

ACTIVITIES AND ISSUES OF A DEVELOPED INFORMATION SYSTEM FOR THE ITALIAN POLAR RESEARCH

A Damiani^{1}, C Rafanelli², E Benedetti², and M Storini¹*

¹*Institute of Physics of Interplanetary Space, National Institute of Astrophysics, Rome, Italy*

*Email: alessandro.damiani@ifs-roma.inaf.it

²*International Center for Earth Sciences c/o Institute of Acoustics "O. M. Corbino", National Research Council, Rome, Italy*

ABSTRACT

Activities performed to develop an information system for the diffusion of Italian polar research (SIRIA project) are here described. The system collects and shares information related to research projects carried out in both the Antarctic (since 1985) and Arctic (since 1997) regions. It is addressed primarily to dedicated users in order to foster interdisciplinary research but non-specialists may also be interested in the major results. SIRIA is in charge of managing the National Antarctic Data Center of Italy and confers its metadata to the Antarctic Master Directory. Since 2003, the National Antarctic Research Program has funded this project, which, by restyling its tasks, databases, and web site, is becoming the portal of Italian polar research. Issues concerning data management and policy in Italy are also covered.

Keywords: Metadata, Data policy, Data management, Antarctica

1 INTRODUCTION

Italy signed the Antarctic Treaty on 18 March, 1981, and achieved consultative status in 1987. It is involved in polar research through the National Antarctic Research Program (PNRA) of Italy, which manages the logistic activities of expeditions. Italy has been operating in Antarctica since 1985 and in the Arctic since 1997. Currently two Italian bases exist in Antarctica: the summer station Mario Zucchelli (74°41' S 164°07' E), located in Terra Nova Bay and Concordia Station (75°06' S e 123°21' E - 3.233 km MSL), at Dome C, which is open all year round and managed in agreement with France. In the Arctic, the station Dirigibile Italia is sited at Ny Ålesund (78°55' N, 11°56' E - 0.010 km MSL), Svalbard Islands, Norway, and is handled by the National Research Council. More than 1500 Italian researchers are involved in polar research, and they are spread over the national territory in almost all research institutes and universities. In spite of the rather late start of Italian activities in the polar regions in comparison with other countries, a great deal of scientific results has been gained, much of it due to an increase in international collaboration. Hence the Italian community is engaged in several different international programs, for example, the International Polar Year (IPY) and the International Heliophysical Year (IHY) (see for example Allison et al., 2007 and Thompson et al., 2008) to name a few. In this way, a huge amount of data from several scientific fields is produced every year.

Technological development of the last few decades has led to new reliable scientific instruments, which are able to perform many complex and precise measurements. Also, the spread of the Internet allows the transmission of information on a planetary scale. If on the one hand this has given a great impetus to science by allowing researchers to obtain the data necessary to understand many phenomena, on the other hand, the production of a huge amount of data gives rise to several issues, in particular where and how all these data can be stored and how a particular data set can be identified. Although modern information systems have proved to be powerful tools with which to manage a large amount of data, a precise cataloguing is necessary in order to give them an effective accessibility. This can be achieved through metadata, a set of structured information that helps the user to determine if a data set meets his needs more effectively than if he examined all the data sets. In addition, it makes it possible for data stored in one location to be visible to users from all over the world¹. Metadata include

¹ See for example the activity of the Open Archives Initiative (<http://www.openarchives.org/>) which promotes interoperability standards that aim to facilitate the efficient dissemination of content.

information on the goodness of the data, who, where, when, and how the data have been obtained, and where they are stored. Thus metadata represent an essential data complement. For this reason, since 2003, the Information System for the Italian Research in Antarctica (Italian acronym SIRIA) has been dedicated to collecting and sharing information (metadata) on polar research projects and data sets connected to Italian expeditions in Antarctica that have been taking place since the eighties. This meta-database involves the eleven research sectors that constitute the National Antarctic Research Program: Biology and Medicine, Geodesy and Observatory, Geophysics, Geology, Glaciology, Physics and Chemistry of the Atmosphere, Sun-Earth Relations and Astrophysics, Oceanography and Marine Ecology, Chemical Contamination, Legal Sciences, and Technology.

2 ISSUES OF DATA SHARING IN POLAR SCIENCES

Because of the difficult working conditions, from the beginning, the polar sciences have exemplified excellent international collaboration, both in logistic and scientific issues. Therefore, each country carries out its activities within a larger team and is always ready to help others by making available its resources and facilities. In addition, the interdisciplinary character of the performed research is an essential element in improving our knowledge of the complex processes governing these regions. Indeed, some of the greatest advances in science have occurred when the knowledge of a scientific field has become available to others also.

In the framework of the polar sciences, the Antarctic Treaty² (section III.1.c) clearly states that “Scientific observations and results from Antarctica shall be exchanged and made freely available.” This simple sentence means that the old fashioned point of view of the researcher who owns the data he has collected must be substituted for a new *modus operandi*.

Therefore, scientists must make their data available to an external audience as soon as possible. A first step might entail releasing the data in conjunction with scientific publications at the time of publication (Reinke, 2000). This would greatly speed up the repetition of experiments with enormous advantages especially for young researchers. In this regard, it is instructive to remember that some important scientific journals use this method. For example, *Science*, one of the most important and famous scientific journals, demands that before publication of a paper, the related data must be made easily available to readers by their deposition in an approved database with an accession number provided for inclusion in the published paper³.

It is necessary to meet the selfish needs of a career with the nobler need to increase human knowledge. This aim is pursued by making the instruments available to realize this increase in developing countries in particular. Hence, the slow growth of science and technology in these countries should be strongly accelerated by a large network of capacity building made available by the international scientific community of the so called developed countries (Damiani et al., 2008). Unfortunately, research institutions find little time to devote to educational and capacity building aspects in other countries.

In Italy, diffusion of scientific knowledge is carried out mainly by specialists in dissemination who are non-specialists in science. The result is an excessive simplification of information that is well-accepted by newspapers or even by schools but not exhaustive enough for university studies. Because of this, students learn about scientific achievements of their country only through publications or university courses. In the first case, the scientific works are often exceedingly specific and not understandable by non-specialists. In the second case, the professors are often not aware of the most recent scientific results not strictly related to their exact field of interest. In this context, scientists directly involved in such research should also be interested in making their results available to a larger audience. Unfortunately, a very successful research career needs only publications; therefore there is no time or call for activities such as dissemination, information systems building, virtual observatories (VO), databases (DB), and other capacities. At present, the international research community is contemplating this problem and seeking a solution. Building a data set citation system⁴ would solve the problem, at least partially (Finney, 2009). In journals, the data set citation would follow the traditional citation

² Information on the Antarctic Treaty is available on the Scientific Committee on Antarctic Research web site at: <http://www.scar.org/treaty/>.

³ See the conditions of acceptance and the database deposition policy of Science available at: <http://www.sciencemag.org/about/authors/>.

⁴ Information on how to cite a data set can be retrieved at: <http://ipydis.org/data/citations.html>.

in order to point out not only the reference paper but also the data set source from where the data came. In this way, a clear and detailed reference would give the data set larger visibility and contribute to enhancing the career of the data set staff similar to a traditional reference.

Nevertheless, the necessity of a new figure of researcher-popularizer is emerging. In order to reach these goals, the researcher should be in position to operate in this way. To be the earliest to publish in a specific field is actually the most important measure of scientific performance for a research worker, but this does not encourage free data exchange. Even if this aspect is unlikely to be overcome, others could nudge the researcher toward a more opened-mind vision. In order to evaluate research work and change this old approach, not only the number of publications but also education and capacity building need to be taken into account. To pursue this aim, it is essential to reduce the distance between researchers and technicians and work together to build information systems that are able to ease access to data. This access will be crucial to the future development of science and society (see for example the activity of the real-time Neutron Monitor database [NMDB] project supported by the Seventh Framework Programme of the European Commission: Steigies et al., 2008).

An accurate preservation of the acquired data also allows future access, which will give precious new information to scientific research. For example, consider the importance of historical data series in climatologic studies. The present work of researchers has thus a higher horizon than that linked to their particular study. It is necessary to make scientists more sensitive to the issue of data preservation, including the need for data storage with appropriate supports that have to be updated periodically following technological evolution and the need for metadata that give basic information on data sets. Under these perspectives, the activity of capacity led by scientists who are expert also in technology and data management is particularly important. These new professional figures, called “X-Men,” are a good opportunity for young researchers, especially in Italy, where the funding for basic research is rather complicated. In this way, the chance to work in a research field would increase for students with scientific background, the numbers of whom are actually decreasing among the Italian universities because of the lack of job opportunities⁵.

In this context, the main difficulty for an information system such as SIRIA is continuous funding over a long period of time. In such a period when, also due to IPYs initiatives, everyone realized the importance of sharing information, the majority of research projects begin to take data management into account. However, after the funding ends, the question of who will guarantee data availability and how this can be accomplished over time arises. For this reason at least, an “official” data holder must be present in every research body, so the existence of initiatives such as SIRIA inside the Italian Nation Antarctic Program remain of vital importance. Because of this data holder, all data need not be physically hosted by the “official” holder. This structure should help research groups share their information by making their know-how available and managing the data directly when required (e.g. managing data of closed projects or of groups that do not have managers themselves). In addition, data quality control should be another duty of SIRIA, in full cooperation with a desirable network of polar data centers.

3 SIRIA PROJECT

The present project, sponsored by the National Antarctic Research Program of Italy, aims at collection, management, and dissemination of information about the scientific research projects carried out in Antarctica and funded by PNRA since the beginning of 1985. It consists of a database of metadata, allowing whoever is interested in polar sciences to get the basic information on the various research projects and results.

From the beginning, SIRIA has worked in close collaboration with the international polar research community. During the second half of the 1990s, The Scientific Committee on Antarctic Research (SCAR) decided to create an Antarctic Data Directory System (ADDS) (Figure 1) in order to share research information. This system

⁵ The Statistical Office of the European Communities (EUROSTAT) on the report “Human Resources in Science & Technology” shows that in Italy the human resources in science and technology as a share of labour force – Total for the 2007 is 35.56% compared with the European mean of 39.25%. In this specific statistic the countries more virtuous are (from first to fourth place) Switzerland, Netherlands, Finland and Denmark all roughly around 50%. Data are available at: <http://epp.eurostat.ec.europa.eu/> on the “Science and technology” section.

contains a central Antarctic Master Directory (AMD)⁶ managed by the Global Change Master Directory of NASA and a network of National Antarctic Data Centers (NADC) that send metadata to the AMD. To manage this system, a Joint Committee on Antarctic Data Management⁷ (JCADM) was established by SCAR in 1997. This committee is still the international organization closest to SIRIA. In order to receive the recommendations of the committee as soon as possible, the Polar Institute of SIRIA is also the Italian member of JCADM.

SIRIA is managed by a Steering Committee (SC); in addition there is a Scientific Referee Committee (SRC) made up of the chiefs of eleven sectors funded by PNRA. The SRC's task is validation of the metadata submitted by the researchers. Finally a Task Force (TF) is responsible for helping scientists fill in the metadata forms and keeping in touch with the different bodies of the project.

At first, the main tasks of SIRIA were:

- Managing the National Antarctic Data Center
- Updating the Antarctic Master Directory
- Preserving the acquired data in the past Italian Antarctic expeditions

At present (July 2008) more than 200 metafiles are held in the database. The SIRIA project constitutes the Italian NADC and, as recommended by SCAR, confers its metadata to the Antarctic Master Directory, the internationally accessible, web-based, searchable record of Antarctic metadata. The majority of Italian metadata is included in the AMD (about 150) and can be retrieved on its web site.

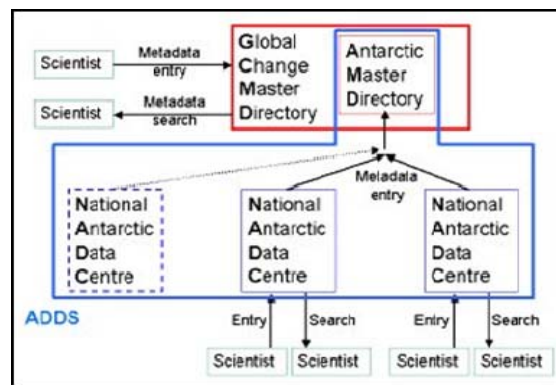


Figure 1. Antarctic Data Directory System (Source: <http://www.jcadm.scar.org/>)

Many metadata standards have been developed to describe geographic data sets. For example the Content Standard for Digital Geospatial Metadata (CSDGM) approved by the Federal Geographic Data Committee, the Directory Interchange Format (DIF) adopted by the AMD, and the ISO 19115 created by the International Organization for Standardization in 2003. The standard adopted by SIRIA is the pre-standard ENV 12657, developed by the European Committee for Standardization (CEN), Technical Committee (TC) 287, which shows similarities with the more recent international standard ISO 19115 (received in Europe as EN ISO 19115:2005). The ENV 12657 is subdivided into several sections; each of these contains mandatory and conditional fields. Hereafter the main sections are briefly described:

- Data set Identification: this section gives information to identify clearly the data set (title).
- Data set Overview: this section gives information to present a comprehensive description of the data set (abstract, purpose of production, references, and related data sets).
- Data set Quality Elements: this section includes information about data quality and accuracy; also the data elaboration techniques are described (spatial and temporal accuracy, completeness, and data source).

⁶ The Antarctic Master Directory portal makes available all the metadata sent by the international community at <http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=amd&MetadataType=0>.

⁷ Information on the activity of Joint Committee on Antarctic Data Management (JCADM) and on the management of the Antarctic data is available at: <http://www.jcadm.scar.org/>. In 2009 JCADM was elevated to a SCAR Standing Committee, namely SCADM (Standing Committee on Antarctic Data Management).

- Spatial Reference System: this section collects information about the spatial distribution of geographical objects (type of reference system).
- Extent: this section gives information about different extension types (geographic extent and temporal extent). Geographic data sets can be described by different extension types: planar, vertical, and temporal extension.
- Data Definition: this section describes the main characteristics of a geographic object in order to facilitate the comparison between two different data sets (object type and attribute type).
- Classification: this section reports keywords (keywords).
- Administrative Metadata: this section gives information about data set storage, format, and distribution (information on organization, point of contact, and data distribution).
- Metadata Reference: this section gathers information about metadata (date of creation and update date).
- Metadata Language: this section indicates the language used to fill in the metadata fields.

Because some fields do not apply to all scientific disciplines, for each sector, the SRC has established the set of metadata elements pertaining to the particular sector, defining a specific metadata format. For simplicity, a Metadata Entry web tool has been realized and can be accessed through the SIRIA web site (<http://nadc-pnra.artov.rm.cnr.it>). The researchers enter a web interface that allows them to fill up and store the metadata directly into a database, after their registration and authentication. While registering, the researcher specifies his sector and is promptly directed to that specific interface. In each field, a customized help describes the characteristics of the field (maximum length, compulsoriness, etc.) and explains its meaning. In addition, a help desk can be contacted via e-mail for any problem or information. Metadata forms can be partially filled out and then saved in a specific text file by choosing the “partial filling” modality. These files can be successively reloaded to be completed or corrected and then saved using the “final submit” modality. The collected metadata are subject to two types of validation checks. First, the TF makes a check on the formalism of the metadata standard. Then, the SRC for each sector validates the metadata for a definitive submission. To use the SIRIA tools, only a web browser is needed. All additional technical requirements may be downloaded following the links in the webpage of the Metadata Entry.

4 WORK IN PROGRESS

Taking into account the above-mentioned demands (see section 2), since 2007 and especially in the last two years, SIRIA has started a restyling process, which has encompassed other activities.

Some changes on the SIRIA web site (Figure 2) have been performed and others are under evaluation. First of all, the SC decided to include in the data center data collected in Arctic regions imitating the SCAR tendency to combine the research of both Polar Regions⁸. The referred data set of the submitted metadata is often hosted not by an international data center but by a smaller information system managed directly by the involved group. Therefore, in order to give major visibility to Italian Antarctic projects, we inserted a new section in the SIRIA web site devoted to giving the users a summary of the data sets produced by the different groups. It consists of external data sets that are not directly managed by the SIRIA staff but often have great scientific importance. In this way, we foster interdisciplinary studies.

The architecture of SIRIA is based on the MySQL Database Management System (DBMS) with a LINUX Operating System and a JAVA applet. In addition, text files are collected in order to send the information to the AMD. Software has been built to convert the information collected through the ENV metadata standard to the DIF standard and to create the XML files accepted by NASA. Currently, we are working to leave the Java applet and endorse a PHP application to obtain better data encodings and format. The simple text file is not adequate to exchange and transfer information. Therefore, we decided to endorse the Extensible Markup Language (XML), the tag language created and managed by the World Wide Web Consortium (W3C) and used by the majority of data centers to encapsulate the data. This will facilitate interaction with the AMD.

In addition, we are working to move the metadata standard from ENV 12657 to ISO 19115. In fact, even though we updated our standard with the changes found in the DIF, which in turn had been updated with the changes in

⁸ St. Petersburg (Russia) hosted the first SCAR/IASC IPY Open Science Conference “Polar Research – Arctic and Antarctic Perspectives in the International Polar Year” from July 8th to 11th 2008.

the ISO, we decided to change standards to avoid problems. In this way we are able to receive the EU dispositions⁹ and also the JCADM recommendation.

We are trying to increase the visibility of the SIRIA web site and expand the target audience. Actually, the SIRIA web site is only in English. We are planning a short version of the web site in Italian and the double language of the metadata core elements (e.g., title, summary, and investigator) in order to stimulate students to access our information system. Moreover, due to the need for administrative information from the PNRA, some additional information will be requested inside the next version of metadata format.

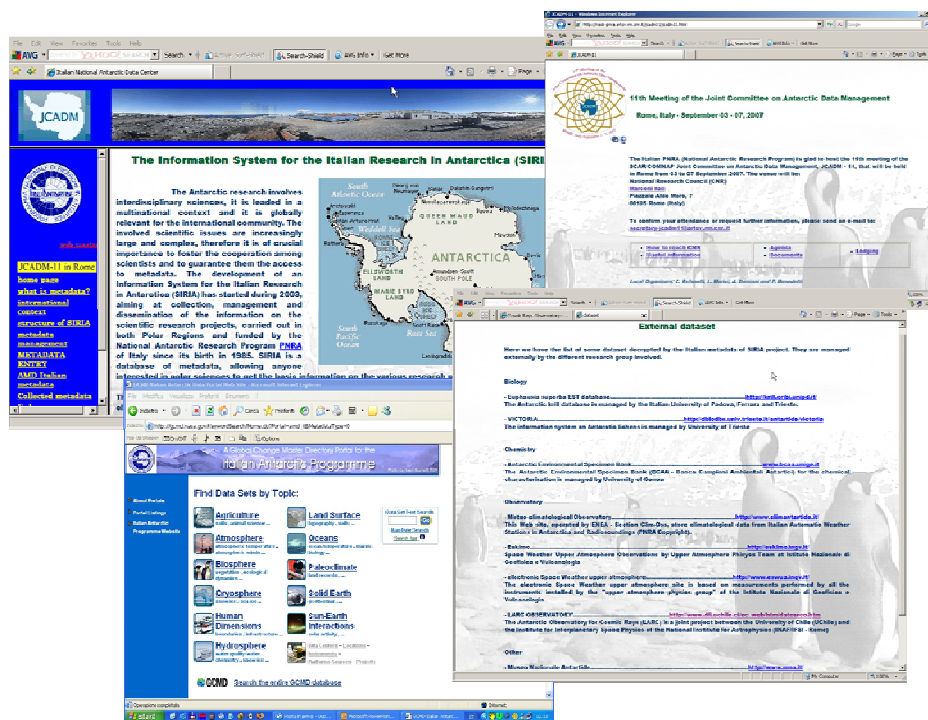


Figure 2. The SIRIA web site: from top left clockwise - home, the JCADM 11 meeting page, the external data set, and the Italian NADC hosted by the AMD

Finally, the SIRIA team will foster educational activities and outreach actions in order to promote interdisciplinary studies in both Polar Regions and to make researchers more sensitive to metadata as an instrument to improve data set accessibility and preservation.

The purpose of all these changes is to create a portal of the Italian polar research that can be interesting and useful both for specialists and students. The challenge is to contribute to the creation of open data access in Italy. A similar philosophy could reap many benefits for the scientific community, e.g., encouragement of different analyses and opinions, easy reproduction of preview results, improvement of education of young researchers, promotion of capacities for developing countries, interdisciplinary research, etc. (see Uhler & Schröder (2007) for a detailed analysis of this topic). In addition, open data access is a more reasonable option in comparison with a closed system. This is especially important in a recession period, such as what we are now experiencing, when central governments are not giving a great deal of funding for research and technology¹⁰. An open data system might seem unrealistic in the near future, but we must take into account the many models and initiatives

⁹ Europe received the ISO 19115 in 2005 as EN ISO 19115:2005 Geographic information.

¹⁰ Europe is still working to achieve the target set by the Lisbon Strategy of devoting 3% of GDP to research and development (R&D) activities by 2010. In 2006 the R&D intensity of the EU-27 was of 1.84% of GDP. The highest R&D intensities, above the 3% target, were achieved just by Sweden (3.73%) and Finland (3.45%). Next countries in the chart like Germany (2.51%), Austria (2.45%), Denmark (2.43%), and France (2.12%) are percent below the Lisbon Strategy target. In the global context, Japan (3.20%) and the United States (2.62%) continued to register higher R&D intensities than the EU-27 average. Data are from "R&D Expenditure and Personnel" Issue number 91/2008 of EUROSTAT.

that are already present in our society, such as a free encyclopedia with articles contributed collaboratively, e.g., Wikipedia; open-source software, e.g., Linux; an open access journal,¹¹ e.g., the Data Science Journal; an open archive under a Creative Commons license,¹² e.g., Earth-Prints¹³, a World Data Center (WDC), a Virtual Observatory, etc. We have already tested the majority of these facilities and appreciate the enormous advantages they give (see for example Lawrence (2001)). The next challenge is to make clear rules for their establishment and development.

5 A DATA POLICY PROPOSAL

Figure 3 shows the collected metadata per sector of the Polar Italian Data Center (updated on July 2008) and the AMD (updated in 2007, courtesy Melanie Meaux). Inside the Italian database, we have ~200 metadata subdivided into eleven research fields. “Geodesy and Observatory” is the largest (21.5% of the total); then are “Oceanography” (19.0%), “Physics and Chemistry of Atmosphere” (17.1%), “Sun-Earth Relations and Astrophysics” (11.2%), and so on. We note some differences in the nomenclature of Italian topics compared with AMD ones. However, these differences do not occur in the typology of the conducted research but only in the methods of subdivision, which, for the Italian DB, depend also on the PNRA administration. Italian topics such as “Legal Sciences,” “Technology,” and “Astrophysics” are submitted with difficulty to the AMD. In fact, it needs to be taken into account that the AMD is a data set devoted to collecting Antarctic metadata related to Earth Science, whereas the Italian data center collects different kind of information also, as it holds all information related to Italian expeditions in Antarctica. Thanks to close collaboration between the Italian NADC and the Central AMD, the need for including “Astrophysics” metadata inside the AMD has been realized. Therefore, a new Antarctic Astrophysics Data Set has become available in the spring of 2008 on the AMD web site. Actually, the majority of what it hosts is Italian astrophysics metadata, but we believe that in a short time, other countries will also decide to insert their similar data. This will foster interdisciplinary research, e.g., studying the character of new Antarctic sites in which to set up telescopes.

It is interesting to consider the number of file per country in the AMD (statistics are not shown here). Currently, almost all countries doing research in Antarctica give their metainformation to the AMD. Generally the countries with an older Antarctic studies tradition have given about one hundred metadata to the AMD. Australia and the USA, however, have given more than one thousand metadata (USA ~ 1200 and Australia ~ 2000). Why does this great difference exist? On the one hand, it is not a surprise to see the USA at the top of this list because of the great amount of its investment in Antarctica. On the other hand, Australia’s contribution could appear peculiar as it is a country that has conducted research in Antarctica for a long time but with funding almost comparable to that of other countries. One reason may be the presence of a good data policy in both countries¹⁴. Analyzing the data policy of the signatories of the Antarctic Treaty is beyond the scope of this article, but it is important to consider some of their policy elements. First, the policies of both the US and Australia point out the importance of metadata as a means of data retrieval and define the necessity of researchers sending these metadata to data centers as soon as possible, usually before the end of the project. In addition, it is clearly stated that within two years of data collection, all data and associated metadata must be

¹¹ Open access movement: Putting peer-reviewed scientific and scholarly literature on the internet. Making it available free of charge and restrictions (e.g., copyright, licensing). For further information see Open Access News web site available at: <http://www.earlham.edu/~7Epeters/fos/fosblog.html>. To find open journals in several subjects see the Directory of Open Access Journals (DOAJ) hosted by Lund University, Sweden available at: <http://www.doaj.org>.

¹² Creative Commons is a nonprofit corporation which provides free licenses and other legal tools in order to increase the amount of creative work available to the public for free and legal sharing, use, repurposing, and remixing. For further information see Creative Commons web site at: <http://creativecommons.org/>.

¹³ Earth-Prints is an open archive created and maintained by Istituto Nazionale di Geofisica e Vulcanologia devoted to collect research results in the fields of Atmosphere, Cryosphere, Hydrosphere and Solid Earth. It is available at: <http://www.earth-prints.org/>.

¹⁴ The USA data policy “Guidelines and Award Conditions for Scientific Data” available at <http://www.nsf.gov/publications> provides the guidelines from the Office of Polar Programs at the USA National Science Foundation.

The “Australian Antarctic Program (AAP) Data Policy” is available at the Australian Antarctic Data Centre at: <http://data.aad.gov.au/>.

submitted. Finally, a third element is the presence of “disadvantages” for scientists who do not execute these duties. These disadvantages could be, for example, a postponed evaluation or denial of new grant money.

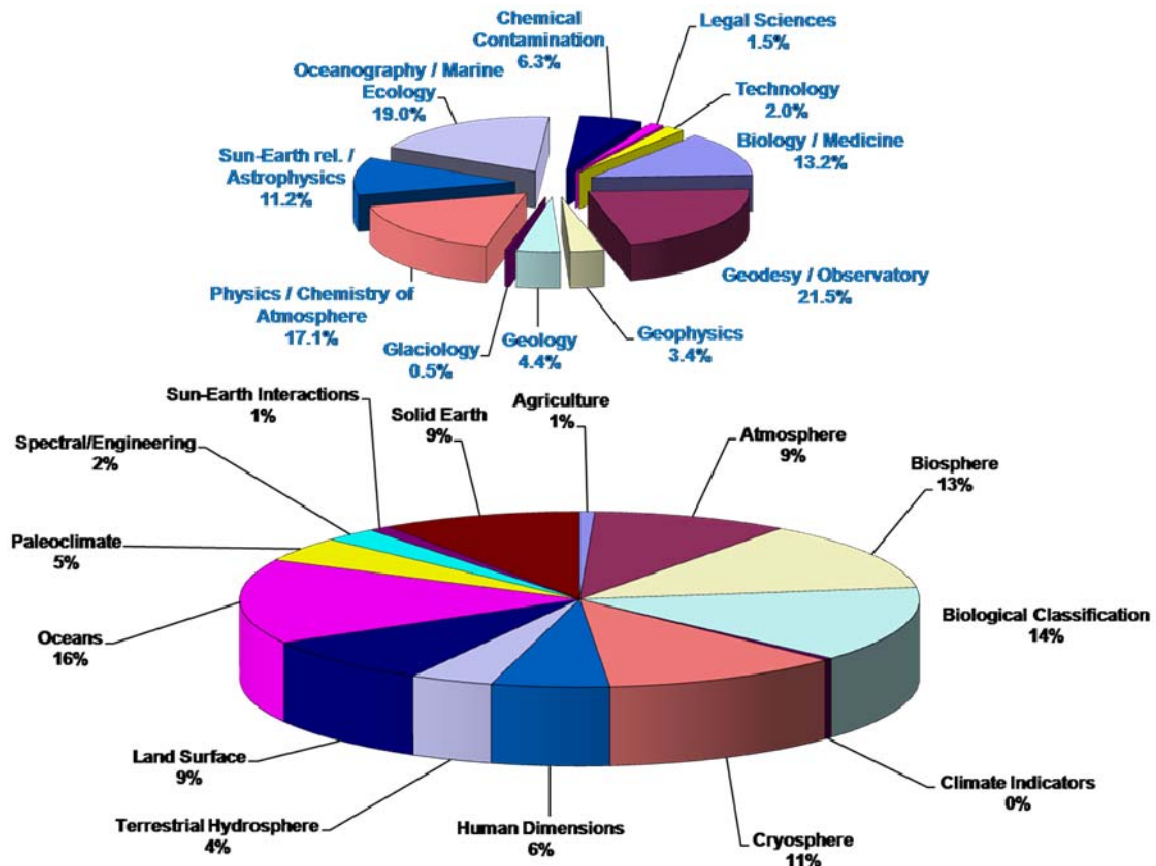


Figure 3. Collected metadata per sector. Upper panel: Polar Italian Data Center. Lower panel: Antarctic Master Directory (courtesy Melanie Meaux)

Figure 4 shows the growth of Italian metadata in SIRIA during the past few years. It is interesting to note the sudden increase just after the end of 2007 when a proposal referring to PNRA projects spread: only PIs that sent the metadata of their projects to the data center were able to obtain funds for future research. Today this proposal is still under evaluation.

However, a detailed data policy for projects funded by PNRA has not yet been organized. We have some recommendations for these projects:

- data must be submitted within 6 years project completion.
- metadata must be submitted by the end of the project

In particular the first recommendation is not in line with the international references, whereas the second is sometimes neglected. Even if in the majority of other countries a period of data publication exclusivity is given to the researchers who collect the data (see for example the NERC data policy Handbook¹⁵ of the Natural Environmental Research Council of UK), it is easy to understand that six years is too long to have sole use of research data. Nevertheless, important and useful information systems are present within the Italian polar community and are available also through the “External Data Set” page of SIRIA, for example, the Antarctic Environmental Specimen Bank, the Meteo-climatological Observatory, the LARC Observatory for cosmic rays, the VICTORIA information system on Antarctic lichens¹⁶, and so on. The electronic Space Weather upper

¹⁵ The NERC Data Policy Handbook is available at: <http://www.nerc.ac.uk/research/sites/data/policy.asp>.

¹⁶ The above mentioned data sets are available as followings: Antarctic Environmental Specimen Bank at <http://www.bcaa.unige.it>, the Meteo-climatological Observatory at <http://www.climantartide.it/>, the LARC

atmosphere (see <http://www.eswua.ingv.it/>) project (Romano et al., 2008), a system to standardize historical and real-time observations for different instruments, has significant relevance even if at present it is just for atmospheric and space weather purposes. Only the Italian Antarctic Museum (<http://www.mna.it/>) is devoted to collecting data sets primarily on rock and meteorites and secondarily on biology and chemistry. This constitutes a portal for data sets coming from projects funded by the PNRA. It is necessary to create a large portal to access all Italian polar research for the establishment of a common Italian data management for polar research. We believe SIRIA could take charge of this role.

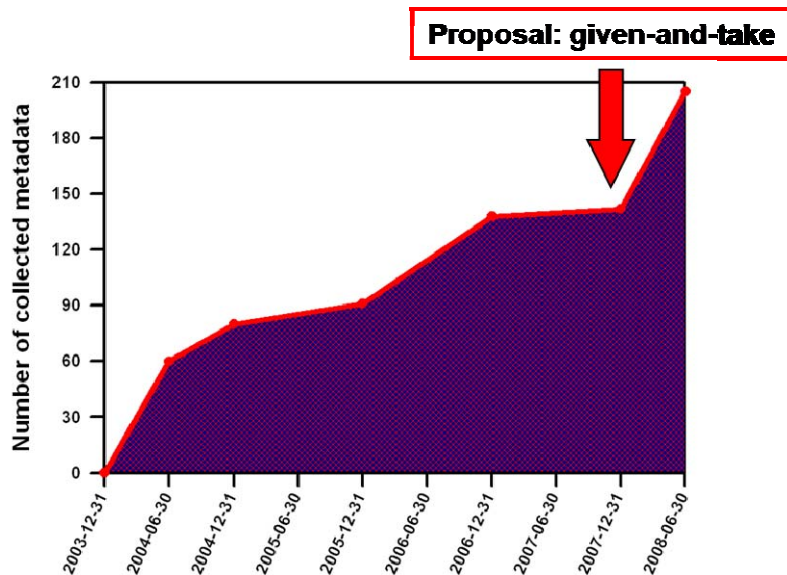


Figure 4. Italian metadata growth

However, the many problems slowing down the establishment of a data policy for polar research in Italy must be taken into account. Some issues apply only to Italy, others are also common abroad.

People working in every research field know very well that to be the earliest to publish on a specific topic is the measure of scientific performance. Therefore, scientists want to be absolutely sure they have completely exploited their data before making them available for public access. In addition, especially for international projects, the funds often come not only from the PNRA but from different sources, and this aspect contributes to complicating the modality of an extensive data policy.

An additional problem is the lack of a polar institute in Italy. This factor makes it more difficult to establish a data policy because scientists belonging to different research institutions often have different laws and regulations and it is not easy to meet the differing needs. A single research institution, however, has the possibility to create a data management plan more specific for its requirements and should be able to instill new behaviors in its employees. A good example of this can be found in the NERC Data Policy Handbook, which illustrates that since 1996 the institution considers it a necessity to have data available to support its mission of research, survey, and monitoring (Thorley, 2006).

A further problem is data acknowledgment. Since the first International Geophysical Year (IGY) in 1957, the need for free and open data exchange has been emphasized. The WDC system was established to answer this need by assuring the availability of data to all scientists. Today after fifty years, some problems are not yet solved. Especially in Italy, research groups are often not very large. Because of funding problems, the role of data acknowledgment is very important for these groups to have external visibility. Because of this, PIs are often not inclined to send their data to the WDCs. On the one hand, the importance of these data centers is recognized;

Observatory for cosmic rays at http://www.dfi.uchile.cl/ec_web/htm/datosrco.htm, the VICTORIA information system on Antarctic lichens at <http://dbiodbs.univ.trieste.it/antartide/victoria>.

on the other hand the groups see their visibility, essential for their survival, disappearing. This is probably an old fashioned point of view but the problem is absolutely real. It could be solved simply by giving attention to the people which collect the data. An extensive acknowledgment regarding not only the Data Center but also the original data “owner” would help to solve the problem.

Finally, special attention should be paid to data preservation. This issue is highly connected with the possibility of long term funding. As already explained in section 3, it is very difficult to keep stable funding to manage an information system, but this funding is essential in insuring data preservation and the ability to access the data in the future.

The SIRIA project intends to propose a data policy for the Italian polar research community, and the model should be the IPY data policy¹⁷. The IPY data management must ensure the security, accessibility, and free exchange of relevant data (ICSU note), and these objectives must be received also by the Italian polar community. In this way the data should be open, free, and complete with metadata also being available.

The insertion of a data management plan and proportionate funds, estimated to be less than 5 % of the total budget for each PNRA, should be very helpful in creating a common data archive.

In order to reach these goals, there are two possible approaches. The first is to designate a data center that collects and hosts all data such as the NERC Designated Data Centres. The second is to use the so called *system of systems* (see for example Maier (1999) or the technical report of Morris et al., 2004) in order to create a network of DBs for implementing a common information space and sharing resources through interoperability systems and universal standards. Due to the anisotropy of the polar research community in Italy, the latter approach seems more appropriate. In this regard, the GIIDA project (kick-off January 15, 2009) of the Department of Earth and Environment of the Italian National Research Council, which aims to design and develop a multidisciplinary infrastructure for the management, processing, and evaluation of environmental data, is important. SIRIA will take part in GIIDA, and the acquired know how will be useful also for a desirable data management plan inside PNRA.

6 CONCLUSION

An information system of the Italian polar research called SIRIA has been funded by PNRA since 2003. In addition to the primary objectives of managing the National Antarctic Data Center, updating the Antarctic Master Directory, and preserving the data acquired in the past Italian Antarctic expeditions, the project is actually under a restyling process, i.e., a new web site, inclusion of Arctic projects, a new metadata standard (ISO 19115), new data encoding (XML), new educational initiatives, and so on. Other activities have been included, especially during the last two years, in order to create a real portal to Italian polar research and expand SIRIA’s audience. In this short paper, the activity of SIRIA has given us also the chance to discuss some data sharing issues inside the polar sciences. We have compared the Italian point of view with various international initiatives and have given some considerations to the possible realization of a common data policy inside the polar research community of Italy.

ACKNOWLEDGMENTS

The SIRIA project is under PNRA funding. A.D. is grateful to Dr. M. Kanao and the Scientific Organizing Committee of IGY+50 Conference for the invitation to present the SIRIA PROJECT.

REFERENCES

NOTE: ALL URLS IN FOOTNOTES WERE RETRIEVED FROM THE WORLD WIDE WEB ON JANUARY 14, 2010.

¹⁷ The International Polar Year 2007-2008 Data Policy is available at: [http:// www.ipy.org/](http://www.ipy.org/).

- Allison, I., Béland, M., Qin, D., Sarukhanian, E., Carlson, D., & Smith, C. (2007) International Polar Year 2007-2008. *WMO Bulletin* 56, pp. 244-249.
- Damiani, A., Benedetti, E., Storini, M., & Rafanelli, C. (2008) Italian polar data center for capacity building associated with the IHY. *Advances in Space Research* 41, pp. 223-226.
- Finney, K. (2009) Data and Information Management Strategy (DIMS) 2009-2013. *SCAR Report* 34, ISSN 1755-9030.
- Lawrence, S. (2001) Free online availability substantially increases a paper's impact. *Nature* 411, pp. 521.
- Maier, M.W. (1999) Architecting principles for systems-of-systems. *Systems Engineering* 1 (4), pp. 267-284.
- Morris, E., Levine, L., Meyers, C., Place, P., Plakosh, D. (2004) *System of Systems Interoperability (SOSI): Final Report*. Technical Report CMU/SEI-2004-TR-004 ESC-TR-2004-004.
- Reinke, M. (2000) Data Policy and Data Publishing in Science. *International Workshop on Advanced Data Storage/Management for High Performance Computing CLRC*. Daresbury Laboratory, UK.
- Romano, V., Pau, S., Pezzopane, M., Zuccheretti, E., Zolesi, B., De Franceschi, G., & Locatelli, S. (2008) The electronic Space Weather upper atmosphere (eSWua) project at INGV: advancements and state of the art. *Annals of Geophysics* 26, pp. 345-351.
- Steigies, C.T. for the NMDB team (2008) NMDB: real-time database for high resolution neutron monitor measurements. *Geophysical Research Abstract* 10, EGU2008-A-04279.
- Thompson, B.J., Davila, J.M., Gopalswamy, N., Haubold, H., Rabello-Soares, M.C., Stamper, R., & Webb, D.F. (2008) First implementations stages of the IHY – *Progress. Space Research Today* 171, pp. 25-34.
- Thorley, M. (2006) Developing the NERC Data Policy. *DCC and DPC Joint Workshop: Policies for Long-Term Curation and Preservation*. Wolfson College, University of Oxford, UK.
- Uhlir, P.F., and Schröder, P. (2007) Open data for global science. *Data Science Journal* 6, pp. OD36-OD53.

(Article history: Received 31 January 2009, Accepted 30 December 2010, Available online 10 February 2010)