Evaluation of Information in Biodiversity Information Systems: a Model to Capitalize on Experts and Amateurs

Didier Sébastien, Olivier Sébastien, Noël Conruyt, Mamy-Haja Rakotobe

Evaluation of Information in Biodiversity Information Systems: A model to capitalize on experts and amateurs

Didier Sebastien, Olivier Sebastien, Noël Conruyt and Mamy-Haja Rakotobe
IREMIA, LIM-EA2525, Université de La Réunion, Saint-Denis, France
didier.sebastien/olivier.sebastien/noel.conruyt/mamy-haja.rakotobe@univ-reunion.fr

Abstract. For observation of coastal and ocean ecosystems, the acquisition and processing of scientific data in databases have led to the proliferation of specialized Information Systems, as it is the case for monitoring biodiversity. Given the various data that they contain, these Biodiversity Information Systems (BIS) provide complex applications focused on the needs of domain experts, but are often closed to amateur's field contributions. However, such a new open and interoperable Biodiversity Information Service could attract more and more hobbyists whose skills would be useful to integrate, in order to share, improve and keep information up to date with people. A part of the solution could be to render service to amateurs by tagging the quality of their contributions, in order to clearly indicate the origin and quality of provided information.

This paper presents a Scientific Information Evaluation (SIE) model instantiated as a Web Service (WS), which is a contribution to the enhancement of BIS for Citizen Science. Crowdsourcing and evaluation by peers in communities of practice enhance this BIS because the WS is managed by adding qualitative metadata (certificates) on the regular primary data found in cards, as a global evaluation and a conversation thread. This WS module has been tested in the frame of the NEXTIC project at University of Reunion Island for data acquisition of the Herbarium of Reunion Island. It reveals the central role of such a module for the community to gain confidence about the BIS. It is to be ported to the marine domain.

Keywords: *biodiversity information system, data evaluation, modeling, crowd-sourcing*

1 INTRODUCTION

The important loss of world biodiversity has led Eco-Informatics experts [F. Recknagel, 2009] 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 systematicians) who work on species monitoring in their everyday tasks: produce inventories of specimens, model, describe, classify and identify them, obtain a better visibility on biodiversity studies, plan coordinated actions between institutions and communicate their results. They focus on the management of taxa and specimens [S. J. Mayo *et al.*, 2008]. In this context, Web Services (WS) are more and more used to manage virtual biodiversity [K. D. Fook, A. M. V. Monteiro, G. Câmara, 2007], leading to the modeling of Biodiversity Information Services for interoperability of these data [N. Conruyt *et al*., 2010]. Such Web services can be found in portals such as EOL (www.eol.org/), GBIF (www.gbif.org), ALA (www.ala.org.au). Nevertheless, the quality and completeness of these data cannot be guaranteed in BIS, because this content needs continuous updates and validation from experts. For some years, amateurs were not allowed to contribute to the system by adding their own data in order to share them with the scientific community. Today, due to biodiversity crisis and lack of specialists, the biodiversity field attracts a parallel community of non-specialists that gathers important sets of data. These data were generally rejected only because experts did not provide them. Nevertheless, crowdsourcing has arisen with the advent of Citizen Science [J.C. Tweddle *et al.*, 2012] and Web 2.0. Indeed, there is a social demand that cannot be eluded for engaging citizens in the knowledge construction about the richness of the environment and their territories [A. Cosquer, R. Raymond, A.-C. Prevot-Julliard, 2012]. Conservation in the context of biodiversity loss needs the acquisition of qualitative data from every skilled and motivated amateur. To gain from these public and opened data, it would be useful to let amateurs have a controlled access on the BIS for them to participate to this endeavor. A solution to acquire such qualitative information is to support a scientific data validation policy, which is often made by an administrator in BIS. However, because the quantity of data expert-administrators have to validate is colossal, they cannot afford to treat amateurs' resources. In this new context, we propose a Scientific Information Evaluation (SIE) model.

To introduce our contribution, we will first detail our model through its prerequisites and general process, then we will present an example of implementation realized in the frame of the Nextic research project.

2 GENERAL MODEL

2.1 Context and prerequisites

Our model relies on modular information systems' architectures and more specifically Biodiversity IS, like MABIS developed in a precedent work [D. Sebastien *et al.*, 2009]. MABIS stands for Modular Architecture for Biodiversity Information System. This stratified model relies on several Web applications that are components interconnected

together by webservices. **Fig. 1** presents the four layers used by the different users. The first layer manages the main entities of the IS through simple online application defined as modules. The second one introduces meta-modules for data management. The third layer is composed by softwares, i.e. installed applications, dedicated to information analyze and knowledge treatment. The fourth, and last layer, allows the diffusion and the valorization of the information contained in the BIS. In this frame, the evaluation system of information can be integrated through a module of the second layer interconnected with the different modules of the first layer, for the management of main entities.

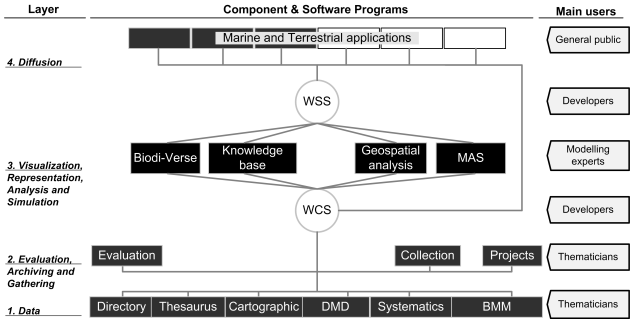

Fig. 1. MABIS general architecture

In order to ease the use of each module is available under three functioning modes (**Fig. 2**). Main mode allows users to 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. Thus, the Scientific Information Evaluation (SIE) module should be able to provide its services through different functioning modes.

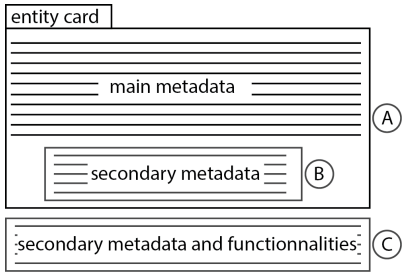

Fig. 2. The three declinations of modules

Now that the context of the SIE and its prerequisites have been defined, we introduce its general model in the following part.

2.2 Model

Our aim is not to provide a methodology to analyze data, for instance by error elimination [K. R. Popper, 1999], but to set up a workflow that induces the community to evaluate, comment and follow the enhancement of a card until it reaches the best acknowledgement level. The 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 who can be the general public or decision makers. The enhancement process is presented in **Fig. 3**. When a card describing an entity is created every user have the possibility to evaluate it. In order to focus the attention of the community on a new card or on a deeply modified entry, the authors have the possibility to emit a request for an evaluation. The evaluation consists in two aspects: a level of certification conveying the achievement's degree of the card for the evaluator, and a thread of comments allowing a discussion/debate among the community of users. Theses elements further the enhancement of the card that can be evaluated and commented again, through a virtuous circle, until it reaches the best certificate.

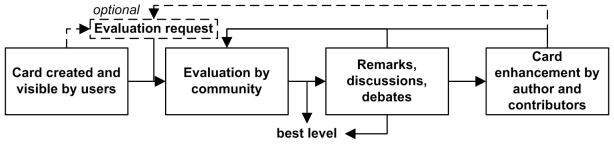

Fig. 3. Evaluation of scientific information

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 trustworthiness that gives a more precise idea of the data acquisition condition. The certificates can be considered as a succession of steps toward the validation level. It is represented as a simple and easy to read icon that sums up the state of the card. **Fig. 4** presents an example of representation that can be used for the different certificates, from the "not evaluated" state to the "validated at present time" state.

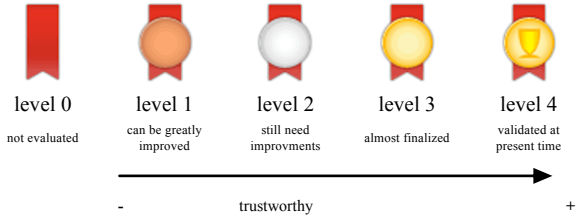

Fig. 4. Levels used for evaluation

In order to reduce the work of data evaluation by experts, two participative evaluation systems are offered. 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. This evaluation is presented as an icon showing the lowest certificate delivered by the expert of the highest level about the card. The algorithm is chosen to be the less permissive, in order to ensure that the quality of selected data is, at least, equal of the displayed certificate.

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 secured and specialized than the precedent one; however, a certificate coming under an important participation of users gives a good basis of evaluation if experts have no time to consider the card. An icon that shows the average level of certification given by voters also sums up this notation based on folksonomy [T. Vander Wal., 2007]. Thus, the algorithm tend to offset the inexperience of the amateurs by considering that the multiplicity of evaluations will reflect a result close to the effective state of the card.

Authoritative and community certifications are displayed side by side as this last one can reflect an appreciation about the last changes on the card. They 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.

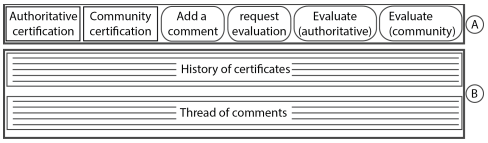

Fig. 5. Banners for deported information of the SIE

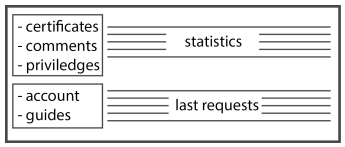

Fig. 6. SIE main mode model

In order to smooth the connections between the different elements in the workflow, information (certificates and comments thread) can be inserted under several forms through the functioning modes of the SIE module. Thus, it is relevant to provide directly in a module dedicated to a type of entity the main information and functionalities, like the certificates, the possibility to add a comment, an alert to focus the attention of the community and request an evaluation, a way to post an evaluation, the history of certificates and the thread of comments **Fig. 5** sums up this list as a set of banners.

The SIE's main mode (**Fig. 6**) is gathers all detailed information about the evaluation of entities. It allows to:

- present general metadata and statistics 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 remote mode of the SIE is of course not related to a specific representation, but can

often be used in other modules, for instance to sort entities by certification level in other Web applications' main mode. This general model could be extended, under specific conditions, from biodiversity IS to other types of IS.

2.3 Managing users interactions

The goal of this section is to deal with the way to involve non-specialists to use a system that rely on scientific content. Indeed, experts and amateurs do not share the same background in term of information classification, vocabulary or even rigor. That is why two levels of use are identified: the casual one, targeted to general users and the expert one. Let us focus first on their similarities and then on their specific aspects relying on the following table describing the main points that separate those groups.

	Experts	**Amateurs**
Group size	Small	Big
Ability to self-diagnosis	High	Generally low
Technical support provided	Very high	Basic
Tools needed	Many	Few
Content submissions	Average	Very high
BIS environment mastering	high	low

Common points of users interaction levels are rather related to the human-computer interface than the BIS-related activities. Indeed it consists in providing support to create and edit content online on one hand and to enable file upload for further sharing. Those activities rely on mechanisms that are nowadays trivial because they are already in use on popular Web 2.0 services such as Youtube (video sharing), Flickr (picture sharing) or Wikipedia. Basically, interface usability can be assessed thanks to criteria, like Nielsen's ones [Nielsen J., 1994]. As far as mobile devices are concerned, dedicated recommendations are proposed by main Operating System builders as Google for the Android environment (https://developer.android.com/design/index.html) to achieve the same goal. Thus, BIS general operations are not so different to the ones accomplished in other application, provided users have the minimal required level of understanding the device.

Specific points of the interface are related to the activities that are not shared by experts and amateurs. Actually, the latter should not be aware of the features made available to the former, as they may not be accessible enough for them. That is why screen layout is to be set slightly differently according to user profile. The main point that guide interaction design is the objective of the service. Tasks listed on the interface show a restricted set of tools for non-specialists. This can be viewed as a "write-only" function, to upload a new case. The full set of features appears only to researchers. Practically, scientists have access to the ontology modules and can edit links between content. Of course, they are to solve the cases brought by the general users. These interactions can be seen as "read-write".

3. IMPLEMENTATION

In order to experiment our model of Scientific Information Evaluation, we implemented it in the frame of a research project: NEXTIC. This project is dedicated to the valorization of natural collections. As NEXTIC relies on MABIS architecture, the SIE model has been instantiated in the form of a module gathering two modes: main and deported. We worked with a designer and optimized the ergonomics of the graphic user interface (GUI) in order to turn several texts into icons representing information and functionalities. **Fig. 7** presents the appearance of the deported mode dedicated to the SIE inserted in other modules GUI.

Fig. 7. Elements for the evaluation injected in the cards

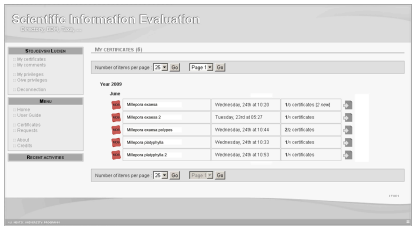

Fig. 8. Main mode of the SIE implementation

The main mode of the module is a web application centralizing the workflow for all entities of all types (**Fig. 8**). It also helps the administrators and the community by presenting statistics and pinpointing major differences between authoritative and community evaluations. The SIE has been connected to one module of the BIS, dedicated to the management of multimedia files. Preliminary results confirm that allowing amateurs on the BIS should enhance the growth of the users community.

4. CONCLUSION AND PERSPECTIVES

In this paper, we introduced a model of scientific information evaluation in Biodiversity Information Systems. This model is based on two kinds of certifications and a workflow to frame the exchanges of users. The main idea is to ease the communication between the community members in order to collect a maximum of contributions and advices about the card describing entities. By opening the system to amateurs, and supporting two kinds of certifications to be able to sort the information, reactivity and returns can be improved. The gradual certifications constitute a more precise evaluation than a Boolean one. Furthermore, it is more easy to automate the administration of the system, by letting the user share their privileges up to their own ones.

We implemented our model as an element of a modular architecture of biodiversity information system, declined in two Web applications. The deported mode provides the essential information for each card whereas the main mode gathers all the certifications in order to pinpoint statistics and evaluation's requests. In future work, we will analyze the usage made by contributors by focusing on their experience, exchanges between them, activity on the module and a comparison between authoritative and community information for the same cards.

ACKNOWLEDGEMENT

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 Data and Immersive services. It is supported by the PO-FEDER 2007-2013 measure.

We are also grateful to Mr. Laurent Cochet and Mr. Lucien Stojcevski, engineers on NEXTIC project, for the implementation information provided to our research team.

REFERENCES

A. Cosquer, R. Raymond, A.-C. Prevot-Julliard. (2012). Observations of everyday biodiversity: a new perspective for conservation?, *Ecology and Society* 17(4): 2. http://dx.doi.org/10.5751/ES-04955-170402.

N. Conruyt *et al.* (2010). Moving from Biodiversity Information Systems to Biodiversity Information Services, *Information and Communication Technologies for Biodiversity and Agriculture*, Ed. by L. Maurer and K. Tochtermann, ISBN: 978-3-8322-8459-6, Shaker Verlag, Aachen.

K. D. Fook, A. M. V. Monteiro, G. Câmara. (2007). Web Service for Cooperation in Biodiversity Modeling, *Advances in Geoinformatics*, Clodoveu A. Davis Jr., Antônio Miguel V. Monteiro (Eds.), Springer Berlin Heidelberg, pp. 203-216.

S. J. Mayo *et al.* (2008). Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis, *Kew Bulletin* vol. 63: 1–16, Springer Netherlands, pp. 1-16, doi: 10.1007/s12225-008-9014-1.

Nielsen J. (1994). Usability inspection methods, *Conference companion on Human factors in computing systems*, ACM, p. 413-414.

K. R. Popper. (1999). All life is problem solving, Routledge Eds.

F. Recknagel. (2009). Ecological informatics: Current scope and future directions, Proc. of *Information Technologies in Environmental Engineering*, Springer Berlin Heidelberg, pp. 3-22, doi:10.1007/978-3-540-88351-7.

D. Sebastien *et al.* (2009). Biodiversity Information Systems evolution: The MABIS model to gather several communities on an adaptable environment, *International Journal On Advances in Systems and Measurements*, Vol. 2 no. 4, issn: 1942-261x.

J.C. Tweddle *et al.* (2012). Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK. *Natural History Museum and NERC Centre for Ecology & Hydrology for UK-EOF*, Available online: *www.ukeof.org.uk*

T. Vander Wal. (2007). Folksonomy, *online posting*, Vol. 7.