



## **ENVIRAD – An integrated database of environmental radiological monitoring information**

Gonzalez<sup>a</sup> S.A., Silva<sup>a</sup> T.A., Conti<sup>a</sup> C.C., Osvath<sup>b</sup> I., Ramadan<sup>b</sup> H., Bartocci<sup>b</sup> J.

<sup>a</sup> *Institute for Radioprotection and Dosimetry (IRD/CNEN), 22780-160, Rio de Janeiro, RJ, Brazil*

<sup>b</sup> *Marine Environment Laboratory (MEL/IAEA), 98012, MC, Monaco*

*gonzalez@ird.gov.br*

---

### **ABSTRACT**

With the increasing use of databases for environmental radiological monitoring of soils, food, water and sediments, among others, there is a demand for integrating the availability of this type of information, aiming at a data management tool for greater searchability for the most varied purposes. ENVIRAD is a new database proposing to integrate a set of previous databases, providing researchers with a tool for a more robust analysis of environmental studies. ENVIRAD, as developed so far, works with the integration of the following databases: MARIS, GLOMARD, GEORAD and SIGLARA databases, enabling queries and demonstrating great potential in the use of data integration for environmental samples.

***Keywords:* database integration, radiological monitoring, environmental database**

---

## 1. INTRODUCTION

The objective of environmental radiation protection is to reduce the risk of exposure to ionizing radiation of animals and human beings, as well as helping to protect the environment from radioactive contamination. It consists of a set of measures taken to restrict the exposure of professionals and the public to radiation in the environment resulting from the release of effluents by authorized practices, accidental contamination or deposition of radioactive material. Actions include the monitoring and control of releases; the radiation fields present in the environment and the levels of radioactivity accumulated in the various environmental compartments. This is done through regular monitoring of the source of the effluents and the environment and the environmental radiological evaluation, using recommendations of the organizations competent in radiation protection and especially the requirements established by the country's regulatory authority [1].

The main organizations that establish guidelines and recommendations for radiation protection are the International Committee on Radiological Protection (ICRP), an independent scientific and technical committee, and the International Atomic Energy Agency (IAEA). In Brazil, the National Nuclear Energy Commission (Comissão Nacional de Energia Nuclear - CNEN), created on October 10, 1956, is a regulatory authority which is responsible for establishing standards and regulations in radioprotection, as well as licensing, overseeing and controlling nuclear activity in Brazil [1].

Nuclear or radiological emergencies result from radionuclides being released into the environment. In order to respond to this kind of emergency in real time, a large network of laboratories is dedicated to the analysis of environmental samples in order to determine the most diverse factors, such as: water potability, soil quality and air pollution. More specifically in the nuclear area, the analysis of samples of various types is aimed at determining the presence of nuclear and radioactive material and their concentrations, in order to determine if the quantities are within acceptable environmental limits so that there is no risk of contamination to plants, animals and human beings.

In the radiological environmental monitoring field, a database allows the knowledge of the radiation levels of the monitored site and can be used as a reference for environmental impact studies and, where appropriate, proceed with decontamination. In environmental analysis laboratories, sample results are recorded on different storage media, whether spreadsheets or databases. The data

are presented for the most varied matrices and in different units and codes. Given this variety, a comparative study among labs and samples may be difficult, slow and subject to failure.

The IAEA has two laboratories in the environmental field: Seibersdorf Laboratory, responsible for terrestrial samples, and the Marine Environment Laboratory (MEL), as the name suggests, for marine samples. These laboratories support the laboratories of Member States to improve the quality of analytical measurement and determination of radionuclides. They also include the development of guidelines and standards in the field of radiation protection for the environment and the development of radioecological methods to be used in radiological impact assessment. MARIS is a database of samples of radioactive contamination in the marine environments (open sea and coastal zones), specifically in seawater, particulate matter, biota and sediment. It was developed by the IAEA for the Marine Environment Laboratory. Before MARIS, the GLOMARD (Global Marine Radioactivity Database) was previously used for this purpose. The sources of all the data are published scientific papers, reports and databases created within institutes or scientific programs in the Member States and can be used with appropriate citation of sources and references [2].

CNEN, through the Institute for Radioprotection and Dosimetry - IRD, has two databases: GEORAD (database of environmental samples in the Brazilian environment) and SIGLARA (database of radionuclide levels in typical Latin American food) [3, 4].

Despite a significant number of radiological environmental monitoring databases, an integrated structure of these databases has not yet been established. Database integration is critical for spatially distributed information to be available in a clear and easy way to researchers and decision makers.

There are currently 25 databases worldwide in the area of environmental radiological monitoring, and no integration or interaction of these databases has been identified [5].

The goal of data integration is to provide unified access to multiple data sources. It gives users the illusion of interacting with a single source [6]. This integration of heterogeneous data sources benefits companies and researches. However, data integration is not an easy task. Finding the common schema, combining different fields, choosing the better approach and filtering elements becomes difficult due to heterogeneity [7] [8].

Data integration is the combination of two or more sources giving the users a unified view of these data [9]. It can be accomplished in two ways: physically and logically. The physical way, in a form of a global database, referred to as a data warehouse technique. The logical way is only virtual and not in the form of a global database [10].

With the increasing use of databases of environmental radiological monitoring in soils, food, oceans and others, there is a greater demand for the integration of these databases in order to prevent and to reduce the risk of contamination, and also, to provide scientist and researchers with a better understanding and a global view of the environment.

ENVIRAD presents a proposal of integration of several databases and, as a result, provides the researcher or the decision maker a better analysis of the monitored environment.

Two major strategies have been identified for designing integrated databases: a top-down and a bottom-up approach. The top-down approach is suitable for homogeneous distributed database management system (DBMS). The task involves integration into a single database. The bottom-up approach is appropriate in multidatabases systems [11].

The integration process used for ENVIRAD is the bottom-up methodology and prepared to integrate four databases: MARIS, GLOMARD, GEORAD and SIGLARA.

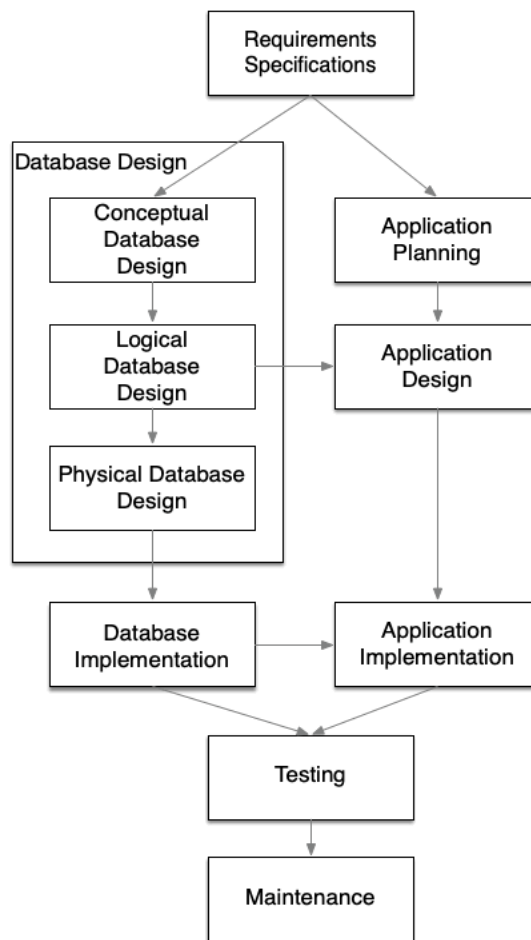
**Table 1:** Databases integrated on ENVIRAD

	<b>Databases Integrated on ENVIRAD</b>			
	<b>MARIS</b>	<b>GLOMARD</b>	<b>GEORAD</b>	<b>SIGLARA</b>
Institute	MEL/IAEA	MEL/IAEA	IRD/CNEN	IRD/CNEN
Type of Samples	Marine Environment	Marine Environment	Brazilian Environment	Latin American food
Origin of Samples	(a)	(a)	Bibliographic Research	(b)
Structure Integrated	Yes	Yes	Yes	Yes
Data Uploaded	Yes	Yes	Waiting for owner's approval	Waiting for owner's approval
(a)	MEL - Marine Environment Laboratory ; ASPAMARD - Asia Pacific Marine Radioactivity Database; BSH - Bundesamt für Seeschifffahrt und Hydrographie; HAM - Historical Artificial Radionuclides in the Pacific Ocean and its Marginal Seas; HELCOM - Helsinki Commission; JAEA-JASPER - Japan Atomic Energy Agency (JAEA); RADNOR - Radioactive dose assessment improvements for the Nordic marine environment: Transport and environmental impact of technetium 99 (99Tc) in marine ecosystems; RWS - RWS Centre For Water Management.			
(b)	Environment Radiological Labs: Argentina, Brazil, Chile, Costa Rica, Cuba, Ecuador, Mexico, Paraguay, Peru, Uruguay and Venezuela.			

## 2. MATERIALS AND METHODS

In order to properly design a database, as shown in Figure 1, it is necessary to have a full understanding of the requirements of the application and, also, to translate these requirements into a conceptual model [12].

**Figure 1:** Database design process [12]



Conceptual data modeling is how one drives into a logical database design [13]. Conceptual design is the process by which the application is determined and, therefore, creates entities and the relations between them. This type of diagram (entity-relationship) has been shown to be a good tool

during the requirements survey and the logical design step to establish communication between database designers and end users [11, 14].

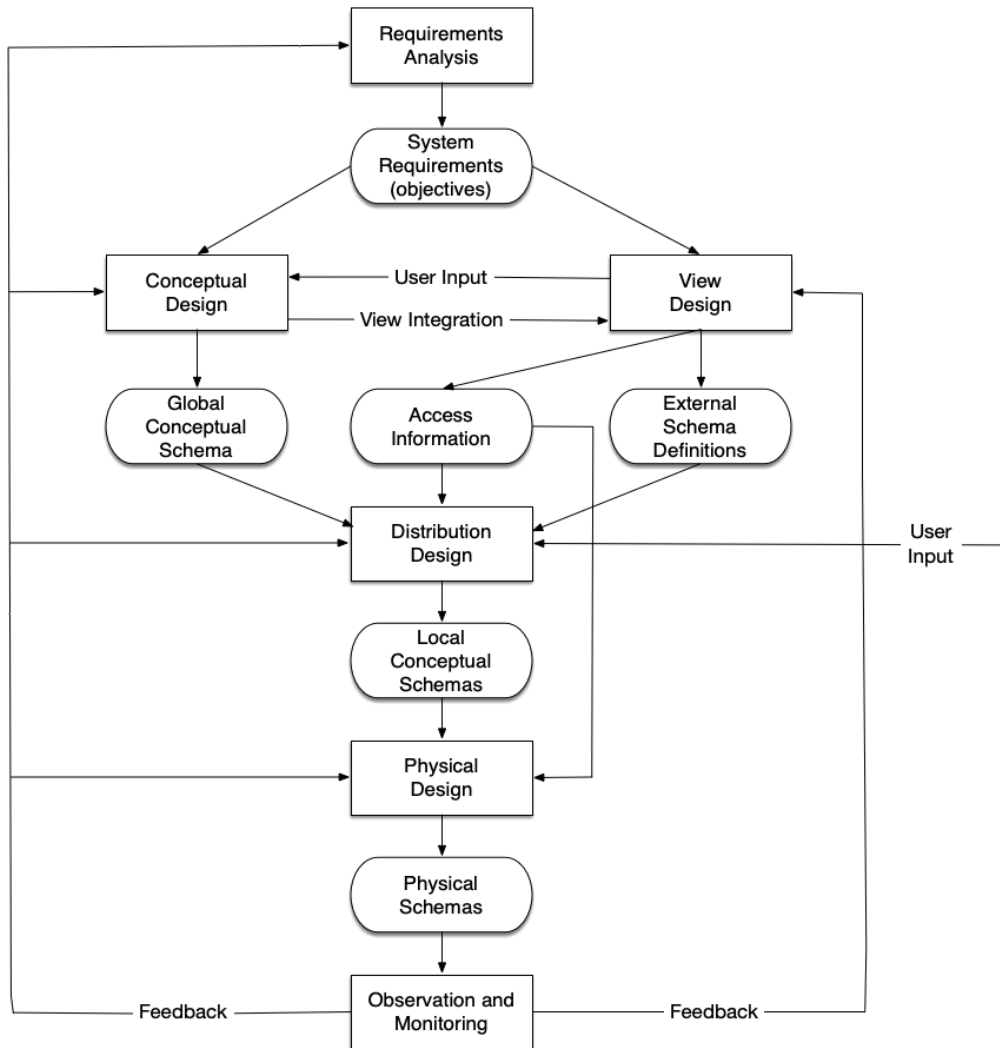
Data abstraction is the suppression of details of data organization and storage, highlighting its essential elements. Its objective is the improvement of data understanding. There is yet another important definition, the database schema, which is the description of the database, different from the database itself. Database description is specified during database design [15].

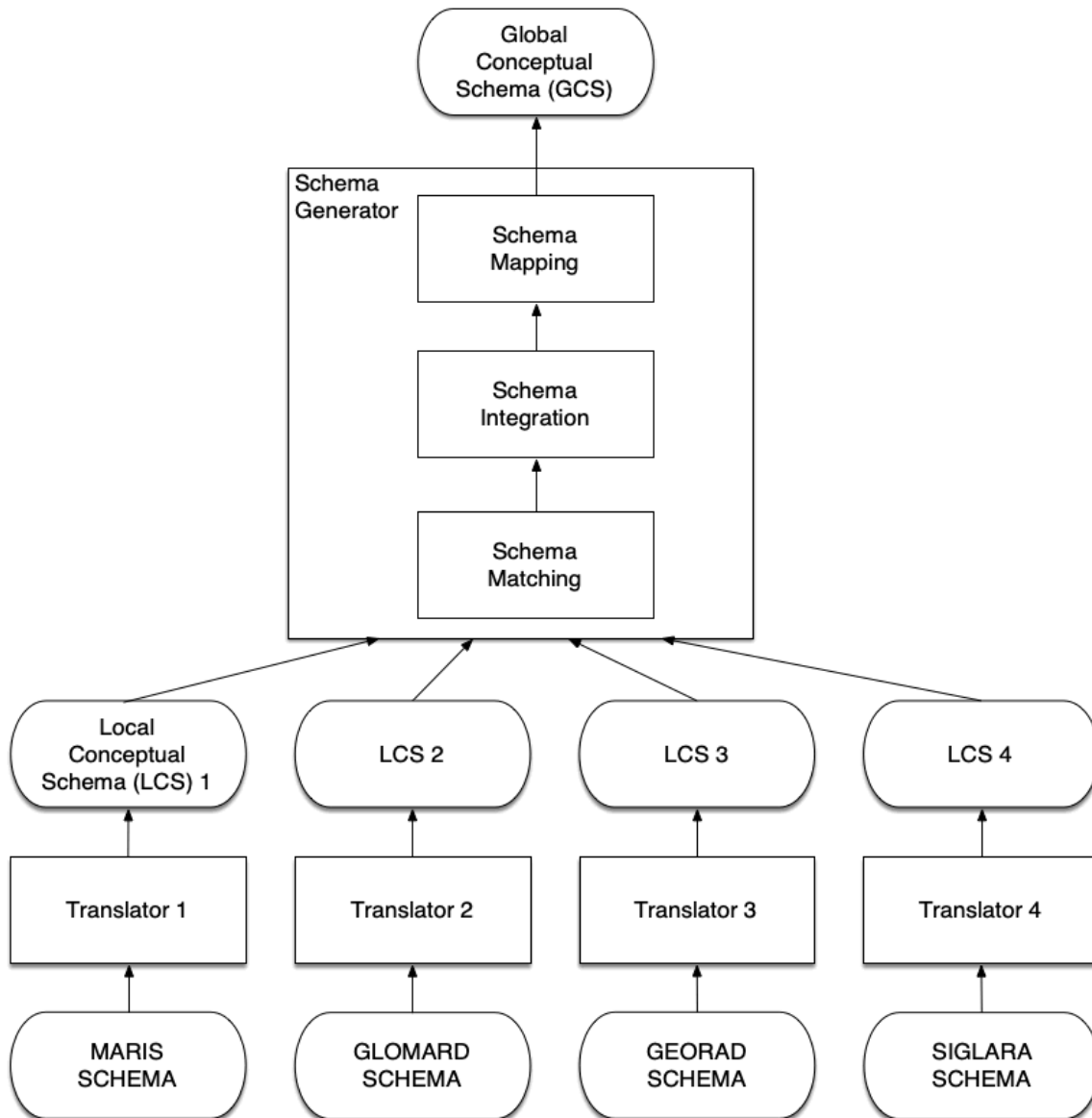
There are two approaches for designing a distributed database: top-down and bottom-up. The top-down approach, as shown in Figure 2, is suitable for integrate, homogeneous, distributed databases, and also a good approach to design the database model from scratch. Bottom-up design, as shown in Figure 3, is the process to integrate heterogeneous databases, either physically or logically, into a cohesive multidatabases [14].

ENVIRAD architecture is based on a global schema and the database integration methodology used was the bottom-up approach. Four databases were involved in this project GLOMARD, MARIS, GEORAD, SIGLARA which were integrated into a single cohesive database.

The bottom-up methodology is based on creating a Global Conceptual Schema (GCS) and, then, mapping the Local Conceptual Schema (LCS) into the GCS. Data warehouses use this approach, but it is not restricted to them. Other data integration methodologies may follow the same strategy [14]. The steps of this methodology are described in the following subsections.

**Figure 2:** Top-down design process [14]



**Figure 3:** Database Integration Process – bottom-up [14]

### 2.1. Global conceptual schema – GCS

The conceptual model was developed in the MySQL Workbench™ software encompassing MARIS and GLOMARD databases from MEL. This choice was made to guide a data structure for environmental monitoring of projects of the IAEA's Member States. GEORAD and SIGLARA, from IRD, were the databases used for final validation of the structure created.



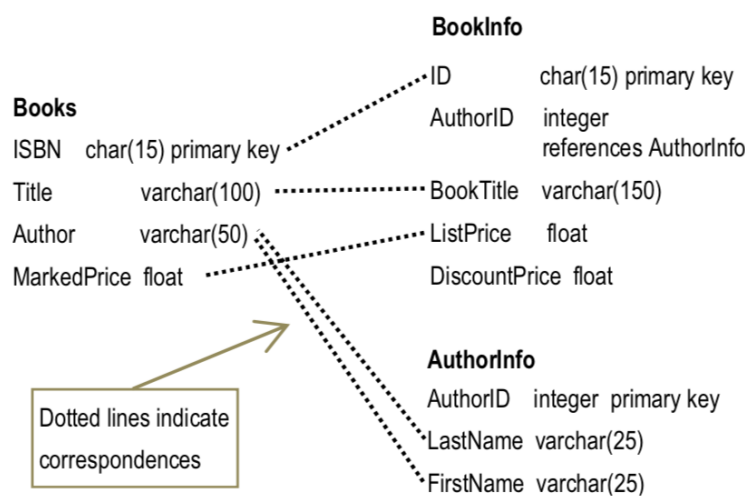
The bottom-up approach was used in this stage of the project. The software MySQL Workbench™ was used for the reengineering of the GEORAD and SIGLARA databases. Since these were MySQL™ databases, the tables, relationships and fields were obtained with this software.

For the MARIS and GLOMARD schemas, since they were developed using Microsoft SQL Server™, obtaining the tables, relationship and fields was much simpler because this DBMS has a feature that shows all the details regarding tables and relationships.

## 2.2. Schema matching

Once the GCS is defined, the correspondence between each local databases and the GCS (target) must be defined [14]. Schema Matching, as shown in Figure 4, is the process of generating correspondences between fields of these schemas [16]. Four group of schema matching were done in this project.

**Figure 4:** Example of Schema Matching [16]

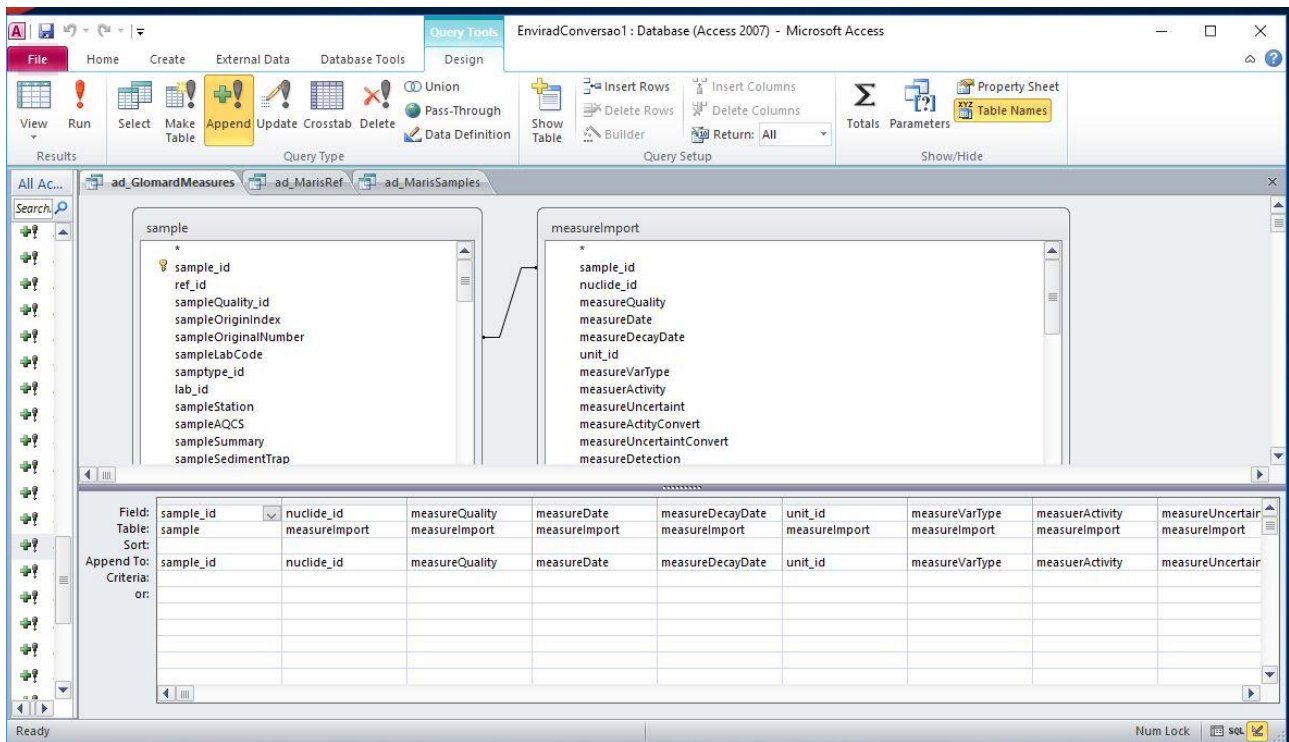


Each target database was associated with an Open Database Connectivity (ODBC). There was three MySQL ODBC™ (ENVIRAD, GEORAD, SIGLARA), and two Microsoft SQL Server ODBC (MARIS, GLOMARD).

For each source database the schema matching Microsoft Access™ (MsAccess) was used, connecting one specific ODBC from the source database with target ENVIRAD ODBC. Each table,

and their fields from the source would correspond to ENVIRAD table and fields, as shown in Figure 5. MsAccess™ is a simple to handle data integration conflict.

**Figure 5:** Microsoft Access™ (MsAccess) - Correspondence between fields



### 2.3. Schema integration

According to the bottom-up methodology, this part of the Schema Generator is not necessary once the GCS has been already defined [14]. In this project, the GCS was built after analyzing four databases, performing a reengineering, merging all source databases and interviewing the end users that deal with radiological sampling.

### 2.4. Schema mapping

MARIS and GLOMARD databases were developed in Microsoft SQL Server™ and the migration and conversion of the data to the new structure was done using Microsoft Access™ software and a MySQL Connector™ / ODBC source to, through Microsoft Access™, create update queries with the

role of reading the tables from the Microsoft SQL Server database, convert and include them in the MySQL™ database.

The application was created using Scriptcase™ software, PHP code generator development tool (open source language). This software also generates Web pages written in PHP in several languages by simply translating the fields from the database tables. In the case of ENVIRAD the system is available in English, Portuguese and Spanish, other languages can be added afterwards.

The ENVIRAD allows the query from one or more databases to be incorporated into it, listing the concentrations of the samples and the values registered for the radionuclides.

ENVIRAD has been developed using open source software and it is available at [envirad.ird.gov.br](http://envirad.ird.gov.br).

### 3. RESULTS

ENVIRAD was developed in Scriptcase™, a PHP code generator, and MySQL™ database. It is available in three languages: English, Portuguese and Spanish. Any other language can be added to the code, if necessary, through the Scriptcase™ tool, without changing the code itself, but only by generating a new version.

The system is available on the Internet for the IAEA Member States. In addition to that, laboratories from IAEA Member States may have their own databases included in ENVIRAD upon registration and making the database available. During the registration process, the "owner" of the data can update and consult their records. Afterwards, the data will be made public, allowing others to make queries, but not update it.

Figure 6 presents the ENVIRAD's conceptual model; Figure 7 presents the querying page and; Figure 8 presents the query results page.

Figure 6: ENVIRAD Conceptual Model

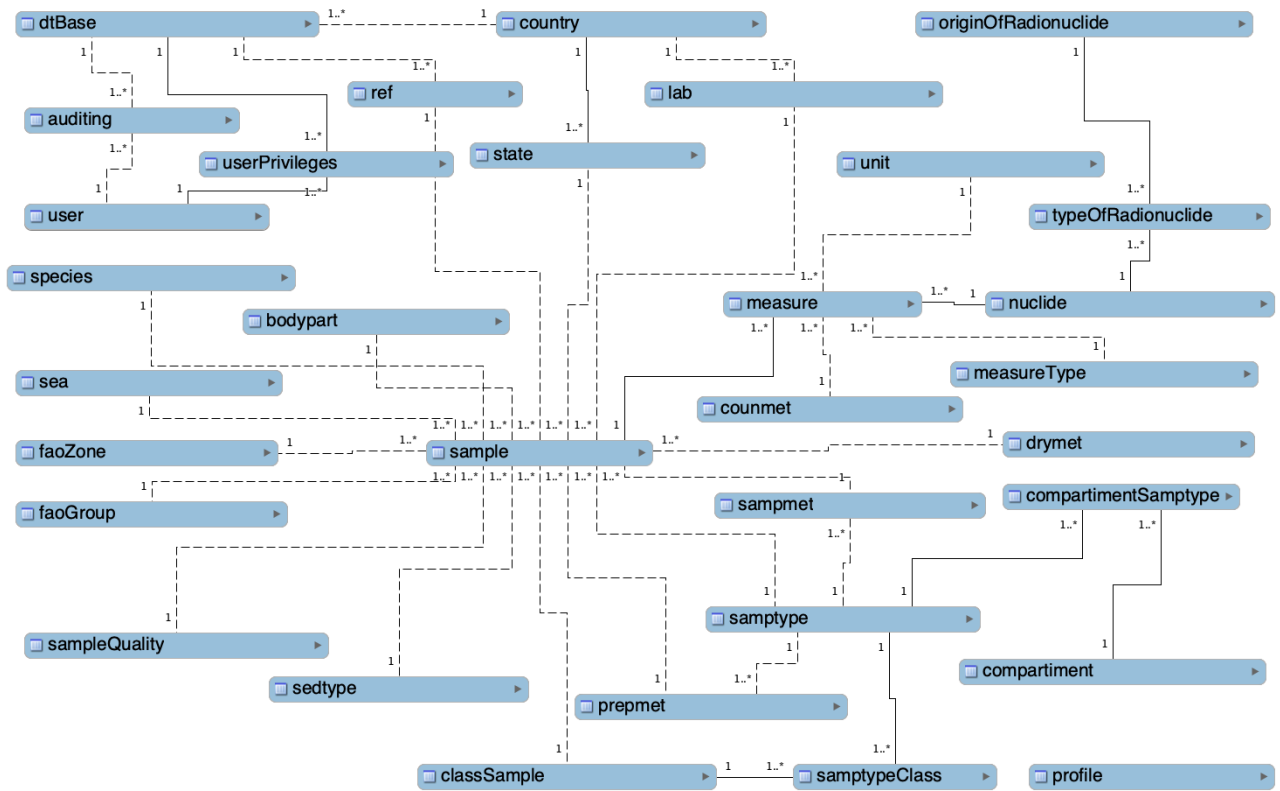
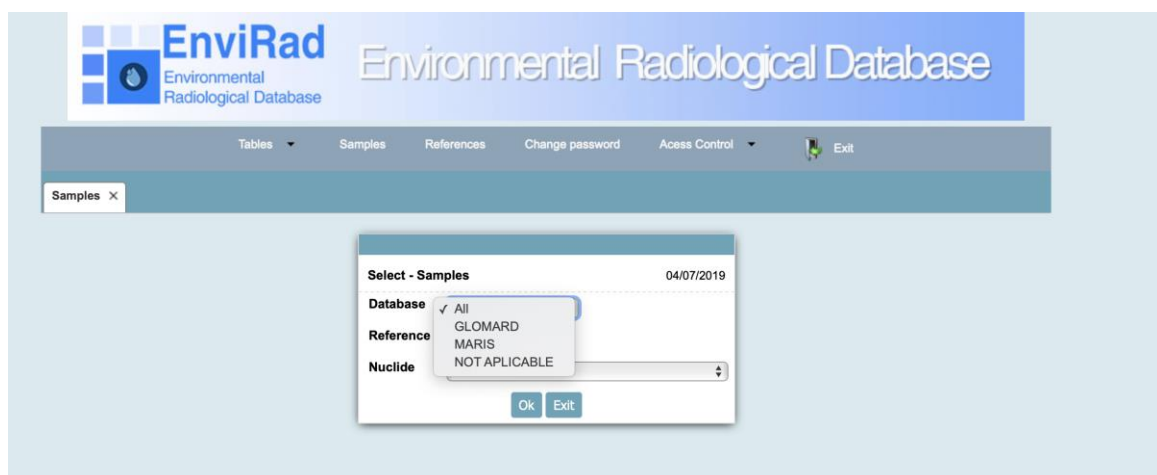


Figure 7: Querying page



**Figure 8:** Query Result showing the samples and the concentration

The screenshot displays the EnviRad web application interface. At the top, the logo and name 'EnviRad Environmental Radiological Database' are visible. Below the navigation bar, the 'Samples' tab is active, showing a form for editing or adding a sample. The form is divided into several sections: General Information, Laboratories, Sample Conditions, Coordinates, and User Input. The 'Sample Conditions' section is currently selected, showing fields for Reference (138), Sample Quality (GOOD), Original Index (58300), Laboratory Sample Code (58300), Note, Country (MONACO PRINCIPALITY), State (Not Available), Class (NOT APLICABLE), Begin Mission, and End Mission. Below this, the 'Concentration' section is visible, featuring a table with one data row for 137Cs - CAESIUM. The table columns include Nuclide, Quality, Measure Type, Date, Decay Date, Activity, Uncertainty, Unit, Activity Converted, Conversion Unit, and VarType. The activity is 14.400000 Bq/m3 with an uncertainty of 1.584000. At the bottom, there is a 'Go to' field and a 'View' dropdown set to 10.

Nuclide *	Quality	Measure Type	Date mm/dd/yyyy	Decay Date mm/dd/yyyy	Activity *	Uncertainty *	Unit *	Activity Converted	Uncertainty	Conversion Unit	VarType	Deta
137Cs - CAESIUM	Good	MEASURE			14.400000	1.584000	Bq/m3	14.400000	1.584000	Bq/m3		=

The Global Conceptual Schema was created after reviewing four radiological monitoring databases, making it possible to create a schema that could be used by others Member States.

Data integration is a challenging task, but the analysis of several databases and the creation of a model that caters to most radiological monitoring databases has been developed and will allow researchers easier access to radiological and integrated information.

The system is a useful tool for the researchers and decision makers. It is possible for any Member State to join this project and would greatly contribute to the system with their own data, improving the database and enabling a better view of the whole.

It has been observed, in some cases, a need for the integration of heterogeneous databases, in which all interested researchers can be part of the general structure with their databases, thus promoting the availability of a vast body of knowledge to all interested parties. For those who do not have the structure and/or the knowledge to develop a database, ENVIRAD will aim to provide a basic structure for all stakeholders. It has also been identified a need for database import functionality

through Microsoft Excel™ spreadsheets, visualization of georeferenced data through Google Earth™, and the importing of heterogeneous databases, where the user will map the fields of their databases with the fields of ENVIRAD database. Those subjects will be addressed at a later time.

The inclusion of the GEORAD and SIGLARA data is still pending, waiting for permission.

#### 4. CONCLUSION

Database integration is not an easy task but it is a powerful tool. It gives the user a unified view of all the databases included in the system. It has proved to be an excellent tool for the decision makers and researchers in the radiological monitoring environment. Without the use of an application like ENVIRAD, it is very hard to search in all the databases available in the subject.

In this study, an integrated database system was build based on a methodology that has been proven to give solid results. A web application was created and it is available for all the IAEA Member States that want to join this project. Additional databases that join this project would give more power to the application.

This project is not finished. There are new functionalities that would give the users of Microsoft Excel™ the possibility of exporting their data and importing them in the ENVIRAD application.

#### REFERENCES

- [1] LAURIA, D.C., N.S.F. MARTINS, and R. ZENARO, **Monitoração Radiológica Ambiental**. 1 ed. Vol. 01. 2007, Rio de Janeiro - RJ: Instituto de Radioproteção e Dosimetria - IRD.
- [2] INTERNATIONAL ATOMIC ENERGY AGENCY (IAEA). **The marine information system**. 2016, May 2019-08-16; Available from: <https://maris.iaea.org/Home.aspx>.
- [3] SILVA, T., et al., **GEORAD: Sistema de Gestão de Dados de Radioatividade no Meio Ambiente**, in IX Latin American IRPA Regional Congress on Radiation Protection and Safety - IRPA2013. 2013, Sociedade Brasileira de Proteção Radiológica - SBPR: Rio de Janeiro - RJ - BRAZIL. p. 10.

- [4] M E NASCIMENTO, L., A. C M FERREIRA, and S. GONZALEZ, **Design and development of a geo-referenced database to radionuclides in food**, in 9th Brazilian Congress on Metrology (Metrologia 2017). 2018. p. 6.
- [5] NASCIMENTO, L.M.E.D., **Desenvolvimento de um Banco de Dados Georeferenciado de Radionuclídeos em Alimentos Típicos na América Latina - SIGLARA**, in Radioecologia. 2014, Instituto de Radioproteção e Dosimetria (IRD). p. 111.
- [6] KROGSTIE, J., A. OPDAHL, and S. BRINKKEMPER, **Conceptual Modelling in Information Systems Engineering**. 2010: Springer. 360.
- [7] MATUSZKA, T. and A. KISS. **Geodint: Towards Semantic Web-Based Geographic Data Integration**. 2014. Cham: Springer International Publishing.
- [8] BRAZHNIK, O. and J.F. JONES, **Anatomy of data integration**. Journal of Biomedical Informatics, 2007. 40(3): p. 252-269.
- [9] LENZERINI, M., **Data Integration: A Theoretical Perspective**, in Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, ACM, Editor. 2002: Madison, Wisconsin. p. 233-246.
- [10] RACHMAN, M.A.F. and G.A.P. SAPTAWATI, **Database integration based on combination schema matching approach (case study: Multi-database of district health information system)**, in 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). 2017. p. 430-435.
- [11] ÖZSU, M.T. and P. VALDURIEZ, **Principles of distributed database systems**. 2 ed. 1999: Prentice-Hall, Inc. 657.
- [12] OSMAN, M., A. HEGAZI, and M.O. HEGAZI, **A Conceptual Foundation and Integration Database Designing Model**. Journal of Computer Science, 2014. 10: p. 376-381.

- [13] TEOREY, T.J., et al., **Database Modeling and Design: Logical Design**. 5 ed. 2011: Elsevier Science.
- [14] ÖZSU, M.T. and P. VALDURIEZ, **Principles of Distributed Database Systems**. 3 ed. 2011: Springer Publishing Company, Incorporated. 864.
- [15] ELMASRI, R. and S. NAVATHE, **Fundamentals of Database Systems**. 6 ed. 2010: Addison-Wesley Publishing Company. 1200.
- [16] BERNSTEIN, P., J. MADHAVAN, and E. RAHM, **Generic Schema Matching, Ten Years Later**. PVLDB, 2011. 4: p. 695-701.