain can potentially answer more questions than the individual models alone, allowing users to address complex questions in a variety of different contexts. It is still a challenge to set up a computing infrastructure where models can be easily plugged and played. However, the proposed “Model Web” (Geller & Turner, 2007) envisages exactly such an environment, and encourages the practical development of a distributed, multidisciplinary network of independent, interoperating models (and data stores) communicating
with each other using Web services. Beyond the simple sharing of information, the Model Web conceives increasing access to models and their outputs, and aims to facilitate greater model-model interaction, resulting in webs of interacting models, databases, and websites (Nativi et al., 2011).

2.1.2.1 eHabitat, a web processing service for ecological modeling

The largest potential benefit from the Model Web is likely to be the practical and easy re-use of basic modeling components for different purposes. We believe that the granularity of the models expected to interact with each other is a critical factor in any operational version of the ModelWeb. A higher granularity is likely to generate more reusable elementary services, greater control for the users composing those services and thus, ultimately, more complex and useful modeling chains. It is with these issues in mind that the eHabitat service was exposed as a habitat-modeling Web processing service, following a Service Oriented Architecture (SOA) model.

The first version of eHabitat (eHabitat 1.0) was designed as a proof of concept to compute habitat similarities for a given Protected Area using only three predefined thematic maps (Dubois et al., 2011). In short, eHabitat would allow end-users to select a protected area, use a predefined set of thematic maps which are characterizing the ecosystems of the area (typically elevation, slopes, percentage of tree cover, climatic variables, vegetation indicators, etc.) and the modeling service would compute the probabilities to find similar combinations of these variables outside of the protected area.

However, the need to monitor ecosystems outside of protected areas, whether terrestrial or marine, is stronger than ever, if only to assess connectivity between protected areas and the external pressures caused by competition for land and water. Providing the scientific community with the means to compute such maps of habitat similarities anywhere on the globe, using their own thematic ingredients, is therefore an interesting option. eHabitat was thus designed to follow a Service Oriented Architecture (SOA) and the underlying model and code used for computing the habitat similarities can be made available as dispersed services/processes over a network (Figure 6).

Figure 6. Data flow in the eHabitat modeling service showing the use of a variety of environment thematic layers made available through catalogues for computing habitat indicators.
Detailed in the uses cases, the current version of eHabitat (eHabitat 2.0) allows for an arbitrary number of thematic maps along with the definition of an area of interest, which serves as a bounding box constraint for further processing (GEO AIP3, 2011). Technically, the WPS was mainly built using an Open Source development environment, i.e. R (R Development Core Team, 2010) for the algorithm which computes the Mahalanobis distances, and Python for setting up the Web Services. Data processed by eHabitat 2.0 are obtained in raster grid format from Web Coverage Services (WCS). An XML request is thus submitted to the WPS indicating where to find the boundaries of a protected area from a Web Feature Service (WFS) as well as the URL addresses where the selected thematic maps may be found. The ehabitatprocess.py script is executed by a PyWPS (Cepicky and Becchi, 2007) instance, which further processes the input data using OWSLib and GDAL (a generic raster library supporting a large variety of different formats) and sends it to R. The results returned from the R code are processed to generate different output formats depending on the required further use of those outputs. For example, if the user will perform further processing on the output, a GeoTiff or netCDF dataset may be requested. For visualization, simple PNG images may be sufficient, or for visualization by other Web mapping clients, the user may request that the output is published as an OGC-WMS, using GDAL, the Mapscript Python package and the Python Image Library (PIL). A schema showing the architecture of eHabitat 2.0 is shown in Figure 7.

Figure 7. Architecture of eHabitat 2.0. The WPS receives an XML request with the ID of a protected area available from the World Database of Protected Areas (IUCN-WCMC, www.wdpa.org), a number of thematic maps as WCS addresses and a calculation bounding box. The PoHS is computed for the protected area within the defined bounding box and the final results released in GeoTIFF, png or OGC-WMS format.
2.2 Provision of park boundaries and of web based editing tools by WCMC

Collecting, managing and presenting maps and statistics on the global system of protected areas as an input to policy makers has rested on the shoulders of the International Union for the Conservation of Nature (IUCN) and the United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) since 1962. Much has changed over the decades regarding the technology, the number of protected areas and even the institutions themselves. What began as a type-written list of protected areas (the UN List) presented to the United Nations every 4 years has developed into a complex spatial and relational database produced monthly that supports numerous regional and international processes and is the source of the official statistics on how much of the world’s land, oceans, forests and other realms are protected.

Changes in technology, primarily the wide adoption of the World Web Web, as well as advances in geographic information systems (GIS) have caused us to re-think our processes of data acquisition, dissemination and quality control. New social networking services have also changed the way people interact with information and with each other. This has created an incredible opportunity to engage with new constituencies in new ways to support conservation.

2.2.1 Official data collection of protected area boundaries

The EuroGEOSS project contributed to the development of tools for governments to submit official park boundaries online. These tools provided government technicians the ability to upload their geospatial data, automatically validate their information against a set of pre-programmed scripts and visually inspect the boundaries and other protected area information all within the protectedplanet.net website. These tools represented a very large technological challenge given the variety and complexity of protected area information that is submitted by governments. Geospatial data formats are traditionally a very unstable data protocol therefore a lot of effort was spent making sure the data were ‘clean’ when they entered the system. Open source geospatial tools (GDAL, GEOS, PROJ, PostGIS) were used to format and validate the data into a form that could be inspected by the government technicians. Once the inspection had occurred, the data can be fed live into the protectedplanet.net website (Figure 8).

2.2.2 Citizen data collection of protected area boundaries

As well creating a system for official supply of protected area data, UNEP-WCMC also created tools to improve these boundaries by the web ‘community’. The community could edit both the general information about the park and the boundary information. Protectedplanet.net did not want to duplicate successful initiatives on the Internet, instead work with them. These include allowing descriptions from Wikipedia and photographs from Panoramio (Figure 9). Users of protectedplanet.net can also either rank or correct the Wikipedia articles and Panoramio photographs.

In the last 10 months, protectedplanet.net has seen 2000 protected areas updated by the web community. These edits have occurred in parts of the world where previous data quality was low. As part of the protectedplanet.net system, the rough location of the community user is stored. From this we have discovered that approximately 70% of users live in the same country as the protected area they are updating. Users have also ranked 100,000 photographs of protected areas and added 2500 Wikipedia articles.
A large challenge associated with a crowd sourcing approach to data collection is the integration of the data into official channels. Protectedplanet has gone some way into providing the tools for governments to incorporate the community data and in 2012 will invest in making it even easier for governments, combined with protected area experts to validate and mainstream the information. The success of a crowd sourcing initiative is obviously linked to having a big enough crowd in the first place. For protectedplanet.net, people that are committed to editing the boundaries of the protected areas is essential. Although protectedplanet.net is a relatively popular website, it will require 100 times more visitors per day at the current rate of editing to review all the protected areas. It is therefore necessary to employ marketing strategies to both reach the protected areas network and increase the general popularity of the site. For this UNEP-WCMC and IUCN are partnering with Google and ESRI to drive more web traffic to the site by embedding the data into their systems which then in turn link back to protectedplanet.net. It is clear from UNEP-WCMC’s first foray into crowd sourcing through protectedplanet.net, environmental data collection are getting increasingly democratized. With mobile phones becoming ubiquitous in the developing world, the unstoppable speed and coverage of the Internet infrastructure and the expectation of web users to be able to ‘contribute’, finding systematic ways to collect the information has never been more important.
2.2.3 Data publishing of the World Database on Protected Areas

A key output from UNEP-WCMC in WP4 was the provision of protected areas data using open geospatial consortium (OGC) web services. These services are now accessible in both web feature service (WFS) and web map service (WMS) protocols (http://maps.protectedplanet.net/geoserver/ows?service=wms&version=1.1.1&request=GetCapabilities). As well as the protected area boundaries, these web services also provide access to all information about the protected areas described in the data standard document. A key challenge of delivering the WDPA in this format was the size of the data set and performance of the services. The WDPA consists of 180,000 protected areas represented by over 2 billion coordinates. This volume of data presents difficulties in both downloading the raw data across the web and performing special querying on the fly. Therefore as part of the service, we set up pre-cached maps and data feeds to improve the speed of delivery.

2.2.4 Data standards

EuroGEOSS enabled UNEP-WCMC to consolidate and create updates to the standard for protected areas data collection. Consisting of 15 standard fields and adhering to ISO 19115 metadata standards, the new standard ensured a consistency of data collection from the 200+...
countries currently supplying data to UNEP-WCMC. The latest data standard can be found here: http://www.wdpa.org/PDF/WDPA%20Data%20Standard.pdf. The challenge with creating the data standards was reducing around 50 patchy pieces of information about the protected areas down to a manageable size (15). Choices were made based on data coverage, reviewing existing needs by governments and through consultation with the protected area community.

2.3 Web services and catalogues for species occurrences by GBIF

The main role of GBIF in WP4 was the provision of species data via web services.

2.3.1 Provision of species occurrences by GBIF

The DOPA draws on the existing GBIF REST based web services for on-the-fly access to species occurrence data. Specifically, the GBIF Occurrence service (http://data.gbif.org/ws/rest/occurrence) can be dynamically accessed by the DOPA to obtain reported occurrences of species within protected areas (or other arbitrary bounding boxes). As participants from EuroGEOSS (e.g. BirdLife, RSPB) and others publish data to the GBIF network, it becomes automatically available through the web services.

In addition, GBIF established OGC Web Map and Web Feature Services for GBIF-mediated African biota occurrence data. (MySQL scripts and report available5). These services have also been registered in the EuroGEOSS broker. The OGC service access points are available at: GBIF WMS African Biota: http://ogc.gbif.org/wms?service=WMS&version=1.1.1&request=GetCapabilities

GBIF WFS African Biota: http://ogc.gbif.org/wfs?service=WFS&version=1.1.0&request=GetCapabilities

In a related development (with support from the GEO Secretariat), GBIF coordinated development of a freely distributable, standards-based, Open Source, web-based GIS client as an Early Product of GEO BON. Supporting viewing and download services, it functions primarily as a means for visualizing information from disparate sources delivered through standards-based web services. It features a demonstration application on African Protected Areas featuring several web services from EuroGEOSS and other participants, and highlighting the DOPA. Further information is available at http://tools.gbif.org/gpaaamp-demo including access to source code and a video demo prepared for GEO Plenary in Beijing, Nov 2010.

2.3.2 Development of a metadata catalogue and supporting activities

Catalogue and network

EuroGEOSS enabled GBIF to begin the implementation of a distributed metadata catalogue system. There were several steps involved. A survey of metadata infrastructure in WP4 partner institutions was followed by scoping of the design of the GBIF metadata catalogue system based on recommendations of a dedicated GBIF task group6. A review / evaluation of catalogue systems ultimately lead to the development of a modular, lightweight system based on separate, Open Source components (Solr7, OAICat8). A dedicated server was established at GBIF Secretariat for hosting the GBIF metadata system. The catalogue is available at:

---

5 http://code.google.com/p/gbif-geospatial/source/browse/#svn/trunk/ogc-processing/script
6 http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=2d85d0e8c76408129024c072d6
7 http://lucene.apache.org/solr/
8 http://alcme.oclc.org/oaicat/
http://metadata.gbif.org/. It provides a search interface (simple and advanced) to the harvested metadata. Metadata is stored in the catalogue in its original, received format (EML, DIF, DC and ISO19139 are currently supported) and mapped to a common search model.

Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)\(^9\) was chosen as the standard network protocol for sharing metadata between online catalogues. This enabled harvesting of metadata from participating GBIF catalogues via OAI-PMH and forwarding of aggregated metadata to the EuroGEOSS broker (see: http://metadata.gbif.org/catalogue/OAIHandler?verb=ListMetadataFormats).

The GBIF Registry is under continual enhancement to enable integration of the metadata catalogue with the other main components (datasets, taxon names) of the GBIF portal.

**Metadata formats**
A GBIF metadata profile based mainly on Ecological Metadata Language (EML)\(^10\) was developed to better support community needs, especially for species names datasets (e.g. checklists) and natural history collections, and for multiple natural languages. A metadata editor was incorporated in the GBIF Integrated Publishing Toolkit (IPT)\(^11\) to enable data publishers to prepare and publish metadata conforming to the GBIF EML profile. However, metadata is accepted in any standard format from participating catalogues and the GBIF central catalogue provides a transformation service between metadata formats (from EML, DIF, DC to ISO19139) to provide biodiversity metadata in ISO19139 format to the EuroGEOSS broker.

**Metadata content**
Promoting the publication of high quality and complete metadata remains a challenging task for GBIF (as for many other networks). As an incentive, GBIF issued a number of awards (GBIF Small Grants for Metadata Catalogues\(^12\)) to encourage uptake/development of metadata catalogues amongst GBIF Participants and their connection to the GBIF network and on to EuroGEOSS. This resulted in small numbers of metadata but was deemed a success in raising awareness in the community of the need for metadata. It was possible for GBIF to auto-generate a minimal EML metadata document for each of the 10,000+ data resources in the GBIF data cache based on an analysis of the data (e.g., to deduce temporal, spatial and taxonomic coverage) and other information provided by the provider serving the data to GBIF. In addition, recognising the need to educate and provide guidance on best practices, GBIF released a suite of tools and associated documentation for publishing data and metadata (http://www.gbif.org/informatics/discoverymetadata/publishing/).

### 2.4 Provision of important bird areas and species range maps by Birdlife International
The primary role of BirdLife International within WP4 has been that of a data provider. The original plans for BirdLife International were to provide a range map for each of the 10,000 species of bird. Prior to the EuroGEOSS project, BirdLife had developed maps for 7,000 species (including the 2,000 globally threatened species).

The development of these range maps is based on a methodology designed by BirdLife using a variety of data sources to produce a range map that broadly equates to the known distribution of a species. Digitization of source data from base layers, bird atlas data, journals and reports as well as harmonization of the species range map data with the network of Important Bird Area data and

---

\(^9\) http://www.openarchives.org/pmh/

\(^10\) http://knb.ecoinformatics.org/software/eml/

\(^11\) http://code.google.com/p/gbif-providertoolkit/

\(^12\) http://www.gbif.org/communications/news-and-events/showsingle/article/nine-metadata-grants-awarded-by-gbif/
delivery of these were all expected areas of work within the various streams of the EuroGEOSS. There was also the intention to provide these data via web services (WFS, WMS) directly from a server hosted externally to BirdLife.

BirdLife Internationals initial objectives were to:

1. Complete a comprehensive global geospatial dataset of the distribution of the world’s birds;
2. Provide the EuroGEOSS partners with access to these data via web services for integration into the Digital Observatory of Protected Areas (DOPA);
3. Develop the functionality to display these species range maps for all species on the BirdLife website (Datazone);
4. Develop the process and functionality for provision of range maps to third parties via the BirdLife website for their own analyses;
5. Train and educate partners in information management techniques and applying GIS to their workflow

The expected benefits for biodiversity assessment of this work are the improved knowledge of the distribution of individual species and the increased availability of this information to decision makers and researchers. In addition to this the training of conservationists in Africa in the use of these data and various information management tools makes sure that they’re better enabled to use this data in their conservation efforts.

2.4.1 Release of species range maps and fact sheets

The digitization of 3,000 species range maps (together with quality control on attributes and geometry) required the creation of a small dedicated team to create the range maps using ArcGIS software and models together with specific datasets and databases.

Maps were produced using datasets covering altitude (DEMs) and landcover together with information on Important Bird Area boundaries, species point localities from WorldBirds and data extracted from GBIF. Some species are habitat and altitude specific and querying both habitat and DEM datasets assisted in defining the species range map, coupled with information gleaned from published journal articles and books a more refined map was produced. For more wide ranging species, only information on country distributions and the occurrence in IBAs were used to define the range map.

BirdLife now has several formats for storing the species data – 1) a single shapefile for each species stored on an internal server for ease of editing, 2) a global species file geodatabase for analytical work and 3) a PostgreSQL database for delivery of the maps via web services (objective i)

Initial investigation, development and delivery of the species to EuroGEOSS via web services (WFS and WMS capabilities) were performed using freely available and open source software, specifically UMN MapServer and GeoServer as proprietary software was deemed too expensive and unsustainable for the purposes of this project. Ultimately GeoServer was chosen over MapServer for its ease of use and configuration. (objective ii)

With the knowledge gained from provision of the range maps via web services, the BirdLife Datazone functionality has been updated to display a species range map on each species factsheet using WMS. Previously, only the globally threatened species had a map displayed on the
factsheet. These were small, static GIFs that were updated to reflect any change to species Red List status.

A more dynamic system for displaying a map for each species factsheet was more desirable. This has been achieved by linking the PostgreSQL database to GeoServer and through use of other open source software like OpenLayers as the map client to display the range maps on a factsheet. These are simple maps that can be zoomed and panned, with the range maps overlaying Google Maps backgrounds, they are a definite improvement over a static graphic. The textual species information for the factsheets are taken from the Datazone MySQL database (objective iii)

Given the completion of a global distribution dataset of the world’s birds and increasing requests to obtain the data, it was necessary to provide a method whereby external audiences mainly from academia and general public could obtain the data easily. A data access policy was created whereby BirdLife International provides access to the dataset via a registration process on the BirdLife website. Once the user has provided personal details and an intended use of the data, we can then let them download the data via a web link (objective iv)

BirdLife International together with staff from the RSPB provided the participants from 7 African Partners (Botswana, Burundi, Egypt, Malawi, Rwanda, South Africa and Zambia) with a week-long information management workshop in Rwanda in 2010.

Participants had skills of varying abilities and the course covered many aspects of data management that they needed to help with their work, from surveying and monitoring of Important Bird Areas, to using the World Bird Database, entering monitoring data and producing reports and analyzing the data in Excel. GIS training was also provided to show the value of integrating GIS within the workflow (creation of maps for inclusion into reports), as well as the importance of creating data from GPS units (a field trip was undertaken whereby the participants could practice using a GPS and the knowledge taken from the workshops) (objective v)

A main obstacle encountered was how to secure a WFS layer using credentials such as username and password and still make this useable to an external user within a web based framework. Although this is easily accomplished using GeoServer with data held locally, it caused a few problems for externally accessing the data via web services. In response to this, we have removed the security for the species maps. BirdLife International would still like some sort of registration process to be implemented before wider access to the maps can be permitted. This enables us to track and monitor data requests and the uses and potential misuses of the data.

The key benefits from this work relate to the provision of all species range maps data to the EuroGEOSS systems and processes, and to the provision of maps into the BirdLife Datazone. Figures 10 and 11 provide an illustration of a single species WFS output from the EuroGEOSS range maps database.
Figure 10. Illustration of a single species WFS output from the EuroGEOSS range maps database:

The following Figure 11 illustrates the same species but as a WMS:

Figure 11. Illustration of a single species WMS output from the EuroGEOSS range maps database
A good practical example of the range maps being used within the EuroGEOSS project has been the work undertaken on eHabitat as described in section 4.

The main obstacles encountered during the project have related to the provision of the data via web services. Although we have the services available, the dataset is much too large (nearly 3 gigabytes of data) to be delivered as a global dataset using web services without crashing the server. It is possible, however, to query and retrieve subsets of species without overburdening the server.

Another major obstacle being the unresolved issue of provision of this data to the EuroGEOSS partners using credentials such as username and password. Initial efforts using the basic security functionality through GeoServer resulted in problems accessing the data via the external server birdlaa5. Subsequently we have dropped the security and access to the data is now “freely” available. This is not the ideal as BirdLife International have a data requests policy such that we monitor who is accessing the data and for what purpose.

Enhancing the BirdLife Datazone functionality to display a species range map on each species factsheet has been facilitated through knowledge sharing from helpful staff at the JRC. This has been an excellent development for disseminating a global atlas of species ranges to a wide external audience. Feedback from the audience will be useful in improving the species range maps over time. The following (Figure 12) provides an illustration to the updated Datazone factsheets:

![Screen capture of Birdlife’s datazone factsheet](image_url)
2.5 Increased data collection by RSPB and validation of modeling scenarios by the RSPB

76 key datasets for UK birds have been digitized and provided to the NBN, 63 of which have been included in GBIF (the difference is due to some datasets having a high degree of sensitivity). The total number of observations in these datasets is almost 3 million, a large proportion of these coming from the UK Big Garden Bird Watch Citizen Science project. (objective i)

Grants were provided to two African countries (Zimbabwe and Zambia) to carry out data digitization into Southern Africa Birds. The Zambian Ornithological Society entered over 18,000 bird observations, while BirdLife Zimbabwe entered over 220,000 observations. (objective iii)

An intermediate data warehouse for observations from the 32 WorldBirds systems spread globally was successfully developed. This warehouse was an adapted version of the WorldBirds software. With Cornell Laboratory for Ornithology, a protocol was developed and tested to import data from this warehouse to the Avian Knowledge Network (AKN) for inclusion in the AKN to GBIF regular update. By passing the data this way, two databases could be updated in one go. WorldBirds data currently number over 3 million observations from over 100 countries.

Prior to the EuroGEOSS project, UK IBA data had been updated at the RSPB and new boundaries provided to BirdLife International. However, due to internal issues, the latest assessments and monitoring data had not been included. During the last months of the EuroGEOSS project, data are being released publicly for inclusion in the Data Zone, along with any boundary amendments for Important Bird Areas (sites).

a. Web services
The WorldBirds system was rolled-out across Africa, with two countries having independent systems (Botswana and Kenya) and the rest assembled as regional hubs (North Africa, East Africa, West Africa, Southern Africa, Indian Ocean Islands). All African nations, with the exception of the UK Overseas Territory of St Helena, are included. Common Bird Monitoring (CBM) functionality was developed for three methodologies: transect counts, point counts and timed counts. To date, Botswana and Uganda are using the CBM screens for their yearly projects, with Portugal in Europe doing a yearly survey having incorporated 6 years of data, and several other countries are running trials, including Burkina Faso, Kenya, Mauritania, Nigeria and Rwanda. BirdLife Botswana have successfully used their participation in the CBM program to leverage funds from government for annual survey work. (objective ii)

b. Citizen Science
A large-scale analysis of Citizen Science data was carried out, using Danish data (comparing data from DOFBasen, the WorldBirds component, and the systematic Danish Common Bird Monitoring scheme) and British data (comparing Citizen Science BirdTrack data with the UK Breeding Bird Survey). Preliminary findings were presented at the 18th European Bird Census Council conference in Cáceres in 2010, and a paper has been submitted for publication (review yet to be received).

With BirdLife International staff, participants from 7 African Partners were trained in a week-long workshop in Rwanda in 2010. The course covered all aspects of data management, from surveying and monitoring, to Common Bird Monitoring, to participation in Citizen Science.

Throughout the EuroGEOSS project, support and training has been provided to BirdLife Partners participating in the Common Bird Monitoring programme, both face-to-face during scheduled project trips and via email/telephone.
3 ENHANCED MODELLING CAPABILITY DEVELOPED BY WP4 FOR EUROGEOSS

The number of interoperable research infrastructures has increased significantly with the growing awareness of the efforts made by the Global Earth Observation System of Systems (GEOSS). One of the Social Benefit Areas (SBA) that is benefiting most from GEOSS is biodiversity, given the costs of monitoring the environment and managing complex information, from space observations to species records including their genetic characteristics. But GEOSS goes beyond simple data sharing, to encourage the publishing and combination of models, an approach which can ease the handling of complex multi-disciplinary questions. In the following section, we will illustrate these concepts by describing eHabitat, a basic Web Processing Service (WPS) for computing the likelihood of finding ecosystems with equal properties. When chained with other services providing data on climate change, eHabitat can be used for ecological forecasting and becomes a useful tool for decision-makers assessing different strategies when selecting new areas to protect. Further enhanced by a discovery broker service, eHabitat can access virtually any kind of thematic data that can be considered as useful when defining ecosystems and their future persistence under different climatic or development scenarios.

3.1 A short introduction to ecological modeling with eHabitat

3.1.1 Using Mahalanobis distances for computing ecological similarities

In this section we describe and illustrate the mathematical method underlying eHabitat, and in section 4.1.2 we describe how this algorithm may be exposed as a Web service. The main idea behind eHabitat is to provide a service allowing end-users to find areas that have similar ecological properties to a reference location. This approach is typically used for ecological niche modeling in which a spatial prediction model for a given species is computed from a set of environmental parameters (see e.g. Clark et al., 1993, Knick and Dyer, 1997, Rotenberry et al., 2006). In this context, Geographic Information Systems (GIS) have proven to be very useful tools for conservation because of the ease of handling various thematic layers and using multi-criteria decision trees for extracting information. To compute the probability of habitat similarity (PoHS) to a reference location for each pixel of the domain under study, a popular approach is used based on the Mahalanobis distance (Mahalanobis, 1936). Numerically, the covariances and the variances of the set of (ecological) indicators within the set of pixels define how much the vector can deviate from the average and still have a high suitability. For a pixel \( i \) the Mahalanobis distance \( D \) is defined as:

\[
D_i^2 = (X_i - m)^T C^{-1} (X_i - m) \tag{1}
\]

where \( X_i \) is the vector of indicators from this pixel, \( m \) the vector of the mean values and \( C^{-1} \) the inverse covariance matrix of the indicator variables at the pixels of interest. The use of the inverse of the covariance matrix makes the Mahalanobis distance scale-independent, i.e., it is not affected by the different scales of the measurements. Because of the use of the inverse covariance matrix, highly correlated indicators will have less individual effect on \( D \) than uncorrelated indicators which could be considered to be more salient in the characterization of the region of interest. When the indicators used to generate the mean vector and covariance matrix are normally distributed, then \( D_i \) is distributed approximately according to a \( \chi^2 \) distribution with \( n-1 \) degrees of freedom, and so we can convert \( D_i \) into probability values (p-values), ranging from 0.0 representing no similarity to 1.0 for areas which are identical to the mean of the PA. If the indicators are not normally distributed, the conversion is still useful as it rescales the unbounded \( D \) values to a 0.0 to 1.0 range. This p-value can be seen as the probability that a pixel outside the investigated area has a
similar set of indicators to those found to the selected area; i.e., it can be interpreted as representing the similarity between that pixel and the indicators in the reference area. Figure 13 illustrates the use of Mahalanobis distances for identifying areas that are presenting ecological characteristics similar to those found in a protected area, the Kafue National Park, in Zambia. A set of ecological variables are used as input data and a map of probabilities to find similar values taken by these variables can be generated.

Figure 13. Map of habitats that are similar in the Zambezian ecoregion to the protected area of Kafue in Zambia, and associated scale of similarity.

### 3.1.2 Ecological forecasting with eHabitat

In the following example, we will illustrate the method described above by means of a basic case study mapping the likelihood of finding similar climatic conditions to those found in a selected Protected Area.

Three key variables are used here to characterize the climatic properties of the site:

- the bio-temperature (the annual average of the temperature after values below freezing to zero);
- the total annual precipitation derived from a mean value of total annual rainfall computed over a number of years;
- the ratio between the annual potential evapotranspiration (PET) and the total annual precipitation.

These three variables are actually those used by Holdridge (1947) to define life zones, i.e. areas with same biological characteristics. Depending on the relative values of the three variables, a site can be approximately classified within one of 38 defined classes (e.g. tropical rain forest, boreal desert, warm temperate dry forest, etc.). For the case illustrated here, we derived the three climatic...
variables from the WorldClim (Hijmans et al., 2005) data base\textsuperscript{13} which provides gridded maps of current and future climate variables at different lat-long resolutions, i.e., 10 minutes, 5 minutes, 2.5 minutes and 30 arc seconds (Figure 14).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig14.png}
\caption{Holdridge’s lifezones as a function of different climatic variables and their use for mapping bioclimatic similarities of a selected area, as illustrated here for the Nanda Devi Biosphere Reserve in India.}
\end{figure}

\textsuperscript{13} http://www.worldclim.org
The dataset for the current climate is produced by interpolating the records from climate stations with a spline interpolation method. Note that the PET was obtained using the equation of Thorntwaite (1948); the equation is simple and frequently used when dealing with large scale computations.

In the following, we show the results of the computation of bioclimatic similarities for an area between Spain and France hand drawn by the end user of a specific web client designed for this purpose\(^\text{14}\). Figure 15 shown the probabilities to find similar ecological conditions using today’s climate (top of Figure 15) and the forecasted conditions predicted for 2050 (bottom of Figure 15). The resolution of the outputs shown here is a 2.5 min grid. By computing such maps for two dates and contrasting the results with those observed today, one can 1) identify the areas with similar current climatic conditions/habitats to those found on the selected site 2) compare these similarities with those obtained with forecasted data, (in this case for 2050).

The bottom screen shot of Figure 15 shows that there is a significant loss of continuity in the climatic conditions, potentially leading to a fragmentation of the habitat.

The approach presents a number of limitations. Ecologically, when the monitored area presents a complex set of highly variable environments, such as a mountain near a lake, computing statistics over the heterogeneous environments which result does not make much sense. There are however a number of ways to go around these problems, for example by carefully stratifying the area into more homogeneous environments. Environments which exhibit particularly low variability within a habitat may also create some numerical challenges for the interpretation of the results. Still, these obstacles are intrinsic to the algorithmic implementation and do not jeopardize the broader idea of the eHabitat service as an elementary component designed to be used for multi-purpose modeling on the internet. It can easily be seen that the approach has huge potential for further development on 3D datasets, which will allow for the modeling of more complex environments like marine protected areas.

\(^{14}\) http://ehabitat-wps.jrc.ec.europa.eu/ehabitat/
Figure 15: Screen capture of a Web graphical user interface of eHabitat showing the probabilities of finding similar combinations of the climatic variables of Holdridge (bioT, Precipitation, PET) to those found today in the orange polygon designed by the end-user in the year 2000 (top figure) and in 2050 (bottom figure). Blue colors show high similarities with current conditions, red and yellow colors show, respectively, medium and low similarities with current conditions.
3.1.3 **Ecological niche modeling with eHabitat**

Species distribution models (SDMs) are typically used for prediction of the potential habitat of a species, based on observations of the species and a set of environmental indicators that are assumed to describe the niche of the species. There is a range of such models used in ecology (Guisan and Zimmerman, 2000) and a relatively common method is based on the Mahalanobis distance to create environmental suitability maps (ESM) (Clark et al., 1993; Knick and Dyer, 1997; Rotenberry et al., 2002). Another method is the MaxEnt method (Phillips et al., 2006) which is based on the creation of pseudo-absence locations. Common to these and other models is that they can be used to generate suitability or similarity maps which can help in identifying regions where a given species is more likely to be observed. These models can therefore be also used to define those locations where species could migrate should their current habitats become unsuitable because of human activities leading to, for example, deforestation or climate change. In the ecological world, many services are Web Map Services, where maps can be overlaid in a web client. These services are, however, usually deployed for improving the visualization of the environmental context of the analyzed information rather than for delivering new products which can be further processed.

Building on the ecological forecasting scenario described in the previous chapter, one can easily imagine the potential offered by such a habitat modeling service for the ecologists which could be chained with other services offering ancillary information on predicted flooding, population and agricultural pressures or land prices. A first effort in this direction has been described by Best et al. (2007) with OBIS-SEAMAP\(^1\), the Ocean Biogeographic Information System - Spatial Ecological Analysis of Megavertebrate Populations, a spatially referenced online database, aggregating marine mammal, seabird and sea turtle observation data from across the globe. Another milestone was the setting up by Nativi et al. (Nativi et al., 2009, GEO AIP-2) of an ecological niche modeling framework built around the popular OpenModeller\(^2\) (Muñoz et al., 2011). This modeling framework successfully employed a service oriented architecture even though at the time OpenModeller was still a stand-alone application, by making the modeling kernel accessible through external interfaces like SOAP and SWIG (Simplified Wrapper and Interface Generator).

There are some challenges in using species distribution modeling for predicting the effects of climate change (Sinclair et al., 2010) that we will not cover here.

3.1.3.1 **Data and associated Web Services**

**Important Bird Areas (IBAs)**

A site is recognized as an Important Bird Area (IBA) only if it meets certain criteria, based on the occurrence of key bird species that are vulnerable to global extinction or whose populations are otherwise irreplaceable.\(^3\) IBAs are identified by Birdlife International, an international partnership of conservation organizations focused on the protection of birds. IBAs are not always parts of the existing network of protected areas, but recognition of a non-protected area as an IBA will usually increase its priority for being protected in future conservation plans. By 2009 around 11,000 IBAs in 130 countries had been identified.

Boundaries of IBAs are now accessible through a web service put in place by Birdlife International. This makes it easy either to download them through a web client, but also accessible for other web services, such as the one we will describe below.

---

\(^1\) [http://seamap.env.duke.edu/](http://seamap.env.duke.edu/)


\(^3\) [http://www.birdlife.org/action/science/sites/](http://www.birdlife.org/action/science/sites/)
Species ranges

In addition to the IBAs, Birdlife International also serves a collection of predicted species range maps. These have been developed from a variety of data sources, including up-to-date and historic observations and literature ranges. The species maps are offered as polygons that are separated by season (breeding, non-breeding, passage, resident), origin (native, reintroduced, introduced, vagrant) and presence (extant, probably extant, possibly extant, possibly extinct, extinct post 1500) (Figure 16).

![Species range maps](image)

Figure 16. Screen capture of one of the WP4 web client using eHabitat WPS showing the overlay of species distribution maps provided by Birdlife International. The maps distinguish areas of seasonal occurrence (resident, breeding, non-breeding, passage).

Species occurrences

There are many organizations that are observing species around the world. BirdLife International Partners are involved in co-coordinating this work in many countries. For species observations, however, the best internet provider is the Global Biodiversity Information Facility to which the RSPB and BirdLife International also provide their data. In addition to the search/browse interface on the GBIF Data Portal (http://data.gbif.org) which allows a user to construct and submit complex, filtered queries to the data cache, several REST-based web services for machine to machine access are available (http://data.gbif.org/tutorial/services). For example, the Occurrence service (http://data.gbif.org/ws/rest/occurrence) can return records for a taxon occurring within a particular geographic bounding box while the Occurrence density web service...
EuroGEOSS, a European approach to GEOSS
FP7 Project nr 226487

D4.8. Assessment and Validation Results of the Biodiversity Operating Capacity developed in EuroGEOSS

(http://data.gbif.org/ws/rest/density) provides counts of occurrence records by one-degree cell. Output formats include the international KML (Keyhole Markup Language) OGC standard used by the popular Google Earth application as well as other mapping systems.

Finding information about a species relies on using its correct name. This is, however, not straightforward as many species may be known by multiple scientific and vernacular names. At the same time, the same name may refer to different species. These and other issues affect accessibility and interoperability when dealing with species information. While a complete solution to this is not available, GBIF, however, in collaboration with many others, including the Encyclopedia of Life (http://eol.org/) and the Catalogue of Life Partnership (http://www.catalogueoflife.org/), is developing the Global Names Architecture (GNA, http://www.gbif.org/informatics/name-services/global-names-architecture/), an informatics infrastructure and associated standards for providing unified discovery and access to information about taxon names, thereby enabling the development of a taxonomic "backbone" to underpin biodiversity informatics.

3.1.3.2 The case of the Black Harrier (Circus maurus)

In the following, we will illustrate the use of the above described web services for modeling the possible impact of climate change on an African bird, the Black Harrier (Circus maurus).

From the IUCN web site of red list species, one will find for Circus maurus¹⁸ that it is restricted to southern Africa, where it is concentrated in the Western Cape (its core range), and occurs in the Eastern Cape, the Northern Cape and Free State (where it is irruptive in both areas), in South Africa, and is also found in Botswana and Lesotho (non-breeding birds), with a tiny isolated population in northern Namibia (less than 50 birds including about five pairs). An unknown proportion migrates between South Africa and Namibia. It is widespread and can be locally common within its breeding range, with high concentrations of breeding pairs (up to 10 pairs/0.7 km²) at suitable sites, such as West Coast National Park in South Africa. The total population is estimated at 1,000-1,500 individuals, although the number of mature individuals is likely to be less than 1,000. Although there has been some debate on its status and trends, with breeding birds now known to irrupt into grasslands in some areas, populations have probably been stable overall and its range has not changed markedly during the 20th century. Recently, the population is thought to have experienced a slight decline due to the loss of some patches of renosterveld vegetation in Overberg. Future declines may occur in reaction to a predicted decrease in rainfall in the western parts of its range.

In using the web client for ecological niche, querying for the Black Harrier, one will find not only a link to the IUCN description but also the species range maps from BirdLife and 92 occurrences for his species provided by GBIF (Figure 17).

¹⁸ http://www.iucnredlist.org/apps/redlist/details/106003406/0
Figure 17. Screen capture of the ecological niche modeling web client using eHabitat WP showing the species distribution map (red) of *Circus maurus* in Southern Africa. Web2.0 will provide links to image archives of the bird. GBIF occurrences are indicated with yellow squares (hidden here by the range maps).

3.1.3.3 Bioclimatic area of the Black Harrier (*Circus maurus*)

We used here again the HADCM3 climate model and the eHabitat WPS for computing the Mahalanobis distances using the Holdridge data obtained at the 96 locations where GBIF reported the bird species. Figure 18 shows, as expected, a good matching between the species distribution model and the map of probabilities to find similar bioclimatic conditions using this time the information found at the locations of the species occurrences rather than over a selected area. The figure shows also the predicted locations for 2080 to find the current Harrier’s bioclimatic area and a dramatic loss of similar bioclimatic conditions can be observed.
Figure 18. Screen capture of the ecological niche modeling web client showing the species occurrences (red points) of *Circus maurus* as reported by GBIF overlaid on the output of the bioclimatic map of similarities. Dark blue areas are more alike to the conditions of the bird’s habitat than pink and orange areas. The upper figure shows the current conditions, the one below shows the probabilities to find similar conditions in 2080.
3.1.3.4 Use of third party web services for decision making

By further accessing other data available through web services, more information potentially useful for decision making can be accessed and incorporated with the other sources of information. Using a population density map as shown in Figure 19, one will quickly realize that possible mitigation strategies against climate change will be difficult in the concerned areas. Areas suitable in the future for black Harrier are overlapping with areas showing a high population density.

![Figure 19](image_url)

Figure 19. Screen capture of the ecological niche modeling web client showing probabilistic species distributions of *Circus maurus* for the year 2080. These maps are overlaying a map of human population density provided by CIESIN. Dark brown areas show areas with high population density, yellow show areas with low population densities.
3.1.4 eHabitat WPS and the EuroGEOSS brokering approach

The Group on Earth Observations (or GEO) is coordinating international efforts to build a Global Earth Observation System of Systems (GEOSS) (GEO, 2009). The aim of GEOSS is to build a public infrastructure to link together existing and planned observing systems around the world and support the development of new systems where gaps currently exist\(^\text{19}\). The infrastructure that coordinates access to the systems, applications, models, and products is the GEOSS Common Infrastructure (GCI). To demonstrate the added-value of GEOSS and enhance the GCI functionalities, the GEO Architecture and Data Committee (ADC) launched the GEOSS AIP Initiative\(^\text{20}\).

In December of 2010, the third phase of GEOSS AIP (AIP-3) was concluded: it developed scientific scenarios for several of the Societal Benefit Areas (SBAs) recognized by GEOSS; also cross-disciplinary pilots were considered. The Biodiversity and Climate Changes domain is part of them. Due to the multidisciplinary nature of this domain, the achieved pilots required the setting up of a multidisciplinary infrastructure. For GEOSS AIP-3 Biodiversity & Climate Change pilots, one of the objectives was to continue the successful experimentations developed by GEOSS in the framework of IP3 and the two previous AIP phases (Nativi et al, 2009).

The following describes a use scenario based on the eHabitat model utilizing a distributed discovery service (i.e. a Discovery Broker) for accessing Biodiversity and Climate Change datasets. Thus, end-users can select the ingredients available on the Internet for modeling the PoHS, an obvious enhancement to the existing modeling capacity of eHabitat 1.0.

The scenario architecture of eHabitat in AIP-3 included the following advanced components developed by the two EC-funded projects:

1) the EuroGEOSS Discovery & Access Broker services;
2) the EuroGEOSS/GENESIS Semantic Discovery Broker which extends the Discovery Broker by underpinning semantically-enabled queries;
3) WorldClim data served through an OGC WCS interface (developed by EuroGEOSS) to allow assessing the impact of changing climatic variables in protected areas;

The eHabitat scenario architecture benefits from the SOA brokering approach (developed in the frame of EuroGEOSS) by implementing the "Catalogue" service through a Discovery Broker which is further coupled with another pair of effective components, the Access Broker and the Semantic Discovery Broker (Figure 20).

![Figure 20. The brokering framework used by the AIP-3 eHabitat pilot -and developed by EuroGEOSS](http://www.earthobservations.org/geoss.shtml)

\(^{19}\) [http://www.earthobservations.org/geoss.shtml]

\(^{20}\) [http://www.earthobservations.org/geoss_call_aip.shtml]
The broker implements an extended SOA approach. This Brokering SOA (Nativi and Bigagli, 2009) is depicted in Figure X and the above indicated components further discussed hereafter.

3.1.4.1 Discovery Broker

Referring to Figure 20, the AIP-3 framework builds on a Discovery Broker with GeoRSS support. It allows interaction with a plethora of heterogeneous services and data models characterizing multi-disciplinary scenarios. Moreover it serves to lower the present GCI entry-barrier by providing users with a homogeneous discovery framework to heterogeneous resources (biodiversity, climate change, etc.) through the addition of “expert” brokering services which hide the heterogeneity of the underlying systems (Figure 21). This solution prevents the eHabitat client from having to “learn” and implement many heterogeneous (and sometimes immature) information technologies.

Figure 21. the extended SOA archetype (Brokering SOA) introduces a new component: the broker which is in charge of implementing all the mediation and binding tasks which are currently the responsibility of the service consumer.
3.1.4.2 Discovery Augmentation Broker (Semantic discovery augmentation)

The Discovery Broker is extended by a Semantic Discovery Broker—a "smart" service which adds semantic intelligence to queries (Santoro et al., 2011). This additional service allows for improved means of finding potential input data by "relaxing" the constraint of using exact keyword match for queries. The Semantic Discovery Broker expands the original query submitted by the user with a set of additional keywords. Such keywords are retrieved from a set of semantic repositories (e.g. controlled vocabularies, gazetteers, etc.). Using the Semantic Discovery Broker it is possible to start a search with the "habitat" keyword, and complete the search finding datasets which match "air temperature" and "atmospheric precipitation".

The query expansion strategy can be also user-assisted; in this case the user can navigate the network of concepts available in the federated semantic repositories and select the concepts of interest for the query as shown in the screen capture of the Graphical User Interface (GUI) that was developed for testing and interacting with the Semantic Discovery Broker (Figure 22).

![Figure 22: Snapshot of the GUI developed for testing interactions with the Semantic Discovery Broker.](image)

The Semantic Discovery Broker implements a workflow engine to:

i) browse semantic repositories;

ii) retrieve concepts of interest, expanding query terms;

iii) formulate a set of catalogue queries, one for each recognized concept (multilingualism\(^{21}\) is supported, as well);

iv) forward the generated set of queries to the Discovery Broker and search for resources (i.e. datasets and models) related to such concepts;

\(^{21}\) The number of supported languages depends on the utilized semantic repositories
v) collect matching results (i.e. data and models along with the related services to access them).

In the AIP-3 demonstration, the Semantic Discovery Broker federates the following services:

- GENESIS SKOS Repository: this repository publishes an alignment of three thesauri: GEMET, GEOSS SBA, and INSPIRE Themes;
- EuroGEOSS Discovery Broker.

### 3.1.4.3 Access Broker

Implementing the Brokered SOA approach, the EuroGEOSS project prototyped a flexible Access Broker component. It provides access to normalized data, making use of existing data pre-processing services, and can be easily extended to accommodate future ones, when available. The system architecture was designed in order to satisfy the following main requirements:

i) the framework should not supplant but complete existing access systems/services;

ii) support the existing Community of Practices (CoPs) in using their own pre-processing components/services; and

iii) be compliant with INSPIRE transformation services implementing rules.

As a practical example of the Access Broker use, we can consider the “Common Environment Data Access” use case. Frequently, in order to exploit discovered data (e.g. to ingest data into a modeling software), these must all be on a common environment – that is, all data must have the same format, CRS, spatial extent/resolution. In heterogeneous environments such SDIs, discovered data are typically provided on different environments and users have to download data and pre-process them to fit their needs. The Access Broker provides users with access to data on a common environment independently from their specific origin environments, moving the task of adapting data from users to system.

### 3.1.4.4 Scenario Steps

The eHabitat scenario is composed of three main phases: discovery (with semantic augmentation) of the required input datasets, access and execution of the eHabitat model fed with the discovered datasets. In this first prototype of the infrastructure we used two different web-based client applications: the Semantic Discovery Broker GUI, and the eHabitat client.

Figure 23 depicts a sequence diagram of the augmented discovery steps for the eHabitat pilot (phase one). In the AIP-3 demonstration we used the user-assisted query expansion strategy to find the input datasets for the eHabitat WPS. As depicted in the diagram, the user starts by inserting one keyword and searching which matching concepts are available on the utilized SKOS Repository; this first operation is called GetConcepts. The result is a set of concepts, related to the initial keyword, which are presented to the user as nodes of a graph. Each node can be further expanded to find related concepts. This sequence of expansions is represented in Figure 22 by the loop named Browse Concept Network.

During the browsing, the user can select one or more concepts of interest. After selecting all the concepts of interest, the user submits the query to the Semantic Discovery Broker. The submitted query is further elaborated by the Semantic Discovery Broker, which:

- uses the SKOS Repository to get all available translations of selected concepts;
- generates a set of traditional geospatial queries;

---

22 A concept is a well-defined term on a controlled vocabulary (thesaurus)
- submits the generated query to the Discovery Broker:
- collects the results and presents them to the user “clusterized” by the matched concept.

Figure 23: Sequence diagram of the augmented Semantic Discovery.

After the first phase, the user the user *drags and drops* the information – based on ISO 19115 metadata model (ISO, 2003) – about the discovered datasets from the Semantic Discovery Broker GUI to the eHabitat client (see Figure 22).

The phases two and three of the scenario (data access, ingestion and execution of the eHabitat model) is described in Figure 24. When the user submits the *Run* command to the eHabitat WPS, the model execution starts. Figure 24 depicts how the eHabitat WPS can transparently retrieve the input datasets, using the Access Broker, from different data providers. In fact, the eHabitat model requires that input layers respect all the same encoding format, CRS, spatial resolution, etc. This would constraint the user to select datasets only from remote services which are known *a priori* to
support the needs of eHabitat model. This is a very strong limitation. Otherwise, the eHabitat WPS should be able to manage a quite complex set of pre-processing steps in order to normalize the remote resources according to its needs.

Introducing the Access Broker removes the constraints above. In fact, as depicted in Figure 24, the eHabitat WPS requests input datasets to the Access Broker. It is now up to the Access Broker to access data from the remote service(s), and call appropriate transformation services (which can be defined by the user) in order to return normalized datasets to the eHabitat WPS. After the Access Broker returns the input datasets to the eHabitat WPS, this can proceed with the model computation and returns the output to the client.

Figure 24: Execution of the eHabitat model.
3.2 Multi-disciplinary modeling in EuroGEOSS: conclusions and further considerations

The Model Web is a concept for a dynamic model infrastructure composed of models interacting with one another via Web services. Its main benefits are improved means to discover, access, reuse and chain models and datasets for multiple purposes. The prototype framework described here (i.e. the AIP-3 eHabitat framework) is designed to illustrate these benefits: end-users are enabled to address more complex problems than when using models independently, repeatability of research is increased and components can be re-used for multiple purposes. The Digital Observatory for Protected Areas (DOPA) of which eHabitat is a component is becoming a platform for integrating observations and models relating to trends in the world's ecosystems, species, and genetic resources. Multi-disciplinary information integration is recognized by the scientific community as essential for the understanding of complex issues such as the response of biodiversity to global changes. This calls for the further development of flexible and scalable systems allowing integration with existing (and heterogeneous) services and data systems. At the same time, it calls for well-orchestrated and documented workflows and chains of analytical steps which apply international and disciplinary standards for achieving interoperability across different disciplinary systems and resources (i.e. data, services and models). The adoption of an extended SOA approach (i.e. Brokered SOA) realized the necessary scalability and flexibility which should in future allow relatively easy interoperability with a set of important environmental services and data systems -e.g. climatology, meteorology, oceanography, hydrology, etc. The adoption of standard Web services to publish the eHabitat service and products encourage other disciplines to use them, bringing these services and products into international initiatives, such as the EC INSPIRE and the GEO GEOSS. Multi-disciplinary interoperability requires for service and data composition rather than technical interoperability. Thus, semantic interoperability must be pursued. The experimental Semantic Discovery Broker described here demonstrated the potential of such an approach.

Because eHabitat can take a broad range of data types from uncontrolled sources, it is exposed to a many different types and levels of uncertainties. When chained to other services, eHabitat would be an additional component further propagating uncertainties from a potentially long chain of model services. This integration of complex resources, such as data and models brings ever increasing challenges in dealing with uncertainty. For future developments, we are building on the lessons learnt from the UncertWeb (www.uncertweb.org) project which promotes and develops tools and standards for quantifying and communicating uncertainty in a distributed, interoperable Model Web (Cornford et al., 2010, Bastin et al., 2011). eHabitat will adopt the open source implementations of encoding standards, service interface profiles, discovery and chaining mechanisms and open source implementations developed in UncertWEB. Our first observations have been presented in Skøien et al. (2011a, b,c).

To build on the experimental framework of the GEO-AIP3 and develop an operational infrastructure, we need to further address scalability, quality of service, and functionality obstacles; such as:

1. Computational scalability is necessary to manage and process a very large amount of datasets -e.g. high resolution remotely observed datasets. Cloud Computing seems to be a possible solution. Actually, the brokering services (i.e. EuroGEOSS Discovery, Access and Semantic Discovery Brokers) comprised by the AIP-3 eHabitat pilot already run on a Cloud Computing capacity.

2. Service chaining needs to be achieved at a high level in order to lower the present entry barrier for modelers. The Composition as a service (CaaS) approach promises to provide the
technology to implement that. Some experimentation with a CaaS implementation for eHabitat is under development in the framework of UncertWeb.

3. Too often data accessibility is intended as data download. To be genuinely useful to and usable by scientists, the discovered datasets must be accessed in a "consistent" way, for example by specifying a Common Grid Environment (CGE). Datasets defined on a CGE are characterized by a common covered spatio-temporal dimension, spatio-temporal resolution, Coordinate Reference System (CSR) and encoding format. The Brokered SOA promises to transparently deliver such functionality, either through filtering or through transformation services. A prototype\(^\text{23}\) was developed in the framework of the EuroGEOSS project.

The benefits offered by elementary Web services in a Model Web context are clear in the case described here: because of the simplicity of the service, the strength of eHabitat WPS is that it can be reused for a large number of purposes by allowing end-users to select any area of reference and to identify their own data sets as input variables which characterize the phenomenon of interest. In the examples shown here, habitats defined by biophysical layers are considered, but because there is no theoretical limitation to the number or type of variables that can be used for computing the PoHS, the potential applications of the Mahalanobis distance model are practically unlimited because there is virtually an infinite number of possible uses. There is a broad range of interdisciplinary possibilities, from socio-economic modeling and ecological forecasting to the optimization of environmental monitoring networks when the reference area is limited in space to a measuring station. Similarly, the simpler the service, the easier it is to chain it with others. An increasing granularity of the ModelWeb environment comes at a cost, however: if the flexibility of the interoperable models and data services is increased through a proliferation of modular, elementary services, the setting up of a chain of services for more complex modeling requires a very clear documentation of the existing services as well as an easy means to discover and evaluate the right service in a possible ocean of other similar services. The setting up of such a framework within which end-users can select their own "ingredients" has been successfully prototyped in the context of the GEOSS AIP (Architecture Implementation Pilot) initiative.

\(^{23}\) [http://184.73.174.89:8080/sdi-gi-dac-1.0.1/geoportal/index.html](http://184.73.174.89:8080/sdi-gi-dac-1.0.1/geoportal/index.html)
4 FUTURE DEVELOPMENTS

JRC
Benefiting from the critical data infrastructure for biodiversity that has been developed through EuroGEOSS, JRC will further expand the analytical services proposed in DOPA to allow end-users to easily access a large variety of fundamental indicators on the value of protected areas at the global scale. Figure 25 illustrates such effort by showing a carbon storage calculator for protected areas developed on the top of the other services for modeling species distributions and habitats irreplaceability. Main challenge is the provision of these services to large groups of end-users considering the heavy computation load required by each web processing service. Therefore, for the time being, modeling services of DOPA will be restricted to only a few registered users and most information pre-computed for the main web clients of DOPA considering the need to provide the biodiversity indicators for more than 100,000 protected areas.

Figure 25: Screen capture of a web client for DOPA integrating species observations provided by the web services of GBIF and Birdlife (lower left box), boundaries of protected areas provided the web services of WCMC, and outputs on habitat irreplaceability and carbon storage of a selected park as produced by the web services from DOPA.

WCMC
Work will continue on protectedplanet.net in 2012 with a strong focus on making the data behind the website as interoperable as possible. This will involve extending the existing OGC web services with complimentary application programming interfaces. Web sites will be able to embed widgets and deep link protected areas information directly on their websites. The infrastructure behind protectedplanet.net will also be abstracted to allow others to build protected areas websites using a similar codebase. As an increasing number of users access protectedplanet.net new tools designed to engage citizens in scientific endeavors will be created.
GBIF
GBIF seeks to make biodiversity information freely and universally available in support of science, society and a sustainable future. The GBIF data portal provides unified access to a global network of data publishers and provides a number of REST based web services which are proving a robust, reliable and well performing component of the DOPA network. In contrast, GBIF faces a difficult challenge to provide OGC WMS and WFS services for its point occurrence data. With over 320 million records currently and ambitions to scale into the billions, it is not possible to allow arbitrary queries of the indexed cache. One possible solution under consideration is to provide a limited number of services based on the most common queries of the data. The GBIF portal also offers geospatial data as one-degree cell density overlays for Google Earth and placemarks for Google Earth. GBIF is also interested in offering computing capacity for the mining of GBIF mediated data based on a new infrastructure built around Hadoop/Hive/HBase which offers the possibility of implementing customized, automated and repeatable work flows.

With a metadata catalogue system supporting a distributed design and based on open standards and implemented using Open Source applications now in place, the GBIF network is well positioned to continue to develop and enhance the system. Foremost, is the need to encourage the publication of high quality complete metadata. GBIF continues to address these issues through training and outreach activities and development of supporting documentation.

GBIF has just commenced its 3rd 5-year phase and has recently published the GBIF Strategic Plan 2012-2016: Seizing the Future ([http://www.gbif.org/orc/?doc_id=2792&l=en](http://www.gbif.org/orc/?doc_id=2792&l=en)) which recognizes that the core business of GBIF is “building and maintaining a global biodiversity research infrastructure to enable the free flow of biodiversity information”. As such, GBIF is committed to serving the needs of its Participants and ensuring access to metadata, data and services.

GBIF is unique as an organization with a global-level focus and mandate for the management of primary biodiversity data. Positioned at the data/science interface it has a crucial role to play in support of science-based policy directed at conserving biodiversity in support of human and planetary well-being. GBIF is a Participating Organization in GEOSS and active in GEO BON (Steering Committee; Implementation Working Groups) and can provide a data foundation for the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, [http://www.ipbes.net/](http://www.ipbes.net/)) due to its success as a global network.

BirdLife International
Future developments would be to further collaborations with EuroGEOSS / JRC to incorporate Important Bird Area (IBA) and species range maps into other projects such as DOPA and any ideas that may ensue from this.

Harnessing the web services technology utilized in the species range maps, BirdLife will display maps for the IBAs and Endemic Bird Area (EBA) datasets in a similar manner on the Datazone.

Training BirdLife partners is always beneficial especially with new developments to processes and technologies employed in providing tools to assist them with monitoring species and sites that they are found. The Rwanda workshop was a successful opportunity to teach staff how to best collect, process and analyze their data, a major obstacle however is the provision of geospatial tools to help them collect and analyze spatial data especially with proprietary software such as ArcGIS. Over time, this situation may well improve given the advances and availability of open source geospatial tools.
RSPB

The key benefits from this work relate to the provision of raw data to the EuroGEOSS systems and processes. Both systematically collected data (e.g. structured surveys such as Common Bird Monitoring) and ad hoc observer records add to our knowledge of the natural environment and how it is changing, and without base datasets, complex systems and modeling are of little value.

The main obstacle encountered during the project has related to the link between RSPB and AKN, and AKN and GBIF. There are currently issues with the transfer of data, and other options (such as direct provision of RSPB data to GBIF) are being investigated and will be implemented regardless of whether or not the EuroGEOSS project is ongoing.

In the future, the RSPB will continue to provide data from the WorldBirds project to initiatives such as the AKN and GBIF, for ongoing transmission to biodiversity services, and will continue to work with BirdLife Partners to facilitate the digitization of recent and historic bird observations.
REFERENCES


International Conference Environmental Informatics”. W. Pillmann, S. Schade and P. Smits (Eds), pp: 46-58, 5-7 October 2011, Ispra, Italy


Skøien, J.O., G. Dubois, J. De Jesus (2011c). Forecasting biomes of protected areas. Procedia Environmental Sciences 7: 44–49


Furthermore, this report has been based on a number of presentations made at international conferences and workshops and published in the papers indicated below. Readers may find in this material additional information. The list below is indicating only the main dissemination events and the main web sites and services developed for EuroGEOSS

Main Web sites and services

- EuroGEOSS and the Biodiversity WP4: http://www.eurogeoss.eu/about/Pages/ WP4.aspx
- The Digital Observatory for Protected Areas (DOPA): http://dopa.jrc.ec.europa.eu/
- The GBIF Data Portal: http://www.gbif.org
- Protected Planet: http://www.protectedplanet.net/
- Birdlife Important Bird Areas: http://www.birdlife.org/datazone/site
- eHabitat modeling service: http://ehabitat://jrc.ec.europa.eu/