SIRIA: the Information System of Italian Research in Antarctica

C. Rafanelli[1], M. Castorina[2], O. Salvetti[3], E. Benedetti[4], A. Damiani[1], M.G. Di Bono[3], M. Martinelli[3], E. Piervitali[1]

1 - CNR – ISAC Sez. Roma - Area della Ricerca "Roma – Tor Vergata" - Via del Fosso del Cavaliere, 100 – 00133 Roma
2 - ENEA - Centro Ricerche Casaccia - Via Anguillarese 301 - 00060 S. Maria di Galeria (Roma)
3 - CNR – ISTI - Area della Ricerca – via. G. Moruzzi, 1 – 56100 - Pisa
4 - CNR – ISAC Sede - Area della Ricerca – via P. Gobetti, 101 – 40129 - Bologna
Metadati-PNRA@isac.cnr.it

Abstract
The Italian Research in Antarctica Information System Project (Italian acronym SIRIA) is devoted to the collection and dissemination of information on Italian research in Antarctica. It is focused to preserve scientific data of PNRA (Antarctic Research National Programme), dating from as long ago as 1985. In addition SIRIA constitutes the Italian contribution to the Antarctic Master Directory, the international database of Antarctic data set descriptions, endorsed by SCAR.
This work presents the aim and the activity of the SIRIA Project.

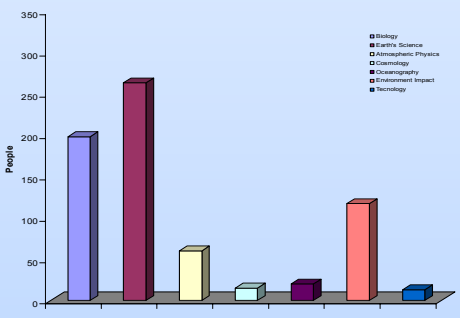

Fig.1 People involved in Antarctic researches in 1988

Introduction
Antarctic science is inter-disciplinary in character, multi-national in execution, and globally relevant. The Antarctic Treaty System states that scientific observations and results from Antarctica have to be exchanged and made freely available. In support of this topic, the Scientific Committee on Antarctic Research (SCAR) and the Council of Managers of National Antarctic Programs (COMNAP) have established the Joint Committee on Antarctic Data Management (JCADM). It is responsible for the Antarctic Master Directory (AMD), the internationally accessible, web-based, searchable record of Antarctic data set descriptions, the *metadata*. As part of the Antarctic Data Directory System the creation of National Antarctic Data Centres has been established, to catalogue datasets and provide information on datasets to the scientists and others with interest in Antarctic sciences.
The Italian SIRIA Project, devoted to collecting and disseminating information (metadata) on Italian research in Antarctica, constitutes the Italian contribution to the Antarctic Master Directory.
This work presents the aim, the structure and the activity of the SIRIA Project. It is focused to preserving scientific data of PNRA (National Programme Antarctic Research), managing the Antarctic Master Directory, recovering scientific data of Antarctic expeditions, dating from as long ago as 1985.
The Antarctic activity is subdivided into 11 scientific sectors: Biology and Medicine, Geodesy and Observatory, Geophysics, Geology, Glaciology, Physic and Chemistry of Atmosphere, Sun-Earth relationship and Astrophysics, Oceanography and Marine Ecology, Chemical Contamination, Legal Sciences and Technology. SIRIA Project collects information on scientific researches of all these sectors. The international standard used is CEN TC 287 and has been developed according to the requirement of so different scientific disciplines. In the Italian data base all the PNRA activities will be recorded, in Antarctica, in Artic and everywhere the PNRA had funded works; the Antarctic metadata will be selected and sent to the AMD.

Italian National Program of Antarctic Research
In 1985 the Italian Parliament authorized our country to develop a research program in Antarctica to acquire the right to enter into the Antarctic Treaty. The Ministry for University and Research implemented the Italian National Program of Antarctic Research (Italian acronym PNRA). The scientific activity was developed by Consiglio Nazionale delle Ricerche (CNR), involving Universities and other Scientific Institution, as the Istituto Nazionale di Geofisica e Vulcanologia (INGV) and the Osservatorio Geofisico Sperimentale (OGS). The logistic activity was developed by the Ente Nazionale per le Energie Alternative (ENEA).
In 2003 the PNRA Consortium was constituted, composed by the above institutions.
From 1985 to 2004, nineteen scientific expeditions were organized in Antarctica, one permanent base at Terranova Bay in the Ross sea region was realized and another base, located in Dome Concordia, has been constructed in cooperation with France. Several international research projects and technology activities have been developed.
During the first years of activity, the PNRA included a lot of scientific disciplines as Earth's Sciences, Biology, Atmospheric Physics, Cosmology, Oceanography, Environmental Impact and Technology. In the following years the researches were oriented forward multidisciplinary studies, focused on global phenomena of atmosphere, biosphere and geosphere.

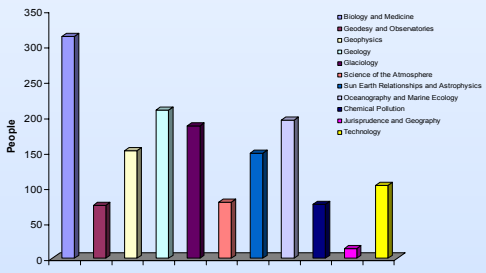

Fig.2 People involved in Antarctic researches in 2003

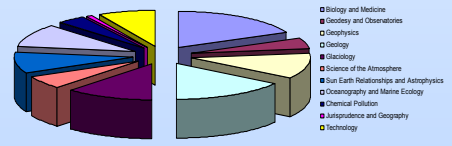

Fig. 3 Research groups for each scientific discipline as percent value

In 1988 the people involved in Antarctic researches were about 700 (fig.1), mainly concentrated in Earth's Sciences (more than 250), biology (about 200), Environmental impact (about 100). In this moment there are 11 scientific disciplines: Biology and Medicine, Geodesy and Observatory, Geophysics, Geology, Glaciology, Physic and Chemistry of Atmosphere, Sun-Earth relationship and Astrophysics, Oceanography and Marine Ecology, Chemical Contamination, Legal Sciences and Technology. In the recent years the number of PNRA researchers and technicians has considerably increased. In fig. 2 the situation for 2003 is shown. About 1550 are the people involved; more than 300 work in Biology sector, more than 200 in Geology, about 200 in Oceanography and Marine Ecology, about 190 in Glaciology and so on. The research groups for each scientific discipline are reported, as percent value in fig. 3. Of course the scientific sectors of "Biology" and "Geology" are composed by the higher number of working groups (17% and 16% respectively).

Metadata
Metadata provide descriptive information that characterize a set of quantitative and/or qualitative collected measurements. The metadata set-up is the best tool for documenting data and instruments involved in data collection. The fundamental metadata parameters give information about *when*, *where*, and *how* data have been obtained and *who* has collected them. Metadata or "data about data" describe the content, quality, condition, and other characteristics of data and constitute a resource for scientists and others with an interest in the sciences. Detailed descriptions of the dataset are necessary to make data effectively usable by other scientists, worldwide located, also in the future. The metadata are considered to be a set of attributes that help users to determine if a particular data set meets their needs. The scientific community of the PNRA participates in carrying out the Italian Antarctic meta-database. Such duties are performed inside the SIRIA Project.

Fig. 4 Home page of SIRIA project

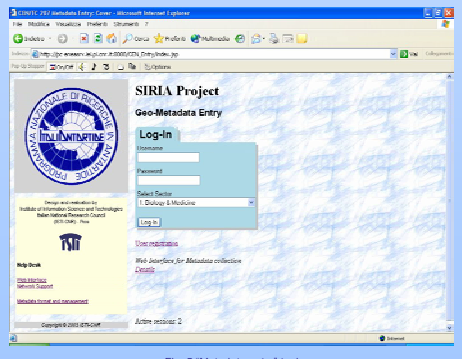

Fig. 5 "Metadata entry" tool

SECTION	DESCRIPTION
Dataset identification	this section gives information to clearly identify the dataset (title)
Dataset overview	this section gives information to present a comprehensive description of the dataset (abstract, purpose of production, references, related dataset)
Dataset quality elements	this section includes information about the data quality and accuracy; also the data elaboration techniques are described (spatial and temporal accuracy, completeness, data source)
Spatial reference system	this section collect information about the spatial distribution of geographical objects (type of reference system)
Extent	geographic datasets can be described by different extension types: planar, vertical and temporal extension. This section gives information about these different extension types (geographic extent, temporal extent)
Data definition	in this section the main characteristics of the geographic object are described in order to facilitate the comparison between two different datasets (object type, attribute type)
Classification	in this section some keywords can be reported (keywords)
Administrative metadata	this section gives administrative information to request the dataset: where it is preserved and how it can be ordered (information on organization, point of contact, data distribution)
Metadata reference	this section collect information about metadata (date of creation, update date)
Metadata language	in this section the language used to fill in metadata fields is indicated.

Tab. 1 Sections of CEN/TC 287

CEN Standard
Standardization represents a major concern in metadata collecting. In fact, a standard format allows us to describe datasets concisely and to easily compare datasets from different scientific disciplines as well.
The standard endorsed by the SIRIA project is the CEN/TC 287 (European Committee for Standardization / Technical Committee 287). According to CEN/TC 287 the basic metadata elements include the following principal sections (Tab. 1) and a set of associated attributes (fields).

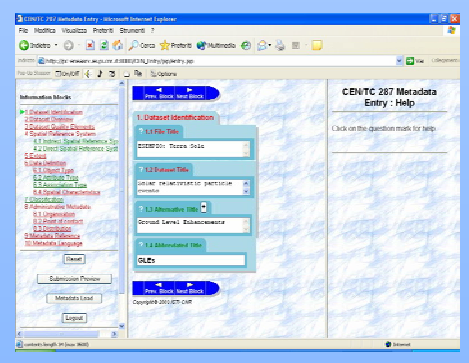

Fig. 6 The first page of the standard CEN/TC 287

Metadata submission and management
The SIRIA project, of course, privileges internet access for better cooperation and also provides international acknowledgment for scientists and their organisations. For this reason, the development of a metadata information system is particularly important for Antarctic science, which has an inter-disciplinary character and international relevance. The information system of Italian research in Antarctica has been developed to manage and disseminate data collected during the measurement campaigns.
The home page of SIRIA project, available at: http://nadc.isti.cnr.it:8080/SIRIA, is shown in Fig.4. It gives different information: aims and history of the project, staff, useful links etc. It is also possible to enter the *Geo-metadata Server*, the polar metadata-base, to retrieve and submit metadata.
The "Metadata Entry", a tool available only for registered users, is a temporary archive where raw metadata are stored before their validation, by the scientific supervisor. Then they are included in the metadata-base.
The standard CEN/TC 287 endorsed by SIRIA has been developed for geographic metadata. However, some scientific data of PNRA do not have a geographic reference. So it is difficult to describe these data using CEN/TC 287. For each PNRA sector, a customized interface has been realized taking into account the specific needs of the specific matter. After the log-in in the "metadata entry" tool, one of eleven interfaces can be chosen to submit metadata, as shown in fig. 5.
Fig. 6 shows the first page of the standard CEN/TC 287. It represents the "dataset identification" section with the relative fields: "file title", "dataset title", etc. In the left-hand column the complete list of CEN sections is reported. On the right-hand side an on-line help for each field is available.
In addition an *help desk* can be used for any problem or information, in particular for help or suggestions about dataset identification and metadata filling in: Metadati-PNRA@isac.cnr.it

Functionalities and technical aspects
The database of officially approved metadata, as above indicated, and the web interface are installed on the CNR - ISTI server in Pisa (Italy) and managed by local task force.
The general structure of the system is shown in figure 7.
The access to the server is both for storing/updating and inquiring.
In the first case, a registration and authentication procedure for users is required for restricted members of PNRA scientific projects. The user inserts metadata using the available forms and saving, in a multi-session way, such information in a specific area operating as a buffer. It is indicated as *area A* in fig. 8. The metadata saved as "final submit" are reviewed by the scientific supervisor for a validation process and then definitively stored in a distinct area of the server (*area B* in the same figure). Finally by an automatic mapping procedure, under ISTI responsibility, the files are moved in the relational database. This is the database available for inquiring.
In the second case, the interested people use the web interface to inquire the metadata collection catalogue. The on-line metadata searching, according to complex specific criteria, is based on an "*Incremental Resource Discovery*" mechanism.
To use SIRIA tools just a web browser is necessary, because the functionality of the system needs the most common applets implemented on personal computers. A *Java* applet manages the geographic area selection and specific information storage; a *JavaScript* code checks the inserted data, types and constraints; a *Java Servlet* and *Java Server Pages* (JSP) technologies are implemented for dynamic production of web pages and *business logic* development. Finally a *Cascading Style Sheets* (CSS) have been utilized for the graphical visualization of the forms.

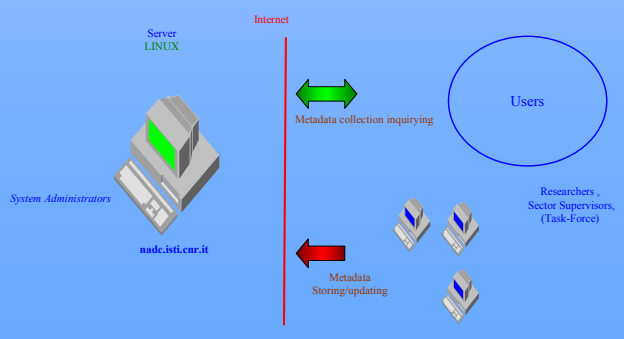

Fig. 7 General system architecture

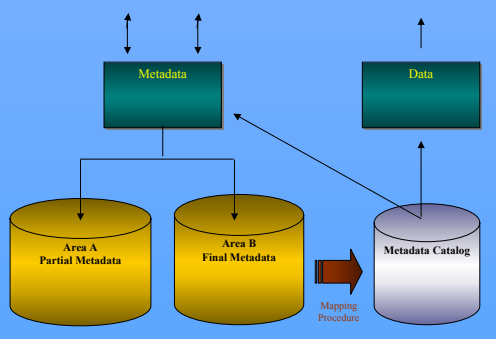

Fig. 8 Metadata organisation on the server

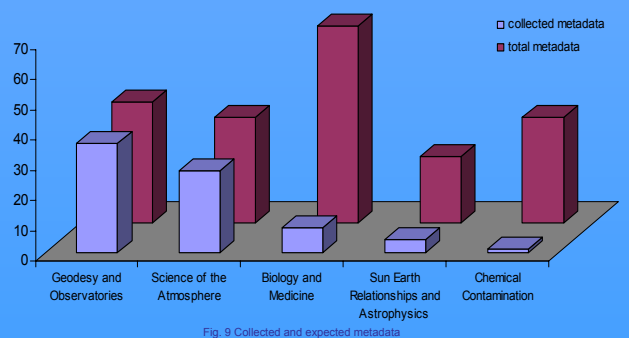

Fig. 9 Collected and expected metadata

Collected metadata: preliminary results
During the first year metadata collection activity has been devoted to 5 PNRA sectors: Geodesy and Observatories, Science of the Atmosphere, Biology and Medicine, Sun Earth Relationships and Astrophysics and Chemical Contamination. To reach this goal the staff of the project took part in different scientific workshop (Roma, Messina, Venezia) where several PNRA researchers were present. The information system has been presented and the modality of use has been explicated.
Presently, we have collected about 80 metadata as shown in fig. 9. In this figure for each sector collected and expected metadata are reported.
With reference to the sectors Geodesy and Observatories and Science of the Atmosphere, the recording work has been almost completed: respectively 36 (90 %) and 27 (77 %) metadata have been collected.

Concluding remark
This work presents the activity of the SIRIA project, the metadata Information System of Italian Research in Antarctica.
The first year of activity was spent in realizing and testing the system according to the specific needs of the workers involved. Presently five research sectors are collecting metadata; the remaining sectors will be operative during the second year of activity.

Acknowledgements
This paper and the SIRIA Project are under PNRA funding. The authors are grateful to the late Mario Zucchelli for his encouragement to start the SIRIA project.

References
Cirocchi G., Gatta S., Panciera L.,Seta E., 2000: "Metadati, informazione di qualità e conservazione delle risorse digitali", *Rivista italiana di biblioteconomia e scienze dell'informazione*, 40, 309-327
Colacino M., Rafanelli C., 2003: "Italian Research on the Antarctic Atmosphere", Annals of Geophysics, 46, 59-67
De Robbio A., "Metadati per la comunicazione scientifica", *Seminario ICCU "I metadati: seminario nazionale". Roma, 3 aprile 2001*
http://www.cenorm.be/cenorm/index.htm
http://www.comnap.aq
http://www.fgdc.gov/standards/standards.html
http://www.jcadm.scar.org
http://www.pnra.it
http://www.provincia.tn.it/urbanistica/siu /metaasp/main.htm
http://www.rete.toscana.it/sett/sit
http://www.scar.org
Timpf S., Raubal M., Kuhn W., 1996: "Experiences with metadata", *7th Int. Symposium on Spatial Data Handling, SDH'96, Delft, the Netherlands (August 12-16, 1996)*, IGU, 12B.31-12B.43
http://www.aib.it/aib/commiss/cnur/dltridi.htm