

A CASE STUDY OF INPA'S BIO-DB AND AN APPROACH TO PROVIDE AN OPEN ANALYTICAL DATABASE ENVIRONMENT

J. L. Campos dos Santos^{1 2}, R. A. de By², and C. Magalhães¹

¹ The National Institute for Amazonian Research (INPA)
André Araújo Avenue, 1756, P.O. Box 478, 69.083-000, Manaus, AM, Brazil

¹ E-mail: lcampos@inpa.gov.br

¹ E-mail: celiomag@inpa.gov.br

² International Institute for Aerospace and Earth Sciences (ITC)
Department of Geoinformatics
Division of Spatial Information Theory and Applied Computer Science
Hengelsetraat 99, P.O. Box 6, 7500 AA Enschede, The Netherlands

² E-mail: santos@itc.nl

² E-mail: deby@itc.nl

KEY WORDS: Data model, Database, Analytical tools, Biodiversity systems, Data integration

ABSTRACT

The National Institute for Amazonian Research (INPA) is the most prominent scientific institute in the Amazon rain forest. Due to its diversity in research subjects, it believes at large in automated systems as important tools, in fact, as the crucial mechanism in the decision making process for policy development for environmental preservation and biodiversity conservation. In this paper, we present INPA's initiative towards an institutional Biodiversity Information System. We discuss the main issues identified in a case study of the biological collection data, which include: data integration among research fields, system complexity, combination of multi-disciplinary skills, software platforms, data policy, data management and terminology. Also, we present a classification of information about INPA's collection. INPA in the past defined an Entomological Data Model and had implemented a prototype as a testbed.

Further, we present a new approach in biodiversity information systems by providing an environment with an open analytical database architecture, comprising three levels: the component, analytical, and information level. To conclude, we make observations about the computer technology available for implementing this architecture.

1 INTRODUCTION

Biodiversity Information Systems (BIS) are important tools in the decision making process for sustainable development and biodiversity conservation. The integration of database systems for sharing data, analysis techniques and results among scientific teams, can be a useful approach for understanding more about the earth's ecosystem, and consequently, to produce useful information in general. Environmental questions always involve a number of subjects. There is need for data and functionality to be made available in an integrated manner. Scientists from different teams often cannot fully use the data gathered by other teams due to poor metadata and poor context information. Hardware and software incompatibilities further endanger multidisciplinary data analysis. This in turn constitutes a fundamental risk for adequate environmental monitoring.

New trends in database technology are promising instruments for data integration and analysis, especially, Knowledge Discovery in Database (KDD), a new dimension in data exploration, integrated with Geographical Information Systems (GIS). In the biodiversity scenario, GIS can play an important role in providing geographic data and maps for environmental modelling and statistical analysis. This facility is one of the requirements of INPA's Bio-DB project, which utilises concepts of location, spatial distribution and relationships, and spatial objects.

INPA's Bio-DB was a case study exercise of implementing an institutional BIS. For that, we used the EDMI Model [16]. An Open Analytical Database Architecture (ADB), with its conceptual and logical

levels, with an emphasis in database technology, GIS functionality, Integration layer, and Web facility can be applied to achieve the desired functionality for BISs.

2 TOWARDS AN INTEGRATED INFORMATION SYSTEM: INPA's OBJECTIVE

Goycochea carried out the first assessment of requirements for a botanical collection database [6]. Sonderegger et al stated that INPA aims to develop an information system of Amazon biodiversity that should include collection records and research information from INPA's scientific fields [16, 13]. The framework of such a system concentrates on data integration among the predominating research fields at INPA. Since INPA's researchers will be the main users and contributors, questions related to their requirements, problems with data integration, and the relationships between research fields must be defined.

The development of an integrated institutional information system has been given a high priority. If well-designed, and implemented using scalable technology, information sharing among research teams will be eased and data management could also become more cost-effective.

Allkin observes that institutions throughout Europe, the USA and Asia invested large sums in attempts to develop BIS but these systems often failed and did not meet the expectations fully [2]. The reasons lie mainly in poor design and poor project management rather than in technical and financial limitations. In Brazil, INPA houses a wide range of research and development projects among many departments that, in some cases, have little data interchange. Allkin also observes that the common factor uniting the activities is geographical location rather than an agreed set of commonly held goals [2]. The complexity of an information system and the cost of its development are proportional to the complexity of the information stored and the functional requirements. INPA is aware of how costly system implementations can be.

The Amazon Surveillance System (SIVAM) Project is about to sponsor this INPA initiative and finance the development of such a system [15]. During the development, a combination of skills, in computer science and biology, and professional experience of the software development cycle is crucial. As emphasized by HMSO [7], the success of a huge initiative depends upon the integration of system development with the institute's strategic management and planning cycle. One thing must be above all considered: justification for such an information system and good system design require a clear, precise specification of who will be the users, and a thorough analysis of their needs.

3 AN INTEGRATED APPROACH TO MANAGING COLLECTION DATA

Goycochea [6] proposes a common approach to the collection of information across all collections. There is merit in this approach, since it is feasible in the short run, it is a manageable task, and there does exist a demand for such a system, which was already put forward in the institutional collection plan. This strategy may be sufficient to satisfy curatorial needs, however, an assessment to identify the demand for data access outside of INPA has yet to be completed.

Goycochea also suggests the use of international terminological standards, as this would improve the usefulness and accessibility of the information stored at INPA and at the same time help communication with different research teams world-wide. One output needed is a terminological standard extended to include terms, that refer to the Amazon region. A standard terminology must be independent from the software platform adopted.

Common sense dictates to consider various existing software for collection management within INPA, before an integrated solution is adopted. Obviously, the use of a single software package would facilitate maintenance and would probably allow considerable improvement in data sharing. It is mandatory that the future system is open for data migration in case a better system is developed in the future.

The systems in these categories that have been considered are *Specify* and *Brahms*. The system *Specify* is standard software for diverse disciplines (botany, zoology, fossil invertebrates, fossil plants and herpetology) to manage collection information [17], while *Brahms* is a database application designed to help curators, botanists and others working with botanical data with the assembly, storage, processing and publication of data in their field [4]. Both systems are now mature enough and INPA has adopted both

for their convenience. Brahms is used by other important herbaria in the region, and provides data standards and data dictionaries for the Amazon basin as part of its installation set-up. This tremendously facilitates the data interchange among plant collections in the region. Since both systems can export data, there is not an apparent disadvantage to adopt them in different collections.

3.1 Information association within the collections

The kinds of information associated with collections have been classified by Allkins [2]. It is presented in Table 1. The classification attempt to identify who might be responsible for generating and managing information, who are the potential users and what is the most appropriate delivery method. This classification does not consider the use of image as information sources.

Type of information	Who generates?	Who manages?	Who uses?	Appropriate delivery mechanism?
1. Curatorial Records Unique number, id, entry date, loan records, supplementary collections made, where stored etc.	Curatorial staff	Curatorial staff	Curatorial staff and INPA Management	Reports / INPA Intranet
2. Specimens Information obtained by visual inspection of individual specimen by staff or specialists.	Research staff / visitor / specialist visiting or receiving loans	Specialist	1. Systematics in INPA in Brazil and beyond 2. INPA staff in all Departments and projects identifying material	Access to all staff / Loans / Gifts duplicates for identification
3. Identification Identity of material – on original label or subsequent determination label	Collector or Systematic specialist	Specialist / Curatorial staff	All users	Checklists Diverse media
4. Label data Other information written on label(s) by collector. Geographical data (lat./long.), habitat descriptions, morphological details, etc.	Collector	Curatorial staff	1. Systematists 2. Ecologists 3. GIS 4. Conservationists 5. Etc.	Published checklists and maps. Books, CDs, internet or intranet.
5. Derived from specimens Information obtained from accumulation, comparison and analysis of data from many specimens.	Research staff / Systematic specialist / Analyst	Research staff / Systematic specialist / Information Dept.	Scientists, agronomists, foresters, conservationists, popular audiences	Field manual, monographs, Highly diverse

Table 1: classes of information associated with collections

3.2 Information dissemination from the collections

The interest of INPA is to distribute and disseminate information for a broader audience. The information from its collections will need to be analysed, summarised and presented in an appropriated way. How these data are analysed and presented will depend largely on one or more steps or agents involved in the process. Analysis and summary of information contained within the collections shall be done by specialists, who retrieve information about a subject for a wider audience. A good example has been applied in the Ducke Reserve Project [8], that implemented the information delivery wanted by INPA, as can be noticed in Figure 1.

3.3 A Tool-Kit approach: PNE experience

In the context of software variety it is important that a system is able to exchange data with another application. Allkin emphasises that the fundamental requirements for data exchange are primarily “the definitions and adherence to appropriate data standards” and secondly “transfer mechanisms between software packages” [1, 2].

The Project PNE (Plantas do Nordeste) is based in Northeast Brazil and it co-operates with more than 25 research institutions and NOGs, comprising 14 research projects in biodiversity and applied botany. To maximise information dissemination and identify priority audiences it has created the Plant Information Centre (PIC). PNE utilises the following criteria for adopting software solutions:

- Generic software that features data independence,
- Software that supports data quality and integrity maintenance,
- Mature products ,
- Software with functionality for data exchange,
- Software that ensures data security.

Three of the programs used by PNE are presented in Figure 2, illustrating how different programs can be adopted to serve a multi-disciplinary project. *Brahms* is used in some PNE projects to manage the herbarium data. *Alice* is used by all PNE projects to store information about plant species. Also stores taxonomic data. *Alice* is designed for managing data about taxa and is not capable for managing collection data. It can publish as reports, books, CDs, Web Pages and *Delta* is used to generate identification guides (electronically or not) from pre-processed species descriptions. These descriptions are available in the *Alice* database, from which subsets can be exported into *Delta*. *Brahms* can export data into the *Alice* system. Data can be exchanged among all PNE projects since they use the defined data standard (PNE-DDS).

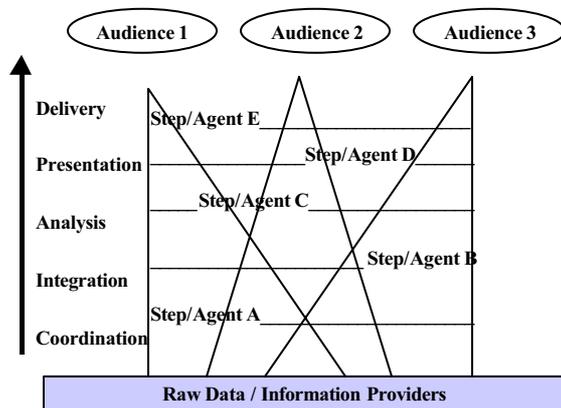
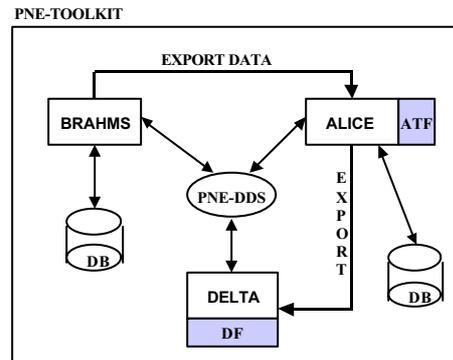


Figure 1: Scope of Information Delivery



Legend:

PNE-DDS: Defined Data Standards
 ALICE-ATF: Alice Transfer Format
 DELTA-DF: Delta Format

Figure 2: The PNE-Toolkit Approach

4 AN OPEN ANALYTICAL DATABASE ARCHITECTURE

We envisage an Open Analytical Database Architecture (ADB) that can contribute to improve BISS development. The architecture is comprised of three levels: a component, analytical and information level. The component level holds the applications, repositories, data catalog, modules for a specific purpose like data extraction, visualization facilities and GIS. The analytical level holds the main functions for data integration and analysis, providing functions for simulation, geoprocessing, and aggregation. The information level consists of facilities for user-data interaction in a Web environment to access and maintain data, and produce information from it. We now discuss each level in more detail.

4.1 Component Level

This level deals with source data, originating from experiment sampling. This kind of data may exhibit different degrees of quality and is generally classified as:

1. Raw/sensor Data: Data sets obtained directly from aircraft reconnaissance, ground-based sampling, satellite imagery, inventories etc.;
2. Calibrated Data: Results from calibration operations of the raw/sensor data;
3. Validated Data: Results from quality assurance (QA) and quality control (QC) procedures over calibrated data established by a project's research team.

The main objective is to store data necessary for analysis. The idea is that each data source/sampling will correspond to a database at this level, called Component Data Repository (CDR).

The structure of such a repository is defined by its Component Schema, a conceptual schema that is implemented via the CDR Mapping Module. We assume that every scientific team is responsible for a

CDR and works independently. Figure 3 presents a diagram that illustrates the architecture of the component level.

The component level must also store the associated metadata. This metadata is important for further analysis and integration. We use the term metadata here in the same sense as [18], that is, information on what is stored and how it is stored. We remark that the component repository will have metadata like temporal data, geo-references, sampling method identification, measure device specification, data provider identification (who collected the data, what type of data, when was data collected), etc. This data is called context information. Organisation the collected data in a CDR imposes the need for a CDR Catalog, which can be browsed and searched before data is needed.

4.2 Analytical Level

Based on the information collected at the Component Level, scientific team can use this data as the object of study. The idea is to build analytical databases that can use data from several CDRs. These ADBs have their own database schema, and can generate their own data using analytical tools.

In many domains, especially computer science, designers are required to make complex associations between entities/objects in the real world and conceptual representations for them. Any model is a surrogate object, a conceptual representation of real phenomena. The scope of a model can include the names for entities/objects represented, descriptions/attributes that represent properties for the entities/objects, formal descriptions of the behaviour of the model, and an interpretation relating the part of the model (and its output) to real world properties that entities/objects represent. A model helps to predict the behaviour of the original universe of discourse based on behaviour of the model. Scientific teams can define computational modelling and simulation environments to extract knowledge from large multidisciplinary data sets. It is considered important for scientific teams, to have environments for programming, pure simulation and model-building as tools to extensively explorer databases and generate derived data.

There are at least three important kinds of derived data: data derived from simulation, data derived from geo-processing and aggregated data.

The users at this level are the **Data Brokers**. To extract knowledge from the data, they may need correctly integrated analytical tools such as simulation systems, GIS, visualising tools. We call attention to a new set of methods that may be suitable for application at this level and which are generically known as data mining and knowledge discovery methods. These methods are used to identify potentially useful, and ultimately understandable patterns in large data sets. This level could also be used to record and document the results from research conducted by scientific teams.

4.3 Information Level

This level addresses the dissemination of data and research results stored at the two above levels of the architecture to a broader audience. This can be done by means of a customised database schema, that is, an information schema, that addresses specific aims, policy makers. Access to this schema can be made through available public interfaces. Web technology and Java applications may provide facilities like searching, information retrieval, and scientific visualisation.

4.4 GIS Capability in an ADB Environment

Biodiversity research has a distinct need to represent and manipulate geographic data. Not only its presentation but also statistical analysis should be supported as well. Access a good visualization is an important feature in any information system. The two data types, tabular and geometric data, must be processed in such a way that users, expert or not, are allowed to obtain generalized and specific views of the data

We assume that the GIS has to provide summary information of a geographic region. The architecture of Figure 4 allows to incorporate GIS as part of the Component Level. At the Information Level, we have the users who access a Web Server. The architecture facilitates the access to tabular data, which originates from any of the CDRs at the lower level. As we mentioned earlier, the CDRs can have different DBMS platforms (e.g., relational and object-oriented).

The GIS functionality should handle geometric data and metadata from the GIS Function Module, which manages the GIS CDR and GIS I/O. As experienced by [11], geometric data can be stored in individual GIS formats, whereas metadata about geometric data can be stored in a relational DBMS. The idea of having the GIS I/O is to deal with GIS data for import/export functions to other GIS. This could also allow users to access the geodata and metadata in the Web environment.

We believe this approach may answer requests like *“Display a map with the distribution of certain animal or botanic species in a certain area of the Amazon Basin.”* Also, the result can be used for further queries and map generation. The proposed Visualization Module must be able to generate charters, maps, images, and HTML scripts.

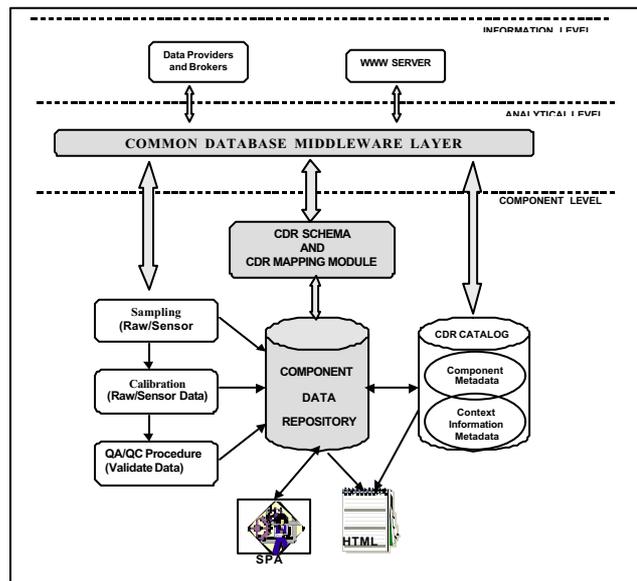


Figure 3: Component Level

5 CONSIDERATIONS ON AVAILABLE TECHNOLOGY

Current state-of-the-art database and user interface technology can support the proposed architecture, and is currently in use in systems within the LBA [12], while others have been presented as having a good potential to be applied in these systems.

5.1 CDR and CDR catalog

In our view, the paradigm that best fits the CDR is that of object server, i.e., a CDR as an interactive entity that supplies data stored in it upon requests of clients. We can consider the clients as the databases of the Analytical Level. In an ideal environment, the clients do not need to be aware of where the CDR are, but they do need to know what information is available from the CDR, how this information can be used (formats, structures, etc.), and what can be done with this information to help obtain knowledge.

The role of provider of this meta-information is played by the CDR Catalog. This catalog must be available to every Data Broker to let it define its Analytical Databases. Thus, whenever a CDR is read, it publishes its information in its CDR Catalog, so that the Data Brokers can use it via the Information Level.

Depending on these characteristics, the structure of a CDR has seven main components: the data sets provided by the CDR, a local CDR catalog, a retriever, a translator, a server, HTML scripts and specific purpose applications (SPAs).

To implement the Component Level, we consider the adoption of Common Object Request Broker Architecture (CORBA), a standard for open, distributed systems [14]. This standard allows the implementation of object servers that retrieve data from several sources (i.e., flat files, DMBS etc.) and

pass this data to clients in a network environment using a common format. CORBA was successfully applied in the system WWW-UDK [10], and we noticed that systems within SIVAM and INPA's BioDB present similar characteristics.

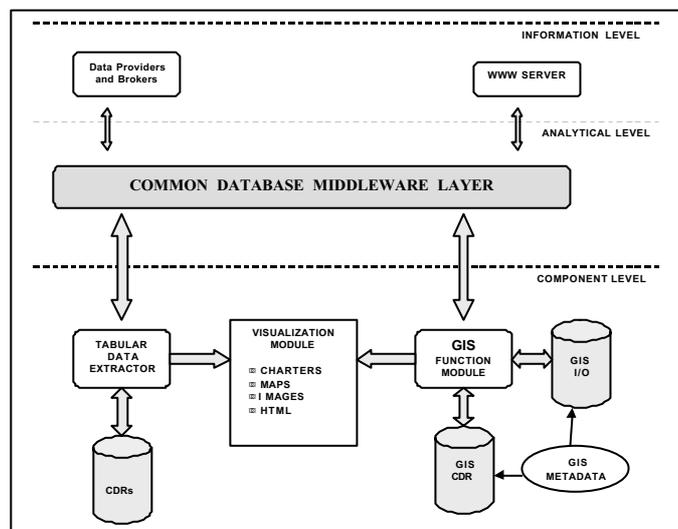


Figure 4: GIS functionality integrated by a common middleware layer

5.2 Analytical Databases

We view an ADB as a scientific notebook that will contain primary data as well as derived data, generated by Data Brokers in the corresponding ADB. Thus, these two kinds of data must be maintained in a DBMS. Due to the fact that geo-referenced data is present in almost all scientific experiments that are carried out in the Amazon region, special attention must be paid to spatial data storage formats. These observations lead us to using a *spatial database* [5].

There are two classes of DBMS that seem to be more adequate: *extensible relational DBMS* [5] and *object-oriented DBMS* [3]. It is our opinion that despite the great modelling flexibility of the second class, the first class tends to be more appropriate since there are several implementations that offer a good reliability and full DBMS functionality. Also at this level, the analysis of stored data must be considered. Tools that can use stored data as well as generate and store derived data must be present. We believe that there are two main types of analytical tools that can be used: *modelling and simulation tools* and *GIS analytical functions*. To integrate these analytical tools with the ADB environment we consider two options:

1. We can implement the needed functions as extensions to the primitive functions of some DBMS, as long as the DBMS allows these extensions, and,
2. We can implement import/export interfaces between the analytical tools and the DBMS.

Extensions to DBMS functionality require much effort to implement but, in some cases, there are packages of such functions available for specific DBMS. We may mention Oracle Spatial, DB2 Spatial Extender, and the SpatialWare DataBlade Module from Informix. These extensions allow users to store, access, manage, and manipulate spatial data in the same database besides traditional data.

The second, more flexible, option does not deprive scientific teams from their favourite analytical tools. This is particularly important for using modelling tools, which, in many cases, are systems previously developed and that work well for the desired goals. To implement this integration we suggest again the association of object servers with the analytical tools, based on the CORBA standard.

Another important discussion is about using heterogeneous DBMS at the analytical level. Although this would not pose a technical problem, we feel that using the same DBMS for several ADB is more convenient: experience can be shared between implementers of the ADB, the sharing of integrating solutions between ADBs, the sharing of extension modules between ADBs (in the case of extensible DBMS), and the ease of integrating ADB in database schema of the information level.

Another possibility to help with the analysis of data at this level is the application of *Data Mining and Knowledge Discovery* methods. The application of these methods for scientific experiments is an active subject of research in computer science. This approach is a promising one for extracting knowledge from large data sets of environmental information.

5.3 Dissemination of Information and Public Interfaces

The access of information stored in a scientific database presumably helps research, since it is possible to solve problems using data from multiple sources. Databases like Genome, Microbial Germplasm, Natural Fungus Collection and Nature Conservancy are being developed using relational technology. The users of these databases are mostly scientists. A number of difficulties exist in locating and accessing data efficiently. They range from the syntax of the query language to the way the language is used during interaction, which usually comprises command line interfaces.

Tool developers have built a variety of toolkits for constructing browser-based interfaces to databases. These facilities can be categorised into one of three classes: interoperability language, schema based and Graphical User Interface (GUI) based environment.

When applications will be made available in the environment, the use of Java seems to be appropriate, since with this language one can enhance the overall performance of the system by transferring some processing tasks to the front-end site, to be achieved by Java applets. Another important aspect with Java is its portability. It is architecture neutral language, its compiler generates Java Virtual Machine codes instead of machine specific to the computer system one is using. Java system itself is portable, the new compiler is written in Java and the runtime libraries are written in C with a clean portability boundary [9].

6 CONCLUSIONS

In this paper, we presented INPA's initiative toward an integrated information system and an open database architecture for BISs. The architecture presented is flexible and can address the main concerns in the integration of BISs because:

- a) It is based on the independence of data providers and data brokers, since they are, in general, distinct teams with specific goals;
- b) It does not force data providers or data brokers to use any specific tool, instead, it creates opportunities for them to use the tools that best fit their tasks;
- c) It permits a gradual development of an ADB for BIS, because CDRs can be added as they are needed and can be developed concurrently;
- d) It is potentially scaleable, that is, new CDRs can be added later to the system, which can grow incrementally, even after the end of an experiment;
- e) It can ease the development of the ADB, because the job of developing a huge database is broken into smaller parts that can be further integrated.

There are some potential disadvantages of this architecture, that must be investigated:

- a) the need of database experts to help in the development of CDRs and ADBs;
- b) the need for specific integration solutions for the CDRs and ADBs, that is, depending on how idiosyncratic an external tool is that must be integrated, much effort might be required to solve a specific problem;
- c) projects may not have a "global" database with all its data stored in it;
- d) some of the approaches presented are new and would imply deeper research and training.

A fundamental aspect of this architecture is the middleware layer used to integrate different systems that can be used in the ADB architecture. It is of crucial importance that the ADB designer teams have a good understanding of the technology used. The integration of different data sources on CDRs, the integration of CDRs in the CDR catalog and the integration of external analysis tools in the ADBs are heavily dependent on it. Note that at the analytical level, different DBMSs could be used in different ADBs, but we think that to use the same DBMS is, as discussed, the best option.

It must be noted that the integration problem is not only a technical matter. In addition, there must be an administrative effort to define standards and policies to make this difficult process work correctly and enforce correctness of the information to be made available.

With respect to the database technology discussed, most of it can be classified as current state-of-the-art. An ADB represents a good opportunity to apply this technology in a practical and rather important problem, which will serve as motivation for future development in the field of database technology.

ACKNOWLEDGMENTS

We are grateful to our colleagues Jürg Sonderegger, Altigran Soares, Paulo Petry with whom we had useful discussions on the subject and to Nívia Oliveira for leading the prototype development.

REFERENCES

1. Allkin, R., 1997. Data management within collaborative projects. In "Plant diversity in Malesia III", Dransfield, J., Coode, M. J. E. and Simpson, D. A. (eds). pp. 5-24. Royal Botanic Gardens Kew, London, 1997
2. Allkin, R., 1998. Collection Information Management at INPA. INPA Internal Report (28 July 1998).
3. Blaha, W.; Premerlani, W.; 1997. Object-Oriented Modeling and Design for Database Applications, Prentice Hall, September 1997.
4. Brahms., 1999. Botanical Research And Herbarium Management System – BRAHMS, Department of Plant Sciences, University of Oxford, February 1999. <http://www.brahms.co.uk/draft/home1>
5. Elmasry, R.; Navathe, S., 1994. Fundamentals of Databases System; In Proceedings of 7th International Symposium on Spatial Data Handling (SDH'96); M. J. Kraak & M. Molenaar, eds., IGU; 4B. pp. 15-29; 1996. 2nd ed, Benjamin-Cummings; 1994.
6. Goycochea, A. P., 1988. Reporte de la consultoría para "diagnosticar a situação atual do gerenciamento das coleções botânicas do INPA e recomendar procedimentos para sua modernização", 37p. April 1998
7. HMSO., 1990. Strategic planning for Information Systems: ensuring that the business benefits. Central Computer and Telecommunications Agency. HMSO, London 199 pp. 1990.
8. Hopkins, M., 1999. Projeto Flora da Reserva Ducke. <http://www.inpa.gov.br/projetos/ducke/index.html>
9. JAVA., 1999. Java's Portability; In <http://tech8.informatik.uni-hamburg.de/Dokumentation/java/>
10. Kramer, R. et. Alli. WWW-UDK: A Web-based Environmental Meta-Information System. *ACM SIGMOD Record*, Vol. 26, No. 1, March 1997.
11. Koschel, A et alli, 1996. A Federation Architecture for an Environmental Information System incorporating GIS, the World Wide Web, and CORBA". Proceedings of the 3rd International Conference/Workshop Integrating GIS and Environmental Modelling, Santa Fé, New Mexico, USA (January 1996). http://negia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/nikolai_ralf/fedarch.html.
12. LBA., 1997. LBA-DIS Recommendations Report, LBA-DIS WG, College Park – Maryland, October; 1997
13. Oliveira, N. M., 1999. Projeto e Implementação do Modelo de Banco de Dados Entomológico do INPA. Monografia de Projeto Final, Departamento de Ciência da Computação, Universidade do Amazonas, 1999.
14. OMG., 1997. Object Management Group. A Discussion of the Object Management Architecture, January 1997.
15. SIVAM., 1999. Amazon Surveillance System. <http://www.sivam.gov.br>
16. Sonderegger, J., Petry, P., Campos dos Santos, J. L., Alves, N. F., 1998. An Entomological Collections Database Model for INPA. In the Proceedings of the 17 International Conference on Conceptual Modelling (ER'98), Singapore, 1998.
17. Specify., 1999. Specify 1.1 is Here, University of Kansas. <http://www.usobi.org/specify/projdesc.html>
18. Tomasick, A. Simon, E., 1997. Improving Access to Environmental Data Using Context Information. *ACM SIGMOD Record*, Vol. 26, No. 1; March 1997.