HYDROSCOPE: experience from a distributed database system for hydrometeorological data

A. Sakellariou,[a] D. Koutsoyiannis,[a] D. Tolikas[b]
[a] *Division of Water Resources, Hydraulics and Maritime Engineering, National Technical University of Athens, 5 Iroon Polytechniou, GR-15700 Zografou, Greece*
[b] *Department of Hydraulics and Environmental Engineering, Aristotle University of Thessaloniki, Thessaloniki, Macedonia, Greece*
University of Thessaloniki, Thessaloniki, Greece

Abstract

HYDROSCOPE is a Greek nation-wide research programme with 14 participating organisations aiming at the development of a national distributed database system for meteorological, hydrological and hydrogeological information. Available modern computer technologies such as powerful workstations, high speed wide-area networks, relational distributed database management systems, client-server architecture, graphical user interfaces and fourth generation programming languages have been used throughout the project. In parallel, a human network consisting of scientists and engineers from all participating organisations, universities, ministries, research centres and public agencies, has been set up. In this paper we present the two-year experience from the project, both at technical and human level, the achievements made and the problems encountered, as well as the perspectives up to the year 2000.

1. Introduction

Over the last decade, resource handling issues and environmental concerns begun to play an increasingly important role in every aspect of human activities. Thus, availability of reliable information and processing tools appeared to be a mandating necessity for water resources management, environmental impact assessment and policy formulation in the public and private sector.

Advances in both hardware and software technologies allow the development of modern information management systems. Relatively new concepts such as relational distributed database model, client-server architecture, visualisation and windowed environment, multi-platform development, object-oriented programming and expert system integration, attempt to satisfy today's information management requirements.

HYDROSCOPE employs a mix of the above mentioned technologies, in order to come up with a solid and viable answer to the modernisation of the hydrometeorological information management in Greece. It also addresses many administrative idiosyncrasies, inherent in the public sector in this country.

2. System Description

HYDROSCOPE is a joint research project aiming at the development of a national databank for hydrometeorological information (Tolikas et al.[3]). Among the partners in this project are ministries, public agencies, research centres and universities (Table 1). The National Technical University of Athens, Division of Water Resources, is the main partner responsible for the co-ordination and monitoring of the project, as well as for the system design.

Technically speaking, HYDROSCOPE is a distributed database system based on a high-speed wide area network (WAN; see Figure 1). The network includes at the moment one node for every partner in the project. Every node is consisted from a host computer (typically a Hewlett-Packard 9000/7XX workstation) acting as database server, and a local area network linking several MS-Windows client PCs and optionally, additional workstations and peripherals. The logical union of these local databases with the aid of specific mechanisms, form the distributed system. We use digital data lines and routers to link all the sites (i.e. nodes) and provide global connectivity.

An important goal also, was to create a human resources co-operation network within the public sector, to promote scientific research and operational efficiency while keeping up with the technological advancements. It is more than obvious that even the most advanced and efficient systems are useless if there is no provision for user acceptance. This can be especially true in technology-suspicious environments commonly found in this type of organisations.

At the time of this writing, Phase I of the project has just been completed. During this first phase, all the necessary infrastructure has been established. The equipment has been put in service and all necessary client/server applications have been developed and tested. In the forthcoming Phase II, most of the available data will be entered to the system. In addition, there are plans for further improvements, which will facilitate the multi-level access of the database, integrate artificial intelligence and further enrich the set of available processing tools.

3. Administrative aspects

3.1 Partnership issues

Profile The participating organisations fall to one of the following three categories as indicated in Table 1.

Table 1: Partners in HYDROSCOPE

Universities, Academic Institutions	Ministries, Public Agencies	Research Centres
National Technical University of Athens (NTUA), Dept. of Water Resources, Hydraulic and Maritime Engineering Aristotle University of Thessaloniki, Dept. of Hydraulic and Environmental Eng. (AUT/DHEE) Aristotle University of Thessaloniki, Dept. of Energy (AUT/DE) University of Athens (UA), Dept. of Applied Physics	Ministry of Agriculture (MA) Ministry of Environment, Physical Planning and Public Works (MEPPW) Ministry of Industry, Energy and Technology (MIET) National Meteorological Service (NMS) Public Power Corporation (PPC) Water Supply and Sewage Corporation of Athens (WSSCA)	National Observatory of Athens (NOA) National Research Centre for Physical Science "Demokritos" (NRCPS "D") Centre for Renewable Energy Sources (CRES) Hellenic Agency for the Local Development and Local Government (HALDLG)

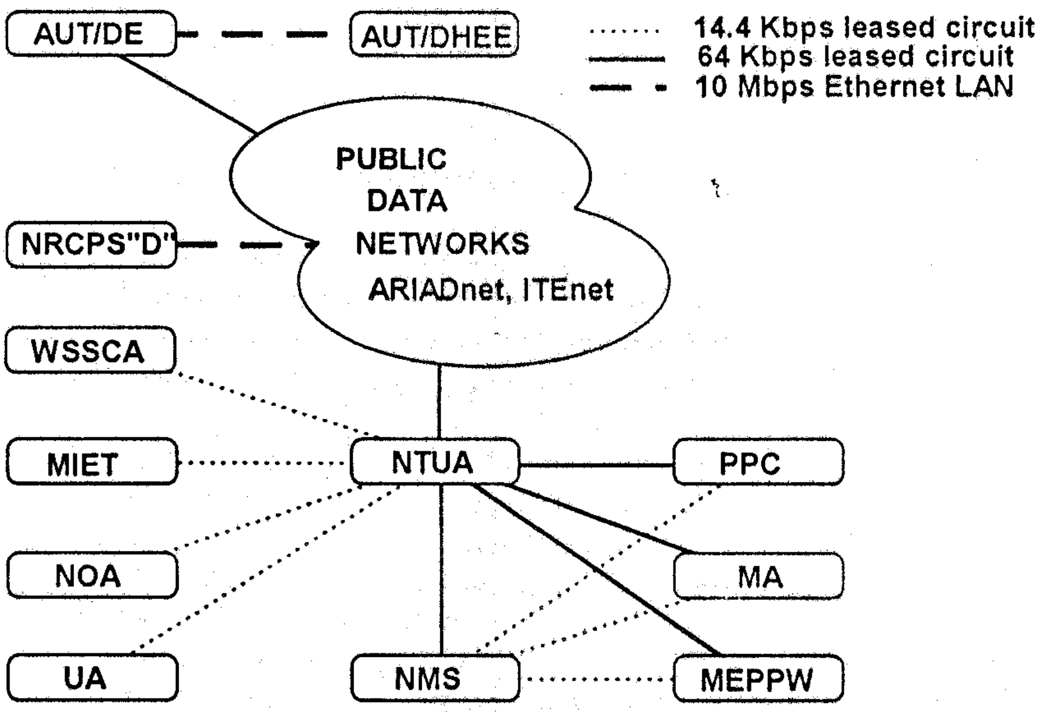

Figure 1 The HYDROSCOPE WAN topology (see Table 1 for abbreviations)

Situation until now Collection and archiving of hydrometeorological data in Greece, is performed by the organisations shown in Table 1, second column.

Owing to lack of co-ordination and assignment of responsibilities, ad-hoc and sporadic activities regarding hydrometeorological monitoring have evolved to less or more organised units within these organisations.

The main volume of the collected data, accounting for more than 450 million measurements, is currently stored in paper form (logbooks, strip charts etc.). It includes hourly and daily measurements of more than 25 variables, dated from 1890 up to now.

With HYDROSCOPE now in place, data retrieval and processing will be easier, much less time consuming, reliable and efficient. For the partners personnel it simply requires access to the local area network with a PC or a workstation.

Flexibility and diverse interests Regarding the collection and archiving of data, each agency has its own objectives and point of view, which in turn is translated into specific needs. For instance the Public Power Corporation is focused in dam studies and construction, while Ministry of Agriculture is interested in land use and extreme phenomena prediction and avoidance. On the other hand, the National Meteorological Service's main duty is to compile and publish weather reports and conduct climatic studies. Thus the diversity and multitude of user needs should be addressed in order to provide a real-world solution. The latter has been taken into account early in the design of HYDROSCOPE and is expected to show its results shortly after the system will be put in operational use.

Autonomy and exchange of information A basic design goal of the system is to ensure autonomy, while promoting data accessibility (Tolikas et al.[3]; Papakostas et al.[1]). Since a strict technical approach would probably lead to a centralised database system, this could seriously endanger the existence of the whole project. Competition, conflicts and overlapping authorities among public agencies along with insufficient and ambiguous legislation would never allow the creation of such a centralised system. In contrast, the distributed database approach employed by HYDROSCOPE (explained in the next section), satisfies most user needs while maintaining the "ownership" of data. This is accomplished thanks to the distributed nature of the system, which enables each public agency to have complete control over the physical storage and administration of its data. At the same time a comprehensive security and accounting sub-system establishes the rules of a fair co-operation among the partners and provides the necessary options for future transaction billing.

Experience, know-how and expertise of potential users An other significant question which arises is: Are potential users technologically up-to-date and ready to adopt new concepts and, in certain cases, innovative methodologies? We have found that official daily duties and bureaucratic procedures, in concert with the lack of internal training programmes, increase the knowledge gap between academic and public institutions, thus leading to a negative answer to this question. It was therefore inevitable for us (a) to promote user participation in every phase of the project (analysis, design, implementation, testing, quality assurance, documentation) and (b) to implement carefully designed training programmes to bring potential users up-to-date.

3.2 Project management

Structure Management of the project was organised hierarchically with the Steering Committee being the heading unit. A four-member Co-ordination Secretariat was assigned the duty to co-ordinate and monitor the project on a day to day basis. Other administrative units include the Scientific Committees and the Executive Council (a sort of partners assembly). The actual work was undertaken by the Workgroups, one for each partner.

Motivation One of the most challenging tasks throughout this project, was to invent motivation policies and provide incentives to get the best results from the available human resources. Job satisfaction, a term almost unknown to the public sector, plays an important role in a project like this one. Most of the partners' personnel responded positively.

Funding and administrative overheads The project is funded by the EC (70%) and Greek authorities (30%). Cash-flow proved to be a major problem for the management of the project, since significant delays were noticed regarding funding, with an immediate impact in the progress of the actual development. Project management had also, among other duties, to report to the supervising authority during the course of the programme. Regular reports included yearly forecasts and budgets, quarterly statements by partner, category of expenses and so on. Owing to the different accounting practices of partners, it is easily understood the complexity and the overhead imposed in every single transaction.

4. Technical aspects

4.1 Design issues

Distributed database model as a diplomatic solution To accommodate autonomy and "ownership" of data, it was imperative for us to design the system around the distributed database model. The implementation of the distributed database system has been based on the INGRES Relational Database Management system. In HYDROSCOPE, we have exploited in full its distributed database capabilities, enabling the users to access a set of apparently networked databases as if it was only one. User location is thus irrelevant to the system, since, assuming the same privilege level is granted, any user from any node on the wide-area network has transparent access to the data residing to his own and every other node. Insertions, updates and deletes in the database are under the administrative control of the "owning" organisation. Remote execution of this type of transactions is rejected by the system (Papakostas[1]), thus ensuring operational autonomy and avoidance of institutional conflicts. On the other hand, local or distributed queries can freely execute in the system, the only limitation being the maximum number of rows of the resulting table which will travel across the network. The purpose of this limit is twofold: to maximise the available bandwidth of the communication channels for a number of concurrent users, and at the same time to discourage potential attempts to transfer massively data from a node to another bypassing the primary security mechanisms. This upper limit can be adjusted by the system administrator according to network traffic patterns and available channel bandwidth. In addition, specially implemented remote procedure calls minimise traffic and processor load.

Client/Server architecture Distinction between client and server applications apart from being state-of-the-art in today's software systems, seems to be a reasonable choice when dealing with a large and ambitious database system, like HYDROSCOPE. Server applications for instance can embody all the security and integrity features devised during the design phase, without prior notification of client (or user) application developers. At the same time features such as the distributed access, user-defined data types and software triggers can transparently be shared among many application programmers. Client/server architecture enables also the multi-platform development and operation, since client tools run even in PC environment under MS-Windows.

Graphical Environment One of the application design goals was to provide the user with a rich set of integrated client applications employing leading-edge technologies in man-machine interaction and user-interfacing. Pursuing this goal, we developed all the client applications exclusively in windowed environments (OSF-Motif for UNIX stations, MS-Windows for PC clients), following specific standards in user-interface design in each of these environments. According to our experience, the overhead imposed by this decision, is easily justified by the increased user-satisfaction. The tool we have used for the application development is INGRES Windows-4GL, an integrated environment with an event-driven object-oriented graphical 4GL, a visual editor, a debugger, code sharing facilities, and version control system suitable for multi-user development.

New methodologies and techniques In the course of the project, we had the challenge to come up with specific solutions to certain types of problems:

1. User-defined data types: Hydrometeorological database applications need sometimes increased flexibility in the definition of data types. For instance the geographical co-ordinates is a data type not implicitly supported by any commercial database management system. Fortunately INGRES provides the means for the system programmer to define and integrate any kind of data type in the heart of the database server. The system programmer has also to define mathematical operations that refer to the newly created data type. The whole operation consists of writing code in C and link it to the database server executables.
2. The Differential-Linear Data Storage (DLDS) technique (Tsakalias[4]) is a method specially devised to overcome certain limitations of the relational data-

base model when dealing with time series data. It attempts a) to standardise the representation of a time-series in a relational database environment; (b) to implement a generalised scheme of a time-series with a varying time step; (c) to cope with the intermittent nature of time-series; (d) to create a set of SQL composite queries that substitute the typical simple queries in a manner that they are consistent with the data interdependence within a time-series; and (e) to provide a set of procedures for typical calculations of a time-series. In addition, the technique takes advantage of the property often exhibited in certain time intervals of hydrological series to have a constant (e.g. dry spells in rainfall) or approximately linearly varying magnitude (e.g. runoff during dry periods) thus providing a notable reduction (up to 60 - 90%) in the storage space and in the network load (in the case of a distributed data base).

3. OPSIS (Tsakalias[5]) is a sub-system which establishes a representation of hydrological data series as objects with standard characteristics, and a protocol that hydrological procedures should follow when acting on such objects. It performs a variety of typical hydrological and statistical computations through a consistent and user-friendly interface. Also, it provides many data visualisation alternatives which enable the user to perform interactively data checking and analysis.

4. PINAX is an embedded expert system module that performs homogeneity control on hydrological data sets. It also identifies nonstationarities and detects outliers of a time series. In addition, it performs curve fitting under nonstationary conditions (e.g., it identifies stage-discharge curves for nonstable river beds). The system uses a novel approach, quite simpler and more effective than a pure statistical approach.

Network Communications A distributed database system has usually severe requirements for speed and bandwidth. Although carefully designed after extensive simulation, HYDROSCOPE requires 64 Kb/s digital links among the nodes with large databases, while 14.4 Kb/s analog lines provide an adequate service level for all the other connections (Figure 1). Hi-speed digital lines were first introduced in Greece while HYDROSCOPE was in its second year of development. Thanks to the Public Telephone Company, we were given the opportunity to become the test bed for the hi-speed digital lines. In this way, we had the option to explore without significant charges, for about one year, this new communication service.

4.2 Implementation issues

System integration Trying to find the optimal mix of equipment for the project needs, we realised that a multi-vendor solution was appealing. To be able to implement the system as it was originally planned, we had to deal with and integrate several hardware components (workstations, PCs, file servers, printer servers, routers, modems, LAN's). We also paid much attention to provide software compatibility among different communication components (protocols, drivers etc.) mainly in the PC environment.

Productivity, Metrics Monitoring programmers' productivity was a task that revealed many interesting aspects. Once again classical software engineering estimations (Shooman[2]), have been confirmed. We have found that individual programmer's productivity varied in our case from 15:1 up to 25:1. That means that an experienced programmer produced about twenty times the output of a poor programmer. The results for run-time performance are more astonishing, reaching a ratio of 50:1, which is contributed mainly to inadequate training. We experienced also productivity problems due to the fact that software development was decentralised, taking place in five different sites.

5. Conclusion and future development

HYDROSCOPE, despite any administrative or technical problems encountered, finally managed to reach the initially set targets both at technical and human level. HYDROSCOPE now has an operational infrastructure including computing and data communications equipment as well as software applications for hydrometeorological information management.

Forthcoming phase II of the project will include the massive data entry and processing. Apart this, new developments will take place such as Geographical Information System integration, on-line telemetry network design and development, and expert system integration. Additional equipment installation is planned, to provide third party access through a variety of paths.

Acknowledgement: HYDROSCOPE project was approved by the Greek General Secretariat of Research and Technology (GSRT) and funded by the EC in the framework of STRIDE HELLAS research programme. Additional funding was provided by certain partners (MIET, MEPPW, MA, WSSCA, NMS, PPC), GSRT and Ministry of Education.

References

1. Papakostas, N., Nalbantis, I. & Koutsoyiannis, D., Modern computer technologies in hydrologic data management, *Proceedings of the 2nd European Conference in Water Resources Technology and Management*, Lisbon, Portugal, 1994.
2. Shooman, M. L., *Software Engineering: Design, Reliability and Management*, McGraw-Hill, 1988.
3. Tolikas, D., Koutsoyiannis, D. & Xanthopoulos, Th., HYDROSCOPE: Un système d'informations pour l'étude des phénomènes hydroclimatiques en Grèce, *6ème Colloque International de Climatologie*, Thessaloniki, Grèce, 1993.
4. Tsakalias, G. & Koutsoyiannis, D., Hydrologic data management using RDBMS with Differential-Linear Data Storage, in Hydrosoft/94, *Proceedings of the 5th Int. Conf. on Hydraulic Engineering Software*, Porto Carras, Greece, 1994 (this volume).
5. Tsakalias, G. & Koutsoyiannis, D., Opsis: An intelligent tool for hydrologic data processing and visualisation, *Proceedings of the 2nd European Conference in Water Resources Technology and Management*, Lisbon, Portugal, 1994.