D.4.1: report on user requirements for the EuroGEOSS biodiversity operating capacity

Title	D.4.1: report on user requirements for the EuroGEOSS biodiversity operating capacity
Creator	Éamonn Ó Tuama, Gregoire Dubois, Andrew Cottam, Ian May, Ian Fisher
Creation date	06/11/2009
Date of last revision	21/12/2009
Subject	Interoperability, User Requirements, Initial Operating Capacity
Status	☐ Draft ☒ Final
Publisher	EuroGEOSS
Type	Text
Description	D.4.1 Report
Contributor	Max Craglia, Lorenzino Vaccari
Format	PDF
Source	
Rights	☒ Public ☐
Identifier	D4 1-v08
Language	En
Relation	WP4, related to WP1, WP2, WP3
Coverage	Not applicable

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

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

TABLE OF CONTENTS

TABLES

FIGURES

ACRONYMS AND ABBREVIATIONS

Abbreviation	Name
ABCD	Access to Biological Collections Data
AEGOS	African-European Georesources Observation System
AJAX	Asynchronous JavaScript And XML
AML	Arc Macro Language
APAAT	African Protected Areas Assessment Tool
BDP	Biological Data Profile
BioCASe	Biological Collections Access Service protocol
CIESIN	Center for International Earth Science Information Network
CMS	Convention on Migratory Species
CNR	Consiglio Nazionale delle Ricerche - Italy
CSW	Catalogue Service for the Web
DiGIR	Distributed Generic Information Retrieval
DOPA	Digital Observatory for Protected Areas
EBA	Endemic Bird Area
EC-JRC	European Commission-Joint Research Centre
EML	Ecological Metadata Language
ENM	Ecological Niche Model
FGDC	Federal Geographic Data Committee
GBIF	Global Biodiversity Information Facility
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GML	Geography Markup Language
GNA	Global Names Architecture
GROMS	Global Register of Migratory Species
HRI	Habitat Irreplaceability Index
HTML	HyperText Markup Language
IBA	Important Bird Area
IUCN	International Union for the Conservation of Nature

ISO	International Organization for Standardization
KML	Keyhole Markup Language
LGTG	LSID GUID Task Group
MDG	Millennium Development Goals
MIFTG	Metadata Implementation Framework Task Group
NASA	National Aeronautics and Space Administration
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
OGC	Open Geospatial Consortium
PA	Protected Area
RSPB	Royal Society for the Protection of Birds
SOS	Sensor Observation Service
SRI	Species Irreplaceability Index
TAPIR	TDWG Access Protocol for Information Retrieval
TDWG	Taxonomic Database Working Group
UNEP-WCMC	United Nations Environment Programme - World Conservation Monitoring Centre
UN-MDG	United Nations - Millennium Development Goals
USGS	U.S. Geological Survey
VMAP	Vector Map
WBDB	World Biodiversity Database
WCS	Web Coverage Service
WCS-T	Web Coverage Service - Transaction
WDPA	World Database of Protected Areas
WFS	Web Feature Service
WFS-T	Web Feature Service - Transaction
WMS	Web Map Service
WPS	Web Processing Service
WRI	World Resources Institute
WWF	World Wildlife Fund
XML	eXtensible Stylesheet Language

1   INTRODUCTION

This document is Deliverable D4.1 for EuroGEOSS Work Package 4 (WP4), a report on user requirements and the existing data infrastructures at partner institutions relevant to biodiversity. In collaboration with Work Package 2 (WP2) (Multi-disciplinary Interoperability), it identifies gaps in the systems currently available at these institutions in relation to data, services and models.

In Section 2, each of the five project partners – EC-JRC, GBIF, UNEP-WCMC, BirdLife International, RSPB - reports on the current status at its own institution in regard to i) current informatics systems, ii) gaps in data, models and services, and iii) user requirements, and the responses of the partners to the questionnaire *Task 2.1: Questionnaire for User Requirements, v02*, provided by WP2, are presented in a summary table.

Section 3 introduces the concept of the Digital Observatory of Protected Areas (DOPA) through building a user scenario that identifies three scenario events: simulation of a new park; use of e-Habitat; e-Habitat, climate change modelling and ecological niche modelling.

In Section 4, Annex 1 provides an overview of the indicators used in the African Protected Areas Assessment Tool (APAAT) while Annex 2 provides a list of recommendations of a task group convened by GBIF to advise on the implementation of a metadata framework for its network. These recommendations will inform deployment of the metadata catalogue system, deliverable D4.2 of WP4: an operational distributed metadata catalogue system for biodiversity data.

2   SYNTHESIS OF THE EXISTING SITUATION

In general, there is a need for all the partners to deploy standards-based web services, in particular and where appropriate, Open Geospatial Consortium (OGC) Web Services, and accompany their datasets with INSPIRE compliant metadata.

Based on the results of the EuroGEOSS initial survey, it is clear that the partners vary considerably in their requirement to deliver and consume services. Most partners either plan or already provide some form of metadata but there is a need to adopt the INSPIRE guidelines. OGC web services are better supported with some of the partners already deploying (at least at the prototype stage) Web Map and Web Feature Services (WMS, WFS) or planning to roll-out these and other OGC services: Web Coverage Service (WCS), Web Processing Service (WPS), and Catalogue Service for the Web (CSW). Data types include both raster and vector for such features as species point occurrences and polygon outlines of species distribution maps and protected areas. The main type of data encoding requirement is XML, either Geography Markup Language (GML) or Darwin Core. Geographic (Plate Carree) projection based on the WGS84 datum is the common map projection requirement. Modelling procedures are already employed by some partners but there is a need to deploy these as web services. Work flow procedures are not seen as a priority by any of the partners.

A key feature of the DOPA will be the automatic generation of indicators to characterise the state of protected areas. These indicators are based on irreplaceability in terms of species composition and uniqueness of habitat, and on exposure to pressure from human sources. The Species Irreplaceability Index (SRI) is calculated for each protected area with respect to all the species of a given taxon (amphibians, mammals and globally threatened birds). The method used accounts for both high species richness of common species but also gives an increased weight to PAs which have a high number of species found in very few other PAs (i.e. endemic or threatened species).

The Habitat Irreplaceability Index (HRI) compares the characteristic habitat of a PA according to climate, terrain, land cover and human population to the availability of similar areas within the same biome in order to generate an indicator of the irreplaceability of the PA habitat: the more irreplaceable a PA habitat is, the higher its ranking in a prioritisation scheme. Boundary pressure on a PA is quantified by calculating, for a range of buffer sizes, the percentage of human influenced pixels around the PA, and applying a weighting based on the closeness of the buffer to the PA boundary.

A description of the existing situation at each of the project partner's institutions follows.

2.1 EC-JRC

The MONDE action of the Institute for Environment and Sustainability (IES) of the Joint Research Centre (JRC) of the European Commission is monitoring natural resources in developing countries, among which protected areas in Africa. To assess the state of protected areas, efforts of MONDE have focused on monitoring habitats at the continental scale using mainly remote sensing techniques and on gathering information on species richness. By assessing the uniqueness of habitats and the level of threat to species, priorities in conservation strategies can be defined in the light of anthropogenic pressures like agricultural activities or climate change. In the frame of EuroGEOSS, the experience of MONDE in developing indicators at the continental level for protected areas in Africa will be ported to other continents.

2.1.1 Current systems

MONDE has developed the African Protected Areas Assessment Tool (APAAT), which can be considered as the first consistent, continent-wide assessment of the state of protected areas in Africa, based on (i) their value for conservation and (ii) anthropogenic pressure. More information can be found at http://bioval.jrc.ec.europa.eu/PA/ and in Hartley et al. (2007).

The APAAT is an online information system based on a Geographic Information System (GIS) and satellite-derived data developed to aid decision makers assess the state and pressure of 741 protected areas (PAs) in Africa. PAs are characterized according to indicators of their irreplaceability in terms of species composition and uniqueness of habitat, and exposure to pressure from anthropogenic sources. The preparation of these indicators followed the workflow summarized below (Figure 1)
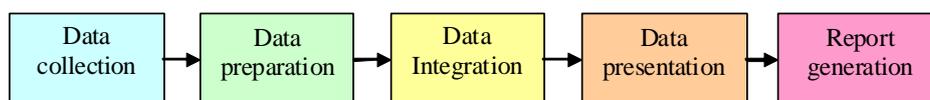
Figure 1: Flowchart of the data processing in the APAAT

2.1.2 Data collection

The APAAT largely relies on data provided by third parties. Datasets have been made available by the following organisations and projects: African Mammals Databank, BirdLife International, CIESIN, Geoland, Global Amphibian Assessment, IUCN, JRC, USGS, NASA, UNEP, UN-MDG, VGT4Africa, VMAP, WDPA, WCMC, The World Bank, WorldClim, WRI and WWF. The core data for the biodiversity indicators used in the current version of the APAAT are:

- Protected Areas from the 2006 version of http://www.unep-wcmc.org/wdpa
- Global Amphibian Assessment from http://www.globalamphibians.org/
- African Mammals from http://www.gisbau.uniroma1.it/amd/homepage.html

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

- Globally threatened birds from http://www.birdlife.org/

Data have been collected through the internet using standard downloads and/or on written requests made to the data providers. All data processed by the APAAT are stored in a PostGIS database, a geospatial extension for PostgreSQL.

2.1.3 Data preparation, integration, presentation and report generation

Much of the data preparation and analysis is carried out within a GIS, ArcGIS, using 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 accomplished in AML. In addition to the typical Unix text processing commands, the following libraries are used:
- PostgreSQL, to access and extract data in the PostgreSQL database;
- GDAL, a geographic data extraction library with command line tools for the manipulation of spatial imagery data, specifically we use gdal_translate to convert between data formats and gdalwarp to project between different map projections;
- ALCHEMY, a program to convert between graphic image formats, especially good for postscript to bitmap format (i.e. vector to raster).

Other home made command line tools are:
- Covariance.exe, to compute an inverse covariance matrix for a set of environmental variables
- Profile.exe, to compute zonal stats on a stack of variables

More information on the technical developments can be found in Nelson & Hartley (2008).

2.1.4 Current gaps in the system

2.1.4.1 Gaps is the data collection process

Although GIS are essential tools for inventorying, correlating and analyzing relationships between species and habitats, the effectiveness of conservation practices can only be assessed through regular assessments. This means that the whole process of collecting, processing and sharing data (products) needs to be automated as much as possible through the use of web services (Dubois et al;, 2009). The new system, known as the Digital Observatory for Protected Areas (DOPA), will be developed as part of EuroGEOSS.

2.1.4.2 Gaps in the data preparation, integration, presentation and report generation processes

The current version of the APAAT is suffering from the following drawbacks:

- No OGC web services are provided while much data could be distributed using WMS, WFS and WCS.
- JRC-MONDE does not currently provide a metadata catalogue of resources used and produced;
- data produced are not really reusable by others (jpg & text) as the APAAT is a web site and not an interoperable web service;
- the information made available cannot be easily found by others (absence from catalogues);
- the system needs re-engineering to become open source and independent from any software provider (the core system is depending on ArcGIS);

- a few key programs, i.e. Covariance.exe and Profile.exe, could be more efficient if replaced by codes derived from StarSpan[1] and R[2], respectively.

2.1.4.3 Current gaps in the data, services and models

2.1.4.3.1 Gaps is the data collection and services

The whole value of the APAAT is derived from the collection and the merging of a very large variety of data covering very different themes, ranging from species distribution maps to socio-economical indicators. Most of the data used for Africa by the APAAT are available on the Internet but their format often needs reprocessing and no web services allow the automatic exchange of data. Applying the concept of the APAAT to other continents would require an access to the same thematic layers but for other regions. An issue for intercontinental comparisons is therefore the harmonisation of the data between continents. We hereafter list the information currently used by the APAAT as well as the one needed to port the system beyond the African continent. In the next version of this report (due in Month 14) we will report on the extent of success in filling these gaps.

Base GIS Layers

Table 1: List of the required GIS base layers

Theme	Source	Available as Web Service
Maps of Virtual Earth	*TBC*	Yes
Country boundaries for the world	*TBC*	No
Ecoregions of the world	*TBC*	No
Boundaries of PA (IUCN classification)	UNEP-WCMC	Yes

Species Layers

Biodiversity indicators of the APAAT are based on species distribution maps defined mainly by IUCN and on the IUCN red list ranking. Gaps can obviously be found in the species mapped. No world maps of reptiles have been found. Table 2 lists the species distribution maps covering the globe.

Table 2: List of the required species world distribution maps with IUCN red list ranking

Theme	Official Source / Alternative	Available as Web Service
Mammals on the IUCN Red List (5 488)	IUCN	No
Amphibians on the IUCN Red List (6000)	IUCN	No
Birds on the IUCN Red List	Birdlife / IUCN	No
Reptiles on the IUCN Red List	NA	NA
Observation and specimen records from GBIF network	GBIF	Yes

Environmental Data

Environmental data are used mainly by the APAAT to model habitats on a 1km x 1km resolution grid. These include percentage tree cover, percentage herbaceous cover, percentage barren cover, elevation in metres, slope in degrees, aridity index, percentage water body presence, normalised difference vegetation index (NDVI), normalised difference water index (NDWI). The habitats uniqueness of each protected area has been computed only for Africa.

---

1   http://starspan.casil.ucdavis.edu/doku/doku.php
2   http://www.r-project.org/

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

Two of the indicated themes (% tree cover and aridity index) could eventually be provided by EuroGEOSS project partners. In order to provide indicators for PAs beyond the Africa continent, the Digital Observatory for Protected Areas will require new environmental datasets. In the framework of EuroGEOSS, the main efforts will focus on two regions (Africa & Europe). Note that the NDVI, NDWI and percentage of water body presence as currently used in the APAAT are derived from the Spot satellite monitoring Africa (VGT4Africa). In the framework of EuroGEOSS, the percentage of tree cover and NDVI maps are expected to be provided by web services to be put in place by JRC-FOREST, while maps of aridity are expected to be obtained from JRC-DESERT.

Table 3: List of required environmental world thematic layers available with a 1 km x 1km grid resolution

Theme	Official Source / Alternative	Continent	Available as Web Service?
% Tree Cover	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
% Tree Herbaceous cover	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
% Barren cover	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
Elevation	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
Slope	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
Aridity index	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*	• World cover • Africa • Europe • Asia • Oceania • North America • South America	• *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC* • *TBC*
% Water body presence	• *TBC* • *SpotVGT*	• World cover • Africa	• *TBC* • *TBC*

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

	• TBC	• Europe	• TBC
	• TBC	• Asia	• TBC
	• TBC	• Oceania	• TBC
	• TBC	• North America	• TBC
	• TBC	• South America	• TBC
NDVI	• TBC	• World cover	• TBC
	• SpotVGT	• Africa	• TBC
	• TBC	• Europe	• TBC
	• TBC	• Asia	• TBC
	• TBC	• Oceania	• TBC
	• TBC	• North America	• TBC
	• TBC	• South America	• TBC
NDWI	• TBC	• World cover	• TBC
	• SpotVGT	• Africa	• TBC
	• TBC	• Europe	• TBC
	• TBC	• Asia	• TBC
	• TBC	• Oceania	• TBC
	• TBC	• North America	• TBC
	• TBC	• South America	• TBC
Climatic data	• TBC	• World cover	• TBC
	• TBC	• Africa	• TBC
	• TBC	• Europe	• TBC
	• TBC	• Asia	• TBC
	• TBC	• Oceania	• TBC
	• TBC	• North America	• TBC
	• TBC	• South America	• TBC

Threats to Protected Areas
Threats to protected areas are mainly assessed in the APAAT through the use of information on population density and agricultural activity.

Table 4: List of required thematic layers for assessing threats to Protected Areas

Theme	*Official Source / Alternative*	*Available as Web Service*
Population density	*To be defined (CIESIN?)*	
Agricultural activities	*To be defined*	
Road network	*To be defined*	

Ideally, the Digital Observatory should use IUCN's Threats Classification Scheme Version 3.0, but a very simplified version will be adopted first using the scheme below:

1 Population Pressure
      1.1 Urban developments
      1.2 Tourism & recreation areas
2 Agriculture & aquaculture activities
3 Energy production & mining
4 Transportation & service corridors (Fragmentation assessment of habitats)
5 Unrests
6 Habitat alteration & Climate Change
      6.1 Fires
      6.2 Dams & water management/use
      6.3 Droughts
      6.4 Temperature extremes
      6.5 Storms and flooding
      6.6 Others
7 Geological events

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

7.1 Volcanoes
7.2 Earthquakes/tsunamis

The above listed thematic information should ideally be made available through web services to allow for an automatic threat assessment. For example, thematic maps on energy production and mining in Africa could be obtained through AEGOS[3]. The conversion of these maps into potential threats indicators will be the subject of more research. The need for ground validated data is also essential and the means to collect information from remote locations needs to be put in place.

Country-Profiles
As well as creating a report for each Protected Area, the APAAT presents reports for each country based on existing country level datasets that have been downloaded and reformatted. These tabular datasets are available at the bottom of the protected area report menu.

The databases used for reporting are:
1. Environmental Performance Index (http://www.yale.edu/epi/)
    Esty, Daniel C., Marc A. Levy, Tanja Srebotnjak, Alexander de Sherbinin, Christine H. Kim, and Bridget Anderson (2006). Pilot 2006 Environmental Performance Index. New Haven: Yale Center for Environmental Law & Policy.
2. Millennium Development Goals (http://mdgs.un.org/unsd/mdg/default.aspx)
    The Millennium Development Goals (MDGs) were adopted by the UN in resolution 55/2 in September 2000. The eight MDGs - which range from halving extreme poverty to halting the spread of HIV/AIDS and providing universal primary education, all by the target date of 2015 - form a blueprint agreed to by all the world's countries and all the world's leading development institutions.
3. World Development Indicators (http://www.worldbank.org/data)
    The World Development Indicators (WDI) publication is the World Bank's premier annual compilation of data about development.
4. Earth Trends Indicators (http://earthtrends.wri.org/)
    World Resources Institute. 2006. EarthTrends: Environmental Information. Washington, D.C.: World Resources Institute.
In the APAAT, the whole process of generating country profiles for Africa is controlled by one AML script which calls many GAWK scripts to process existing CSV and other tabular data into HTML and which insert the relevant inline images to the HTML pages. The scripts need to become platform independent as well as independent from propriety software.

2.1.4.3.2   Gaps is the models and services

As a result of the data collection process, the APAAT is proposing for each assessed protected area, a radar plot (Figure 2) indicating the species richness of the area and the value of associated habitat using some indicators on the uniqueness of these species and habitats, as well as the pressure on the protected area considering agricultural activities and distance from populated areas.

---

3   http://www.aegos-project.org/

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

Figure 2: Example radar plot of all indicators for a given protected area (red), against the average of all protected areas in the country (grey).
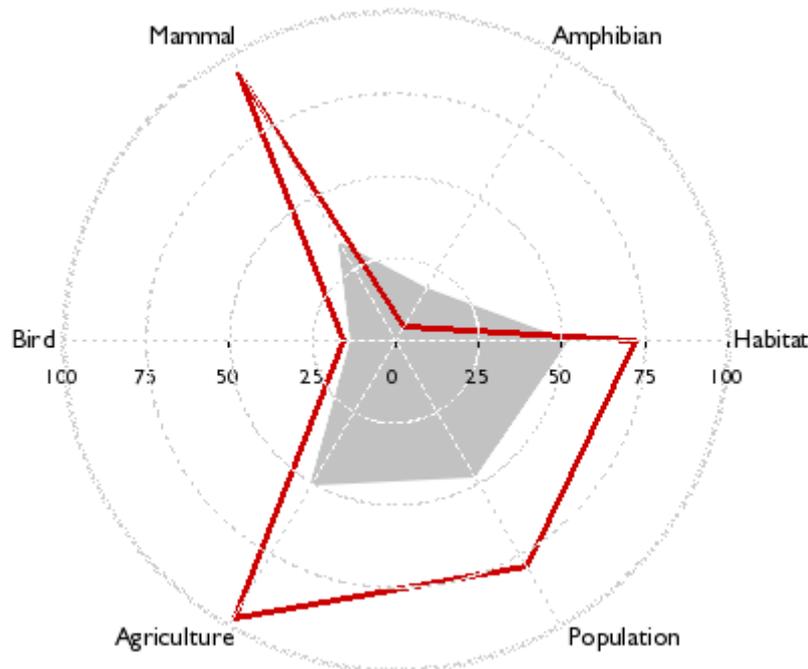
The plot is derived from indicators that can be seen as the outcome of models. These indicators are discussed in Annex 1. For the time being, these indicators are neither made available in an exchangeable and reusable format, nor can they be easily generated automatically for new areas.

2.1.5 Main user requirements

The main user requirements for JRC-MONDE within the project are:

- To allow automated assessments of protected areas;
- To provide interactive tools for assessing biodiversity, habitats, and threats to habitats in unprotected areas;
- To extend the geographic coverage of the current system to the whole globe;
- To improve the reuse of the information provided;
- To improve the indicators used for the assessment of protected areas through the development of means to collect ground truth validated data;
- To provide a habitat modelling service that can characterize habitats in protected areas and be coupled with climate change models for assessing changes to habitats.

2.2 GBIF

2.2.1 Introduction

The Global Biodiversity Information Facility is an intergovernmental organisation tasked with helping to build a global informatics infrastructure through promoting participation and working through partners, mobilising biodiversity data, promoting data exchange standards, building an informatics architecture, promoting capacity building, and catalysing the development of analytical tools. It is not, of itself, a data provider or aggregator but rather provides a dynamic index of data published via the GBIF network by GBIF Participants.

2.2.2   Current Informatics System

The GBIF informatics network consists of numerous geographically dispersed data publishing nodes that are integrated through a data portal maintained by the GBIF Secretariat in Copenhagen. To date, over 190 million species occurrence records are available from over 8000 datasets from over 300 providers. Rather than act as a central repository of data, GBIF maintains a data cache and an associated index. This means that data providers retain full control over, and responsibility for, their own data. But by maintaining an index of a subset of the most commonly used record elements selected from the superset of all elements provided by the data providers, GBIF helps users to search efficiently for biodiversity data. It is inherently difficult to maintain full dynamic access to a network comprising hundreds of data serving nodes, any of which may go offline temporarily and thus become unavailable. GBIF has succeeded in developing a centralised data cache for its network with a strategy for its periodic updating. A centralised cache also ensures the fastest possible response time for services built on the network.

To enable interoperability amongst the nodes, the GBIF network relies on a set of community agreed standards. These are -

1. Darwin Core[4] and ABCD Schema[5] as federation schemas for sharing species primary occurrence data.
2. The DiGIR[6], TAPIR[7] and BioCASe[8] protocols for enabling access to, and transport of data.

The main components of the network are:

1. data providing nodes
2. a central registry of the nodes and their access points on the web
3. a central caching and indexing system for the data published by the nodes
4. a data portal frontend providing unified access to the data from all nodes on the network
5. web services and API for programmatic access to data on the network

The core data that the GBIF network serves are:

1. taxon names
2. taxon occurrence information, i.e., primary species occurrence data, either -
    1. specimen records from natural history collections
    2. observational records

Through its indexing activity, GBIF effectively builds its own internal core data model. The mandatory fields required for indexing are:

1. Scientific Name
2. Institution Code
3. Collection Code
4. Catalogue Number

---

4 http://rs.tdwg.org/dwc/index.htm
5 http://www.bgbm.org/tdwg/codata/schema/
6 http://digir.sourceforge.net/
7 http://www.tdwg.org/activities/tapir/
8 http://www.biocase.org/products/provider_software/index.shtml

But certain other fields are also highly desirable:

1. Geospatial location - where specimen was collected or observation made
2. Collection date - when specimen was collected or observation made
3. Higher Taxon Information (to avoid ambiguities with homonyms etc. during indexing)
4. Basis of Record (whether specimen or observation)
5. DateLastModified (to reduce load during the indexing process)

GBIF currently indexes some 30 fields from Darwin Core (and their equivalent in ABCD).

Interfaces to the cached data

The GBIF data portal (http://data.gbif.org) provides unified access to the data published on the network. Four main routes into the data are provided – via species, country, datasets, and occurrences. The latter comprises a graphical interface that enables a user to construct complex filtered searches of the data. All result sets can be further refined through reviewing individual records, taxonomy and classification, constituent countries and datasets, and the final selection downloaded in spreadsheet format (Tab Delimited, CSV, Excel), Darwin Core XML, one-degree cell density overlays for Google Earth or placemarks for Google Earth. A list of species in results is also available.

To enable programmatic access to the data, GBIF provides a set of REST based web services (http://data.gbif.org/tutorial/services):

- *Taxon data service* (http://data.gbif.org/ws/rest/taxon) provides a range of options for viewing information on the names and classifications used by the different datasets that are accessible through the GBIF portal. Data are returned using the TDWG Taxon Concept Schema.

- *Occurrence record data service* (http://data.gbif.org/ws/rest/occurrence) provides a range of filters for selecting occurrence records. The currently supported response formats include TDWG Darwin Core records and KML (for use with Google Earth).

- *Occurrence density data service* (http://data.gbif.org/ws/rest/density) returns summary counts of occurrence records by one-degree cell for a single taxon, country, dataset, data provider or data network. The currently supported response formats include a simple XML format and KML (for use with Google Earth).

- *Dataset metadata service* (http://data.gbif.org/ws/rest/resource) returns metadata relating to the datasets accessible through the GBIF portal. The currently supported response format is a simple XML structure.

- *Data provider metadata service* (http://data.gbif.org/ws/rest/provider) returns metadata on the data providers sharing data through the GBIF portal. The currently supported response format is a simple XML structure.

- *Data network metadata service* (http://data.gbif.org/ws/rest/network) returns metadata on the data networks connected to the GBIF portal. The currently supported response format is a simple XML structure.

2.2.3 Gaps in data, services, models

2.2.3.1 Data

GBIF's near-term priorities concern the mobilisation and access to richer data types, including names data, that go beyond the simple point occurrence data currently indexed and are being coordinated through two specific areas in the GBIF work programme -

- DIGIT: focuses on catalysing the technical and sociological mechanism to assist Participants to mobilise the volume, depth and density of primary biodiversity data to allow useful and credible analyses using the portal and web services, in particular, by developing the network's capacity to discover a greatly expanded array of primary biodiversity data types beyond specimen and observational records.
- ECAT: focuses on developing a Global Names Architecture[9] (GNA) comprising a comprehensive index of known organism names cross-referenced to a directory of authoritative taxonomic data.

2.2.3.2 Services

Data portal: the GBIF data portal cannot serve all needs. Different countries, regions and thematic communities have specific requirements. With this in mind, GBIF is developing a Nodes Portal Toolkit that can be customised for particular purposes.

OGC Web Services: there is a need for GBIF to provide robust, production level WMS and WFS services.

Metadata: GBIF does not currently provide a metadata catalogue of resources on its network because of the original focus on a data model based on Darwin Core and the DIGIR protocol. Metadata about resources published on the network derived through the protocols (DiGIR, TAPIR, BioCASe) and indexing process is available through the resultset pages returned from querying the cache and via REST services (see section above) but this is not expressed in a standard metadata format. GBIF, recognising the need to develop a metadata catalogue system for its own network and enable interoperability with metadata initiatives of other networks, convened a special task group to advise on requirements and make recommendations. The report of the task group – the Metadata Implementation Framework Task Group (MIFTG) - is currently under review and will be made available soon [URL to be supplied]. The main recommendations of the task group are listed in Annex 2.

Persistent identifiers for biodiversity resources: GBIF has identified the provision of identifiers for biodiversity objects as one of the central challenges to developing a global bioinformatics infrastructure. One of the stated goals in the GBIF strategic plans document "*GBIF Plans 2007 – 2011 from prototype towards full operation*" (http://www2.gbif.org/strategic_plans.pdf) is to consolidate the underlying enabling infrastructure and standardisation for global connectivity of biodiversity data and information through an activity to "develop a system of globally unique identifiers and encourage their use throughout biodiversity informatics". GBIF convened a task group, the "LSID GUID Task Group" (LGTG) to explore the issues, review requirements and offer recommendations on the way forward, with particular reference to the GBIF network, that will enable GBIF to provide architecture leadership and best practices for implementation.

---

9   http://www2.gbif.org/PosterGNAA0low-res.pdf

2.2.4 Main User Requirements

Placing user requirements in the broadest context, the recommendations of the recent eBiosphere[10] international conference are particularly relevant. Several initiatives were identified as vital contributions to a strategic road map for Biodiversity Informatics, including:

- Creating durable, global registries for the resources that are basic to biodiversity informatics (e.g., repositories, collections);
- Completing the construction of a solid taxonomic infrastructure;
- Creating ontologies for biodiversity data;
- Developing an approach to the citation of published data and information services;

GBIF, itself, through interactions[11; 12] with the broader biodiversity community, has identified the following needs in tools, services and data.

Tools

- Visualisation tools to support integrating species point occurrence data with secondary data resources, including species' distributions. Examples of the latter include the Convention on Migratory Species (CMS) Global Register of Migratory Species (GROMS) polygon-based species' distribution maps and the UNEP WCMC/IUCN World Database on Protected Areas (WDPA). Such tools have several potential uses, including: i) validation, quality control (e.g., taxonomic issues, identification of spatial resolution, geospatial coordinate errors); ii) gap analysis of available point occurrence data, sample sizes, etc. throughout an estimated species' distribution; iii) development of species checklists/species richness for an area of interest.

- Data discovery system

Services

- Web services to allow initial visualisation of available species information followed by download to allow more complex analyses using desktop tools;
- Mechanisms to enable on-line editing and certification of species' distribution maps;
- Use of feedback tools such as 'Get Satisfaction' and 'User Voice' to help assess functionality of applications and meet user needs.
- Services for Invasive Alien Species (maps, time scales, reference images)
- Services to support reporting requirements of the parties to the Convention on Biological Diversity;

Data

- need to document spatial resolution and sampling effort in metadata; the latter is not routinely captured in GBIF mediated data thus precluding species abundance analyses;
- scalability of computing capacity to handle exponential growth in data volumes and reduce processing times;
- integration of GBIF mediated species occurrence data with environmental data;
- integration of GBIF mediated species occurrence data with IUCN Red List data

---

10 http://www.e-biosphere09.org/page/workshop.html

11 Workshop on Informatics Tools for Large-Scale Spatial Analyses and Mapping held at the Global Biodiversity Information Facility Secretariat, Copenhagen, 22-23 September 2009

12 GBIF Participants Survey, 2009

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

2.3   UNEP-WCMC

2.3.1   Organisation and Informatics Goals

UNEP-WCMC is the biodiversity monitoring arm of the United Nations Environment Programme and is responsible for a range of biodiversity-related services that support conservation decision-makers. The organization synthesizes, analyses and disseminates biodiversity knowledge to provide authoritative, strategic and timely information for conventions, countries, organizations and companies. This information is used in the development and implementation of their policies and decisions.

UNEP-WCMC collate and manage a wide range of biodiversity datasets including species distributions and trade, ecosystem extent and state, climate and climate change and world protected areas. Since 1981 UNEP-WCMC, through its Protected Areas Programme, has been compiling information about protected areas and making it available to the global community. The World Database on Protected Areas, as it is known, is a joint project of UNEP and IUCN that brings together protected area data from national governments, non-governmental organizations, academic institutions, international biodiversity convention secretariats and many others.

UNEP-WCMC's main informatics goals with respect to the WDPA is to improve the overall content, completeness and quality of the WDPA and to improve access to the WDPA, and derived and related information, to decision makers and other potential users. See www.unep-wcmc.org for further details.

2.3.2   Current systems

The WDPA, as it stands, comprises the following components:

- A spatial database that holds the WDPA data, and that is managed internally at UNEP-WCMC
- A set of tools and applications for validating, importing and managing data provided by the stakeholders in the WDPA
- A set of data standards that describe the core attributes of the WDPA and promote the exchange of data
- A public facing website (http://www.wdpa.org) that is a public interface to the WDPA and allows users to map and download the data
- A Web Feature Service (WFS) that delivers the WDPA data to the EuroGEOSS partners directly
- A set of tools and utilities for working with the WDPA and producing value-added analyses, such as summaries of protected areas coverage through time.

The core content of the WDPA contains information on the protected area name, area, country and IUCN category. In addition there are complimentary databases that record management effectiveness (i.e. how effective the protected areas actually are in terms of preserving biodiversity).

2.3.3   Gaps in data, services and models

The WDPA has been in development for over 25 years and now contains data on 194 countries. The quality, accuracy and timeliness of the data obviously varies widely depending on the data provider, but additionally there remain geographical gaps in the data that need to be filled. These geographic gaps are prioritised by UNEP-WCMC and contacts made with the national authorities. In addition to these geographic gaps, there are countries which only have protected area data as points rather than polygons and these are also a priority to fill.

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

There are also information gaps with respect to how effective the protected areas actually are. Information on the level of protection afforded the PAs is captured in the IUCN category (the IUCN categories are described here). However, this is only populated for a small proportion of the PAs and UNEP-WCMC are promoting and enabling the data providers to provide this information.

Another method of identifying change in protected areas and monitoring how protected they really are is to look at land-cover change and level of degradation within them. This is another area where UNEP-WCMC want to develop tools aimed at capturing change through time, supporting time-series analyses and models. These kinds of tools would be heavily dependent on the availability and interpretation of remotely sensed imagery and services.

2.3.4   Main user requirements

The main user requirements for the WDPA within the project are:

* To improve the geographic coverage and completeness of the WDPA, including increasing the proportion of polygon data
* To improve the quality of the WDPA in terms of spatial and attribute accuracy
* To improve the assessment of the level of management and of protection afforded to protected areas
* To create new content showing land cover change and ecosystem/biodiversity degradation within protected areas, supporting time-series analysis and modelling

The other requirements for the WDPA, in terms of improving the management of the data and the efficiency of getting new content into the WDPA are covered by other projects within UNEP-WCMC.

2.4   BirdLife International

2.4.1   Organisation and Informatics Goals

BirdLife is an international NGO representing a Partnership of over 110 conservation organisations globally, using birds as the unifying focus of work. BirdLife strives to conserve birds, their habitats and global biodiversity, working with people towards sustainability in the use of natural resources.

See www.birdlife.org for further details.

As part of BirdLife's remit, they maintain data on behalf of the Partnership. These data cover species distributions, point localities, and conservation status, Endemic Bird Areas (EBAs), and Important Bird Areas (IBAs) (those locations of specific conservation value for birds). These data must be maintained at a high quality, readily available to Partners and, increasingly, available to the public and other organisations in easy-to-interpret formats.

2.4.2   Current systems

The main applications are:

* World Biodiversity Database (WBDB) - an on-line database maintained by Partners; WBDB holds information on IBAs and the species found within them. This is a hosted service, developed using php and MySQL.
* World Bird Database - an off-line tool used internally to maintain information on the conservation status of all 10,000 bird species. The database is built on MS Access 1997.
* Data Zone - the public delivery mechanism of IBA, species, and EBA data through www.birdlife.org/datazone. Currently an ASP application, hosted externally, although there are plans to move to a hosted php/MySQL application.

- GIS data - held internally, with intention to provide via the Data Zone; GIS data cover species point localities, range polygons, and polygons for both IBAs and EBAs. Held in ArcGIS Shapefiles/Geodatabase format.

2.4.3 Gaps in data, services and models

At present, not all species have their ranges mapped (60% completed), and about 10% of IBAs are only available as approximate point locations.

There are currently no services delivering these data either internally or externally. This delivery is intended to be developed as part of the EuroGEOSS project.

There are no models available or currently required.

2.4.4 Main user requirements

The main user requirements are:
- to maintain systems for storing information and making it available to the Partnership;
- to complete the species range polygon dataset and the IBA inventory dataset;
- to enable improved analyses;
- to present BirdLife data to the public in a variety of formats;
- to provide a consolidated global view of the conservation status of birds.

2.5 Royal Society for the Protection of Birds

2.5.1 Organisation and informatics goals

The Royal Society for the Protection of Birds (RSPB) is an environmental charity, and the UK partner of BirdLife International. RSPB speaks out for birds and wildlife, tackling the problems that threaten the environment. It relies upon memberships and donations to fund its work, with over 1 million members, a staff of 1,300 people and 13,000 volunteers, and more than 200 nature reserves in the UK and abroad.

See www.rspb.org.uk for further details.

In the context of EuroGEOSS, the main informatics goal is to encourage the public to contribute their data to conservation, to make those data available to a wide variety of organisations, and to promote the use of informatics tools by the public to help them engage in practical protection of their environment.

2.5.2 Current systems

The main RSPB application relevant to the EuroGEOSS project is WorldBirds - a suite of Citizen Science databases that is used to collect birdwatchers' information. Data represent a visit to a location, and the birds and conditions recorded during that visit. More than 130 countries are covered by the WorldBirds core model, with another 15 having independent systems (that are not part of the EuroGEOSS project). The databases are hosted externally, using php and MySQL.

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

2.5.3   Gaps in data, services and models

Data are collected opportunistically and provided by the public under no particular timescale or minimum quantity. Not all countries are covered, and the quality of data vary between users - the system does promote data collection minimum standards (location, species, date, observer).

There are currently no services delivering these data either internally or externally, apart from the WorldBirds application itself (www.worldbirds.org to visit the map portal that links all participating countries). Delivery is intended to be developed as part of the EuroGEOSS project.

There are no models available or currently required.

2.5.4   Main user requirements

The main goals of WorldBirds are:

- the collection of bird records to enable scientific analyses of trends and distributions;
- the provision of conservation information to BirdLife Partners and others;
- an increase in capacity of Partners, in terms of membership, human resource, and funds;
- an engendering of community engagement and awareness in bird conservation, with more members of the public actively engaging in environmental projection;
- the collation of all public observations into a single database to enable delivery to the Avian Knowledge Network, and from there into the Global Biodiversity Information Facility (GBIF) and other applications;
- the development of enhanced uses of the data, including incorporation of other datasets to enrich the bird data held.

2.6   Summary of the WP2 questionnaire.

2.6.1   WP2 questionnaire

The following table provides a summary of responses of WP4 partners (EC-JRC, GBIF, UNEP-WCMC, BirdLife Int., RSPB) to the questionnaire *Task 2.1: Questionnaire for User Requirements*, v02 provided by WP2 (Nativi 2009).

Table 5: Summary of responses to questionnaire on user requirements

	EU-JRC	GBIF	UNEP-WCMC	BirdLife International	RSPB
Does your system/tool use International Standards for Metadata (e.g. ISO 19115 application profiles, Dublin Core application profile, …)? a) if Yes, which ones? Since when?	Plan to adopt the simplified INSPIRE implementation for protected areas. OGC web services being served will contain sufficient information to match with the	No.	Not at present. The World Database on Protected Areas is a collation of information from a large number of countries and we have a metadata schema for all of the constituent datasets based on Dublin Core.	World Bird Database (WBDB) has no metadata currently. GIS data conform to metadata standard 19115.	WorldBirds systems. No metadata are collected for these Citizen Science data and there are no plans to use a standard. Data attributes maintained in bespoke structure.

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

	Simple Dublin Core standard		However, the data is not comprehensive or complete for many countries since the data have been collated over a long period of time (>20 years). The metadata schema and content is currently under review.		
b) if No, do you use other Best Practices for metadata (e.g. Darwin, THREDDS, …)? Which ones?		Metadata is derived through Darwin Core federation schema and DiGIR, TAPIR and BioCASe protocols			
Do you have plans to use International Standards for metadata? Which ones?		GBIF is implementing a metadata catalogue for its network and will accept all metadata in its original format. The main standards in use are: FGDC, ISO 19115/19139, EML			
Does your system/tool use International Standards for expressing and formalizing Quality Metadata? a) if Yes, which ones? Since when?	Only basic schema validation (?) (INSPIRE, ISO 19115, Dublin Core)		No quality metadata used. A simple hand-written quality description about each country datasets is given.	No	No
b) if No, do you use other Best Practices for Quality metadata? Which ones?		No.			
Do you have plans to use International Standards for Quality metadata? Which ones?		No.			
Does your system/tool use	WMS, WFS, WCS, WPS,	Not at full production level.	Yes. WMS, WFS, WCS,	WBDB - none.	No

International Standards for Web Services interfaces (e.g. OGC Web Services: WMS, WCS, WFS, CSW, WPS, SOS, …)? a) if Yes, which ones? Since when?	CSW(?)	The GBIF Integrated Publishing Toolkit incorporates GeoServer and can provide OGC WMS and WFS. A simple GML application profile for a subset of Darwin Core fields is available. The GBIF data portal currently offers a simple beta OGC WMS for queries based on country, species and datasets but the service is not exposed via a capabilities file end point.	KML.	GIS data via Data Zone currently no application, but intention is to have a WMS service.	
b) if No, do you use other Best Practices for Web Service interfaces (e.g. DiGIR, OPeNDAP, …)? Which ones?		The GBIF portal also offers geospatial data as one-degree cell density overlays for Google Earth and placemarks for Google Earth. GBIF supports Best Practices in its community by promoting such protocols as DiGIR and it replacement TAPIR, and in response to a widespread need for lowering barriers to data publishing, GBIF developed a simple Darwin Core text archive format.			
Do you have plans to use International Standards for Web Service interfaces? Which ones?		GBIF supplies a set of REST web services for accessing the data cache. GBIF is planning to deliver robust WMS and WFS services for a subset of the data cache (depending on user requirements).			
If you use OGC Web Services, which version (e.g. OGC WMS 1.1 or WMS 1.3	WMS version 1.0.0 to 1.3.0 WFS version	Beta OGC service uses WMS 1.3. a) The data cache is	OGC WFS 1.1 OGC WMS v1.0.0, v1.1.0, v1.1.1, v1.3.0 OGC WCS		

or other)? a) Which software and hardware to you use to supply these services (server)?	1.0.0 (GML 2 and 3) WCS version 1.0.0 and 1.1.0 The web services are provided by MapServer 5.4.2 WPS 1.0.0 (pyWPS implementation) CSW using GI-CAT (?)	currently stored in MySQL and GeoServer is used for map and WFS output; this was achieved by developing a custom plugin for GeoServer. GBIF is planning to use PostgreSQL/PostGIS for geospatial services and processing. GBIF has at its disposal database servers with 2xQCP and 128G RAM, along with high end application servers for mapping services.	1.0.0, 1.1.0, and 1.1.1 a) Windows Server 2000; ArcGIS Server 9.3 Hardware is an HP ProLiant DL380 G5, 2x Quard, 3.0 GHz, 8Gb RAM		
b) Which software and hardware to you use to consume these services (client)?	Services will be consumed by clients like OpenLayers, AJAX functions (etc)	b) The GBIF data portal uses a simple OpenLayers based client to display maps generated by the OGC WMS, and custom JS mapping solutions which generate WMS requests.	b) Web browsers through custom web applications		
c) Are these services available across the Internet? If Yes where?		c) GBIF does not currently offer its WMS for consumption by third parties at a production level, but some people have overlayed maps for demonstration purposes.	c) Yes, but they are not in the public domain and the licensing would not cover use beyond EuroGEOSS.		
Which type of data do you handle with the implemented Web Services?	Vector, Raster, Spatial Database (Vector), Numerical	The OGC WMS provides species occurrence records as counts per 1 degree cell which at highest resolution resolve to actual point occurrences.	For the World Database of Protected Areas we publish both point and polygon information on protected areas. Where we have polygon data, the polygon represents the boundary of the protected areas and where we have point data in most cases the point represents the centre point	None at present; species point and polygon data and location (Important Bird Area) polygon data planned.	

			of the protected area (i.e. not label point but the geometric centre). However, the data quality is not usually sufficient to determine what the point relates to.		
Which map projections do you use?	Lat/Long WGS84 (EPGS:4326)	Geographic (Plate Carree) projection based on WGS84 datum	Many. Mainly WGS84, Mollweide, Web Mercator	Geographic WGS 1984	WorldBirds uses a standard Latitude/Longitude WGS84 projection.
Which type of data encoding do you handle with the implemented Web Services (e.g. binary:CF-netCDF, XML:GML, binary:HDF, XML:GeoSciML, ….)?	XML:GML3	The main (non-OGC) encoding used by GBIF is XML based on Darwin Core and ABCD schemas for species occurrence data, and on Taxon Concept Schema (TCS) for taxon concept descriptions.	XML:GML	None	None
Have you encountered any problems with implemented international standards/best practices in term of limits or ambiguities of specification, of performance, of product metadata definition... ? Please describe them and how you solved them	Not yet	WMS and WFS do not seem to scale well for real-time querying of millions of point data. There is also concern about the file sizes that would be generated through transporting the data as GML.	No problems known		No
Does your system/tool use Modelling procedures? a) if Yes, which ones? Since when?	JRC-MONDE will develop a habitat modeller: various thematic (environmental) map layers can be selected and mixed together for habitat modelling using Mahalanobis distances. Polygons can then be used to define the region of interest and compute the likelihood to find habitat that are	Yes. One of the main application areas of GBIF data has been development of ecological niche models.The GBIF data portal provides a niche modelling service to demonstrate the integration of species occurrence data with other applications such as modelling tools.	No. There are a wide range of uses of protected area information that have been made and many of these uses have been published, some have involved modelling. Within UNEP-WCMC we are not currently using the WDPA for modelling, but we intend to start looking at it in the	No	No

	similar to the one found in the polygon outside of the polygon but in the same eco-region. Biodiversity indicators are also modelled by counting species and through other means to be further explored. The same habitat modeller will also be used for simulating habitats according to different climate change models.		future.		
b) if No, do you have plans to use Modeling procedures? Which ones?					
Does your system/tool use Workflow procedures? a) if Yes, which ones? Since when?	JRC-MONDE will not use workflow procedures in the data portal but a tab interface will allow navigating through different services.	The primary function of the GBIF data portal is to provide species occurrence data that can be consumed by other parties who want to use them in their own workflow system. In the process of indexing data, GBIF applies a set of validation routines to check for taxonomic and geographic correctness, and to build particular indexes, e.g., species lists for a protected area.	In relation to the WDPA, we have a set of validation procedures which are run on new datasets when they arrive which flag any data quality issues with the data, so a decision can be made on any necessary cleaning. These are the only automation routines we run on the WDPA, but GBIF use automation routines on the WDPA to produce a new products which is the species/protected area overlay analysis.		No
b) if No, do you have plans to use Workflow procedures? Which ones?					

The following, existing standards based web services will be included in the EuroGEOSS brokering framework that will enable the initial operating capability of the Digital Observatory for Protected Areas :

- <u>UNEP-WCMC OGC WFS for WDPA</u> : this service provides the World Database of Protected Areas as an OGC Web Feature Service.
  Online address :
  http://downloads.wdpa.org/arcgis/services/EuroGEOSS/EuroGEOSS_WDPA/MapServer/ WFSServer

- <u>GBIF REST Service : occurrence record data service</u> - this service provides a range of filters for selecting occurrence records. The currently supported response formats include TDWG Darwin Core records and KML (for use with Google Earth).
  Online address : http://data.gbif.org/ws/rest/occurrence

- <u>GBIF REST Service : occurrence density data service</u> - this service returns summary counts of occurrence records by one-degree cell for a single taxon, country, dataset, data provider or data network. The currently supported response formats include a simple XML format and KML (for use with Google Earth).
  Online address : http://data.gbif.org/ws/rest/density

- <u>GBIF REST Service : taxon data service</u> - this service provides a range of options for viewing information on the names and classifications used by the different datasets that are accessible through the GBIF portal. Data are returned using the TDWG Taxon Concept Schema. See the documentation.
  Online address : http://data.gbif.org/ws/rest/taxon

- <u>GBIF REST Service: dataset metadata service</u> - This service returns metadata relating to the datasets accessible through the GBIF portal. The currently supported response format is a simple XML structure.
  Online address: http://data.gbif.org/ws/rest/resource

- <u>GBIF REST Service: data provider metadata service</u> - This service returns metadata on the data providers sharing data through the GBIF portal. The currently supported response format is a simple XML structure. Online address: http://data.gbif.org/ws/rest/provider

- <u>GBIF REST Service: data network metadata service</u> - This service returns metadata on the data networks connected to the GBIF portal. The currently supported response format is a simple XML structure.
  Online address: http://data.gbif.org/ws/rest/network

2.6.2 Summary of other questionnaires

2.7 Requirements of different user categories/organizations

- EC-JRC
  - <u>Current systems</u> : the African Protected Areas Assessment Tool (APAAT), is based on a desktop GIS and associated set of tools for statistical analysis and report generation ; data from third parties is collected through the internet using standard downloads and/or on written requests made to the data providers
  - <u>Data collections</u> : datasets used in the APAAT come from several organisations and projects: African Mammals Databank, BirdLife International, CIESIN, Geoland,

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

Global Amphibian Assessment, IUCN, JRC, USGS, NASA, UNEP, UN-MDG, VGT4Africa, VMAP, WDPA, WCMC, The World Bank, WorldClim, WRI and WWF.

- o Current gaps : no OGC web services are provided; no metadata catalogue of resources used and produced; data produced are not readily reusable by others; information made available cannot be easily found by others; need for system to be re-engineered to become open source and independent from any software provider; need for more efficient scripting/statistical analysis packages.
- o Main user requirements : general requirement for whole process of collecting, processing and sharing data (products) to be automated as much as possible through the use of web services ; automated assessments of protected areas; interactive tools for assessing biodiversity, habitats, and threats to habitats in unprotected areas; extension of geographic coverage of the current system to the whole globe; improve reuse of the information provided; improve the indicators used for the assessment of protected areas through the development of means to collect ground truth validated data; habitat modelling service that can characterize habitats in protected areas and be coupled with climate change models for assessing changes to habitats.

- GBIF
  - o Current systems: GBIF network uses community based standards (Darwin Core, ABCD federation schemas; DiGIR, TAPIR and BioCASe protocols to harvest distributed data holdings intoa dynamic cache; the indexed data cache provides access to 190 million point occurrence species records via well documented REST based web services; no production level OGC W*S available; no metadata catalogue yet; existing metadata is derived through Darwin Core federation schema and DiGIR, TAPIR and BioCASe protocols.
  - o Data collections: access to over 8,000 datasets from over 300 providers covering both museum specimen and observational data.
  - o Current gaps: large gaps in taxonomic, spatial and temporal coverage ; need for richer data types beyond simple point occurrence data ; need for gloabl scale informatics system to handle species names ; need for metadata catalogue and quality metadata ; need for robust OGC geospatial web services.
  - o Main user requirements: durable, global registries for biodiversity resources; completion of the construction of a solid taxonomic infrastructure; development of - ontologies for biodiversity data, citation mechanism for published data, visualisation tools, data discovery system, OGC web services, data quality descriptors in metadata; enabling scalability of computing capacity to handle exponential growth in data volumes and reduce processing times; integration of GBIF mediated species occurrence data with other data, e.g., environmental data, IUCN Red List data.

- UNEP-WCMC
  - o Current systems: spatial database for the WDPA data; set of tools and applications for validating, importing and managing WDPA data ; set of data standards that describe the core attributes of the WDPA and promote the exchange of data; WDPA public facing website (http://www.wdpa.org); OGC Web Feature Service (WFS) for access to the WDPA data; set of tools and utilities for working with the WDPA and producing value-added analyses.
  - o Data collections: the WDPA provides data on protected areas in 194 countries including information on the protected area name, area, country, IUCN category and management effectiveness.
  - o Current gaps: quality, accuracy and timeliness of the data; geographical gaps in the data; protected area data as points rather than polygons; information on how effective particular protected areas actually are.

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

- o Main user requirements: improve the geographic coverage and completeness of the WDPA, including increasing the proportion of polygon data; improve spatial and attribute accuracy; improve the assessment of the level of management and of protection afforded to protected areas; create new content showing land cover change and ecosystem/biodiversity degradation within protected areas, supporting time-series analysis and modelling.

- BirdLife International
    - o Current systems: World Biodiversity Database (WBDB); World Bird Database; Data Zone portal (www.birdlife.org/datazone); GIS data held in ArcGIS Shapefiles/Geodatabase format.
    - o Data collections: data cover species distributions, point localities, and conservation status, Endemic Bird Areas (EBAs), and Important Bird Areas (IBAs)
    - o Current gaps: only 60% of species have ranges mapped; approx 10% of IBAs available only as approximate point locations; currently no web services to deliver data.
    - o Main user requirements: maintain information systems; complete the species range polygon dataset and the IBA inventory dataset; enable improved analyses; present BirdLife data to the public in a variety of formats; provide a consolidated global view of the conservation status of birds.

- RSPB
    - o Current systems: WorldBirds - a suite of Citizen Science databases
    - o Data collections: Data in WorldBirds represent a visit to a location, and the birds and conditions recorded during that visit ; more than 130 countries are covered by the WorldBirds core model.
    - o Current gaps: Data are collected opportunistically and provided by the public under no particular timescale or minimum quantity ; not all countries are covered ; quality of data vary between users ; no web servies for delivering data.
    - o Main user requirements: use of bird records to enable scientific analyses of trends and distributions; provision of conservation information to BirdLife Partners and others; collation of all public observations into a single database to enable delivery to the Avian Knowledge Network, and from there into the Global Biodiversity Information Facility (GBIF) and other applications; development of enhanced uses of the data, including incorporation of other datasets to enrich the bird data held.

3   SCENARIOS

3.1   Use Scenarios

In this section, the outcome and conclusions gathered in the previous section are used to identify and describe with detail relevant use scenarios, use cases (optional) and requirements that shall be used by the EuroGEOSS Biodiversity Capacity system(s).

Table 6: Digital Observatory of Protected Areas use scenario

Brief Summary
Develop a Digital Observatory of Protected Areas (DOPA), an interoperable environmental information service, to enable improved assessment of areas of high ecological value, typically Protected Areas.

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

Community Objectives

The use of protected areas to help preserve the planet's biodiversity is a well established practice. However, biodiversity loss does occur in protected areas. Habitat degradation and climate change are the primary contributing factors to loss, yet their impacts are difficult to monitor accurately and forecast.

The purpose of the Biodiversity Operating Capacity planned in the framework of EuroGEOSS is to develop an interoperable environmental information service for enabling improved assessments of areas of high ecological value, typically PAs. This service will become a Digital Observatory for Protected Areas which will mainly rely on its own habitat characterization and modelling tools for ecological assessments and forecasting and on the project partners' information on species distributions for assessing biodiversity. The experience in developing indicators at the continental level for protected areas in Africa will be ported to other continents.

As an interoperable web service generating biodiversity related indicators for selected areas, the DOPA can become a useful decision-making tool in a web processing chain assessing environmental changes due to anthropogenic activities, including climate change. In particular, the development of a habitat modelling web service within the DOPA will allow the community to assess possible consequences to our environment, an issue requiring inter-disciplinary interoperability and advanced modelling from multi-scale heterogeneous data sources.

Actors

The primary actors are the representatives of the partner institutions of EuroGEOSS Work Package 4 who will contribute data and web services to the DOPA, and Work Package 2 partners who will develop multi-disciplinary interoperability and ensure compliance of the EuroGEOSS operating capacity with the principles of GEOSS, INSPIRE, SEIS and GMES.

- o EC-JRC – Gregoire Dubois
- o UNEP-WCMC – Andrew Cottam
- o BirdLife International – Ian May
- o Royal Society for the Protection of Birds – Ian Fisher
- o Global Biodiversity Information Facility – Éamonn Ó Tuama
- o CNR – Stefano Nativi
- o Other: scientists/research biologists (species/habitat modelling); climate data modeller; end users of DOPA – policy/decision makers.

Context and pre-conditions

To assess the state of protected areas, the JRC MONDE activities have focused on monitoring habitats at the continental scale using remote sensing techniques, mainly, and on gathering information on species richness. By assessing the uniqueness of habitats and the level of threat to species, priorities in conservation strategies can be defined in the light of anthropogenic pressures like agricultural activities or climate change.

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

PAs are characterized according to indicators of their irreplaceability in terms of species composition and uniqueness of habitat, and exposure to pressure from anthropogenic sources. In contrast to regular surveys of fauna and flora that can be expensive to conduct, monitoring habitats can benefit from well established earth observation systems and techniques that can

capture most of the environmental parameters that characterise habitats. However, the APAAT largely relies on data provided by third parties collected through the internet using standard downloads and/or on written requests made to the data providers. Much of the data preparation and analysis is carried out within a GIS (Nelson & Hartley, 2008) but, because the effectiveness of conservation practices can only be assessed through regular assessments, the whole process of discovery, collecting and processing data needs to be automated as much as possible and exposed through metadata catalogues and web services.

The whole value of the APAAT is derived from the collection and the merging of a very large variety of data covering very different themes, ranging from species distribution maps to socio-economical indicators. These datasets include: base GIS layers, Species distributions/occurrences layers, and environmental data. For each assessed protected area, the APAAT provides a radar plot to indicate the species richness of the area and the value of associated habitat using some indicators on the uniqueness of these species and habitats, as well as the pressure on the protected area considering agricultural activities and distance from populated areas (see page x). However, impacts on PAs are difficult to accurately monitor and forecast within the current architecture of the APAAT. Moreover, the current structure of the APAAT has provided few opportunities to end-users to assess indicators as these are only available through reports in a graphical form.

The Digital Observatory for Protected Areas aims to become a real time assessment tool relying as much as possible on web services to exchange and process data automatically. The DOPA will also improve the interaction between data providers and data users by allowing users to validate, correct and add to existing repositories of species data, protected area boundaries, threat and habitat information, basic GIS layers, etc.

The APAAT is providing assessments for protected areas in the light of other protected areas of the continent and of the other protected areas of the country. The DOPA is aiming to further extend the use of protected areas to Important Bird Areas (IBAs) as well as to provide the tools and means to obtain an assessment based on biodiversity and threats to any area end-users might be interested in.

Threats to protected areas are mainly assessed in the APAAT through the use of information on population density and agricultural activity. The DOPA will adopt a simplified version of the IUCN's Threats Classification Scheme Version 3.0 which should, ideally, be made available through web services to allow for an automatic threat assessment.

The APAAT also provides reports (country profiles) for each country based on existing country level datasets that are downloaded and reformatted. The DOPA will revise the current approach of preparing country profiles as the amount of information currently made available is far too detailed. Shorter profiles using possibly other more suitable indicators will have to be considered although it is acknowledged that such profiles are necessary to provide a context to the current state of PAs.

Scenario Events
(use cases are derived from scenario steps described below)

step	description

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

Step	Description: The Digital Observatory for Protected Areas - Simulation of new park
01	**Scientist access EuroGEOSS GEO Portal**
01.1	Scientist access EuroGEOSS Portal and searches for "Biodiversity Applications"
01.2	Scientist selects Digital Observatory for Protected Areas (DOPA) client application
01.3	Scientist accesses Digital Observatory for Protected Areas (DOPA) client application
02	**DOPA Client: data discovery and visualisation**
02.1	Scientist uses a catalogue to discover, for a particular region, datasets of interest and their service access points.
02.2	In the DOPA GIS client, the scientist views a world map and selects layers of interest, e.g., protected areas of the world and Important Bird Areas (IBAs)
02.3	Scientist views the indicators (SRI, HRI) for a selected number of PAs. The Indicators are generated automatically in the background as new versions of these datasets become available.
02.4	Scientist selects map layers representing species richness (mammals, amphibians and birds).
02.5	Scientist selects the category of the protected area to be displayed.
02.6	Scientist selects map layers representing species densities (raster overlay of maps of number of mammals, birds, and/or amphibians).
02.7	Scientist selects the vulnerability levels (least concerned, near threatened, vulnerable, endangered, critically endangered) of the species to display to identify critical areas.
02.8	Scientist requests statistics for species densities and level of vulnerability for each protected area.
02.9	Scientist selects a PA on the basis of classes of values of indicators.
03	**DOPA Client: defining a new area**
03.1	Scientist loads/draws a polygon of a **new area of interest** (guided possibly by maps of species densities) (Figure 3).
03.2	Scientist selects modelling service that recalculates the indicators of all PAs in the associated country and in the ecoregion for the newly added polygon(s) as well as for all the other areas. The indicators for all the other protected areas in the same ecoregion and country are recalculated (Figure 4).
03.3	Scientist submits request to modelling service.
03.4	Scientist views modelling service outputs (updated indicators for area of interest).
03.5	Scientist requests printed report on new area of interest.
Step	**Description: Using E-Habitat in the Digital Observatory for Protected Areas**
01	**Scientist access EuroGEOSS GEO Portal**
01.1	Scientist access EuroGEOSS Portal and searches for "Biodiversity Applications"
01.2	Scientist selects Digital Observatory for Protected Areas (DOPA) client application
01.3	Scientist accesses Digital Observatory for Protected Areas (DOPA) client application
02	**DOPA Client: data discovery and visualisation**

02.1	Scientist uses a catalogue to discover, for a particular region, datasets required to compute the habitat, and their service access points.
02.2	In the DOPA GIS client, the scientist views a world map and selects layers of interest.
02.3	Scientist selects one or more protected areas in a region of interest.
02.4	Scientist submits a request to the Habitat Irreplaceability Indicator modelling service for PAs of interest using the new thematic layers/datasets as inputs.
02.5	Scientist views the outputs of the Habitat Irreplaceability Indicator modelling service.
02.6	Biodiversity Indicators of PAs are compared among themselves for the various habitats scenarios
Step	**Description: E-Habitat, Ecological Niche Modelling and Climate Change Modelling in the Digital Observatory for Protected Areas**
01	**Scientist access EuroGEOSS GEO Portal**
01.1	Scientist access EuroGEOSS Portal and searches for "Biodiversity Applications"
01.2	Scientist selects Digital Observatory for Protected Areas (DOPA) client application
01.3	Scientist accesses Digital Observatory for Protected Areas (DOPA) client application
02	**Scientist selects species of interest**
02.1	Scientist submits query to brokering service to discover suitable presence datasets for a particular species
02.2	The Broker mediates the request, forwarding it to the OGC WFS that is used to publish the data.
02.3	Scientist selects one or more datasets returned by the query.
02.4	Scientist receives the selected datasets via the Broker which passes the request to the WFS.
03	**Scientist selects environmental datasets (actual)**
03.1	Scientist submits query to brokering service to discover suitable environmental datasets
03.2	The Broker mediates the request, forwarding it to the OGC WCS that is used to publish the data.
03.3	Scientist selects one or more datasets returned by the query.
03.4	Scientist receives the selected datasets via the Broker which passes the request to the WCS.
04	**Scientist selects environmental datasets (projected)**
04.1	Scientist submits query to brokering service to discover suitable projected environmental datasets derived from Climate Change modelling
04.2	The Broker mediates the request, forwarding it to the OGC WCS that is used to publish the data.
04.3	Scientist selects one or more datasets returned by the query.
04.4	Scientist receives the selected datasets via the Broker which passes the request to the WCS.
05	**Scientist generates Ecological Niche Model (ENM)**
05.1	Scientist uses the ENM client application to submit selected species and environmental datasets to the ENM Server

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

05.2	Scientist uses the ENM client application to select an algorithm for the model and to set the algorithm parameters.
05.3	The ENM client sends a request to the WPS Server to generate the ENM.
06	**Scientist generates projection of the Ecological Niche Model**
06.1	Scientist uses the ENM client application to submit the projected environmental datasets to the ENM Server
06.2	Scientist uses the ENM client application to select the particular ecological niche model to be projected.
06.3	The ENM client application sends a request to the WPS Server to project the ENM.
06.4	The ENM client submits a transaction request to the WCS-T to publish the generated projection.
06.4	Scientist uses the ENM client application to retrieve the generated projection.
06.5	The Broker mediates the request, forwarding it to the OGC WCS-T that is used to publish the projection.
06.6	Scientist selects the generated projection from the datasets returned by the query.
06.7	Scientist receives the selected projection via the Broker which passes the request to the WCS-T.

Post-Conditions

The Climate Change Modelling scenario can be validated by splitting a species presence dataset into two regions, generating the ENM from one region and checking how well it predicts the distribution of the species in the second region.

Based on the effect of Climate Change on the habitat of the species in question, a new Protected Area can be designated taking into account corridors that would allow migration from the old to the new area.

Special Requirements

References

Figure 3: Hand drawn/uploaded polygon and request for biodiversity (species) indicators
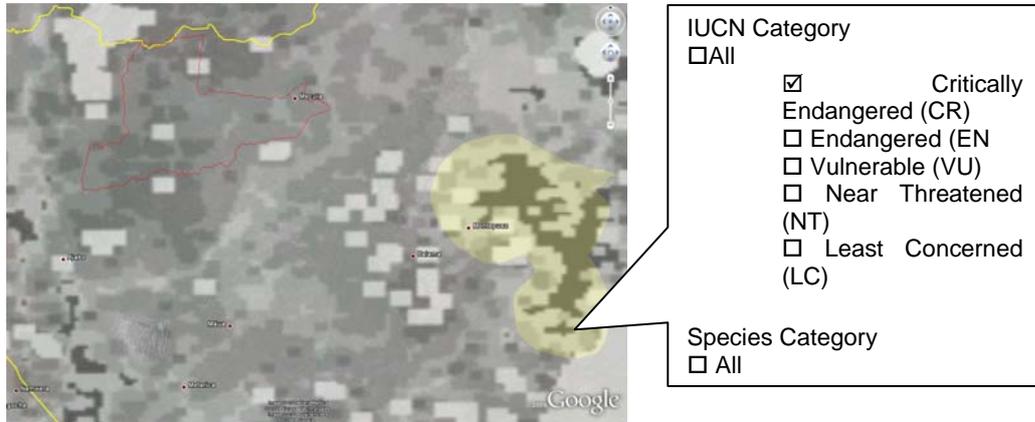

Figure 4: Hand drawn/uploaded polygon and automatic computation of biodiversity indicators (habitat irreplaceability, species richness relative to other PAs, etc)

3.2   Requirements

3.2.1   Metadata

Metadata, conforming to INSPIRE guidelines, should accompany all biodiversity datasets delivered through the EuroGEOSS geo portal. However, different metadata standards, and not just ISO 19115, must be accommodated to serve the needs of particular communities, e.g., Ecological Metadata Language.

3.2.2 Discovery services

There is a need for a central biodiversity metadata catalogue system that harvests metadata from other distributed catalogues and provides a virtual view of all metadata on the network. The central metadata catalogue should provide search/browsing functions and must, in turn, interface to the EuroGEOSS portal clearinghouse to enable inter-disciplinary search across Forestry, Drought and Biodiversity resources, and provide access points for data and associated services.

3.2.3 Access services

Data access should be provided through use of standards-based web services and support the full chain of collecting, processing and sharing data/products.

3.2.4 Data quality

To establish fitness for purpose when considering prospective use of data, data quality statements should be included in metadata, in particular for geospatial, taxonomic and temporal coverage. For example, accuracy and precision should be noted for geo-location measurements, and sampling techniques and protocols should be described. In addition, adoption of community-agreed controlled vocabularies and ontologies will help to provide more accurate descriptions of the data.

ANNEXES

Annex 1: Indicators used in APAAT

Species irreplaceability index (SRI)

For both mammal and amphibian taxa, two different products were calculated. Firstly, in order to provide a general overview of the spatial distribution of species richness, endemism, and threat over the whole continent, a 1km resolution map was calculated using the geographical distributions of all species for the two different taxa. This was done, by giving a value of 1 to each 1km pixel in which a given species was found, and adding together all such maps for all African species of that taxon. This method was repeated separately for amphibians and mammals. Furthermore, a similar map for endemic species was calculated. We defined endemism to be the 25% of species from a taxon with the smallest geographical range. These species were selected (232 out of 930 amphibians, and 70 out of 280 mammals), and the range maps were added together to give the number of endemic species in each 1km pixel. These two maps were not produced for birds, since the IBA database for birds does not contain species distribution maps. Finally, the IUCN red list category of each species was used to produce a map for threatened species (of amphibians and mammals), using the same method of summing species ranges. We defined a threatened species to be listed as vulnerable, endangered, or critically endangered in the IUCN red list.

Secondly, an irreplaceability index (RI) was calculated for each protected area, with respect to all the species of a given taxon (amphibian, mammals and globally threatened birds). This was done by counting how many protected areas a species occurs in (*n*), and adding *1/n* to the RI of each of those protected areas. The same procedure was carried out for all species in a given taxon. The RI of a given PA is defined as:

$$RI_P = \sum_{i=1}^{t}\left(\frac{1}{n_s}\right)$$

Where **p** is a protected area, **n** is the number of PAs a species **s** is found in, and **t** is the total number of species in the taxon database. Note that the conservation value of all species is equal here. This method of measuring biodiversity was chosen because it accounts for both high species richness of common species, but it also gives an increased weight to PAs which have a high number of species found in very few other PAs (i.e. endemic or threatened species). It is important to identify such protected areas, because the success of these PAs is essential to the survival of the species they contain, and furthermore essential to achieving the goal of reducing the rate of current biodiversity loss. Furthermore, this measure provides a basis for the inter-comparison of protected areas, since the RI for a given PA, is a reflection of how that PA's species composition is repeated in other protected areas.

Data collection limitations are the uneven sampling and the different mapping scales. There is a bias towards easily accessible areas and unsuitable areas within the extent of species of occurrences, especially for amphibians. Typically, the mapping scales for amphibians are furthermore very different from large mammals. The effect this would have on a PA's RI would be to give more influence to species with smaller ranges.

Habitat irreplaceability index (HRI)

Typically a protected area contains a specific habitat that can be characterised by climate, terrain, land cover and human population. With such a characterisation, it is possible to identify similar

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

areas within the same biome. Naturally some areas will be more similar to the PA habitat than others, and for some PAs there may be large areas of similar habitat and yet others PA habitats may be unique. If, for each PA habitat, we can identify and rank areas based on their habitat similarity, we can create an indicator of the irreplaceability of the PA habitat (Habitat irReplaceability Index or HRI), and the more irreplaceable a PA habitat is, the higher its ranking in a prioritisation scheme. A useful method for ranking areas in terms of their similarity to a defined habitat is based on the Mahalanobis distance metric. Mahalanobis distances provide a powerful method of measuring how similar a set of conditions is to an ideal set of conditions, and can be very useful for identifying which regions in a landscape are most similar to some "ideal" landscape. The smaller the distance, the more similar the conditions are.

In the APAAT, nine raster maps (Figure 4) have been used to characterise each PA. The variables describe the PA habitat in terms of climate, terrain and land cover. Human population is used later on in the process of creating the HRI for each PA.

The use of the HRI still need to be explored in depth before it can be validated. Among the various issues, the main ones are the propagation of uncertainties in generating the HRI, the choice of the variables, the sampling support (i.e. the size of the area) and the choice of the Mahalanobis distances rather than the Maximum Entropy. Moreover, the HRI has been computed to each Protected Area in Africa that have an area of at least 1,000 hectare but networks of connected PAs (transboundary parks such as the WAP park complex of Benin, Niger, and Burkina Faso) have not been considered.

Threats to Protected Areas

Threats to protected areas are obtained from the assessment of boundary pressure on a PA derived from information on agricultural activities and from a park pressure index derived from the population density around the PA.

Boundary pressure on a PA is assessed by quantifying the amount of agriculture found in the immediate vicinity of the protected area. The assumption here is that the more agriculture found immediately next to the boundary of the PA, then the more pressure there is likely to be on land cover conversion within the boundary of the PA. This is considered a threat to the habitat and consequently the species found within the protected area.

The GLC2000 Africa dataset was used to identify land cover classes which contained a high proportion of anthropogenic influence. These classes were selected on the assumption that they represent areas which contain existing human settlements or agricultural activity. Since the interpretation of land use information from satellite derived products is notoriously problematic, the spatial accuracy of such a classification may have varying degrees of accuracy across the continent.

In order to quantify the boundary pressure on a protected area, the percentage of human influenced pixels was counted within an expanding buffer zone around the PA. This percentage was calculated for a range of buffer sizes, at 1km intervals, between 1km and 30km distance away from the boundary of the PA. Next, a weighting was applied to each buffer in the range, such that higher weight was given to the buffers closer to the PA boundary and lower weight to those further away. This weighting was calculated using a bi-square function and applied to the boundary pressure score for each buffer in the 1 to 30km range. Finally, for each PA, the sum of all the weighted scores was divided by the sum of all the bi-square weights, to create a final score of anthropogenic pressure.

Therefore, the equation for calculating the anthropogenic boundary pressure on a protected area is defined as

EuroGEOSS, a European
approach to GEOSS
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

$$BP_p = \frac{\sum_{n=1}^{n=30}(h_n W_n)}{\sum_{n=1}^{n=30} W_n}$$

$$W_n = \left(\frac{1-n^2}{r^2}\right)^2$$

where BP_p is the boundary pressure score for a given protected area, h is the percentage of high human influenced pixels with a buffer of size n, W_n is the weighting for buffer size n, and r is a constant equal to 30, the maximum buffer size.

The threat to habitat indicator is quantified by estimating the population pressure in the surrounding area. The premise here is that high population density within and around the PA implies:

5. High pressure for land use conversion (conversion to agricultural use) in the buffer zone around the PA and within the PA itself.
6. Higher levels of fishing and hunting in and around the PA.
7. Higher risk of small scale deforestation (thinning) for timber and fuel.
8. Higher risk of intentional burning and clearing.

Not all these factors may be applicable to a single PA, but if one or more of these risks are present then we suggest that this Population Pressure Index (PPI) is a reasonable indicator of these factors.

Estimating the neighbouring population around each PA requires information on ambient population counts (population per administrative unit) and a definition of 'neighbourhood' or zone of influence around each PA. Typically, these zones are represented by a buffer based on Euclidean distance from the location of interest. In this case, this would create a distance buffer zone around each PA (for example 50km) and then the population within this distance would be estimated. A high population density would mean a high PPI score and high pressure.

However, distance is not a suitable measure for describing the potential zone of influence. Two points may be equidistant from a location, but if one point lies on a main road to the location, and the other does not and is also situated behind a natural barrier (river, mountain range), then the influence of that second point on the location is much less than that of the first point. If we replace Euclidean distance with economic distance, i.e. the cost in terms of financial or time resources, then we will be able to generate a much more appropriate measure of the zone of influence around a location, based on the potential accessibility of that location. Hence, a cost distance function has been used as in Nelson (2008) to derive the population pressure.

The development of the DOPA will allow reconsidering the proposed indicators even if their main concepts will be retained. In particular, the HRI is a key indicator for biodiversity and the approach currently implemented will be developed in the last phase of EuroGEOSS as a web processing service currently defined as e-Habitat (Figure 5). E-Habitat, coupled with other data and model based web services, could be used for ecological forecasting considering various climate changed scenarios. This will facilitate more robust and accurate quantification of changes and threats within and around areas of ecological value and so evidence based policy. The DOPA will provide access not just to data but also to analytical models made understandable and usable by scientists from

different disciplinary domains. The results from modelling should also indicate the usage of data (user applications and frequency/popularity) and identify gaps, errors and uncertainties which must be fed back to the data providers.

This requirement to assess uncertainty propagation in the habitat models as well as in the land cover data that are used will be addressed, however, in another European Project (UncertWEB, in negotiation) and the uncertainty framework proposed integrated within the DOPA.
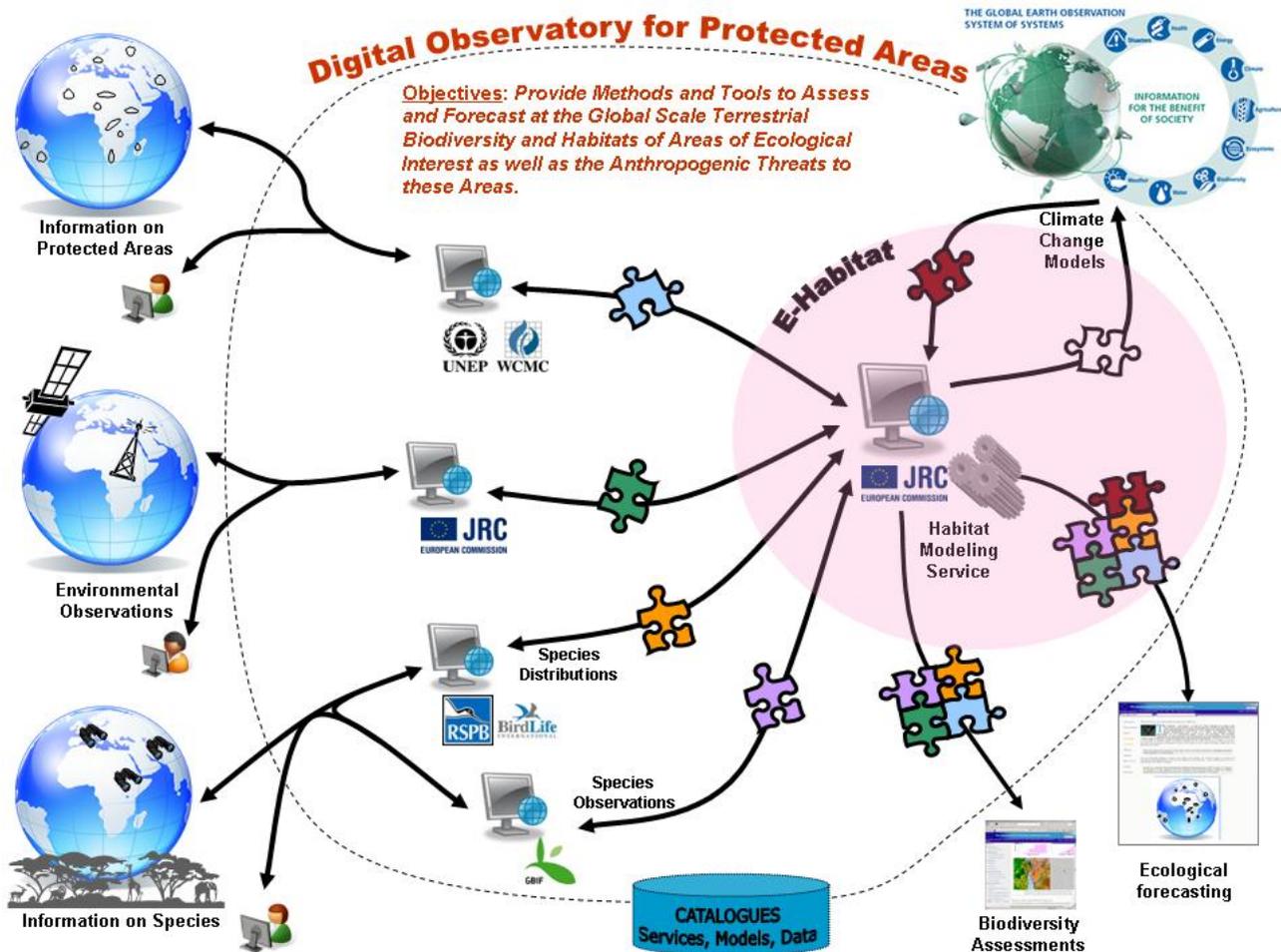

Figure 5: The Digital Observatory for Protected Areas (DOPA) is an interoperable web service version of the APAAT which will also allow interactions with a climate change model web service.

	EuroGEOSS, a European approach to GEOSS	D4 1-v08
	FP7 Project nr 226487	D.4.1: report on user requirements for the EuroGEOSS biodiversity operating capacity

Annex 2: GBIF metadata task group recommendations

The GBIF Metadata Implementation Framework Task Group was convened in 2009 to provide GBIF with guidelines on implementing a distributed metadata infrastructure for its network. The group made the following recommendations.

R1.   GBIF should adopt, or adapt, existing technology where it meets most of the needs of the catalogue project.

R2.   GBIF should seek to collaborate on any new development in order to maximize impact of its development resources.

R3.   Any software developed should be made available as open source

R4.   GBIF should develop a comprehensive list of metadata specifications pertinent to various communities and make sure the metadata catalogue supports these, and keep it updated

R5.   GBIF should promote metadata best practices (e.g., through training activities, web, etc.)

R6.   Metadata should be able to describe multiple types of primary biodiversity data.

R7.   Metadata should support data discovery, interpretation, and analytical reuse

R8.   Metadata should support search/browse by space, time, taxa, and theme

R9.   Metadata should support search/browse by name of provider/name of organization

R10.   Metadata should support search by related publications

R11.   GBIF should accept metadata in multiple formats that are in common use.

R12.   GBIF should provide crosswalks to enable retrieval of metadata in multiple standards commonly in use in order to aid interoperability (e.g., provide conversion among all of EML, BDP, ISO 19115).(See related recommendations R27 - support of multiple metadata models; R28 – ability to return original, contributed format)

R13.   When approached for a recommendation about which metadata standard to use, GBIF should recommend a standard most appropriate for the data being described.

R14.   GBIF minimum requirements for metadata provision should be trivial, but GBIF should accept very detailed metadata in any of the standard formats. The minimal acceptable metadata record might only include the Identifier, Title, Creator, Contact, Metadata Publisher, and Abstract.

R15.   GBIF should highly recommend that metadata additionally include geographic coverage, temporal coverage, taxonomic concepts, methods, data quality (linked to domain specific controlled vocabularies), provenance, thematic keywords, structured entity and attribute descriptions, measurement units using a controlled vocabulary, physical format of the data, distribution information, access control, and intellectual rights.

R16.   GBIF should recommend that a full, detailed, and high quality metadata record is in the best interest of scientific advance, and that providers should provide more complete metadata than the minimum requirements and recommended fields .

R17.   Required metadata *must* be provided in English, with an optional additional translation to one or more other languages. The optional translation should be provided in a format determined by the standard being used. Recommended metadata fields *should* be provided in English, but *may* be provided in any language as determined by the contributor.

R18.   GBIF should develop conventions or solutions to indicate that one metadata record represents an alternate-language translation of another, and/or that two or more fields in a document represent multiple translations. GBIF should do this in conjunction with other initiatives working on metadata standards.

R19.   Each metadata record and data object should possess a location-independent, globally unique identifier which can be used to retrieve the metadata/data object and serves to differentiate each version of the object (i.e., the ID is idempotent).

R20.   Each regional node must replicate metadata to other regional nodes when record changes occur using a GBIF-prescribed replication protocol.

R21.   Each regional node should also provide a harvesting interface that exposes metadata via their unique identifiers.

R22.   GBIF should choose one or more regional nodes with adequate technical infrastructure on each continent to serve as a metadata replica in that region

R23.   GBIF should develop a 'virtual portal' that uses the regional nodes for failover (in the event of network or node outage) and load balancing across the regional nodes

R24.   GBIF needs a registry to maintain list of regional nodes and their relevant service endpoints

R25.   GBIF should develop and pursue an implementation plan that builds this infrastructure in an incremental fashion.

R26.   The metadata catalogue system must support multiple metadata models natively

R27.   The metadata catalogue system must be able to return the original contributed metadata object

R28.   The metadata catalogue system must support unique versioning of metadata and data objects using globally unique identifiers to differentiate revisions

R29.   The metadata catalogue system must support replication and harvesting of metadata (and data) from providers

R30.   The metadata catalogue system must support search and discovery

R31.   The metadata catalogue system must support metadata in XML serializations

R32.   The metadata catalogue system must support input from multiple metadata editors

R33.   The metadata catalogue system must support international language documents and queries

R34.   The metadata catalogue system should support conversion from one metadata model to another and ability to return these alternate formats on request

R35.   The metadata catalogue system should be redistributable under an acceptable open source license

R36.   The metadata catalogue system should support sorting of search results

R37.   The metadata catalogue system should support logical queries and filters on individual metadata fields from multiple standards

R38.   The metadata catalogue system should collect access log statistics on all operations that create, read, update, or delete records

R39.   The metadata catalogue system should maintain a summary of holdings

R40.   The metadata catalogue system should enforce access control restrictions on non-public metadata for read and write by metadata editors

R41. The metadata catalogue system should register with one or more node registries to advertise services available.

R42. The metadata catalogue system should expose one or more standard query APIs for programmatic access by client applications

R43. The metadata catalogue system should provide attribution and branding for original metadata providers

R44. The metadata catalogue system may expose metadata records to other search engines (e.g., provide site index for Google, Yahoo)

R45. The metadata catalogue system may provide bookmarkable queries

R46. The metadata catalogue system may provide subscription services to new metadata records (e.g., RSS feed on query)

R47. The metadata catalogue system may provide thesaurus services for searching and access by other editors/clients

R48. GBIF should create a web-based editor for the GBIF portal for individuals to register their datasets. This should collect the mandatory and recommended list of fields

R49. GBIF should support editors that have the following criteria

The following are criteria that should be used to evaluate the suitability of any editors. It is not expected that any given editor satisfies all criteria:

1. XML input from other editors/sources that are already in place.

2. Ability to edit entity-attribute information.

3. Support auto-capture of metadata elements from the data.

4. Support multiple schemas. Provide a validation service to those schemas.

5. Capacity to validate as you edit.

6. Require partially edited records to be saved and kept for later edits.

7. Ensure fields such as creator, contact, etc. can be easily replicated to reduce effort.

8. Copy from existing records to reduce editing effort. Create author/node-specific profiles including validation rules such as spatial extents.

9. Interface must be well designed for the audience it targets.

10. Control access to records including the ability for the editor to specify access of individuals, groups or public.

11. Metadata editors should support internationalization of the user interface and underlying software components.

12. Metadata editors should support input of metadata in multiple natural languages.

13. Off-line editing should be possible. There should be support for mobile devices for in-field editing and creation of records.

14. Consider open source versus commercial for encouraging the deployment of tools.

15. Shippable desktop application or a service-based tool (e.g., web site)

16. XML input from other editors/sources that are already in place.

It is recognised that most tools will not support all the criteria but this list may give guidelines to continued development and improvement of metadata editors.

**EuroGEOSS, a European
approach to GEOSS**
FP7 Project nr 226487

D4 1-v08

D.4.1: report on user requirements for the EuroGEOSS
biodiversity operating capacity

R50.   GBIF should support any metadata editor that outputs metadata that are valid according to the previous accepted list.

R51.   GBIF maintains a list of recommended tools against the feature set.

R52.   Providers should use controlled vocabularies in any metadata field for which an appropriate vocabulary exists, and should use a multi-lingual thesaurus when appropriate.

R53.   The GBIF vocabularies registry is a valuable service, but should be extended to include a canonical identifier for each vocabulary, and should work to be consistent with other vocabulary registries (e.g., oasis, info, srw).

R54.   Providers should reference the canonical identifier for a vocabulary when listing it in a metadata document (e.g., in the keywordThesausus field in EML).

R55.   GBIF should create an applicability statement identifying which vocabularies are most appropriate for particular fields in particular metadata standards (e.g., use ISO country code in 'country' field).

R56.   The GBIF vocabulary registry should support registration of new and existing vocabularies by third parties.

REFERENCES

Alameh, N., 2009. *AIP-2. Use Cases Engineering Report. GEOSS Architecture Implementation Pilot, Phase 2*, GEO. Available at:
        http://www.ogcnetwork.net/system/files/Renewable_Energy_AIP_SBA_ER-FINAL-1.pdf.

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

Hartley, A., A. Nelson, P. Mayaux, and J. Grégoire, 2007. *The Assessment of African Protected Areas.* Luxembourg: Office for Official Publications of the European Communities, 2007. EUR 22780 EN

Nativi, S., 2009a. *T2.1 Use Scenario Template*, EuroGEOSS. Available at:
        http://www.eurogeoss.eu/wp/wp2/Documents/UseScenarios-EbroDrought_100809.doc.
Nativi, S., 2009b. *Task 2.1: Questionnaire for User Requirements*, EuroGEOSS.

Nelson, A. and A. Hartley, 2008. *Development of the website on the Assessment of African Protected Areas.* Luxembourg: Office for Official Publications of the European Communities, 2007. EUR 23520 EN

Nelson, A., 2008. *Travel time to major cities: A global map of Accessibility.* Office for Official Publications of the European Communities, Luxembourg. DOI:10.2788/95835, ISBN:978-92-79-09771-3.