

GAIA-CLIM Report / Deliverable D7.2

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



A Horizon 2020 project; Grant agreement: 640276

Date: 31st Aug 2016

Lead Beneficiary: NERSC

Nature: ODRP

Dissemination level: PU





Work-package	WP 7 (Project Management)
Deliverable	D7.2
Title	Data Management Plan
Nature	ODRP
Dissemination	PU
Lead Beneficiary	Nansen Environmental and remote Sensing Center, Norway (NERSC)
Date	31 st Aug 2016
Status	Draft
Authors	Anna Mikalsen(NERSC) Corinne Voces (NUIM), Richard Davy (NERSC)
Editors	Peter Thorne (NUIM)
Reviewers	
Contacts	Richard.davy@nersc.no , corinne.voces@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

<u>Version</u>	<u>Updated by</u>	<u>Date</u>
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	25/08/16

Table of Contents

D.7.2: Data management plan commensurate with Pilot on Open Research Data, second version, September 2016.....	5
1. Project brief description	5
2. Outline of GAIA-CLIM's policy for data management	5
3. Pilot on Open Research Data.....	8
4. Dissemination and Exploitation of Results.....	8
5. Preservation of value added data products.....	9
6. Primary source datasets currently envisaged to be used within GAIA-CLIM.....	10
7. Scientific research data should be easily:	11
1. Discoverable	11
2. Accessible.....	11
3. Assessable and intelligible	11
4. Usable beyond the original purpose for which it was collected	12
5. Interoperable to specific quality standards	12
Summary.....	13
Annex 1: Specific limitations and/or conditions of background material covered within the GAIA-CLIM consortium agreement	14
Annex 2: Updated list of contributing observations	15
1. GRUAN	15
2. TCCON.....	16
3. ACTRIS-2.....	17
3a. EARLINET.....	18
4. WCCOS	19
4a. SHADOZ	21
5. NDACC.....	21

D.7.2: Data management plan commensurate with Pilot on Open Research Data, second version, September 2016

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 is establishing sound 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) Temperature, Water Vapour, Ozone, Carbon Dioxide, Methane and Aerosols. GAIA-CLIM is adding value by:

- 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.

Further details on GAIA-CLIM can be found at www.gaia-clim.eu

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

The purpose of this Data Management Plan (DMP) is to set up a coherent approach to the collection and use of data records managed within the GAIA-CLIM project. It shall promote 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. This DMP version reflects the evolution of the project at the time of writing, based on the current plans of the consortium in relation to data produced and/or collected during the lifespan of the project and progress to date. The third and final version of this DMP, due at the end of the project, month 36, will provide a final agreed record of the data management needs and policies of GAIA-CLIM in respect of data dissemination, and sharing widely project findings and recommendations in line with the H2020 Pilot on Open Research Data to which GAIA-CLIM is a signatory.

This DMP shall ensure that:

- There is 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. 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 constitutes 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 is not envisaged to directly collect primary data, i.e. make measurements for the sole purpose of the project. Rather it will provide added value and additional metadata to existing and forthcoming 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 will retain only that primary data used in its mission, which will constitute a small subset of the primary data made available by the contributing observing networks. It is not the intention of GAIA-CLIM to create or curate a comprehensive archive from underlying primary observational networks, nor would it be practical to do so. Contributing networks retain primary IPR and may well, in future, revise their data, data formats, metadata etc. beyond the lifetime of GAIA-CLIM. We note the recent C3S tender (Copernicus Climate Change Service Volume II, June 2016), which shall perform this task under Lot 3 if awarded to one or more consortia. As a long-term sustainable service this constitutes a more appropriate mechanism to address this issue, providing a sustainable long-term resource for multiple use cases.

Sharing of value added products derived from this project is vital 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 will focus primarily upon supporting the scientific findings of the project, actively working to create appropriate linkages and maximize the availability and utility of the data and tools produced once the project activities have concluded. Conclusion of the project is defined as the date upon which the final deliverables have been achieved, when the final version of published data will be made available in a long-term repository. The release of the data and associated products into the public domain is the *de facto* policy, ensuring usability and long-term availability of data to the scientific and public audiences alike. Collaboration with internal and external project partners will be an ongoing priority to ensure this takes place.

GAIA-CLIM shall primarily use metrologically reference quality measurements that are traceable and have well quantified uncertainty estimates. A full current listing of envisaged contributing observations is available at the end of this document under Annex 2. This list has been revised from the list provided in D7.1, in light of project activities carried out up to the time of writing under task 1.2, which has enabled a comprehensive mapping of current observational capabilities and

classification of the reviewed datasets. Further updates to the list of contributing data streams may accrue over the remainder of the project lifetime and shall be documented in the final DMP, should they arise.

Importantly, GAIA-CLIM will only make 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 shall thereby either implicitly or explicitly agree to release that portion of their data we have utilised according to this data management plan; otherwise their data will not be used in GAIA-CLIM. The usage of satellite data has to follow the data policies prescribed by the satellite operators, although GAIA-CLIM will only use those data where the rights for re-use and re-distribution in the 'virtual observatory' can be attained. In reality, this constitutes the vast majority of satellite data. Furthermore, re-analysis and Numerical Weather Prediction (NWP) data may also become part of the forthcoming 'virtual observatory'. Such data will generally arise from within the consortium (ECMWF and MO partners under WP4) and no restrictions are envisaged.

At the time of writing, the 'virtual observatory' is still in the process of being developed and tested with pilot users and its fully articulated data policy does not exist yet, as it is yet to be made public. However, once in place, this data policy will be made explicit and shall be in compliance with the H2020 Pilot on Open Research Data (s. next section) and this Data Management Plan.

Project parts dealing with enhancing existing primary data streams are:

- 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'
 - a. Creation of a collocation database between EO measures and reference-quality measurements.
 - b. Adoption of WIGOS and ESA-CCI standards for observational metadata in the virtual observatory. The discovery and observational metadata will be realised in a MongoDB environment.
 - 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, 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 participates in the H2020 Pilot on Open Research Data. Knowledge generated during the project will be shared openly. Any milestones, deliverables or technical documents produced which were deemed Public in the Grant Agreement will, following appropriate internal-to-project review procedures involving at least an expert and a management-based review, be published online and made discoverable. Peer-reviewed publications will by policy be 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 will be paid).

4. Dissemination and Exploitation of Results

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

- 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 will be entirely open and available to use for any application area. Significant efforts will be made to build an interface that is easy to use and which makes data discovery, visualization and analysis effortless. The 'virtual observatory' work package includes 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.

During the lifetime of the project the GAIA-CLIM website (www.gaia-clim.eu) will act as the public interface of the project. It shall provide an overview of the activities by work package (WP), as well as an open portal to disseminate information on project developments such as publications, peer-

reviewed journal articles and publicly-released deliverables. This includes the latest Gaps Assessment and Impact Document (GAID), providing the user with a living document with several versions published throughout the life of the project to enable iterative feedback¹. The web site allows all partners in the consortium to author and edit content, including WPs. Updates on a project blog from each WP on a quarterly basis are shortly to commence, that will enable appropriate communication and public propagation of results as they occur.

5. Preservation of value added data products

User engagement and outreach are key facets of the GAIA-CLIM objectives to build upon pre-existing capabilities of the partners and produce and disseminate tangible added value products at the end of the project. The 'virtual observatory' will constitute the primary means of dissemination of project results. Significant efforts have been taken 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, while optimizing the use of reference data. The objective is to develop an interface that uses software tools to deliver products in standard formats, such as NetCDF that are compliant with CF (Climate and Forecast) conventions. Such formats are self-describing and provide a definitive description of what the data in each variable represents, and 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 would support the relevance and sustainability of the facility 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. It is important to stress that at this juncture the Virtual Observatory is solely a demonstrator project and therefore the GAIA-CLIM data, documentation and facilities will be retained in frozen mode in the state they exist in at the end of the project. The tools and data GAIA-CLIM expects to retain and make available consist of:

- Metadata collected under Work Package 1 relating to networks and their measurement maturity. This shall include:
 - 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 from underlying networks (including new measurements able to be created under methods arising from WP2).

¹ Current version of GAID :http://www.gaia-clim.eu/system/files/workpkg_files/640276_Living%20Gap%20Assessment%20Document%20v2_1Mar.pdf

- Co-location uncertainty information arising from a variety of approaches including statistically based and dynamically based estimation (WP3 and 4)
- Tools to convert from geophysical to TOA radiance (WP4)
- Capabilities to visualize, subset and analyse the co-location database (WP5)

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.

6. Primary source datasets currently envisaged to be used within GAIA-CLIM

As outlined above in the GAIA-CLIM policy for data management, this DMP has been updated from the previous version D7.1. Version 2 of the DMP reflects the state of project activities at end of month 17. The review of the existing observing non-satellite capabilities under WP1 at the global scale has allowed us to identify 57 networks currently operating in the different continents, on different spatial domains and measuring different ECVs using one or more measurement techniques. Subsequently, the datasets initially envisaged to contribute primary data streams to be used in GAIA-CLIM ‘virtual observatory’ have been updated here based upon the analysis within Task 1.2.

The contributing networks foreseen at this time are as follows:

1. GRUAN
2. TCCON
3. ACTRIS-2
 - a. EARLINET
4. WCCOS
 - a. SHADOZ
5. NDACC

Further details on these networks, their governance and their data policies are given in Annex 2. As primary data collectors they shall assess data quality, integrity, originality and content prior to publishing. Whilst GAIA-CLIM activities under Work Package 2 may lead to changes in how the data are processed by the underlying networks, GAIA-CLIM is a user of these primary data products and not a provider.

As described previously, GAIA-CLIM will respect 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 would 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 shall only use the open delayed-mode data.

7. Scientific research data should be easily:

1. Discoverable

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier, DOI)?

Data and metadata will mainly be made available through the 'virtual observatory' facility. This online tool will make the data discoverable and also provide mapping, comparison and visualization functions. Data versioning, source locations, and any DOIs from the primary data sources will be retained. The possibility of creating data and software DOIs for the 'virtual observatory' at this stage still requires further investigation. For instance, DOI-registration works well for static data sets but remains mostly unexplored for regularly updated (changed) data. While it is possible to preserve DOIs for datasets to which new data is appended (as has been done for TCCON), a decision for or against the usage of DOIs depends on the final envisaged operational mode of the 'virtual observatory', which needs to be developed during the project and is dependent upon follow-on support opportunities. The 'virtual observatory' facility will be hosted by EUMETSAT and made discoverable.

2. Accessible

Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses?

As GAIA-CLIM participates in the Pilot on Open Research Data, knowledge generated during the project is shared openly. Any milestones, deliverables or technical documents produced are, following appropriate internal-to-project review procedures, published online and made discoverable. Commensurate with the Pilot on Open Research Data, all work explicitly produced by GAIA-CLIM will be open. However, GAIA-CLIM work in many cases will build upon pre-existing capabilities of the partners. In a restricted subset of these cases, Intellectual Property Right (IPR) restrictions relate to these background materials as articulated in the Consortium Agreement (cf. Annex 1). The policing of this aspect is the responsibility of the Technical Coordination Group.

The 'virtual observatory' facility will be entirely open and available to use for any application area. However, following the results of the user survey, the 'virtual observatory' will contain online applications. The underlying software will be openly shared to the extent useful, but GAIA-CLIM will not provide software usage support for users. This is beyond the scope and resources of the project.

Peer-reviewed publications will by policy be to journals that are either open access or allow the authors to pay for the articles to be made open access.

3. Assessable and intelligible

Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review?

Research is undertaken within GAIA-CLIM to improve observational traceability for a number of broadly used methods of observation and the quantification of the co-location mismatch uncertainties. The software resulting from GAIA-CLIM that shall constitute input to the 'virtual observatory' shall be shared openly and without restriction and shall be well documented.

The novel approach of GAIA-CLIM is 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.

4. Usable beyond the original purpose for which it was collected

Are the data and associated software produced and/or used in the project useable by third parties even long time after the collection of the data?

Data served will be available for any use regardless of whether it is within the currently envisaged end-uses or otherwise. Significant efforts will be made to build an interface that is easy to use and which makes data discovery, visualization and analysis effortless. All software that underlies the 'virtual observatory' and is created using GAIA-CLIM resources shall be made available if deemed useable as a stand-alone. The 'virtual observatory' work package includes a specific task dedicated to documenting the steps required to transition this facility from a research to an operations framework in support of Copernicus services.

Once the project is completed, the 'virtual observatory' and its underlying software will remain available, but in a "frozen state" with the aim of becoming further developed and integrated into the emerging Copernicus Climate Change Service and Copernicus Atmospheric Service. If continued in this way, Copernicus data and software distribution policies will be applied in the long-term.

5. Interoperable to specific quality standards

Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc.?

The project will only deal with both EO and sub-orbital (including in-situ and ground-based remote-sensing) data, which are available for academic use without restriction to simplify issues over dissemination of added value products derived by the project. These added value products will be made available immediately after they are produced and quality controlled without restriction. Since these added value products are constructed solely from sources with policies allowing re-use, we can serve them using a free and open policy for academic and research purposes. Data are to be accompanied by a conversion tool that will enable likely two different output formats that are in broad use within the recognised primary stakeholder communities, e.g. CF-compliant NetCDF. The data will be made available along with reading routines and visualisation tools through the 'virtual observatory' facility, which will allow data discovery and data usage for calibration and validation of level 1 and level 2/3 EO observations. The expectation is that new software written will use open-source software to the extent possible and useful and use of existing software shall have a preference for using programming languages that are open source or have open source compilers available such as e.g., C++, Fortran or python.

Summary

This Data Management Plan presents the data management policy that will be used by the GAIA-CLIM partners in their collection and use of data records during the lifespan of the project. This DMP version has evolved to reflect the plans of the project partners to date, and there will be a third and final version produced at the conclusion of the project. Since there will be no primary data produced under the GAIA-CLIM project the data management policy relates to meta-data 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 Earth Observation 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. TCCON

Data set reference and name

Total Carbon Column Observing Network (TCCON)

Data set description

TCCON is a network of ground-based Fourier Transform Spectrometers that takes direct solar absorption spectra at about 20 sites around the globe. From these, column averaged mole fractions of trace gases (CO₂, CH₄, N₂O, HF, CO, H₂O, and HDO) are inferred with a retrieval software. The HF and HDO retrievals are uncalibrated and hence preliminary. Each site contributes their dataset as an extending series for the current version of the retrieval software. Data are updated monthly and are publicly available no later than one year after the measurement; however, many sites choose to release their data much sooner.

Standards and metadata

TCCON products are calibrated against in-situ WMO values². In this way, the long-term stability is checked continuously. All data are delivered with an extensive metadata collection describing the measurement procedures and processing.

Data sharing

Data is openly accessible at the TCCON data server and hosted at the Carbon Dioxide Information Analysis Center (CDIAC)³ at Oak Ridge National Laboratory, USA. The data is made freely available to everyone. Acknowledgement and/or co-authorship in case of heavy use cases is expected. The data are stored in NetCDF format and each file has a DOI assigned to it (one per site and retrieval version). It is envisaged that each dataset will be described in a data publication paper.

Archiving and preservation (including storage and backup)

² Atmospheric measurements are only useful if calibrated to a common reference scale. For this reason the atmospheric in-situ measurements are anchored to the WMO Mole Fraction Scale. For the remote sensing measurements, this is not possible. Therefore the aircraft profiles, based on in-situ measurements are applied. Calibrations have been performed at several sites by high-flying aircrafts, covering the altitude region from about 200 m to 12 km. Internal test measurements by calibration gas cells ensure to detect any change in the alignment of the instruments.

³[TCCON data archive available at : http://tcccon.ornl.gov/](http://tcccon.ornl.gov/)

Archiving and preservation are ensured by the World Data Center (WDC) for Atmospheric Trace Gases⁴ standard implemented by CDIAC. WDCs have a responsibility to make available a detailed description of the data available to other WDCs and to the scientific community; all data held under WDC auspices must be accessible to such visitors. In the near future, the data will be mirrored at the PANGAEA⁵ data center, hosted by the Alfred-Wegener-Institute in Bremerhaven/Germany.

3. ACTRIS-2

Data set reference and name

ACTRIS-2 (Aerosols, Clouds, and Trace gases Research InfraStructure) Integrated Activity (IA)⁶

Data set description

ACTRIS-2 is a European Project aiming at integrating European ground-based stations, equipped with advanced atmospheric probing instrumentation for aerosols, clouds, and short-lived gas-phase species. ACTRIS-2 will have the essential role to support building of new knowledge as well as policy issues on climate change, air quality, and long-range transport of pollutants. The networks provide consistent datasets of observations, which are made using state-of-the-art measurement technology and data processing. Many of the stations from the different networks are co-located with or close to remote-sensing and in-situ instrumentation. The data is available through the ACTRIS data portal⁷.

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- Global Atmosphere Watch (GAW), The European Monitoring and Evaluation Programme (EMEP) and other EU-funded projects (such as ENVR^{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.

ACTRIS data sets are measured with the corresponding recommended methodology. Furthermore, the data are qualified as ACTRIS data sets only if they comply with the additional requirements based on standard operating procedures (SOP), complies with the procedures and recommendations

⁴ <http://cdiac.ornl.gov/>.Further information on WDC:
<http://cdiac.ornl.gov/wdca/wdcinfo.html#introduction>

⁵ <http://www.pangaea.de/>

⁶ <http://www.actris.eu>

⁷ Go to <http://actris.nilu.no> to access the latest map (updated automatically) with additional information about the ACTRIS sites and variables or download data from the ACTRIS Data Centre.

provided by the ACTRIS community and the measurement data has been submitted to the relevant data repository.⁸

Data sharing

The ACTRIS Data Centre web portal allows searching and analysing atmospheric composition data from a multitude of data archives through a single user interface. For some of the databases, the interface enables data download. The focus of the portal is validated data, but preliminary or rapid delivery data is also available. 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. The data repositories provide free and open access to all data resulting from the activities of the ACTRIS infrastructure network, complemented with data from other relevant networks which are hosted and provided with the permission of each organization or data provider contributing to the data archive. Use of this data implies an agreement to reciprocate.

Archiving and preservation (including storage and backup)

The ACTRIS database is maintained by the Norsk Institutt for Luftforskning (NILU). 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. Currently, a long term strategy for providing access to data and other related services is being explored through the establishment of ACTRIS as a European Research Infrastructure.

3a. EARLINET

Data set reference and name

EARLINET, the European Aerosol Research Lidar Network (EARLINET)

Data set description

EARLINET is the first European aerosol lidar network. It was established in year 2000 with the main goal to provide a comprehensive, quantitative, and statistically significant data base for the aerosol distribution on a continental scale. EARLINET is supported by ACTRIS-2 an H2020 funded project and is also part of WMO- GAW.

EARLINET aims at building a quantitative comprehensive statistical database of the horizontal, vertical, and temporal distribution of aerosols on a continental scale⁹. The goal is to provide aerosol data with unbiased sampling, for important selected processes, and air-mass history, together with

⁸ , 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/>

⁹ Specific data policy for EARLINET can be found at <http://www.earlinet.org/index.php?id=127>

comprehensive analyses of these data. The contributing stations are distributed over most of Europe, using advanced quantitative laser remote sensing to directly measure the vertical distribution of aerosols, supported by a suite of more conventional observations. Special care is taken to assure data quality, including inter-comparisons at instrument and evaluation levels. A major part of the measurements is performed according to a fixed schedule to provide an unbiased statistically significant data set.

Standards and metadata

EARLINET database represents a comprehensive collection of data which is harmonized to provide uniform datasets. Measurements are performed systematically at all the stations on a fixed measurement scheduling. Data are stored in Netcdf-Format and using a detailed file structure. The files can be stored as single profile file (containing aerosol optical profile(s) at a fixed time) and as a time series file (containing aerosol optical profile(s) at different times).

Data sharing

EARLINET data are freely available at EARLINET-ASOS database to individual users upon acceptance of the network data usage policy. Preliminary data are only available to EARLINET consortium members. Usage of data from the EARLINET database should follow an agreement to reciprocate and acknowledge the network

Each EARLINET site has a PI responsible for deployment, maintenance and data collection¹⁰. The PI has priority use of the data collected at the site. The PI is entitled to be informed of any other use of that site data

Archiving and preservation (including storage and backup)

The EARLINET data base is a core data base in ACTRIS and is maintained by NILU (see archiving of ACTRIS above).

4. WCCOS

Data set reference and name

World Calibration Centre for ozone sondes measurements (WCCOS)

Data set description

The WMO-GAW ozone observing system comprises more than 100 stations worldwide that measure total column ozone and ozone profiles in the troposphere and stratosphere. Since 1996, the WMO has attributed the role of the world calibration center for ozonesondes (WCCOS) to the Research Center at Forschungszentrum Jülich, Germany¹¹. The facility works to promote understanding and

¹⁰ All stations frequently updated are available at Quicklooks: <http://www.meteo.physik.uni-muenchen.de/~stlidar/quicklooks/European-quicklooks.html>

¹¹ Institute of Energy and Climate Research, Jülich Forschungszentrum: http://www.fz-juelich.de/iek/iek-8/EN/Expertise/Infrastructure/WCCOS/WCCOS_node.html

quality assurance of ozone sondes used in GAW/GLONET focusing on ozone sonde precision, accuracy and long term stability. The primary goals of the WCCOS are, to establish well documented SOPs, and to assess differences in instrument manufactures and in variations of SOPs in use.

Data from the WMO-GAW ozone observing system are collected and stored at the [WMO World Ozone and Ultraviolet Radiation Data Centre \(WOUDC\)](#) at Meteorological Service of Canada a branch of Environment Canada, Toronto. There are over 500 registered [stations](#) in the archive from over 150 [contributors](#). Metadata such as station lists with locations, are available from the WOUDC Website.

Standards and metadata

The primary objectives of the GAW quality assurance system are to ensure that the data in the World Data Centres are consistent, of known and adequate quality, supported by comprehensive metadata, and sufficiently complete to describe global atmospheric states with respect to spatial and temporal distribution. The GAW quality assurance (QA) system impacts all aspects of atmospheric chemistry observations, including training of station personnel; assessment of infrastructures, operations and the quality of observations at the sites; documentation of data submitted to the WDCs; improvement of the quality and documentation of legacy data at the WDCs.

Data sharing

All data is freely available to the scientific community through one of the six World Data Centres of WMO-GAW. Each centre is responsible for archiving the relevant measurement parameters. They collect, document and archive atmospheric measurements and the associated metadata from measurement stations world-wide. All WOUDC data are made available by using the [data search / download](#), or anonymous FTP access. WOUDC has registered Digital Object Identifiers (DOIs) for both Ozone and UV data. Use of the WOUDC data are governed by the [World Meteorological Organization \(WMO\) data policy](#) and [WMO Global Atmosphere Watch \(GAW\) data use policy](#).

The WMO facilitates the free and unrestricted exchange of data and information, products and services in real or near-real time on matters relating to safety and security of society, economic welfare and the protection of the environment. [Resolution 40 \(Cg-XII\)](#): WMO policy and practice for the exchange of meteorological and related data and products including guidelines on the relationships in commercial meteorological activities.

Use of data obtained from one of the WMO GAW World Data Centres (WDC) is subject to the following statement endorsed by the WMO Executive Council and the Commission for Atmospheric Sciences (EC/CAS) panel of experts working group on environmental pollution and atmospheric chemistry [WMO, 2001a].: For scientific, educational and policy related use, access to these [GAW] data is unlimited and provided without charge. By their use you accept that an offer of co-authorship will be made through personal contact with the data providers or owners whenever substantial use is made of their data. In all cases, an acknowledgment must be made to the data providers or owners and to the data centre when these data are used within a publication.

Archiving and preservation (including storage and backup)

The Open Geospatial Consortium web services provide enhanced access to the archive, which provides standards-based-on-demand access to the archive.

4a. SHADOZ

Data set reference and name

Southern Hemisphere ADditional OZonesondes (SHADOZ)

Data set description

SHADOZ is designed to improve data discrepancy of ozonesonde stations operating in the tropics, subtropics and southern hemisphere in general , by coordinating launches, supplying additional sondes in some cases, and by providing a central archive location.

Standards and metadata

Since 1998, ozonesonde data have been collected and made available through this website, currently displaying data for 13 current sites and 4 former sites. In addition to the data and graphical display of the sounding, the air parcel history of each sounding is depicted at the SHADOZ Web site. These are based on 5-day back trajectories run with the National Aeronautics and Space Administration (NASA) Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model.

Data sharing

Access to SHADOZ data is free to the public. Whenever substantial use is made of their data, user shall accept that an offer of co-authorship will be made through personal contact. In all cases, an acknowledgement must be made to the data co-investigators and SHADOZ when these data are used in publication.

Archiving and preservation (including storage and backup)

SHADOZ data is archived by the NASA Goddard Space Flight Centre. The individual station PIs are responsible for retaining information on the original data processing, launch procedures, and scheduling. The SHADOZ homepage (<http://croc.gsfc.nasa.gov/shadoz/>) includes the contact information for the PIs.

5. 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 stations¹² 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 measured¹³.

Standards and metadata

NDACC is organized in several working groups¹⁴, 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 Networks¹⁵. 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.¹⁶ However, many NDACC investigators have agreed to make their data publicly available immediately upon archiving. The public record is available through anonymous ftp¹⁷. 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

¹² 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/>.

¹³ The observational capabilities can be displayed by a chart on: http://www.ndsc.ncep.noaa.gov/obs_chart/obs-capabil-2015-10-14.pdf

¹⁴ Split into instrument and theme working groups: <http://www.ndsc.ncep.noaa.gov/organize/>

¹⁵ The NDACC protocols are accessible under <http://www.ndsc.ncep.noaa.gov/organize/protocols/>

¹⁶ 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.

¹⁷ Regular NDACC data: <ftp.cpc.ncep.noaa.gov/ndacc/station>

investigators PI(s). Rapid delivery data, which will likely be revised before entry in the full database, is also available for some instruments¹⁸.

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 table¹⁹ and clicking on the station name will take the user to the associated public data site.

¹⁸ 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.

¹⁹ http://www.ndsc.ncep.noaa.gov/data/data_tbl/