TASK QUARTERLY 15 No 2, 235–245

DEVELOPMENT OF SERVICES AND INFRASTRUCTURE FOR DATA ACQUISITION AND INFORMATION EXCHANGE

MŚCISŁAW NAKONIECZNY[1], ANTONI NOWAKOWSKI[2] AND JACEK WYRWIŃSKI[3]

[1] *TASK Academic Computer Centre, Narutowicza 11/12, 80-233 Gdansk, Poland*
mnak@task.gda.pl

[2] *Department of Biomedical Engineering, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland*
antowak@biomed.eti.pg.gda.pl

[3] *Institute of Oceanology of the Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland*
wyrwa@iopan.gda.pl

(Received 2 March 2011)

Abstract: Focusing on the local solutions used by the institutions organizing the INFOBAZY (Infobases) conference, we briefly examine the development of IT infrastructure for science in Poland, including resources, networks and a high performance supercomputer centre. This development is presented from the perspective of the last twenty years – although the first INFOBAZY conference was organized in 1997, the IT infrastructure has been developed since 1991, when the first modern network server was installed in Warsaw. This presentation demonstrates both the enormous progress in the field and the state of the art today.

Keywords: informatics infrastructure, scientific databases

1. Introduction

It has been twenty years since the first modern network server was installed, laying foundations of the development of IT infrastructure for science in Poland. This event, the importance of which no one could have foreseen in 1991, constitutes the true beginning of the integration between the Polish, European and world science. The organizers of the INFOBAZY conference were privileged to partake

in the development of modern IT infrastructure for science since its beginning, particularly in the Tricity, but also in the rest of Poland, participating in the works of KBN (State Committee for Scientific Research) since 1993, and in Europe. Prof. Antoni Nowakowski represented Poland in the Information Society Technology Committee during the 5$^{\text{th}}$ Framework Programme of the European Union. For seven years he has also been the head of a team of experts on IT infrastructure for science in KBN/MNiSW (Ministry of Science and Higher Education).

One of the milestones in the informatization of Polish science was, among others, the enactment of the *Programme for the Development of IT Infrastructure for Science in Poland* in 1995, according to which until 2000 each Polish scientist should have access to the Internet and IT resources, while five supercomputer centres, located in Warsaw (ICM), Krakow (Cyfronet), Poznan (PCSS), Gdansk (TASK), and Wroclaw (WCSS) should play a prominent role. Another milestone was the enactment of a successive *Programme* in 2000 and the establishment of the PIONIER consortium, which constituted the basis for initiating the development of IT infrastructure for science based on the state-of-the-art optical and wireless technologies.

One of the most important elements of these *Programmes* was the understanding of, and the emphasis on, educating society; the teaching of the most recent information technologies, which were becoming widely accessible, but were still in need of promotion and popularization. Here, it is worth quoting Minister Małgorzata Kozłowska [1] speaking to the participants of the first INFOBAZY conference in 1997: "The initiators and organizers of this conference aim for it to become a plane for the exchange of experiences and the presentation of achievements and problems encountered by scientists and academics who generate the IT resources available through academic computer networks. [...] I hope that this conference, along with such conferences as POLMAN and INFOFESTIWAL, finds its place among other cyclical events organised under KBN auspices."

The following plenary report from the 6$^{\text{th}}$ INFOBAZY Conference, entitled "Science, European Projects and Information Society", which took place in 2011, is presented on the 20$^{\text{th}}$ anniversary of the introduction of modern digital computer networks into Polish science and on the 15$^{\text{th}}$ anniversary of the introduction of integrated funding of IT resources available through supercomputer centres. We note with great pleasure that the words of Minister Kozłowska still apply today – there is a continued need for the INFOBAZY conference and the conference measures up to expectations. As time has shown, much larger events, *i.e.* POLMAN and INFOFESTIWAL, exhausted their mission, whereas fields such as IT resources and databases are still needed and alive, which this conference proves.

In the subsequent part of the report, we focus on the development of the TASK Academic Computer Centre. TASK is one of the five supercomputer centres in Poland, boasting a harmonious cooperation with the remaining centres. Also, using the example of the Institute of Oceanology of the Polish Academy of Sciences

(IO PAN), which is one of many scientific institutions operating in the region, and a relatively small one, we would like to demonstrate how scientific initiatives can be effectively developed. Finally, we would like to briefly reminisce about previous conferences and to devote a moment to the prospects of scientific databases.

2. TASK Academic Computer Centre

The academic year 2011/12 is another year the TASK Academic Computer Centre (CI TASK) functions as a supercomputer centre. The expansion of the Centre's resources that has been carried out for eighteen years now and consists in the implementation of the state-of-the-art and fastest network technologies as well as the continuous increase in the supercomputing power allow to implement new IT services necessary for the development of an information society. The Centre constitutes an important element of the IT infrastructure for science in Poland; it manages one of the largest metropolitan computer networks in the country, and it is one of five national high-performance computing centres.

This dynamically developing Centre, which continues to expand its potential, is an inseparable partner of academia, strengthening its leading position in the field. At present TASK (Tricity Academic Computer Network) cooperates with 118 local networks at universities and research centres, as well as with 31 networks in halls of residence. TASK also supports educational units, public buildings, hospitals, units of local government, companies cooperating with scientific units as well as several cities in the region of Pomerania.

The infrastructure of the network, based from the start on fibre-optic cables, amounts to over 250 km. A broadband network and the technologies it utilizes enable fast, remote access to the resources of the computational servers (high-performance computers) and to several other services. The networks dedicated to science utilize 10 gigabit Ethernet (10 Gb/s) for their main interconnects. The backbone PIONIER network is the basis for the communication of all metropolitan area networks (MANs). TASK is connected to the national PIONIER network by means of three interconnects (2×10 Gb/s each) passing through the cities of Torun, Szczecin and Olsztyn. The PIONIER network enables broadband transmission of data with the high throughput required for certain research projects, *e.g.* in the field of geography and oceanography (transmission of satellite pictures); interactive HDTV; as well as for setting up virtual intercity networks used by cooperating participants of various research projects which require a guaranteed wide transmission band. The access to European and international academic networks is realized through the GEANT network, which is connected to the PIONIER network by multiple interconnects with a throughput of 10 Gb/s each (Germany and Czech Republic).

We point out that TASK is not only a vast and fast Internet network, but also a modern IT Centre which can boast a world-class supercomputer, Galera Plus (all generations of supercomputers utilized by CI TASK were ranked high in

the list of TOP 500 fastest computers in the world). This 2264-processor, 10896-core supercomputer with a computational performance of 102 Tflops currently functions as the basic computational server in our IT Centre. It is extensively used for advanced computations and computer simulations by national and international scientists. The functionality of the Centre's supercomputers is supported by the LUSTRE system for archiving and exchanging files and an IBM tape storage system with a storage capacity of 2.5 PB.

Approximately 650 users (scientists, doctoral and postgraduate students) hold computational grants needed to access the high-performance computers. Users have access to over 30 software packages supporting their scientific research. The wide range of installed high-performance software can be applied in diverse fields, both in teaching duties and for basic and applied research. Apart from scientific software (such as Abaqus, Gaussian, MSC.Software, Ansys), our HPC centre provides compilers, libraries and supporting tools enabling users to create their own applications. The Centre also provides a local licence server for packages such as Matlab, Mathematica, ArcInfo and Abaqus, which is utilized by external users. A similar server, with national scope, is also available for the MSC.Software package.

We stress that such dynamic development of our Centre in recent years was made possible by funding from the *Programme Innovative Economy*. The Centre takes part in six large EU projects, carrying them out in cooperation with many national units. In 2009–2012 the Centre was involved in the following projects:

MAYDAY EURO 2012

This project is focused on the identification of objects or dangerous events, transmitted from CCTV equipment or other multimedia stream sources in various fields, such as medicine, art or industry. This problem will be solved by designing algorithms using the computing power of a computing cluster. This is particularly important in view of the 2012 UEFA European Football Championship in Poland.

Polish grid infrastructure PL-Grid for e-Science

Carrying out scientific research necessitates using advanced IT technologies. In order to realise the new paradigm of carrying out research known as e-Science, a grid infrastructure is required, which allows research groups from environments that lack computer centres to comfortably access computer resources.

Pomeranian Digital Library

The aim of this project is to create a digital library, which would enable online preservation and presentation of the oldest written texts and other relics, kept by Pomeranian libraries.

Integrated Ocean Data and Information Management System

The aim of this project is to create a National Centre for Oceanographic Data. To this effect, an Integrated Ocean Data and Information Management System must be created, which constitutes the direct aim of the project. CI TASK provides the IT infrastructure and computing power necessary for the realization of this project.

PLATON – Science Services Platform

The aim of this project is to expand the national teleinformatic infrastructure (PIONIER network) with applications and services supporting the research and development of Polish research groups devoted to Innovative Economy. The direct aim of the project is to implement various modern teleinformatic services.

NewMAN – Extension of 21 Local Teleinformatic Scientific Networks

This project is carried out by members of the PIONIER Consortium. Its aim is to provide 21 MANs with the state-of-the-art network equipment using MPLS to carry out advanced teleinformatic services and to set up dynamic virtual connections and dedicated on-demand channels to be used by universities and scientific institutions in national and international projects.

Participation in the aforementioned EU programmes obligated us to implement several new solutions encompassing the following services:

- multimedia services related to image-, sound- and text-processing,
- distributed computing,
- aggregation of computing power,
- on-campus computing,
- oceanographic databases,
- library bases,
- distributed archiving,
- videoconferences,
- eduroam,
- interactive scientific television,
- on-demand network throughput.

Apart from the projects funded from EU funds, we carry out investments related to the expansion of the Centre's infrastructure, necessary to expand the functional computing power of computing servers and to develop new network services for the academic and research environment. The development of the Internet results in continual progress made possible by the emergence of services which could not have been utilized earlier due to technological limitations of network throughput or inadequate computing power of servers. Currently, broadly defined multimedia services, which replace text and graphical information and digitally encoded audio and video contents from other sources, are becoming progressively more popular and significant. Sending and using multimedia content generated by intelligent terminals and servers, made possible by broadband network, is becoming an independent service, one of the most important on the Internet.

A modern fibre-optic network significantly influenced the continuing cooperation between universities and the integration of academia. This was crowned with a recently implemented videoconferencing service encompassing all universities in the Tricity, whose authorities belong to the Vice-chancellors Council of Pomorskie Province. The application of modern videoconferencing systems transmitting full-HD image and broadband sound allows many users to carry out virtual

meetings almost as if they were having face-to-face conversations, despite the fact that they are many kilometres apart.

As part of pilot works, video is streamed from webcams located in selected locations in the Tricity. In this way, the entire world can use a web browser to watch video streams from webcams installed in the National Sailing Centre in Gdansk, the marina and Skwer Kościuszki (Kościuszko Square) in Gdynia, the Hippodrome in Sopot, and in the Old City in Gdansk.

Being highly experienced in video streaming and carrying out videoconferences, CI TASK substantially and technically supported the multimedia campaign "Citizen Solidarity" organized as part of the celebrations of the 30$^{\text{th}}$ anniversary of the Solidarity movement. Using the Centre's resources, a videoconference telebridge was set up, connecting Plac Solidarności (Solidarity Square) in Gdansk and the marina in Gdynia. Furthermore, an experiment in augmented reality was performed in front of the plaque of strikers' 21 demands next to the historic gate to the Gdansk Shipyard. It was possible to watch both communication installations "live" in a web browser.

TV TASK is a new service, developing dynamically over the last several months. Owing to a well-equipped television studio connected to the Internet, the viewers can learn the secrets of a supercomputer server room or watch the coverage of events taking place in the Tricity. Our TV producers prepared video materials for the workshops "Hoist the Sails of Science – Study Tour" and the Baltic Science Festival, among others. These can be watched on the CI TASK website.

With the aim of carrying out marine research, the Centre was expanded with a Marine Internet Laboratory (MILA). This laboratory, equipped with a sailing vessel, carries out research, among other things, on measuring Internet connectivity at sea. It also supported IO PAN in carrying out the hydrological and aerosol research in the Bay of Gdansk and on the Polish Baltic coast.

In the scope of scientific activities for the general public, the Centre co-organises a series of monthly meetings with outstanding Polish scientists in Kawiarnia Naukowa (Scientific Café), held under the patronage of the Vice-chancellors Council of Pomorskie Province as part of the Baltic Science Festival.

Aiming to popularise science, the Centre is open to visits of school and university students and researchers from the entire region, allowing them to get acquainted with the most recent world achievements in the field of modern computer clusters and the development of IT networks in the Tricity and the rest of Poland. Posters presenting the history of the Centre, and thus the history of the development of the Internet network and supercomputers in the Tricity, are on display next to the server room. Last year, 2500 interested visitors saw the Galera supercomputer.

2011 was an important stage in the realization of many national projects in the framework of the *Programme Innovative Economy*, which concerned the progress in the development of databases and IT resources for science in digital

form and the implementation of advanced applications and teleinformatic services for academia. The 6$^{\text{th}}$ INFOBAZY Conference 2011 also constitutes a part of the activities of CI TASK. With the approval of the Minister of Science and Higher Education, this Conference acts as a forum for presenting the progress of individual projects. This year, its central theme is *Science, European Projects, Information Society*. Thus, the Conference acts as a forum for exchanging experience, presenting new databases, implementing advanced applications and teleinformatic services for academia, and, above all, for presenting the progress of European projects and the vision of further development of services and of dissemination of information.

The current edition of the Conference is particularly focused on those European Projects which substantially support, both financially and thematically, the *Programme for the Development of IT Infrastructure for Science*, as well as on the *Applications and Services Necessary for the Development of Information Society*. Particularly important are the visions of the development of services and of dissemination of information, as well as the presentation of current achievements in creating databases and making them accessible to science and society, including national and local authorities.

3. IO PAN as a co-organizer of the INFOBAZY conference

The Institute of Oceanology of the Polish Academy of Sciences (IO PAN) in Sopot was established in 1983 as a continuation of the Marine Station (1951) and of the Oceanology Department of the PAN Geophysics Institute (1970).

The mission of the Institute is to carry out basic research on marine environment and to broaden the knowledge of its phenomena and processes. This mission has been performed by the Institute for 50 years.

Figure 1. The head office of IO PAN in Sopot

The strategic research of IO PAN is focused on:

1. The role of the ocean in shaping the climate and the consequences of climate change in European seas (research on the transport of solar radiation, the exchange of radiative energy in the water-atmosphere system, photosynthesis processes, mass and energy between the sea and the atmosphere and on thermohaline circulation).
2. Natural and anthropogenic change of the Baltic sea environment (research on hydrodynamic and biological processes, migration of chemical substances, biochemical processes in marine environment and on organic compounds in bottom sediment).
3. Changes in ecosystems by the shore of shelf seas (research on biodiversity in the functioning of inshore ecosystems and the fauna of Spitsbergen, functioning of ecosystems in marine sites of the NATURA 2000 program).
4. Genetic and physiological mechanisms of the functioning of marine organisms and the basics of marine biotechnology.

This diverse and multidisciplinary subject matter requires a wide range of research platforms (such as the Institute's own vessel Oceania, other research vessels, moorings – long-term measuring buoys, seashore measurements, measurement laboratories) and diverse measuring (meteorological, optical, hydrodynamic, acoustic, chemical, *etc.*) equipment.

Figure 2. The research vessel Oceania

The Institute gathers diverse measurement data, such as satellite photos, vertical profiles of seawater (up to 5 km), temporal-spatial profiles (multikilometre transects); gathering samples of phyto-and zooplankton, water and bottom sediment samples, meteorological and aerosol profiles, LIDAR and acoustic data, spectrophotometric data, seashore data and many others. The obtained data are very diverse and heterogeneous, they are gathered in series over many years in the vast basins of the Baltic Sea and Nordic Seas.

Figure 3. The head office of the Institute of Oceanology, a co-organizer of the INFOBAZY conference in 2008

Managing, processing, maintaining and making these data available to the public has always been a priority for the Institute. This was reflected in the addresses at the previous conferences that the Institute co-organized.

The limited financial resources of the Institute hampered the efforts to organize a genuine data-serving system, even though the Institute has organized the *Regional Oceanographic Database* for many years. The financial problems did not go unnoticed, which resulted in the securing of serious EU funds, giving hope for their prompt resolution – for instance, the European grants *Integrated Ocean Data and Information Management System* and *SatBałtyk – A Baltic Environmental Satellite Remote Sensing System*, which will be discussed during this Conference. Using the gathered data and database tools, the Institute actively participates in many international programmes, such as ALKEKONGE, AWAKE, BIOCOLOR, HABITAT MAPPING, LITUS, DESAMBEM and many others.

4. INFOBAZY – summary

The concept of central subsidization of scientific databases in Poland was part of the *Programme* enacted in 1995. It was carried out in 1996, fifteen years ago, and the corresponding regulations and information were included in

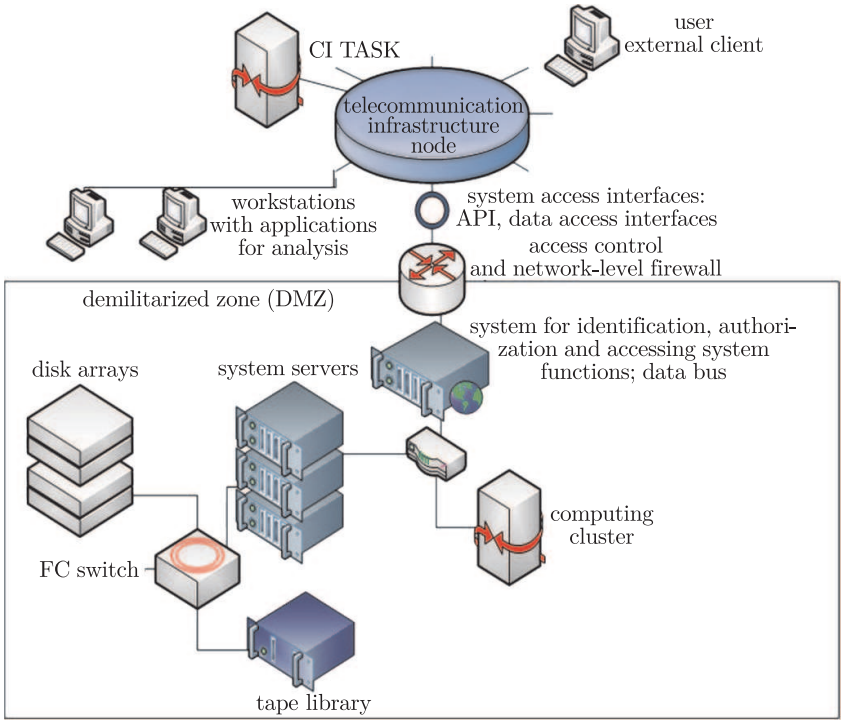

Figure 4. General schematic of IT structure at IO PAN

the proceedings of the first INFOBAZY conference [1]. The principal aim was to ensure Internet access to the subsidized databases. Initially, the number of proposals exceeded 150, in practice, however, in the first year 15, and in the subsequent year 135 projects were subsidized. The INFOBAZY conference, along with a central repository of scientific databases, managed by CI TASK, known as INFOBAZA (Infobase) were among the outcomes of these activities. Unfortunately, this initiative was not subsidized on a regular basis, and although formally it exists until this day, the database is far from up to date.

Over the course of time, the Ministry became less fond of the campaign for the central subsidization of databases, as the number of proposals was vast, and the capabilities of the program – rather modest. In consequence, the central financing programme for databases no longer exists, but certain databases are still being developed, new ones spring into existence, some disappear from the Internet, generally speaking – the field is alive! This is best proved by the fact that the INFOBAZY conference still exists and by the existence of CODATA – an international organization, whose Poland is an active member of CODATA also actively supports the INFOBAZY conference, and one of the sessions is devoted to the memory of Prof. Andrzej Bylicki, a long-time chairman of the CODATA Polish National Committee, a former vice-president of this organization, who passed away in 2010.

A significant step in the direction of creating new scientific databases was the commencement of a programme subsidized from EU funds, and in particular from the European Social Fund (ESF), *i.e.* the *Programme Innovative Economy*. Many of the presentations at this conference are the results of research made possible by the funding provided by this programme, which marked a qualitative change – the amount of funds significantly exceeded the former funding from the Budget, thus promoting the emergence of projects whose scale would have been inconceivable at the first stage of the programme for the development of IT infrastructure for science. One of the examples of such projects is the aforementioned *Integrated Ocean Data and Information Management System*.

During previous conferences, we were concerned with how to secure funds for the creation of databases. This problem remains current. Therefore, the INFOBAZY conference is still held under the patronage of the Ministry of Science and Higher Education and a representative of the Ministry is always present at the Conference to advise us on the possibilities of further investment in the development of IT infrastructure for science, whose fundamental element is a wide platform for the exchange of scientific data. These are the aims of the current Conference as well.

References

[1] 1997, Conference Materials from INFOBAZY'97, Gdansk