Deliverable D10.1: Definition of the ACTRIS Data Management Plan

Cathrine Lund Myhre, (NILU), Lucia Mona (CNR-IMAA), Markus Fiebig (NILU), Giuseppe D'Amico (CNR-IMAA), Francesco Amato (CNR-IMAA), Ewan O'Connor (FMI), Jacques Descloitres (CNRS-ICARE), Ann Mari Fjaeraa (NILU), Thomas Hamburger (NILU), Anne Hirsikko (FMI), Paolo Laj (CNRS-LGGE)

Work package no	**WP*10***
Deliverable no.	**D10.1**
Lead beneficiary	**NILU**
Deliverable type	☑ R (Document, report) ☐ DEC (Websites, patent fillings, videos, etc.) ☐ OTHER: please specify
Dissemination level	☑ PU (public) ☐ CO (confidential, only for members of the Consortium, incl Commission)
Estimated delivery date	**Month *6***
Actual delivery date	***01/11/2015***
Version	
Comments	

ACTRIS (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRAIA-2014-2015, Grant Agreement number: 654109

Aerosol, Clouds, and Trace Gases Research InfraStructure

ACTRIS Data Management Plan

Cathrine Lund Myhre, (NILU), Lucia Mona (CNR-IMAA), Markus Fiebig (NILU), Giuseppe D'Amico (CNR-IMAA), Francesco Amato (CNR-IMAA), Ewan O'Connor (FMI), Jacques Descloitres (CNRS-ICARE), Ann Mari Fjaeraa (NILU), Thomas Hamburger (NILU), Anne Hirsikko (FMI), Paolo Laj (CNRS-LGGE)

23 October 2015

Contents

1 Introduction to ACTRIS Data Centre

ACTRIS-2 (Aerosols, Clouds, and Trace gases Research InfraStructure) Integrating Activity (IA) addresses the scope of integrating state-of-the-art European ground-based stations for long-term observations of aerosols, clouds and short lived gases. ACTRIS-2 is a unique research infrastructure improving the quality of atmospheric observations, developing new methods and protocols, and harmonizing existing observations of the atmospheric variables listed in Appendix I.

The overall goal of the ACTRIS Data Centre is to provide scientists and other user groups with free and open access to all ACTRIS infrastructure data, complemented with access to innovative and mature data products, together with tools for quality assurance (QA), data analysis and research.

The numerous measurement methodologies applied in ACTRIS result in a considerable diversity of the data collected. In accordance with these requirements, the ACTRIS Data Centre consists of three topical data repositories archiving the measurement data, which are all linked through the ACTRIS data portal to provide a single access point to all data. Hence, the ACTRIS Data Centre is founded on 3 topical data repositories:

- Near-surface aerosol and trace gas data are reported to EBAS: http://ebas.nilu.no/
- Aerosol profile data are reported to the EARLINET Data base: http://access.earlinet.org/EARLINET/
- Cloud profile data are reported to the Cloudnet data base: http://www.cloud-net.org/data/

In addition, ICARE contributes with the production and provision of satellite data that complements the ACTRIS ground-based data: http://www.icare.univ-lille1.fr/catalogue.

Generally, the ACTRIS Data Centre and data management activity aim to work in accordance with the ENVRI Reference Model, hosted at www.envri.eu/rm.

2 ACTRIS data set descriptions

ACTRIS data sets are atmospheric variables listed in Appendix I, measured with the corresponding recommended methodology. Furthermore, the data are qualified as ACTRIS data sets only if they comply with the additional requirements specified in section 2.1 -2.3**.** The list of variables are expected to increase during the progress of ACTRIS, particularly secondary data products. During ACTRIS-2, e.g. the aerosol and cloud databases will be augmented with new classification products developed through the combination of existing sensors with additional instrumentation; and products providing information about aerosol layering and typing, together with advanced products derived from long term series or special case analyses. In addition, new parameters utilising these products will also be prepared, andstandardized preprocessed lidar data and NRT optical property profiles will be available.

2.1 Aerosol and trace gas near-surface data sets

Aerosol and trace gas near-surface data are qualified as ACTRIS data only if

- The atmospheric variables are included in the list in Appendix I
- The applied procedures comply with the standard operating procedures (SOP), and measurement recommendations and guidelines provided by the ACTRIS near-surface community. See section 4.1of this document for more details.
- The measurement data are submitted to the topic data base EBAS by using the reporting templates and procedures recommended by the ACTRIS near-surface community, and available at http://ebas-submit.nilu.no

Datasets fulfilling the requirements above qualify for the "ACTRIS" near-surface data set label. The types of variables are expected to expand during ACTRIS-2. The data can in addition be associated with other programs and frameworks such as GAW, EMEP, and national EPA etc. The data originator determines other project associations.

Standard collection and reporting procedure for aerosol and trace gas near-surface measurement data:

- Deadline for reporting data is 31 July of the following year from the reported measurements
- Data are submitted to a dedicated ftp-server at the data centre
- An auto-generated e-mail is sent to the data submitter to confirm that the data is received
- After submission, the data undergo an automatic format, NASA-Ames 1001, and metadata check, followed by manual inspection.
- If the data file is accepted, data are imported to EBAS, and feedback is given to the data originator. If there are suspicious data (e.g. suspicious data points/outliers) or format errors (in e.g. metadata, formats, etc.) the data originator is contacted and asked to assess, correct, and re-submit data.
- Data originators are asked about their project affiliation with collaborating networks and frameworks (EMEP, GAW-WDCA etc.)
- Trace gas data is made available to GAW-WDCGG; aerosol data are made available to GAW-WDCA.
- Near-real-time (NRT) data collection is set up and the raw data are auto-processed to hourly averages

2.2 Aerosol profile data sets

Aerosol profile data are qualified as ACTRIS data only if

- The atmospheric profile variables are included in the list in Appendix I.
- The applied procedures comply with the recommendations and procedures provided by the ACTRIS profile community, harmonised with EARLINET. See section 4.2 of this document for more details.
- The data are reported to the EARLINET DB in accordance with the reporting procedures (available at http://www.earlinet.org/).

Standard collection and reporting procedure for aerosol profile data:

- Data originators have the possibility to use, in addition to their own quality-assured method, the common standardized automatic analysis software developed within EARLINET, namely the Single Calculus Chain (SCC), for analysing their own lidar data to obtain optical properties from raw data, and passing through preprocessed data.
- New data shall be uploaded to the EARLINET DB within 3 months after measurement by data originator as preliminary data.
- Preliminary data shall be made accessible to the public as soon as possible, and automatically by the database 1 year after the measurement.
- All data will pass an approval process within 2 years after being measured. The approval is undertaken by an internal group of experts. During the ACTRIS2 project automatic QC procedures will be implemented and applied starting from these previous experiences.

At the beginning of ACTRIS-2 project, the aerosol vertical profile database contain aerosol optical properties profiles. During the ACTRIS-2 project, it will be augmented with more products, providing also information about the layering, and typing. In addition, standardized preprocessed lidar data and NRT optical properties profiles will be available.

2.3 Cloud profile data sets

Cloud profile data are qualified as ACTRIS data only if

- The atmospheric profile variables are included in the list in Appendix 1
- The processing applied complies with the procedures and recommendations provided by the ACTRIS community harmonised with Cloudnet.
- The data are reported to the Cloudnet DB in accordance with the reporting procedures

Standard collection and reporting procedure for cloud profile data

- Utilise the Cloudnet processing scheme.
- Preliminary data is accessible immediately to the community and public on insertion into the Cloudnet DB, together with a statement their appropriateness and validity for use.
- All data undergoes an approval process for final publishing, cognisant with full periodic calibration assessment and approval by expert panel.
- Selected variables are provided in NRT for the purposes of assimilation and NRT evaluation of NWP model data.

2.4 ACTRIS Secondary data products, combined data and project data tools

ACTRIS Secondary data are derived from the primary ACTRIS data described in section2.1-2.3, by e.g. averaging, filtering of events, interpolation of data etc. ACTRIS secondary data sets and project data tools can also include codes, algorithms and software used to generate ACTRIS primary or secondary data. Whereas primary datasets are regularly updated mainly due to the collection of new measurements and extension of the time series, secondary datasets are normally not updated. Secondary datasets are usually the result of targeted analysis, special studies, case studies, or processed for model experiments, including work performed under ACTRIS Joint Research Activities, and Transnational Access.They are “single purpose”, i.e. made for one specific purpose, as opposed to primary data which are documented as to serve as many purposes as possible.

2.4.1 Advanced products based on aerosol and trace gas near-surface data sets

Advanced products based on aerosol and trace gas near-surface data sets will be developed in collaboration with joint research activities and in accordance with other scientific requests during the project. Standard advanced products can include typically aggregated data such as daily, monthly or annual means of selected variables. Furthermore, the potential of long-term high quality ACTRIS-2 data for understanding of trends in atmospheric composition shall be further developed. A methodology will be put in place to analyse and produce regularly site-specific and regional trends. Suitable near-surface variables are particle size, and particle optical properties. Additionally, online QA tools and products will be offered for checking the consistency of the data sets in terms of ratios between specific trace gases, and closure tests between aerosol variables from different instruments.

2.4.2 Advanced products based on aerosol profile data sets

Advanced data products will be designed time by time following the specific needs as they are results of specific studies. Advanced data are stored and made freely available at EARLINET database as advanced products. These are the results of devoted (typically published) studies. Standard advanced products include climatological products from long-term observations. Further advanced products can be the results of JRA as microphysical aerosol products based on inversion of multi-channel lidar data, and microphysical aerosol products from combined lidar and sun-photometer observations. In particular, ICARE will automatically process raw lidar data from the EARLINET DB, combined with coincident AERONET data, using the GARRLiC (Generalized Aerosol Retrieval from Radiometer and Lidar Combined data) algorithm to retrieve vertical profiles of aerosol properties.

2.4.3 Advanced products based on cloud profile data sets

Advanced data products are prepared automatically by the Cloudnet processing scheme include model evaluation datasets, and diurnal/seasonal composites. In addition, advanced classification and products will be available from certain sites, and from campaigns, where additional instruments and products are combined.

2.4.4 Data sets resulting from combined activities with external data providers

The ICARE data centre routinely collects and produces various satellite data sets and model analyses that are used either in support of ground-based data analysis or in combination with ground-based data to generate advanced derived products. These data sets will be channelled to the ACTRIS portal using co-location and extraction/subsetting tool.

2.5 The ACTRIS user community

The ACTRIS user community can be classified as primary users (direct users of ACTRIS data, data products and services) and secondary users (using results from primary users, e.g. from international data centres). These are both internal and external users. In general, the user community can be summarized into five groups:

1) **Atmospheric science research community.** Together with atmospheric chemistry and physics, this also includes climate change research and meteorology, as well as multidisciplinary research combining these aspects (such as air quality, and climate interactions with links between aerosols, clouds and weather).
2) **Research communities in neighbouring fields of research.** These are environmental and ecosystem science, marine science, geosciences/geophysics, space physics, biodiversity, health and energy research. These communities will benefit from ACTRIS through the long-term provision of high-quality data products and through the enhanced capacity to perform interdisciplinary research.
3) **Operational observation and data management.** This community includes international data centres and international programmes to which ACTRIS contributes via the provision of long-term and consistent high-quality data products. Many research programmes and operational services (such as the Copernicus Atmosphere Monitoring and Climate Services) use ACTRIS to produce reliable data.
4) **Industry and private sector users**. These benefit from the services and high quality standards of the ACTRIS Calibration Centres, and from the free and open access to data products.

5) **Legislative / policy making community**. This include the user groups within climate, air quality and environmental issues including actors from local organisations, through national governments, to international conventions and treaties (including IPCC and UNFCCC, and UNECE-CLRTAP via the link to EMEP). This user community uses ACTRIS research results to define, update and enhance knowledge for decision making, policy topic preparation and drafting response and mitigation policies.

3 ACTRIS data set references and names

ACTRIS works towards establishing traceability for all applicable variables. In collaboration with partners in the ENVRI^{plus} project, ACTRIS is working towards use of digital object identifiers (DOIs), in order to assure proper attribution is given to data originators adequately reflecting their contributions.

Generally, ACTRIS data set names aim to be compliant with CF (Climate and Forecast) conventions. In the case where no standard CF names are defined, an application will be sent to establish these.

3.1 Aerosol and trace gas near-surface data set references and names

The near-surface data set names are listed in Appendix I. For most near-surface variables, ACTRIS data are traceable from the final data product back to the time of measurement. Traceability is implemented by a series of data levels leading from curated, instrument specific raw data to the final, automatically and manually quality assured data product. Processing steps between data levels are documented by SOPs.

All submissions of near-surface data passing quality assurance are uniquely identified in the EBAS database with a unique dataset identity numbers, ID-numbers. In case of updates, a ID-number is generated, and previous data versions are kept available upon request while the latest version is served through the database web-interface. Defined requests from the data holdings are identified in the web-interface by unique URLs that allow external links to the data.

3.2 Aerosol profiles

The aerosol profile data set names are listed in Appendix I. The EARLINET database is a version controlled database. The use of SCC allows the full traceability of the data: SSC converts individual instrument raw signals into standardized and quality-assured pre-processed lidar data. The SCC tool will be used to develop a harmonised network-wide, open and freely accessible quicklook database (high-resolution images of time-height cross sections). The standardized pre-processed data will also serve as input for any further processing of lidar data, within the SCC as well as in other processing algorithms (e.g., combined retrievals with sun photometer, combined retrievals with Cloudnet).

All aerosol profiles passed through quality check inspections manual and/or automatic leading to bi-annual final publication of quality checked data collection with DOI assignment. The DOI is assigned through the publication on the CERA database. In case of updates, only the latest version of data is available at http://access.earlinet.org and a new collection of data (with new DOI) is published. Previous data versions are kept available.

3.3 Cloud profiles

The cloud profile data set names are listed in Appendix I. The common use of the Cloudnet processing scheme ensures full traceability of the data from raw individual instrument measurements through to a combined standardised and quality-assured processed data set. The Cloudnet processing scheme ensures harmonisation of products across a relatively heterogeneous network. All quicklooks are open and freely accessible at http://www.cloud-net.org/quicklooks/

It is envisaged that publication of curated datasets with DOI assignment will commence as soon as possible. Currently, only the latest data version is available through http://www.cloud-net.org/data/ due to the large data volume requirements.

4 ACTRIS Standards and metadata

ACTRIS standards and metadata systems are well-developed, with parameter/variable standardization already existing in most cases. If this is not the case, ACTRIS, as a leading community in this field of atmospheric science, will work in collaboration with WMO-GAW, EMEP and other EU-funded projects (such as ENVRI^{plus}) in order to set the standards and foster interoperability between both the large variety of data products developed with ACTRIS itself, and with respect to external data centres.

4.1 Standards and metadata for aerosol and trace gas near-surface data

All aerosol and trace gas near-surface data sets are archived and provided in the NASA-Ames 1001 format.

4.1.1 Regular quality-assured data

Standards, SOPs and recommendations for each near-surface variable measured within ACTRIS are listed in the table below.

Variable	Reference SOP and recommendations
Particle light scattering coefficient	GAW report #200
Particle light absorption coefficient	GAW report #200
Particle number concentration	Wiedensohler et al., Atmos. Meas. Tech., 5, 657-685, 2012, doi:10.5194/amt-5-657-2012
Particle number size distributions (fine fraction)	Wiedensohler et al., Atmos. Meas. Tech., 5, 657-685, 2012, doi:10.5194/amt-5-657-2012
Particle number size distributions (coarse fraction)	ACTRIS protocol in preparation
Cloud condensation nuclei number concentration	ACTRIS protocol in preparation
Liquid Water Content	ACTRIS protocol in preparation, see also Guyot et al., Atmos. Meas. Tech. Discuss., 8, 5511-5563, doi:10.5194/amtd-8-5511-2015, 2015.
Particulate organic and elemental carbon mass concentrations (OC/EC)	EMEP/CCC (2014) Manual for sampling and chemical analysis. Chapter 4.22 (Last rev. February 2014). URL: http://www.nilu.no/projects/ccc/manual/index.html. See also Cavalli et al., Atmos. Meas. Tech., 3, 79-89, 2010, doi:10.5194/amt-3-79-2010
Particulate size-resolved chemical composition (organic & inorganic size-resolved mass speciation)	ACTRIS protocol in preparationSee also Ng, et al., Aerosol Science and Technology, 45:770-784. 2011, DOI:10.1080/02786826.2011.560211 and Fröhlichet al., Atmos. Meas. Tech., 6:3225-3241, 2013, doi:10.5194/amt-6-3225-2013.

Variable	Reference SOP and recommendations
Particulate levogluocsan mass concentration	Yttri et al,. Atmos. Meas. Tech., 8, 125–147, 2015, Further ACTRIS recommendations in preparation.
Volatile Organic Compounds (VOCs)	ACTRIS-FP7 Deliverable D4.9:Final SOPs for VOCs measurements http://www.actris.net/Portals/97/Publications/quality%20standards/WP4_D4.9_M42_30092014.pdf
NO_{xy}	ACTRIS-FP7 Deliverable D4.10: Standardized operating procedures (SOPs) for NOxy measurements http://www.actris.net/Portals/97/Publications/quality%20standards/WP4_D4.10_M42_140919.pdf

Metadata: A comprehensive metadata system and description of each ACTRIS near-surface variable is implemented in the topic data base EBAS. All ACTRIS near-surface variables are reported to EBAS by using the reporting templates recommended by the ACTRIS near-surface community, harmonized with GAW-recommendations. The templates ensure that the measurements are reported in accordance with the procedures for the employed instrument, and include all the necessary metadata required to precisely describe the measurements, including uncertainty/percentiles. In this way, all ACTRIS near-surface data are accompanied by a sufficient documentation of the measurements to have in-depth information on the quality of the data. Information about the reporting procedure and metadata items are open accessible and available through http://ebas-submit.nilu.no. Metadata are interconnected with GAWSIS and the ACTRIS data center handling of metadata is INSPIRE and WIS-ready.

4.1.2 Near-real-time (NRT) data

Near-real-time (NRT) data flow is offered to the data originators as daily quality check for selected variables, with the possibility for an alert system for outliers, instrumental failures and inconsistencies.NRT data collection and dissemination is available for the near-surface ACTRIS observables as identified in Appendix I.

Participating stations submit their data as annotated raw data in hourly submissions starting and ending at the turn of an hour. As an exception, 3-hourly submissions are accepted if indicated by limited connectivity with the station. The raw data are auto-processed to hourly averages, while periods with obvious instrument malfunctions are disregarded. Special sampling conditions or transport episodes are not flagged. The processed NRT data are available through the EBAS web-interface or through auto-updated custom FTP extracts.

4.2 Standards and metadata for aerosol profiles

Aerosol profiles data are archived and provided in netCDF format. All published EARLINET data are in CF (Climate and Forecast) 1.5 compliant format. A migration for all the data to this convention is planned.

Variable	Reference SOP and recommendations
Aerosol backscatter coefficient profile	Bockmann et al., Appl. Opt. 2004
Aerosol extinction coefficient profile	Pappalardo et al., Appl. Opt. 2004
Lidar ratio profile	Pappalardo et al., Appl. Opt. 2004
Ångström exponent profile	Pappalardo et al., Appl. Opt. 2004
Backscatter-related Ångström exponent profile	Bockmann et al., Appl. Opt. 2004
Particle depolarization ratio profile	ACTRIS-FP7 Deliverable D2.7, see also Freudenthaler et al., Tellus, 2008
Planetary boundary Layer	Matthias et al., JGR 2004

Metadata: All aerosol profile data are accompanied by respective metadata reporting information about the station, the system, and the timing of the measurements. Aerosol profile data sets reported to the ACTRIS data centre can be the results of regular operation of the EARLINET network, but also related to specific campaigns and joint research activities. Homogeneous and well-established quality of data originating from different systems is assured through a rigorous quality assurance program addressing both instrument performance and evaluation of the algorithms. Information about the QA program are summarized in Pappalardo et al., AMT, 2014 and are open and freely available at http://www.atmos-meas-tech.net/7/2389/2014/amt-7-2389-2014.html ACTRIS-2 improvement of the SCC is a step forward to complete harmonization of the aerosol profiles data quality. During ACTRIS-2, protocols and quality check procedures will be further optimized, in particular for new products, in NA2 and data QC tools will be developed in NA2 in collaboration with the data centre, checking the data optical properties consistency and through the comparison with climatological data. The SCC and all QC tools will be available to all potential users of ACTRIS data, both internal and external.

4.3 Standards and metadata for cloud profiles

4.3.1 Quality-assured data

Cloud profiles are archived and provided in netCDF format, with CF–compliant metadata.

The base-line SOPs and recommendations for Cloudnet variables is given in Illingworth et al., (2007), with updates given in ACTRIS-FP7 Deliverable D5.10

Variable	Reference SOP and recommendations
Cloud and aerosol target classification	Illingworth et al., BAMS, 2007
Drizzle products	ACTRIS-FP7 Deliverable D5.7, see also O'Connor et al., JTECH, 2005
Ice water content	Hogan et al., JAMC, 2006
Liquid water content	Illingworth et al., BAMS, 2007
Liquid water path	MWRNET, http://cetemps.aquila.infn.it/mwrnet/ see also Gaussiat et al., JTECH, 2007
Higher-level metrics	ACTRIS-FP7 Deliverable D5.10

Metadata: Cloud profile data are accompanied by metadata describing the station, instrument combination and supporting ancillary measurements, and processing software version. Metadata describing instrument calibration history will be implemented within ACTRIS-2. Harmonization and rigorous quality control for data originating from different instruments and instrument combinations is achieved through the common use of the Cloudnet processing software, summarised in Illingworth et al. (2007). All metadata is propagated through to every cloud product derived from the measurements; this requirement will be mandated for all new products derived during ACTRIS-2. The Cloudnet processing scheme, and the interface description for generating new products, is freely available for all potential users of ACTRIS data, whether internal or external.

4.3.2 Near-real-time (NRT) data

All cloud NRT data is processed in the same manner as for quality-assured data, together with all accompanying metadata. However, subsequent instrument calibration may require reprocessing to generate a revised product which uses the updated calibration values.

5 Sharing of ACTRIS data sets and data products

5.1 Access to ACTRIS data sets and data products

The ACTRIS Data Centre compile, archive and provide access to all ACTRIS data, and the ACTRIS data portal (http://actris.nilu.no) is giving free and open access to all data resulting from the activities of the ACTRIS infrastructure, including advanced data products resulting from ACTRIS research activities. Every dataset created within ACTRIS is owned by the ACTRIS partner(s) who created this dataset. *The ACTRIS Data Policy (*http://actris.nilu.no/Content/Documents/DataPolicy.pdf) regulates the sharing and use of ACTRIS data, see section 5.3.

The ACTRIS data portal (http://actris.nilu.no) provide access to ACTRIS data sets. This is a virtual research environment with access to all data from ACTRIS platforms and higher level data products resulting from scientific activities. The portal is structured as a metadata catalogue, searching the topical data bases, enabling data download from the primary archive and combination of data across the primary data repositories. The metadata catalogue is updated every night, providing access to all recent ACTRIS data. All data are archived in the topical data repositories, to [1)]maintain access to last version of data, [2)]avoid duplications and [3)]keep full traceability of the data sets.

The cooperation of ACTRIS with EUDAT, has already started and will proceed through ENVRI[PLUS], providing a further instrument for discovering the ACTRIS data sets.

5.1.1 Aerosol and trace gas near-surface data repository

The ACTRIS data repository for all aerosol and trace gas near-surface data is EBAS. http://ebas.nilu.no. The web portal is set up on a dedicated linux server running in Python program language. EBAS is an atmospheric database infrastructure where open access to research data has developed over almost 45 years and the data infrastructure is developed, operated, and maintained by NILU - Norwegian Institute for Air Research. The main objective of EBAS is to handle, store and disseminate atmospheric composition data generated by international and national frameworks to various types of user communities. Currently, EBAS is a data repository for ACTRIS, and also hosts the World Data Centre of aerosols under WMO Global Atmosphere Watch (GAW) and data from European Monitoring and Evaluation Programme (EMEP) under the UN Convention for Long-Range Transport of Air Pollution (CLRTAP), among other frameworks and programmes.

No embargo times apply to these data; all data is reported to EBAS as early as possible, and no later than 31 July the following year of the measurement. The data sets are made available to all users as soon as possible after quality control and quality assurance.

5.1.2 Aerosol profile data repository

The ACTRIS data repository for all aerosol profile data ishttp://access.earlinet.org. The aerosol profile database is hosted, maintained and operated by CNR-IMAA (National Research Council-Institute of Methodologies for Environmental Analysis) where the Single Calculus Chain for the automatic processing of lidar data for aerosol optical properties retrieval was designed, optimized and operated for the whole network. CNR-IMAA hosts different advanced products developed by EARLINET in the past for providing access to external users (volcanic eruption products, satellite validation datasets and NRT EARLINET subsets).

Aerosol profiles data are regularly published (every 2 years) on the CERA database, following the first database publications of EARLINET database. This assures the discoverability of the data through the association of a DOI to the data and the archiving on CERA, a recognized official repository.

5.1.3 Cloud profile data repository

The ACTRIS data repository for all cloud profile data ishttp://www.cloud-net.org. The cloud profile database is currently hosted, maintained and operated by the University of Reading, but is in transition to FMI (Finnish Meteorological Institute). The database provides the capability for both in-house processing of instrument data, and collection of on-site processed data through distributed use of the Cloudnet processing scheme. Both NRT access (e.g. model evaluation) and full quality-assured archived data access is available for internal and external users.

No embargo is applied to data quicklooks, available in NRT when possible. An embargo is generally only applied to data when a site is in testing mode (new instrumentation or re-calibration of existing instrumentation). Otherwise all data sets are immediately available in NRT-mode (no QA) or as soon as quality control/assurance has been applied. During the course of ACTRIS-2 quality-assured archived datasets will be published in a recognized official repository with an associated DOI.

5.2 Access to secondary data and combined data products

ACTRIS secondary data sets are stored in dedicated catalogue in the ACTRIS Data Centre or specified in the ACTRIS topical databases to provide long term access for all users. Access to these data sets and products is made available through the ACTRIS data portal: http://actris.nilu.no.

The ICARE Data and Services Centre is hosted by the University of Lille in partnership with CNRS and CNES. ICARE routinely collects various data sets from third party data providers (e.g., space agencies, meteorological agencies, ground-based observation stations) and generates a large number of derived products. All data sets are available for download at http://www.icare.univ-lille1.fr/catalogue through direct FTP access or web-based services, upon receipt or upon production, some of them in NRT. In addition, ICARE provides visualisation and analysis tools (e.g., http://www.icare.univ-lille1.fr/browse), and tools to co-locate and subset data sets at the vicinity of ground-based observation networks (http://www.icare.univ-lille1.fr/extract). Existing tools will be fine-tuned to meet specific ACTRIS requirements. Access to selected data and services will be facilitated through the ACTRIS portal.
No embargo is applied to data quicklooks. Most data sets are freely available for download upon registration. Some restrictions in data access or data use may be inherited from original data providers or algorithm PIs for experimental products generated at ICARE.

5.3 The ACTRIS Data Policy

The ACTRIS Data Policy regulates the sharing of ACTRIS data and includes information on dissemination, sharing and access procedures for various types of data and various user groups. The ACTRIS Data Policy is publically available from the ACTRIS web site, from the ACTRIS Data Centre, and here: http://actris.nilu.no/Content/Documents/DataPolicy.pdf

The 1st version of the ACTRIS Data Policy was established under ACTRIS-FP7, June 2012. The 2nd version was approved by ACTRIS-2 SSC, September 2015.

6 Archiving and preservation of ACTRIS data sets

The main structure and installations of the ACTRIS Data Centre is located at *NILU - Norwegian Institute for Air Research*, Kjeller, Norway. NILU hosts EBAS archiving all near-surface data sets, in addition to the ACTRIS Data Portal. The other installations are the EARLINET DB at *National Research Council - Institute of Environmental Analysis* (CNR), Tito Scalo, Potenza, Italy, the satellite data components at *University of Lille*, Villeneuve d'Ascq, France, and the cloud profile data at *Reading University*, Reading, UK. There will be a transfer of the installation from Reading University to FMI (Finnish Meteorological Institute) by May 2016.

6.1 Aerosol and trace gas near-surface data

EBAS is a relational database (Sybase) developed in the mid-1990s. Data from primary projects and programmes, such as ACTRIS, GAW-WDCA, EMEP, AMAP, are physically stored in EBAS. All data in EBAS are, in addition, stored at a dedicated disk in the file tree at NILU. This include all 3 levels (0-1-2) of data.

The complete data system is backed up regularly. This includes incremental back up of the data base 6 times per week, and one weekly back up of the full data base to a server in a neighbor building to ensure as complete as possible storage of all data for future use in case of e.g. fires or other damages to the physical construction. File submission is conducted by ftp. A separate ftp area is allocated to incoming files, and all activities herein are logged on a separate log file, and backed up on 2 hour frequency. An alert system is implemented to ensure warning messages if there are problems during file transfer from the data originators to the data centre.

Ca 455 separate new comprehensive files including meta data with annual time series of medium to high time resolution (seconds to week) is expected per year. A significant growth in this number is not expected on annual scale. In total this will sum up to ca 10GB/year from ca 150 000 single column files, including both raw data and auxiliary parameters.

EBAS is based on data management over more than 40 years. Last 10 years there has been a European project-type cooperation from FP5 to Horizon2020, with and EMEP and GAW programmes since 1970's as the fundament. Sharing visions and goals with the supporting long-term policy driven frameworks have ensured long-term funding for the core data base infrastructure. Currently, a long-term strategy for providing access to all ACTRIS data and other related services are explored through the establishment of ACTRIS as a RI. For this reason, ACTRIS is applying a position on the next ESFRI (European Strategy Forum on Research Infrastructures) roadmap for Research Infrastructures.

6.2 Aerosol profiles

The storage infrastructure is composed by two servers and a SAN (Storage Area Network). One server hosts the EARLINET PostgreSQL database and the other one is used to interface both end-users and data submitters to the EARLINET database. This last server is connected to an internal SAN on which the data submitted by the user are safety stored. A daily back up of the EARLINET database is made automatically and it is stored on the SAN.

The current size of the PostgresSQL EARLINET database is about 1GB. The total amount of data submitted (NetCDF EARLINET files) is about 900MB. An estimation of the growing rate of the database at this rate is 100-200MB/year. However a significant growth in number of files to be collected is expected because of: the use of the Single Calculus Chain for the data submission, the inclusion into the ACTRIS aerosol profiles database of new products (pre-processed data, NRT optical properties, profiles, aerosol layers properties and multi-wavelength datasets), increases of the number of EARLINET stations and

increase of EARLINET h24 stations. We estimate that at the end of ACTRIS2 project, the ACTRIS aerosol profile database could growth at a rate of about 12-15 GB per year.

The EARLINET database is maintained by the National Research Council of Italy with long term commitment for archiving and preservation. The archiving on CERA database is a further measure for assuring the availability of the data through redundancy of the archive. Further developments in terms of specific services will be developed in ACTRIS 2 as aerosol profiles quality check tools and processing through the SCC. Long term strategy for providing access to data and other related services is explored through the establishment of ACTRIS as a RI and for this reason ACTRIS is applying a position on the next ESFRI (European Strategy Forum on Research Infrastructures) roadmap for Research Infrastructures.

6.3 Cloud profiles

The Cloudnet database is a file-based database, due to the nature of the typical use-case and data volume. The infrastructure comprises an FTP server for incoming data streams, rsync server for outgoing data streams, processing server, webserver, with data storage distributed across a series of virtual file-systems including incremental backups. Due to the data volume, most sites also hold a copy of their own processed data, effectively acting as a second distributed database and additional backup.

The current size of the database is about 10 TB and the volume is expected to grow by close to 0.5 TB per year with the current set of stations and the standard products. However, there will be a significant increase in volume when the planned move to multi-peak and spectral products is undertaken; this is in addition to a slight increase arising through the creation of new products. The transfer of the database to FMI will ensure the long-term commitment for archiving and preservation. Publication of QA datasets will aid dataset preservation.

7 ACTRIS Data Centre– Organisation and personal resources

The ACTRIS Data Centre involves personal with broad and complementary background and competence. In total, more than 25 persons are involved in the data management, on full or part time.

A crucial structure of the ACTRIS data centre is the use of topical data centres run by scientists with expertise in the relevant field. This ensures not only proper curation of the data, which makes the data usable by both, experts and non-experts, but also a close connection to the data provider and user communities. A topical data centre run by scientists with data curation expertise serves as identifying elements built jointly with the data provider community, and as connecting element between data providers and users. The fundamental structure of the data centre is based on efficient use of complementary competence. This includes involvements of senior scientists, young scientists, engineers, programmers, and data base developers. A data centre serving several related communities, e.g. scientific and regulatory ones, are facilitating exchange and collaboration between these. Additionally, involvement of senior scientists working actively within various scientific communities is another pre-requisite, to ensure the links to various scientific user groups, for distribution of data products, and user oriented development of the data centre.

The ACTRIS data portal acts as umbrella for the topical data centres allowing search, download, and common visualisation of the data archived at the topical data centres. Maybe even more important, it will also connect ACTRIS with other European and international research data centres by allowing the same services for the data stored there by making use of latest inter-operability specifications. Also at the administrative plain, the ACTRIS portal represents the infrastructures in the relevant bodies working an unifying data management, and relays new developments to the whole infrastructure.

Appendix I:List of ACTRIS variables and recommended methodology

ACTRIS Aerosol particle variables Variable name	Recommended methodology	Validated data	NRT	Typical time res.	Higher timeres. available
Near-surface aerosol particle variables					
Particle light scattering coefficient	Integrating Nephelometer	X	X	1h	X
Particle light backscattering coefficient	Integrating Nephelometer	X	X	1h	X
Particle number size distribution	Mobility particle size spectrometer (e.g. differential mobility particle size, scanning mobility particle sizer) or Optical particle size spectrometer (e.g. optical particle counter, optical particle sizer) or Aerodynamic particle size spectrometer (e.g. aerodynamic particle sizer)	X	X	1h	X
Particle light absorption coefficient	Filter Absorption Photometer (e.g. Particle Soot/Absorption Photometer, Multi-Angle Absorption Photometry, Aethalometer)	X	X	1h	X
Particle number concentration	Condensation Particle Counter	X	X	1h	X
Cloud condensation nuclei number concentration	Condensation Cloud Nuclei Counter	X	X(later)	1h	X
Hygroscopic growth factor	Hygroscopicity Tandem Differential Mobility Analyzer	X		1h	X
Particulate organic and elemental carbon mass concentrations (OC/EC)	Filter sampling + evolved gas analysis with optical correction for charring (thermal-optical analysis)	X		1d-1week	
Particulate size-resolved chemical composition (organic & inorganic size-resolved mass speciation)	Aerosol Mass Spectrometer, Aerosol Chemical Speciation Monitor	X		1h	X
Particulate levogluocsan mass concentration	Filter sampling + offline methodology	X		1d-1week	

ACTRIS near-surface trace gas variables				
Variable	**Recommended methodology**	**Validated data**	**NRT**	**Approx. time resolution**
NMHCs (C2-C9 hydrocarbons) **See detailed list*	on-line: GC-FID, GC-MS, GS-FID/MS, GC-Medusa, PTR-MS off-line traps: ads-tubes off-line: steel canisters + glass flasks, combined with the on-line instruments in laboratories	X		1 h-2/week
OVOCs (oxidised volatile organic compounds as aldehydes, ketons, alcohols,) *See detailed list of the compounds at the end of the document*	on-line: GC-FID, GC-MS, GS-FID/MS, GC-Medusa, PTR-MS off-line traps: ads-tubes, DNPH-cartridge-HPLC	X		1 h-2/week
Terpenes (biogenic hydrocarbons with a terpene-structure) **See detailed list at the end of the document*	on-line (GC-FID, GC-MS, GS-FID/MS, GC-Medusa) and off-line traps (ads-tubes)	X		1 h-2/week
NO	NO-O_3 chemiluminescence	X	X	1 min - 1 h
NO2	indirect: NO-O_3 chemiluminescence coupled to photolytic converter (Xenon lamp (PLC) or diode (BLC)), direct: cavity ring down spectroscopy (CRDS), laser induced fluorescence (LIF), Cavity Attenuated Phase Shift Spectroscopy (CAPS)	X	X	1 min - 1 h
NOy (NO, NO2, NO3, N2O5, HNO2, HNO3, PAN, organic nitrates and aerosol nitrates sum of oxidized nitrogen species with an oxidation number >1, both organic and inorganic.)	indirect: NO-O3 chemiluminescence coupled to gold converter	X	X	1 min - 1 h

ACTRIS Aerosol particle variables				
Variable name	**Recommended methodology**	**Validated data**	**NRT**	**Approx. time resolution**
Column and profile aerosol particle variables (remote particle observations from ground)				
Aerosol backscatter coefficient profile	Backscatter lidar / Raman lidar/High spectral resolution lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Aerosol extinction coefficient profile	Raman lidar / High spectral resolution lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Lidar ratio profile	Raman lidar / High spectral resolution lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Ångström exponent profile	Multiwavelength Raman lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Backscatter-related Ångström exponent profile	Multiwavelength backscatter lidar / Raman lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Particle depolarization ratio profile	Depolarization backscatter lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Particle layer geometrical properties (height and thickness)	Backscatter lidar / Raman lidar/ High spectral resolution lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Particle layer optical properties (extinction, backscatter, lidar ratio, Ångström exponent, depolarization ratio, optical depth)	Multiwavelength Raman lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)
Aerosol optical depth (column)	Sun/sky photometer	x	x	
Planetary boundary layer height	Backscatter lidar / Raman lidar/ High spectral resolution lidar	X		0.5 h, 2+1 measur. per week + special events + CALIPSO overpasses (2.5 h)

ACTRIS cloud variables				
Variable	**Recommended methodology**	**Validated data**	**NRT**	**Approx. time /height resolution**
Column and profile cloud variables (remote observations from ground)				
cloud/aerosol target classification	cloud radar, lidar/ceilometer, NWP model or radiosonde (optional: microwave radiometer)	X	X	30 seconds / 60 metres
drizzle drop size distribution	doppler cloud radar, lidar/ceilometer, NWP model or radiosonde (optional: microwave radiometer)	X	X	30 seconds / 60 metres
drizzle water content	doppler cloud radar, lidar/ceilometer, NWP model or radiosonde (optional: microwave radiometer)	X	X	30 seconds / 60 metres
drizzle water flux	cloud radar, lidar/ceilometer, NWP model or radiosonde (optional: microwave radiometer)	X	X	30 seconds / 60 metres
ice water content	cloud radar, lidar/ceilometer, NWP model or radiosonde (optional: microwave radiometer)	X	X	30 seconds / 60 metres
liquid water content	cloud radar, lidar/ceilometer, microwave radiometer	X	X	30 seconds / 60 metres
liquid water path	dual- or multi-frequency microwave radiometers (ceilometer useful for identifying clear-sky)	X	X	30 seconds
rainrate	drop-counting raingauge or disdrometer preferable to tipping bucket raingauges	X	X	30 seconds
Near-surface cloud variables				
Liquid Water Content	In-situ cloud-microphysical sensors	X		5 min

Detailed list of trace gases included in ACTRIS - *Alkanes, Alkenes, Alkynes*

Alkanes	ethane	2-methylhexane	**Alkenes**	ethene	**Alkynes** ethyne
	propane	n-heptane		propene	proypne
	2-methylpropane	2-2-4-trimethylpentane		trans-2-butene	1-butyne
	n-butane	3-methylheptane		1-butene	
	2-2-dimethylpropane	n-octane		2-methylpropene	
	2-methylbutane	n-nonane		cis-2-butene	
	n-pentane	n-decane		1-3-butadiene	
	cyclopentane	methyl-cyclohexane		3-methyl-1-butene	
	methyl-cyclopentane	n-undecane		2-methyl-2-butene	
	2-2-dimethylbutane	n-dodecane		trans-2-pentene	
	2-3-dimethylbutane	n-tridecane		cyclopentene	
	2-methylpentane	n-tetradecane		1-pentene	
	3-methylpentane	n-pentadecane		cis-2-pentene	
	cyclohexane	n-hexadecane		1-hexene	
	n-hexane			isoprene	
	methyl-cyclohexane				
	2-2-3-trimethylbutane				
	2-3-dimethylpentane				
	2-2-dimethylpentane				
	2-4-dimethylpentane				
	3-3-dimethylpentane				
	3-methylhexane				

Detailed list of trace gases included in ACTRIS - *OVOCs, Terpenes, Aromatics*

OVOCs	methanol	methylethylketon	**Terpenes**	alpha-thujene	**Aromatics**	benzene
	ethanol	methacrolein		tricyclene		toluene
	isopropanol	methylvinylketon		alpha-pinene		ethylbenzene
	n-propanol	glyoxal		camphene		m-p-xylene
	n-butanol	methylglyoxal		sabinene		o-xylene
	methyl-butanol	butylacetat		myrcene		1-3-5-trimethylbenzene
	formaldehyde	acetonitrile		beta-pinene		1-2-4-trimethylbenzene
	acetaldehyde			alpha-phellandrene		1-2-3-trimethylbenzene
	n-propanal			3-carene		
	n-butanal			alpha-terpinene		
	pentanal			m-cymene		
	hexanal			cis-ocimene		
	heptanal			p-cymene		
	octanal			limonene		
	decanal			beta-phellandrene		
	undecanal			eucalyptol		
	benzaldehyde			gamma-terpinene		
	acrolein			terpinolene		
	acetone			camphor		