# eHabitat: a Contribution to the Model Web for Habitat Assessments and Ecological Forecasting.

G. Dubois[a]*, J. Skøien[a], J. de Jesus[a], S. Peedell[a], A. Hartley[b], S. Nativi[c] , M. Santoro[c], G. Geller[d],

[a] European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, 21027(VA), Italy – (gregoire.dubois, jon.skoien, jorge.de-jesus, stephen.peedell) @jrc.ec.europa,eu
[b] UK Met Office Hadley Centre, Exeter, United Kingdom – andrew.hartley@metoffice.gov.uk
[c] CNR - National Research Council of Italy, IMAA  – santoro@imaa.cnr.it, stefano.nativi@cnr.it
[d] NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA – gary.n.geller@jpl.nasa.gov

**Abstract – In striving to improve the predictive capabilities of ecological forecasting we face three basic choices: develop new models, improve existing ones or increase the connectivity of models so they can work together. The latter approach of chaining different interoperable models is of particular interest, as technical developments have made it increasingly viable to combine models that can answer more questions than the individual models alone, allowing users to address complex questions, often of a multi-disciplinary nature. This concept of a Model Web encourages the setting up of a dynamic network of interoperating models, communicating with each other using standardized web services. It is the purpose of this paper to introduce the potential contribution of eHabitat to the Model Web. eHabitat is conceived as a Web Processing Service for computing the likelihood of finding ecosystems with equal properties. By developing eHabitat according to Model Web principles, end-users can define the thematic layers for input to the model from various sources. These input layers are discovered using standards-based catalogues, which are a fundamental component of Model Web and generic Spatial Data Infrastructures. eHabitat integrates data ranging from remote sensing data to socio-economical indicators, thus offering a huge potential for multi-disciplinary modelling.**

Keywords: ecological forecasting, interoperability, SOA, web services, Model Web

## 1. INTRODUCTION

The Digital Observatory for Protected Areas (DOPA) is a biodiversity information system currently developed as a set of interoperable web services at the Joint Research Centre of the European Commission in collaboration with other international organizations, including currently GBIF, UNEP-WCMC, Birdlife International and RSPB. Largely derived from an assessment tool of Protected Areas in Africa (Hartley et al., 2007), DOPA (Dubois et al., 2009, 2010) is also designed to assess the state and pressure of protected areas (PAs) and to prioritize them accordingly, in order to support decision making and fund allocation processes. DOPA uses a wide range of climatic, environmental and socio-economic data to compute a set of indicators for each protected area, quantifying its value with regards to species diversity, irreplaceability of habitats, and threat from population pressure and agricultural boundary pressure. These indicators can be further summarised to show how each PA performs, for each indicator, in comparison to other protected areas from the same country or the same ecoregion. This approach was used by Hartley et al. (2007) to produce the first consistent, continent-wide assessment of 741 protected areas in Africa, based on their value for conservation and anthropogenic pressures.

*Corresponding author

### 1.1 The Habitat Irreplaceability Index (HRI)

To assess the uniqueness of the ecosystems found in each PA, Hartley et al. (2007) proposed to rank the relative importance of each PA by using a habitat irreplaceability index (HRI). The HRI assumes that the analysed area contains a single habitat that can be characterised by a number of geophysical variables and, by analysing the likelihood of finding similar combinations of these variables outside of the investigated area, one can assess the uniqueness of the habitat (Figure 1).
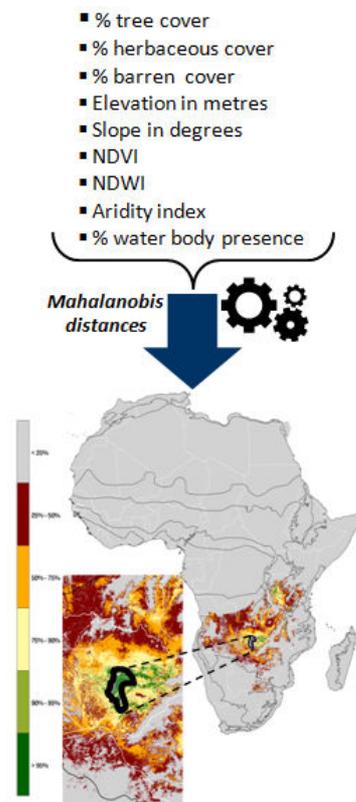


Figure 1.  *p*-value grid (insert, lower left) showing probabilities, in the Zambezian ecoregion, to find similar habitats to the one found in the Kafue national park (Zambia). The HRI is computed here using the 9 thematic layers indicated on top. NDVI = Normalized Difference Vegetation Index, NDWI = Normalized Difference Water Index (NDWI).

For each PA larger than 1,000 hectares, the variables of Figure 1 were combined using a GIS-based niche modelling approach (see e.g. Jennes 2003; Rotenberry, Preston and Knick, 2006) in order to compute a mean value of each variable and the inverse covariance matrix. An ecoregion in which the PA is situated can be selected and the Mahalanobis distance *D* computed for each

1 km$^2$ pixel falling in the ecoregion. $D$ is computed using the square root of

$$D^2 = (\mathbf{x}\text{-}\mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x}\text{-}\mathbf{m})$$

where $\mathbf{x}$ is the vector of data, $\mathbf{m}$ the vector of the mean values of the independent variables and $\mathbf{C}^{-1}$ the inverse covariance matrix of the independent variables.

When the predictor variables used to generate the mean vector and covariance matrix are normally distributed, then $D$ is distributed according to a $\chi^2$ distribution with $n$-1 degrees of freedom, and we can so convert $D$ into $p$-values. These probability values range from 0.0 representing no similarity, through to 1.0 for areas which are identical. If the predictor variables are not normally distributed, then we can still make the conversion because it rescales the unbounded $D$ values to a 0.0 to 1.0 range.

The $p$-value grid is used as an input to the HRI for the PA. A pixel represents a potential replaceable habitat for the PA if the pixel is outside the PA and its $p$-value is 95% or better. In addition, further conditional rules can be used to filter out unsuitable locations, should similar habitats be found, for example, in areas where population density is not negligible. The total area of the pixels meeting these criteria can be summed and divided by the area of the PA to generate the HRI value. The smaller the HRI, the lower the area of similar habitat outside the PA, with a HRI of 0 suggesting that the PA habitat is unique and therefore irreplaceable. Conversely, a HRI greater than 1.0 suggests that there are potentially similar habitats with a total area that is greater than the PA. An illustration of a $p$-value grid is shown in Figure 1.

## 2. FIRST STEPS TOWARDS EHABITAT, A WEB PROCESSING SERVICE (WPS) FOR ECOLOGICAL MODELLING

### 2.1 eHabitat 1.0, a WPS based on PyWPS and R

The need to monitor ecosystems outside of protected areas, whether terrestrial or marine, is stronger than ever, if only to assess connectivity between protected areas and the external pressures coming from the competition for land and water. Providing the scientific community with the means to compute the HRI anywhere on the globe using their own thematic ingredients is therefore an interesting option. eHabitat was therefore conceived in the framework of DOPA as a Web Processing Service (WPS) i.e. a web service following a Service Oriented Architecture (SOA) model where pre-programmed geospatial calculations and/or computation models are available as dispersed services/processes over a network. Note that instead of computing the HRI which involves a restrictive criteria of a $p$-value of 95%, eHabitat computes simply the PoHS, that is the probability of habitat similarity.

The first version of eHabitat (eHabitat 1.0) was designed as a proof of concept to compute the HRI for a given protected area using only 3 thematic maps. Technically, the WPS was mainly built using an Open Source development environment, i.e. R (R Development Core Team, 2010) for the mathematical codes to compute the Mahalanobis distances, and Python for setting up the Web Services. Data processed by eHabitat 1.0 were obtained from Web Coverage Services. An XML request is thus submitted to the WPS indicating where to find the boundaries of a protected area from a Web Coverage Service (WCS) as well as the addresses of where to find the 3 thematic maps. The `ehabitatprocess.py` script that runs a PyWPS (Cepicky and Becchi, 2007) further processing the input data using GDAL

API and send it to R. The results returned from the R codes would be processed to generate a GeoTiff image which is then transformed using GDAL, the Mapscript python package and the Python Image Library (PIL). A schema showing the dataflow in eHabitat 1.0 is shown in Figure 2.



Figure 2. Flowchart of eHabitat 1.0. The WPS receives an XML request with the ID of a protected area available from the World Database of Protected Areas (IUCN-WCMC) and 3 thematic maps in a raster format. The HRI is computed for the protected area within a defined bounding box and the final results released in a GeoTIFF and png format.

### 2.2 eHabitat in the Model Web

Better results from ecological modelling can be obtained either by improving existing models or by developing new ones. Chaining different interoperable models is a third alternative that is most interesting as such a chain would answer more questions than the individual models alone, allowing users to address complex questions of different nature. This concept of a Model Web (Geller & Turner, 2007) encourages practically the setting up of a distributed, multidisciplinary network of independent, interoperating models (and datastores) communicating with each other using web services. Beyond the simple sharing of information, the Model Web conceives open-ended systems of interoperable computer models and databases communicating via SOAs.

The setting up of the eHabitat WPS is a first step in this direction. End-users can, for example, use the outputs from other climate change model services and use these as input data in eHabitat. As a result, eHabitat can become an ecological forecasting service for assessing climate change impact in protected areas. This example using eHabitat is further discussed in Skøien et al. (2011). The use of a WPS for computing the HRI offers a large range of interdisciplinary possibilities that can go from socio-economic modelling, ecological forecasting to the optimisation of environmental monitoring networks when the reference area is limited in space to a measuring station. The full potential of eHabitat depends thus very much on the ability of end-users to find and use other thematic information as input for eHabitat.

**2.3 eHabitat 2.0. A WPS using catalogues and discovery brokers**

The Group on Earth Observations (or GEO) is coordinating international efforts to build a Global Earth Observation System of Systems (GEOSS). This emerging public infrastructure is interconnecting a diverse and growing array of instruments and systems for monitoring and forecasting changes in the global environment. In order to provide the GEOSS Common Infrastructure (GCI) with new functionalities and capacities, the GEO Architecture and Data Committee (ADC) launched the Architecture Implementation Pilot (AIP) Initiative. The third phase of AIP (AIP-3), completed at the end of 2010, increased GEOSS capacity to support 9 strategic Societal Benefit Areas (SBAs): Disaster Management, Health, Energy, Climate, Water, Weather, Ecosystems, Agriculture and Biodiversity. From these SBAs, biodiversity is possibly the one needing the largest efforts in setting up multidisciplinary and cross-disciplinary infrastructures that will allow the research activities required to understand the impact of our environmental policies on the loss of species diversity.

In the AIP-3 context, the EC-funded EuroGEOSS and GENESIS projects have developed a multidisciplinary use scenario dealing with Climate Change and Biodiversity SBAs. Through a catalogue, end-users can select the ingredients available on the internet for modelling the PoHS The use scenario proposed for AIP-3 was conceived to understand how environmental and climatic factors impact on ecosystems found in a given protected area. Based on the previous AIP-2 experience (GEOSS AIP-2, 2009; Nativi et al., 2009) the EuroGEOSS and GENESIS projects enhanced the existing modelling capacity of eHabitat 1.0 with:

1) a discovery broker service which underpins semantics enabled queries: the EuroGEOSS/GENESIS Discovery Augmentation Component (DAC);

2) environmental modelling components (i.e. OGC WPS instances) implementing algorithms to predict evolution of PAs ecosystems;

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

A conceptual flowchart of the enhanced infrastructure of eHabitat is shown in Figure 3. The "eHabitat" use scenario is detailed in the engineering report (GEO AIP-3, 2011) and accessible through the GEO Portal.

# 3. FORTHCOMING DEVELOPMENTS

The use of open standards for spatial data and of open source programming languages for the development of the core functionalities of the system are expected to encourage the participation of the scientific community beyond the current partnerships of DOPA and to favour the sharing of such an observatory which could be installed at any other location. The following points are indicative of research directions linked to the further development of DOPA's key web processing service.

**3.1 Going around the statistical limitations of the PoHS**
The PoHS approach presents a number of limitations encountered when the monitored area presents a complex set of highly variable environments, like a coral reef or a mountain

near a lake. Still, there are a number of ways to go around these problems that should not jeopardize the efforts in further encouraging the developments of eHabitat. Stratifying carefully the area into more homogeneous environments is a possibility. Environments with particularly small variability within a habitat can also create some challenges in interpreting the results. This will strongly restrict and limit the PoHS, to a much larger degree than what is useful for forecasting.
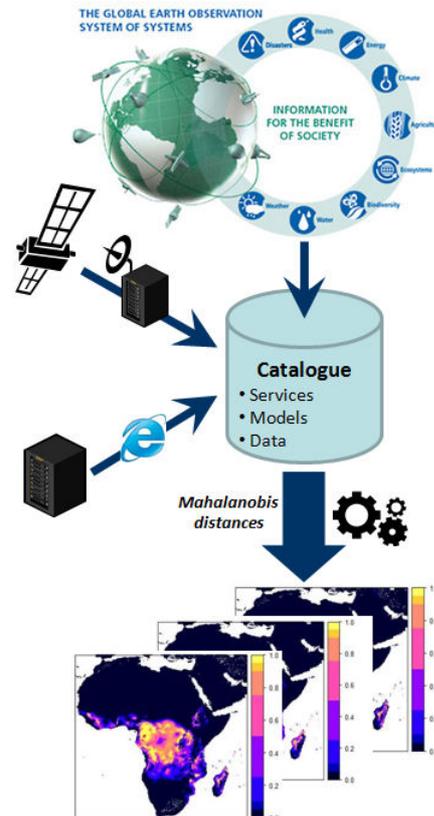


Figure 3. Design of the enhanced architecture of eHabitat as developed in the AIP-3. In contrast to eHabitat version 1.0, eHabitat 2.0 benefits from means to discover and use other data, models and services.

**3.2 From the trees to the abysses: the 3D case of eHabitat**
We started only recently to work on the porting of eHabitat to 3D modelling in view to allow for improved means to capture the complexity of ecosystems at all scales. The increased dimensionality is particularly needed for assessing marine protected areas but the approach could also be used to assess very small scale habitats, like a full tree. The use of Mahalanobis distance itself is practically invariant to the number of dimensions, but the need for 3D modelling capacities requires the development of improved visualisation means of the outputs, something which can also give computational challenges.

**3.3 Setting up an operational capacity for eHabitat**
To port the concept of eHabitat from an experimental framework to one that is operational and open to the modelling community, a number of computational obstacles need to be solved:

1) The various thematic maps selected through a catalogue service often need to be pre-processed (cropped, projected, gridded, aggregated/disaggregated) instantly before being used as input data, as all layers must have the same extent, projection and resolution. Such services are not always available as WPS.

2) The computational back-end of eHabitat needs to be powerful enough to deal with potentially many and/or very large datasets as it is often the case with high resolution remote sensing data. Similarly, chained systems need to be able to cope with asynchronous requests.

3) End-users may want to generate simulations for better assessing uncertainties of their models, a step that is computationally very expensive.

### 3.4 Increasing the computational capacity of chained services

The linking of various modelling services offers a huge potential for innovation provided that these services have the ability to correspond almost synchronously with each other, something that requires enormous calculation power. Like the electricity grid, cloud computing provides via the internet on demand computing services to connected devices, providing so to end-users potentially unlimited computational capacities when needed. Ideally we would link eHabitat with cloud computing services to join more efficiently the Model WEB and become freely accessible to academic users.

### 3.5 The Model Web and uncertainties

Because these input data for eHabitat can a broad range of sources and vary between geophysical data captured through remote sensing and socio-economical indicators, the PoHS is exposed to a broad range of different types and levels of uncertainties. Chained to other services, e-Habitat would be an additional component further propagating uncertainties from a potentially long chain of model services. This integration of complex resources, such as data and models brings ever increasing challenges in dealing with uncertainty. For such a system to be used for policy or decision making, users must be provided with information on the quality of the outputs, since all system components will be subject to uncertainty. UncertWeb (www.uncertweb.org) will create the Uncertainty enabled Model Web by promoting interoperability between data and models with quantified uncertainty, building on existing open, international standards. UncertWeb will thus develop open source implementations of encoding standards, service interface profiles, discovery and chaining mechanisms and open source implementations, and generic tools to realize a "Model Web" taking uncertainty in data and models into account.

### REFERENCES

J. Cepicky and L. Becchi (2007), Geospatial processing via Internet on remote servers – PyWPS, *OSGeo Journal* 1 (May 2007): 5p

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

G. Dubois, A. Hartley, A. Nelson, P. Mayaux, 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

GEOSS AIP-2 Engineering Report (2009), "The Impact of Climate Change on Pikas Regional Distribution", available at http://www.ogcnetwork.net/system/files/FINAL-pikas_AIP_SBA_ER.pdf

GEOSS AIP-3 Engineering Report (2011), "eHabitat", available at http://www.ogcnetwork.net/system/files/CCBio-eHabitat-ER-v2.0-FINAL.pdf

G. N. Geller, and W. Turner (2007), The model web: a concept for ecological forecasting, *Geoscience and Remote Sensing Symposium*, 2007. IGARSS 2007. IEEE International, 2469 – 2472, 23-28 July 2007.

A. Hartley, A. Nelson, P. Mayaux, and J.-M. Grégoire (2007), *The assessment of African protected areas*, JRC Scientific and Technical Reports, EUR 22780 EN, 70 pp, Office for Official Publications of the European Communities, Luxembourg.

J. Jenness. Mahalanobis distances (mahalanobis.avx) extension for ArcView 3.x. Jenness Enterprises, 2003, http://www.jennessent.com/arcview/mahalanobis.htm, (accessed 12 Jan 2011)

S. Nativi, P. Mazzetti, H. Saarenmaa, J. Kerr, E. O. Tuama (2009), Biodiversity and climate change use scenarios framework for the GEOSS interoperability pilot process, *Ecological Informatics*, **4**, 23-33.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

J.T. Rotenberry, K. L. Preston and S. T. Knick (2006), GIS-based niche modeling for mapping species' habitat, *Ecology*, **87**, 1458-64.

J. Skøien, P. Truong, G. Dubois, D. Cornford, G.B.M. Heuvelink, G. Geller (2011), Uncertainty Propagation in the Model Web. A Case Study with eHabitat, (This Volume)