

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Mobile System Applied to Species Distribution Modelling

Álvaro Silva, Pedro Corrêa and Carlos Valêncio

*Nokia Institute of Technology and University of São Paulo, University of São Paulo,  
Instituto de Biociências, Letras e Ciências Exatas, UNESP – Univ Estadual Paulista  
Brazil*

### 1. Introduction

Species distribution based on ecological niche modelling has been used in several areas of ecology. It uses mathematics techniques which are applied to weather statistics and other physical factors which can affect the geographic extension of species in its ecological niche (Soberón& Peterson, 2005).

Based on known localization data (or absence) occurrences of individual species and relating them to environmental variables (such as relief, climate, humidity, etc), it is possible to predict the probability that a region will be favourable for those species survival.

This process depends on the quality of the gathered data in the field. Iwashita (Iwashita, 2008) presents a research concerning the errors influence on the collection point position. There are, however, human factors that can also influence the quality of these points. Here, the main problem is related to automate the presence and absence data gathering process and how to share that data with other scientists.

In order to solve these problems, this chapter presents a mobile system which supports the data gathering and modelling for the distribution of ecological niche. This solution involves Service Oriented Architecture applied to mobile systems. Beyond this, it proposes a new approach to help scientists choose an area to gather field data by previewing the models available for the researcher.

### 2. Distribution species modelling

Distribution species modelling is a way of analysing data applied mainly in biology which uses advanced geographic information systems (Peterson, 2001).

To understand what this modelling represents, it is necessary to understand the concept of ecological niche: according to Hutchinson (1957) ecological niche is defined as “a space with n-dimensional volume where each dimension represents the interval of environmental conditions or necessary sources for the species survival and reproduction”.

In Peterson ( 2001), ecological niche is defined as a group of ecological condition in which a species is capable of maintaining population without immigration.

According to these concepts the ecological niche is nothing more than a determined region where the group of factors favours the species survival. Environmental features that influence species survival can be temperature, humidity, salinity, pH, feeding sources, luminous intensity, predatory pressure, population density, among others. Environmental factors are limited and remain relatively constant on the interval related to these animals timeline (Bazzaz, 1998).

The ecological niche is divided between realized and fundamental. Fundamental niche is defined as a group of environmental conditions necessary for species survival without considering the predators influence. Realized niche is where the species really occurs (Malanson et al., 1992). You can say that realized niche is a sub-group of the fundamental one.

Predictive modelling of species distribution is mainly concerned with the ecological niche modelling. It proposes a solution based on artificial intelligence for foreseeing a probable geographic species distribution of species.

The distribution modelling of ecological niche plays an important role in ecology. Among main applications is the environmental preservation areas planning (Austin, 2002; Guisan& Zimmermann, 2000; Sohn, 2009). Choosing a preservation area requires knowledge about a species ecological niche. With predictive modelling it is possible to identify statistically these areas.

Another area wherein modelling is a driving force, is in climate change research (Peterson et al., 2001; V. Canhos et al., 2005) which aims to identify how living creatures are affected by global warming.

More applications can be found: species replacement in nature, species and habitat management, biogeography and others (Guisan& Zimmermann, 2000).

One model visualization is shown on Figure 1, extracted from Sohn(2009) which made the modelling for cook-of-the-rock (*rupicolarupicola*) bird in the Brazilian Amazon region.

## 2.1 Collecting and modelling process

Basically, predictive modelling of ecological niche occurs in three phases: collecting, modelling and analysing the actual models. Figure 2 shows an IDEF0 diagram used to represent the business plan modelling. In this figure, it can be realized that to obtain a validated model, the sequence has to be followed: first collect the data, then use the entries in the model creation phase which are the data collected and the environmental variables, that are processed with the support of a number of predictive algorithms, thus generating the model and a set of indicators for this model. Based on these indicators it is possible to validate the generated model, evaluating its quality.

The first stage of the entire process is to obtain data about a species' presence or absence to be studied. Both are basically latitude and longitude.

Presence data represents the species' incidence or abundance outcomes in a given position or area. Absence data is given when there was a search for that species in some known region and even in an individual finding (Engler&Rechsteiner, 2004). Absences may occur due to the following (Philips et al., 2006):

- There was a species, but it could not be detected;
- The habitat is suitable, but for historical reasons the species is absent

- The habitat is not suitable.

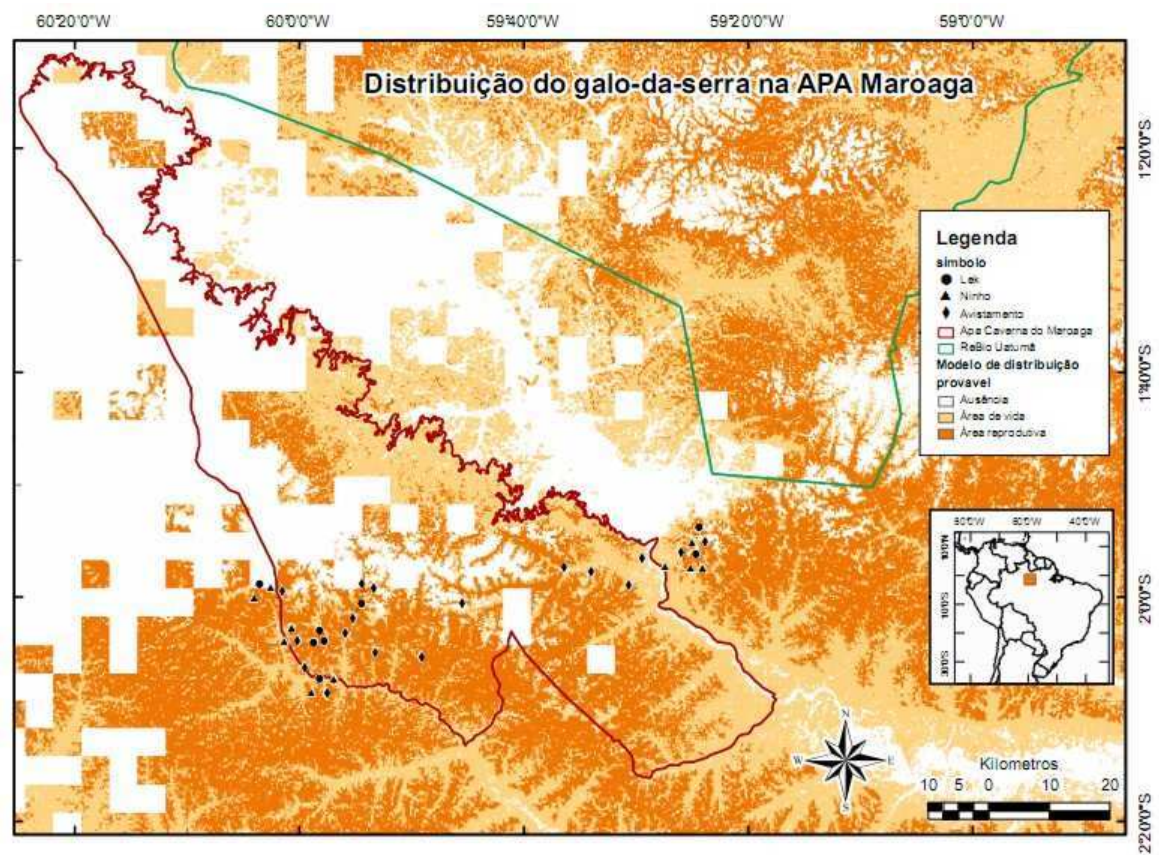


Fig. 1. Model visualization for cook-of-the-rock in Amazon area.

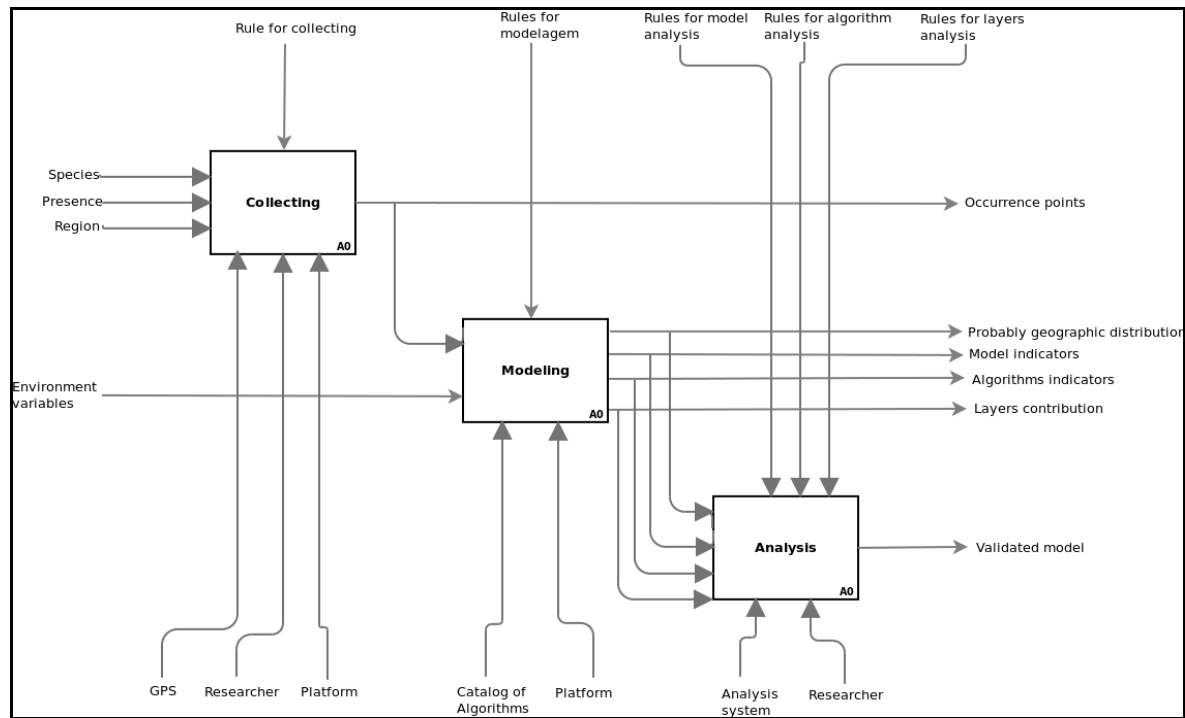


Fig. 2. IDEF0 diagram for modeling process.

The collecting can be separated into another process which follows the steps below:

1. Selecting the area to be studied: This is the stage where it determines the area used to collecting. The species is used as one of the inputs for this phase due to the fact that a species is usually found historically in a particular geographic region (Philips et al., 2006). This region can be, for example, the Amazon region or throughout Australia.
2. Choosing the spatial resolution: Spatial resolution is the scale used for the collection. This is an important step that needs to be emphasized. The choice of resolution can influence the model interpretation, since some patterns can occur in a given resolution, but on another scale may not be noticeable (Guisan&Thuiller, 2005). There are some techniques that help identify the best resolution to use in modelling as in Isaaks and Srivastava (Isaaks&Srivastava, 1989), which suggests the use of variograms to determine the sampling interval.
3. Restore occurrences in other data bases: It is about getting the data of the occurrence from other sources such as museums, zoos and environmental agencies. There are some entities that maintain a database with large numbers of environmental collections such as IABIN/PTN, GBIF, and Ornis.
4. Determination of observation points: Determine how many and where the points are positioned to observe the species. Based on the chosen spatial resolution and the known occurrence points, is intended to promote the best possible distribution of these points.
5. Determining the time of collection: It is the strategy to turn into a more efficient observation. Some species are nocturnal and others diurnal, and some climatic conditions affect these habits as intense heat, etc. Thus, it is necessary to determine these times.
6. Observation: In this phase, techniques are applied to the census of the species studied. These techniques aim to check both the presence and absence. The absence (or record scratch as it is called) (Engler&Rechsteiner, 2004) is a prominent factor due to the difficulty in obtaining such data (Philips et al., 2006). Depending on the species it can be used, for example, recorders that play back the sound of the female or male to attract a species that is in the region, as in Sohn(2009). It is also necessary to determine the time it will make the observation and the number of attempts.
7. Registration: The presence or absence data is stored electronically. Basically each record contains the following information: listener identification, latitude, longitude, type of observation (sighting, nest location, hearing, etc.) and registration date. Among these, the latitude and longitude are the ones that deserve greater emphasis. It uses a GPS to get these coordinates.

### 3. Architecture propose

With the modelling and collection processes being understood, an architecture can be elaborated to solve the problem described above. Here, it will be proposed and specified an architecture of a mobile system applied to species distribution modelling.

The architecture requires a generic basis to be used on any mobile application that needs to perform the species distributions modelling. Furthermore, it should also be able to adapt to constant changes in this area, such as allowing the use of new algorithms and data formats. Thus, the SOC paradigm is quite appropriate, and for this reason, this paper presents an SOA solution.

Basically, the process described in ANDREI, et al. (Andrei et al., 2004) was followed, which describes one for designing a service-oriented architecture, which combines the standards for e-business and concepts of service-oriented computing to solve problems encountered in industry. The steps are shown below:

- 1. Domain decomposition
- 2. Goal service-model creation
- 3. Subsystem analysis
- 4. Service allocation
- 5. Component specification
- 6. Structure enterprise components using patterns
- 7. Technology realization mapping

The steps above will be discussed on next sessions, except the steps 5 and 6. The step 5 will be performed on steps 1 and 3, while the step 6 suggests the use of standard IBM runtime, which are not applied to the pattern of collaboration.

3.1 Domain decomposition

From a business point of view, the domain consists on a series of functional areas. For this case one can observe two main areas: the system of collecting and modelling, as shown in Figure 3.



Fig. 3. IDEF0 diagram for modelling process.

For these functional areas, the following use cases can be identified, in general:

- UC1: Collect presence and absence data
- UC2: Create view for localization
- UC3: Create model
- UC4: View model

On Figure 4 is shown the use case diagram for this system.

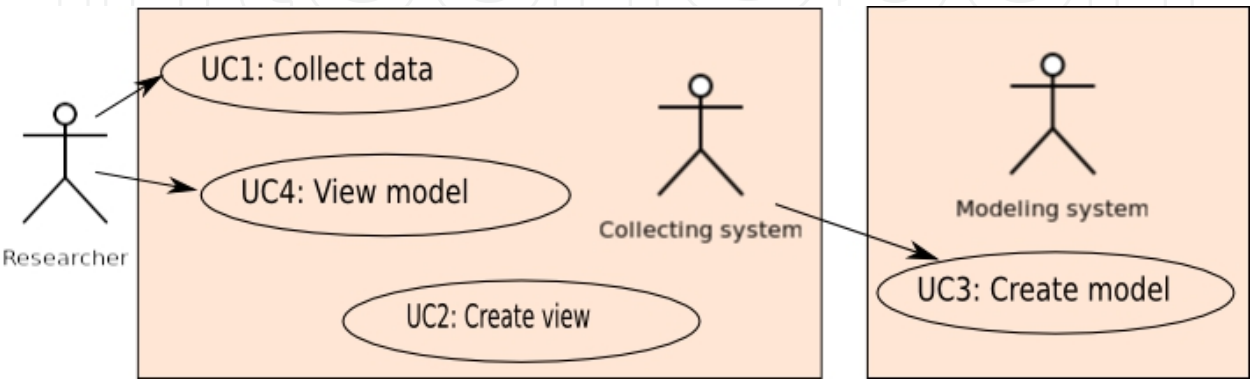


Fig. 4. Business use case diagram.

This business use cases are strong candidates for business services that will be exposed.

Once the business use cases are defined it is necessary to define the inputs and outputs of each service. As the project continues to be developed, the functional areas that were identified will be directed to a subsystem. It can be explained with the fact that the areas are a business sense, while subsystems are technology notions.

For each step described by ANDREI (Andrei et al., 2004), a set of IBM corresponding patterns are applied. Basically, these standards are divided into four parts: the highest standards, the application standards, runtime patterns and Product mappings.

This step will apply the highest standards. It is necessary to choose the pattern that best fits the system proposed on this paper. Since this system needs to provide an interaction among researchers from the field data collection and from other areas, the standard chosen was the Collaboration. Their application to the system proposed in this work can be seen in Figure 5.

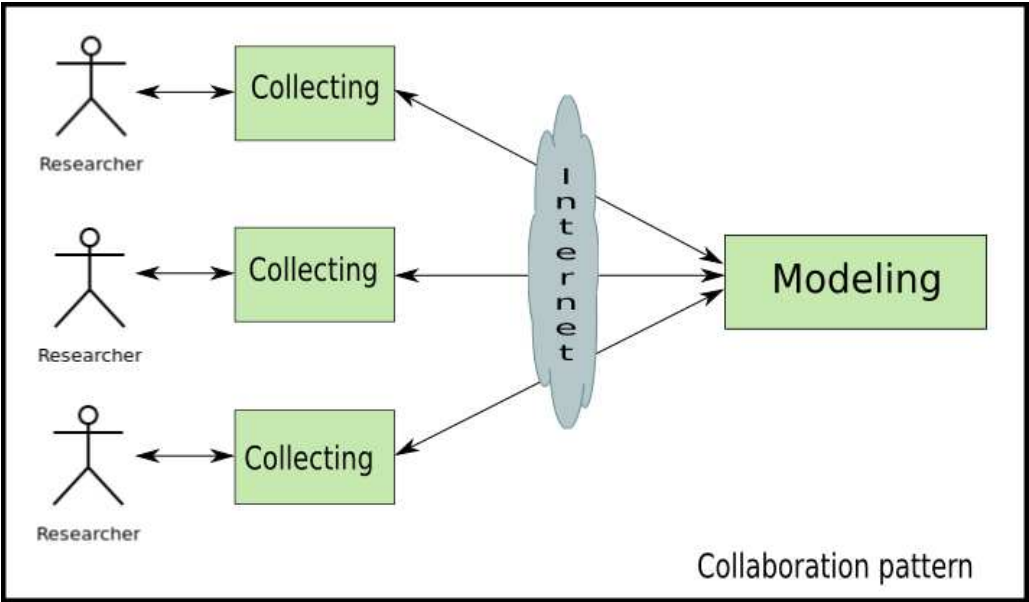


Fig. 5. Collaboration business pattern diagram.

3.2 Goal service-model creation

This second step is to create a model to identify how the services identified are complete with respect to the business. The following is the goal-service model in a nested notation, the mobile system for modeling species distribution:

- 1. Create accurate species distribution model
    - a. Create model
  - 2. Aggregate data from distributed teams
  - 3. Collect data in an automated way
    - a. Collect presence and absence data
    - b. Provide a friendly interaction experience
- Manage project
- Generate view for localization
- View model

In this model the objectives are shown in regular text, service in *italics* and other necessary services that were not discovered during domain decomposition are shown in **bold**. It is observed that the services meet the main objectives of such system. At this stage it was also possible to identify a new service: managing the project. With this the user can add a new project for the species they are studying.

3.3 Subsystem analysis

At this stage of architectural design, business use cases are refined into system use cases. The subsystems are composed by business components and technical components.

High-level business use cases that have been identified in previous steps will be part of the subsystem components interface. The subsystems identified are Modelling and Collecting. The collecting subsystem is responsible for providing to the researcher the tools needed to create a species distribution model. The modelling subsystem is responsible for generating the model based on data provided by the collection subsystem. The components are shown in Figure 6.

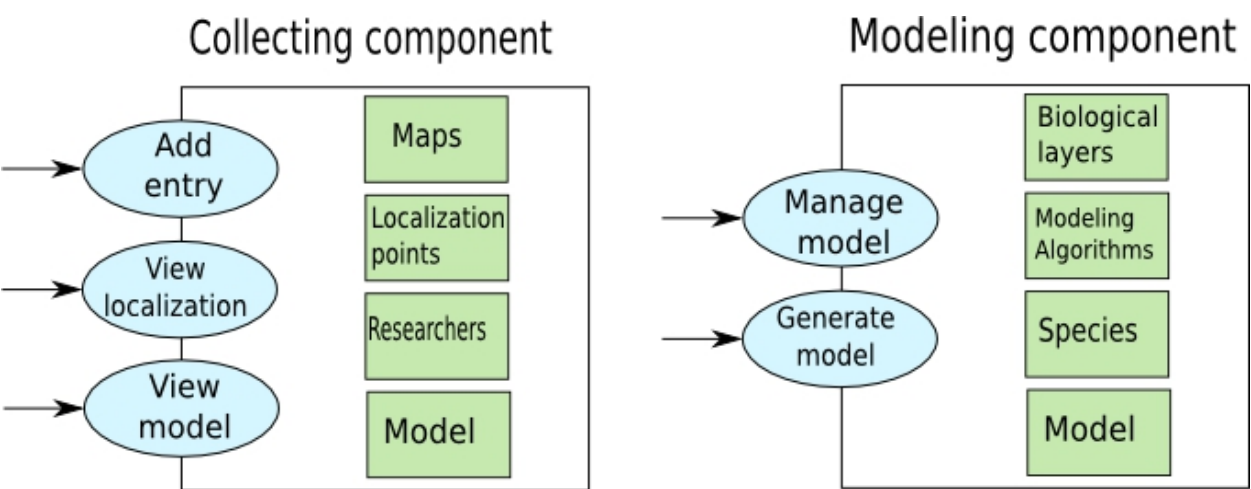


Fig. 6. Collect and Modelling components.

Based on this analysis the services below were identified:

- Localization service: Responsible for providing GPS data (latitude, longitude, altitude and so on).
- Maps service: Provides maps to be viewed on the mobile device. The answers are made in terms of small tiles of defined size. It means the service provides small sections of the map that the client wants. This service is responsible for storing maps to view offline. Provides both maps with satellite imagery, maps and biological layers.
- Occurrence points service: Provides an interface to store the presence or absence points by species.
- Species distribution modelling service: It is used to request the generation of the model for the desired species. In the case of mobile device, this service uses other remote services to generate the model.

The Figure 7 shows the IBM patterns applied to structure these services.

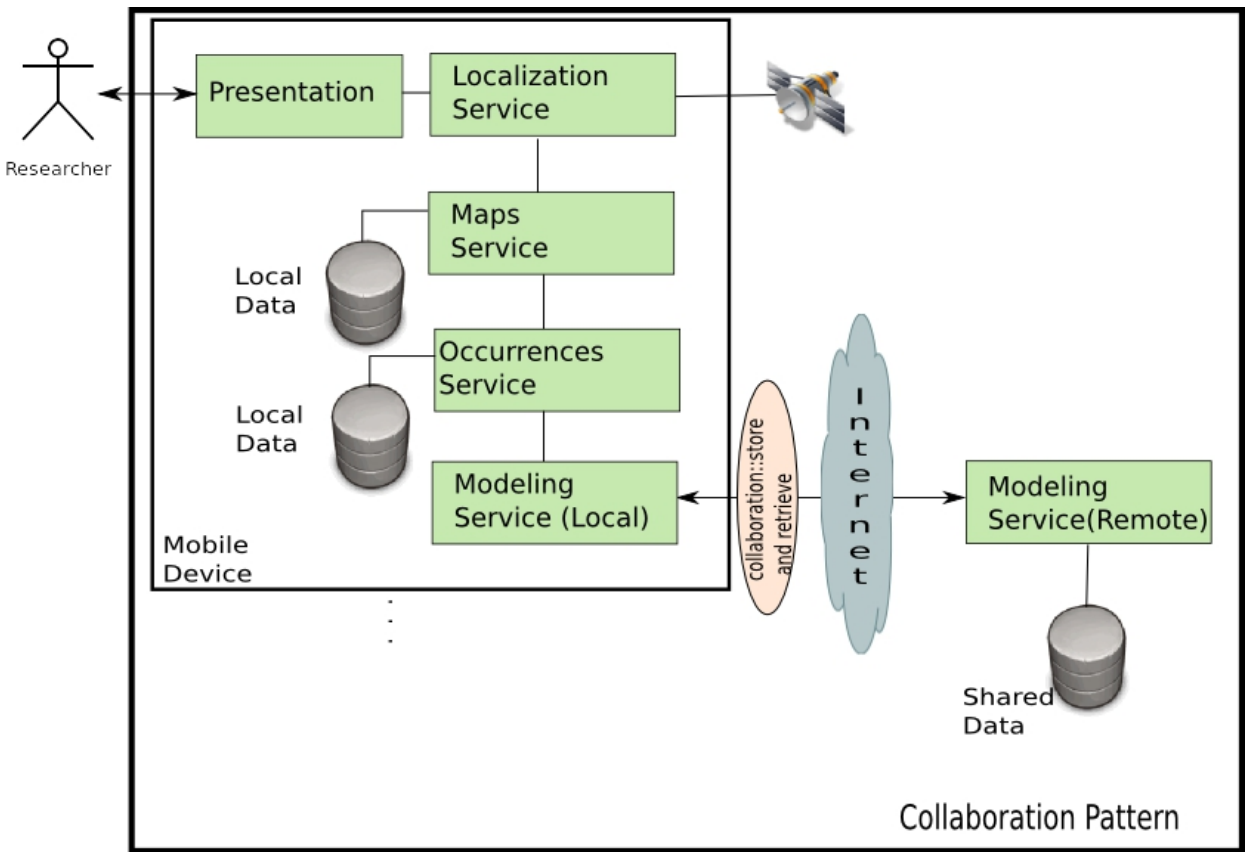


Fig. 7. Architecture with patterns applied and services for collaboration.

The type of collaboration that is used in this system is asynchronous, due to researchers only be able to add their data when there is a network available. According to the standard business collaboration of IBM, there are two types of application pattern: Store and Retrieve and Real-time. This system will use the first one, Collaboration:Store and Retrieve, because it meets the fact that the data does not need to be made available in real time.

This application pattern is further divided into two lines: community and directed. The first is used when data is available for a group of independent users who work together in some common interest. The second is used when data are available for a specific person or a closed and known group. The second one will be used due to, in certain cases, the presence data of some species should be kept in a secure and restricted access.

3.4 Service allocation

The purpose of this step is to ensure that all services will be hosted in a specific place and they all return a business value. The service allocation comes to who (what component) will provide and manage the implementation of each service.

For the proposed system all services will be available on mobile devices to be used by any application. Only the modelling service will be available on an external server with more processing power, however, there will be a customer-service model that will interface with this external service.

The main local services offer the following public interface:

- Location: Implement an easy way to retrieve the geolocation data. Abstracts the use of hardware resources so they are high level and simple. Simply, have the following methods: getLongitude, getLatitude, getAltitude.
- Maps: Service that provides maps to be displayed in applications. It also stores the maps when the device is online. It provides both satellite maps at various scales and biological layers. Provides through tiles (small parts of the map) to be lighter for the mobile device. Main methods are: getTile, and getLayerlistLayer. It also has some signs to notify the applications that are connected to them, such as the signal newLayerDownloaded.
- SpeciesDistributionModel Service: Provides an interface for any application to generate a species distribution model. This service is the one that connects the remote services to building the model, since it is not generated locally. If different models for the project registration in service. It also accesses other local service, which provides the points of presence and absence of a species to feed the model. Its main methods are: registerProject, getModel, addPresencePoints, addAbsentPoints, removePresencePoints, removeAbsentPoints, addServer, addLayer, removeLayer.
- Occurrence Points Service: responsible for storing the points of presence and absence of species. Main methods are: registerSpecies, add (bool presence, speciesId, location), remove (bool presence, speciesId, location).

### 3.5 Technology realization mapping

In this step, the technology used to implement the system will be specified. This technology can change according to other implementations.

On the server side, the services were implemented in C++, used as a CGI application that runs as a server in Apache Web Server. The interface is provided using SOAP (it could be another lightweight patterns). The library used to provide and manipulate data to generate a predictive species model is the Openmodeller running on a Linux machine.

The client side is an application developed in Python and uses the Qt graphics library, specifically the component QML (QtMarkup Language), which allows the creation of complex graphical interfaces using multiple facilitators to the developer. For the integration between Python and Qt, the library for python PySide was used, which lets the Qt Library be used, that is in C++, through Python. Another important library to be mentioned here is the QtMobility, which allows access to device resources such as access to the network card, GPS device, battery information, among others.

The services are registered on the client side services framework in Qt Mobility. These services are exported and available to use in any application, thus enabling its use by third-party applications. Once it uses only libraries provided by the Qt framework and the Qt Mobility, these services can run on any mobile or desktop device supported by the framework. Each service uses an XML that is used to register it as shown below:

```
<?xml version="1.0" encoding="utf-8" ?>
<SFW version="1.1">
<service>
<name>SpeciesDistributionModel</name>
<filepath>species_distribution_model</filepath>
```

```

<description>This service provides an easy way to load and generates an species
distribution model, based on some location data.</description>
<interface>
<name>org.usp.sdmm.SpeciesDistributionModel</name>
<version>0.1</version>
<description>Interface to generate a species distribution modeling.</description>
</interface>
</service>
</SFW>

```

#### 4. Client prototype

Below we present some features of the proposed system. The case study of the system has been performed in Amazonas state in partnership with National Institute for Researches of Amazon (INPA) (INPA, 2010).

##### 4.1 The system importance

The usage of mobile phones for performing data gathering offers the following benefits to collecting and modelling ecological niche distribution:

- Collection automation: A user needs to select only a button on the geographic interface to identify a point of presence or absence of a species. Automatically the system calculates geographic coordination through GPS (Global Positioning System) and stores that data. The user will not have to take notes in other sources or electronic spreadsheet. This also guarantees reliability on the collected data.
- Accompanying changes in the model: With this system it is possible to follow the mobile phone through the model evolution and extent. More data can be added, including data from other field researchers.
- Better use of a researcher's distribution: With the simplification of the system more areas can be analysed by researchers. The researcher can work looking for species in different areas and still transfer data to all other researchers to access.
- Convenience: The mobile device is more compact thereby reducing the number of devices necessary to conduct the research.
- Usability: A more friendly and intuitive interface will be proposed in this document. With this system the user will gain some other facilities to perform data collection such as selecting only one button on the graphic interface to identify the presence or absence of species.
- Faster GPS in regions with GSM (Global System for Mobile Communications) networks: Modern phones do not use the traditional geo positioning system, but the Assisted GPS (A-GPS). It is a system that uses a server for helping to minimize the Time To First Fix (TTFF) and improving the robustness of the positioning. The accuracy of location is less than 3.1 meters and the TTFF are less than 5 seconds, working even in situation of critical satellite signs (Schreiner, 2007). The A-GPS uses any available networks such as the GSM network of the operator or even a wireless network in urban areas. This type of technology use case is important for species modelling which live in urban environments, such as rats and mosquitoes for example, when identifying possible diseases routs (Santana et al., 2008).

## 4.2 Technical restrictions

There are two main technical restrictions for the proposed system. They are:

- Network availability: One of the biggest challenges this project faces is the issue of communication between the cellular and the service provider. Most use cases for a system such as this involve communication networks with intermittent network access so the system shall work most of the time in off-line mode. The main objective of the application is automating the data collected and sending to the server. However it is not intended that data collection occurs at the point of collection, but as soon as there is an available network. Working in off-line mode, the application will be able to store the current coordinates, show saved models and analyse these models.
- Processing power of the devices: The predictive modelling of the species distribution requires high performance computing (Santana et al., 2008). The generation of a model can take between hours and days to prepare.

## 4.3 Business restriction

This type of research in Brazil needs to comply on some rules concerning the transfer of biodiversity data. This data needs, in many cases, to be kept secret. These measures are necessary for environmental property protection.

This kind of concern is important for Brazil, since 20% of the total number of species on the planet is in it. In august of 2009 the Ministry of Science and Technology published 'the concierge 693' which standardizes the manipulation and distribution of research data about biodiversity. This concierge makes the following observations:

- Management and authorship of data must be published.
- Usage condition and the access to data must be protected in the database.

## 4.4 Usability

The system needs to have easy browsing, be intuitive and have few steps necessary for performing tasks. The kind of device that is being proposed for use is a Smart phone, which has a limited screen size. It means that the icons need to be arranged in such to be very easy to use. Figure 1 shows the screen of this tool's main menu in a mobile device.



Fig. 8. Main Menu in a mobile device demonstrating the usability with the thumbs.

This interface shows how efficient it is for small devices that have touch screens. Usage of the thumbs allows more comfortable experience. Figure 2 shows the initial activities with the screens skeleton flowchart.

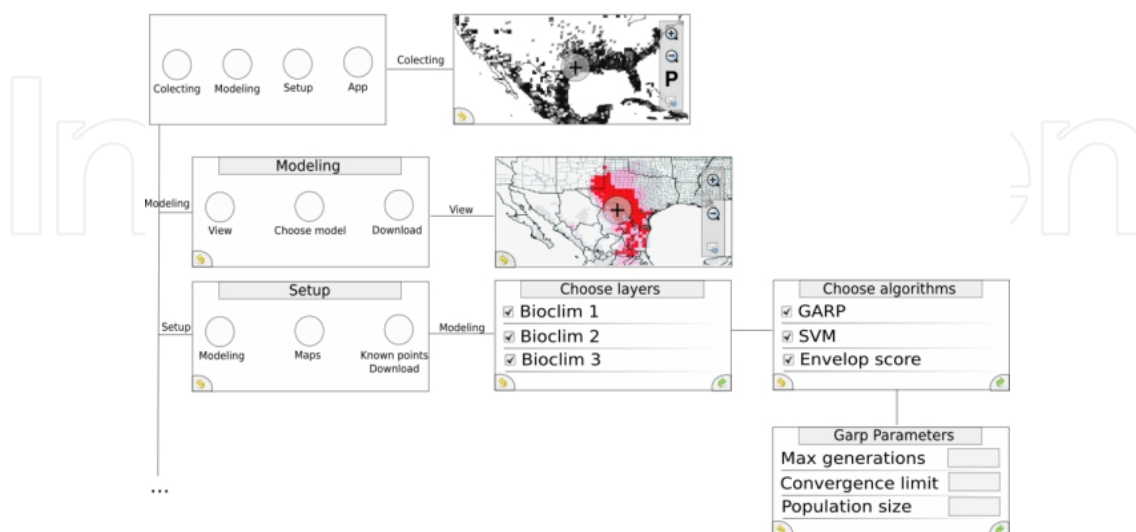


Fig. 9. Flowchart for some screens of the system.

## 5. Conclusion

This section presents the conclusions and final considerations based on the topics discussed in this chapter.

This article presented an architecture based on services to data collecting applied on species distribution modelling using mobile systems. In order to define the architecture, first of all, a follow up was made with researchers of the INPA Brazilian institute and it was based on real case. This phase made possible the process formalization for modelling and collecting data using IDEF0 diagrams. Based on this diagram, we can identify the collecting step really needs automation due to a lot of manual work.

Also, architecture was proposed. It uses a Service Oriented Architecture approach for mobile systems. This architecture is out to be satisfactory for species distribution modelling, since the data resources are by nature distributed. For the devices, an advantage using SOA is the ability for several applications to collaborate securely in collect and model species data. But, on the other hand the use of some standards in communication with the services can be expensive, as the patterns that use XML, which parser can be really expensive

This architecture can be used both to collect data for species distribution modelling, as for similar mobile systems. It is also due to the characteristics of SOA, as services decoupling.

Finally, was modelled a prototype for the application which takes under consideration a friendly user interaction. This prototype is simple to be used in field and reliable for the biological data. It proved very satisfactory to data collecting and generating models that can be used during the collecting phase. Another feature for this architecture is the collaboration that allows users from many sources share environmental data. One of the cons of this prototype is that the features are only basics, and it is necessary to add more features.

This area still has many computational challenges and they can become in future works. During this research it was identified some of those challenges. One of them is to use augmented reality and sensors network to improve the data collecting. Using the sensor networks is possible to identify the species while it is possible to use a mobile phone with a camera and using augmented reality to help the user to find the points and restore some data from the sensor.

## 6. References

- Austin, M. P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, In: Ecological modelling Elsevier , pp. 101-118.
- Andrei, M., Ang, J., Arsanjani, A., Chua, S., Comte, P., Krogdahl, P., et al. (2004). Patterns: Service-oriented architecture and web services (1fs Edition ed.). New York: IBM WebSphere.
- Bazzaz, F. A. (1998). Plants in changing environments: Linking physiological population, am community ecology. Am community ecology .
- Engler, R., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absense data. Journal of Applied Ecology , 41 (2), 263-274.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological modeling , pp. 147-186.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. Ecology Letters , 8 (9), 993-1009.
- Hutchinson, G. E. (1957). Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology , pp. 415-427.
- Isaaks, E. H., & Srivastava, R. M. (1989). Applied Geostatistics (2 ed.). Oxford, Oxford: University Press.
- Iwashita, F. (2008). Sensibilidade de modelos de distribuição de espécies a erros de posicionamento de dados de coleta.
- Malanson, G. P., Westman, W. E., & Yan, Y.-L. (1992). Realized versus fundamental niche functions in a model of chaparral response to climatic change. Ecological Modelling [ECOL. MODEL.] , 64 (4), pp. 261-277.
- Peterson, A. T. (2001). Predicting species geographic distributions based on ecological niche modeling. Cooper Ornithological Society .
- Peterson, A. T., Sánchez-Cordero, V., Soberón, J., Bartley, J., Buddemeier, R. W., & Navarro-Sigüenza, A. G. (2001). Effects of global climate change on geographic distributions of Mexican Cracidae. Ecological modelling , 21-30.
- Philips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling , 190 (3-4), 231-259.
- Santana, F. S., Siqueira, M. F., Saraiva, A. M., & Correa, P. L. (2008). A reference business process for ecological niche modelling. Ecological informatics 3 , 75-86.
- SCHREINER, K. (2007). Where We At? Mobile Phones Bring GPS to the Masses. IEEE Computer Graphics and Applications , 27, 6-11.
- Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. p. 10.

Sohn, N. (2009). Distribuição provável, uso de hábitat e estimativas populacionais do galo-da-serra (*Rupicola rupicola*) com recomendações para sua conservação. Universidade Federal do Amazonas.

IntechOpen

IntechOpen

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen