ITALIAN PHYSICAL SOCIETY

CONFERENCE PROCEEDINGS

VOLUME 69

8th Workshop Italian Research on Antarctic Atmosphere

edited by M. Colacino and G. Giovanelli
Bologna, 20-22 October 1999

Area della Ricerca, CNR

Centenario 1897-1997

Società Italiana di Fisica
Bologna - Italia

Conference Proceedings Vol. 69
"Italian Research on Antarctic Atmosphere"
M. Colacino and G. Giovanelli (Eds.)
SIF, Bologna, 2000

A GEO-DATA SERVER FOR ANTARCTIC DATA

L. Fortunati [1], O. Salvetti[2],
V. Costa[1] , R. Demontis[1] , A. Galligani[2], C. Carlesi[2], S. Biagioni[2]

[1] Institute CNUCE, Italian National Research Council,
Via S. Maria 36, 56126 Pisa, Italy
[2] Institute for Information Processing, Italian National Research Council,
Via S. Maria 46, 56126 Pisa, Italy

Abstract

This project is the natural evolution of the South Pole System for Antarctic Data Exchange [1] and has as main objectives the organisation, homogenisation, standardisation and diffusion of all the data acquired by the National Program for Research in Antarctica.

A main activity is to design and implement a Geo-Data Server [2, 3, 4], constituted by a system storing consolidated data for remote consultation.

Main features of Geo-Data Server are the definition of an Antarctic cartographic and image database, according to the international standards, a relational catalog, containing meta-information about the data archived in Geo-Data Server and a high-level interface to access the system.

At present, a simulation prototype is under construction [5], composed of a UNIX workstation and three high-performance PCs networked, where the data managed are supplied by two Data Centres of South Pole System for Antarctic Data Exchange ('GEO' Centre, ENEA Casaccia, for remote sensing data, and 'TER' Centre, Dept. of Earth Science, University of Siena, for cartographic data).

In this paper, the first results of the approach followed are shown.

PNRA Structure and Information System

The PNRA (National Program for Research in Antarctica) was organised in 5 main disciplinary sectors:

1. Geological Evolution of the Antarctic Continent and Southern Sea
2. Global Change
   2a. Glaciology and Paleoclimate
   2b. Antarctic Atmosphere, Sea-Ice-Atmosphere Interaction
   2c. Biology, Environmental Contamination and Medicine
3. Observatories and Geographical Information
4. Management Methodologies and Regulations for the Presentation of the Environment.

Each sector/sub-sector was associated with a Working Group of the National Scientific Committee for Antarctica (CSNA).

The South Pole System for Antarctic Data Exchange and the Data Centres
An integrated multi-disciplinary information system, known as the South Pole System [6], was planned within the framework of the National Programme for Research in Antarctica and promoted by ENEA (National Agency for Alternative Energy). The system consisted of a series of heterogeneous data banks located in Data Centres, geographically distributed throughout the Italian territory, each one dedicated to a specific discipline (Fig. 1).

Code	Site	Data	URL
AGO	IFSI-CNR Frascati	Sun-Earth Relationship and Astrophysical Research Data	http://sunserv.ifsi.rm.cnr.it/~ago/AGOindex.html
BAT	IGM-CNR Bologna	Bathymetric data	http://terra.bo.cnr.it/
BIO	University of Padua	Land Biology data	http://caos.bio.unipd.it/antartide/
CHI	University of Venice	Environmental Impact data	http://polosud.unive.it/
FIS	ING Rome	Earth Seismology data	http://bounty.ingrm.it/Welcom_fis.html
GEM	OGS Trieste	Marine Seismology data	http://www.ogs.trieste.it/pnra/
GEO	ENEA/Casaccia	Satellite Remote Sensing Data	http://terranova.iei.pi.cnr.it/GEO/geo-cas.html
LID	IROE-CNR Firenze	LIDAR Data	http://lid.iroe.fi.cnr.it/
MAR	IST-CNR Messina	Marine Ecology Data	http://aistalas.ist.me.cnr.it/south-pole/
MET	ENEA/Casaccia	Meteo stations data	http://terranova.iei.pi.cnr.it/MET/index.html
OCE	IFA-CNR Rome	Oceanographical data	http://terranova.iei.pi.cnr.it/OCE/oce-rm.html
TER	University of Siena	Earth Science and Glaciology data	http://www.dst.unisi.it/PNRA/pnra.html

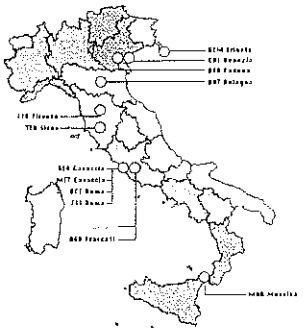

Fig. 1 – Data Centres of the South-Pole System.

The main objective of the South Pole System was to make all the data and information concerning the Italian Antarctic expeditions available to the scientific community, through the creation of a series of specialised networked Data Centres. In particular, the system provided:

- local and remote access to all the information
- compatibility with international standards
- homogeneous user-friendly interfaces
- the possibility to perform comparative and interdisciplinary analyses.

The System consists of a set of heterogeneous databases located in 12 geographically distributed Data Centres, each one containing data specific for discipline.
The Data Centres were connected via telematic networks (Internet) and the "brain" of the system was at the European Space Agency (ESA-ESRIN) in Frascati (Fig. 2).

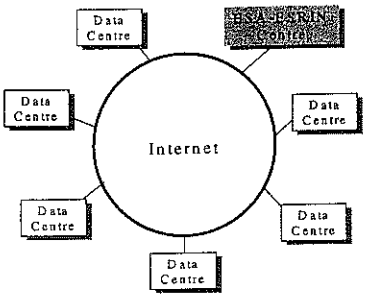

Fig. 2 – Network of the Data Centres.

The ESA-ESRIN Centre contains the meta-information about the research activity performed inside the PNRA, structured as a subset of CEOS IDN. Its content constitutes the South-Pole Directory and is part of the Italian System for Antarctic Data Exchange (Fig. 3). It was planned to contain both on- and off-line all the main scientific results collected during the Italian Antarctic expeditions sponsored by PNRA. It is a multidisciplinary directory that provides the user with the existing research activities and available data description as well as the procedures for data releasing. The Directory adopted the Directory Interchange Format (DIF) of the National Space Science Data Centre (NSSCD).

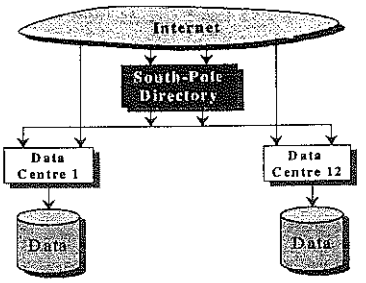

Fig. 3 – Meta-information about the PNRA research activity is contained in the South-Pole Directory.

Besides, a user-friendly system interface was designed in such a way to define a common model able to cover both internal aspects (i.e. data access and retrieval by local users) and external aspects (remote access) [7, 8].

After the implementation of the system and the population of the archives, the situation presents some particular aspects that lead to the following observations:

- non co-ordinate evolution of the various databases: each Data Centre makes use of its own data models, generally without standardised criteria, and consequently producing poor integrability of data between the Data Centres;
- proprietary use of the databases: the database is projected and implemented mainly for use inside the Data Centre;
- integration only limited to meta-information: only at this level data from Data Centres may be integrated (South Pole Directory);
- poor documentation of data.

The South Pole Directory population presents some faults:

- each Data Centre can modify the layout of its site;
- each Data Centre can self-define keywords (usable for search operations) creating conflicts;
- the Data Centres can insert data in the databases without creating the associated DIF files.

The fundamental properties of the system are therefore not satisfied.

In fact, the user finds:

- inconsistency of the user interface among the various centres or total absence of the sites of some centres;
- deep difference for accessing the different local databases;
- poor efficiency and incapability of exploiting the research instruments;
- lack of available data and of information about them.

The Geo-Data Server

To overcome the previous problems (mainly related to data standardisation and consistency of the database) and to really make all the data and information concerning the Italian Antarctic expeditions available to the scientific community, a centralised structure (the Geo-Data Server) was proposed aiming to reach the following general goals [2, 5]:

- to allow the multi-disciplinary analysis of the data collected within the Italian campaigns in Antarctica and of the subsequent elaboration carried out: scientific information, rightly structured and standardised, can be diffused into various scientific contexts to produce synergy in the study of interdisciplinary problems. This is a strategic interest consisting both in the valorisation of previous scientific experiences and in the aggregation of multiple specialist competencies;
- to dispose of consultation and help tools having a specific functionality for an integrated use of the data provided by the different research groups;
- to allow data integration on spatial basis;
- to aim at an harmonisation with the international context, through the participation in international research activities.

Such an approach can produce the following benefits:

- a research sector having worked traditionally on its data could access the data produced by other sectors in order to define an interdisciplinary research activity context;
- the knowledge that some data exist and are available, together with their exact and detailed documentation, has a remarkable value, especially considering that data production is quite expensive;
- in a national context, the actual South-Pole System will be increased in value, being the information kernel of the current PNRA knowledge;
- in an international context, the PNRA connection with other countries by sharing research data and results should be considered a strategic and relevant factor (see Art III of the Antarctic Treaty and the several recommendations of SCAR on Antarctic data and information availability);
- the alignment and/or the adaptation to standards of data models resident into international databases allows the PNRA information to be shared with international research groups, and vice versa.

It is important here to highlight that the Geo-Data Server has not the goal to replace any already existing disciplinary database located in the Data Centres; it aims to implement a spatial-based integration of the consolidated data contained, giving an additional value to the current information power.

Concrete objectives may be identified synthetically by:

- the definition of a reference data-model usable and extensible for local GIS needs;
- the adoption of a standard cartographic base (area of interest, scale, themes, ...);
- the activation of a network data server based on existing network facilities (Internet);
- the organisation of a catalog of the available data with relative characteristics.

The system thus provides:

- compatibility with international standards;
- local and remote access to all the information;
- the possibility to perform comparative and interdisciplinary search;
- homogeneous user-friendly interfaces in the catalog search operations;
- efficient organisation of the data and the operational flows.

The Geo-Data Server provides also the interface for searching and accessing all available data: this facility is a consequence of the definition of metadata standard and the creation of a data Catalog.
Data availability can be enquired on a geographical area and for a predefined set of attributes.

The Geo-Data Server Archive

In this new philosophy of the Geo-Data Server system, all the Data Centres are both data providers and users. They contribute exclusively with *consolidated data*, that is with data validated, published and officially released by the Data Centres.

All data and their descriptions are processed for integration and homogenisation, and then stored in a unique centralised system, the Geo-Data Server Archive. The integration and homogenisation phase consists mainly of the adoption of a common geographic and temporal reference system and a homogeneous data model and of a data conversion process (Fig. 4).

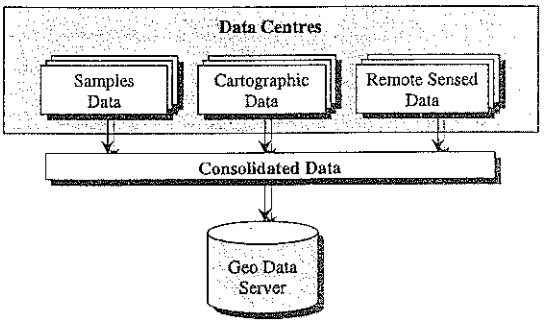

Fig. 4 – Data typology stored into the Geo-Data Server.

The Geo-Data Server Architecture

The Geo-Data Server is made of two parts (Fig. 5):

- a GIS-Host, for data processing and accessing;
- a Disco-Server discovery tools, for metadata storing and classification.

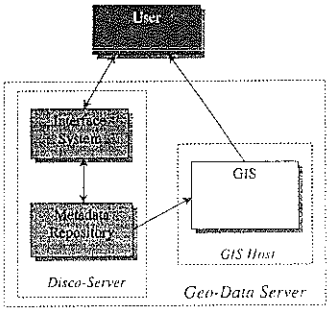

Fig. 5 – Overall architecture of the Geo-Data Server.

Both the two parts communicate with the network by means a WWW server and specific programs providing the user interface for searching and accessing all available data.

The Geo-Data Server is split on two distinct hardware systems: the GIS-Host on a Sun Unix and the Disco-Server on a PC Windows NT (Fig. 6):

- The GIS-Host performs all the tasks involving the data coming from the Data Centres: input, checking, validation, mapping to the defined data model (data conversion of both the geographic and alphanumeric components). These are off-line works carried out by GIS processes that contribute to the population of the Geo-Data Server Archive. The GIS interfaces also the WWW server from which it receives the data extraction requests (from the Geo-Data Server Archive). Specific tasks involving GIS processing are activated in order to extract data and to export them in the requested format. Then, the data downloading is allowed.

- The Disco Server performs a set of tasks synthesised as follow:
  - inputs metadata according to the adopted standard;
  - prepares all the reference backgrounds for the geographic queries, normalise of the keywords and set up the thesauri;
  - performs the queries (on a geographical or textual basis) on the Metadata Catalog and show the results as a list of entries or full metadata information;
  - addresses the GIS Host for data downloading.

  At the base of the query task is the link between the WWW and the DBMS with which all the user requests are processed and the results of the search process are showed.

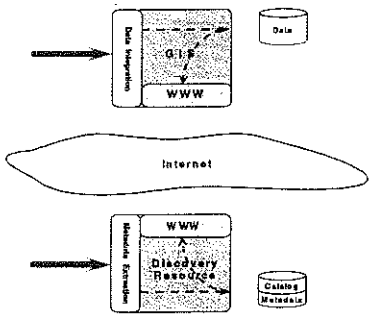

Fig. 6 – Geo-Data Server components and their interaction.

The Data Standardisation Process

The creation of the archive of data involves aspects regarding the standardisation of the data from two points of view: Metadata and Data Exchange. The adopted solution has been selected after a study to evaluate the state-of-the-art in this field.

- ***Standards for Metadata: state-of-the-art***

  The most important metadata standards and their relationships were evaluated (Fig. 7), starting from the one (D.I.F.) that is actually in use and the one (ISO/TC 211) proposed by SCAR. The most interesting standards for our purposes are listed below:

  - ISO/TC 211, the world standard
  - C.S.D.G.M., the US metadata standard

- CEN/TC 287, the European standard
- D.I.F., a very simple US metadata standard (used by CEOS IDN)

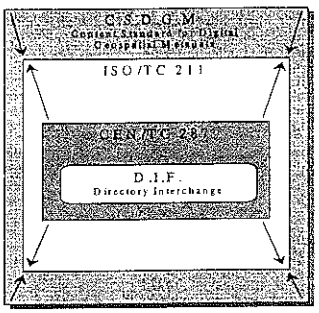

Fig. 7 – General compatibility relationship among the geographical metadata standards.

ISO/TC 211 would be particularly interesting being recommended by SCAR and designed as an international standard, but unfortunately, it is not yet completely defined and its documentation is at present reserved to the development groups only. On the other side, the European standard is a new emerging standard moving to be similar to ISO and compatible with it.

Furthermore, it is also interesting to maintain the maximum compatibility with the older Italian Antarctic systems, particularly with the South-Pole System, and with other GIS systems and directories, like the Global Change Master Directory and the Antarctic Master Directory. For these reasons, has been implemented an automatic tool able to convert D.I.F. to CEN/TC 287.

The relationships between the two standards are shown in Figure 8:

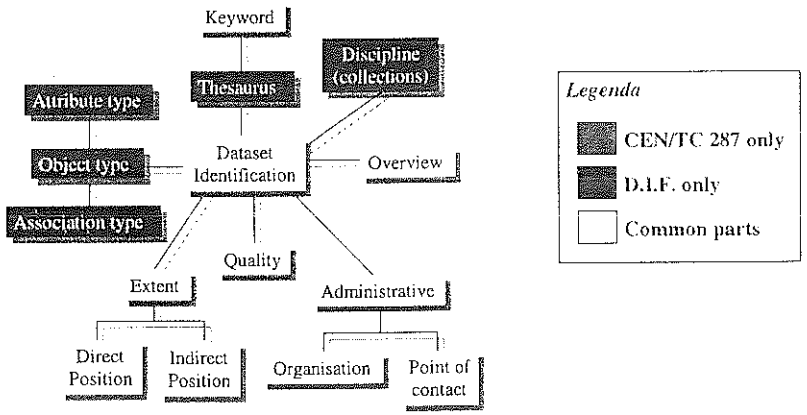

Fig. 8 – Relationships between DIF and CEN/TC 287.

- *Standards for Data Exchange: state-of-the-art*

  This task was performed taking into account a set of data standards [9, 10], related to the following categories:

  - *International standards* (DIGEST, ISO8211, CEN287, OGIS, ...): are standards mainly related to international standardising organisations or pools of heterogeneous organisations working in specific fields (industry/integrator, government agency, hardware technology, software technology, data/information supplier, university, development lab, etc.);
  - *National standards* (EDIGEO, NTF, ...): are standards promoted by national governments for their local administrations;
  - *Proprietary (commercial) standards* (E00, Shapefile, DGN ISFF, DXF, GIF, GEOTIFF, TIGER, ...): are standards defined by the software producers, specific for their software systems produced.

The Adopted Solution

In order to start in the realisation of the system, some fundamental decisions were adopted in the following areas:

- **System Architecture:**

⇨ *Two separate hardware systems for Data and Metadata handling*: in this way, two independent environments, each one having its own functionality and also the possibility to operate autonomously, are available. Moreover, a main aspect is that the system for Metadata can address the system for Data.
The system for Metadata is a PC running Windows NT, with SQL Server as DBMS and Internet Information Service as WWW software. The system for Data is a SUN workstation running Unix, with Arc/Info as GIS software and Apache as WWW software.

⇨ *Scalable components configuration*: the whole system can grow adding more functionality and more datasets.

⇨ *Use of commercial products for hw and sw*: this permits to work with consolidated and stable products, minimising problems arising from failures and maintenance.

- **Standards:**

⇨ *Metadata*: CEN287 is the actual selected standard and will migrate to ISO TC212 when available; this choice permits to start immediately with the implementation of the system, without waiting the complete definition of the ISO standard. All the information stored with this standard will be used to migrate to the ISO standard with a simple conversion task.

⇨ *Data exchange*: waiting for SCAR directives, the choice is to make use of several proprietary standards (initially E00, DXF). In this way, the implementation of the system can proceed without having a degraded environment. In fact, the user data input requirements can be solved simply adding to the system all the needed data formats.

- **User Interface:**

⇨ *Data query*: graphical (geographical) and textual (alphanumerical);
⇨ *Data restitution*: graphical (geographical) and textual (alphanumerical).

Acknowledgement

This research has been supported by the 'Programma Nazionale di Ricerche in Antartide' (PNRA) for the period 1996-98, in the *Telerilevamento, GIS e Banche Dati* Thematic Area.

References

[1] S. Biagioni, C. Carlesi, O. Salvetti: "The South-Pole Information System for the access to distributed and heterogeneous information", VI Workshop on 'Antarctic Atmosphere', Vol. 51, Ed. M. Colacino et al., Firenze, 1995.

[2] Fortunati L.: "Geo-data server on WWW", Proceedings of the Joint European Conference and Exibition on Geographical Information, Barcelona - March 27-29, 1996, pp. 96-99.

[3] L. Fortunati, C. Magnarapa: "Server di dati geografici su World Wide Web", AM/FM 7ª Conf./Esp. It. 'L'Infrastruttura di Informazione geografica Italiana ed Europea', Pisa, 20-21 Giugno, pp. 189-195, 1996.

[4] S. Biagioni, C. Carlesi, L. Fortunati, O. Salvetti: "A Geo-Data Server for the South-Pole Information System", Proc. 2nd AGILE Conference on Geographic Information Science, Rome, 1999.

[5] S. Biagioni, C. carlesi, V. Costa, L. Fortunati, A. Galligani, O. Salvetti: "A Web-Based Geo-Data Server for Antarctic Data", Proc. 3rd AGILE Conference on Geographic Information Science, Helsinki, pp. 45-48, 2000.

[6] S. Biagioni, C. Carlesi, G. Leone, O. Salvetti: "National Antarctic Research Programme: the South Pole Information System", ERCIM NEWS, Paris (France), N. 22, July 1995.

[7] S. Biagioni, C. Carlesi, O. Salvetti: "The South-Pole Information System: Design and WWW Access Features", Fifth International W3 Conference, Paris, June 1996.

[8] S. Biagioni, C. Carlesi, O. Salvetti, V. Spoto: "A New Graphic User Interface to the South-Pole Information System", VII Workshop on 'Antarctic Atmosphere', Vol. 62, Ed. M. Colacino et al., Bologna, pp. 393-398, 1998.

[9] Fortunati L., R. della Maggiore, C. Magnarapa, "State of the art of geographical data formats", Esprit III Programme - Project no. 8017 - GEO2DIS/CNR/SR/WP131/01/V1.0 - 21/07/94.

[10] Fortunati L., R. della Maggiore, U. Sassoli, L. Raynal, H. Deheinzelen, "Synthesis of selected data sources", Esprit III Programme - Project no. 8017 - GEO2DIS/CNR/SD/WP131/02/V1.0 - 21/07/94.