

The Istituto Nazionale di Geofisica e Vulcanologia Data Management System for the Arctic Sciences

Umberto Apponi¹, Vincenzo Romano^{1,2}, Giorgiana De Franceschi², Paola Del Carlo³, Ilaria Isola³, Patrizia Macri², Giovanni Muscari², Leonardo Sagnotti², Luca Spogli^{1,2}, Stefano Urbini²

1. *Spacearth Technology S.r.l., Via di Vigna Murata 605, 00143 Rome, Italy*

2. *Istituto Nazionale di Geofisica e Vulcanologia, Environment Department, Via di Vigna Murata 605, 00148-Rome, Italy*

3. *Istituto Nazionale di Geofisica e Vulcanologia, Volcanoes Department, Via della Faggiola, 32-56126 Pisa, Italy*

Abstract: The brokering approach can be successfully used to overcome the crucial question of searching among enormous amount of data (raw and/or processed) produced and stored in different information systems. In this paper, authors describe the Data Management System the DMS (Data Management System) developed by INGV (Istituto Nazionale di Geofisica e Vulcanologia) to support the brokering system GEOSS (Global Earth Observation System of Systems) adopted for the ARCA (Arctic Present Climate Change and Past Extreme Events) project. This DMS includes heterogeneous data that contributes to the ARCA objective (www.arcaproject.it) focusing on multi-parametric and multi-disciplinary studies on the mechanism (s) behind the release of large volumes of cold and fresh water from melting of ice caps. The DMS is accessible directly at the www.arca.rm.ingv.it, or through the IADC (Italian Arctic Data Center) at <http://arcticnode.dta.cnr.it/iadc/gi-portal/index.jsp> that interoperates with the GEOSS brokering system (<http://www.geoport.org/>) making easy and fast the search of specific data set and its URL.

Key words: Brokering, ARCA (Arctic Present Climate Change and Past Extreme Events), IADC (Italian Arctic Data Center), GEOSS (Global Earth Observation System of Systems), DMS (Data Management System), data, metadata, INSPIRE (Infrastructure for Spatial Information in Europe), ISO (International Organization for Standardization) 19115 and 19119.

1. Introduction

The GEOSS (Global Earth Observation System of Systems) brokering approach [1] is based on a standard C-S (Client Server) architectures plus an added component, the broker [2] that guarantees and implements the semantic interoperability. The C-S architecture [3] is defined as “a computing model in which the server hosts, delivers and manages most of the resources and services to be consumed by the client. This type of architecture has one or more client computers connected to a central server over a network or Internet connection.”

This standard C-S architecture (Fig. 1) has two main stakeholders, the user (i.e. Service Consumer or Service Requestor) and the provider. The bus is an important part of the structure supporting protocols and

standard interfaces that are necessary to link users and providers.

To improve data searching facility at user level, especially when data are stored in multiple sources or repositories, the broker is added to the standard C-S architecture. This solution facilitates the management of different systems/nodes, e.g. by avoiding changes of their pre-existing standard for data transfer, and adapts to the producer service protocols and interfaces with the users' one.

The GEOSS brokering infrastructure adopted by ARCA (Arctic Present Climate Change and Past Extreme Events) [4] facilitates the search of data through a metadata catalogue furnished by the IADC (Italian Arctic Data Center) system. Users can obtain the URL to access specific data set, which are stored in a single node, e.g. the DMS (Data Management System) at INGV (Istituto Nazionale di Geofisica e Vulcanologia).

Corresponding author: Umberto Apponi, master, engineer, main research fields: electronics and informatics.

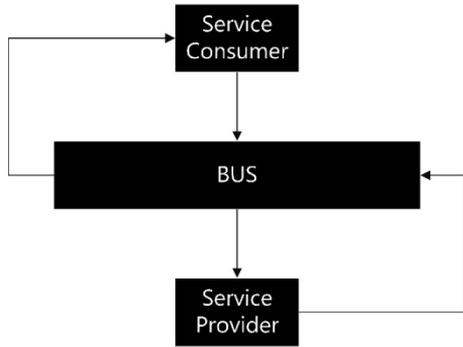


Fig. 1 The standard client server architecture.

The DMS is detailed in Section 2, where GEOSS as well IADC are briefly described. The standards adopted for the INGV metadata and the description of the different data set are given in Section 3. Concluding remarks are reported in Section 4, where the similarity between the data infrastructure implemented within ARCA and that one adopted by SCADM (Standing Committee on Antarctic Data Management) is highlighted.

2. Infrastructure Description

The infrastructure developed within ARCA contains three main sections (Fig. 2):

- The Nodes: servers in which ARCA data are stored. The DMS is the INGV node containing several datasets from different research areas.
- The IADC system: it is managed by CNR and collects metadata catalog for GEOSS. The user can search data with this portal that communicates with GEOSS in a transparent way.
- The brokering system: the GEOSS portal offers a single Internet access point for searching data, images and analytical software packages. It connects users to existing databases and portals, providing reliable, up-to-date and user-friendly information.

2.1 GEOSS and IADC

The GEOSS was developed by the GEO group to access, search and discover data. GEOSS is composed by four elements [5]:

- GEO portal;
- GEOSS Clearinghouse;

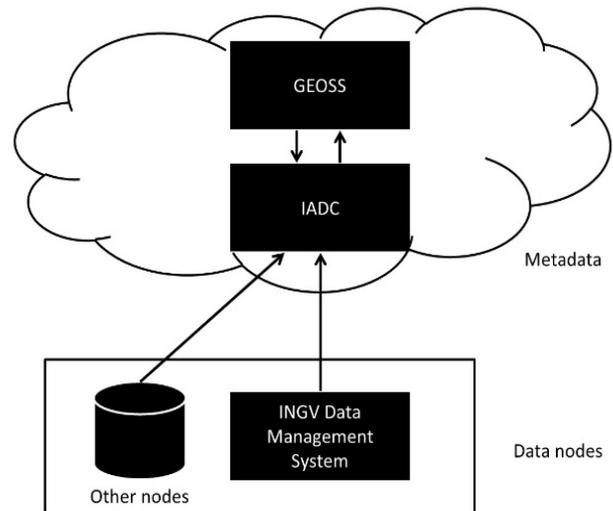


Fig. 2 ARCA Infrastructure description with connections between information components.

- GEOSS Components and Service Registry;
- GEOSS Standards and Interoperability Registry.

A user that wants to search information can directly access the GEO portal that provides a web interface for a fast and simple queries creation. The portal activates the GEOSS Clearinghouse that is the engine for managing and distributing components and services provided by 102 nations and 106 institutions (contributors). Information on components and services by contributors are included in the so-called GEOSS Components and Service Registry that defines the name and the way to manage them.

The GEOSS Standards and Interoperability Registry enables the contributors to configure their systems in order to implement the interoperability between different kinds of information systems. The interoperability concept is particularly stressed when multidisciplinary data are considered, taking into account the main standards to qualify and organize the information stored in several and various systems, used in the framework of national and international projects [6].

2.2 GEOSS and IADC

The main scope of the INGV DMS is the widest data dissemination among the communities of reference, letting users to reach data by querying the IADC portal,

where metadata information is stored [7]. The data dissemination is then further amplified through the link between IADC and GEOSS. The structure of the INGV DMS node has been developed taking into account both the requirements from the data owner and from the IADC system guided by GEOSS. These requirements can be summarized as follows:

- The metadata have to be compliant with INSPIRE and ISO 19115 standards (see Section 3);
- Two levels of data policy have to be implemented (free and restricted) decided together with the data owner depending on the type of data and/or product available;
- A user-friendly interface with simple graphics and template has to be developed to advantage the navigation and searching data on the node.

The DMS can be thought as two different sub-systems, the operation type, i.e. the download and the upload, identifies each of them. The researcher feeds (uploads) the server at INGV with his/her own data. The downloader (User) has two options: to obtain the desired information accessing directly the INGV DMS or by receiving the URL after the access to IADC. The flow chart in Fig. 3 is a representative of the actions to obtain the data/information for which the user is looking .

A web interface at <http://arca.rm.ingv.it>, represented in Fig. 4, has been developed to guarantee the user-friendly accessibility for DMS. The HW-SW characteristics of the DMS are summarized in Table 1.

The Debian Operating System (version 3. 16. 7) is installed on the DMS server machine, running on Intel Xeon Processor E5-2620 with 8 GB of RAM (Random Memory Access) memory.

The MySQL Database is used to store the web site’s management information (e.g. users’ credentials). The main services started on the server are SSH and FTP that allow uploading dataset from the owner/researcher. One couple of credentials are assigned to each of them together with a personal folder to upload further information, e.g. publications related to the data. The

owner can distribute another kind of credential to the user/downloader in order to provide a restricted access to the data. The server machine uses Apache 2. 4. 2 web server to manage web clients’ requests.

3. Metadata for the Infrastructure

The identification of the metadata fields is the main goal in order to define the proper information for qualifying the data. This activity has been carried out investigating the European commission regulations and national laws of reference. The most important are the “COMMISSION REGULATION (EC) N 1205/2008 of December 3, 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata” and the DM November 10, 2011, that Italy transposed for the metadata rules of EC n°1205/2008. By taking into account the above mentioned regulations and laws

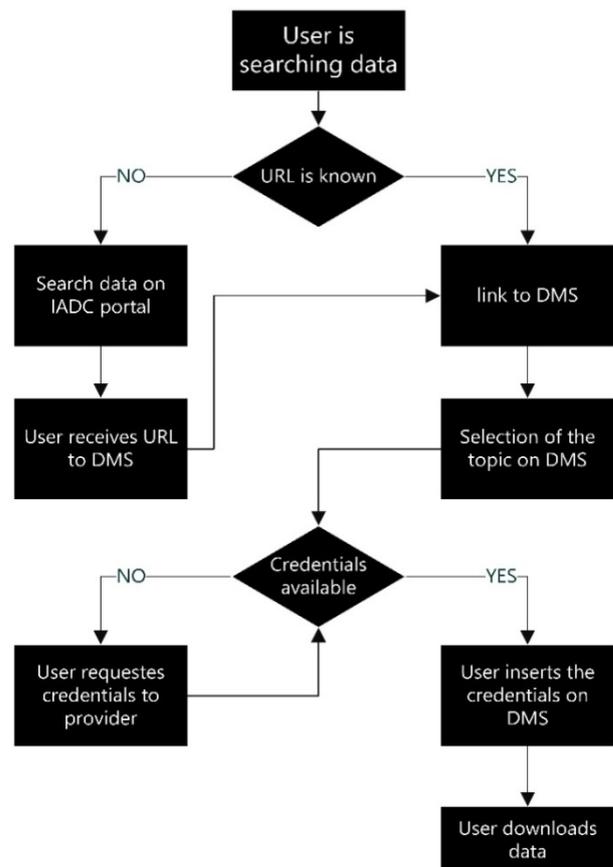


Fig. 3 Flow chart of the user searching operations on DMS.



Fig. 4 Data management system’s web interface.

Table 1 Hardware and software characteristics of INGV data management system.

Characteristics	Implementation
CMS	WordPress 4.3.2
Processor	Intel Xeon E5-2620
RAM	8 GB
Operating System	Debian 3.16.7
Storage	100 GB
Webserver	Apache 2.4.2
Services	FTP, SSH, Apache 2.4.2

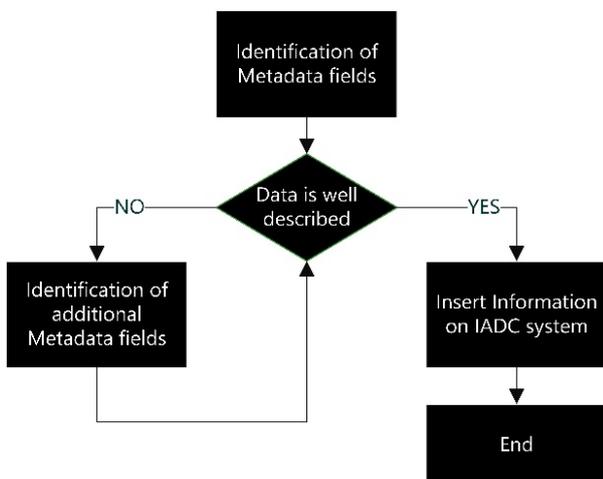


Fig. 5 Flow chart of the activities did by INGV in order to define metadata fields.

together to the INSPIRE METADATA guide [8], INGV metadata field compliant with ISO 19115 and ISO 19119 standards are finally obtained. For a better data description, other metadata fields have been provided to the IADC system. The flow chart in Fig. 5 describes the INGV activities for metadata.

3.1 Dataset on Data Management System

The datasets stored on the DMS server have been divided by topics, in order to provide to the user a general overview of their contents.

Going onto details, the dataset can be classified as:

- Ice Monitoring: collection of elaborated satellite images to investigate the ice front variation observed at major outlet glaciers. Data are available from the U.S. Geological Survey web site <http://glovis.usgs.gov/>. Detailed 3D maps of the bedrock morphology in the study areas were produced through the analysis of Radio Echo Sounding data, trying to evaluate the physical condition at ice/bedrock interface (wet and dry analysis) and its effect on calving process, ice flow

velocity and ice sheet mass balance. The variation of ice flow speed in the analyzed outlet glaciers was also considered. These maps were built exploiting the CRSIS (Centre for Remote Sensing of Ice Sheet, Kansas University; <https://data.cresis.ku.edu/>) public datasets [9];

- **Atmosphere:** The data uploaded on the ARCA DMS are related to the INGV activities for longwave down welling irradiance (LW) at the ground and column water vapor (PWV). These measurements were carried out from the atmospheric observatory at Thule Air Base (76.5° N, 68.8° W), Greenland, during the period from February 2 to March 12, 2014, and are part of a larger dataset dedicated to the study of atmospheric parameters in relation to the Earth energy balance and the temperature at the surface. The column water vapor or Precipitable Water Vapor (PWV) is measured at Thule Air Base (web site at <http://www.thuleatmos-it.it/> [10]) by means of a ground-based millimeter-wave spectrometer (GBMS). The GBMS observes rotational lines of atmospheric molecules emitting between 230 and 280 GHz, with a spectral pass band of 600 MHz. The PWV is calculated from direct measurements of atmospheric opacity (τ) at the indicated frequencies by using the linear relation $PWV = a + \tau \cdot b$, where a and b depend only on frequency. Measurements have a temporal resolution of 15 minutes and are carried out continuously when the GBMS is operated (usually only during wintertime) [11, 12].

- **Tephrochronology:** The dataset contained in the site refers to samples of marine sediments studied in order to identify tephra (volcanic particles) useful for stratigraphic and chronological information. Fifty samples of sediments were sampled from the GS191-01 PC core taken during the 2014 PREPARED oceanographic cruise held in the Arctic Sea (western of Svalbard Islands, Norway) by researchers from OGS. Cores were stored in OGS and University of Trieste facilities (Trieste, Italy). Samples were collected with a sampling rate of 10 cm, starting from the sea bottom

(top of the core) and for an interval of 5 m. A further 12 samples were selected based on preliminary magnetic susceptibility analysis along the total extent of the core. The material collected were studied preliminarily under the stereomicroscope to assess the different components and test the possible presence of volcanic particles. Given the very minute size of the recovered fragments, frequently $< 50 \mu\text{m}$, many of the samples were analyzed in alcohol solution using a petrographic microscope. This permitted to observe the samples in polarized light and easily distinguish glassy shards from other fragments very similar in color and shape (e.g. quartz or silic mineral fragments). Samples containing volcanic fragments were then mounted with epoxy resin and prepared as standard polished thin sections for SEM-EDS and EMPA analyses [13];

- **Speleothems Isotopes:** high-latitude Holocene paleoclimatic variations (in the medium latitude-paleoclimatic archives) identified from geochemical proxies from cave carbonate deposits (speleothems) [14, 15]. In Fig. 7, the isotopes of Oxygen and Carbon are analyzed and plotted as a function of distance from the top of the speleothem;

- **Paleomagnetism:** The data refer to the paleomagnetic and rock magnetic properties measured on marine sedimentary cores, retrieved from the upper slope between the northwestern Barents Sea and the southwestern Svalbard margin (Storfjorden and the Kveithola trough-mouth fans). Paleomagnetic and rock magnetic measurements were carried out on u-channel samples at 1 cm spacing, with specific instrumentation installed in the magnetically shielded room of the INGV-palaeomagnetic laboratory (Fig. 8). These data provide the opportunity to investigate the secular variation of the geomagnetic field in the geological past (known as paleosecular variation-PSV). Further analyses are in progress on other sedimentary cores collected during the Arctic expeditions supported by the Italian projects PNRA-CORIBAR and ARCA [16].

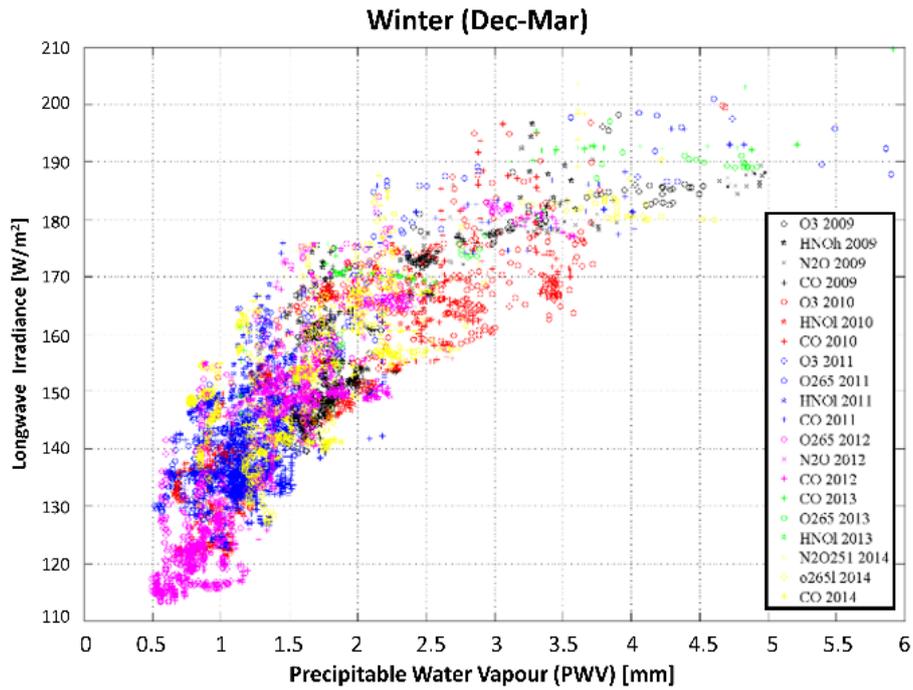


Fig. 6 Incoming longwave radiation in clear-sky conditions measured at the ground as a function of the column content of water vapour at Thule Air Base. Different colors and symbols indicate the different year and frequency (in the range 230-280 GHz, only for PWV data) at which measurements were carried out.

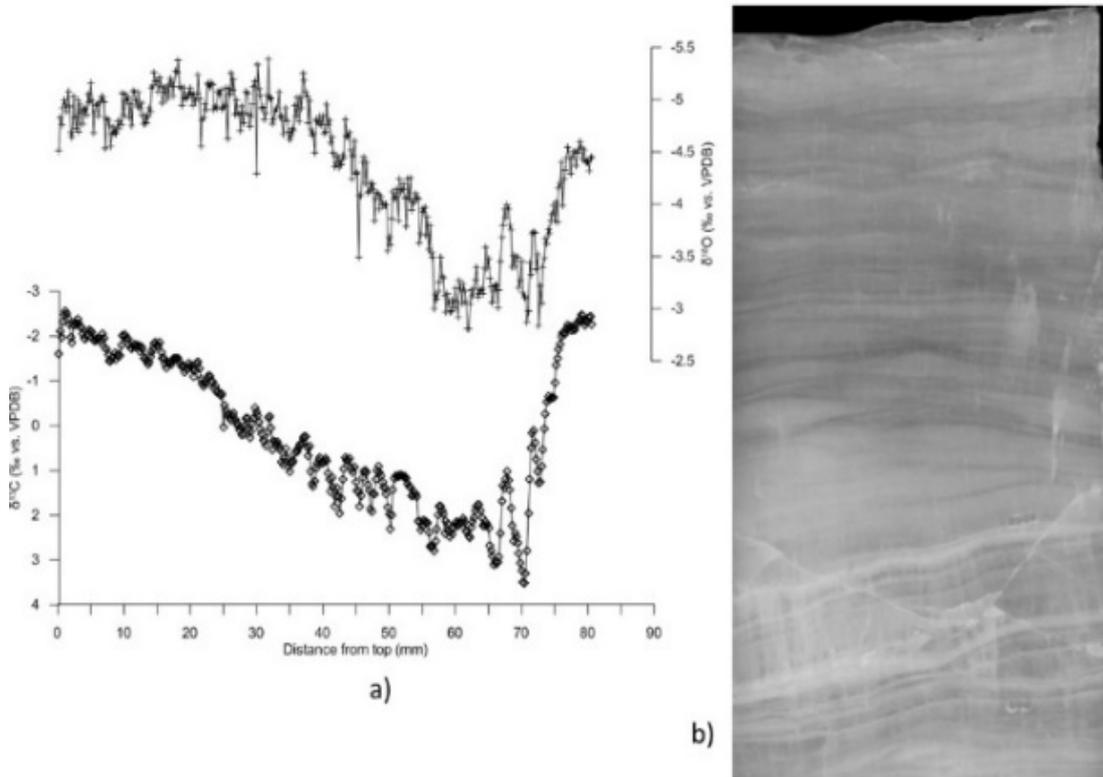


Fig. 7 (a) Graphic of the distance from the top of the flowstone for oxygen and carbon isotopes, respectively (top and bottom); (b) Flowstone's sampled image.

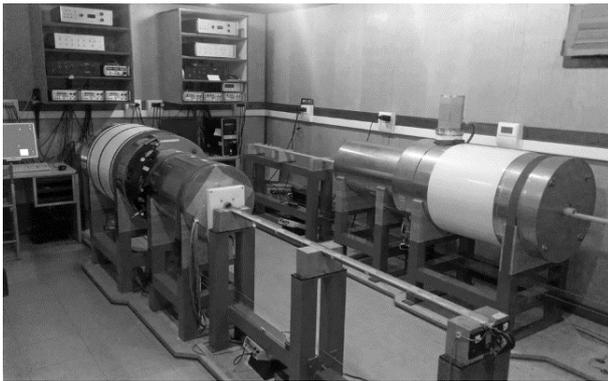


Fig. 8 2G DC squid cryogenic magnetometers installed in the magnetically shielded room of the paleomagnetic laboratory at the INGV.

4. Concluding Remarks

The innovative aspect of the ARCA infrastructure is the management of a great number of heterogeneous systems and nodes (data repositories) in easy way, through the broker component that allows user to search data stored in multiple sources. The INGV-DMS node has been developed to contribute to the ARCA ICT infrastructure, hosting Arctic data from INGV research activities. It helps user to discover and access the information, through a simple web interface and credentials control access preserving data providers on the use of the data. Currently the DMS contains six datasets but this is able to include and manage further data set from different disciplines, providers, Institutions dealing with Arctic sciences.

The ARCA infrastructure is similar to that one adopted in the framework of the Standing Committee on Antarctic Data Management (SCADM) (www.scar.org/data-products/scadm). The latter has two main components: the Antarctic Master Directory (AMD) and the Global Change Master Directory (GCMD, <http://gcmd.nasa.gov/>). The AMD, hosted at GCMD, is a repository for the metadata provided by Contributors. Ad hoc GCMD services allow GEOSS the capability to perform queries to the Earth Science metadata collection of NASA, including Antarctic metadata. So using the GEOSS portal to which ARCA contributes, both Antarctic and Arctic data are reachable,

benefitting all the scientific stakeholders interested in the field.

Acknowledgements

The DMS has been developed in the frame of the ARCA project (2015-2016), funded by the Italian Ministry of Instruction University and Research.

References

- [1] Nativi, S., Craglia, M., and Pearlman, J. 2012. "The Brokering Approach for Multidisciplinary Interoperability: A Position Paper." *International Journal of Spatial Data Infrastructures Research* 7: 1-15.
- [2] Nativi, S., Khalsa, S., Domenico, B., Craglia, M., Pearlman, J., Mazzetti, P., et al. 2011. "The Brokering Approach for Earth Science Cyber Infrastructure." *NSF Earth Cube White*.
- [3] Berson, A. 1996. *Client/server Architecture*. New York: McGraw-Hill Inc..
- [4] ARCA Group in Consiglio Nazionale delle Ricerche 2016. "Arctic Present Climate Change and Past Extreme Events Web Site." Consiglio Nazionale delle Ricerche. Accessed December 1, 2016. <http://www.arcaproject.it/index.php/it/>.
- [5] Group on Earth Observation 2016. "Global Earth Observation System of Systems Web Site." Group on Earth Observation. Accessed December 1, 2016. https://www.earthobservations.org/gci_gci.shtml.
- [6] Romano, V., Spogli, L., Salvati, A., Rafanelli, C., and Cafarella, L. 2013. "The IDIPOS Project: is a Multidisciplinary Data Infrastructure for Weather and Space Weather Feasible?" *Annals of Geophysics* 56 (2): R0224.
- [7] Consiglio Nazionale delle Ricerche 2016. "Italian Arctic Data Center Web Site." Consiglio Nazionale delle Ricerche. Accessed December 1, 2016. <http://arcticnode.dta.cnr.it/cnr/about.php>.
- [8] Team Metadata and European Commission Joint Research Centre 2013. "INSPIRE Metadata Implementing Rules Technical Guidelines Based on EN ISO 19115 and EN ISO 19119." Accessed December 2, 2016 http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf.
- [9] US Department of the Interior 2016, "U.S. Geological Survey Web Site." Department of the Interior. Accessed December 1, 2016. <http://glvis.usgs.gov/>.
- [10] Danish Meteorological Institute 2016. "Atmospheric Observatory at Thule Air Base Web Site." Danish

- Meteorological Institute. Accessed December 5, 2016. <http://www.thuleatmos-it.it/>.
- [11] Fiorucci, I., Muscari, G., Bianchi, C., Di Girolamo, P., Esposito, F., Grieco, G., et al. 2008. "Measurements of Low Amounts of Precipitable Water Vapor by Millimeter Wave Spectroscopy: An Intercomparison with Radiosonde, Raman Lidar, and Fourier Transform Infrared Data." *Journal of Geophysical Research: Atmospheres* 113 (D14).
- [12] Muscari, G., Di Biagio, C., di Sarra, A., Cacciani, M., Ascanius, S. E., Bertagnolio, P. P., et al. 2014. "Observations of Surface Radiation and Stratospheric Processes at Thule Air Base, Greenland, during the IPY." *Annals of Geophysics* 57 (3).
- [13] San Lucchi, R. G., Caricchi, C., Del Carlo, P., Di Roberto, A., Macri, P., and Sagnotti, L. 2016. "Lithostratigraphic Reconstruction of a nearly 20 m-Long Piston Core Collected in the Sediment Drift Area West of Svalbard." Accessed December 1, 2016. http://www.arcaproject.it/documents/Final_Conference/abstracts/ARCA_Lucchi_3.pdf.
- [14] Drysdale, R. N., Zanchetta, G., Hellstrom, J. C., Fallick, A. E., Zhao, J. X., Isola, I., et al. 2004. "Palaeoclimatic Implications of the Growth History and Stable Isotope ($\delta^{18}\text{O}$ and $\delta^{13}\text{C}$) Geochemistry of a Middle to Late Pleistocene Stalagmite from Central-Western Italy." *Earth and Planetary Science Letters* 27 (3): 215-29.
- [15] McDermott, F. 2004. "Palaeo-Climature Reconstruction from Stable Isotope Variations in Speleothems: A Review." *Quaternary Science Reviews* 23 (7): 901-18.
- [16] Sagnotti, L., Macri, P., and Lucchi, R. G. 2016. "Geomagnetic Palaeosecular Variation around 15 ka ago from NW Barents Sea Cores (South of Svalbard)." *Geophysical Journal International* 204 (2): 785-97.