GAIA-CLIM Report / Deliverable D7.3

Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:
Data Management Plan (DMP)

A Horizon 2020 project; Grant agreement: 640276
Date: 22 March 2018
Lead Beneficiary: NERSC
Nature: ODRP
Dissemination level: PU

Work-package	WP 7 (Project Management)
Deliverable	D7.3
Title	Data Management Plan
Nature	ODRP
Dissemination	PU
Lead Beneficiary	Nansen Environmental and remote Sensing Center, Norway (NERSC)
Date	22 March 2018
Status	Final
Authors	Anna Mikalsen(NERSC), Peter Thorne (NUIM)
Editors	Peter Thorne (NUIM)
Reviewers	
Contacts	anna.mikalsen@nersc.no, peter.thorne@nuim.ie
URL	http://www.gaia-clim.eu/

This document has been produced in the context of the GAIA-CLIM project. The research leading to these results has received funding from the European Union's Horizon 2020 Programme under grant agreement n° 640276. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view

Document version control

Version	*Updated by*	*Date*
V 1.0	Anna Mikalsen	Public Deliverable 7.1, 31/7/15
V 1.1	Corinne Voces	7/6/15
V 1.2	Richard Davy, Peter Thorne	14/07/15
V 1.3	Corrine Voces, Richard Davy, Peter Thorne	03/08/16
V 1.4	WP leads, Richard Davy	Public Deliverable 7.2, 25/08/16
V 1.5	Anna Mikalsen, Peter Thorne, Fabio Madonna	01/03/18
V1.6	Anna Mikalsen, Peter Thorne, Fabien Carmanati, Joerg Schulz	Public deliverable 7.3, 22/03/18

Table of Contents

D.7.3: Data management plan commensurate with the Pilot on Open Research Data, third (final) version, March 2018

Project Name: Gap Analysis for Integrated Atmospheric ECV Climate Monitoring (GAIA-CLIM)

Funder: European Commission (Horizon 2020)

Grant Title: No 640276

1. Project brief description

The Gap Analysis for Integrated Atmospheric ECV Climate Monitoring (GAIA-CLIM) Project endeavoured to establish improved methods for the characterisation of satellite-based Earth Observation (EO) data by surface-based and sub-orbital measurement platforms for six of the GCOS atmospheric Essential Climate Variables (ECVs), namely, Temperature, Water Vapour, Ozone, Carbon Dioxide, Methane, and Aerosols. GAIA-CLIM added value by:

- Objectively assessing and mapping existing measurement capabilities
- Improving traceability and uncertainty quantification on sub-orbital measurements;
- Quantifying co-location uncertainties between sub-orbital and satellite data;
- Using traceable measurements in data assimilation; and
- Providing co-location match-up data, metadata, and uncertainty estimates via a 'virtual observatory' facility.

The novel approach of GAIA-CLIM was to demonstrate comprehensive, traceable, EO Cal/Val for a number of metrologically mature ECVs, in the domains of atmospheric state and composition, that will guarantee that products are assessable and intelligible to third-party users.

Further details on GAIA-CLIM's project outcomes can be found at www.gaia-clim.eu

2. Outline of GAIA-CLIM's policy for data management

GAIA-CLIM has been a member of the Open Data Pilot under H2020. The project promoted the processing and sharing of data openly in support of project aims of enhancing the long-term value of EO data for the scientific community. The purpose of this Data Management Plan (DMP) is to document the collection and use of data records that were managed within the GAIA-CLIM project. This third and final version of the DMP, updated from previous versions D7.1 and D7.2, reflects the final status of the project in relation to data produced and/or collected. It provides a final agreed record of the data management policies of GAIA-CLIM in respect of data dissemination.

This DMP ensured that:

- There has been a coherent and evolving approach as to what, specifically, is required on management of data by the consortium throughout the lifetime of the project, and after its completion.
- Project findings are publicly available both during and after the project (including the virtual observatory, which will continue to be accessible on EUMETSAT servers after the end of the project). In addition, that they represent a lasting legacy to the contributing observing networks, leading to improvements in data traceability and comparability of EO measurement systems.
- Data preservation strategies are in place in support of long-term use of project outcomes.
- Data usage by GAIA-CLIM respects conditions of use, policy on access, and intellectual property rights of the primary data collectors, including authorship and acknowledgement that accurately reflects the contributions of those involved.

It should be stressed that GAIA-CLIM constituted a rather particular case in terms of data management as covered by the guidance pertaining to the preparation of DMPs under the H2020 Pilot on Open Research Data. The project did not directly collect primary data, i.e., make measurements for the sole purpose of the project. Rather, it provided added value and additional metadata to existing measurements (by optimizing the value of multiple sources of primary data to enable traceable characterization of EO data) taken by both consortium members under separate funding support and by third party institutions. Therefore, and in line with the project objectives, as used in this document, the term 'project data' refers to metadata and / or value-added products, i.e. secondary data products arising from primary data created, hosted, and managed by existing networks and stations, in order to improve global capabilities to use non-satellite data to characterise space-borne satellite measurement systems.

In this context, it is important to stress that GAIA-CLIM has retained only that primary data used in its mission, which solely constitutes a small subset of the primary data made available by the contributing observing networks. It was never the intention of GAIA-CLIM to create or curate a comprehensive archive from underlying primary observational networks, nor would it have been practical to do so. Contributing networks retain primary Intellectual Property Rights (IPR) and may well, in future, revise their data, data formats, metadata etc. We note in particular that the C3S 311a Lot 3 activity[1], instigated in 2017 (and led by GAIA-CLIM participants), constitutes an operational service for accessing the baseline and reference in-situ data holdings. This is a more appropriate mechanism to address the issue, providing a sustainable long-term resource for multiple use cases.

Sharing of value added products derived from GAIA-CLIM is important to the long-term study of EO sensor performance, validation of satellite-derived data products, and to maximize their value in climate applications. In this regard, this DMP focusses primarily upon supporting the scientific findings of the project, which has been actively working to create appropriate linkages, and maximize the availability and utility of the data and tools produced after the project. The release of the data and associated products into the public domain has been the *de facto* policy, ensuring usability and long-term availability of data to scientific and public audiences alike. Collaboration with internal and

[1] See http://www.ciao.imaa.cnr.it/index.php?option=com_content&view=article&id=220&Itemid=288 for further information.

external project partners will remain ongoing to ensure this takes place on a best endeavours-basis post-project cessation.

The main way of actually serving data from the GAIA-CLIM project is through its 'virtual observatory' tool[2], developed and hosted by TUT and EUMETSAT. The virtual observatory is intended to provide the user with access to both metadata and observational data from different ground-based reference networks with co-located satellite data. GAIA-CLIM has used measurements of metrologically "reference-quality" (from GRUAN and NDACC networks) that are traceable and have well quantified uncertainty estimates and measurements (from AERONET), for which no sufficient evidence has yet been provided to assess the reference-quality status, but which are close to that status. A full listing of contributing observations is available at the end of this document under Annex 2. Importantly, GAIA-CLIM only made use of those primary observations to which no academic restrictions to use, re-use, and re-distribution currently apply. The providers of primary data from these networks have thereby either implicitly or explicitly agreed to release that portion of their data which we have utilised according to this DMP. Similarly, GAIA-CLIM has only used satellite data available for re-use and re-distribution. Furthermore, re-analysis and Numerical Weather Prediction (NWP) data are also part of the virtual observatory. Such data arose from within the consortium (ECMWF and MO partners under WP4) without restriction. However, GAIA-CLIM work in many cases built upon pre-existing capabilities of the partners. In a restricted subset of these cases, IPR restrictions relate to these background materials as articulated in the Consortium Agreement (cf. Annex 1). A list of co-located datasets available in the virtual observatory is given in Annex 3.

The virtual observatory data policy is made explicit and is in compliance with the H2020 Pilot on Open Research Data (s. next section) and this DMP.

Project parts that dealt with enhancing existing primary data streams were:

- Preparation and assessment of reference-quality non-satellite data (including in global assimilation systems) and characterisation of key satellite datasets:

  a. Assessment of several new satellite missions, using data assimilation of reference-quality non-satellite measurements, targeting temperature and humidity (under work package 4).

  b. Development of infrastructure to deliver quantified uncertainties for reference-data co-locations with satellite measurements (under work packages 3 and 5).

  c. Development of capabilities for preparation, monitoring, analysis, and evaluation of reference-quality data (under work packages 2 and 5).

  d. Development of a general methodology for using reference-quality non-satellite data for the characterisation of EO data (under work packages 4 and 5).

- Creation and population of a virtual observatory:

[2] The virtual observatory is available from: http://gaia-clim.vo.eumetsat.int/vo/#/

a. Creation of a collocation database between EO measures and reference-quality measurements.

b. Adoption of ISO/WIGOS and ESA-CCI standards for observational metadata in the virtual observatory.

c. Preparation of data to enable comparisons, including relevant uncertainty information and metadata for users to understand and make appropriate use of the data for various applications.

d. Creation of data interrogation and visualization tools for both non-satellite and satellite observing capabilities, building upon existing European and global infrastructure capabilities offered by partners and in-kind collaborators.

e. Planning for the potential transition of the resulting virtual observatory from research to operational status in support of the Copernicus Climate Change Service (C3S) and Copernicus Atmospheric Monitoring Service (CAMS).

3. Pilot on Open Research Data

GAIA-CLIM participated in the H2020 Pilot on Open Research Data. Knowledge generated during the project has been shared openly. Any milestones, deliverables, or technical documents produced, which were deemed public in the Grant Agreement, have all been published online and made discoverable. Peer-reviewed publications have all been submitted to journals that are either open access or allow the authors to pay for the articles to be made open access (for such instances, the additional charges have been paid).

4. Dissemination and Exploitation of Results

In order to maximize the benefit and usability of project findings, GAIA-CLIM incorporated a strong focus on user interaction throughout the life cycle of the project. The virtual observatory, as a key outcome of the project, is the primary means of dissemination of data and associated project findings through which end-users are able to access, visualize, and utilize the outputs of the project. The virtual observatory built upon and extended a number of existing facilities operated by project partners, which already undertook subsets of the desired functionality, specifically:

- the Network of Remote Sensing Ground-Based Observations in support of the Copernicus Atmospheric Service (NORS);
- the Cloud-Aerosol-Water-Radiation Interactions (ICARE) Project;
- the US National Oceanic and Atmospheric Administration (NOAA) Products Validation System (NPROVS).

The resulting virtual observatory facility is entirely open and available to use for any application area. All downloaded data are provided in NetCDF-4 format and use the Climate and Forecast (CF) metadata convention. Significant efforts have been made to build an interface that is easy to use and which makes data discovery, visualization, and analysis user-friendly. The virtual observatory work package

included a specific task dedicated to documenting the steps required to transition this facility from a research to an operational framework with a view to constituting a long-term infrastructure (see deliverable D5.8 Transition roadmap for the virtual observatory[3]).

The GAIA-CLIM website[4] represents the public interface of the project and shall remain available for at least five years after the end of GAIA-CLIM (on a best endeavours maintenance basis from NERSC). It provides an overview of the main results, activities by work package, as well as an open portal to disseminate information on project outcomes such as publications, peer-reviewed journal articles, and project deliverables. This particularly includes access to:

- The Library of (1) smoothing/sampling error estimates for key atmospheric composition measurement systems, and (2) smoothing/sampling error estimates for key data comparisons,
- The Product Traceability and Uncertainty (PTU) documents developed by WP2.
- The full list of gaps in knowledge and observing capability identified within the scope of GAIA-CLIM;
- The virtual observatory, and
- The 'GRUAN Processor' tool.

5. Preservation of value added data products

The value-added tools and data products GAIA-CLIM has retained and made available consist of:

- Metadata collected under work package 1 relating to networks and their measurement maturity. This included:
    - Station location metadata to ISO standard 19115;
    - Measurement system metadata;
    - Visualisation capabilities;
    - 3D-tool design for the visualisation of existing measurements online;
    - Measurement maturity assessment metadata;
    - Observational metadata following WIGOS and ESA-CCI standards.
- Selected primary data that meets co-location criteria and is deemed reference quality or almost reference quality from underlying networks. (work package 2)
- Co-location uncertainty information arising from a variety of approaches, including statistically based and dynamically based estimation (work packages WP3 and WP4).
- The "GRUAN Processor" tools to convert from geophysical to Top of the Atmosphere (TOA) radiance (WP4).
- Capabilities to visualize, subset, and analyse the co-location database (work package WP5).

Most of the above products and capabilities have been hosted and preserved at the www.gaia-clim.eu domain. Those tools and capabilities preserved elsewhere are described in subsequent sub-sections.

[3] D5.8 Transition Roadmap for the Virtual Observatory: http://www.gaia-clim.eu/system/files/document/d5.8.pdf

[4] www.gaia-clim.eu

Were follow-on support available, then new data and services may be able to be appended and the facilities made operational. These new data and capabilities would be subject to the data policies of the provider and funder at that time. GAIA-CLIM shall undertake solely to preserve and make available the data and functionalities created during the project lifetime.

5.1 Virtual observatory

The virtual observatory constitutes the primary means of dissemination of project results. The virtual observatory facility is entirely open and available to use for any application area. Data versioning, source locations, and any DOIs from the primary data sources are retained. Significant efforts were undertaken to collaborate with existing European and international programs with similar aims in order to produce a facility which makes data discovery, visualization, and analysis easy and useful for the end users, while optimizing the use of reference data. The objective was to develop an interface that uses software tools to deliver products in standard formats, such as NetCDF that are compliant with CF conventions. Such formats are self-describing and provide a definitive description of what the data in each variable represents, as well as the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction and display capabilities. In turn, this may support the relevance and sustainability of the facility, or aspects thereof, into the future, as well as effective data preservation.

The final data products will be kept beyond the lifetime of the project through the virtual observatory, which shall be hosted by EUMETSAT. It is important to stress that the virtual observatory is solely a demonstrator project and therefore the GAIA-CLIM data, documentation, and functionalities will be retained in "frozen" mode in the state they existed in at the end of the project with the aim of becoming further developed and integrated into the emerging Copernicus or any other service. If continued in this way, data and software distribution policies of the respective service will be applied in the long-term.

The virtual observatory has been designed and developed as a traditional client-server application. Visible to the user is the graphical user interface (GUI) that allows the user to interact with the components of the virtual observatory. The GUI is used to send queries to retrieve data that result in graphical displays. The data themselves are stored in a non-relational database that holds all the data including the uncertainties of the measurements and the co-locations available in the virtual observatory. The non-relational data base is very versatile, allowing easy addition of new data of various types to the virtual observatory and makes extensions in the future relatively simple, should the virtual observatory be further developed and made operational.

5.2 Metadata discovery tool

The metadata discovery and visualization tool is planned to be maintained by CNR over long-term, including annual updates of the metadata and the GUI. The discovery and observational metadata was realised in PostgreSQL and then linked to a MongoDB via GeoServer platform, all of which are open source. The 3D tool uses the Cesium JavaScript library, which is distributed under the Apache 2.0 license agreement. To remain consistent with the conditions of use in our 3D tool, the Cesium logo

and a link to its website are present. The source code of the 3D tool itself is openly available on Github[5], including documentation[6].

5.3 Library of smoothing/sampling uncertainties

Research was undertaken within GAIA-CLIM to improve quantification of the co-location mismatch uncertainties. Several methods have been developed for - and applied to - the quantification of smoothing and sampling issues in a range of atmospheric ground-based measurement techniques, and to estimate the uncertainties that need to be taken into account when comparing non-perfectly co-located ground-based and satellite measurements with different spatio-temporal smoothing and sampling properties. The resulting software and Look-up tables that constitute input to the virtual observatory have been documented and shared openly without restriction. The actual guiding material and data files are hosted by BIRA and available for download by anonymous/guest ftp at: ftp://ftp-ae.oma.be/dist/GAIA-CLIM/D3_6/

5.4 GRUAN processor

A stand-alone 'GRUAN Processor' module has been developed based on a core radiative transfer modelling capability built around two existing open-source software packages (EUMETSAT's NWP SAF[7] RTTOV fast radiative transfer model and Radiance Simulator). This software, referred to as the GRUAN Processor, enables the comparison of collocated geophysical fields and simulated brightness temperature between radiosonde and model fields.

The Processor will allow improved NWP-based calibration, recalibration, and validation of satellite instruments thanks to robust channel-by-channel uncertainty estimates. In addition, it is expected to serve as a long-term, semi-automatic, monitoring tool for the Met Office NWP global model. The integration and automation of the Processor in the Met Office system is expected to take place during the fiscal year 2018/2019. It is to the discretion of ECMWF to use their copy of the Processor (installed and used within the scope of the GAIA-CLIM project), and/or its future versions, as an additional monitoring tool in their system.

This work is in line with the Copernicus CAMS and C3S streams and has the potential to become an operational Copernicus service with data and graphic-based monitoring available from the Copernicus portal. The Processor post-processed outputs are publicly available on a demonstrator web-page[8] hosted by NWP SAF.

[5] The GAIA-CLIM metadata discovery and visualization software is open accessible on Github: https://github.com/emanueletramutola/gaiaclim

[6] https://github.com/emanueletramutola/gaiaclim/blob/master/README.md https://github.com/emanueletramutola/gaiaclim/blob/master/LICENSE.

[7] www.nwpsaf.eu

[8] https://www.nwpsaf.eu/GProc_test/ins.shtml

6. Primary source datasets used within GAIA-CLIM

The final contributing networks to the virtual observatory were as follows:

1. GRUAN,
2. NDACC, and
3. AERONET.

Further details on these networks, their governance, and their data policies are given in Annex 2. A list of co-located datasets available in the virtual observatory is given in Annex 3. As primary data collectors, these networks have assessed data quality, integrity, originality, and content prior to publishing. Whilst GAIA-CLIM activities under work package 2 lead to changes in how a subset of the data are processed by the underlying networks, GAIA-CLIM was a user, not at a provider, of these primary data products.

As described previously, GAIA-CLIM has respected the data policies and practices of the data originators/custodians, and the documentation herein should not be taken to imply advocacy for changing their existing policies. Rather, it is important to note that GAIA-CLIM activities and this DMP work alongside and document existing practices that pertain to the source data. Where networks have data policies that place restrictions on near-real-time use, GAIA-CLIM has only used the open delayed-mode data.

7. Summary

This Data Management Plan presents the data management policy that has been used by the GAIA-CLIM partners in their collection and use of data records during the lifespan of the project. This third and final DMP version has evolved to reflect the conclusion of the project. Since there was no primary data produced under the GAIA-CLIM project, the data management policy relates to metadata and added-value products produced in GAIA-CLIM for existing and future measurements. This is in keeping with the over-arching objectives of GAIA-CLIM to provide the necessary methodologies for the characterization of satellite-based EO data using surface and sub-orbital measurements.

Annex 1: Specific limitations and/or conditions of background material covered within the GAIA-CLIM consortium agreement

In concordance with the GAIA-CLIM consortium agreement, the following background is hereby identified and agreed upon for the Project. Specific limitations and/or conditions, shall be as mentioned hereunder:

Describe Background	**Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)**	**Specific limitations and/or conditions for exploitation (Article 25.3 Grant Agreement)**
Satellite Data for the purposes of WP4. (a) Within scope of WMO Resolution 40 (b) Outside scope of WMO Resolution 40	The Data described is third-party background and for the purposes of; (a) it can be used on a royalty-free basis and in compliance with any terms associated with WMO Resolution 40. (b) can be used subject to the terms and conditions of the original owners	
Software and Coding for the purposes of WP4 (which includes associated algorithms)	These will be fully documented and included as part of the Deliverables of the Project and as such will not be subject to any specific restrictions on use, beyond those agreed in the Consortium Agreement.	
Data to be used for Mapping Geographical Capabilities in WP1	The Data described contains third-party Background, which is split into two categories; (a) AMDAR data, the use of such information being subject to the terms and conditions imposed by AMDAR. (b) Information of Airline Communication Equipment and Flight Routes will also form part of the Data, however all this Information is publically available.	
Data, Modelling and Software used for Met Office contribution to WP2	Mode-S data, observations and comparisons with the UKV model will be Met Office owned Background, to which no special conditions beyond those agreed in the Consortium Agreement shall apply.	

Annex 2: Updated list of contributing observations

This annex details specifics of each contributing set of observations sufficient to understand the data and its policies. Each data source is discussed under a consistent set of headings viz:

- Dataset reference and name
- Dataset description
- Standards and metadata
- Data sharing
- Archiving and preservation

Note that GAIA-CLIM shall not act as a mirror repository for these primary data sources as a whole but shall rather retain only those subsets used within the Virtual Observatory. Users should go to the primary source which may, amongst others, reflect: subsequent updates to formatting; new observations taken; or reprocessing based upon improved understanding.

1. GRUAN

Data set reference and name

Global Climate Observing System (GCOS) Reference Upper Air Network (GRUAN)

Data set description

A group of stations coordinated by the GRUAN Lead Centre, hosted by the German Meteorological Service, DWD. Data products that meet necessary conditions of traceability and uncertainty quantification, documentation and publication are served via the US National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NOAA NCEI) , hosting and providing access to one of the most significant archives on earth, with comprehensive oceanic, atmospheric, and geophysical data

Standards and metadata

Data and comprehensive metadata must be undertaken according to stated requirements (documented through a technical document), shared with a central processing facility, and traceable to either SI or community accepted standards. The processing is open and transparent.

Data sharing

Data are shared publicly without restriction or delay via NOAA NCEI.

Archiving and preservation (including storage and backup)

Archived data is fully and openly available, to the extent permitted by US law and subject to valid privacy, confidentiality, security, or other restrictions. The archive is on a secure backed-up service and a copy is retained at the GRUAN Lead Centre. Entire data streams are periodically reprocessed

when new insights on instruments accrue. Such reprocessing always incurs a change in version number and associated documentation.

2. NDACC

Data set reference and name

Network for the Detection of Atmospheric Composition Change (NDACC)

Data set description

The NDACC is composed of more than 70 high-quality, remote-sensing research stations [9] for observing and understanding the physical and chemical state of the stratosphere and upper troposphere and for assessing the impact of stratospheric changes on the underlying troposphere and on global climate. While the NDACC remains committed to monitoring changes in the stratosphere with an emphasis on the long-term evolution of the ozone layer, its priorities have broadened considerably to encompass issues such as the detection of trends in overall atmospheric composition and understanding their impacts on the stratosphere and troposphere, and establishing links between climate change and atmospheric composition. A wide variety of trace gases are measured[10].

Standards and metadata

NDACC is organized in several working groups[11], which are predominantly based on the applied measurement techniques: i.e. Brewer & Dobson, FTIR, Lidar, Microwave, Satellite, Sondes, Spectral UV, Theory, UV/Vis and Water Vapor. To ensure quality and consistency of NDACC operations and products, a number of protocols have been formulated covering topics such as measurement and analysis procedures, data protocol, instrument inter-comparisons, theory and analysis, validation, and Cooperating Networks[12]. Regular working group meetings and instrument inter-comparisons are held to safeguard a continued high standard of the network's products.

Data sharing

All NDACC data is subject to an internal analysis/ verification process, NDACC data more than one year old will then be available to anyone through centralized scientific data archiving and distribution facilities[13]. However, many NDACC investigators have agreed to make their data publicly available

[9] A map of the NDACC stations can be found on: http://www.ndsc.ncep.noaa.gov/; and a list of all the currently active stations, including a description of the instrumentation and data products, is available on: http://www.ndsc.ncep.noaa.gov/sites/.

[10] The observational capabilities can be displayed by a chart on: http://www.ndsc.ncep.noaa.gov/obs_chart/obs-capabil-2015-10-14.pdf

[11] Split into instrument and theme working groups: http://www.ndsc.ncep.noaa.gov/organize/

[12] The NDACC protocols are accessible under http://www.ndsc.ncep.noaa.gov/organize/protocols/

[13] A directory summarizing the operational status of the NDACC is available at: http://www.ndsc.ncep.noaa.gov/data/madir/. Long-term NDACC measurement activities are listed in Sections Ia and Ib, while Sections IIa and IIb include NDACC measurement activities conducted intermittently or during limited duration campaigns.

immediately upon archiving. The public record is available through anonymous ftp[14]. The use of NDACC data prior to its being made publicly available (i.e., for field campaigns, satellite validation, etc.) is possible via collaborative arrangement with the appropriate principal investigators PI(s). Rapid delivery data, which will likely be revised before entry in the full database, is also available for some instruments[15].

In all cases when NDACC data is used in a publication, the authors agree to acknowledge both the NDACC data center and the data provider. Whenever substantial use is made of NDACC data in a publication an offer of co-authorship will be made to the associated NDACC investigator through personal contact with the data providers and/or owners. Users of NDACC data are also expected to consult the on-line documentation and reference articles to fully understand the scope and limitations of the instruments and resulting data, and are encouraged to contact the appropriate NDACC PI (listed in the data documentation on the web page) to ensure the proper use of specific data sets. Those using NDACC data in a talk or paper are asked to acknowledge its use, and to inform the 'Theory and Analysis Working Group' PIs of any relevant publications.

Archiving and preservation (including storage and backup)

All data are released to the public and available on the anonymous ftp site no more than one year after measurement date. Data and comprehensive metadata is accessible via the NDACC data table[16] and clicking on the station name will take the user to the associated public data site.

3. AERONET

Data set reference and name

AErosol RObotic NETwork

Data set description

AERONET[17] is a global network of sun/sky radiometers that is monitoring Aerosol Optical Depth (AOD) and a limited subset of aerosol optical properties for AOD trend analysis, for optical properties characterization, and for the validation of satellite retrievals. Cloud products are also available for a part of the network. At several stations, sun photometer measurements are co-located with Raman lidar measurements.

AERONET collaboration provides globally distributed observations of spectral AOD, inversion products, and precipitable water in diverse aerosol regimes. Version 3 AOD data are computed for three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened and quality controlled),

[14] Regular NDACC data: ftp.cpc.ncep.noaa.gov/ndacc/station

[15] NDACC rapid delivery data: ftp://ftp.cpc.ncep.noaa.gov/ndacc/RD . As soon as the standard verified version is available, the RD data will be removed and the fully verified version will be archived in the NDACC public database.

[16] http://www.ndsc.ncep.noaa.gov/data/data_tbl/

[17] https://aeronet.gsfc.nasa.gov/

and Level 2.0 (quality-assured). Inversions, precipitable water, and other AOD-dependent products are derived from these levels and may implement additional quality checks.

Standards and metadata

The network imposes standardization of instruments, calibration, data processing, and distribution. The measured radiances are automatically sent to the NASA – Goddard Space Flight Center (NASA-GSFC), where they are processed according to the standardized AERONET data analysis technique. Quality control is provided by AERONET during processing and retrieval of the geophysical quantities. Once these data sets are manually inspected, they can then be upgraded to Level 2 (quality assured data product).

Data sharing

AERONET data are distributed via the internet through a web tool and FTP. The primary distribution method of AERONET data is the AERONET web download tool. This tool provides access to most AERONET data products (e.g., AOD and inversions). For special requests and products not available through the download tool (e.g., data for specific instruments), an FTP transfer can be setup to update remote systems.

AERONET uses the ASCII text format for data dissemination. These data are provided within compressed zip files to improve data transfer between systems.

Archiving and preservation (including storage and backup)

The AERONET program provides a long-term and readily accessible public domain database of aerosol optical, microphysical, and radiative properties for aerosol research and characterization and validation of satellite retrievals, and synergies with other databases.

The processing algorithms have evolved from Version 1.0 to Version 2.0 and now Version 3.0. The Version 3 databases are available from the AERONET and PHOTONS websites. Version 2 data may be downloaded from the website through 2018 and thereafter upon special request. New AERONET products will be released as new measurement techniques and algorithms are adopted and validated by the AERONET research community. The AERONET website also provides AERONET-related news, a description of research and operational activities, related Earth Science links, and an AERONET staff directory.

Annex 3: Co-located datasets available in the virtual observatory

ECV / Type	Method	From	To	No of co-locations	No of stations
Brightness temperature / Co-location	ECMWF vs HIRS on Metop-A	01.01.2013	31.12.2016	33 096	16
Brightness temperature / Co-location	UK MetOffice vs HIRS on Metop-A	01.01.2013	31.12.2016	51 816	17
Brightness temperature / Co-location	GRUAN Processor (ECMWF) vs HIRS on Metop-A	01.01.2013	31.12.2016	33 038	16
Brightness temperature / Co-location	GRUAN Processor (UK MetOffice) vs HIRS on Metop-A	01.01.2013	31.12.2016	51 674	17
Brightness temperature / Co-location	ECMWF vs HIRS on Metop-B	01.01.2013	31.12.2016	21 432	16
Brightness temperature / Co-location	UK MetOffice vs HIRS on Metop-B	01.01.2013	31.12.2016	21 720	17
Brightness temperature / Co-location	GRUAN Processor (ECMWF) vs HIRS on Metop-B	01.01.2013	31.12.2016	21 411	16
Brightness temperature / Co-location	GRUAN Processor (UK MetOffice) vs HIRS on Metop-B	01.01.2013	31.12.2016	21 689	17
Brightness temperature / Co-location	ECMWF vs HIRS on NOAA19	01.01.2013	31.12.2016	5 716	16

ECV / Type	Method	From	To	No of co-locations	No of stations
Brightness temperature / Co-location	UK MetOffice vs HIRS on NOAA19	01.01.2013	31.12.2016	23 540	17
Brightness temperature / Co-location	GRUAN Processor (ECMWF) vs HIRS on NOAA19	01.01.2013	31.12.2016	5 708	16
Brightness temperature / Co-location	GRUAN Processor (UK MetOffice) vs HIRS on NOAA19	01.01.2013	31.12.2016	23 497	17
Aerosol optical depth / Co-location	AERONET Sunphotometer vs AATSR on Envisat	01.07.2002	08.04.2012	30 086	527
Ozone / Co-location	NDACC DOAS vs GOME-2	23.01.2007	30.09.2017	12 710	2
Ozone / Co-location	NDACC FTIR vs GOME-2	29.01.2007	21.09.2017	17 409	10