DOI: 10.26314/GARR-Conf22-proceedings-08

ACTRIS aerosol profiling database: new design and new products for a wider use of aerosol lidar data

Claudio Dema, Ermann Ripepi, Giuseppe D'Amico, Pilar Gumà-Claramunt, Lucia Mona

CNR, IMAA, Potenza, Italy

Abstract. CNR-IMAA is leading and handling the aerosol remote sensing database of the ACTRIS European Research Infrastructure. This node is responsible for data curation, access and traceability of the data collected at the ACTRIS aerosol lidar stations distributed over Europe. A complete redesign of ARES is in progress for answering to the need of open and FAIR data. In addition, new ARES features involve the implementation of a new workflow and the provision of new tailored products to face the growing and pressing demand for aerosol vertical profiling information for manifold applications like model assimilation and evaluation, air quality impact investigations and aviation risks management. The new database also answers to some technical needs like quality controls and full traceability. These tasks become challenging when the RI is based on inhomogeneous sensors. The Single Calculus Chain is the solution adopted to ensure homogenous, traceable and quality controlled data.

dataKeywords. Research Infrastructure, FAIR data, Traceability, Interoperability, ACTRIS

Introduzione

CNR-IMAA is leading and handling the aerosol remote sensing (ARES) database of the ACTRIS European Research Infrastructure (Aerosol, Clouds and Trace gases Research InfraStructure Network). This node is responsible for data curation, access and traceability of the data collected at the ACTRIS aerosol lidar stations distributed over Europe (currently 35) (Dema C., Mona L., 2022). A complete redesign of ARES is in progress for answering to the need of open and FAIR data. In addition, new ARES features involve the implementation of a new workflow (allowing also NRT data provision) and the provision of new products to face the growing and pressing demand for aerosol vertical profiling information.

The new database answers to some technical needs like quality controls and full traceability, and to the growing demand of tailored products for different end user communities. A RI should ensure high-quality products by implementing a rigorous quality assurance program starting from raw data up to the final products. At same time, the RI products should be fully traceable in a way that it is always possible to fully characterize all computational steps, starting from the corresponding raw data to the final products. This aspect is particularly important as it plays a fundamental role in making the data FAIR compliant. These tasks become challenging when the RI is based on inhomogeneous sensors as happens for EARLINET (European Aerosol Research Lidar NETwork), the ACTRIS aerosol

remote sensing component (Pappalardo G. et al., 2014). EARLINET is a good example showing the high level of inhomogeneity of the network sensors (lidar systems in this case).
Most of the EARLINET lidar systems are home-made or highly customized. In cases like that, the implementation of a standard, centralized and quality assured scheme for the analysis of raw data is probably the most efficient solution to provide FAIR and quality assured data at RI level. The SCC (EARLINET Single Calculus Chain) is the solution adopted by the ACTRIS (Aerosol, Clouds and Trace gases Research InfraStructure Network) aerosol remote data center to ensure homogenous, traceable and quality controlled data (D'Amico G. et al., 2015).

1. ARES Data Products

Level 0 (raw) data are centrally processed at ACTRIS ARES DC level, generating Level 1 preprocessed signals and (not-fully QC data) optical properties products. On-the-fly QC procedures guarantee the basic quality control on Level 1 optical properties data. When the Level 1 data passes also the physical quality control procedures, is labelled as Level 2 data.
ARES DC will offer also products resulting from the processing at DC itself of Level 2 lidar and photometer data collected at the aerosol remote sensing NFs (National Facilities). Finally, Level 3 climatological products are produced at ARES DC from lidar Level 2 optical properties products.
New tailored products are highly requested for several applications, such as model assimilation and evaluation, air quality impact investigations, and aviation risks management. The development of such new products is mainly based on scientific advances and has then to be implemented into the centralized processing in a standardized way when the required maturity level is reached. This is the case for example of one promising product now under investigation about the identification of intense and potentially aviation-risky occurrence of desert dust/volcanic ash cases.

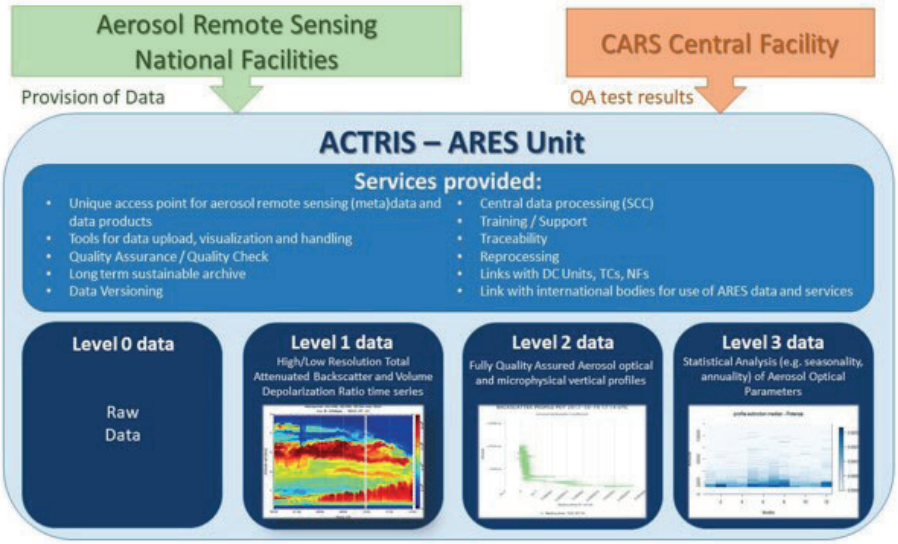


Fig. 1
ARES Data Products and Services

2. ARES Services

The ARES-CNR activities can be grouped according to the following thematic areas:

- Centralized Data Processing. This is provided through the Single Calculus Chain (SCC), the common standardized automatic analysis software for analyzing aerosol lidar data to obtain aerosol optical properties profiles from raw data.
- Data Archive and Storage. Data archive service of ACTRIS aerosol profile Level 1, 2 & 3 data, including off-site backup, documenting provenance, and providing a link to QA / QC data is given. Moreover, the data, the metadata database, the user interface, the web applications, the data curation tasks, including the data ingestion, and the SCC (database, calculus modules, web interface) are all hosted on dedicated servers.
- Recording of metadata in a dedicated RDBMS. The data archive for Level 1 and Level 2 data is suited for a relational database system, keeping all metadata and measurement data in one database. The database supports historic storage of data.
- Traceability. All data products, pre-products, and sample handling protocols, as well as all software and algorithms used in production steps, are version controlled, archived following long-term archive standards, and identified through Persistent Identifiers (PIDs) in each version. The provenance throughout all ACTRIS aerosol remote sensing workflows is provided by the use of standardized provenance scheme, facilitating attribution of entities involved in workflow execution. ARES is implementing the use of DOIs provided by DataCite for collections of data, and Handle PIDs provided through a Local Handle Server installed within ARES infrastructure which communicates with Global Handle Registries, according to the Handle System.
- Harmonization and Data Versioning. ARES Unit ensures that all instances of a specific data production step operated in the ACTRIS network uses the same identified version at any given time. Moreover, ARES relational database is a version controlled database suitable for reprocessing of the data and updates with new quality control procedures if needed. A Versioning System has been implemented directly in the RDBMS by using Data Manipulation Language triggers. New versions are centrally produced if new QC procedures and/or new processing features are released. Additionally, new versions of the files are allowed and centrally handled for fixing file bugs in particular for legacy data.
- Quality Control. When the data are submitted to ARES database, on-the-fly QC procedures are performed directly during the submission phase. The basic QC procedures check that the file content complies with the correct file structure, whilst the advanced or scientific QC check all mandatory products reported in the files in terms of scientific content.
- Interoperability. A THREDDS (Thematic Real-time Environmental Distributed Data Services) Server is used for serving data and metadata in an automated way through the OPeNDAP protocol. In addition to this, ARES provides a REST API for machine-to-machine interaction. The API serves metadata in JSON format and data in NetCDF format.

3. Archiving and preservation of ARES data

All data are stored in the ARES database which is hosted by the CNR ARES data center. The ARES infrastructure is composed by eight virtual machines (VM) and two Storage Area Network (SAN) in synchronous replication for high availability (H.A.). The ARES infrastructure is maintained by the National Research Council of Italy with long term commitment for archiving and preservation. A secondary asynchronous backup system at the CNR headquarters in Rome is being implemented.

References

D'Amico G., Amodeo A., Baars H., Binietoglou I., Freudenthaler V., Mattis I., Wandinger U., and Pappalardo G. (2015), EARLINET Single Calculus Chain – overview on methodology and strategy, Atmos. Meas. Tech., (8), pp 4891–4916, https://doi.org/10.5194/amt-8-4891-2015.

Dema C., Mona L. (2022), ARES: ACTRIS Aerosol REmote Sensing Data Centre Unit, https://www.actris.eu/topical-centre/data-centre/ares-aerosol-remote-sensing-data-centre-unit.

Pappalardo G., Amodeo A., Apituley A., Comeron A., Freudenthaler V., Linné H., Ansmann A., Bösenberg J., D'Amico G., Mattis I., Mona L., Wandinger U., Amiridis V., Alados-Arboledas L., Nicolae D., and Wiegner M. (2014), EARLINET: towards an advanced sustainable European aerosol lidar network, Atmos. Meas. Tech., (7), pp 2389–2409, https://doi.org/10.5194/amt-7-2389-2014.

Acknowledgments

We acknowledge the Aerosol, Clouds and Trace Gases Research Infrastructure Implementation (ACTRIS IMP) and the Environmental research infrastructures building FAIR services for research, innovation and society (ENVRI-FAIR) Projects. The work has been also funded by CIR01_00015 - PER-ACTRIS-IT "Potenziamento della componente italiana della Infrastruttura di Ricerca Aerosol, Clouds and Trace Gases Research Infrastructure - Rafforzamento del capitale umano" and by PER-ACTRIS-IT "Potenziamento della componente italiana della Infrastruttura di Ricerca Aerosol, Clouds and Trace Gases Research Infrastructure", PON "Ricerca e Innovazione 2014-2020".

Authors

Claudio Dema claudio.dema@imaa.cnr.it

Claudio Dema received the M.E. degree in Computer Engineering in 2017. He is currently working at CNR IMAA, within the ACTRIS Aerosol Remote Sensing (ARES) Data Centre. His main activities are: support to the management of ARES; management, analysis and processing of climatological data relating to atmospheric aerosols; design and application of standards and techniques to make scientific data FAIR; software analysis, design and development; design, creation and management of databases.

Ermann Ripepi ermann.ripepi@cnr.it

IT technician of the National Research Council - Institute of Methodologies for Environmental Analysis since 2007. At CNR-IMAA he is Head of the Infrastructure and Net-

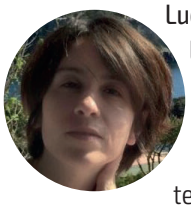
work Division of the ICT Office, APM GARR and ICT Manager of the Territorial Research Area of Potenza of the CNR. In recent years he has mainly dealt with networking, IT security and virtualization environments.

Giuseppe D'Amico giuseppe.damico@imaa.cnr.it

Giuseppe D'Amico (Laura Degree in Physics, Ph. D. Theoretical and Applied Physics) is a CNR researcher since 2006. He has over 15 years of research experience in atmospheric studies using different remote sensing techniques such as lidars, microwave radiometer, cloud radar. He also worked on the development and implementation of data analysis algorithms to deliver quality assured lidar products in NRT for coordinated lidar network like EARLINET (the European Aerosol Research LIdar NETwork).

Pilar Gumà-Claramunt pilar.guma@imaa.cnr.it

Pilar Gumà-Claramunt has a BSc in Environmental Sciences, a MSc in Meteorology and a PhD in Methods and Technologies for Environmental Monitoring. She is currently working at CNR-IMAA, within the ACTRIS Aerosol Remote Sensing Data Centre.



Lucia Mona lucia.mona@imaa.cnr.it

Lucia Mona is Senior researcher at CNR- IMAA. She is the leader of the ACTRIS Aerosol Remote Sensing (ARES) Data Center unit. Her expertise is about exploitation of EARLINET (European Aerosol Research Lidar NETwork) database for comparison/integration with other ground-based and satellite measurements and models and model evaluation/integration studies for peculiar long-range transport cases and for multi-year observation.