International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

SHARE Geonetwork, a system for climate and paleoclimate data sharing

M. T. Melis[1,2] , F. Locci [1] , F. Dessì[1], I. Frigerio[3], D. Strigaro[3], E. Vuillermoz[2].

[1] *TeleGIS Laboratory, Department of Chemical and Geological Sciences, University of Cagliari, Italy, telegis@unica.it*
[2] *Ev-K2 CNR Committee, Bergamo - Italy*
[3] *Geomatic Laboratory, University of Milano Bicocca, Milano, Italy*

Abstract: This study is dedicated to the development of a webGIS service platform for the environment data sharing. This new service is integrated into national and international projects supported by EvK2-CNR Committee and mainly focused on climate change studies. The system is based on the integration of the metadata catalogue Geonetwork and a dedicated database WDBPALEO. The scientific interest of this service is the possibility to access into the same site to the data acquired from the network of highest altitude stations for actual climate analysis and provision and data from paleoclimate samples acquired from ice and sea cores. The service for the first time shares these data in a downloadable standard format for all researchers and describes with the metadata catalogue all the necessary information for the correct use of these data. Since 2005, the Ev-K2 CNR Committee has promoted an integrated environmental project named SHARE (Station at High Altitude for Research on the Environment) focused on the mountain regions as primary indicators of climate change. Originally launched as a system of measurements in environmental and earth sciences in the Himalaya-Karakorum region, SHARE has later expanded its network to Europe (Apennines and Alps), Africa (Ruwenzori) and more recently to South America (Andes). The Italian project NextData, supported by Ministry of Education, Universities and Research has proposed this system for the development of a national service of environment data and metadata sharing dedicated to the study of climate change in the last 2000 years in the Mediterranean basin.

Keywords: Weather Database, Spatial Database, Open Source, High Mountains, SHARE, Geonetwork, paleoclimate, ice cores, marine cores.

1. INTRODUCTION

Scope of this work is to show the design, the development and the implementation of the Environmental Data System (EDS) for the SHARE Geonetwork Web-Platform, in the framework of the NEXTDATA project. The EDS is designed up to now to store data from two sources: high altitude automatic stations and non-polar ice cores and marine sediment cores (Melis et al., 2013).
The EDS is split in two parts: WDB and WDBPALEO database system: the first is dedicated to the storing of high altitude data from Automatic Weather Stations (AWS) of SHARE (Station at High Altitude for Research on the Environment) Network, and the second to paleoclimate data from non-polar ice cores and marine sediment cores.
These two data sources (AWS and cores) would be the base of a timeline history for a global past climate reconstruction. Mountains are sentinels of climate and environmental change and many marine regions provide information on past climate variations. The Italian Project of Interest NextData will favour the implementation of measurement networks in remote mountain and marine areas and will develop efficient web portals to access meteoclimatic and atmospheric composition data, past climate information from ice and sediment cores, biodiversity and ecosystem data, measurements of the hydrological cycle, marine reanalyses and climate projections at global and regional scale. New data on the present and past climatic variability and future climate projections in the Alps, the Himalaya-Karakoram, the Mediterranean region and other areas of interest will be obtained and made available (http://www.nextdataproject.it).

2. STORING APPROACH

The database structure proposed in this paper is one of the main products of the NEXTDATA project and it is called WDBPALEO. The aim of this database is to store spatial and temporal information about ice cores and marine sediment cores, data derived from chemical and physical characterization of the ice cores and a variety of metadata including information about the data provider.
WDBPALEO is a geographical database and the spatial information is defined by a couple of georeferenced coordinates that identify a point where the cores are drilled. All coordinates for every single core are stored in the database in longitude/latitude with WGS84 datum (EPSG 4326).
The WDBPALEO structure is based on WDB (Water and Weather Database System) (Institute T.N.M., 2012) architecture, a database developed by the Norwegian Meteorological Institute for their operability data-centre to stock hydro-meteorological data. This type of structure has been chosen because cores can be compared to a weather station. In fact, these two entities, from conceptual point of view, have two common principle aspects:

1. Both, cores and weather stations are represented geographically with a couple of coordinates that in a GIS environment can be treated as a geometrical vector point;
2. Both record climate data but, more specifically, cores give information about the past climate trend and the weather stations give information about the current climatic system. In fact they store the same type of data, characterized by a numerical value with a parameter related to temporal information.

Furthermore, WDB is released according to the GNU General Public License and so it is completely configurable, customizable and sharable.

2.1. WDB Structure

WDB is an open source (GNU Public License) data storage solution for meteorological, hydrological, and oceanographic data based on the PostgreSQL object-relational database system, running on Linux. Developed by the PROFF program at Norwegian Meteorological Institute (met.no) in 2008 and now WDB it is available with the 1.5.0 version (https://github.com/wdb), based on the ROAD (SMHI-Swedish Meteorological and Hydrological Institute) architecture. The system utilizes PostGIS for GIS (Geographic Information Systems) supports and handles Grids (e.g., forecast fields) and Point (e.g., observation) data.
WDB server architecture is composed by a WDB core (PostgreSQL), a Command Interface (WCI) and a series of loading programs (wdb-fastload, wdb-gribload, ecc) to load data from the Data Storage System.
WDB is designed to be:

1. robust enough to handle high volume of data
2. flexible to handle for example new type of data
3. supporting quality and consistency of data
4. simple to use , easy to maintain and operate.

Loading programs are deployed to load different of type of data, providing a direct connection to WDB, also with the possibility to handle large amount of data using them in batch programs.
The core of WDB is a list of 36 SQL functions to set up the entire schema (schemaDefinition.sql) and functionality of the database system (administration, geometry, parameters, etc.).
The “schema definition” initializes mainly:

- WDB_SCHEMA (usually wdb_int) is the core schema of the WDB system. It contains the core data tables, functions, and views;
- WCI_SCHEMA (usually wci_int) is a schema that contains the internal functions, views, and tables utilized by the WCI;
- WCI is the schema that contains the external functions of the WDB Call Interface;
- test is a schema that contains views, functions and tables that are specific for the testing of the WDB system.
- WDB Command Interface (WCI) is the data retrieval API for WDB, it allows the extraction of entire fields or points (time-series) from the database using an SQL-based function interface. WCI is compatible with C++ and it could be usable in C, Java, Perl, Python, and Fortran as well.
- WCI is a useful programming strategy to retrieve data from WDB to prevent the knowledge of the underlying data model (tables/views) by the user and the use of library files (e.g., changes in a library require that the applications using that library are recompiled and each operating system has its own library). WCI is designed in theory for portability to other operating systems.

3. WDBPALEO

WDB is designed to archive weather data, but its general structure could be also appropriate to stock data derived from analysis about ice cores and sediment cores. From an accurate study about the database architecture, we can identify three areas where information and metadata about data providers, parameters and cores can be archived. The core is the table where the numeric value of core parameters is inserted in. In fact, it has a lot of foreign keys that recalls the three parts previously described. In this way we can archive all data in the table of the database where values of the parameters are derived from chemicals and physicals measurements. In spite of this, paleo-climate data are peculiarly different from weather data: one of the most important differences is the temporal factor. The first weather observations date back to 1654, so WDB allows storing data that are not older than the XVIIth century. This is not the same in the ice and marine core sediments.
Antarctic and Greenland ice cores, but also marine sediment cores provide very useful proxy archives for paleoclimate study. For example the Antarctic Vostok ice core has provided climatic data over the past 420,000 years (Petit J.R., et al. 1999). The deep ice core from Dome C, Antarctica, provides a climate record for the past 740,000 years (Lambert F., et al. 2008). So, the problem of “time field” has been deal with the aim to structure a database that should also archived paleo- climatic data, such as from ice cores and marine sediment data cores.

3.1 ‘Time Field’ problem

Data/time field in PostgreSQL can give some problems, and mainly in the recording of environmental climate data because it uses the Boost library for time data with an inferior limit of about 1400 BC.. Here we present the adopted solution of this problem. First we used the 'timestamp with time zone' field, with a storage size of 8 bytes. This field type can archive temporal data from 4713 BC to 294276 AD. So we had problems when we started to insert data that referred to a period before 4713 BC.
To solve the problem with PostgreSQL time field, we have modified a part of the source code by inserting new functions to write data into IDB, without time limits. This improvement has increased the value of temporal data, which is essential for paleo-climatic analysis. We used a 'real' data type field. Positive numbers are years after Christ and negative numbers are referred to years before Christ. The 'real' field type can archive a 4 bytes information and can be positive, negative and with a precision of six decimals.

3.2 Reference System

The predefined Reference System (RS) in WDB is the 4030 EPSG. This RS is based on the WGS84 ellipsoid but it has an unknown datum. In a geospatial database the information about geographical position is basic. To offer and to work with geo- spatial correct information we worked with the EPSG 4326 that is used by the GPS satellite navigation system and for NATO military geodetic surveying.

4 LOADING DATA

The function wci.write() is used to load data: it executes an SQL statement and loads data point in the database (see Figure 1).

```
SELECT wci.write
(
        'value'           -- gid,
        'placename'       -- text,
        'referencetime'   -- timestamp without time zone,
        'validtimerom'    -- timestamp without time zone,
        'validtimeto'     -- timestamp without time zone
        'valueparameter' -- text,
        'levelparameter' -- text,
        'levelFrom'       -- float,
        'levelTo'         --float
)
```

Figure 1: WDB loading script: wci.write()

Before starting any loading operation the verification that all the metadata information are properly set in the database is required. ”data provider”, and ”placename” has been set in adaptation step, while “value“, “referencetime“, “validtimefrom”, ”validtimeto“, ”levelparameter”, ”levelfrom” and ”levelto” are picked up directly from data files. Using a script developed at met.no (wdb-fastload utility) it is possible to load a large amount of data, that takes a text data format from standard input and copies it into the database.
Raw or validated Data have a well-defined format; hence, before starting loading data, pre-processing is required. The pre-processing core is a Python script that takes the raw or validated file as input and converts it in a wdb-fastload format. In order to make the loading operations simple and fast, we have developed a Python GUI using the program Glade (see Figure 2). With this loading tool is possible adding new data provider in the database (only for administrator users), convert files in wdb-fastload format (now for fifteen data format corresponding to the AWS data logger) and then load the selected and converted file in the database.

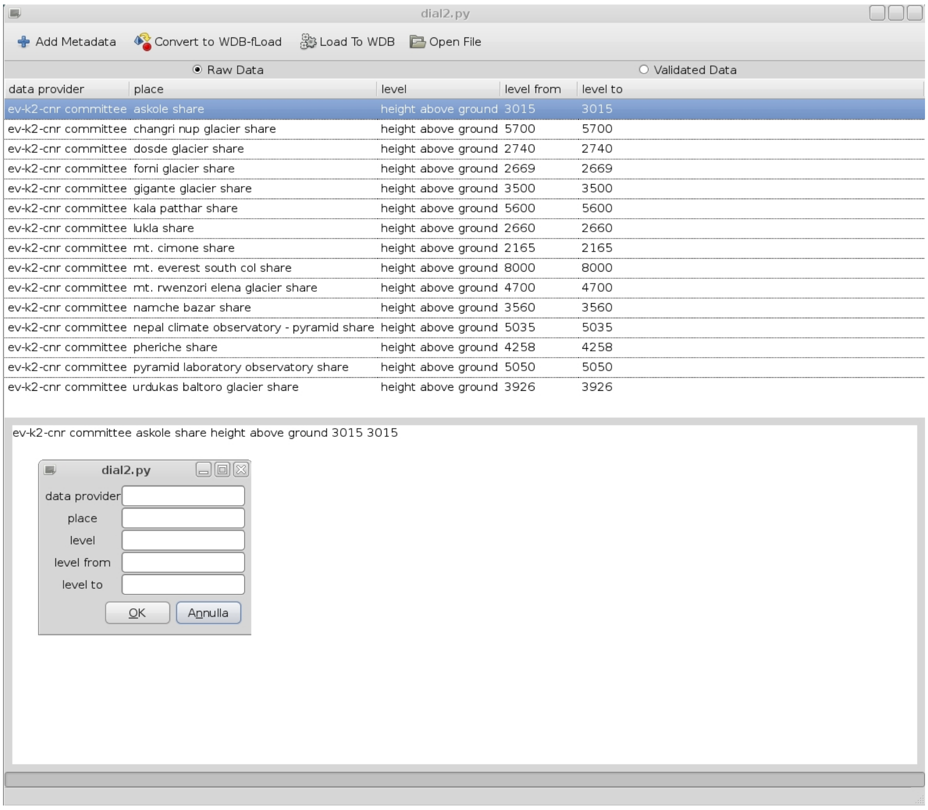

Figure 2: Python GUI for loading data and metadata into the database

The first loading data step concerns the geographic information. In most cases, the papers where the spatial information are written, reported the identical coordinate for different cores drilled in the same place. This problem is due to the fact that the GPS coordinates were taken with poor precision and it refer to the drilling site and not to a single core. To respect the topological rules, in order to insert the geographic data in to WDBPALEO, a GIS operation (shift points) was applied; this function moves overlapped points with the same coordinates in a circle around the original position. At the end of this operation, all points with core name and drilling site attributes were stored.
After this step it is necessary to input the Data Provider. The Data Provider identifies the source of the data; literally, the entity that provides the data. It can be the person in charge of core drilling or the principal investigator about the cores analysis. A DataProviderName that is used to search the data into the database identifies a data provider.
The cParameter in WDBPALEO identifies the characteristic or measurable factor of the value being parameterized. Parameters provide a definitive description of what the data represents, including chemical and physical properties. When cores are drilled they are cut and prepared to be catalogued and stored for analysis. Cores contain many proxy-parameters to help scientists to reconstruct past climates. For example, regarding chemical analysis, the concentration of atmospheric trace gases

such as nitrous oxide (N_2O), methane (CH_4) and carbon dioxide (CO_2) provide information about natural variation and man-made change of atmospheric composition.
After an accurate investigation about the main physical and chemical factors, over about 1000 parameters with a proper measurement unit were selected. To standardize the data, each parameter was defined by a IUPAC name for chemical value and SI (International System of Units) for units of measurements. Afterward, data were stored in the WDBPALEO "parametername" table.
Parameters are expressed by values. For raw numeric value here we intend the number obtained from a specific chemical or physical analysis on ice core sample. For example: -14.62 d18O, -0.30 Tanom or 87.6 ng/l Pb. This step was critical as each raw numeric value of chemical and physical measure are linked to the three parties mentioned above (core, data provider and parameters).
First, before loading data into WDBPALEO, the original txt and xls files were modified in CSV format with a precise formatting. A bash shell script was written in order to simplify the loading procedure. It reads CSV file, where core data were been antecedently prepared with a precise column order and then connecting to the database. Then, it recalls a SQL function that can distribute each attribute inside the right tables fields. This software has the aim to help future users in loading new data. The SQL function used to write data in the database was built with idea to check the data that were going to be insert in the database to avoid redundancy and other errors.
Nowadays 175 non-polar ice cores and 228 marine sediment cores have been loaded into WDBPALEO. (see Figure 3a and Figure 3b).

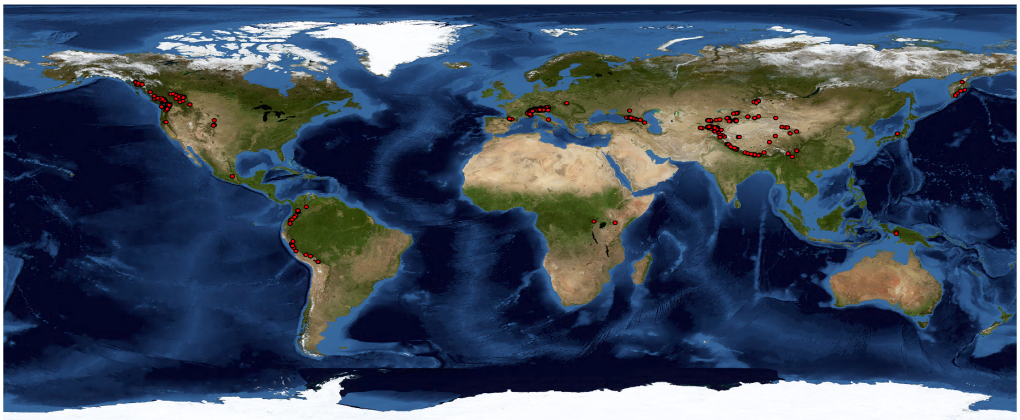

Figure 3a Distribution of non-polar ice glaciers

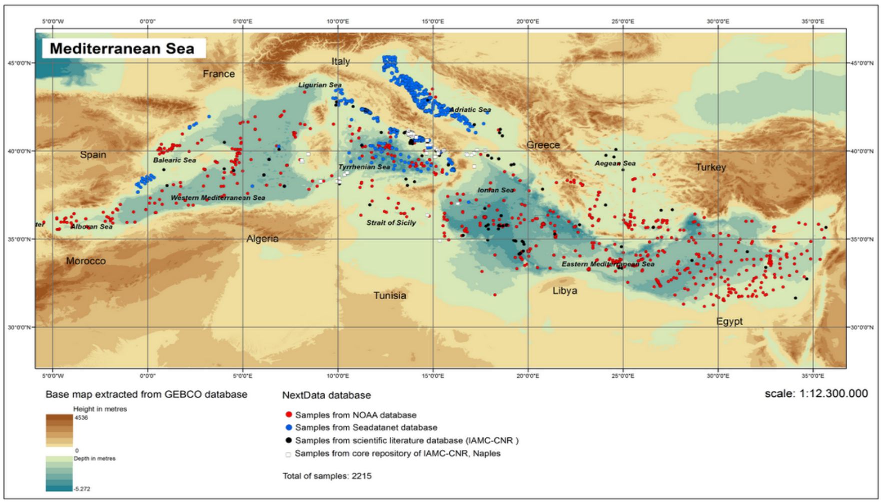

Figure 3b. Marine sediment cores points.

5 Geonetwork: metadata sharing

The Share Geonetwork platform for web services is based on the architecture of Geonetwork Open Source for the development of the data and metadata catalog dedicated to the high altitude research related to the SHARE project (Locci et al., 2013). NexttData project also choose to follow the INSPIRE (Infrastructure for Spatial InfoRmation in Europe) (INSPIRE, 2007), European framework that is an infrastructure for spatial information in Europe that will help to make spatial or geographical information more accessible and interoperable for a wide range of purposes supporting sustainable development. The Share Geonetwork platform for web services is based on the architecture of Geonetwork Open Source for the development of the data and metadata catalog dedicated to the high altitude research related to the SHARE project.

5.1 Geonetwork background

A prototype of the Geonetwork catalogue was developed by the Food and Agriculture Organization of the United Nations (FAO) in 2001 to systematically collect and publish the geographic datasets produced within the organization. Moreover, the World Food Programme (WFP) joined the project and with its contribution the first version of the software was released in 2003. Geonetwork has been developed following the principles of Free and Open Source Software (FOSS) and based on International and Open Standards for services and protocols, like the ISO-TC211 and the Open Geospatial Consortium (OGC) specifications (OCG, 2012). Currently, there is a big community that uses and develops the Geonetwork software. Geonetwork Open Source consists of:

- advanced search discovery tools,
- metadata catalogue, descriptive records for dataset,
- geographic data and map viewer

Improving sharing and access to data and information by the use of this open source system it is possible, also adopting two international standards: DUBLIN CORE for general documents and ISO19139 for geographic data. The use of these standards ensures that metadata can be interpreted by software and users and make data and information resources easier to find through the World Wide Web.

5.2 Metadata storing

In SHARE Geonetwork it is possible to archive the metadata of cores stored in WDBPALEO. To create a better structure, we choose to archive metadata following a hierarchical structure Parent/Child (Figure 4). Starting from the project domain (parent) that contains information about the perforation project such as: Scope work of the project, geographic region and point of contact of principal investigator. After the parent, there is the Drilling campaign. First child have to contain the name of the campaign, the reference time, the methods used for drill, number of Ice core taken from each extraction. The last is the Ice Core. It contains: ID of the Core, device (corer, benne) abstract of the principal paper wrote about that ice core, information about the point of contact and last the spatial information, coordinates, of the ice core.
Metadata are updated whenever new data are loaded or a change on data, on sensors, on instruments and in information related on data occurs. Also, new metadata scheme are developed whenever new type of environmental data need to be added to the web-service platform.
A system of hierarchical privileges, roles, and user groups to manage users and permissions to access, modification and data downloading has been designed. For the access to public information there are no restrictions, while in order to have access to specific information or functionality, an account that will be provided by the system administrator will be required. Also, it will be possible to read the information about a resource, download or browse interactively data for that resource depending on the role of an authenticated user and privileges set for the metadata record. Authenticated users (and depending on privileges) can create, import and edit metadata records. They can also upload and configure data links for interactive map. Each authenticated user is assigned to a particular working group and it is able to view the data within that group. The core of the information system is installed in a server located at the Ev-K2 CNR Committee Centre in Bergamo (main node), but there is the possibility to install relocated subsystems based on the same technology, named focal point of SHARE, which will contain metadata and data connected to the main node. Each focal point has its own region of interest and the system will be able to make a

search on all nodes simultaneously. This distributed system enhances the efficiency in term of speed of access to resources.

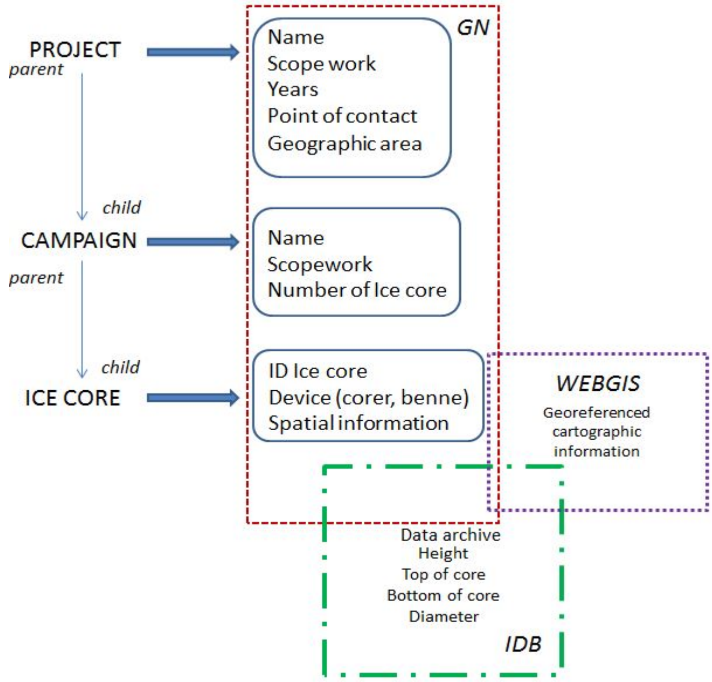

Figure 4. Example of hierarchical structure Parent/Child of metadata for ice cores.

5.3 Interactive maps

Geonetwork open source provides Internet access to interactive maps, satellite imagery and related spatial databases with the possibility of layers management and queries through an integrated WebGIS based on GeoServer software (Pumphrey, 2009). The integration of Google Maps cartographic data in the GeoServer has been developed making available advanced and complete satellite images, street maps, physical maps and hybrid maps. The system was deeply modified to bring the configurations to the publication of the Google's layers: the main issue is related to the change of the cartographic parameters to make them compatible with the publication in Google. The interface of SHARE Geonetwork was then changed in the configuration files and scripts with their dependencies that concern the segment mapping.

6. DISCUSSION AND CONCLUSIONS

The SHARE project and the SHARE Geonetwork in particular, were born from the necessity to improve the knowledge of mountains ecosystem and sharing data and metadata with the scientific community. These information are collected by direct observations from climatic and meteorological high mountain stations. The salient features of this project is access to and collection of high mountains research data in sensitive and remote areas (such as Africa and Asia, Alps and Andes), carried out in collaboration with local and governmental institution for the aim of preserving these fragile ecosystems (Salerno et al., 2010). In this work the original database WDB has been modified and adapted to collect data from ice and marine cores, WDBPALEO. In this way, a single point of access for actual, recent and paleo- climate data is available for the researchers and for all interested community.

A new database structure has been proposed to store and share data derived from chemical and physical cores characterization. Actually, this is the first geodatabase where the raw numeric values derived from the measurement on cores samples were stored and available. Unlike other databases, WDBPALEO allows users to search the numeric value of a specific chemical and physical analysis, starting from a particular data provider name or parameter. This is essential because it permits a rapid

data search and quickly comparison of different core variables. Moreover the absence of storage temporal limitation, allow the upgrade of database with cores from other providers.
The structure to store and share data, from database to Web GIS application, is based on open source software tools and they are free, useful and very capable. The present database is released with GNU license and it could potentially be customized and shared it without limitations.
The added value of this project is the interaction between cores data characterization, their geographic information and metadata. Through the development of web GIS application it is possible to share environmental datasets and furthermore provide easy access for users with lacking GIS knowledge.
The SHARE Geonetwork platform provides: a further impetus to high mountains research, complementing other projects such as GLORIA (Global Observation Research Initiative in Alpine Environments) (Grabherr et al., 2000) and the recent creation of a database for storing, processing and sharing glaciological data in Italy (Nigrelli and Marino, 2012); a methodology to implement a climatological/geographical information system using open source tools.

7. ACKNOWLEDGMENTS

This work has been carried out in the framework of SHARE and Next Data projects funded by the Italian Ministry of University and Research.

8. REFERENCES

Grabherr, G., Gottfried, M., Pauli, H., 2000. Gloria: A global observation research initiative in alpine environments. Mountain Research and Development 20, 190–191.

Lambert, F. et al..: Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. Nature 452, 616–619, 2008.

Locci F., Dessi F̀., De Amicis M., Frigerio I., Strigaro D., Vuillermoz E., Melis M. T., (2013). WDBPALEO 1.0: database per la raccolta di dati paleo-climatologici. 17a Conferenza Nazionale ASITA, Riva del Garda, Italy, 5-7 November, 2013. pp. 885- 892. ISBN 978-88-903132-8-8.

Melis, M. T., Dessì, F., Locci, F., Bonasoni, P., Vuillermoz, E. (2013). Share Geonetwork: a web-service platform for environmental data sharing. Proc. SPIE 8795, First International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2013), 87951V (August 5, 2013); doi:10.1117/12.2027602; http://dx.doi.org/10.1117/12.2027602.

Nigrelli, G., Marino, A., 2012. Dbclim: A web-based, open source relational database for rainfall event studies. Computers and Geosciences 48, 337–339.

Petit, J. R. et al.: Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature 399, 429–436, 1999

Pumphrey, M., 2009. GeoServer 2.2.x User Manual. http://docs.geoserver.org/stable/en/user/.

Salerno, F., Viviano, G., Thakuri, S., Flury, B., Maskey, R.K., Khanal, S.N., Bhuju, D., Carrer, M., Bhochhibhoya, S., Melis, M.T., Giannino, F., Staiano, A., Carteni, F., Mazzoleni, S., Cogo, A., Sapkota, A., Shrestha, S., Pandey, R.K., Manfredi, E.C., 2010. Energy, forest, and indoor air pollution models for Sagarmatha national park and buffer zone, Nepal. Mountain Research and Development (MRD) 30, 113–126.

WDB - https://github.com/wdb