Biodiversity Information Systems Evolution:
The MABIS model to gather several communities on an adaptable environment

Didier Sébastien, Noël Conruyt, Rémy Courdier, Nicolas Sébastien
IREMIA, LIM-EA2525,
Université de la Réunion
Saint-Denis, France
didier.sebastien/noel.conruyt/ remy.courdier/
nicolas.sebastien@univ-reunion.fr

Tullio Tanzi
Département TSI - LTCI UMR 5141 - CNRS
Institut TELECOM - TELECOM ParisTech
Paris, France
tullio.tanzi@telecom-paristech.fr

Abstract—The computerization of scientific data treatment and the relatively recent awareness of the fragility of the natural world environment (Rio Conference "Earth Summit" in 1992) have led to the proliferation of Information Systems dedicated to biodiversity. Given the various data that they contain, they are complex applications, focused on the needs of one type of environmental scientist, closed to amateur’s contributions and unable to support ethological information. Moreover, the data contained in these systems are hardly accessible to non-specialists like the general public or decision-makers.

At the same time, new communication protocols, like webservices ensure a better sharing of information between applications; while immersive representations of three-dimensional virtual worlds, also known as metaverses, allow a more natural assimilation of information. By putting users in a reality reproduction built from information systems, all entities can be represented in a consistent virtual environment [15]. But sharing and turning Biodiversity Information System’s data in a coherent metaverse is not a trivial process. It relies on an adapted architecture and the availability of specific metadata, as several virtual worlds representing several levels of details can be generated.

This paper is a contribution to the enhancement of Biodiversity Information Systems in order to make them more adaptable, usable, and representable for several kinds of users: different types of specialists, amateurs, and the general public. The MABIS modular architecture is presented as an open, efficient and evolutionary model for structuring Biodiversity Information System. The advantages of its architecture allow to ease its development, store ethological data, manage, share and represent complex entities and metadata while ensuring their authentication through an evaluation process.

Keywords: Biodiversity Information System, BIS, MABIS, behavior, immersion, metaverse, data evaluation, webservice

I. INTRODUCTION

The important loss of world biodiversity has led Eco-Informatics experts [18] to develop specific Environmental Management Information Systems (EMIS) called Biodiversity Information Systems (BIS). The objective is to help communities of practice in Biology (thematic experts, also called “thematicians”) who work on this phenomenon in their everyday tasks: produce inventories of species, obtain a better visibility on biodiversities’ studies, plan coordinated actions between institutions and communicate their results. In this context, Information Systems (IS) are more and more used to manage biodiversity data [6].

Most BIS focus on the management of taxa and specimens [9], but the available functions and uses are directed to fit the needs of a specific category of scientific specialists. For the same substrate (environmental information), it is possible to define three different types of users and needs: researchers, managers, and curators. Each of these professions works with different objectives in the environmental field and thus has different needs in terms of BIS functions (see Figure 1).

Figure 1. Main types of scientific professions involved in environmental data management

The researchers’ objective is to make scientific progress by accumulating new knowledge. They accomplish it mainly by discovering new specimens, analyzing them in laboratories (ex situ), and studying them by experimentation (in situ). On the global scale, the objective is to understand the diversity and evolution of the life through time, space and interactions, as well as the mechanisms at the origin of this diversity.

The objective of environment managers is to discover ways to preserve and restore the ecosystems for which they are responsible. To achieve these goals, they need to uncover specific indicators that characterize the environment and they
must accurately adhere to the policies of environmental restoration they have developed.

The curators' aim is to preserve and manage collections. Fulfilling this task requires acquiring new specimens, sorting and arranging them in order to preserve the taxonomists' primary tool and biodiversity hot spot, but also setting up expositions destined for the general public.

As we can see, in these three scientific vocations specimens are manipulated, but apart from sharing a metadata basis, there is a need to associate different types of entities (documents, maps, collections) with different levels of granularity. That is why, even if their field of work is the same, researchers, environment managers and curators will use a specific BIS to manage their data. This however, implies lost time in repeated metadata entries in the absence of data sharing and communication between the communities that work on the same set of taxa.

Moreover, amateurs are not allowed to contribute on the system by adding their own data in order to share them with the scientific community. Indeed, the biodiversity field attracts a parallel community of non-specialists which gathers important sets of data. Most of time, these data are rejected by BIS only because they were not provided by experts. However, to profit from these data, it would be useful to let amateurs have a controlled access on the system.

Another common limitation of BIS is they do not support ethological information as a specific data type. Because ethological studies are not considered as a specific experimentation, specimen’s behavior is not recognized as an independent entity on the system, and researchers studying them have no access to specialized and structured forms, like ethograms, to enter this type of information.

Lastly, biodiversity information systems contain a huge quantity and variety of data that lead to important assimilation’s difficulties, especially for non-specialists of the thematic. Thus, it is often difficult for thematicians to communicate their results to the general public or the decision makers through the classic Graphic User Interface (GUI) of BIS.

Therefore, we propose an open architecture that tries to find a place for each type of users, from the experts to the general public. The aim is to make information systems more adaptable, usable, and representable for several kinds of users; without losing the scientific credibility.

II. THE MABIS MODULAR ARCHITECTURE

A. General architecture

In order to share common data and knowledge between distinct BIS, it is possible to create a complete set of web services, which translates into a great amount of development for each platform. Another possibility would be to set up an adaptable BIS, corresponding to the needs of each type of user, without imposing a complex Graphic User Interface. Relying on this approach, we propose the Modular Architecture for Biodiversity Information System (MABIS). This model uses several Web applications that constitute modules interconnected together by webservices (Figure 2).

These Web applications are structured in four layers:

- The first layer is dedicated to the management of the main BIS entities, i.e. the directory of participants to the project, the thesaurus for explaining the meaning of terms, the Documentary Multimedia Database (DMD) [14] for the studied objects, the systematics module for storing, organizing and presenting specimens and taxa, the cartographic module for georeferencing them, and the Behavior Management Module (BMM) to register ethological data.
- The second layer provides tools to authenticate the information stored in the first layer (evaluation module), to allow curators to archive collections’ data about specimens, and to manage the several types of entities from the first layer in the frame of contextual environmental projects, for example to provide managers the means to monitor their protected areas (follow-up module).
- The third layer provides the software tools to visualize, represent, analyze and treat the data from the first two layers. For instance, the Biodi-Verse software is dedicated to the generation of immersive views for selected data, whereas the Multi-Agent Simulation (MAS) software offers simulation’s possibilities with the codified information on specimens [2],
- The fourth layer is dedicated to vulgarization. It hosts Web portals relying on data supported by the other layers.

Figure 2. MABIS general architecture

More than the advantage of facilitating its development, thanks to its modular architecture, MABIS let each type of user focuses on his immediate task: each module constitutes
an entry point (Figure 3) corresponding to precise expectancies and functionalities. Furthermore, to ensure exchanges between the MABIS architecture and other BIS, two gates allow remote controls on the main layers: the Web Components Services (WCS) gate opens on functionalities offered by the two first layers, whereas the Web Software Services (WSS) gate provides a convenient access to the third layer.

B. The modules

The modules constituting the IS are applications dedicated to the management of a limited set of entities’ types. All the information related to an entity (for example, a document) is gathered on a structured card which can be enhanced with elements from another entity (Figure 4).

Each module has three functioning modes: main, deported, and remote mode. In main mode, users work directly in the environment of the Web application dedicated to the management of the entity he is focusing on. The deported mode is used to provide popup windows and inclusions offering synthetic graphical information’s and functionalities from a module in another one used in main mode. The remote mode means that the Web application is used through its webservices, in order to fully integrate its data and functionalities in another module’s GUI.

Moreover, a distinction can be made between modules, taking into account their complexity. The Web applications of the first and second layers are dedicated to the management of entities. They can be fully used through Web browsers in the three modes and constitute WCS. By contrast, the Web applications of the third layer, provide complex treatment tools on primary data of the first layer, so rely on a main mode requiring a typical installation on computers. Thus, this mode is mainly not working like a SaaS (Software as a Service) application. Only deported and remote modes can be used through Web Browser. These modules constitute WSS.

In order to make easier the assimilation of the Web Components Services, which represent the common entry points of new primary data, we have standardized them to keep the same logic of entities’ management. Indeed each entity’s type has to support common categories to guarantee a good communication between contributors. The simple template represented on Figure 5, established through a human centered development [14], has led the general organization of all the WCS deployed in MABIS.

III. THE EVALUATION OF SCIENTIFIC INFORMATION

The biodiversity field attracts a community of enthusiasts, non-expert of the thematic: amateurs. They gather a huge quantity of data that are often not accepted on BIS because the providers are not specialists. A solution to this loss of information is to support a scientific data validation policy, which is often made by an administrator in Biodiversity Information Systems. However, because the quantity of data expert-administrators have to validate is colossal, they cannot afford to treat amateurs’ resources.

In this context, we propose a Scientific Information Evaluator (SIE) model. This model relies on the MABIS architecture and comes into the form of a SIE module that manages the evaluation of all entities’ cards distributed in the IS. Our aim is not to provide a methodology to analyze data, for instance by error elimination [12], but to deliver a tool that follows the enhancement of a card until it reaches the best acknowledgement level. This module offers several ways to authenticate the scientific aspects of data, to share the evaluation’s work, to simplify the communication between experts, and to ease the identification of the evaluation’s state by end-users. The enhancement process can be represented as in Figure 6.

Figure 5. Template standardizing the WCS

An instance of this template can be seen on Figure 9.

Figure 6. Evaluation of scientific information in MABIS model
Contrary to most validation processes that result in a binary answer (validated/not validated), evaluations result in certificates associated to a precise time. Each certificate corresponds to a level of trustworthy that gives a more precise idea of the data condition (Figure 7). The certificates can be considered as a succession of steps toward the validation level. In order to reduce the work of data evaluation by experts, two evaluation systems are offered.

- level 0: not evaluated
- level 1: can be greatly improved
- level 2: still need improvements
- level 3: almost finalized
- level 4: validated at present time

Figure 7. Levels used for the evaluation

The authoritative certification represents the evaluation of the current version of the entity’s card made by identified specialists of the thematic on the BIS. Depending on their recognized specialization in a discipline, a limited set of experts acquire from the administrator the possibility to deliver certificates that represent their approval to the data. This ability to deliver up to a defined level of certification is entity-specific, and for the taxa, taxa-specific. So the SIE stores a list of privileges that can be associated to profiles of users registered in the directory. The propagation of this certification is regulated by a cooptation system that let specialists share their privileges to other users up to their own level. Thus the administration and the evaluation work are shared. This evaluation based on hierarchy is summed up by a simple icon that represents the lowest certification given by the highest specialist. Experts’ debate, i.e. when two or more specialists at the same highest level deliver different certificates for the same card, is automatically detected and notified by the sign “!”. In this case, users are warned that it could be necessary to read the comments and evaluate themselves the card.

The community certification is the evaluation of the card made by all identified users on the BIS. Because everyone can participate and deliver the certificate they consider deserved, this evaluation is less specialized than the precedent one; however, a certificate coming under an expert’s debate, i.e. when two or more specialists at the same highest level deliver different certificates for the same card, is automatically detected and notified by the sign “!”.

These two certification modes are easy to use (a simple button) and can be used as criteria, in a combined way or not, to sort and research data. Thus, it is easy to work only with, for instance, data defined as relevant by experts.

(Figure 8) adds three inclusions in the graphical interface of cards, to:
- indicates the level of validation, which represents the level of approval of the information. Each level is represented by an icon that sums-up the participative evaluation of a data.
- presents detailed information and an historic about the evaluation of the card (all marks from all voters),
- offers a simple thread that permits authors and evaluators to discuss about the card.

The SIE’s main mode (Figure 9) is the Web application that gathers all information about evaluation of entities. It allows to:
- present general metadata about entities’ evaluation on the BIS (most discussed cards, most heterogeneous evaluation per module, etc.).
- list all evaluations’ requests and focus on those concerning the authenticated user.
- show the logged in user his own evaluation requests, delivered certificates, discussions’ threads (new messages since last connection), and the current evolution of his cards.
- manage the sharing of privileges concerning the authorized level of certification.

The Scientific Information Evaluator’s deported mode

The Scientific Information Evaluator’s deported mode

The remote mode of the SIE is of course not related to a specific GUI, but it is often used in other modules, for instance to sort entities by certification level in other Web applications’ main mode.

IV. THE MANAGEMENT OF ETHOLOGICAL INFORMATION

A. States and actions, a difference between specialists’ approach and system’s structure

Ethology is the science that studies the behavior of living species. Biologists that analyze the behavior of animals (also called “ethologists”) focus their researches on the establishment of lists of states and actions, the determination of their occurrence and linking. By observing several individuals of a taxon, they identify their way of life, gaits, interactions between them and other species, impact on their ecosystem, etc. These observations are usually made during short analyses’ sessions, on a limited number of specimens; so a lot of data is produced in a short duration, often associated to a precise scale of time (minutes or seconds). This kind of scale is hardly supported by common BIS, on
which temporal data associated to a biodiversity project (like a monitoring) uses larger time’s scales (days).

Independently of the structural aspect of BIS that has to evolve to handle the information associated to precise scales of time, the user interface has to be enhanced in order to facilitate the data’s entry process. That is why the best way to integrate this dimension without compromising the human-computer interaction is, in the frame of our architecture, to add a new WCS: the Behavior Management Module (BMM). As we have said, ethologists focus on states and particularly actions (movements) performed by the studied species. Indeed states and actions are linked: particular characteristics of internal (not visible) and external (visible) state lead to specific actions. Without a sustained analysis of a specimen, it is hard to evaluate its state (for instance, his hunger): researchers have to determine it from the actions that will be induced. So even if a state leads to specific actions, for animals, ethologists often deduce the precedent state from the following action. That is why, to follow their methodology of work and processes, our module focuses more on actions. However, by developing tools permitting biologists to gather and represent dynamical data through a formalized approach, this work also contributes to initiate a discussion with experts to enhance their methodology of study.

B. Implementation in the MABIS architecture

In the frame of our architecture, entities from the WCS are intended to be used by the WSS as primary resources. Thus, it is interesting to take into account the use that will be done with the ethological data since the conception of the application. Specimen’s behavior information, by facilitating the extraction of rules leading agents used in MAS, constitutes relevant information to build simulations. Moreover, the description of actions stored in the behavior module can be useful to generate immersive representations of data. That is why, in the implementation of this module, we have anticipated the use that will be made in terms of data treatments and representations.

<table>
<thead>
<tr>
<th>Text</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Action</td>
</tr>
<tr>
<td>Representation</td>
<td>MAS</td>
</tr>
<tr>
<td>body description</td>
<td>list of relevant factors and way to determine their value</td>
</tr>
<tr>
<td>description of body movements</td>
<td>Rules of variation of the values</td>
</tr>
<tr>
<td>animation between static poses</td>
<td></td>
</tr>
</tbody>
</table>

TABLE I. DESCRIPORS FOR STATES AND ACTIONS

Current BIS are able to handle temporal data associated to specimens as a succession of observations describing its general state [2]. If this information could be sufficient in some cases, for instance phenological studies on plants, it is not in many others, like behavioral studies on animals. Finally, only little information can be automatically extracted from Systematics module to generate rules necessary to build Multi-Agent Simulation, as the observations are consigned in generic text fields. So there is a strong need to formalize spatio-temporal metadata about specimen’s evolution and behavior. However, there are several ways to manage behavioral information in BIS. For instance, it is possible to consider either only one default state with actions as variations to this referential, or several states associated to actions that make the links between them. In order to clearly separate actions and states, we chose the second possibility structured through three approaches: textual description, visual representation, and formalized rules describing specimens’ behavior as agents in MAS (TABLE I). The objective is to establish a consistent timeline (Figure 10) made by states and actions, linked to the Systematics module, and ready to provide a formalized support to the immersive representation and ecosystems’ simulation.

![Figure 10. Consistency of timeline offered by the BMM](image)

The way specimens’ states are described also has to change for a more generic aspect. This directly impacts on the general Systematics module’s structure. Instead of describing several times the same state of a specimen as several observations, it is possible to define a set of generic states associated to a taxonomic level. Doing this way, an observation on a specimen becomes the junction between a generic action of the taxon, and a specialized scientific annotation (Figure 11). The process is, of course, the same for states. Indeed the introduction of a BMM in BIS, and its strong relation with the module in charge of taxa and specimens’ management lead to several evolutions of the Biodiversity module. The main improvements are explained in parallel of the descriptions of the three BMM’s sides, but first, we have to take over global aspects of this tool.

![Figure 11. BMM entities’ integration in MABIS](image)

The BMM aims to manage specimen’s behavior in two ways. The first one is to store labeled descriptions of actions and states achieved by specimens in the frame of a behavior process to achieve a goal. Ontologies are constituted by linking actions to a taxonomic rank in order to specialize or generalize it. For instance, the action “eat” can be defined in general for mammalians, and a second occurrence can be added to specialize the action for the Felidae family. Lists of different actions focusing on the same taxa can be gathered on a single view in order to constitute ethograms.

The second one is to store lists of actions related to specimens to constitute spatio-temporal capture sessions that
focus specimens. Indeed ethologists mainly study animal behavior through three methods:

- Focus sampling with continuous recording, which focuses and lists all actions achieved by a specific specimen during a very limited time.
- Scan sampling, which points out all relevant actions achieved by several specimens.
- Ad libitum sampling, which means that the observer records as much as he can of whatever he can see. This method is generally used to get an overview of the specimen’s group.

In all cases, the user has to associate or describe actions made by one specimen or more. Digitally speaking, it is just a matter of presentation, by action or by specimen. After entering the data on the system, it is easy to sort them by using the BMM.

So, in terms of use cases, the BMM answers to several users needs. Firstly, it will help ethologists to provide standardized, easily interoperable by computational request, action cards. Secondly, they will easily constitute and share sampling sessions on the Web. Thirdly, it will allow them to easily communicate with the general public their observations by showing immersive reconstructions through immersive representations. Fourthly, simulation experts, will access to the module as a warehouse of generic and specific models for their MAS, knowing they could possibly represent their results in an immersive way. Fifthly, the collection of 3D models, animated or not, gathered in the frame of the BMM will constitute a warehouse of species’ shapes and movements for computer graphics experts.

As we have explained, states are often deducted from the following actions, so action represents the main point for ethologists. In this frame, we will now focus on how the “action” entity can be integrated through the three approaches in the same module.

C. Three representations for one action

To build a reliable and generic action’s description form, that allows a relevant generation of action’s description cards, we have to consider the definition of “behavior”. For J.B. Watson, the behavior is “the whole of the objectively observable reactions that an organization generally equipped with a nervous system carries out in answer to stimulations of the medium, themselves objectively observable”. Note that this definition is particularly opened, and underlines following key points. Behavior:

- refers to any specimen, whatever its phylum is (not restricted to animals),
- considers the observation’s environment,
- focuses on objective actions.

Regarding these clues, and usual form’s fields, we tried to build a generic action’s description form that will gather main data (Figure 12). This general part contains textual and numerical data that are used by several users (ethologists, simulation experts, computer graphics experts). These fields allow multiteria searches to retrieve specific actions and are highly used by the simulation and the representation engines (especially the duration and the periodicity fields).

In the frame of a BIS’s module, the aim is to formalize the descriptions in order to meet both user’s expectations and system’s requirements, for instance, to avoid future plain text searches. The consistency of the integration of the behavior module in the IS relies on the links each entity (action, state, and sampling session) for each approach (text, simulation, representation) develops with other modules. Given the nature of the entities and their metadata, the systematic and the cartographic modules are particularly important in their description cards.

1) Descriptive aspect

The textual description (Figure 13) of specimen’s behavior constitutes a classic approach of an ethological study. It is dedicated to specialists in Ethology, as the technical vocabulary they use is sometimes unreachable for the general public. That is why visual representation is so important. The general context field describes the conditions that must be gathered, for the specimen and the environment, to produce the action. It is described in a more formal way in the simulation aspect’s part.
2) Representation aspect

This part collects the files made by ethologists and computer graphics experts to graphically represent the actions. The system accepts as inputs several files corresponding to several representations. Traditional graphical representations, in two dimensions, is relevant in the frame of the IS because it allows thematicians to gather all their data on the same system. Indeed, with the affordability of digital camera, thematicians produce more and more videos and photos to illustrate their researches. Furthermore, immersive environment are able to display this information on 3D surfaces as animated textures when tridimensional meshes are not available. That is why we chose to support this usual approach on the graphical action’s description card’s part (Figure 14). However, as we also aim at representing immersive representations, the application gathers all files needed to prepare this process.

Thus, several graphic-representation files are supported:

- The icon is an illustration created by the user to represent a specific action of a species. It can be used either in the IS to graphically sum-up a sampling session, or by the multi-agent simulation engine for its usual 2D visualization, or by the 3D engine as a texture.
- The photos and videos are multimedia documents representing actions, linked to its card. Because biologists usually take several photos and sketches to illustrate the same action, it is important that the system allows them to gather all their pictures by handling multiple file uploading.
- The “Animation bones” field refers to a tridimensional spatio-temporalized description of animation, in the BVH (BioVision Hierarchy) format, which is very used by motion-capture (“mocap”) systems. Indeed, an animation description file can be built with several methods (handmade by a computer graphic expert, or with an automatic acquisition system). The description of action can then be applied to the 3D model of the specimen in order to make it reproduce the movement. This step requires a computer graphic expert to help the thematicians.
- Because it is not always possible to apply an animation to the 3D shape linked with the specimen in the Biodiversity module, a possibility is given to directly upload an animated model that shows the 3D reproduction of an action for the specimen. In this case, the COLLADA format, which stands for “Collaborative Design Activity”, is favored, as it is an open and compatible file.

All files linked to this part are stored and managed by the documentary multimedia database of the IS, and shared by webservices in a transparent way for the user.

3) Simulation Aspect

The Multi-Agent Simulation (MAS) aspect of the Behavior Management Module aims at gathering information concerning taxa behavior for multi-agent simulation. Toward the generation of metaverses, we use the multi-agent paradigm as presented by Ferber [3] to define the different components that take parts in simulation:

- Agents are used to represent specimens. Indeed, agents are autonomous entities that have their own partial perception of the environment they live in. They also interact with other agents and with the environment.
- Environment is used to describe the landscape of the simulation and the global evolution rule of the ecosystem. In multi-agent simulation, the environment is the world agents evolve in. It can adopt many topologies: spatial or non spatial, with metrics, etc. The environment also defines global evolution rules: for example gravity or temperature evolution during a year.
- Objects are used to represent some specimens that do not play a major role in the simulation. In our description, some specimens can impact the behavior of agents but their own behavior is not really interesting considering the simulation scope. So to avoid useless computation, these specimens are not described as agents but as situated objects that interact with agents.

Classic BIS provides information for environment initialization: many layers of geographic information systems are used to describe the topographic landscape and to provide objects or agents localization. Based on this layers principle, we defined the Dynamic-Oriented Modeling [11] which uses the multi-environment approach. This approach enables the splitting of the environment into sub-environments. Each one of these environments contains a specific information layer. For example, a first environment will contain topographic information, another one will be used for messages exchanges between specimens and a third one will describe the vegetal food repartition.

In GEAMAS-NG [17], our MAS platform, we also described a new temporal approach: the Temporality Model [10][16]. This model enables agents to define their own activation times and to link them with periodic behaviors. For example, an agent can define a temporality associated with its reproduction period: this temporality will periodically trigger the agent reproduction behavior. Moreover, the Temporality Model perfectly matches action’s description proposed by the general part of the Descriptive Aspect of the Behavior Management Module.
Toward the building of metaverses, the main issue remains the definition of taxa’s behavior. Indeed, both the Descriptive Aspect and the Representation Aspect of the BMM aims at gathering information on a particular specimen: thematicians must analyze this information to determine the taxon’s generic behavior. To build a taxon’s multi-agent representation, they must define its state, i.e. the key internal (physiological or intellectual) parameters that drive its behavior, its perception capacities, i.e. its sensitivity towards the environment stimuli, and its action capacities, i.e. the way it updates its state and modify its environment.

To define the complete behavior of a taxon, a first step consists in defining the set of actions this taxon can undertake. In the Behavior Management Module the Simulation Aspect is linked with the other aspect and so centered on actions description.

Defining an agent’s action consists in answering three questions:

- What did trigger the action? What are the external or/and internal causes that made the agent undertake the action?
- How does the action impact the agent’s internal state?
- How does the action impact the agent’s environment and other agents?

Answering these questions requires knowledge of the environment the taxon evolves in, and its state. In the Multi-Agent Simulation Aspect of the BMM, we provide basic edition for action definition: having defined the key parameters of the taxon’s state and the parameters of environments, the user can select some of the parameters and propose the action precondition (Figure 10). In the same way, user defines the consequences of the action on the taxon’s state and the environments (Figure 15).

We chose to provide a basic formalism so that thematicians can fill in the Multi-Agent Simulation aspect with a minimal assistance from modelers. The modelers, or the multi-agent system, can then adapt the behavior to more complex formalisms like DEVs [20] or Netlogo [19].

Figure 15. Description’s fields of the behavior’s form dedicated to graphical representations

<table>
<thead>
<tr>
<th>Preconditions:</th>
<th>MAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal state</td>
<td>age</td>
</tr>
<tr>
<td>External state</td>
<td>perceive food</td>
</tr>
<tr>
<td>Internal consequences</td>
<td>age</td>
</tr>
<tr>
<td>External consequences</td>
<td>move</td>
</tr>
</tbody>
</table>

D. General considerations on the proposed solution

The three approaches used to describe the same phenomenon “action” can be identified as equivalent to the Model-View-Controller (MVC) paradigm developed in computer science. The textual description of an action corresponds to model’s presentation; the graphical description’s section is a view; whereas the simulation’s part represents the controller. By reproducing the MVC paradigm in the frame of a BIS module, we provide a stable and adaptive structure to the immersive world’s generation process.

As we have said, each of the three aspects corresponds to a specific profession: ethologist, computer graphics expert, and simulation expert. Even if each of them can limit their use of the BMM at their specific section, the real improvement in the process of porting a part of the IS as a virtual world is reached when their work can be merged in one representation: the meaning given by the ethologist, the quality of representation provided by the computer graphic expert, and the formalism by the simulation expert. To obtain the best result on a specific mirror world, a meeting between the three specialists on a co-design platform [1] is necessary to ensure the correspondence of the three aspects. Indeed, without a good communication between the ethologist and the two other experts, the consistency of the three aspects is not guaranteed. Only the expert in behavior has the scientific knowledge to accurately describe relevant movements done by specimens during specific behaviors.

However ensuring the communication between the experts is not a trivial task. In this way, multi-competent profile experts constitute a real advantage for the system, as they could check data consistency. This should be done at least for entries at the top of systematics hierarchy. Thus, the validated data can then be considered as references for users working at more specialized level of the taxonomy. In this way, it is important to feed the BMM with a set of generic initial values describing very common actions for each kingdom of systematics. These data will provide a frame for a future evolution of the module in terms of error prevention and detection. Moreover, it is important that ethologists use the same labels for equivalent actions and states at different levels of the systematics, in order to ensure the automation of the specialization and generalization process in data representation and simulation.

Alongside these recommendations that contribute to the proper deployment of the application, several improvements of this tool can be set up. More than an automatic comparison between specific and generic data, a comparison between different occurrences of equivalent actions for different taxa would greatly improve scientific knowledge in term of species’ evolution. Another track of BMM technical evolution would put up with the enhancing of the interoperability between several BMM. Indeed the BMM shares its data with other modules and other BIS through webservices. Evolving the BMM to a meta-component able to exchange, compare and analyze data of other BMM would provide a controlled approach on information availability, validity, and contributor’s relative participation. The more structured data is available, the better the quality of simulation and representation will be.

V. THE GENERATION OF BIS’ IMMERSIVE REPRESENTATIONS

Because of the diversity and the huge amount of information contained in BIS, it is very difficult to evaluate the contribution of a user, to understand and keep in mind all the data associated to a specimen or a project. Furthermore,
this information is hardly understandable in that shape to the general public. In the same time, the progress of real time 3D technologies allows to build new representations of information by creating virtual worlds, also called “metaverses”. These metaverses are built up by an aggregation of 3D models which can be seen from any point of view at any time (spatiotemporal representation). These 3D shapes can represent any object, with a variable Level Of Details (LOD). Because metaverses offer visualization close to reality, it is more easy to figure out and understand specific configurations of the IS entities through their realistic representations [7]. Modeling research results through virtual immersive learning environment also constitutes an ideal way to analyze and communicate them with decision-makers and the general public.

However, obtaining 3D representations of information systems, and specially those dedicated to biodiversity management, is not a trivial task since the creation of the virtual world must be automated to support on demand generation. This section is a contribution to achieve this goal through a Biodi-Verse module, by defining a generic process to build metaverses from IS data [15], and establishing a typology of the different virtual world models that can be obtained.

### A. From 2D to 3D BIS representation: General Steps

The principle of porting IS information from a 2D interface to a 3D immersive environment assumes that metadata associated to its entities provide a way to place (through lat-long coordinates) and to represent (3D models) them. If this last point is missing, then a substitute like generic representations (3D icons) can be used. We must precise here that our work aims to produce three-dimensional models that can be obtained.

1. **The user (thematicians, decision-makers or general public)** expresses his request to the system, that is to say, defines the part of the IS he wants to see as a metaverse. In order to limit the processing time, it is necessary to reduce the research’s field by submitting a context of request to the user. An acceptable basis is to focus on the request on a project (i.e. a metaverse representing all the entities of a particular project), or on users (i.e. a metaverse representing all entities belonging to specified users).

2. The description of the metaverse should be obtained by a software, the Virtual World Builder (VWB) [13], which interrogates the BIS (by using, for instance, its webservices). This analyzer works in 2 steps. First, it takes a “photo” of the IS data (depending on the above request) at a precise time by creating an XML World Description (XWD, see Illustration 1) file gathering all the instances and metadata that have to appear in the virtual world. It creates a list of the entities that matches the request’s specifications. Then, the analyzer builds the metaverse architecture that will contain the entities and places them on it. The more accurate the VWB makes this generic XWD and the more accurate to the IS the metaverse will be. The configuration of the analyzer by users allows the software to determine the world’s architecture that will contain the representation of entities.

3. The description of the virtual world can then be combined with 3D models by the 3D engine in order to create the metaverse corresponding to the user visualization request (configuration of the VWB). Depending on the 3D engine, the XWD file is converted in a compatible input format. The specificities of the representation (level of details, representation type) are settled by the user before the metaverse generation, in order to lead and limit the complexity of the 3D world. It can be seen as a second filter, after the initial request.

As in all metaverses, a single click on a 3D entity can show the 2D card of its metadata from the IS, this representation combines the advantages of standard representation with the immersive visualization. We will now focus on the possibilities the Virtual World Builder could offer by defining the different appearances that should lead the generation process.

### B. Typology of immersive representations for IS dedicated to biodiversity

Depending on the nature of entities, and the information linked to them, several representations are possible. In this part we will present the two models of representation and their declinations that correspond to four levels of details.

1. **Tree of rooms**

   In order to represent the maximum of information from BIS, it is necessary to imagine a way to generate a metaverse compatible with all their entities, whatever the nature and the number of metadata associated to them would be. The tree of
rooms is the first model of representation, and the most general one. It consists in a schematic 3D view of several rooms linked by tunnels, forming a tree data structure. Each tunnel gives information about the following room which contains 3D icons of entities’ instances.

The positioning of these rooms and the connections between them are driven by the XWD which is itself built from the user request. Because each entity has a specific nature (contact, document, taxon, specimen, observation, map, definition and project) and belong to at least one user, we assume these two points can be used as default classification parameters for the representation of the entities in virtual worlds. Most of time, instances of entities are gathered in collections and subcollections in the Information Systems’ modules. This third parameter should then be suggested to the user, in order to refine the generic arrangement of the metaverse.

Thus, the user chooses in his request which restrictions could be done (for instance, build a world representing the files shared by a specific user), and the kind of hierarchy that should be used to build the XWD (example: by user, then by nature of entity, then by collection). Note that if the request aims at representing a whole biodiversity project stored in the project management module, the hierarchy is inherent to the project’s structure, so it is not necessary to ask the user for classifications or hierarchy details.

Because the size of each room depends on the number of files contained in it, the general architecture of the metaverse gives a good idea of a project importance, or of users participation in the IS content.

As we can see on Figure 17, there are a lot of lost spaces beside tunnels in the metaverse architecture. But that is one of the advantages of virtual architectures: usability and appropriation can be the main focus of the world’s shape, independently of cost or space optimization’s matters. 3D representation also helps to show several levels of granularity. Thus, if the tree data structure is used to represent groups of entities, relations between them stored in the IS can be represented by wires linking them above the room’s walls (Figure 18). So users can “jump” from a document to all the related ones easily. Of course these links can be deactivated on demand in order to preserve the visibility of documents. In that way, this first level of representation offers several granularity levels corresponding to several data structures: one tree of hierarchical entities for several graphs of related documents.

2) Mirror Worlds

One of the specificities of IS dedicated to biodiversity is that they mainly deal with entities that have a physical materiality in the real world. In that way, it is possible to generate a virtual world trying to reproduce reality with 3D shapes correctly positioned. This kind of virtual worlds are called “mirror worlds” and constitute the second model of representation. Three increasing levels of complexity can be defined and mixed together with this model.

a) Static mirror worlds

Static mirror worlds try to reproduce reality at a precise frozen time: it is a photo of a scene reconstitution. All represented entities need a precise geolocation in order to be placed in the virtual world. Furthermore, to proceed to the automatic generation of the instantiated entities, several descriptive metadata are required to build their three-dimensional shapes. TABLE II. presents a list of entities that could be shown in mirror worlds, and the associated descriptors (other than location) that would greatly help to automate the process of shape customization.

```
<table>
<thead>
<tr>
<th>Entity</th>
<th>User</th>
<th>Specimen</th>
<th>Map</th>
<th>Sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common in most IS</td>
<td>gender</td>
<td>associated photo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not common. Improve virtual objects recognition</td>
<td>real dimensions (width, length, and height)</td>
<td>viewpoint and clipping of associated photos</td>
<td>altitude of the viewpoint</td>
<td>scope</td>
</tr>
<tr>
<td>Generated Object</td>
<td>3D avatar</td>
<td>3D shape or textured pictures adapted to visualization angle</td>
<td>move view at a precise position and zoom</td>
<td>sound played at the defined position</td>
</tr>
</tbody>
</table>
```

TABLE II. ENTITIES VISUALIZATIONS IN MIRROR WORLD AND METADATA NEEDED

This level of details has two main advantages. First, because static mirror worlds tend to reproduce reality, they are directly understandable by all kind of public. Second, they can display a configuration at a precise time and place,
so it is possible to represent and immerse the user in configurations that do not exist anymore.

But mirror worlds also bring new constraints and difficulties compared to the first model:

- Mirror worlds are usually built by placing 3D models on a background map (generally an orthophotographic view of the place) as ground. So it is necessary to have a background map of the place which needs to be represented, at a scale that fits user’s request.
- Depending on the request, it is difficult to make a metaverse adapted to several scales (microscale or macroscale). Without the assistance of a proper Head-Up Display (HUD), users risk to miss existing information.
- It is impossible to represent entities that have not a physical materiality (for example a definition in the thesaurus). If it should imperatively be represented in the virtual world, then it would be compulsory to use the 3D icons like those used at the tree of rooms’ level.

More generally, when an entity’s representation is not available at a level of detail, it is possible to use its representation at the precedent level. For instance, if a generic 3D shape is not available for a specimen, it is possible to use its photo as a planar representation.

b) Scripted mirror worlds

The scripted mirror world is the second level of this model and introduces the notion of time in the generated metaverse. Indeed, BIS contain temporal information. Especially, project management modules contain information like “specimen S of taxon T has been observed (observation O) at time tx and place P (map) by user U”. Several data from a specimen studied in the frame of a monitoring experimentation constitute the script of a chronosequence that can be represented by an interpolated animation: the movement of thematicians and specimens’ representation in the metaverse (Figure 19). Of course, depending the size of the gap between interpolated keys, an important bias can be introduced, but the resulting animations are still interesting for the immersion as long as this uncertainty is clearly notified to user and identified in the 3D scene (opacity modified during interpolation).

With the implementation of a behavior module like the BMM in MABIS, common actions (like eating, sleeping, etc.) could then be represented. The most complex available representation is used to show the action: tridimensional shapes are better than videos, which are better than photos, themselves superior to icons in terms of immersion experience. If none are available, it is then possible to find a substitute at a more generalist level of the systematic hierarchy (Figure 20). In the case of an immersive representation, depending the choices made by the users during the visualization’s configuration, a specific 2D representation (icon) can be given a better importance than a generic 3D representation.

The animation in the virtual world can focus on a specific place or follow a particular specimen among the other inventoried in the area of study. Because this representation allows a more natural visualization of data, it is often easier to analyze than the original script. By enlarging the research scope of the VWB in the IS (from a project’s entities to the whole IS), and adjusting the tuning of the passage of time, researchers can discover which other specimens and thematicians come at the same place but different times. Thus, it is possible to identify inter-species relations and establish collaborations between specialists working on nearby subjects.

So, this level of representation makes a new step toward the representation of reality, which is very useful to handle specific configurations of entities. Showing animated chronosequences is also demonstrative and useful for the analysis of collected data, but, given the focus made by IS dedicated to biodiversity, restricted to the representation of users and specimen (Directory and Systematics modules). However, it uses the information from other modules (e.g. cartographic module).

c) Simulation metaverses

The scripted mirror world has introduced the time factor by representing actions described in the project management module. The third, and highest level of complexity in this second model, keeps the time factor, but instead of just reproducing past actions, it introduces a simulation engine that analyses the scripted data of each module, and particularly those provided by the BMM, in order to
represent possible present and future of specimens’ instances.

There are two ways to simulate entities’ moves: by using a mathematic model, or a multiagent system [2]. In the first case, most of the actions in the simulation are established by determining periodic elements in the script. In the second case, each entity (specimen) becomes an agent in the simulation. The aim is to generate rules to define agent behavior and interactions from its data in the IS. For the moment, this can only be done by simulations’ experts. Automating this step is facilitated by the structure chosen for the Behavior Management Module.

Using a multiagent approach, it is possible to link specimens to the behavior module that returns adapted rules to feed agents in the simulation. Because species behavior can be linked to their taxonomic level, generic rules can be determined to ensure results at several granularity levels. Indeed the BMM module greatly helps BIS that have to generate simulation metaverses.

C. General considerations on the proposed solution

BIS entities can be represented in metaverses as four representations corresponding to two models (Figure 21). The first representation (view) is generic, compatible with almost all IS structures, whereas other representations require a lot more metadata and specific information related to entities. However, each view has its own relevance. The first view provides an easy way to immersively evaluate the IS content according to specific grouping factors, as “user” or “project”. The second view is adapted to the immersion in a frozen scene to analyze its configuration. The third view (second level of the second model) is an animated reality reproduction that tries to represent moves and actions: it is adapted to scene reconstitution. The fourth view is based on a simulation engine, and represents its output as a metaverse in order to facilitate the simulation’s understanding by decision-makers.

Figure 21: Immersive representations of information for the main entities, at each complexity level

If we analyze the request that allows the generation of a metaverse, we can determine that it can be divided into two steps. The first one is always to select the entities that have to be represented in the virtual world. The second step, in the first model, is to build the tree structure, whereas, in the second model, there is no structure to build but a background map to retrieve, on which entities will be placed by exploiting their geographical metadata. That fact means that, from an XWD file, it is not possible to change from a model to another without making new requests to the IS. The XWD needs to be regenerated to swap views, instant rearrangements of virtual worlds are not possible without preserving all metadata (even those not represented) associated to entities.

Thus, depending their needs and the data availability, users can choose the representation that would help them most. However, it is possible to assist them in the choice of the representation that would give them more information by using the evaluation of the metaverse’s quantity of representable information. Indeed, from the initial request, the VWB is able to determine metadata on the different models of virtual worlds that could be generated: number of entities corresponding to the request, availability of metadata (georeferencing data necessary for mirror worlds, quantity of temporal information associated to entities for scripted worlds), or determination of the application’s resources limitations (i.e. generic shapes for the entities that have to be represented). Using these clues, the system can suggest a less complex view if the one asked by user risks to appear empty.

Future evolutions of the application will probably greatly improve the immersion experience in BIS. Prospection fields are numerous and rely on the same technical development that edutainment software [4] have recently received, adapted to the thematic:

• Share the immersion experience on virtual online worlds [7] that could be visited by several thematicians at the same time.
• Add a pattern recognition module to the VWB in order to generate, at the same time of the metaverse, a list of interest points that should be considered by researchers, or by general public.
• Profit from the growing interoperability between BIS to provide a better integration of distributed data with data warehouses (like Google 3D warehouse [5]).

The generation of virtual worlds from IS is indeed a research field that will require numerous studies and propositions, since it is a recent field of investigation.

VI. OVERALL DISCUSSION

The MABIS architecture has been instantiated in the Etic program, and more recently in its last evolution, in the Nextic project: two French environmental initiatives to build a BIS dedicated to the management of tropical information in Mascarenhas Archipelago. The screenshots we presented are extracted from these BIS which permitted us to gather feedbacks in order to improve our model up to the one we have presented here. If the global layered structure is now stable, it is however difficult to obtain feedbacks from anything else than a driven discussion with end-users. Indeed, it is not an easy task to formalize a study to globally analyze an IS that is distributed through several applications without having to evaluate each of the modules’ GUI. That is why our multidisciplinary team is now associated with knowledge engineers that focus on the evaluation of the interactions between the interfaces of modules and the different types of users.

In the same time, we are also trying to improve our innovative modules, like the Biodi-Verse that tries to generate metaverses from BIS information. Now that a
reliable process has been defined and first experimentations started, we are expecting precise analysis for testing and optimizing both BMM and Biodi-Verse applications. One of the aims is to build a protocol to evaluate the relevance of metaverses, depending their structure and level of details, and their utility to the different types of users. Although we have no doubt that the complexity of virtual world generation from BIS is a brake for the non-specialists, in the frame of the Nextra project, we first prefer to focus on experts and amateurs which represent our main users at the present time.

The MABIS model offers a convenient evolution of its own structure through its modular layered architecture. In the frame of the whole BIS, major enhancement like the adding of a new entity, is supported through the development of a new module: a WCS or a WSS, depending the complexity of the new entity and its treatments. In the frame of a single module, in order to provide the three modes (main, deported, and remote) of use, functionalities are fully decoupled from the GUI. Thus enhancement through the adding of functionalities is facilitated.

However, it is important to note that each module does not require all functioning mode. In the frame of the Nextra project, we decided to separate the authentication of users and the provisioning of their data, two features initially gathered by the directory, in two modules. The idea is to keep the provisioning in the directory and extract the authentication part in a new dedicated module supporting several protocols like Lightweight Directory Access Protocol (LDAP), OpenID and Code Access Security (CAS). This evolution is to let other institutions use our BIS through their own authentication server. In this case, the MABIS authentication module does not require a main mode as this functionality will essentially be used through its webservices, in remote mode.

This generalization of webservices in MABIS also allows to expect major enhancements in terms of inter-BIS data sharing. Two cases can be considered:

1. Exchanges between two BIS build on the MABIS structure. As the systems rely on the same architecture, it is convenient to exchange and merge the information from the two platforms. Thus the search engine of all modules should soon integrate a simple checkbox allowing to extend the search of any entities to all referenced MABIS platforms.

2. Exchanges between an information system built on MABIS architecture, and another BIS. In this case, the communication between the two platforms directly depends on the existing webservices and the data models used to describe the entities. As several structures coexist (for instance Dublin Core and METS for documents, or SDD and DELTA for taxa), and because BIS concept is relatively young, it is difficult to choose and exchange among the different possibilities.

Note that the current MABIS architecture does not impose a particular data model for entities as the structure defined is more global. However, even with two MABIS systems based on different data models describing its entities, the existence of homologous webservices ensures the facilitation of exchanges.

In terms of general evolution of the MABIS structure now, we plan to prospect on two points. The first one is to try to implement parts of WSS, like the VWB of Biodi-Verse, in SaaS. Indeed new evolutions of programming languages, like Action Script 3, allow unexplored possibilities to provide on the Web traditionally installed complex software. The second point is to build an abstract presentation layer to virtually gather all the functionalities offered by the WCS and the WSS gates in a consistent, easy to interrogate, interface. These developments are part of a strategy to strengthen the exchanges between MABIS architecture and international initiatives like the Global Biodiversity Information Facility (GBIF).

VII. CONCLUSION AND PERSPECTIVES

We have presented a new model of modular Biodiversity Information System: the MABIS architecture. These modules are components and software offering services through three different modes (main, deported and remote). MABIS provides a better flexibility on a relevant layered structure, in order to gather different types of users usually spread on several systems. Using this model, we have focused on three specialized modules:

- An evaluation component, the SIE, to associate certificates to scientific information stored in the BIS. This tool allows to follow the enhancement of data shared by specialists and amateurs through an authoritative and a community certification process. It allows users to debates on cards information with the aim to improve their content.

- A behavior management component (the BMM), to manage ethological data on specimens and taxa. The main considered entities are action, as a behavior unit, and sampling session, that represents an ordered collection of actions associated to one or more specimens. Three descriptive approaches are supported (textual, visual, and formal) to respond to the needs of three different experts (ethologists, computer graphics experts, simulation experts).

- A Biodi-Verse software, to generate immersive representations of selected MABIS entities. Because information systems dedicated to biodiversity gather sets of data difficult to appropriate, we present several possibilities to represent them through virtual worlds. Using a generic process, it is possible to generate up to four different representations of an IS part, depending on the available resources and the data structure.

The instantiations of MABIS architecture through the Etic program and Nextra project have already given positive feedbacks that also suggest enhancement tracks to the model in terms of data exchanges between BIS. Our future researches will be directed toward these fields.
ACKNOWLEDGMENT
The ETIC program promotes the enhancement of insular tropical environment contents from biological researchers using Information and Communication Technologies. The European Union, the French National Government, and the Regional Council of Reunion funded it in the frame of the DOCUP-FEDER 2002-2006, A9-04 ICT measure.

NEXTIC is the next ETIC project for distributing Biodiversity edited contents on the Web with webservices, then with Semantic and Immersive services. It is supported by the PO-FEDER 2007-2013 measure.

We are also grateful to Mr Laurent Cochet, engineer on Nextic project, for the implementation information provided to our research team.

REFERENCES