

Managing Insular Tropical Environment through Data and Knowledge Bases by using Web Services: a case study on Corals and Herbarium of La Réunion Island

N. Conruyt, D. Sebastien, D. Payet, Y. Geynet, D. Caron, and D. Grosser

*Laboratory of Computer Science and Mathematics, IREMA, University of Reunion Island,
Saint-Denis, 97715, France
(conruyt/didier.sebastien/denis.payet/geynet/caron/grosser@univ-reunion.fr)*

Abstract: Decision-makers who want to manage Insular Tropical Environment more efficiently need to narrow the gap between the production of scientific knowledge in universities or other labs and its pragmatic use by the general public. One key environmental problem is to preserve biodiversity of ecosystems that are under human pressure. As we only protect what we know, a solution is to share expert knowledge about habitats and species on the Web for educating the public about their richness and beauty. Data and knowledge bases are part of a biodiversity information system that we have built to deal with this research enhancement problem, through the ETIC program. They were conceived using Web Services in order that each module communicates its functionalities and information one another and with external systems.

Keywords: Biodiversity Management; Knowledge Bases; Web Services; Information System; ETIC.

1. INTRODUCTION

The biological diversity of the islands in the South West Indian Ocean (Madagascar, Comoros, Mauritius, La Réunion, etc...) is still rich despite the important anthropic pressures, which are increasing from year to year. La Réunion Island is considered as one of the 20 richest world's biodiversity hotspot [Myers *et al.*, 2000]. But its population (also very diverse) is expected to reach the one million mark in just thirty years from the present 700,000. This will be a real problem for the management of the territory, as well as for the protection of fragile species. To face this environmental sustainable development problem, two natural parks have been created (one for the sea¹ and one for the mountains and cirques²). Their missions are to promote the natural and cultural heritage by increasing and sharing knowledge about its biodiversity, and to protect it against destruction. On the other side, the University of La Réunion has accumulated plenty of observations, data, information and knowledge on ecosystems over the past forty years. This information may be found in laboratory checklists, collections, museums, literatures, charts, maps, images, movies and sound files, individual databases, yet is hardly exploited by anyone except the authors themselves. There exists a gap between the missions of these institutions, because the first have to *manage biodiversity objects* whereas the second have to *produce data and knowledge about them* from their research activities. In order to propose new, sustainable practices of biodiversity management, research must not only lead to new knowledge and understanding of the interactions between the functioning of ecosystems and human activities, but also lead to pedagogical transmission of this knowledge through research en-

¹ [http://www.reunion.ecologie.gouv.fr/rubrique.php3?id_rubrique=2, visited on 11/04/08]

² [<http://www.parc-national-reunion.prd.fr/>, visited on 11/04/08]

hancement. Governments and European Commission that are aware of sustainable development make environmental education a priority³, and reef ecosystems and tropical forests may be preserved against damage by creating awareness, sensitivity, and skills while fostering participatory action [Stepath, 2002].

2. ENHANCING BIODIVERSITY RESEARCH WITH ICT

As part of the natural environment, biodiversity has been defined as “the variety of life in all its forms, levels and interactions. It includes ecosystem, species and genetic diversity” [Hunter, 1996]. For our research, we focus on populations of specimens between the taxa⁴ and organ levels of biodiversity research (Figure 1), i.e. *Systematics*. It is the scientific discipline that deals with listing, describing, naming, classifying and identifying living organisms. The originality of our insular tropical biodiversity management method is that we concentrate on natural *objects* that are specimens in the fields (living specimens) and in museums (collection specimens). Experts in biology at the University have studied them intimately for years and are the only persons able to recognize their names that give access to more information. They build their personal or tacit knowledge [Polanyi, 1962] by observing them in the fields and in their laboratories under the microscope, then interpreting them with descriptions. These described objects form the development basis of their formal or explicit knowledge in monographs that makes authority in their domain of speciality. But this expertise that is based on interpretation of objects through a description process is becoming extremely rare because experts are retiring without being replaced. For future biodiversity

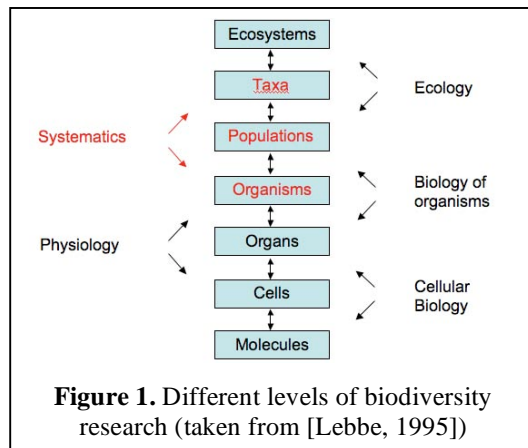


Figure 1. Different levels of biodiversity research (taken from [Lebbe, 1995])

studies relying on species identification, decision-makers, environmental technicians and the general public will only be left with monographic descriptions and collections in museums. Nevertheless, with all the possibilities that the Internet has to offer, research is no longer confined to books dedicated to specialists. New innovative research can easily reach a wide audience, even in developing countries - at the click of a mouse, by offering online biodiversity information. In this context of knowledge sharing, the role of Information and Communication Technologies (ICT), and in particular Artificial Intelligence Techniques, is to give access to different kinds of online information that are structured through *data and knowledge bases* and can exchange information using *Web Services*. Our research work contributes to the vast effort of environmental conservation by introducing *knowledge management with ICT* at the beginning of the process.

3. THE ETIC PROGRAM

ETIC⁵ is a publicly funded project, based on La Réunion Island in the South-West of Indian Ocean, whose goal is to develop innovative ideas and ICT solutions for the management of biodiversity research contents. The program was created in 2004 at University of La Réunion Island for research and knowledge enhancement of Insular Tropical Environments, by using AI techniques such as *Knowledge Engineering* for building expert systems and *Collective Intelligence* for building multi agent systems, and *Information and Communication* tools such as content management systems [Conruyt et al., 2006a]. Indeed, the first step to protect our insular tropical environment is to better educate citizens about its richness because we can only protect what we know! ETIC is based on several thematic projects and

³ [http://ec.europa.eu/research/leaflets/biodiversity/index_en.html, visited on 11/04/08]

⁴ Taxa are the names of ranks in the scientific classification: Species, Genus, Family, Order, Class ...

⁵ [<http://etic.univ-reunion.fr>, visited on 11/04/08]

a collaborative methodology, stressing partnerships between researchers, educators, decision-makers, enterprises, associations and end-users who wish to share and communicate their environmental data and knowledge off and on line. With the help of computer scientists, web designers, programmers and graphics experts, the common goal is to participate in the construction of an *Information System* (IS) for environmental management on the Internet. Content include terrestrial (i.e. Herbarium) and marine (i.e. Corals) biodiversity descriptions about specimens, their geography, ecology, photography, taxonomy and bibliography contextual information in La Réunion Island in the South-West Indian Ocean.

4. THE BIODIVERSITY MANAGEMENT MODULES

The core of the ETIC IS platform is the integration of knowledge bases and databases about biodiversity knowledge and data by using Web Services.

4.1 Knowledge bases

The former applications are instances of a Knowledge Based Management Tool called IKBS. This Iterative Knowledge Base System [Conruyt and Grosser, 2003] lets specialists define an Object-Attribute-Value *descriptive model* of the domain knowledge (input) and describe *cases* (output) based on this ontology (Figure 2). This knowledge acquisition phase can be repetitive because IKBS applies the scientific method of Popper (1973) in biology (conjecture and test) with an iterative process of knowledge management: 1) Observe and familiarize oneself, 2) Represent observations, i.e. make a descriptive model and related descriptions (cases), 3) Build hypotheses from pre-classified descriptions, i.e. generate identification keys (supervised classification), 4) Test and use them with new observations, i.e. identify new specimens, 5) Refine the initial knowledge (new characters, cases and classifications) (Figure 3).

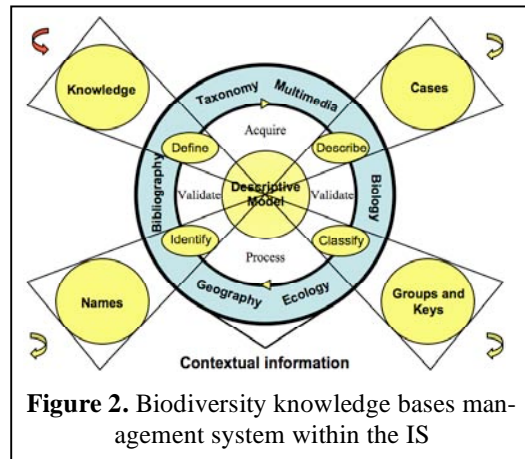


Figure 2. Biodiversity knowledge bases management system within the IS

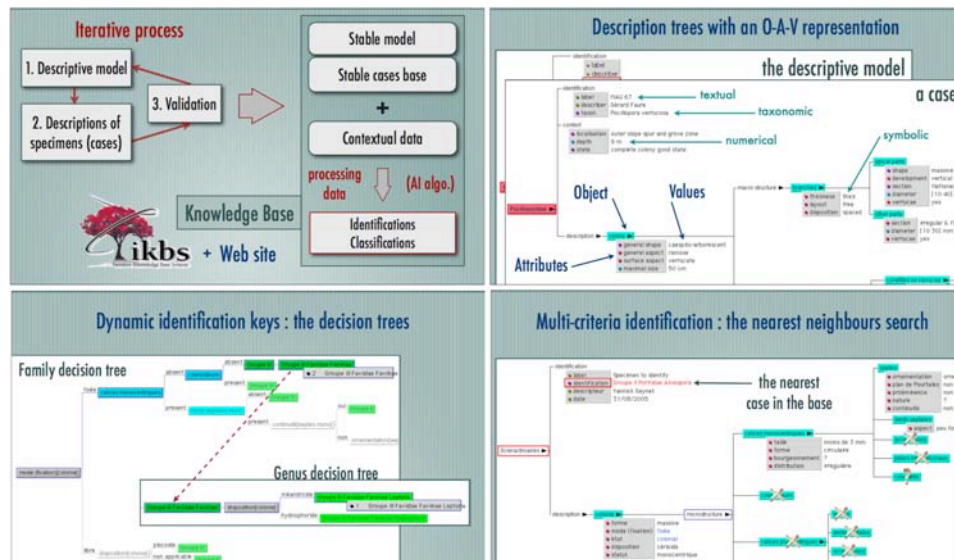
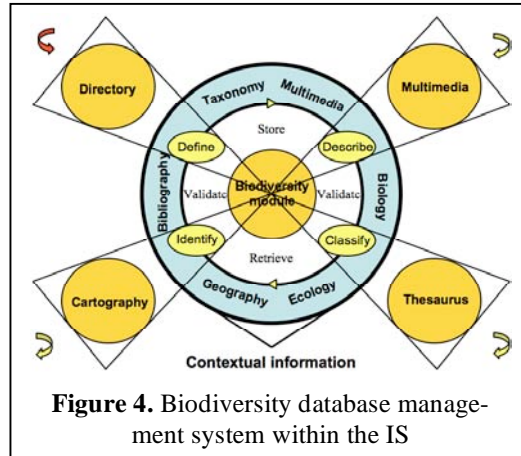


Figure 3. IKBS methodology of Systematics knowledge management

After an automatic classification process based on tree induction of pre-classified cases, end users are able to identify new descriptions with a questionnaire. End-users proceed by photo-interpretation of specimens to obtain a genus name, or by observing microscopic specimen elements under the binocular to identify a species name [Conruyt et al, 2006a].

4.2 Databases

The latter is a *biodiversity module*, i.e. a database of *objects* that stores, organizes and presents scientific data about field observations, collected specimens (samples) and taxa descriptions. Other database modules (i.e. directory, multimedia, thesaurus and cartography) complement this central module: 1) the directory gives access to the *subjects*, i.e. the individual and community researchers with their profiles in a card index, 2) the multimedia database manages all types of *documents* (photo, video, sound, etc.) that can be indexed to specimen objects, 3) the thesaurus will be an illustrated glossary that stores the meaning of Insular Tropical Environment *vocabulary*, 4) the cartography is a tool for *georeferencing* data on a map (made with the GoogleMaps API). All of these modules are linked by Web Services



so as to constitute a modular, interoperable and integrated biodiversity specimen and species database management system (Figure 4). The data entry process in the biodiversity module is organized around the memorization of specimens' information, which is collected in notebooks by biologists when they are inventorying biodiversity. It has been structured in five edition tasks (actions) that fit with the daily work of monitoring specimens in the fields: 1) *Origin* of specimen, where was it found? 2) Short *description*, what was observed? 3) Taxon *identification*, what is it? 4) *Status* of specimen, i.e. sex, nature, state, fertility, development stage, 5) *Label* of specimen if it is to be put in collection. In addition to this *internal specimen information* in the biodiversity module, the surrounding modules manage *external contextual data* such as the location of the specimen (space identification with the Geolocation system), the identity of the subject (who is the observer in the Directory), the image or video of the specimen with associated metadata in the Multimedia Database (Figure 5).

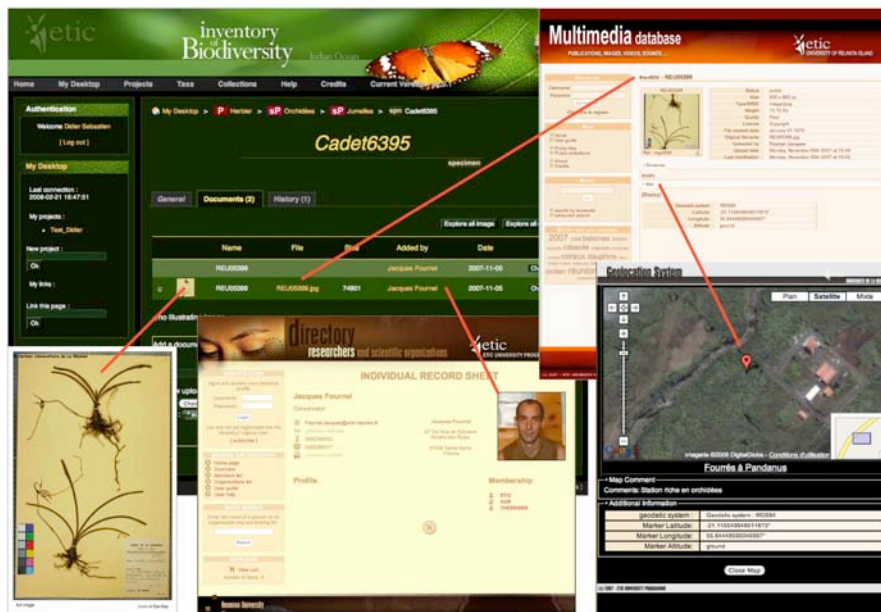


Figure 5. ETIC process of biodiversity data management

5. WEB SERVICES AS SUPPORT FOR INTEGRATION AND OPENESS

In order to make our IS modular and interoperable, we have chosen to design it according to Services Oriented Architecture (SOA). As it is a web platform, we chose the Web Services approach [Curbera *et al.*, 2001], which provides a number of features and benefits.

Among the technologies to implement Web Services, we chose the WS-* technology whose specifications are based on SOAP and WSDL standards:

- SOAP (Simple Object Access Protocol) for exchanging messages is a RPC (Remote Procedure Call) protocol built on object-oriented XML.
- WSDL (Web Service Description Language) for describing: Web Services, their operations, messages used, the types of used data, and the used protocols. The WSDL describes a public interface for access to a Web Service. This description is written in XML and indicates "how to use the service."

5.1 Advantages of using Web Services in ETIC IS

Using Web Services structured the way we developed ETIC IS:

- Firstly, we have chosen this technology because it allows other IS to use our modules independently [Pires *et al.*, 2002]. This way, information stored in ETIC IS is fully open and exploitable by other institutions [Yang *et al.*, 2002].
- Secondly, it modified the way we imagine the connections in the IS itself. It let us develop each module in a heterogeneous way, using the appropriate technology (PHP/JAVA/Flash Action Script). For example, the ETIC Directory is developed in PHP whereas the first version of our Multimedia Database [Sebastien and Conruyt, 2008] was relying on the JAVA technology. Although we used different programming languages, these two modules were able to exchange secured information, thanks to Web Services.
- Thirdly, from a technical point of view, Web Services impose a rigour that permits to update our systems and services without losing the compatibility with older versions. Each new version of a service is indexed with a number, and the WSDL provides the way to implement the new client, but it does not impact on the older versions that remain fully functional.
- Fourthly, from a management point of view, using Web Services was a good choice. Indeed, the ETIC development team often changed. Because of this important turn-over, it was difficult to transmit the development key points from an engineer to his successor. Thanks to the Web Services (and particularly the WSDL declaration), it was not necessary for the new team to fully understand a service source code to use it as a client. It saved a lot of development time.

Thus, using Web Services was of course important to open our IS, but it also helped us to improve its inner structure.

5.2 What are the ETIC services?

Because each module is dedicated to accomplish a precise task, each of them provides some specific functionalities and dedicated Web Services. This way, any other IS can reach ours and decide to use one or the whole of ETIC modules for its own purpose. But contrary to data consultation that is completely open, these clients need to be referenced in ETIC directory if they want to add information in ETIC IS. This security allows us to ensure the traceability of information injected in our system.

The implemented connections between ETIC modules are shown in Table 1. On the five modules deployed now (Directory, Multimedia, Cartography, Biodiversity, IKBS) all use Web Services as client, but only three provide services (Table 2).

The work is still in progress. For example, the development of the Thesaurus (one of the database's modules, see Figure 4) is not presented here. Furthermore, the current version of the Cartographic Browser does not implement a connection to the ETIC directory. In fact, it cannot be used as a standalone module yet. Other modules asking for its services ensure

themselves the client identification process. Moreover, several functionalities supported by the modules don't have a Web Service allowing a distant interrogation of it yet (e.g. to use remotely the multi criteria search of the MDB, see Table 2 below). Only main services are already released. Of course, this will change with the development of our modular IS.

Table 1. Connections between ETIC modules

Web Services		Client				
		Directory	Multimedia	Cartography	Biodiversity	IKBS
Provider	Directory		✓	○	✓	○
	Multimedia	○		✗	✓	✓
	Cartography	○	✓		✓	○
	Biodiversity	✗	✗	✗		✗
	IKBS	✗	✗	✗	✗	

Legend: ✓ : already implemented, ○ : in progress, ✗ : not relevant now

Table 2. Services already functional in ETIC IS

Module	Tasks provided by Web Service
Directory	<ul style="list-style-type: none"> ▪ Find a user by his name ▪ Get the list of members' id ▪ Get a minimum of information on each member of the directory ▪ Get a minimum of information on a specified member ▪ Get all information about a specified member ▪ Authenticate a member with its login/password (secured service)
Multimedia Database (MDB)	<ul style="list-style-type: none"> ▪ Return two links (documents' thumbnail and mini-thumbnail) corresponding to the input array of document id ▪ Return all information (metadata, download link) corresponding to the input array of documents' id ▪ Return two links (to the information card on the MDB and to the file) corresponding to the input array of documents' id ▪ Allow the distant upload of a document on the MDB ▪ Find a document on the MDB (Quick Search)
Cartographic Browser	<ul style="list-style-type: none"> ▪ Verify if a map exists for a specified id ▪ Get all information about specified map ▪ Find a map from its description ▪ Delete maps from the given list

Independently of the fact that we have to develop new Web Services for each module, we also have to increase the integration of the existing services *between* modules. This integration must be the result of focus groups realised with biologists in order to provide them the right information when they need it. This tuning is a Web Design concern because, for end-users, Web Services must be transparent.

6. INTEGRATION OF MODULES

The usefulness of integrating both data and knowledge bases by using Web Services is to make them interoperable for better information (data + knowledge) management between research scientists before Web publication (authentication with e-research or on-line collaborative research), and then facilitate the sharing of their data and knowledge with more general public on the Internet (education with e-learning) [Conruyt *et al.*, 2002]. Our ETIC platform gathers tools for helping decision makers to manage the biodiversity of natural parks (see Figure 6 for the Marine park example). The process starts with Knowledge Bases that are used for the process of describing specimens, classifying them and identifying a name for unknown specimens. Databases give then access to different types of ecological information on known taxa. Indeed, the first question is to recognize the name of what has been observed to make a biodiversity diagnosis. Then, in order to answer questions about the evolution of ecosystems, the next step will be to monitor the dynamics of populations with temporal databases on coral reef (or forest) and simulate their behaviour with Multi-Agent Systems when facing damages such as pollution, pests, over-fishing, etc.

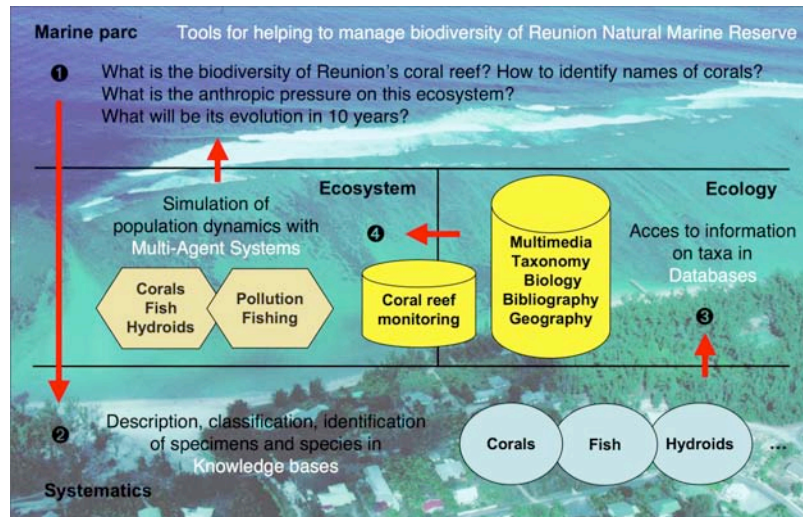


Figure 6. Coral reef biodiversity management process

Web Services have been designed currently to collect the information about actors of the ETIC community (the Directory⁶). Some persons are identified as project leaders in their scientific domains. As they want to publish their data and knowledge on biodiversity with their colleagues, they create their proper organization with authorized members in the directory. They can then share their documents in the multimedia database⁷, edit them together, and also localize them with a geo-referenced system using Google Maps. But their objective is to store information about taxa and specimens in the fields before putting them in collections. The Biodiversity module⁸ is thus used as the core Web Service to inventory species they are working on.

On a technological level, our approach of integration has been clearly made in the *delegation* meaning. When a functional need is identified during the module design, we find out if the missing functionality must be implemented by adding code in the module or as an interaction with another existing module or a new module. This way allows us to easily identify all modules to be deployed, and interactions between modules are explained as the system is built. It is an efficient process when such project is started from an empty module set.

But the next challenge is to apply the opposite approach: *data aggregation*. We have to ask ourselves about the potential benefits that can emerge from new composition and fusion of knowledge from different modules. This is precisely the topic of our future work, through which we hope to formalize a cross ontology that will serve as a basis for analytical tools (including Multi-Agent Systems simulations) in order to improve our understanding of data, and perhaps produce new knowledge.

7. CONCLUSION

In the domain of life sciences, engineers and designers who build information systems have to model expert knowledge at two skill levels. On the one hand, they store their explicit knowledge in databases, because it is made of contextual data (eco-biological, biogeographical, photographic, biographical, taxonomical, etc.) that is found in books. This information can be retrieved more easily in databases with a multi-criteria research. On the other hand, they acquire expert implicit or tacit knowledge in knowledge bases, because this know-how is more complex to elicit: this experimental and intuitive expertise is grounded in their daily work throughout a life practised in the fields. It is better to process

⁶ [<http://etic.univ-reunion.fr/annuaire/>, visited on 11/04/08]

⁷ [<http://etic.univ-reunion.fr/bdm/>, visited on 11/04/08]

⁸ [<http://etic.univ-reunion.fr/biodiv/>, visited on 11/04/08]

case descriptions with induction techniques than to tell the expert to explain what are his rules of thumb. All these interoperable modules are prototypes that are still under development. They have been tested using the Herbarium boards and monographs of Reunion University [Cadet, 1980] and the coral specimens collection and taxa descriptions of the Mascarene Islands [Faure, 1982]. In order to improve the usability of these tools and anticipate biologists' needs, we have used a co-design platform methodology with several focus groups [Conruyt et al., 2006b]. The next step is to deliver a fully operational IS to naturalists who are working in natural parks in Reunion Island and to be interoperable with the "Système d'Information sur la Nature et les Paysages" (www.naturefrance.fr) at a national level and the Global Biodiversity Information Facility Network (www.gbif.org) at an international level.

ACKNOWLEDGEMENTS

The ETIC program is financed by the French National Government, the Regional Council of Reunion and the European Union (<http://www.reunion-europe.org/docup/A9-04-CI.html>).

REFERENCES

- Cadet, T., La végétation de l'île de La Réunion. *Thèse es sciences*, Université de La Réunion. 312 p, 1980.
- Conruyt, N., Geynet, Y., Grosser, D., Santally, M., Senteni, A., Sonah, S., The IKBS-Darwin Research & Learning System for Sharing Biodiversity Knowledge of Tropical Islands. *Proc. of the International Conference on Computers in Education*, pp. 760-764, Auckland, New Zealand, 2002.
- Conruyt, N., Grosser, D., Knowledge engineering in environmental sciences with IKBS. *AI Communications*, 16(4), IOS Press, 267-278, 2003.
- Conruyt, N., Grosser, D., Geynet, Y., Enhancing Insular Tropical Environment contents through ICT and a Co-design methodology: the case of the ETIC program, *Proc. of the 5th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BE-SAI'06)*, 1-1-1-7, Riva del Garda, Italy, 2006a.
- Conruyt, N., Conruyt, P., Sebastien, O., A methodology for Designing E-services from a Co-Design Platform, *Proc. of the 6th International Conference on Knowledge Management (IKNOW'06)*, Graz, Austria, September 6-8, 2006, K. Tochtermann / H. Maurer (Eds.), pp. 226-230, ISSN 0948-695x, J.UCS, Graz, Austria, 2006b.
- Curbera, F., Mukhi, M., Weerawarana, S., On the Emergence of a Web Services Component Model, *Proceedings of the 6th International Workshop on Component Oriented Programming (WCOP'01)*, 2001.
- Faure, G., Recherche sur les peuplements de scléactiniaires des récifs coralliens des Mascareignes. *Thèse es sciences*, Université Aix-Marseille II, 1982.
- Lebbe, J., Systématique et informatique. Systématique et biodiversité, Bourgoin T. (Ed), *Biosystema*, Paris, 13:71-79, 1995.
- Hunter, M.L., Jr. Fundamentals of conservation biology. Blackwell Science, Cambridge, Massachusetts, 482pp., 1996.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B., Kent, J. Biodiversity Hotspots for Conservation Priorities, *Nature*, 403:853-858, 2000.
- Pires, P. F., Benevides, M. R. F., Mattoso, M., Building Reliable Web Services Compositions, *Proc. of the International Conference on Web Databases and Web Services*, LNCS, vol. 2593, pages 59-72, Springer-Verlag, 2002.
- Polanyi, M., Personal Knowledge: Towards a Post-Critical Philosophy; Chicago: The University of Chicago Press, 1962.
- Popper, K.R., La logique de la découverte scientifique. Payot (Eds.) Press, Paris, 1973.
- Stepath, C.M. Awareness and community-based monitoring, *Proc. of the 9th International Coral Reef Symposium*, Vol:2. Bali, 807-812, 2002.
- Sebastien, D., Conruyt, N., Online multimedia database for communities of practice in Biology: a real use challenge, *Proc. of the 3rd International Conference on Internet and Web Applications and Services*, 2008 (to appear).
- Yang, J., Papazoglou, M.P., Web Component: A Substrate for Web Service Reuse and Composition, *Proc. of the 14th International Conference on Advanced Information Systems Engineering (CAISE'02)*, Toronto, Canada, 2002.