CROATICA CHEMICA ACTA CCACAA **70** (1) 435–458 (1997)

ISSN-0011-1643
*CCA-2431*

*Original Scientific Paper*

# Information System on Investigations of the Adriatic Sea*

*Jadranka Pečar-Ilić,*[a] *Ivica Ružić*[b] *and Zoran Skočir*[a]

[a]*Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia*

[b]*Ruđer Bošković Institute, Center for Marine Research, P.O.B. 1016, 10001 Zagreb, Croatia*

Received January 31, 1996; accepted September 2, 1996

Development of the integral referral Information System on Investigations of the Adriatic Sea, with built-in functional and organizational aspects of the performed investigations, is described. According to the selected development method, adapted to the software available CASE (Computed Aided System Engineering) »POSE« & RDBMS (Relational DataBase Management System) »INGRES«, the Information System (IS) has been designed and implemented. The IS design process is described in detail. The system was subdivided into three subsystems: POM (Project Organization Module), EOM (Expedition Organization Module), and SPMM (Sampling and Parameter Measurements Module). The necessary entities and their relationships were defined using the E-R (Entity-Relationship) diagram concept. The integral E-R diagram was transformed into a set of 3NF (Third Normal Form) relation schemes. The SQL (Structured Query Language) command file was generated and started in the INGRES, producing a relational database scheme. The integrity constraints were built into the database by defining 87 SQL procedures and 146 rules for 61 relation schemes. Therefore, the database integrity control could be performed while the data were inserted. The developed application is based on hierarchically organized frames (Menu, Browse and Report), allowing the information flow (»top-down«) between them. The application enables data presentation in space and time under user-defined conditions. The verification of the system was performed using real data from three research projects.

* Dedicated to Marko Branica on the occasion of his 65$^{th}$ birthday.

## INTRODUCTION

Investigations of seas and oceans have a long tradition. They have often been performed within interdisciplinary macroprojects, some of which are internationally coordinated. For this purpose, highly developed and large countries have a whole network of organized services cooperating with a variety of scientific/research institutions and information centers. Smaller and less developed countries conduct these activities mainly through simpler systems. With time, a need has developed to co-ordinate globally such research efforts within international co-operation, where international agencies UNESCO, ICSU, UNEP, WHO, FAO, IOC and WMO play an important role. Constant international exchange of global oceanographic data is based on the IGOOS General Plan. The first phase of the Plan was accepted by the IOC and WMO and was initiated in 1969.[1] Later on, this programme was extended to include the exchange of data obtained by the advanced methods based on the use of satellites, automated equipment installed on special buoys and ships.

For the purpose of storage and elaboration of these global oceanographic data, two world data centers have been established (WDC A within NOAA of the USA and WDC B in the former USSR).[2] Together with other permanent centers for individual disciplines or world regions and a number of National Oceanographic Data Centers (NODC), these two centers collect oceanographic data from Declared National Programmes (DNP), co-operative international expeditions, as well as other oceanographic programmes willing to submit their data. In collaboration with the SCOR and WMO, World Data Centers A and B have produced guides for International Oceanographic Data Exchange (IODE[3]), where standardized formats are described for Reports on Observations and Samples Collected in Oceanographic Programmes (ROSCOP[4]), International Geophysical & Geological Cruise Inventory (IG/GCI[5]), Results On Marine Biological Investigations (ROMBI,[6]) and International Marine Geological Data.[7]

Information on oceanographic investigations enables production of inventories of the available oceanographic data. In order to enable a multipurpose use of such information by different international organizations, the IOC has established the Marine Environmental Data Information referral system (MEDI[8]), which also includes data on the organizational characteristics of potential data sources (similar to INFOTERRA activities organized for the International Referral Center by the UNEP).

Investigations of the Adriatic Sea are mainly carried out by research institutions in Croatia, Italy and Slovenia. However, a significant part of such investigations has also been performed by several research institutions from Austria, France and Germany. Investigations of the Adriatic Sea in Croatia are performed within the scientific activity of four universities, several independent research institutions and some research groups within governmental and nongovernmental institutions.[9] So far, these research activities have not had any adequate information support because NODC, now existing in more than 50 countries in the world, has not yet been established in Croatia. As a result, each individual research organization has developed its own information services according to its needs.

However, some attempts have been made in Croatia to establish information systems for individual projects. The first one was initiated in 1979 for elaboration of data collected within the UNDP project JADRAN II, for which a specific database »ANKETA« was established. For this purpose, SELGEM software was used on the UNIVAC mainframe computer in the Computer Center of the University of Zagreb (SRCE[10–13]) by the Center for Marine Research Zagreb. Later on, in collaboration with UNEP, a bibliographic database with about 3500 records for the period between 1975 and 1990 was established using the UNESCO Library program CDS/ISIS for microcomputers.[14] As part of the UNEP MED POL Phase-II project in 1990, the Center for Marine Research Zagreb produced the first version of a guide for storage of oceanographic data for IBM PC compatible microcomputers (based on dBASE software[15]), which was later adopted and applied for the storage and exchange of data for the sanitary quality of coastal waters within the AlpeAdria project.[16] Data collected for the AlpeAdria project enabled the use of the GIS (Geographical Information System) »MAPINFO« to produce, for the first time, the corresponding thematic maps for the Croatian part of the North Adriatic.

It was only after 1991 that the development of the integral referral Information System on Investigations of the Adriatic Sea, based on international referral activities and guides (MEDI, ROSCOP) as well as on the systems produced so far in the Center for Marine Research Zagreb, was initiated.

In 1992, the object-oriented transformation methodology for relational database design was used to initiate the development of the Oceanographic Information System for the activities of the Institute for Oceanography and Fisheries in Split. Conceptual and logical database models were designed.[17] Later, development of referral database and some specific thematic databases for the Oceanographic Information System (OIS) have been initiated through a joint effort of the Institute of Oceanography and Fisheries and the State Hydrographic Institute of Croatia, also in Split. These were relational databases implemented using RDBMS »ORACLE«.[18] Only a part of the planned very complex OIS has been developed, while some of its segments are still in the development stage, or planned to be developed. From the documentation available about the so-called »Extended model objects-connections of the referral database« only a general framework could be recognized. There is no information how to apply such a general framework. However, it is clear that earlier described conceptual and logical models have not been implemented so far in their original form.[17]

The above mentioned integral referral information system, designed and implemented in the Center for Marine Research Zagreb according to the international guides, integrates functional aspects (definition of all functions relevant to a research project) and organizational aspects of the investigations of the Adriatic Sea (definitions of institutions responsible for individual functions and definition of the role of individual participants in a research project). The information system enables efficient storage and retrieval, as well as presentation of information on investigation results. The data stored in the database of this information system will serve as a catalogue of the research activities on the Adriatic Sea performed to date. The system will also enable spatial and temporal presentation of the existing oceanographic data for different compartments of the Adriatic Sea environment (water, sediment, biota effluents, air *etc.*), as well as presentation of general information about the institutions and researchers responsible for the conditions of the use of these data. This paper presents original solutions to the key problems in the design and implementation phase of the development cycle of this Information System, using modern information technology.

## INFORMATION SYSTEM DEVELOPMENT METHOD

Analysis of the available software has shown that an efficient development of the Information System on Investigations of the Adriatic Sea can be effected by using two optimally connected modern program systems.

POSE (Picture Oriented Software Engineering), developed by the Computer Systems Adviser Inc., is a flexible and modular CASE (Computed Aided System Engineering) system, which is used for the critical phases of the Information System (IS) life (development) cycle.[19]

INGRES (INteractive and Graphic REtrieval System), developed at the Berkeley University in California, is a RDBMS (Relational DataBase Management System[20,21]). In the Center for Marine Research Zagreb, we have installed the INGRES version with character-oriented developing tools for PC486 under the SCO-UNIX operating system.

From the defined approaches for the development of the IS and the availability of software, the appropriate method for the development of the Information System on Investigations of the Adriatic Sea has been defined.[22] This method includes the basic elements of the SSADM (Structured System Analysis and Design Method[23]), which has been adapted with regard to the previously mentioned software available. The development method is primarily based on the »top-down approach«. According to this approach, the conceptual model of the organization's data requirements (graphic representation of the entities and their relationships necessary to describe the investigations of the Adriatic Sea) was formulated first. Then, the conceptual database model was mapped into the logical database model (relational model) and, finally, the relational physical database was designed depending on RDBMS (use of physical options such as partitioning, clustering, organization optimizations and compressed indexes).

Complete development of this IS proceeds per phases of the IS linear life (development) cycle, applying a combination of the »phase and incremental approaches«. Individual steps or processes within the phases are performed many times (incremental approach), but an individual phase is passed only once (phase approach). The final result is a complete model (for example, a logical model of the information subsystem). The »holistic (integral) approach« is applied considering the »relationship between the integral system and its individual components« as a criterion, and the result is an integral system, an integral data model, and a unique database.

The following phases of the IS life (development) cycle can be defined: planning strategy, system and requirements analysis, IS design, and IS implementation.[22] Using the standard procedures, diagramming techniques and models, and applying the so-called »soft approach«, the function and data modelling are performed hand in hand.

## CHARACTERISTICS OF PERFORMING AN OCEANOGRAPHIC RESEARCH PROJECT

We have analyzed the way in which an oceanographic research project is performed and have recognized the following facts:

- Research projects are classified as regional, national, declared national, and international projects, or bilateral and co-ordinating programmes. Many different institutions, such as United Nations agencies, governmental and intergovernmental, non-governmental, professional, educational, industrial/ commercial, private, religious/philosophical, political, and mixed institutions are involved in their sponsorship.
- Projects usually last from one to several years. Participating institutions, such as research, development, military, governmental, medical and educational institutions, are engaged in solving the problems of different areas of interest. Participants in the research (*i.e.* researchers) perform programme defined tasks. Some of the participants are responsible for the research organization and infrastructure maintenance (*i.e.* other persons).
- One or more expeditions are made for research project purposes each year. These expeditions can be classified as:

  - cruises by ships in chosen transects (set of stations located in specific geographic areas, visited according to a defined frequency and time schedule), and
  - standard sampling exits using smaller boats or some other movable supporting devices at individual stations.
- Pursuant to the defined programme, in situ measurements or sampling actions from a specific matrix and depth are carried out by appropriate methods at individual stations. In these cases, specialized stationary or movable supporting devices with appropriate installations are used.
- Samples are collected (except in the cases of *in situ* measurements) using appropriate sampling techniques and methods, according to the specific requirements for individual measurements. Afterwards, samples are treated in an appropriate way, and stored by the responsible institution to be ready for necessary laboratory measurements.
- Measurements of individual parameters or parameter groups are performed by appropriate methods and techniques.
- The measurement results obtained are stored in the records of the responsible institutions and designated database centers.
- The goal of the developed referral IS described in this paper is to provide availability of information on the measurements performed and on the conditions of their use. It contains no individual numerical values of the measurement results.

Based on these facts, individual actions (functions) performed within individual research projects as well as the associated objects are determined. A responsible institution is designated for each action and for the whole project. The elementary actions, together with the associated objects, are presented in Figure 1.

## DEFINITION OF THE INFORMATION SYSTEM STRUCTURE

Analysis of the basic actions, shown in Figure 1, points to the conclusion that the IS structure will be relatively complex. In order to simplify the structure, its function and object decomposition was performed.

Using function decomposition, the system was divided into three relatively independent subsystems (modules) which have a simpler structure than that of the whole system (see Figure 2.). The role of the modules is to achieve the following goals:

1. presentation of the organizational structure of the projects and institutions involved in project sponsorship or research – POM (Project Organization Module)

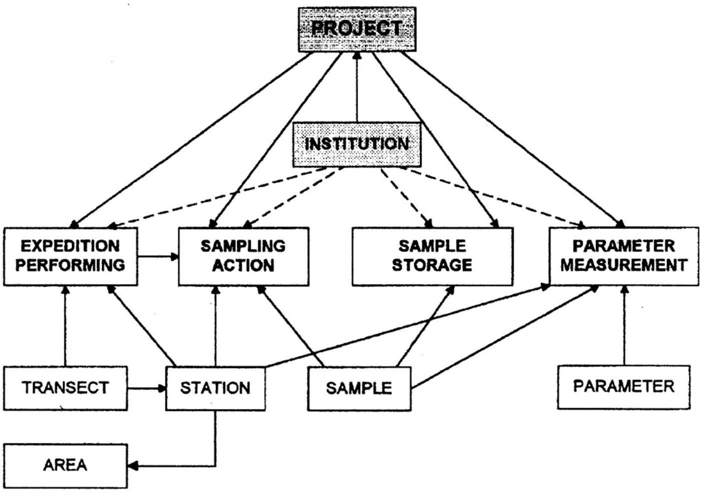


Figure 1. Scheme of elementary actions included in the investigation of marine environment and associated objects necessary for their description (dotted lines represent responsibilities for individual actions).

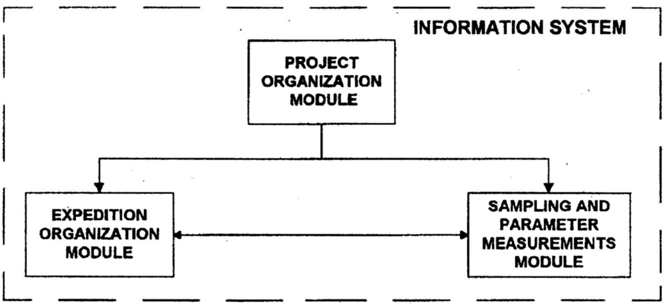


Figure 2. Subsystems of the Information System on Investigations of the Adriatic Sea.

2. presentation of organizational aspects of expeditions made within individual projects – EOM (Expedition Organization Module)
3. presentation of organizational and functional aspects of actions performed, including sampling at the stations, sample storage by responsible institutions, and in situ measurements or parameter measurements with samples – SPMM (Sampling and Parameter Measurements Module)

Using object decomposition, the essential entities and their identifiers are defined. They can be classified into the following entity types:

- GLOBAL_PROJECT – a set of all global projects (national, bilateral, international, *etc.*), which last one or several years and include one or more annual projects. The corresponding identifier is PROJ_INT_CODE.
- PROJECT – a set of annual projects for which all previously mentioned actions are performed (see also Figure 1). The corresponding identifiers are PROJ_INT_CODE and PROJ_YEAR.
- INSTITUTION – a set of institutions participating in the projects, or institutions involved in global project sponsorship. The corresponding identifier is INSTIT_CODE.
- PERSON – a set of all participants (researchers and other persons) involved in annual projects. The corresponding identifier is PERSON_CODE.
- EXPEDITION – a set of all expeditions (cruises and standard sampling exits) performed for annual projects. The corresponding identifier is EXPED_CODE.
- AREA – a set of geographic areas investigated within annual projects. The corresponding identifier is AREA_CODE.
- STATION – a set of stations (standard and specific) located within certain geographic areas. The corresponding identifiers are AREA_CODE and STAT_CODE.
- SAMPLE – a set of samples (standard and specific) which are taken from individual stations using appropriate. methods or techniques. Corresponding identifiers are EXPED_CODE, SAMP_ACTION_CODE and SAMP_CODE.
- MEASURING_PARAM – a set of measured parameters or parameter groups. The corresponding identifier is MEAS_PARAM_CODE.

With these essential entities, all actions of the research project can be described generally. According to the functions of each subsystem (module), other necessary entities and their identifiers are defined. The essential entities and these additional entities for each subsystem are used together for a detailed specification of their structure. The first and the second subsystems (POM and EOM) are connected through the entity type EXPEDITION. This connection is possible since the first subsystem includes the expeditions made within individual annual projects, and the second subsystem presents the organizational aspects of expeditions. Entity types PERSON and INSTITUTION are used by all subsystems. Using these two entities, the responsibilities for the actions performed within individual annual projects are defined. The third subsystem (SPMM) is connected with the first and the second subsystems through the entities PROJECT and EXPEDITION, which enables the optimum system structure design. Actions such as sampling, sample storage, and parameter measurements are performed for a particular expedition within an annual project. The detailed design of their structure using specially defined concepts will be presented later in the paper.

## CONCEPTUAL DATABASE DESIGN OF THE INFORMATION SYSTEM

During the conceptual database design process, the following tasks have been performed:

1. creating of the function decomposition diagram
2. creating of the E-R (Entity-Relationship) diagram
3. checking of the E-R diagram completeness and its validity
4. checking of the E-R and function decomposition diagrams consistency

Due to the complexity of the Information System structure, detailed procedures for the conceptual database design will be presented only for the first subsystem POM (Project Organization Module). This subsystem has been chosen because it has a relatively simple structure but, at the same time, enables to explain the concepts used and the results obtained for each component of the design process.

Two elementary functions that enable the first subsystem (POM) to perform its principal goal and specified tasks are defined. The first function is informing on the organizational structure of the global project. It also includes components such as informing on the organizational structure of an annual project, identification of annual project participants, identification of the researcher responsible for the annual project, and identification of expeditions performed for the annual project. The second function is informing on the organizational structure of the institution. It also includes components such as the institution structure identification (corporation, institute, and department) and the institution data inventory (general information and activities performed).

These functions and their corresponding components are inserted into the POSE-DCD (DeComposition Diagrammer[24]) module, and the initial function decomposition diagram is created. According to the defined tasks of the first subsystem (POM), other functions of lower hierarchy levels are defined using functional decomposition. The final result was the function decomposition diagram for the whole first subsystem (POM), presented in Figure 3. According to previously defined functions for the POM, the necessary entities (some of the essential entities together with the specific entities defined for the first subsystem) and the relationships between them are presented in the E-R diagram, shown in Figure 4. The corresponding E-R diagrams for the other two subsystems are produced in a similar way, using POSE-DMD (Data Model Diagrammer[25]) module.[22]

During the design process of the E-R and function decomposition diagrams, the necessary checking of their completeness, validity and consistency is performed. Validity checking includes removal of not easily recognizable synonyms and incorporation of user requirements. Verification of the defined functions and definition of the required new functions are carried out using the cor-

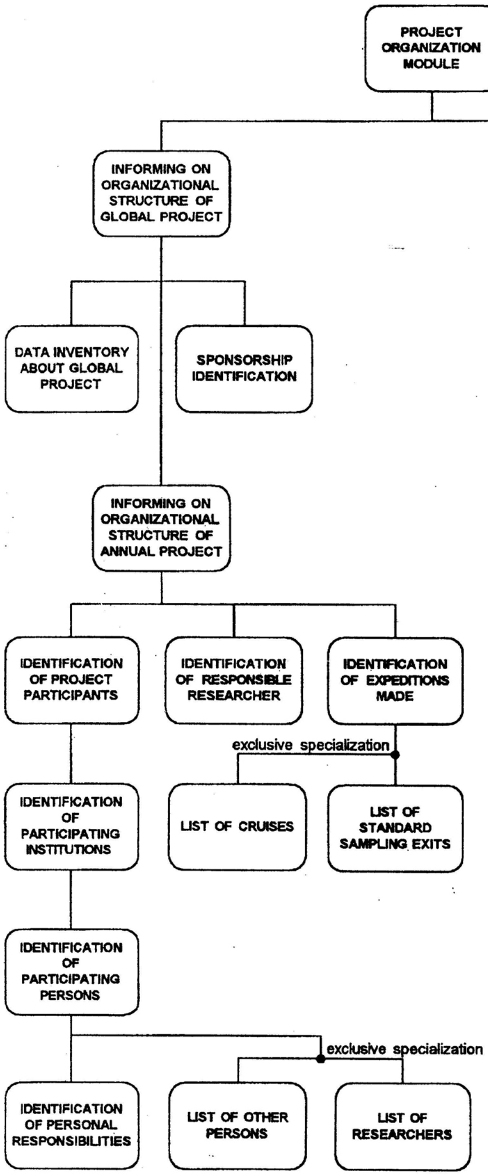


Figure 3. Functional decomposition diagram, showing functional breakdowns for the Project Organization Module, obtained by the POSE-DCD module.

responding data flow diagrams, obtained with the POSE-DFD (Data Flow Diagrammer[26]) module. Establishment of the entities defined for individual

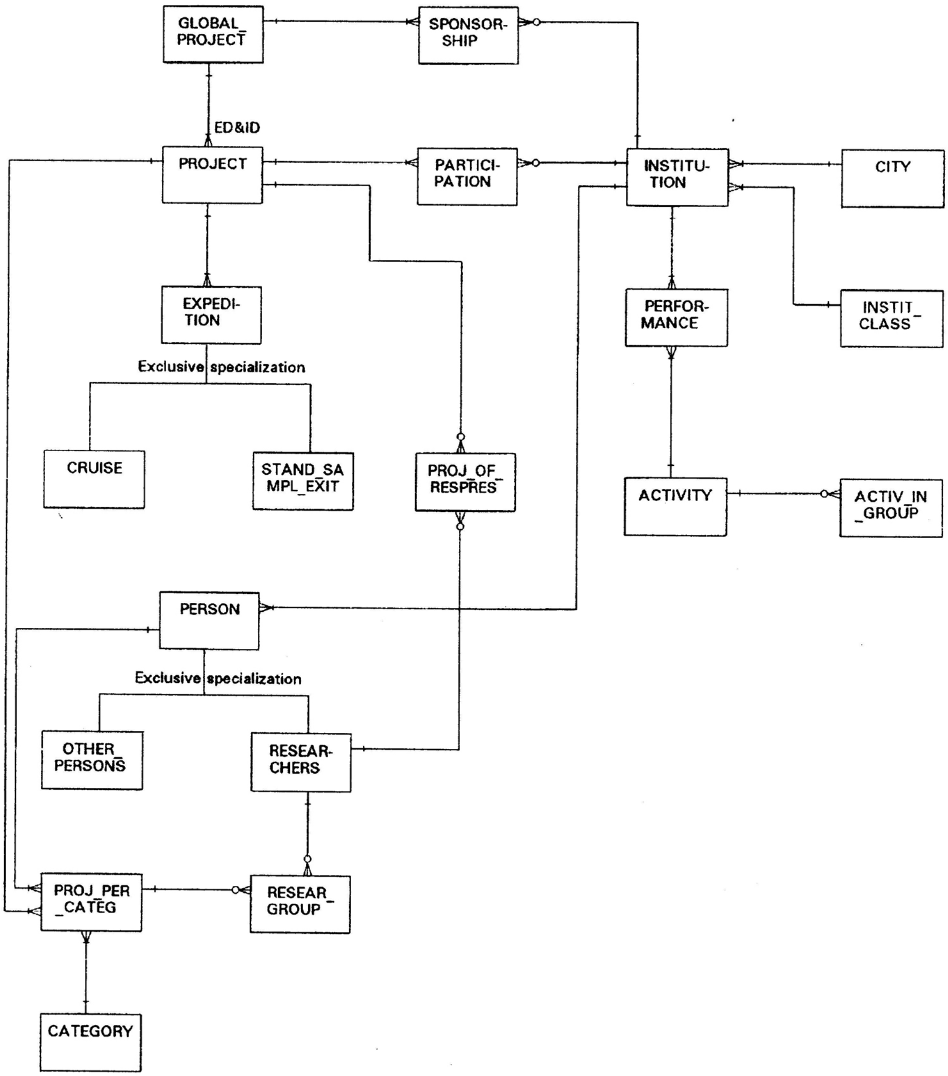


Figure 4. Entity – Relationship Diagram for the Project Organization Module, obtained by the POSE-DMD tool using Martin's notation.

functions is done using the corresponding matrix diagrams obtained with the POSE-PMD (Planning Matrix Diagrammer[27]) module.

As the final result of the conceptual database design process, the integral E-R diagram for the whole system was obtained from the individual E-R diagrams for all three subsystems.[22]

### *Entity and Relationship Types from the First Subsystem (POM)*

There are several entity and relationship types presented in the E-R diagram, shown in Figure 4. The first is the elementary entity type, which can be identified uniquely and exists alone in the data model. Examples of such entities for the first subsystem (POM) are: GLOBAL_PROJECT, INSTITUTION, EXPEDITION, PERSON, *etc.* The second entity type is the »weak« entity type, which can have certain dependencies on other entities. The corresponding relationship types are specified using the following notation in the E-R diagram:[28]

- ED – »Existence-Dependent« relationship (existence of a weak entity depends on the existence of other entities)
- ID – »Identification Dependency« on other entities (a weak entity cannot be identified uniquely by its own attributes, and has to be identified by its relationships to other entities)
- ED&ID – represents both dependencies (most ID dependencies are associated to existence dependencies, although existence dependency does not imply ID dependency)

The entity PROJECT is an example of the weak entity type since it depends (ED&ID) on the entity GLOBAL_PROJECT. It is identified uniquely by the composite key consisting of the global project code and its own internal *year* code.

The third entity type is a composite (intersection) entity formed by other entities (elementary entities or other composite entities). According to Martin's notation,[23,25] relationships with the cardinalities *one to one* (1:1) and *one to many* (1:M) can be drawn immediately on the E-R diagram whereas those with *many to many* (M:N) relationships cannot. We have, therefore, created composite entities for these relationships. On the E-R diagram shown in Figure 4, composite entities have been created to represent all binary relationships with the cardinalities of M:N and one ternary relationship. Figure 5 presents the way in which the composite entity PROJ_PER_CATEG (representing the ternary relationship between the entities PROJECT, PERSON, and CATEGORY) has been created. By detailed analysis it can be shown that the above mentioned ternary relationship cannot be exactly reproduced in terms of two binary relationships between individual pairs of entities participating in the ternary relationship (PROJECT, PERSON and PERSON, CATEGORY).

There are a few cases of an entity type needing additional subgrouping of its entities (producing special hierarchy structures) in order to define more precisely their properties and relationships.[23] The first case is the so-called exclusive specialization, in which the subgruping of the superclass entity is performed through a single subclass and according to the value of only one of its attributes. An example of such structure is the set of subclasses RESEARCHERS, OTHER_PERSONS. It is a specialization of the superclass entity PERSON based on its attribute *per_subtype* (an example from the E-R diagram shown in Figure 4.). This structure implies that each person on the project can be classified as a researcher or as other person, but not as both. The second type is nonexclusive specialization. In this structure, the superclass entity includes attributes common to all subclasses, although each subclass has also its own specific attributes. An example of such structure is the set of subclasses BIOTA_SAMPLING, PLANKTON_SAMPL, SEDIMENT_SAMPL, TAR_BALL_SAMPL, AIRBOR_POL_SAMPL. This is a specialization of the superclass entity SAMPLING_ACTION based on the *action_ty_code* attribute (an example from the E-R diagram for the third subsystem, not shown in this paper). The sampling action is described with common attributes, and the above mentioned actions are also described with their additional and specific attributes (for example BIOTA_SAMPLING is additionally described with the attributes *tissue* and *species*).

### *Entity Types from the Second and Third Subsystems (EOM & SPMM)*

Since the relationships between the objects in the second and third subsystems are very complex, it is necessary to avoid the complexity of primary keys, generated during the creation of such relation schemes. To achieve stability of the relational data model, we introduced, as a compromise solution, the use of composite entities, to which we assigned keys. On the example

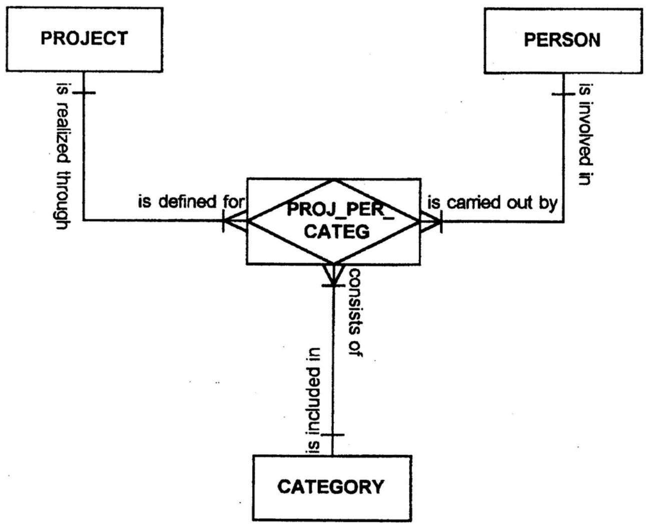


Figure 5. Illustration of the ternary relationship represented by the composite entity PROJ_PER_CATEG, describing the responsibilities of the project participants.

of composite entities STORAGE and SAMPLING_ACTION, we will describe the procedure used to assign the corresponding keys to them. The entities are defined to replace binary and ternary relationships:

1. STORAGE ⇔ rel. between PROJECT and SAMPLE
2. SAMPLING_ACTION ⇔ rel. between EXPEDITION, STATION, and ACTION_TYPE

In these relationships, weak entity types are: PROJECT, STATION, and SAMPLE, which are ED&ID dependent on the entities: GLOBAL_PROJECT, AREA, and SAMPLING_ACTION. Using the standard procedure, these relationships would be replaced by composite entities, which would have composite keys consisting of several key attributes. When such an E-R diagram is transformed into a relational data model, the corresponding relation schemes with complex composite keys will be produced. The primary key of the relation scheme is presented upper case and italic, and a foreign key is presented lower case and italic. Relation schemes are presented in the following way:

1. STORAGE (*EXPED_CODE, AREA_CODE, STAT_CODE, ACTION_TY_CODE, PROJ_INT_CODE, PROJ_YEAR, treat_code, ...*)
2. SAMPLING_ACTION (*EXPED_CODE, AREA_CODE, STAT_CODE, ACTION_TY_CODE*, sampling_date, sampling_remark, ...)

This solution was not satisfactory for providing the stability of the relational model. Therefore, in the next step, we returned to the composite entities STORAGE and SAMPLING_ACTION, and assigned to them their own primary keys *STORAGE_NUMBER* and *SAMP_ACTION_CODE*. In this way, we reduced the number of entities on which they depended. Now, the produced relation schemes have the following form:

1. STORAGE (*STORAGE_NUMBER, PROJ_INT_CODE, PROJ_YEAR, ...*)
2. SAMPLING_ACTION (*SAMP_ACTION_CODE, EXPED_CODE, ...*)

The first relation scheme indicates that sample storage can be identified uniquely with a *storage number* and *annual project* composite key consisting of the *global project* and *year* codes. The second relation scheme shows that the sampling action can be identified uniquely with *action* and *expedition* codes.

## LOGICAL DATABASE DESIGN OF THE INFORMATION SYSTEM

Based on the integral E-R diagram, the logical scheme of the complete database of our IS has been produced. The COMPOSITE LOGICAL DATABASE DESIGN process, producing the logical database scheme (relational database scheme), has been applied. This process includes the following steps:

1. mapping of the E-R diagram into the relational database scheme, consisting of a set of individual relation schemes, using the POSE-DMN (Data Model Normalizer[29]) module;
2. normalization of the relation schemes until they reach 3NF (Third Normal Form[30]), using the POSE-DMN module;
3. generating the SQL/DS (Structured Query Language/Data System) command file for the database structure definition, using the POSE-DBA (Data Base Aid[31]) module into which the normalized E-R diagram has been inserted;
4. adapting of the SQL/DS file to the SQL formats[21] producing the command file, which contains the DLL (Data Definition Language) commands for the relational database scheme and index definition

Normalization of relation schemes is the most important procedure performed during the composite logical database design process. Unsatisfactory relation schemes, between the attributes of which certain dependencies exist, are decomposed by breaking up their attributes into smaller relation schemes with desirable properties.[32] By normalization we ensure a »good« design of relation schemes (normal forms as criteria[30]) by disallowing modification anomalies (problems related to the adding, deleting, or updating of records). The normalized relation schemes are compared to the E-R diagram, which is afterwards modified by adding new entity types (composite entity types) or by removing exceeding relationships (transitive dependencies). The normalization procedure was repeated until full agreement between the E-R diagram and the normalized relation schemes was reached. Finally, a nonredundant relational conceptual database model for the whole system, consisting of 61 relation schemes in 3NF, was generated.[22] As an example, the relation schemes generated for the first subsystem (POM) are presented in Table I, according to standardized notation.[29]

In the POSE-DBA module, the normalized integral E-R diagram is automatically transformed into the SQL/DS command file, containing commands for creating individual tables from the corresponding relation schemes and for the indexes for primary keys of these relation schemes. The final result of the IS design process was the SQL command file, obtained from the SQL/DS command file after it was adapted to SQL formats.[22]

## DATABASE IMPLEMENTATION OF THE INFORMATION SYSTEM

In the second part of the IS developing process, the relational conceptual database model of the Information System on Investigations of the Adriatic Sea was implemented as a centralized database. In the IQL (Interactive Query Language[20]) subsystem of INGRES, the SQL command file was started, and the corresponding tables and indexes were created.[22] As opti-

TABLE I

3NF Relation Schemes for the Project Organization Module

| | RELATIONS | ATTRIBUTES |
|---|---|---|
| 1 | ACTIV_IN_GROUP | =(ACTIV_IN_GR_C, ***act_code***, activ_in_gr_n, gr_ac_rem) |
| 2 | ACTIVITY | =(ACT_CODE, act_name, act_type, act_rem) |
| 3 | CATEGORY | =(CATEGORY_CODE, category_name) |
| 4 | CITY | =(CITY_CODE, city_denom, country_denom, zipcode, remark) |
| 5 | CRUISE | =(EXPED_CODE, cruise_remark, ***ps_code***) |
| 6 | EXPEDITION | =(EXPED_CODE, exped_name, exp_subtype, exped_remark, ***proj_year***, ***person_code***, ***proj_int_code***) |
| 7 | GLOBAL_PROJECT | =(PROJ_INT_CODE, proj_orig_name, proj_type, proj_start_date, proj_end_date, proj_global_rem) |
| 8 | INSTIT_CLASS | =(CLASS_CODE, instit_class, class_remark) |
| 9 | INSTITUTION | =(INSTIT_CODE, ***city_code***, ***class_code***, instit_address, instit_corporate, instit_denom, instit_type, instit_rem, ***person_code***) |
| 10 | OTHER_PERSONS | =(PERSON_CODE, person_title, person_tel, person_fax, person_email) |
| 11 | PARTICIPATION | =(INSTIT_CODE, PROJ_INT_CODE, PROJ_YEAR, partic_rem) |
| 12 | PERFORMANCE | =(ACT_CODE, INSTIT_CODE, perfor_rem) |
| 13 | PERSON | =(PERSON_CODE, instit_code, person_name, per_subtype, person_remark) |
| 14 | PROJ_OF_RESPRES | =(PERSON_CODE, PROJ_INT_CODE, PROJ_YEAR, person_proj_rem) |
| 15 | PROJ_PER_CATEG | =(PROJ_INT_CODE, PROJ_YEAR, PERSON_CODE, CATEGORY_CODE, per_cat_remark) |
| 16 | PROJECT | =(PROJ_INT_CODE, PROJ_YEAR, proj_remark) |
| 17 | RESEAR_GROUP | =(GROUP_CODE, resear_codes, group_remark, ***proj_int_code***, ***proj_year***, ***person_code***, ***category_code*** |
| 18 | RESEARCHERS | =(PERSON_CODE, resear_prof, resear_title, resear_remark) |
| 19 | SPONSORSHIP | =(PROJ_INT_CODE, INSTIT_CODE, proj_signific, spon_rem) |
| 20 | STAND_SAMPL_EXIT | =(EXPED_CODE, ***ps_code***, stand_exit_rem) |

mum table organization, the *unique btree* organization was chosen. This is the table in which duplicate rows and *null value* for its unique primary key are not permitted. Since the *unique btree* table organization was selected, two structural rules, the uniqueness of key and the entity constraint, were automatically provided. The third structural rule is the referential constraint generally applied to relationships. According to this constraint, a relationship can exist only if the participating entity instances have already been recorded in the corresponding relations.[33] Since the referential constraint is not automatically provided by INGRES, we had to perform additionally the following actions:

1. building-in of the integrity constraints as additional knowledge in the database
2. building-in of a set of rules necessary to control this knowledge

These actions were performed primarily to ensure the database state, all data in the database at a particular point in time, to be consistent, which is the case when all state constraints are satisfied. To perform these actions, the *procedures* and *rules* concepts were used.[34] The database procedures are functional groups of statements managed as objects by INGRES, as a part of database definition. Such statements normally include one or more queries to the database. A rule is a user-defined mechanism that invokes a specific database procedure or procedures. Rules and procedures are both stored in the database server, and are invoked automatically when database changes in a specific way. For this reason, they provide database integrity control, so that no integrity control is necessary in applications. We defined 87 procedures and 146 rules for all 61 tables in the database to perform integrity control of the data while inserted into the database.[22] If certain data violated the defined constraint (the domain, referential or general semantic integrity constraint), they were rejected. The user is then informed, through the message on the screen, about the integrity constraint which has been violated.

## APPLICATION DEVELOPMENT AND SYSTEM VERIFICATION

The application development process proceeded gradually. Using the *application flow diagram* and *visual queries* concepts, the hierarchical data presentation, from the most general to more detailed one, was provided. Hierarchically organized frames (Forms, Reports, and Menus) were obtained. Information flow (»top-down«) was established between the frames. Previously defined functions and specified user requirements were implemented in these frames using the visual queries concept.[22] Visual queries are graphical presentations based on master/detail relationships defined between the tables or views from the database.[20] Columns of the same type from individual tables or views are joined, and the corresponding data are produced.

*Parameter Measurement* application frames are shown in Figure 6. They can be described as follows:

- proj_list (General Information about the Global and Annual Projects) presents general information about the projects and participating institutions;

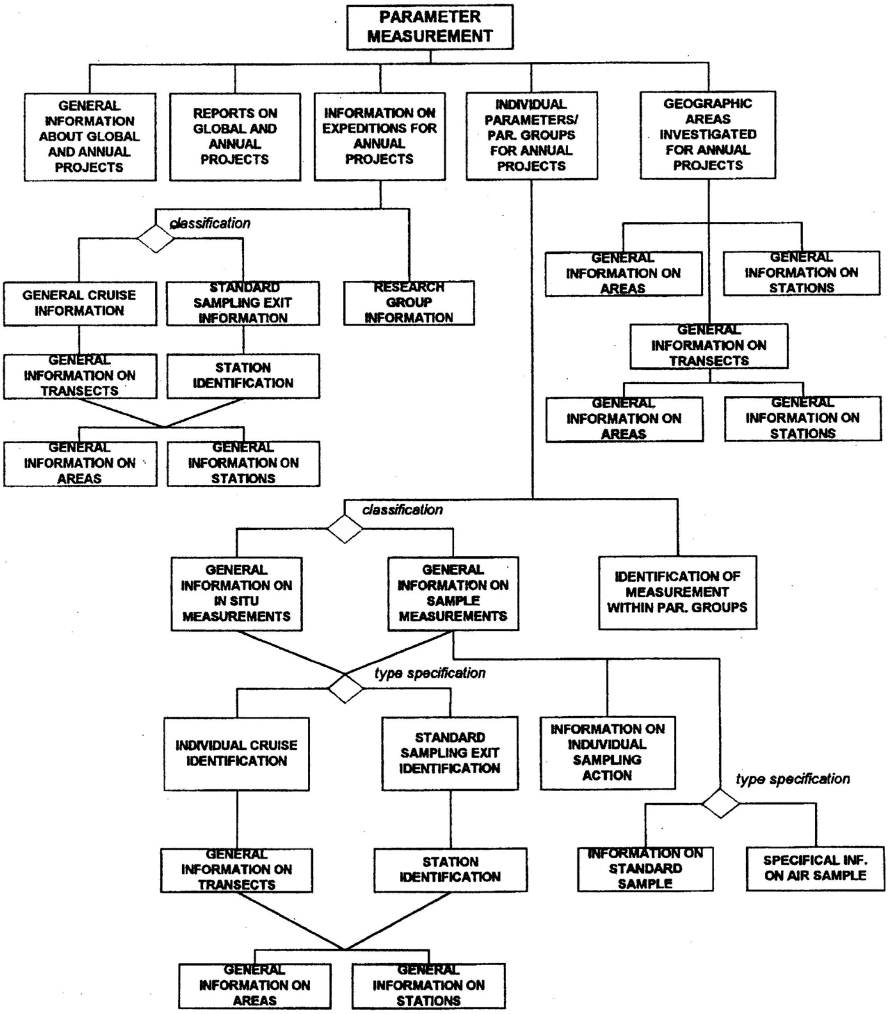


Figure 6. Application flow diagram for the Parameter Measurement procedure, generated by the INGRES ABF (Application By Forms) subsystem.

- proj_report (Reports on Global and Annual Projects) has the same functions as those mentioned for the above frame but can also provide data storage and printing of reports;
- exped_list (Information on Expeditions for Annual Projects) presents the following:

– organizational aspects of the expeditions performed for an individual project (expedition name, starting date, responsible researcher, research group and institution, *etc.*);

– classification of cruises (information about the ships, installations, transsects used, stations within the respective areas visited, *etc.*) and standard sampling exits (information about the stations within the respective areas visited);

- param_list (Individual Parameters/Par. Groups for Annual Projects) presents the following:

– information on parameter measurements for an individual project, including the date of measurement, responsible institution (researcher), classification of measurements (*in situ* or measurements with samples), *etc.*;

– data retrieval by specified parameter and/or, if necessary, the corresponding parameter groups;

– specification of expeditions and stations and presentation of relevant data (also including data produced with the exped_list frame) for *in situ* measurements;

– specification of expeditions and sampling actions, and presentation of data including the matrix, measurement method, classification into standard and air samples, *etc.* (also including data produced with the exped_list frame) in the case of measurements with samples;

– additional information on specified sampling action (responsible institution and researcher as well as sampling method);

– additional information on the specified sample for which measurements are performed (for example, depth from which the standard sample was taken);

- area_list (Geographic Areas Investigated for Annual Projects) presents all investigated areas, transects, and stations visited. It enables data presentation for an individual project and expedition.

Data about areas and stations can be retrieved in several different ways. Retrieving an AREA table is performed by its *area_code* key column or by its WMO (World Meteorological Code) columns. Retrieving a STATION table is performed in the same way as described for the AREA table, and additionally by the geographical co-ordinate columns.

According to the description of application frames, this application primarily provides answers to the following basic questions:

- WHAT – relating to various types of data stored in the tables of the database;
- WHERE and WHEN – relating to the data collected from a specified location at a specified time interval;
- WHO – relating to the institutions (researchers) responsible for the project and individual actions performed, and for the data stored in the database.

**REVIEW OF THE MEASUREMENTS MADE WITH THE SAMPLE** Browse Frame

Measur. Subtype : MEAS_WITH_SAMPLE

Proj. Code : *ALA*
Proj. Year : *1992*

Measur. Number : *3*

| Parameter Type | Parameter Code | Parameter Name | Method | Measur. Date | Person Code |
|---|---|---|---|---|---|
| *INDIVIDUAL* | *FS* | *Faecal streptococci* | *OTH* | | *ALA90PU000001* |

**INFORMATION ON THE EXPEDITION, SAMPLING ACTION, MATRIX, STATION AND SAMPLE**

Expedition Code : *ALA1*
Sampling Action Code : *10*
Resp. resear. Code : *ALA90PU000001*
Sampling Date : *29-jul-1992*

Matrix Code : *SW*
Area Code : *PO*
Station Code: *01*

Sample Code : *1*

Sample Subtype : *COMMON_SAMPLE*

**exped_type(F4) samp_act_info(F5) sample_type(F6) Next(F9) >**

**Help(F1) End(F10)**

Figure 7. An example of the form containing information obtained during the *proj_list* application frame execution.

Therefore, this application enables data presentation in space and time with built-in organizational aspects of investigations under user defined conditions. With this application, the verification of the Information System was performed.[22] An example of the GENERAL INFORMATION ON SAMPLE MEASUREMENTS screen, generated through a set of data presentation forms, is presented in Figure 7. The screen was produced when the *param_list* option from the main menu was selected (corresponding to INDIVIDUAL PARAMETERS/PAR. GROUPS FOR ANNUAL PROJECT, see also Figure 6.).

## DISCUSSION

This paper describes the design and implementation process of the integral referral Information System on Investigations of the Adriatic Sea.

In the first phase, the problems relating to the efficient use and optimal mutual connection between two different software products are solved to provide a complete and successful IS development process. POSE was used for the IS design and INGRES for IS implementation. Most of the effort was put into a detailed description of the IS design. The IS implementation process is described and some of the final results are presented. Examples of the SQL command file and of the procedures and rules are not presented in this paper since they represent a set of standard SQL commands whose formats are well known (*e.g.* table, index, procedure and rule definition).

Detailed results obtained during the IS design process are presented for the first subsystem (POM). In addition, interesting solutions for the key problems that appeared during the IS design process for the other two subsystems (EOM & SPMM) are presented.

Very complex relationships between the objects defined to describe the execution of a research project were successfully mapped into the semantically rich E-R model. Necessary entity types and structures and the corresponding relationships between them were defined. The problem appeared when the integral E-R model for the whole IS was transformed into a strictly mathematical formalism, *i.e.* relational model. The relational model is semantically weaker than the E-R model. The solution was found in using the procedure and rule definitions to increase relational model semantics. In such a way, the structural and semantic integrity were built into the database as additional knowledge. RDBMS INGRES controls this knowledge through a defined set of rules. As the final result, the implemented database has minimal redundancy, a stable structure and data consistency. The application was developed that enables efficient presentation of the collected oceanographic data for a given time and location. Verification of the Information System was performed using the data from three research projects:

MED POL Phase II-International Programme, ASCOP- Adriatic Scientific Co-operative Programme, and AlpeAdria-Regional Programme.

Future development of this IS will be continued, with the main goal of increasing its performances. For this purpose, the graphic environment (X-windows as a generally accepted norm), menus (more user-friendly), and programs enabling spatial and descriptive data imports will be developed. These goals could be achieved by using new software products presented in summer 1995 at the INGRES World Conference (organized within the Computer Associates manifestation CA-World'95 in New Orleans). These new CA-OpenIngres products could satisfy these requirements. The CA product OME (Object Management Extension[35]), which enables completely new data type definitions (including, for example, images, maps, sound and video stored in relational tables), could be of special interest. These new data types can be accessed using SQL statements just like any other data types.

*Acknowledgment.* – The authors gratefully acknowledge the support of the National Science Fund of Croatia.

## REFERENCES

1. Intergovernmental Oceanographic Commission, World Meteorological Organization, *General Plan and Implementation Programme 1977–1982*, Technical Series 16, WMO-No.466, January 1977.
2. International Council of Scientific Unions – Panel on World Data Centers (Geophysical and Solar), *Fourth Consolidated Guide to International Data Exchange through the World Data Centers*, Washington, June 1979.
3. Intergovernmental Oceanographic Commission, *Manual on International Oceanographic Data Exchange (Fourth Edition)*, Manuals and Guides 9., UNESCO 1975.
4. Intergovernmental Oceanographic Commission, *Reports of Observations / Samples Collected by Oceanographic Programmes (ROSCOP)*, UNESCO, March 1974.
5. NOAA Form 24–18 (6–73), *International Geological / Geophysical Cruise Inventory (IG / GCI)*, Revised Edition, June 1973.
6. Intergovernmental Oceanographic Commission, *Results on Marine Biological Investigations (ROMBI)*, NOAA Form 24–31(9–75)
7. International Marine Geological Data Format, *A. Description of Unconsolidated Sediment and B. Descriptions of Hard Rocks*
8. Intergovernmental Oceanographic Commission, *MEDI, Marine Environmental Data Information Retrieval System*, Input Registration form – Format, March 1979.
9. Ministry for environmental protection, physical planning and housing, *National report on environment and development*, Croatia, Zagreb, May 1992 (in Croatian).
10. *MAP-UNEP*, Report of the consultation meeting on data processing, Athens, 1988.
11. J. Cushing and I. Ružić, *Report on the computerized elaboration of a survey about the coastal regions of Croatia*, Center for Marine Research, Ruđer Bošković Institute, Zagreb, December 1979 (in Croatian).

12. J. Cushing and I. Ružić, *Final report on computer processing of Anketa about Coastal Regions of Croatia*, Center for Marine Research, Ruđer Bošković Institute, Zagreb, December 1979 (in Croatian).
13. I. Ružić, *Description of a system designed for storage of oceanographic data collected in the investigation of the Adriatic Sea*, Center for Marine Research, Ruđer Bošković Institute, Zagreb, June 1982.
14. Z. Konrad, I. Ružić, *Bibliography on the Adriatic marine environment*, DRAFT, MED POL, United Nations Environment Programme, Center for Marine Research, Ruđer Bošković Institute, Zagreb & UNEP, Athens, 1993.
15. I. Ružić, *A manual for entering data collected within the framework of the national monitoring programme of Yugoslavia (MED POL) through IBM PC compatibles and storage on flexible 5.25" diskettes (Version 1.0.)*, Center for Marine Research, Ruđer Bošković Institute, Zagreb & UNEP, Athens, 1990.
16. J. Pečar, I. Ružić, *A manual for entering data collected within the framework of AlpeAdria Programme on IBM PC compatible microcomputers (Version 1.0.)*, Center for Marine Research, Ruđer Bošković Institute, Zagreb, August 1991 (in Croatian).
17. M. Glavurtić, *Oceanographic Database Design*, M.Sc. Thesis, Faculty of Electrical Engineering and Computing, University of Zagreb, 1992 (in Croatian).
18. V. Dadić, M. Srdelić, A. Grubišić, *Oceanographic Information System and Databases of the Adriatic Sea,* 17$^{th}$ International Conference ITI'95, Pula, June 13–16, 1995, Croatia, pp. 289–294.
19. *POSE Overview*, Computer Systems Advisers, Inc., N. J., U. S. A., 1990.
20. *Introducing INGRES for the UNIX and VMS Operating Systems*, Relational Tehnology Inc., 1989.
21. *INGRES/SQL Reference Manual for the UNIX and VMS Operating Systems*, Ingres Corporation, 1991.
22. J. Pečar-Ilić, *Information System on Investigations of the Adriatic Sea*, M.Sc. Thesis, Faculty of Electrical Engineering and Computing, University of Zagreb, May 26., 1995 (in Croatian).
23. V. Strahonja, M. Varga, M. Pavlić, *Basic Methodology of Information System Development*, in Information System Design (Methodological Guide), Croatian Agency for Informatic & INA-INFO, Zagreb, 1992 (in Croatian).
24. *POSE Reference Guide for Decomposition Diagrammer*, Computer System Advisers, Research Pte Ltd., 1990.
25. *POSE Reference Guide for Data Model Diagrammer*, Computer System Advisers, Research Pte Ltd., 1990.
26. *POSE Reference Guide for Data Flow Diagrammer*, Computer System Advisers, Research Pte Ltd., 1990.
27. *POSE Reference Guide for Planning Matrix Diagrammer*, Computer System Advisers, Research Pte Ltd., 1989.
28. Peter P.S. Chen, *Database Design Based on Entity and Relationship* in S. Bing Yao (Ed.), *Principles of Database Design*, Volume I: Logical Organizations, 1985.
29. *POSE Reference Guide for Data Model Normalizer*, Computer System Advisers, Research Pte Ltd., 1990–91.
30. D. Maier, *The Theory of Relational DataBases*, Computer Science Press, 1985.
31. *POSE Reference Guide for Data Base Aid*, Computer System Advisers, Research Pte Ltd., 1990–91.

32. Z. Skočir, *Information Analysis Proceeding and Formal Description of Telecommunication Network*, ITA 9(69–114), Zagreb, 1990 (in Croatian).
33. J. Pečar, *Relational Database Integrity*, B. Sc., Department of Telecommunications, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, 1990.
34. Z. Skočir, *Further Investigation and Development of Information System for Telecommunications Network*, ITA, 13, 329–336, 1994.
35. *The Real World Solution For Database Processing*, CA-OpenIngres, Computer Associates International, Inc., 1995.

## SAŽETAK

### Informacijski sustav o istraživanju Jadranskog mora

*Jadranka Pečar-Ilić, Ivica Ružić i Zoran Skočir*

Opisan je razvoj integralnog referalnog Informacijskog sustava o istraživanju Jadranskog mora s ugrađenim funkcijskim i organizacijskim aspektima provedenih istraživanja. Prema odabranoj metodi razvoja, prilagođenoj s obzirom na programske sustave CASE (Computed Aided System Engineering) »POSE« i RDBMS (Relational DataBase Management System) »INGRES« izvršeno je projektiranje i implementacija Informacijskog Sustava (IS). Detaljno je opisano projektiranje IS. Izvršena je dekompozicija sustava na tri podsustava: POM (Project Organization Module), EOM (Expedition Organization Module) i SPMM (Sampling and Parameter Measurements Module). Definirani su potrebni entiteti i njihovi odnosi koristeći dijagrame entiteti-veze. Integralni dijagram entiteti-veze pretvoren je u skup 3NF (Third Normal Form) relacijskih shema. Generirana je komandna datoteka SQL (Structured Query Language), a u sustavu INGRES je njezinim pokretanjem kreirana struktura baze podataka (relacijske sheme i indeksi). U bazu podataka ugrađen je osnovni integritet definirajući 87 SQL procedura i 146 SQL pravila nad 61-nom relacijskom shemom. Na taj način, omogućena je kontrola integriteta podataka unesenih u bazu. Razvijena aplikacija zasniva se na hijerarhijski organiziranim okvirima (Menu, Browse i Report) između kojih je omogućen slijed informacija (»top-down«). Aplikacija omogućuje prostorni i vremenski prikaz podataka uz definirane korisničke uvjete. Verifikacija sustava izvršena je s podacima prikupljenim u tri istraživačka projekta.