Deliverable 4.2: ACTRIS Data Management Plan

Authors

Core team

Cathrine Lund Myhre[1], Richard Rud[1], Markus Fiebig[1], Cathy Boonne[2], Guillaume Brissebrat[2], Nicolas Pascal[2], Bénédicte Picquet-Varrault[2], Claudio Dema[3], Lucia Mona[3], Giuseppe D'Amico[3],, Francesco Amato[3], Sergio Ciamprone[3], Ewan O'Connor[4] (FMI), Simo Tukiainen,[4] Anniina Korpinen[4], Michael Schulz[5], Jonas Gliss[5], Carlos Pérez García-Pando[6], Sara Basart[6]

1. NILU - Norwegian institute for air research
2. CNRS – AERIS Atmosphere and Service Data Pole
3. CNR – IMAA Institute of Methodologies for Environmental Analysis
4. FMI - Finnish Meteorological Institute
5. MetNo - The Norwegian Meteorological Institute
6. BSC - Barcelona Supercomputing Center

Contributors at unit level:

Rodrigo Guzman (CNR/AERIS), Massimo Chiari (CNR/AERIS), Massimo Chiari (INFN), Anja Claude (DWD), Olivier Favez (INERIS), Romy Fösig (KIT), Silvia Henning (TROPOS), Robert Holla (DWD), Robert Wegener (FZ Jülich), Tuija Jokinen (Univ. Helsinki), Griša Močnik, (IJS), Anke Mutzel (TROPOS), Jakub Ondráček (CAS ICPF), Dominik van Pinxteren (TROPOS), Jean-Philippe Putaud (JRC), Stefan Reimann (EMPA), Stéphane Sauvage (Univ. Lille, IMT), Rainer Steinbrecher (KIT), Alfred Wiedensohler (TROPOS), Doina Nicolae (INOE), Aldo Amodeo (CNR)

Work package no	WP4
Deliverable no.	D4.2
Lead beneficiary	NILU
Deliverable type	X R (Document, report) ☐ DEC (Websites, patent fillings, videos, etc.) ☐ OTHER: please specify .....
Dissemination level	X PU (public) ☐ CO (confidential, only for members of the Consortium, incl. Commission)
Estimated delivery date	30/11/2019
Actual delivery date	29/11/2019
Version	27 November 2019
Reviewed by	WP4 leader
Accepted by	Sanna Sorvari Sundet
Comments	The ACTRIS Data Management Plan (DMP) is a living document, and will be updated regularly in accordance with the implementation and development within ACTRIS. The latest version will always be available here: https://github.com/actris/data-management-plan/blob/master/DMP/ACTRIS-DMP.md and the one that is valid and should be used.

Preface

Data Management Plan – General Definition

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).

"The Open Research Data Pilot In Horizon 2020"

ACTRIS Data Management Plan (DMP) is a living document, and will be updated in accordance with the implementation and development within ACTRIS. The latest version will always be available here: https://github.com/actris/data-management-plan/blob/master/DMP/ACTRIS-DMP.md and the one that is valid and should be used. This document is the first full version of the ACTRIS DMP, the Deliverable 4.2 within ACITRS PPP.

The ACTRIS DMP is based on a template from the Digital Curation Centre that ensures that the DMP comply with the FAIR principles, and structured in the recommended way by EU and answering to the questions raised in the "Open Research Data Pilot In Horizon 2020". The ACTRIS DMP is a an online machine-actionable DMP that is very easy to extend and update. The ACTRIS DMP serves as the work plan for the data centre, and includes the links internally between the data centre units, and to national facilities (NF) and topical centres (TC). Accordingly, there has been a extensive processes, involving a large part of the ACTRIS community.

The work with the revised ACTRIS DMP were intensified at a meeting at NILU January 2019, and the first outline and draft of the DMP were presented and discussed at ACTRIS PPP meeting the first week of May in Rome 2019. Since then, the link to the DMP has been available to all within ACTRIS PPP to facilitate contributions and feedback. Interested have been encouraged to follow the processes through the link, and interact with the ACTRIS DC team. In total at least 38 persons has contributed directly from all DC units and TCs and been to the present version of the DMP, to take the complexity of ACTRIS data production into account. During 2019, there have been numerous smaller meetings and virtual meetings to work on the DMP. Recently, the DMP has been discussed at 3 various technical community meeting in October and November 2019 (Kuopio (simulation chamber data), Prague (In-Situ data), Rome (aerosol and cloud remote sensing data), at the GA in Helsinki, and at separate WS for the data centre team in November. The DMP will continue to be updated with further details and in line with the decisions and implementation during ACTRIS implementation phase.

ACTRIS – Aerosol, Clouds and Trace Gases Research InfraStructure

Data Management Plan

Version 27 November 2019

Contents

1. Introduction to The ACTRIS Data Centre and ACTRIS Data Management Plan....7

1.1 The mission, overall goal and structure of the ACTRIS Data Centre....9

1.2 The overall goal and structure of ACTRIS Data Management Plan....12

2. ACTRIS data and ACTRIS data levels....13

3. Data summary of the ACTRIS data centre....14

3.1 ACTRIS In situ data centre unit (In-Situ)....15

3.2 ACTRIS Aerosol remote sensing data centre unit (ARES)....17

3.3 ACTRIS Cloud remote sensing data centre unit (CLU)....19

3.4 ACTRIS trace gases remote sensing data centre unit (GRES)....20

3.5 ACTRIS Atmospheric simulation chamber data centre unit (ASC)....23

3.6 ACTRIS data and services (ACCESS)....25

4. Data Management at the ACTRIS data centre....27

4.1 ACTRIS access and service policy....27

4.2 Introduction and overview of ACTRIS Data Management architecture....29

4.3 Findable: Making data findable, including provisions for metadata [FAIR data]....48

4.4 Accessible: Making data openly accessible [FAIR data]....51

4.5 Interoperable: Making data interoperable [FAIR data]....53

4.6 Reuseable: Increase data re-use [FAIR data]....54

5. Allocation of resources....56

6. Data security....57

6.1 Archiving and preservation of In-Situ data....57

6.2 Archiving and preservation of ARES data....58

6.3 Archiving and preservation of CLU data....58

6.4 Archiving and preservation of GRES data....59

6.5 Archiving and preservation of ASC data....59

6.6 Archiving and preservation of ACCESS data....59

7. Ethical aspects....60

8. Appendix....61

Appendix 1: List of ACTRIS variables from observational platforms and associated recommended methodology....61

Appendix 2: List of ACTRIS level 3 data products....61

Appendix 3: ACTRIS In situ data centre unit (In-Situ) data life cycle....61

Appendix 4: ACTRIS Aerosol remote sensing data centre unit (ARES) data life cycle and workflow diagram....66

Appendix 5: ACTRIS Cloud remote sensing data centre unit (CLU) data life cycle and workflow diagram 68

Appendix 6: ACTRIS trace gases remote sensing data centre unit (GRES) data life cycle and workflow diagram 69

Appendix 7: ACTRIS Atmospheric simulation chamber data centre unit (ASC) data life cycle and workflow diagram 72

Appendix 8: Data lifecycle and workflow for ACCESS Data Centre Unit 73

Appendix 9: Format and external data sources for level 3 variables 74

Appendix 10: ReOBS workflow diagram 78

Appendix 11: Satellite data subsets workflow diagram 79

1. Introduction to The ACTRIS Data Centre and ACTRIS Data Management Plan

The Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS) focuses on producing high-quality data for the understanding of short-lived atmospheric constituents and their interactions. These constituents have a residence time in the atmosphere from hours to weeks. The short lifetimes make their concentrations highly variable in time and space and involve processes occurring on very short timescales. These considerations separate the short-lived atmospheric constituents from long-lived greenhouse gases, and calls for a four dimensional distributed observatory. The Research Infrastructure (RI) ACTRIS is the pan-European RI that consolidates activities amongst European partners for observations of aerosols, clouds, and trace gases and for understanding of the related atmospheric processes, as well as to provide RI services to wide user groups (See the Stakeholder Handbook for more information).

ACTRIS data are data from observational or exploratory National Facilities complying with the procedures established within ACTRIS.

ACTRIS observational platforms are fixed ground-based stations that produce long-term data based on a regular measurement schedule and common operation standards. These platforms perform measurements of aerosol, clouds, and reactive trace gases from the Earth surface throughout the troposphere up to the stratosphere by applying state-of-the-art remote-sensing and in situ measurement techniques under consideration of harmonized, standardized, and quality controlled instrumentation, operation procedures and data retrieval schemes. The sites are strategically located in diverse climatic regimes both within and outside Europe, and many of them contribute to one or several European and international networks, such as EMEP, NDACC, or GAW, and are possibly partly shared with other environmental infrastructures, such as ICOS, SIOS, ANAEE or eLTER.

Laboratory platforms and mobile platforms that perform dedicated experiments and contribute data on atmospheric constituents, processes, events or regions by following common ACTRIS standards are considered ACTRIS exploratory platform. In addition to these,atmospheric simulation chambers are ACTRIS exploratory platforms too. These chambers are among the most advanced tools for studying and quantifying atmospheric processes and are used to provide many of the parameters incorporated in air quality and climate models. Atmospheric simulation chamber data contribute to better predict the behavior of the atmosphere over all time scales through a detailed understanding of the physical and chemical processes, which affect air quality and climate change. Atmospheric simulation chambers are among the most advanced tools for studying and quantifying atmospheric processes and are used to provide many of the parameters incorporated in air quality and climate models.

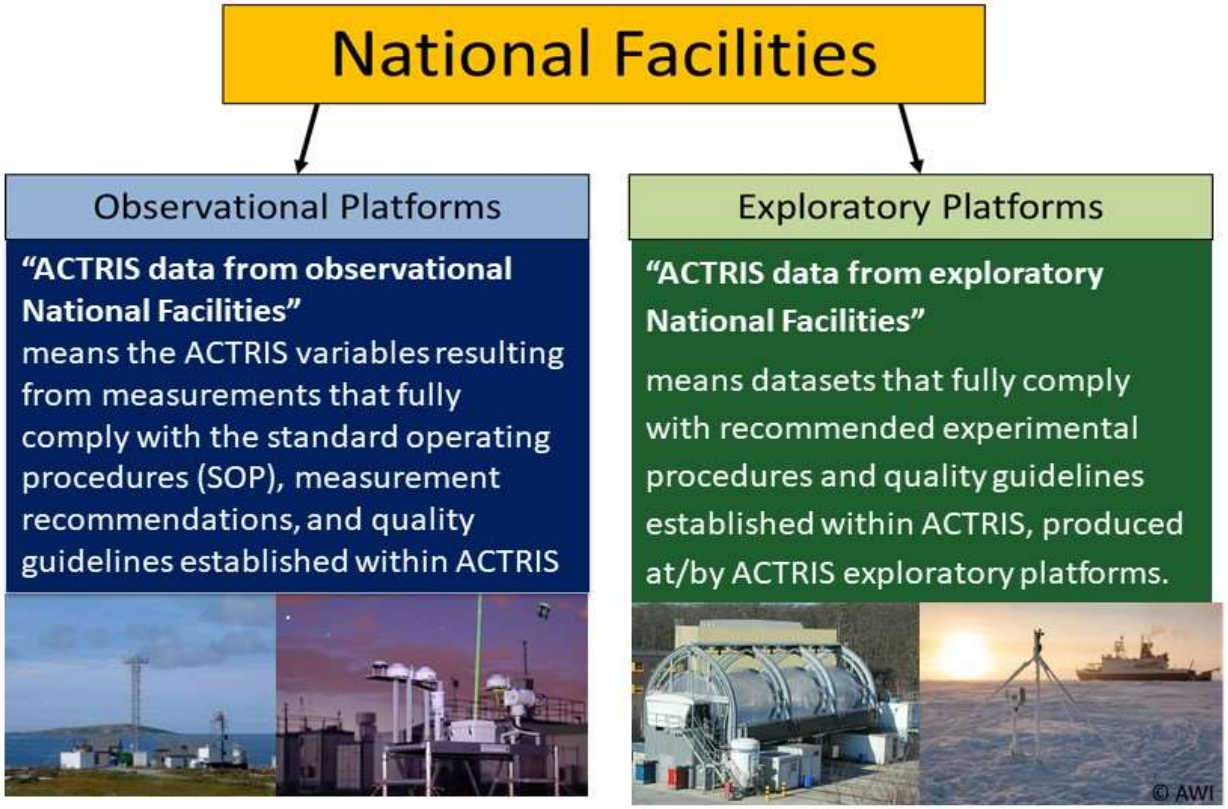


Figure 1: Overview of the types of National Facilities providing data to ACTRIS Data Centre

ACTRIS is a unique RI improving both the quality of and access to atmospheric observations, developing new methods and protocols, and harmonizing existing observations of the atmospheric variables listed in Appendix 1. Appendix 1 includes an updated list of all ACTRIS variables associated to recommended measurement methodology.

1.1 The mission, overall goal and structure of the ACTRIS Data Centre

The mission of the ACTRIS Data Centre (DC) is to compile, archive and provide access to well documented and traceable ACTRIS measurement data and data products, including digital tools for data quality control, analysis, visualisation, and research. As a tool for science, the highest priorities for the ACTRIS DC are to maintain and increase the availability of ACTRIS data and data products relevant to climate and air quality research for all interested users.

The overall goal of the ACTRIS Data Centre (DC) is to provide scientists and other user groups with free and open access to all ACTRIS data, complemented with access to innovative and mature data products, together with tools for quality assurance (QA), data analysis and research. ACTRIS data and products should be **f**indable, **a**ccessible, **i**nteroperable and **r**eusable (FAIR), and the data centre work towards fulfilling the FAIR principles. The numerous measurement methodologies applied in ACTRIS result in a considerable diversity of the data collected. In accordance with these requirements, the ACTRIS DC will be organized in 6 Units, with clear links and procedures for interaction between the data centre Units, National Facilities (NFs) and topical centres (TCs). The ACTRIS DC will be coordinated by the ACCESS unit leader and all data is linked through the ACTRIS data portal serving as the access point to all data and related information. The units and short names are:

- ACTRIS data and services access unit (ACCESS)
- ACTRIS In situ data centre unit (In-Situ)
- ACTRIS Aerosol remote sensing data centre unit (ARES)
- ACTRIS Cloud remote sensing data centre unit (CLU)
- ACTRIS trace gases remote sensing data centre unit (GRES)
- ACTRIS Atmospheric simulation chamber data centre unit (ASC)

During the ACTRIS implementation phase (expected 2020-2024), the Central Facilities will be constructed and their services tested. The ACTRIS Central Facilities host selection was a part of ACTRIS PPP, and the following consortium is selected to host the ACTRIS Data Centre, and the various units with services to data producers and data users.

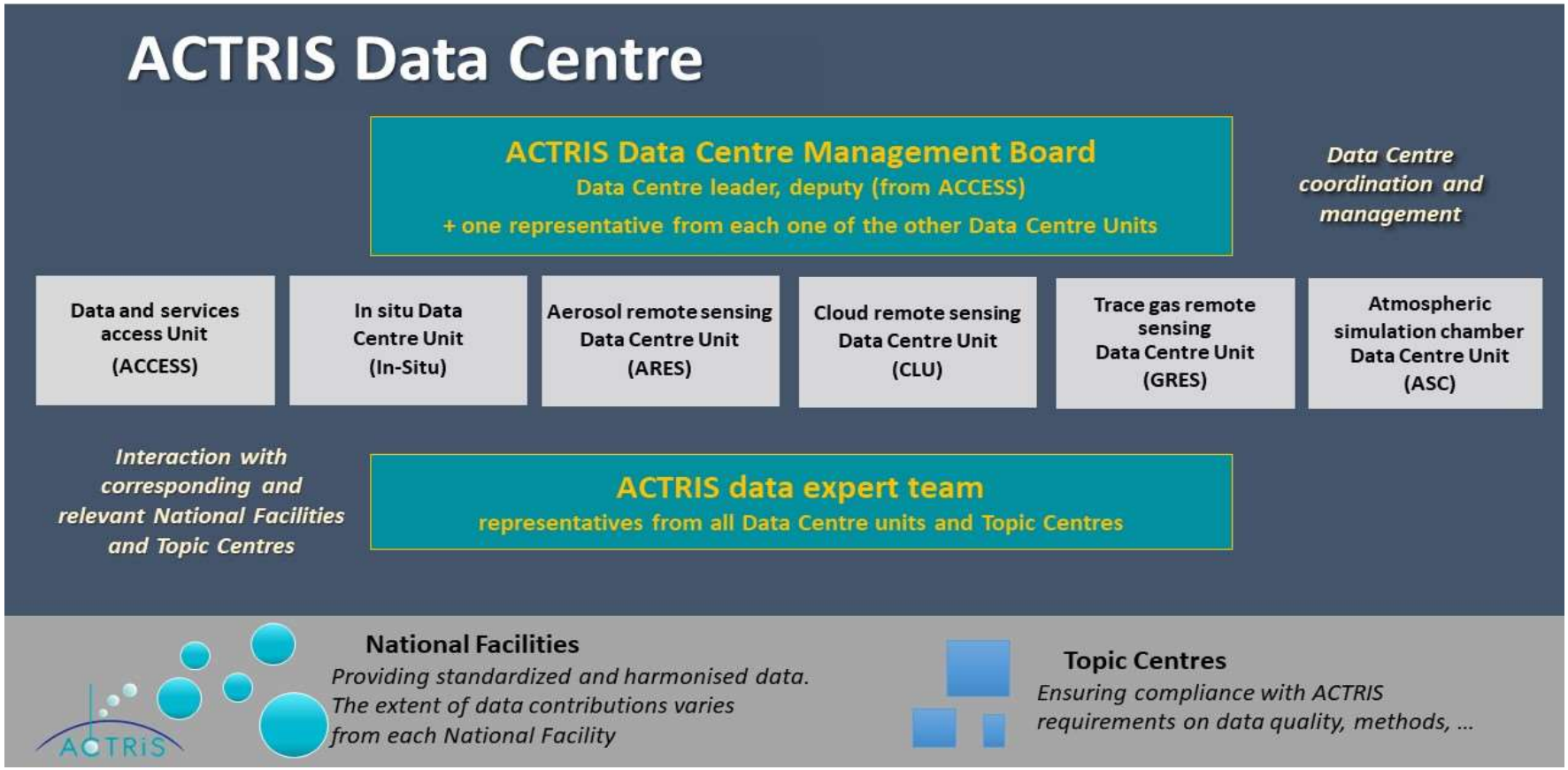


Figure 2: Architecture of the ACTRIS Data Centre

Name of Central Facility and associated Unit	Hosting institution and contribution	Main activities
ACTRIS data and services access unit (ACCESS)	NILU (lead), CNRS, MetNo, BSC	ACTRIS web interface for data, services and tools, called "The ACTRIS Data Centre". Main activities are discovery and access to ACTRIS data and data products, digital tools provided by the topical centres and the data centre units, documentation, access to software and tools for data production. Offer visualisation of ACTRIS data products. Data production of selected Level 3 data and synergy data products. The data centre will offer bridge to external data bases and sources.
ACTRIS In-Situ data centre unit (In-Situ)	NILU	Data curation service for in situ data: all aerosol, cloud and trace gas in situ data. This comprises inclusion of data in the data base EBAS, archiving and documentation. Support for centralized data processing, harmonization, traceability, quality control and data product generation. Training and online tools for QA, QC. The activity enables RRT and NRT delivery.
ACTRIS Aerosol remote sensing data centre unit (ARES)	CNR (lead), CNRS	Aerosol remote sensing data processing and curation. This includes centralized processing, traceability, harmonization and data versioning, quality control, data archiving in EARLINET DB, data provision and documentation. The activity enables RRT and NRT delivery. Tutorial activities. Production of level 3 data for climatological analysis and new products.
ACTRIS Cloud remote sensing data centre unit (CLU)	FMI	Data curation service for cloud remote sensing data. Support for centralized cloud remote sensing data processing, traceability, harmonization, automated quality control and product generation, and data archiving. Enables RRT and NRT delivery. Production of level 3 data for NWP model evaluation.
ACTRIS Atmospheric simulation chamber data centre unit (ASC)	CNRS	Data curation service for atmospheric simulation chamber data. This includes standardized process for data submission, quality control, inclusion of data in the AERIS data base, search metadata creation and provision and archiving.
ACTRIS trace gases remote sensing data centre unit (GRES)	CNRS	Data curation service for reactive trace gases remote sensing data. This comprises standardized process for data submission, quality control, inclusion of data in the AERIS data base, metadata creation and provision and archiving. Production of level 3 data for

Name of Central Facility and associated Unit	Hosting institution and contribution	Main activities
		climatological analysis, and added values products (quicklooks, links to EVDC - ESA Atmospheric Validation Data Centre).

Table 1: Short description of the ACTRIS DC units and the research performing organizations leading and contributing to the units.

1.2 The overall goal and structure of ACTRIS Data Management Plan

The ACTRIS Data Management Plan (DMP) should document the key elements of the ACTRIS data management life cycle, and the plans for the data collected, processed and/or generated. The goal of the DMP is to describe the present situation and the operational ACTRIS data center. Furthermore the DMP should also describe the technical solutions agreed, that are currently under implementation, and outline the strategy and development needed towards making ACTRIS data FAIR at ACTRIS Data Centre Level.

The ACTRIS DMP is a "living" online document which is set up to be machine-actionable as a part of the FAIR data ecosystem. The DMP should be a hub of information on ACTRIS FAIR digital objects. The goal is to make the ACTRIS DMP accessible for all stakeholders (repository operators, funders, researchers, publishers, infrastructure providers etc.) by making it available and accessible for both humans and machines. We currently use GitHub as the platform for collaboration on the DMP, this enables all actors working with or within ACTRIS to directly contribute and suggest changes to the document. Furthermore, the ACTRIS Data Management Plan should follow the glossary of terminology and definitions used in ACTRIS.

2. ACTRIS data and ACTRIS data levels

ACTRIS data are data from observational or exploratory National Facilities complying with the procedures established within ACTRIS. ACTRIS data comprises ACTRIS variables resulting from measurements at National Facilities that fully comply with the standard operating procedures (SOP), measurement recommendations, and quality guidelines established within ACTRIS. The ACTRIS atmospheric variables are listed in Appendix I, associated to the corresponding recommended methodology.

There are 4 levels of ACTRIS data: * **ACTRIS level 0 data:** Raw sensor output, either mV or physical units. Native resolution, metadata necessary for next level. * **ACTRIS level 1 data:** Calibrated and quality assured data with minimum level of quality control. * **ACTRIS level 2 data:** Approved and fully quality controlled ACTRIS data product or geophysical variable. * **ACTRIS level 3 data:** Elaborated ACTRIS data products derived by post-processing of ACTRIS Level 0 -1 -2 data, and data from other sources. The data can be gridded or not.

Additionally to these data products which are completely under the control of ACTRIS with established procedures and standards, the ACTRIS DC will also produce additional data products of interest of the the scientific and user communities. These are **ACTRIS synthesis product:** data products from e.g. research activities, not under direct ACTRIS responsibility, but for which ACTRIS offers repository and access.

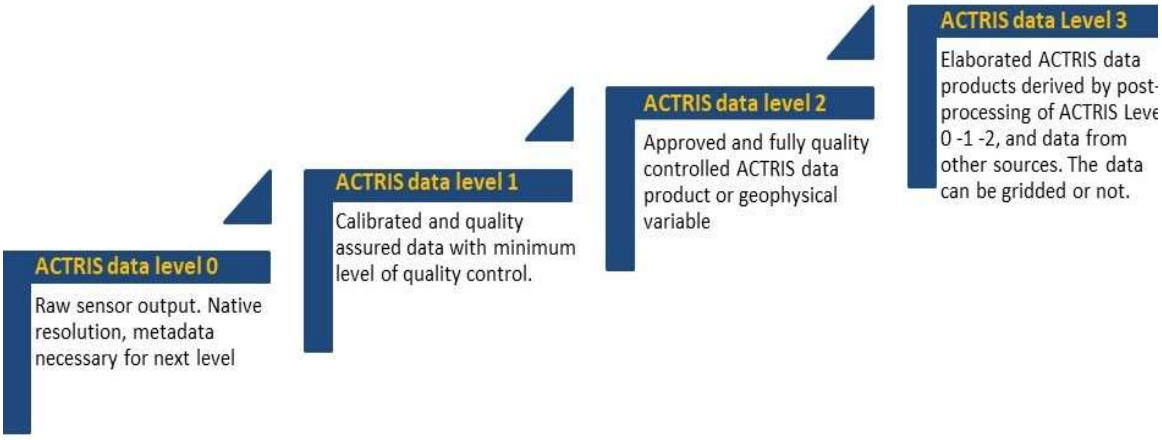


Figure 3: ACTRIS data levels

The list of ACTRIS variables are expected to increase during the progress of ACTRIS. Level 3 data products are expected to increase in quantity and number of variables because of the expected increase in ACTRIS data synergistic usage with other datasets. Additionally the expected technological and methodological developments fostered by ACTRIS itself will increase the ACTRIS observational capabilities and therefore the number and quality of observable atmospheric related variables (Level 1 and Level 2 products).

3. Data summary of the ACTRIS data centre

The purpose of the data collection/generation

Primary goal of ACTRIS is to produce high quality integrated datasets in the area of atmospheric sciences and provide services, including access to instrumented platforms, tailored for scientific and technological usage. The purpose of the data collection and generation of data products in ACTRIS is to provide open access to aerosol, cloud and trace gas in situ and remote sensing measurements of high quality, benefiting a large community of scientists involved in atmospheric science and related areas as well as policy makers, the private sector, educators and the general public.

See the Stakeholder Handbook for more information.

The relation to the objectives of the project as stated in Stakeholder Handbook

The primary goal of ACTRIS is to produce high quality integrated datasets in the area of atmospheric sciences and provide services, including access to instrumented platforms, tailored for scientific and technological usage. The main objectives of ACTRIS are:

- to provide information on the 4D-compositon and variability and of the physical, optical and chemical properties of short-lived atmospheric constituents, from the surface throughout the troposphere to the stratosphere, with the required level of precision, coherence and integration;
- to provide information and understanding on the atmospheric processes driving the formation, transformation and removal of short-lived atmospheric constituents;
- to provide efficient open access to ACTRIS data and services and the means to effectively use the ACTRIS products;
- to ensure and raise the quality of data and use of up-to-date technology used in the RI and the quality of services offered to the community of users, involving partners from the private sector; and
- to promote training of operators and users and enhance linkage between research, education and innovation in the field of atmospheric science.

Management of ACTRIS data relates to measuring atmospheric composition and the ability to predict the future behavior of the atmosphere over all time scales. High quality observational data harmonized across the countries and continents facilitates this, and needs to be supported by:

- Documentation of archiving procedures and access to level 0 -> level 3 data produced by the National Facilities (NFs), Topical Centres (TCs), and Central Facilities (CFs)
- Documented and traceable processing chain of level 0 data
- Documented, traceable processing and long-term archiving and preservation of all ACTRIS level 1 to level 3 data and data products
- Access to ACTRIS data, data products, and digital tools through a single point of entry
- Documentation of data, data flow, citation service, and data attribution, including version control, data traceability, and interoperability,
- Data curation and support for campaigns and dedicated research projects and initiatives, external or internal to ACTRIS.

Main users of ACTRIS data and software

ACTRIS will produce data and data products essential to a wide range of communities as described in detail in the Stakeholder Handbook, section “Users” including:

- Atmospheric science research communities world-wide
- The climate and air-quality, observational/ experimental/ modelling/ satellite communities, national and international research programmes and organisations;
- Environmental science research communities and communities from other neighboring fields: hydro-marine, bio-ecosystem, geosciences, space physics, energy, health, and food domain, to study interactions and processes in across different disciplines;
- Instrument manufacturers and sensor industries for development, testing, prototyping and demonstration;
- Operational services, National weather services, climate services for model validation, weather and climate analysis and forecasting;
- Space agencies for validation and the development of new satellite missions;
- National and regional air quality monitoring networks and environmental protection agencies for air quality assessments and validation of air pollution models;
- Policy makers and local/ regional/ national authorities for climate, air-quality, health and atmoshperic hazards related information for decision making and policy development.
- Copernicus atmospheric monitoring service (ECMWF)
- Science community working on air quality, climate change and stratospheric ozone depletion issues

3.1 ACTRIS In situ data centre unit (In-Situ)

The In-Situ data centre unit provides data curation service for aerosol, cloud and trace gas in situ data, as well as archiving of tihis data using the EBAS database. This comprises tools for harmonized data submission and meta data templates, inclusion of data and meta data in the data base, traceability, harmonization and data versioning, quality control, archiving, documentation and data provision. Training and online tools for QA, QC are offered. The activity enables RRT and NRT data compilation and delivery and provides tutorial activities. Furthermore, support for centralized data processing, harmonization, and data product generation, both level 2 and level 3 is offered and further implemented during the implementation phase.

The types and formats of data generated/collected

The ACTRIS In-situ data centre unit is supported by the EBAS database infrastructure. In situ data submitted to ACTRIS need to be formatted in the EBAS NASA-Ames format (ASCII file) by the data originator. There are exsisting instructions and templates for each instrument/group of instruments. The EBAS NASA-Ames format is based on the ASCII text NASA-Ames 1001 format, but contains additional metadata specifications ensuring proper documentation from the EBAS-Submit documentation website as well as tools for file-generation (*beta*) and file-submission.

ACTRIS in situ data is also available in netCDF 4 format through the EBAS Thredds Server, following the CF 1.7 convention and the Attribute Convention for Data Discovery 1-3 (ACDD).

Re-use of existing data

The ACTRIS data user interface will include access to aerosol and trace gas in situ legacy data resulting from ACTRIS pre-projects (for In-Situ CREATE, EUSAAR, ACTRIS-FP7. These will also be included as a part of the ACTRIS In Situ data centre unit. Legacy data resulting from ACTRIS pre-projects will be available in the same format as current products.

The origin of the data

The origin of the data is derived from instrument raw data, either through online or offline observations.

The expected size of the data

Type	Number of annual datasests (end 2019)	Number of annual datasets (min by 2025)	Number of annual datasets (max by 2025)
ACTRIS in situ aerosol data	60	50	120
ACTRIS in situ cloud data	0	35	105
ACTRIS in situ trace gas data	27	30	60

Table 2: Number of annual datasets

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
ACTRIS in situ aerosol data	18 000 MB	15 000 MB	50 000 MB
ACTRIS in situ cloud data	0 MB	1 GB	3 GB
ACTRIS in situ trace gas data	300 MB	200 MB	400 MB

Table 3: Data volume

Data utility

According to IPCC AR5, aerosol particles in the atmosphere are the most prominent source of uncertainty of climate predictions. Depending on their properties, they can have a warming as well as cooling effect on climate by scattering and absorbing solar radiation, and they can extend brightness and lifetime of clouds. Volatile organic compunds (VOCs) are one source of precursors for aerosol particles by forming condensable vapours during oxidation and decay in the atmosphere. By interaction with nitrogen oxides (NO_x), themselves a pollutant emitted by combustion, decaying VOCs form ozone (O_3), another pollutant. All of these, particulate matter, VOCs, NO_x, and ozone, have adverse effects on human health. Data on concentrations and properties of these compounds stored in the ACTRIS DC In Situ unit address all of the named effects: * **In Situ** data feed into the IPCC assessment reports in order to quantify and reduce the uncertainty of climate change. * **In Situ** data are the basis of national and international assessment reports of air quality. * **In Situ** data feed into and validate operational air quality prediction products, e.g. by Copernicus.

Outline of data life cycle (workflow and workflow diagram)

Detail on the data life cycle and workflow (workflow diagrams for data production) for in situ observations can be found in *Appendix 3: ACTRIS in situ aerosol, cloud and trace gas data lifecycle and workflow (draft)*.

3.2 ACTRIS Aerosol remote sensing data centre unit (ARES)

The ARES data centre unit provides a data curation and data processing service for aerosol remote sensing data coming frm lidar and photometer obsrevations. This includes centralized data processing, data storage, recording of metadata in a dedicated RDBMS, traceability, harmonization and data versioning, quality control, documentation and data provision. The unit allows for RRT and NRT data provisioning and offers support and training activities. Furthermore, level 3 data production for climatological analysis and the delivery of new data products will be further implemented and offered during the implementation phase.

The main goal is providing access of high quality and document datasets of the aerosol optical properties vertical distribution in the whole troposphere and upper stratosphere with short time resolution. This long term dataset collected at continental scale allows: - investigation of the relationship between near-surface processes (as pollution or air quality issues) and atmospheric aerosol contents; - addressing the challenging issue of direct and indirect effects of aerosol in the climate change.

The types and formats of data generated/collected

The ACTRIS ARES data centre unit is built on the heritage of the EARLINET Data base infrastructure and integartes the photometer aerosol data processing. Aerosol remote sensing data submitted to ACTRIS need to be compliant to a specific format established by the ARES unit centralized processing suite. All further data levels are produced by the ARES processing suite. ARES provides data compliance with NetCDF4, following Climate Forecast (CF) 1.7 conventions.

- ARES Level 1 data products consist of high and low resolution total attenuated backscatter and volume depolarization ratio time series provided in NRT or RRT. provided by the photometer observation are also available. Additionally ARES provides columnar information and synergistic lidar/photometer products as vertical profiles of aerosol microphysical properties as Level 1 data.
- ARES Level 2 data products contain fully quality assured aerosol extinction, backscatter, lidar ratio, Angstrom exponent and depolarization ratio vertical profiles and full quality controlled columnar information and aerosol microphysical properties profiles.
- ARES Level 3 data products are retrieved from the level 2 data and provide statistical analysis (including seasonality and annuality) of the most important aerosol optical parameters.

Re-use of existing data

The ACTRIS data user interface will include access to aerosol remote sensing legacy data resulting from ACTRIS pre-projects (for ARES EARLINET, EARLINET-ASOS). These will also be included as a part of the ACTRIS ARES data centre unit. Legacy data resulting from ACTRIS pre-projects will be available in the same format as current products.

The origin of the data

The origin of the data is derived from instrument raw data provided by the data originators in a common format (NetCDF).

The expected size of the data

Type	Number of annual datasets (end 2019)	Number of annual datasets (min by 2025)	Number of annual datasets (max by 2025)
ACTRIS aerosol remote sensing data	28	30	52
GRASP/GARRLiC	6	30	52

Table 4: Number of annual datasets

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
ACTRIS aerosol remote sensing data	4 GB	2.5 TB	20 TB
GRASP/GARRLiC	3.2 GB	40 GB	50 GB

Table 5: Data volume

Data utility

Atmospheric aerosols are considered one of the major uncertainties in climate forcing, and a detailed aerosol characterization is needed in order to understand their role in the atmospheric processes as well as human health and environment. The most significant source of uncertainty is the large variability in space and time. Due to their short lifetime and strong interactions, their global concentrations and properties are poorly known. For these reasons, information on the large-scale three-dimensional aerosol distribution in the atmosphere should be continuously monitored. It is undoubted that information on the vertical distribution is particularly important and that lidar remote sensing is the most appropriate tool for providing this information. ARES data products are particularly useful for addressing important issues like validation and improvement of models that predict the future state of the atmosphere and its dependence on different scenarios describing economic development, including those actions taken to preserve the quality of the environment.

- ARES Level 1 data are particularly interesting for several applications such as model assimilation and monitoring of special/critical events (volcanic eruptions, dust intrusions, ...).
- ARES Level 2 data allow for an optimal and complete optical and microphysical characterization of atmospheric aerosol. This is the fundamental starting point for any study regarding the assessment of aerosol in many atmospheric processes (climatology, climate change, Earth radiative budget, aerosol layer characterization, long range transported aerosol processes).
- ARES Level 3 data are climatological products providing statistical analysis of aerosol optical parameters. These products are useful for the characterization of different sites all over Europe as well as to underline seasonalities, annualities and trends.

Outline of data life cycle (workflow and workflow diagram)

Details of the data life cycle and workflow (workflow diagrams for data production) for aerosol remote sensing observations can be found in Appendix 4: ACTRIS aerosol remote sensing data lifecycle and workflow (draft).

3.3 ACTRIS Cloud remote sensing data centre unit (CLU)

The CLU data centre unit provides data curation and data processing service of cloud remote sensing data. This includes centralized processing, traceability, harmonization and data versioning, quality control, data provision and archiving, and documentation. The activity enables RRT and NRT data compilation and delivery, and participation in training. Furthermore, data product generation of level 3 data for forecast and climate model evaluation, climatological analysis and new products is offered and further implemented during the implementation phase.

The types and formats of data generated/collected

The ACTRIS CLU data centre unit is making use of the Cloudnet database infrastructure. Cloud remote sensing data submitted to ACTRIS need to be in a specified format compliant with the centralized processing suite. CLU provides data compliant with netCDF 3 and netCDF4 formats as much as possible, and following CF 1.7 convention. Level 0 data submitted to ACTRIS CLU are required to be in a specified format compliant with the centralized processing suite. All further data levels are produced by the CLU processing suite.

Re-use of existing data

The ACTRIS data user interface will include access to cloud remote sensing legacy data resulting from ACTRIS pre-projects (for CLU Cloudnet). These will also be included as a part of the ACTRIS CLU data centre unit. Legacy data resulting from ACTRIS preprojects will be available in the same format as current products.

The origin of the data

Data is derived from instrument raw data, coupled with thermodynamic profiles from NWP model.

The expected size of the data

Type	Number of annual datasests (end 2019)	Number of annual datasets (min by 2025)	Number of annual datasets (max by 2025)
ACTRIS cloud remote sensing data	11	15	25

Table 6: Number of annual datasets

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
ACTRIS cloud remote sensing data	15 TB	50 TB	150 TB

Table 7: Data volume

Data utility

Clouds are highly variable in time, space, and in their macro- and microphysical aspects. This variability directly impacts radiative transfer and the hydrological cycle, and the accurate representation of clouds is fundamental to climate and numerical weather prediction. CLU products are particular valuable for investigating the response of cloud microphysical processes to changes in other atmospheric variables (aerosol-cloud-precipitation interaction), evaluating and developing the parametrization schemes used to represent cloud in climate and numerical weather prediction models, and for validating satellite products used in data assimilation.

CLU level 2 data are utilised by a large community of atmospheric scientists and operational agencies, with products permitting both process studies and model parametrization CLU level 3 comprises climatological products for climate and forecast model evaluation, together with seasonal and diurnal composites enabling the characterisation of cloud properties across Europe.

Outline of data life cycle (workflow and workflow diagram)

Details on the data life cycle and workflow (workflow diagrams for data production) for remote sensing observations can be found in Appendix 5: ACTRIS cloud remote sensing data lifecycle and workflow (draft).

3.4 ACTRIS trace gases remote sensing data centre unit (GRES)

The ACTRIS trace gases remote sensing data centre unit (GRES) is supported by the AERIS data base (https://gres.aeris-data.fr). The GRES data centre unit provides data curation service for reactive trace gases remote sensing data. This includes data conversion processing, standardized process for data submission, quality control, inclusion of data in the data base, search metadata creation, data provision and archiving. In addition, data product generation of level 3 for climatological analysis and added values products (quicklooks, links to EVDC-ESA Atmospheric Validation Data Centre) is offered and implemented during the implementation phase.

The ACTRIS-GRES unit is structured in one unique database including measurements issued from five types of instruments: - FTIR: Fourier Transform Infra-Red Spectrometry, - UVVIS: Ultra-Violet and Visible spectroscopy including - UV-VIS ZS (zenith-sky) SAOZ (Sytème d'Analyse par Observation Zénithale) spectrometer, - UVVIS MAX-DOAS (Multi-AXis Differential Optical Absorption Spectroscopy instruments), - PANDORA instruments. - LIDAR DIAL: Differential Absorption Lidar.

The types and formats of data generated/collected

Level 2 and level 3 trace gases remote sensing data produced are either profiles (O3) or columns (O3, C2H6, HCHO, NO2, NH3 …) products. The level 2b data are processed from the consolidation of level 2a data using quality assurance and quality control procedures. The level 3 data are produced from level 2b data, trace gas profiles or columns, and correspond to monthly averaged climatologies as well as coincident data with satellite overpasses. Level 2 and level 3 trace gases remote sensing data submitted to ACTRIS need to be in GEOMS HDF data format (Generic Earth Observation Metadata Standard, http://www.ndsc.ncep.noaa.gov/data/formats) following the appropriate GEOMS template for FTIR, UVVIS and LIDAR measurements (https://evdc.esa.int/tools.data-formatting-templates/). The GEOMS

data format allows the necessary requirements to setup the ACTRIS data curation service for trace gas remote sensing data. The level 2 and level 3 data will be also converted in NetCDF (https://www.unidata.ucar.edu/software/netcdf/)) version 4 format following the CF 1-7 (Climate Forecast) conventions and be disseminated. The Climate and Forecast conventions are metadata conventions for earth science data. The conventions define metadata that are included in the same file as the data making the file "self-describing". Level 0 and 1 data submitted to ACTRIS GRES are required to be in a specified format compliant with the centralized processing suite. All further data levels are produced by the NF processing suite.

Re-use of existing data

The ACTRIS data user interface will include access to reactive trace gases remote sensing legacy data data resulting from AERIS project (for NDACC data) . These will also be included as a part of the ACTRIS GRES data centre unit. Legacy data resulting from AERIS project will be available in the same format as current products.

The origin of the data

The L2 data are derived from instrument raw data, through offline observations. All the data processing is performed by NFs.

The expected size of the data

Type	Number of annual datasets (end 2019)	Number of annual datasets (min by 2025)	Number of annual datasets (max by 2025)
ACTRIS-GRES FTIR	276	624	3744
ACTRIS-GRES SAOZ	2900	7250	14500
ACTRIS-GRES MAX-DOAS	14600	14600	14600
ACTRIS-GRES PANDORA	37230	7665	10220
ACTRIS-GRES LIDAR DIAL	400	100	200

Table 8: Number of annual datasets

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
ACTRIS-GRES FTIR	2,5 GB	25 GB	150 GB
ACTRIS-GRES SAOZ	0,6 GB	6 GB	15 GB
ACTRIS-GRES MAX-DOAS	600 GB	3 TB	3 TB
ACTRIS-GRES PANDORA	1,7 TB	6 TB	10 TB
ACTRIS-GRES LIDAR DIAL	0,4 GB	1 GB	2 GB

Table 9: Data volume

Data utility

The data of GRES could be used to monitoring the evolution of key stratospheric gas trace such ozone under the effect of anthropogenic emissions, climate change and natural events. The retrieval of previous data by GRES will allow homogeneous data series to compute trends. Data could be also used in support of validation of satellite measurements deployed by international space agencies as well as models simulations.

Outline of data life cycle (workflow and workflow diagram)

Detail on the data life cycle and workflow (workflow diagrams for data production) for trace gases remote sensing data can be found in Appendix 6.

3.5 ACTRIS Atmospheric simulation chamber data centre unit (ASC)

The ASC data centre unit provides data curation service for data obtained from experiments in atmospheric simulation chambers (ACTRIS exploratory platforms). This includes tools for harmonized data submission and meta data templates, inclusion of data and metadata in the database, traceability, harmonization and data versioning, quality control, archiving, documentation and data provision. The ASC unit is structured in three pillars:

- The Database of Atmospheric Simulation Chamber Studies (DASCS) provides access to experimental data (level 2 data), typically time-series of measured parameters during an experiment in a simulation chamber.
- The Library of Analytical Resources (LAR) provides quantitative analytical resources that include infrared spectra and mass spectra of molecules and derivatives (level 3 data).
- The Library of Advanced Data Products (LADP) provides different types of mature data products (level 3 data): rate constants of reactions, quantum yields and photolysis frequencies of trace gas compounds, secondary organic aerosol (SOA) yields, mass extinction/absorption/scattering coefficients and complex refractive index of aerosols, growth factors of aerosols and modelling tools. The detailed list of ACTRIS level 3 data products is given in Appendix 9.

The types and formats of data generated/collected

The ACTRIS ASC data centre unit is making use of EUROCHAMP database (https://data.eurochamp.org/) which is hosted by AERIS infrastructure. Data submitted to the DASCS pillar have to be provided by NFs in a standard format, called "EDF format" (EUROCHAMP Data Format) which is based on an ASCII text format and contains additional metadata in a header. These data are completed with rich metadata which are available from the website and give access to a technical description of the chambers (size, volume, walls, irradiation system), the experimental protocols used for the generation of the data, and an "auxiliary mechanism" which provides the chamber-dependent parameters affecting the observations. Currently, work is being conducted with regards to providing tools for access and download of data also in the netCDF 4 format, compliant with the CF 1.7 convention. This will be implemented during ACTRIS implementation phase.

Level 3 data provided in LAR are IR and mass spectra in JCAMP-DX format which is the standard format recommended by IUPAC for spectra. It is a 2D graphic format based on ASCII format. Metadata are attached and made available through the ACTRIS data user interface. These data are provided by NFs.

Level 3 data provided in LADP are of different types and have thus different formats. However, each type of data is provided in a harmonized format. Most of them are provided as a unique value with metadata attached.

Re-use of existing data

The ACTRIS data user interface will include access to atmospheric simulation chamber legacy data resulting from ACTRIS pre-projects (for ASChere EUROCHAMP-1, -2, and EUROCHAMP-2020). These will also be included as a part of the ACTRIS ASC data centre unit. Legacy data resulting from ACTRIS pre-projects will be available in the same format as current products.

The origin of the data

Data provided in DASCS and LAR pillars are derived from instrument raw data and data provided in LADP are produced from L2 data processing. All the data processing is performed by NFs.

The expected size of the data

Type	Number of annual datasets (end 2019)	Number of annual datasets (min by 2025)	Number of annual datasets (max by 2025)
ACTRIS-ASC DASCS	200	50	300
ACTRIS-ASC LAR	20	10	50
ACTRIS-ASC LADP	70	50	200

Table 10: Number of annual datasets

Type	Data volume (end 2019)	Data volumes (min by 2025)	Data volume (max by 2025)
ACTRIS-ASC DASCS	1,2 GB	1,5 GB	2,4 GB
ACTRIS-ASC LAR	67 MB	76 MB	120 MB
ACTRIS-ASC LADP	26 KB	200 KB	500 KB

Table 11: Data volume

Data utility

Atmospheric simulation chamber data contribute to better predict the behavior of the atmosphere over all time scales through a detailed understanding of the physical and chemical processes which affect air quality and climate change. ACTRIS-ASC unit gives access to different types of data and data products essential to a wide range of communities. Many of these parameters are incorporated in air quality and climate models.

- Level 2 data provided in DASCS are of high interest for a large community of users in atmospheric science research and related areas, as well as the private sector. In particular, they are largely used for modelling activities to develop and/or validate chemical schemes of atmospheric models.
- Level 3 data provided in the LAR are of high interest for a large community of users in atmospheric sciences, analytical chemistry and related areas, as well as the private sector. Indeed, quantitative chemical analysis of infrared spectra for complex mixtures requires access to standards for the calibration of instruments. However, as the chemical species formed by these processes are often very complex (and not commercially available), their spectra are not available in the "classical" databases of analytical chemistry, or are not useful due to their low resolution. To tackle this issue, the EUROCHAMP consortium has developed its own Library of infrared spectra and has made it freely available to the scientific communities.
- Level 3 data products provided in the LADP are especially useful for researchers working on atmospheric observations, as well as atmospheric model development and validation. It includes products for the development of chemical mechanisms in atmospheric models (e.g. rate coefficients, photolysis frequencies, SOA yields, vapor pressures, etc.), products for the retrieval of

satellite data and for radiative transfer modelling (e.g.), and tools to generate oxidation schemes which are very useful to interpret field measurements as well as laboratory studies.

Outline of data life cycle (workflow and workflow diagram)

Detail on the data life cycle andA preliminary version of the data workflow (workflow diagrams for data production) for Atmospheric Simulation chamber data can be found in Appendix 7. The definition of this workflow is still under progress and a finalized version will be available in 2020.

3.6 ACTRIS data and services (ACCESS)

ACTRIS Data Center is a distributed data centre, and ACTRIS data and services access unit (ACCESS) is responsible for organising access to measurement data from the topic data centre units, and documentation of procedures as support to observational and exploratory NFs. The ACCESS unit provides the ACTRIS web interface for data download, services and digital tools as well as performing data production of Level 3 data, and synergy data products.

The ACTRIS access web interface is called "The ACTRIS Data Centre" and incldues a meta data catalogue. The main activities are **Discovery and access** to ACTRIS data and data products, overview of digital tools provided by the topical centres and the data centre units, documentation, software and tools for data production. **Visualisation** of ACTRIS data products. **Data production** of Level 3 data and synergy data products. The data centre also offers **bridge to external data bases and sources**.

The ACTRIS ACCESS unit offer acecss to elaborated aerosol, cloud and trace gas data products, issued of advanced multi-instrument synergistic algorithms, long term reanalysis, modelling and satellite data and sources. These can be produced within the ACCESS unit, topic data centre units, topic centres, or extrnal contributions. The list of ACTRIS level 3 data products is detailed in the Appendix II, and consiste of three main categories:

I. **Level 3 data solely based on data from ACTRIS observational platforms**

II. **Level 3 data and tools from multi-source data integration services, employing external gr ound based measurement data**

III. **Level 3 data products involving regional and global model data**

The types and formats of data generated/collected

The objective is that most of the level 3 data generated will be in NetCDF data format and have metadata compliant to the NetCDF CF Metadata Conventions. This format and metadata are widely used in the atmospheric science community, and is supported by a lot of standard visualization and analysis tools. Nevertheless, the collected data can come from external sources accordingly, non standard formats may also be used. In these cases, they will be rather kept in their original format.

Re-use of existing data

The generated products and online services available from ACTRIS-preproject use existing ACTRIS L0-1-2, satellite and model data.

The origin of the data

The origin of the data is derived from ground-based and satellite observations, retrieval algorithms and model simulations.

The expected size of the data

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
GRASP-AOD	0 GB	50 GB	20 TB
ReOBS	2 GB	30 GB	100 GB

Table 12: Number of annual datasets

Type	Data volume (end 2019)	Data volume (min by 2025)	Data volume (max by 2025)
GRASP-AOD	0 GB	50 GB	20 TB
ReOBS	2 GB	30 GB	100 GB

Table 13: Data volume

Generated (on-demand services)

Product	Typical dataset per day	Typical volume per day
Satellite data subsets	100	100 MB
Transport modelling products for assessment of source regions	...	
Colocation service of data from contributing networks	400	400 MB
Model Evaluation Service	30	300 MB
NWP Model Evaluation Service	120	100 MB

Data utility

Data from ACTRIS is contributing to better prediction of the behavior of the atmosphere over all time scales through a detailed understanding of the physical, optical and chemical properties of aerosols, clouds and trace gases in the atmosphere, as well as data for improving knowledge of processes which affect air quality and climate change.

ACTRIS data are very diverse, covering numerous measurement methodologies resulting in a considerable diversity of the data collected. In accordance with these requirements, the ACTRIS DC is organized in 6 Units.

The ACCESS unit utilize data from all the 5 topical DC units, and produce level 3 products combining various data and also models prodcuibf new and value added data producst. Accordingly, the ACCESS data utility can cover all the more specific data utility sections described in sections 3.1 – 3.5.

Outline of data life cycle (workflow and workflow diagram)

Detail on the data life cycle and workflow (workflow diagrams for data production) for level 3 data can be found in Detail on the data life cycle and workflow (workflow diagrams for data production) for level 3 data can be found in Appendix 8

4. Data Management at the ACTRIS data centre

ACTRIS data and products should be findable, accessible, interoperable and reusable (FAIR), and the data centre works towards fulfilling the FAIR principles. This chapter is describing the operational ACTRIS Data Flow from National Facilities (NF) to users now, decisions that are currently under implementation, and the work and solutions that will be implemented during the implementation phase (2020-2025). The section starts with a brief introduction to ACTRIS Access strategy, then introduction to the data management system in ACTRIS in section 4.1, including detailed descriptions of data flows within each unit (4.1.1.-4.1.5). This is followed by sections describing detailed solutions and implementation plans making ACTRIS data and products findable (4.2), accessible (4.3), interoperable (4.4), and reusable (4.5).

4.1 ACTRIS access and service policy

ACTRIS is offering access to a large variety of high-quality services offered by ACTRIS facilities, to a wide range of users and needs, for scientific, technological and innovation-oriented usage. Accordingly, ACTRIS has developed an access strategy to give clear guidelines and describe the general principles for access provided by ACTRIS to Users. When the ACTRIS services are in operation, the Users will access the ACTRIS services through a single entry point, as shown in Figure 4 below.

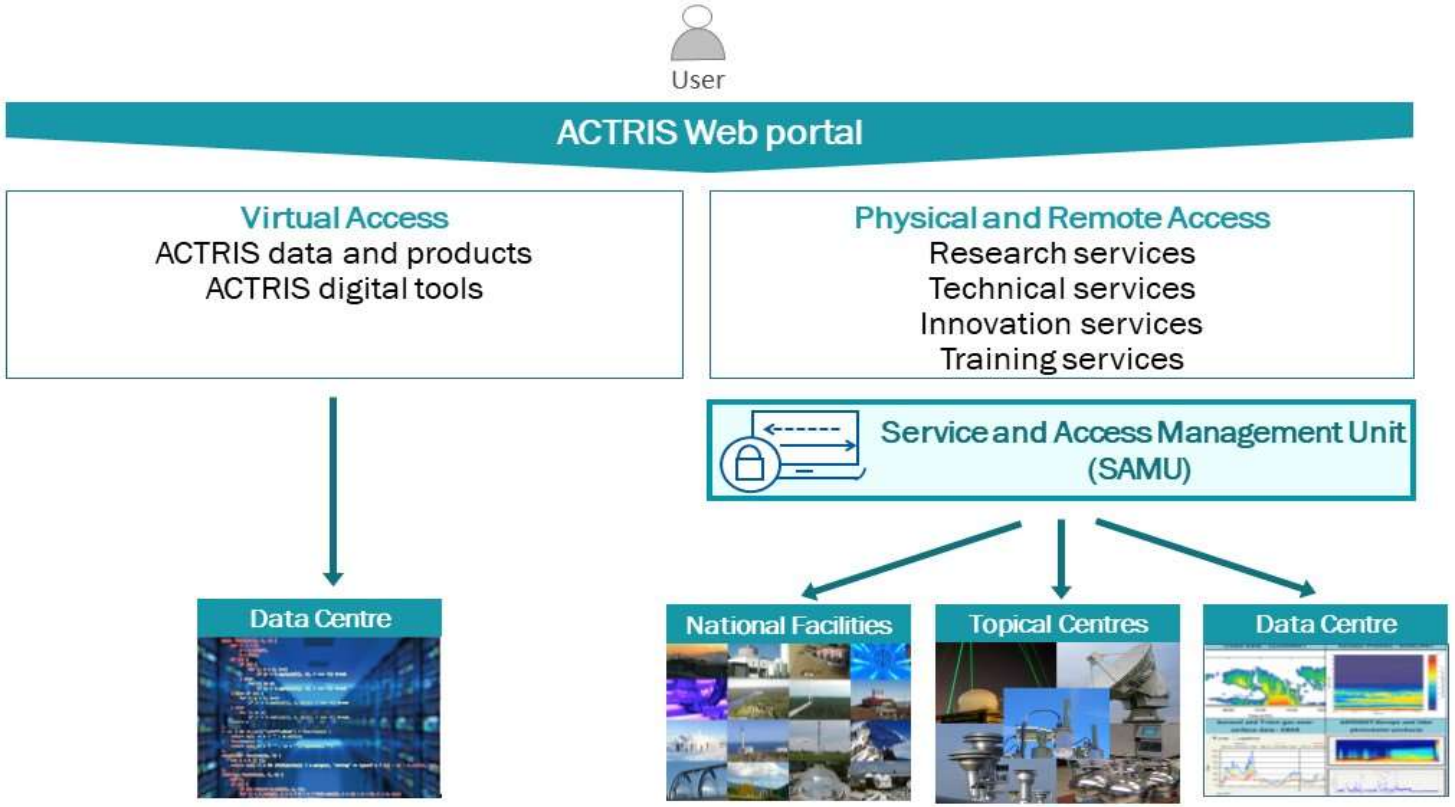


Figure 4: Overview of ACTRIS Access strategy

Virtual access is wide access to ACTRIS data and digital tools and does not require a selection process. Virtual access to ACTRIS data and digital tools is free access, and is given in compliance with the ACTRIS data policy for data from ACTRIS labelled NFs. Competitive access is Physical or Remote access to the ACTRIS Facilities, including access to specific services offered by the Data Centre and shall be managed by the SAMU and requires a selection process. This can e.g. be data centre services for comprehensive research campaigns or large volume data delivery tailored for specific purposes.

4.2 Introduction and overview of ACTRIS Data Management architecture

ACTRIS Data Management is handled by the individual data centre unit:

- ACTRIS In situ data centre unit for all aerosol, cloud and trace gas in situ data - In-Situ
- ACTRIS Aerosol remote sensing data centre unit - ARES
- ACTRIS Cloud remote sensing data centre unit - CLU
- ACTRIS Trace gases remote sensing data centre unit - GRES
- ACTRIS Atmospheric simulation chamber data centre unit – ASC
- ACTRIS data and service access unit - ACCESS

An overview of the elements in the data flow is shown in Figure 5.

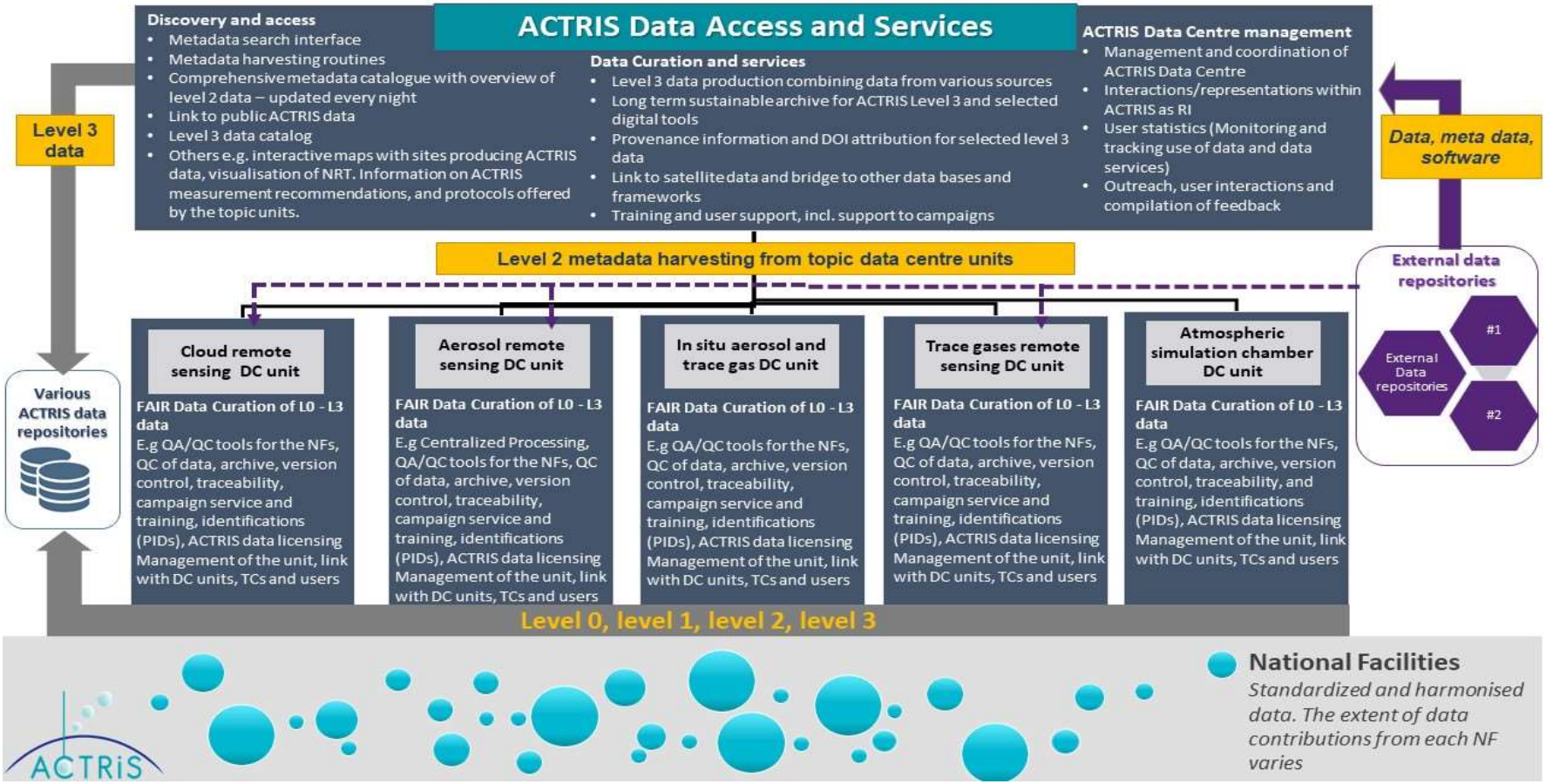


Figure 5: Overview of ACTRIS Data Centre components, links and main activities for the various units

4.2.1 ACCESS role and data management

Access to quality controlled data from the topic data centre units is organised by the ACTRIS data and service access unit (ACCESS). The ACCESS unit includes a frequently updated metadata catalogue for identification of data, and links to measurement data stored by the topical data centre units. ACCESS also produces level 3 data, and organizes the catalogue of ACTRIS level 3 data, produced either by the topical data centre units or within ACCESS. ACCESS is structuring and compiling access to services, tools and documentation, and maintaining and further developing the web interface called “ACTRIS Data Centre” (currently https://actris.nilu.no).

The tasks are summarized in Figure 5 above and include the organization of ACTRIS level 3 data.

All data centre units are providing metadata, and interfaces for access to data and metadata indexed in the current ACTRIS metadata catalogue, except for ASC. Index to ASC data is under implementation, and with the aim to be ready within first part of 2020. The metadata is used to identify and access data through the ACTRIS Data Centre web portal. The metadata catalogue is regularly updated, at least every night through automatic procedures. ASC unit has developed their own metadata catalogue and data and metadata is currently available through EUROCHAMP Data Centre portal.

Figure 7 the current technical architecture and the interface used between the topical data center units, as well as ACCESS and the ACTRIS Data Centre web interface with access for users.

- We use ISO19115 with WIS profile as starting point for metadata exchange. The profile will be extended with ACTRIS specific metadata.

The current setup is a web portal with a database that collects metadata from In Situ, ARES, CLU and GRES via custom web services, but currentyl machine-to-machine access is not possible. Implementation of ASC is under development and will be ready during 2020, and in the future, the aim is to harvest all ACTRIS metadata in to a single metadata catalogue, providing discovery metadata for all ACTRIS data using ISO19115 with the WIS metadata profile enabling machine-to-machine access of ACTRIS metadata.

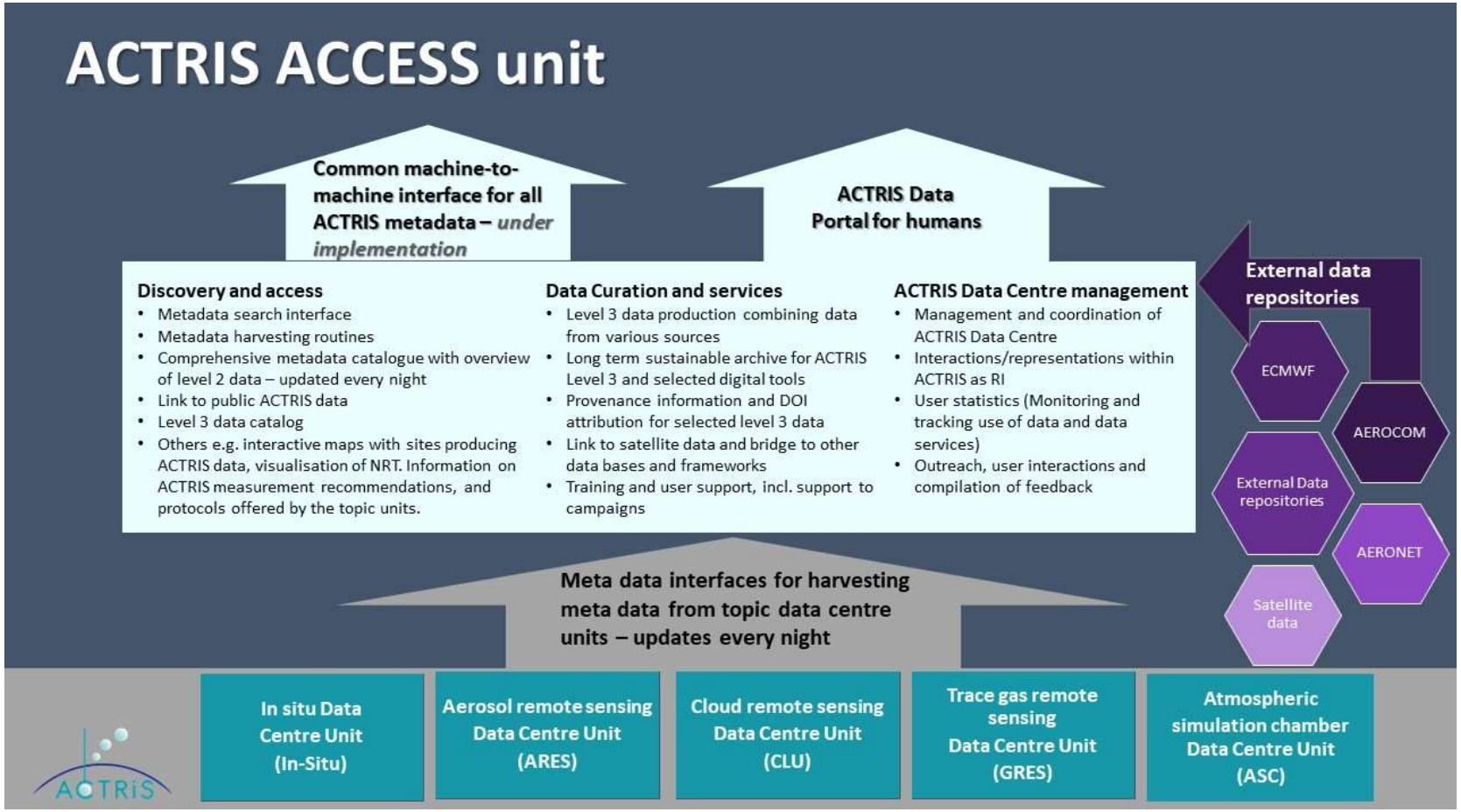


Figure 6: Overview of the tasks of the ACCESS unit

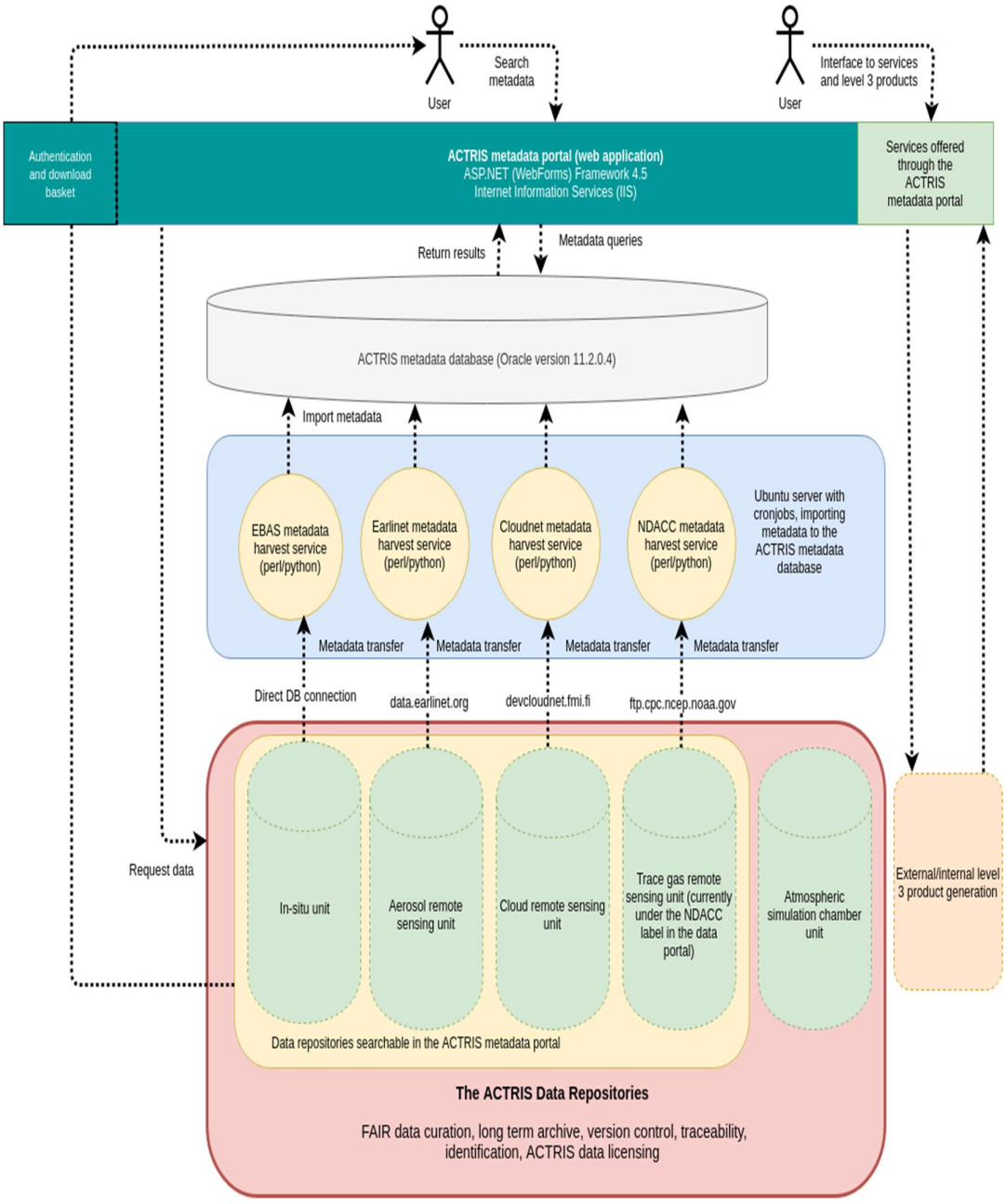


Figure 7: Technical architecture of the ACTRIS meta data portal

As visualized in Figure 6 ACCESS organizes the level 3 data. The collected and generated level 3 datasets will be extended during the implementation phase, and the complete list of variables under implementation is included in Appendix 2. Details of the level 3 data production in operation is included in Appendix 9.

Overview of when data is made available (level 2 data)

DC unit	Submission deadline	Date when data is made available by the DC unit	Provsion of NRT data	Comment
In Situ	31th of May	31th of June	hourly	
GRES	Within 4 months after measurement	Within 4 months after measurement		There is not a specific date for data submission and availability for GRES and ASC unit. Example: for FTIR data, NF will deliver data every 1 to 3 month; and 15 days later the data will be available by the DC unit.
ARES				
ASC			Not applicable	
CLU	Automatic	1 day after submission		

Table 14: Overview of when data is made available

4.2.2 In-Situ dataflow and data management

The data management of ACTRIS in situ aerosol, cloud, and trace gas variables (listed in Appendix 1) follows a common workflow (see Appendix 3 for details). The workflow is separated into 2 branches: * **Online observations**: Measurement done directly on sample air stream immediately after sampling, measurement reported by instrument while sample passes through or immediately after. Instrument QA by on- and off-site comparisons to standard instruments / primary standards. RRT data provision is possible and default. * **Offline observations**: Measurement done on sample medium in which sample is collected. Sample analysis usually disconnected from sample collection in time and location. Sample handling is documented by a series of reports, leading to final data product. QA on sample handling (e.g. field blanks) and analysis (e.g. round-robin). Rapid delivery for data possible.

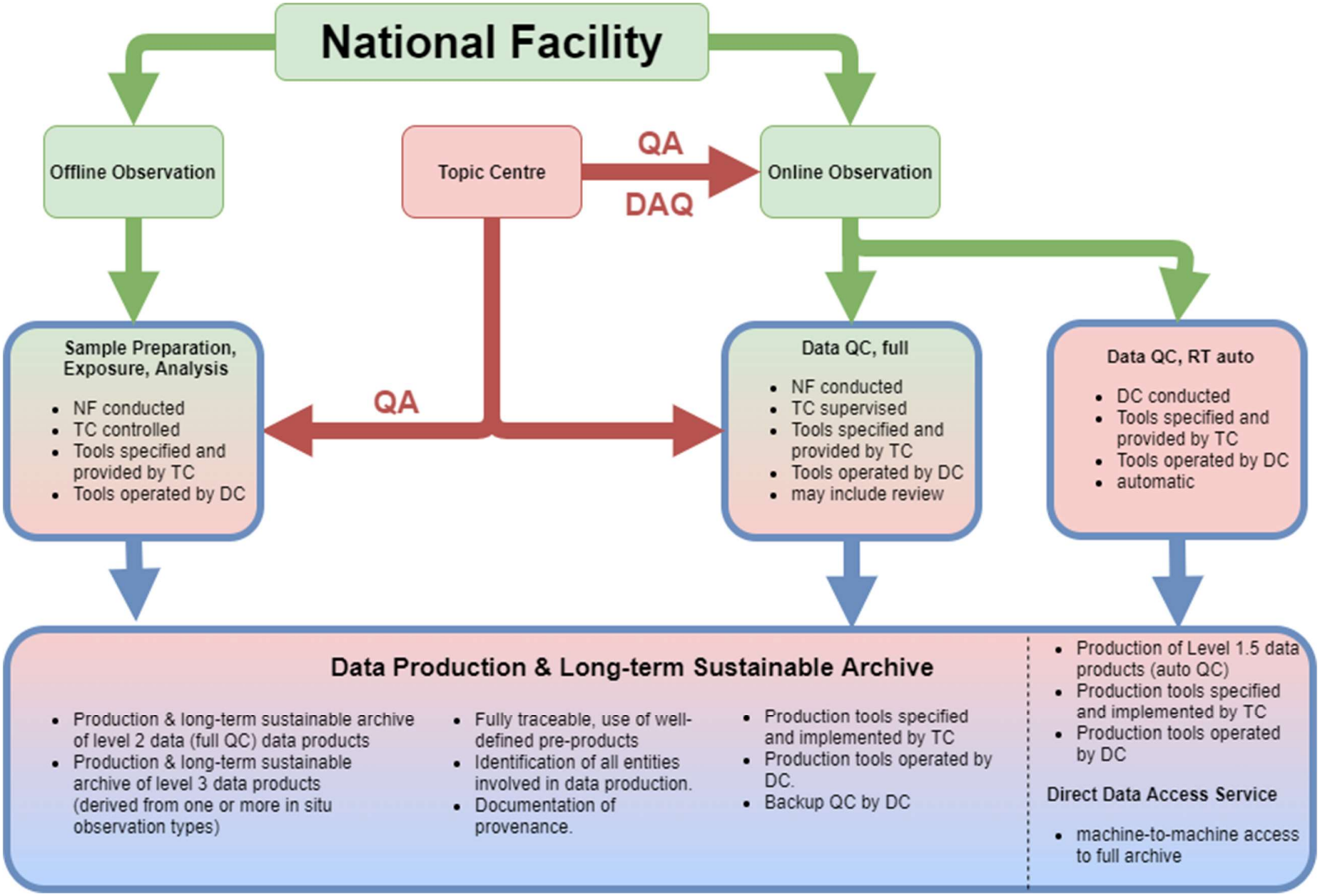


Figure 8: Simplified workflow of the ACTRIS In Situ data centre unit, focussing on distribution of responsibilities and services to users.

If an offline analysis process has been sufficiently streamlined, it may be described by the online workflow.

ACTRIS In situ concretises the ACTRIS data levels as follows: * **Level 0**: by default, raw data as produced by the instrument, all main and instrument status parameters provided by instrument, brought to a well-defined data format. If needed to limit raw data volume, level 0 may be a condensed version of instrument raw data. Discovery, use, provenance, and configuration metadata attached, including all information needed for further data production, as far as known at stations level (e.g. QC metadata). Instrument model specific, time resolution native to instrument, temperature and pressure conditions as provided by instrument. * **Level 1**: Contains physical property observed, including measures of uncertainty. Instrument status parameters, QC measurements and invalid data removed, quality control and post-calibrations applied. Time resolution native to instrument. Content specific for measured property, standard conditions of temperature and pressure. * **Level 2**: Contains data averaged to a homogeneous time resolution, typically 1 h. Includes measure of atmospheric variability, content specific for measured property.

4.2.2.1 General Characteristics of In Situ Data Production

All In Situ data products, level 1 (RRT) and level 2 (fully QCed), are archived in the In Situ data repository, hosted in NILU's EBAS database, and made available through ACCESS. In Situ produces selected level 3 products from these (Appendix 9).

As a general rule in ACTRIS in situ data QC, only data subject to instrument malfunctions are flagged as invalid. Data with episodes, local contamination, etc. are flagged as such, but not as invalid, to avoid limiting data use only to certain use cases. Data level designations are used to distinguish between data having received automatic QC, and data having undergone manual QC. When calculating time averages in data production, any non-invalidating flags occurring in the source data during the averaging period are copied to the average result, while invalid data points in the source data are excluded from the average. Temporal coverage flags are applied to the average.

The content of workflow tasks and the responsibilities for them are specified in Appendix 3, separately for each In Situ instrument type. Certain responsibilities common to both online and offline observations are by default distributed between NF, TC, and DC as follows:

NF: * Conducts / participates in on-site, off-site QA measures, and round robin exercises as specified by TC. * Reacts to feedback on instrument / observation / analysis operation and data quality from both TC and DC within 1 week, and takes corrective action within 1 month.

TC: * Maintains operating procedures for each In Situ instrument type in his responsibility. * Defines QA/QC measures necessary for an instrument type / observation. * In collaboration with DC, specifies data and metadata items contained in QA/QC measure reports for each instrument / observation type. * Specifies actions to be executed in each workflow task box for each instrument / observation type. * Documents QA measures conducted by TC using the tools jointly specified / provided by TC and DC.

DC:

DC Core services: * Archives all level 0, 1, and 2 data occurring during workflow execution. * Archives level 3 data produced by In Situ. * Links data to relevant QA/QC data. * Operation of data production and QC tools for ACTRIS in situ data, administration of data production workflow ensuring homogeneous data products, e.g. via a business workflow tool connecting NFs, TC, and DC. * Archive for documentation of QA measure results throughout ACTRIS, setup of infrastructure, and standards of operation, including identification of documents * PID identification of all objects in ACTRIS workflows executions, incl. data (pre-) products, software, humans, organisations, instruments, including versioning, DOIs for level 2 data products. * Document provenance throughout all ACTRIS workflows by use of standardised provenance scheme, facilitating attribution of entities involved in workflow execution. * Training events for data submitters to all data centre units * Reacts on requests typically within 1 week / 1 month. * Deadline for publication of data, end of August or 1 month after closing the last issues. * Documentation, procedures, tutorials and tools, guidance and helpdesk available to NFs * Access to ACTRIS In Situ level 0, 1, 2, 3 data. * Access to ACTRIS level 2 legacy data archived in the ACTRIS data repositories, will be accessible via the ACTRIS web entry point. * Bridge to external ground-based observational data relevant for ACTRIS * Climatology products for ACTRIS variables at National Facilities across Europe. * Interoperability and link to other RIs and initiatives. * Knowledge transfer and training on the use of data products and tools * Monitoring task execution in unit, representing unit in ACTRIS bodies. * Monitor workflow execution across In Situ TCs and DC, maintain and update workflow elements. * Support to regional and global networks and related initiatives. ACTRIS will support international frameworks in the field of air quality and climate change, e.g. GAW including GALION, EMEP, and GCOS, and further utilize and add value to satellite based atmospheric observation.

DC Added-value services: * Access to Software, digital tools and user support for processing of ACTRIS data tailored for analysis and research * Aerosol surface in situ data – combination of variables and instruments. Production and distribution of surface in situ level 3 products (closure, model comparison, full-range PNSD, PM mass from size distribution, key particle optical properties at dry state). * Contribution to collocation service of data from regional and global networks. Benchmark data products adding complementary data from GAW and EMEP together with ACTRIS data. * Contribution to Satellite data – combined with ground based ACTRIS data. On-demand distribution of satellite data collocated with ACTRIS ground-based observations * Contribution to Optimal interpolation and Gap filling tool.

DC Campaign services: * Provision of digital tools and data services during observation campaigns. * Data curation and archive of campaign data. * Digital tools and products for campaign support. * Campaign dashboard service.

4.2.2.2 Online In Situ Data Production

Already at the station, the raw data stream from the instrument is transcribed to a homogeneous level 0 format, and annotated with discovery, use, provenance, and configuration metadata. The level 0 data are transferred to the ACTRIS DC at a RRT schedule (latest 3 h after measurement, normally 1 h). At this point, the In Situ online workflow splits into 2 branches: 1. **RRT data production:** incoming level 0 data are auto-QCed for outliers, spikes, calibration data, and automatically identifiable instrument malfunctions, and flagged accordingly, yielding level 0b. From there, levels 1b and 1.5 (final RRT data product) are produced. RRT data are offered through a data subscription service. 2. **Fully QCed data production:** data are manually QCed for outliers, spikes, calibration data, episodes (e.g. atmospheric transport, local / regional

contamination), and instrument malfunctions. Tools for manual data QC are provided centrally. Manual QC is assisted by automatic pre-screening of data, similar to the auto-QC for RRT data. There are 2 options for organising the QC process, both are applied at least annually: 1. **TC review**: data QC is conducted by NF and supervised by TC, and follows its own sub-workflow. 2. **NF review**: data QC by an identified person under the responsibility of the NF.

From the fully QCed level 0 data, i.e. level 0a, levels 1a and 2 (final data product) are produced.

All fully QCed data are to be submitted to the In Situ DC unit on an annual basis by 31 May of the year following the calendar year to be reported. If the TC review option is used, NFs need to submit their initial QCed version to the review process by 31 March of that year, where the review process is typically supervised by the TC.

The content of workflow tasks and the responsibilities for them are specified in Appendix 3, separately for each In Situ instrument type. The following responsibilities specific to online observations are distributed between NF, TC, and DC as such:

NF: * Operates the instrument according to TC recommendations. * Conducts data QC as specified by TC, documents QC as specified jointly by TC and DC, participates in data QC review, if applicable. * Conducts on-site QA measures and calibrations as specified by TC, and documents them using the tools specified and provided by TC. * Uses data acquisition and handling software as provided / specified by TC. * Maintains infrastructure for RRT data transfer.

TC: * Implements and maintains data production and QC software for each In Situ instrument type in his responsibility to NF. * Supervises on-site QA measures and calibrations, provides specification and tools for documenting these in a traceable way. * Conducts off-site QA measures on instruments as required by instrument type. * In collaboration with DC, specifies data and metadata items contained in data levels 0, 1, and 2 for each instrument type. * Implements and maintains software executing task boxes in data production workflow. * In collaboration with DC, specifies sub-workflow implementation for data QC review, including review procedures and rules, if applicable. * Conducts and supervises data QC review, if applicable.

DC:

DC Core services: * NRT, RRT data production. * Data submission & curation service of online ACTRIS in situ data. * Secondary data QC before publication of data (syntactic and semantic correctness of metadata, syntactic correctness of data, parameter dependent completeness of metadata, completeness and correctness of framework associations).

DC Added-value services: * Contribution to services co-ordinated by other ACTRIS partners (source apportionment submicron organic aerosol, VOC source attribution, cloud occurrence at cloud in situ NFs). * PM retrieval @GAW sites globally. * Alert Service for National Facilities on instrument malfunctions.

4.2.2.3 Offline In Situ Data Production

In the offline workflow for ACTRIS in situ data, data production is centred around the sample medium, following its way through the workflow: * **Sample Medium Pre-Treatment:** Pre-heating, Impregnation, Weighing in, Pre-cleaning. * **Sample Medium Exposure:** Transport to field station, Exposure in sampling device, Transport to lab * **Sample Preparation:** Weighing out, Sample medium extraction, Sample medium apportioning. * **Sample analysis.**

Again, the content of workflow tasks and the responsibilities for them are specified in Appendix 3, separately for each In Situ offline instrument type. The following responsibilities specific to offline observations are distributed between NF, TC, and DC as such:

NF: * Conduct sample medium pre-treatment, sample medium exposure, sample preparation, and sample analysis in accordance with procedures defined by TC. * Document all sample handling steps in the documentation system specified and implemented by TC, and operated by DC. * Evaluate sample according to TC guidelines, using the tools specified and implemented by TC, and operated by DC * Respond and act on quality control feedback by DC. * Participate in quality assurance measured defined and conducted by TC, e.g. round-robin exercises.

TC: * Specify guidelines and implement documentation system for sample medium pre-treatment, sample medium exposure, sample preparation, and sample analysis. * Specify guidelines and implement algorithm for sample evaluation. * Specify guidelines for data quality control. * Specify procedures for, implement, and conduct quality assurance measures as appropriate for observation method, document them, and store results in QA measure database operated by DC.

DC:

DC Core services: * Operate documentation system for sample medium pre-treatment, sample medium exposure, sample preparation, and sample analysis specified and implemented by TCs. * Operate archive for documentation of sample handling steps. * Operate algorithm for sample evaluation. * Implement and operate quality control step for offline level 2 data.

4.2.3 ARES dataflow and data management

At the present, the ACTRIS aerosol remote sensing component is a highly inhomogeneous in terms of instrumentations: most of the lidar systems are home-made or highly customized. In cases like that, the implementation of a standard, centralized and quality assured scheme for the analysis of raw data is the most efficient solution to provide FAIR and quality assured data at RI level.

The SCC (EARLINET Single Calculus Chain) is the solution adopted by the ACTRIS (Aerosol, Clouds and Trace gases Research InfraStructure Network) aerosol remote data center to ensure homogenous, traceable and quality controlled data. Main concepts at the base of the SCC are automatization and full traceability of quality-assured aerosol optical products.

The ARES DC also compile aerosol optical and physical properties (profile and column) from combined lidar + photometer observations collected at NFs. The GARRLiC (Generalized Aerosol Retrieval from Radiometer and Lidar Combined data) retrieval will be used for this, which synergistically inverts coincident lidar and radiometer observations, starting from SCC products and AERONET-ACTRIS processing stream products. These processing streams are fully controlled by ACTRIS.

The data curation workflow is suitable for the provision in NRT and RRT, following the same steps and procedures of the standard processing. NRT/RRT delivery of not fully quality assured data can be possible as long as a NF provides raw data to the DC in NRT/RRT.

Raw data collected at the NF in the original acquisition data format are transcribed in a homogeneous and agreed netCDF data format to the aerosol remote sensing processing suite at the ACTRIS DC, being the ARES Level 0 data. All information needed for the steps forward in the processing chain is annotated into the file or in a dedicated database (SCC database). It is recommended that raw data should be centrally stored and it should be under the responsibility of the NF to keep a local backup.

Level 0 data are centrally progressed at ACTRIS ARES DC level, generating Level 1 (not-fully QC data) preprocessed signals and optical properties products. On-the-fly QC procedures guarantees basic quality control on Level 1 data.

Off-line QC procedures are run systematically after the outcomes from related TC (namely CARS) are available and transferred to the ARES DC unit. The data originator and the CARS TC units receives feedback of the outcome of the QC. This feedback mechanism potentially allows to discover and address instrumental issues, with links to the TC. All the data compliant to all the QC requirements (both pre-processed and processed data) are made available as Level 2 data.

ARES DC offers also products resulting from the processing at DC itself of Level 2 lidar and photometer data collected at the aerosol remote sensing NFs. Finally, Level 3 climatological products are produced at ARES DC from lidar Level 2 optical property products.

Product Type	Availability (Typical)
Level 1	RRT / NRT
Level 2	1 year
Level 3	1 year

Table 15: ARES Data Products Availability

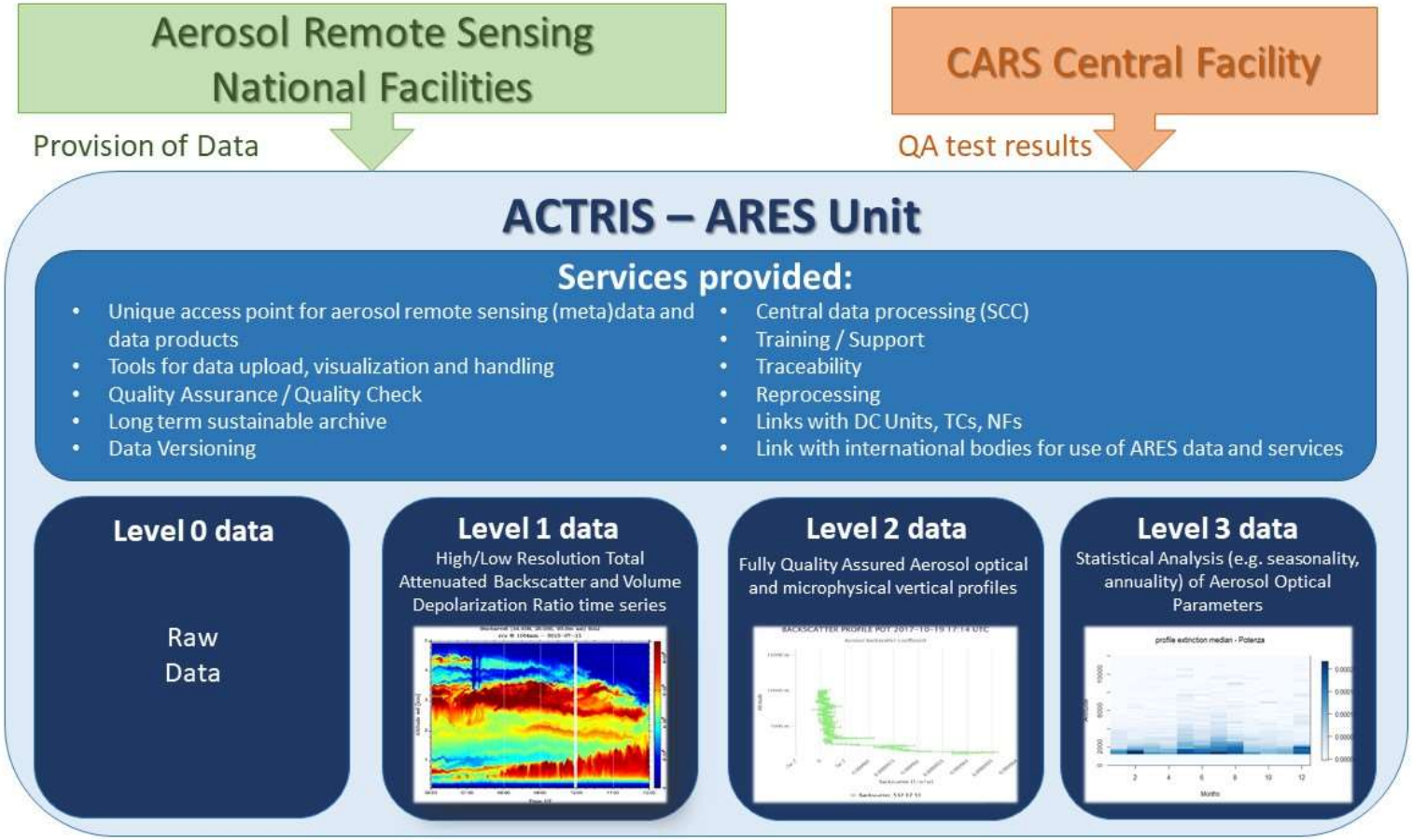


Figure 9: ARES Data Products Availability

4.2.4 CLU dataflow and data management

Modern cloud remote sensing instruments can produce vast amounts of raw data. This data first need to be stored locally at the measurement site. Then, the data are transferred to FMI servers for processing and archiving. Currently, FMI offers an FTP access point to establish the file transfer, but it is site operators responsibility to maintain regular data flow to FMI.

It should be noted that, technically, it is also possible to execute the first processing step already on site, and only transfer the processed measurement files, that are much smaller, to FMI for further processing. It is currently unclear if this option will be used in the operational ACTRIS processing or not.

At FMI, the raw measurement files from various instruments are processed to obtain more standardized netCDF files with a common metadata structure. In this stage, we also screen out noisy data points and apply possible calibration factors. This first processing step is applied to cloud radar and lidar measurements, but the microwave radiometer (MWR) data are processed elsewhere. FMI only receives the calibrated and processed Level 2 MWR files needed in the further processing steps.

After receiving and processing the raw data (and receiving MWR files), we generate all Level 2 cloud products with our in-house processing suite. All processed data are stored in netCDF files, which are archived on FMIs data servers. From the processed files, we generate a metadata database which is synchronized with the master metadata repository hosted by the ACCESS unit. All of our metadata is available as a JSON format via restful http API. The actual metadata standard is yet to be decided, but it must comply with the netCDF conventions because we use the netCDF file format. All data files encounter regular back-ups.

A general overview of the links between national facilities, CLU, and the corresponding topical centre, CCRES, are illustrated in Figure 9.

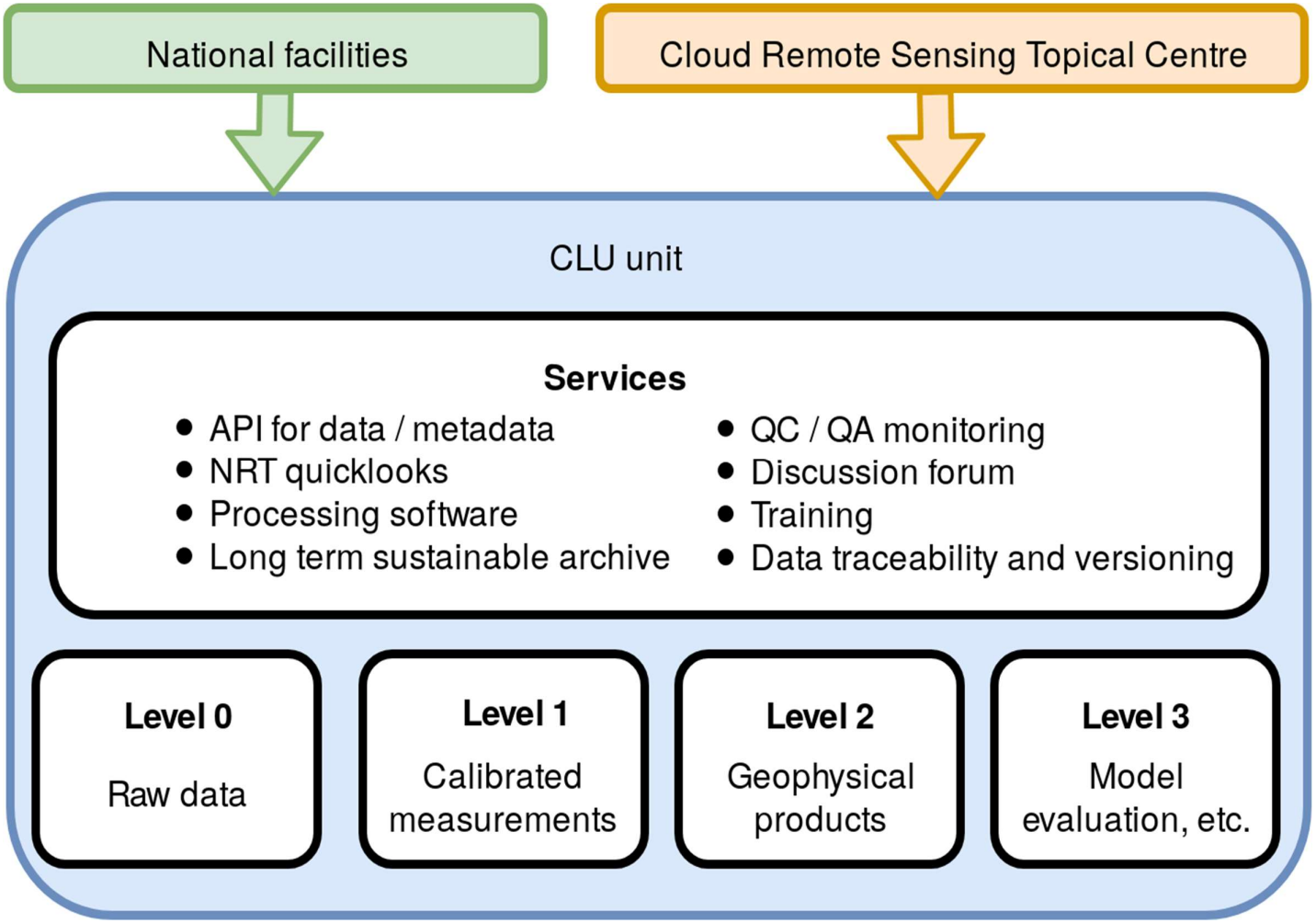


Figure 10. CLU data products and services

4.2.5 GRES dataflow and data management

Data provided in GRES unit are L2 and L3 data produced from L0 and L1 data processing performed at NFs level (see Figure 10). These data have to be provided by NFs in GEOMS HDF format then converted within GRES DC in NetCDF format in order to be disseminated through ACTRIS DC, These data have to be completed with rich metadata. NFs are also in charge of providing tools to facilitate the generation and the handling of the data. The GRES unit is in charge of: - creating and maintening the metadata catalogue, - providing a free and open access to metadata, data and tools developed by NFs through user-friendly web interfaces, - providing an open access to documents about description of algorithm retrieval by types of data and description of data quality assurance and control procedures using in NF's for the data production, - developing tools to convert L2 and L3 data into NetCDF format and to ensure the completeness of data provision process, - developing quicklooks for level 2 and 3 data in order to a more comprehensible understanding of the data, - offering links to EVDC-ESA Atmospheric Validation Data Centre), - assuring long-term archiving of L2 and L3 data. Jointly with NFs and TCs, it also contributes to the elaboration of the data workflow,

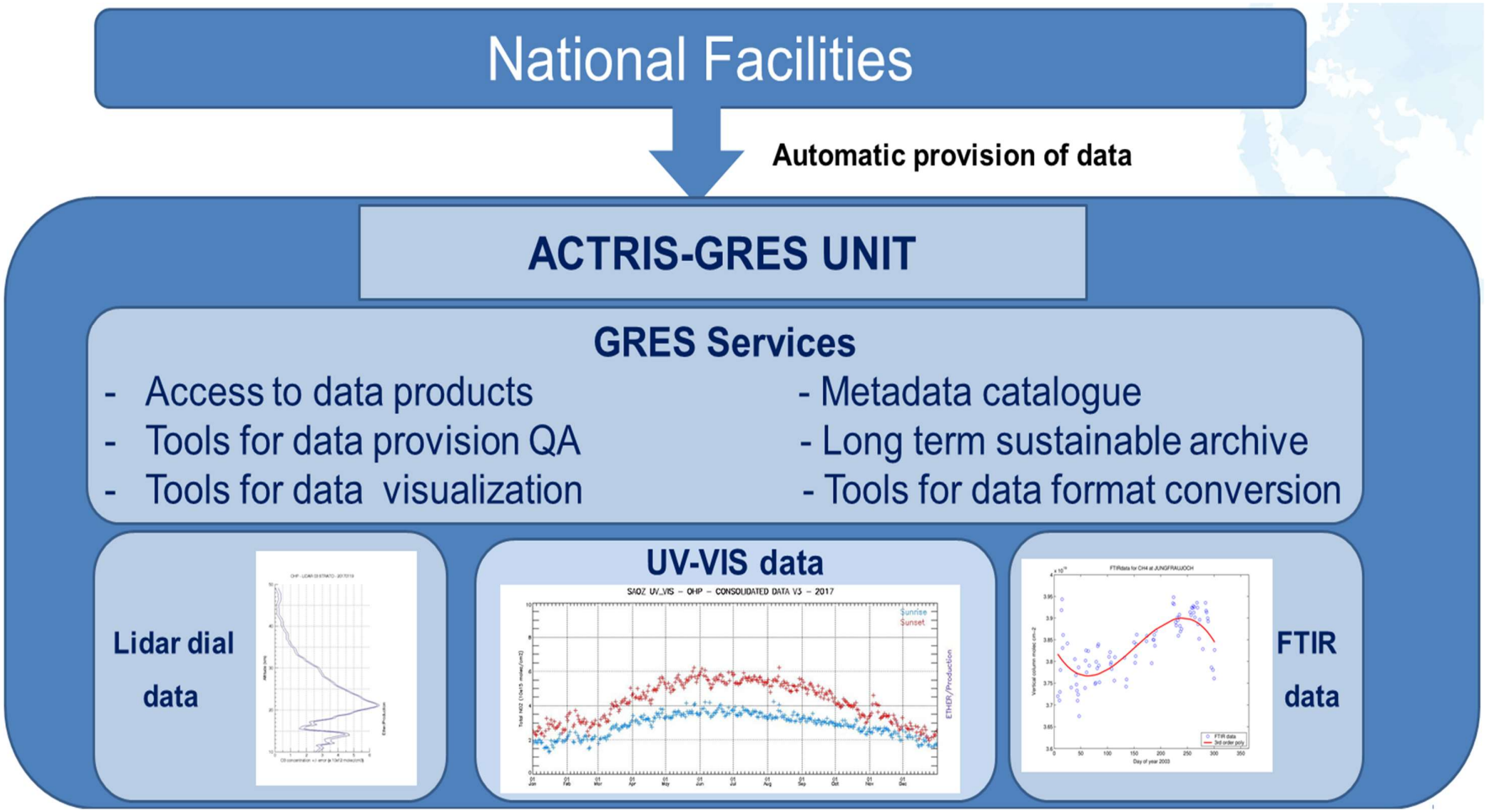


Figure 11: GRES data products and services

4.2.6 ASC dataflow and data management

Data provided in ASC unit are L2 and L3 data produced These data are produced from L0 and L1 data processing performed at NFs level (see Figure 12). These Datadata have to be provided by NFs in a standard formats and to be completed with rich metadata (see section 3.5). NFs are also in charge of providing tools to facilitate the generation and the handling of the data. The ASC unit is in charge of i) providing a free and open access to data and tools developed by NFs through user-friendly web interfaces, ii) developing data visualization tools, iii) developing tools to ensure the quality and the completeness of the data provision process, iv) creating and maintaining the metadata catalogue, and finally v) assuring long-term archiving of L2 and L3 data. Jointly with NFs and TCs, it also contributes to the elaboration of the data workflow.

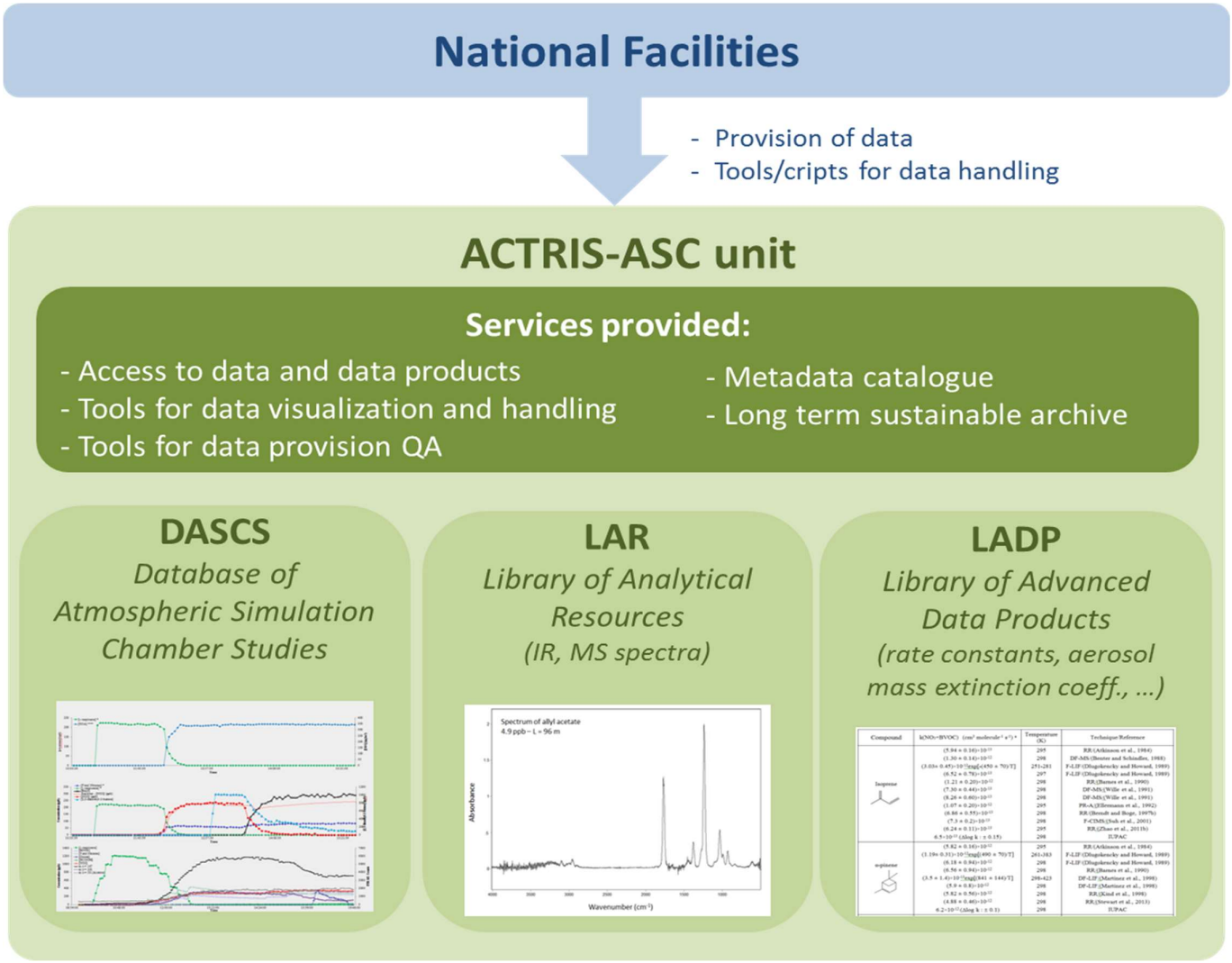


Figure 12: current overview of ASC unit

4.3 Findable: Making data findable, including provisions for metadata [FAIR data]

4.3.1 ACTRIS variable names and implementation of vocabulary

Generally, ACTRIS data set names aims to be compliant with CF (Climate and Forecast) metadata conventions. In the case where no standard CF names are defined, an application will be sent to establish these. The names used are in appendix 1. Currently there is no search model used by the ACCESS unit (ACTRIS Data Centre web interface). Still search keywords are implemented to varying degrees on the individual data center unit level (e.g. search keywords are used for the EBAS ISO19115 records). The ACTRIS data center will in the future use a controlled set of vocabularies for search keywords like Global Change Master Directory (GCMD) or similar, and semantic search will be implemented to facilitate use of ACTRIS variable across scientific disciplines and domains.

ASC unit has developed a user-friendly web interface which includes searching tools based on the metadata catalogue for the three pillars, DASCS, LAR and LADP. Relevant searching criteria have been defined for each pillar.

Standard vocabulary might not always be used, but in all cases they should be mapped to standard vocabulary if existing by the DC ACCESS unit.

Data center unit	Vocabulary name	Comment
In Situ	IUPAC, CF-1.7	
ARES	CF-1.7	
CLU	CF-1.7	
ACCESS	Defined by primary repository	
ASC	CF-1.7	
GRES	CF-1.7	

Table 16: List of vocabularies

4.3.2 Metadata standards and meta data services

ACTRIS will harvest metadata from a large range of observations employing methodologies provided by multiple data centre units covering different types of data both in terms of size, time coverage and metadata. The ACCESS unit aims at providing discovery metadata in a common format for all ACTRIS level 2 data, using a common standard that is WIS compliant such as ISO19139 or ISO19115. A decision about the standard is not taken, and under consideration. In any case, exceptions may occur in cases where the selected metadata standards do not meet the need to describe the data. The present situation is shown in Table 18.

Future efforts will further develop the system shown in Figure 7 and make it possible for the ACCESS unit to harvest all metadata from the different data centre units and collect this in a central ACTRIS metadata catalog and provide this through a commonly used protocol for Metadata harvesting like OAI-PMH or similar. A decision about the standard is not taken, and under consideration. The present situation is shown in Table 18. ACTRIS data should be described with rich metadata. Currently metadata services are offered on data centre unit level, but the aim is to offer all ACTRIS level 2 data through a centralized metadata service.

There might be instances where standards do not cover the need for describing the data at the data centre unit. In this case, ACTRIS Data Centre will still try to provide metadata in a way that is similar to the agreed formats and standards and at the same time push for an extension of the specified standard.

ACTRIS aiming at following the INSPIRE directive for metadata formatting. Present standard(s) for metadata is at the ACCESS unit level. A decision is needed if data centre units should provide metadata according to specific standards, as well as providing metadata from the ACTRIS DC to the ENVRI cluster, EOSC etc.

Tables below show the status by July 2019.

Data centre unit	**metadata service**	**end-point**	**standard**
In Situ	OAI-PMH	https://ebas-oai-pmh.nilu.no/oai/provider?verb=ListIdentifiers&metadataPrefix=iso19115	ISO 19115-2, CF-1.7,ACDD
ARES	ISO via Thredds server, JSON via REST API, HTTP via Apache Server	https://login.earlinet.org:8443/thredds/catalog/earlinedbscan/catalog.html , https://data.earlinet.org/api/services/ , https://data.earlinet.org/	ISO 19115-2 , ECMA262-3, CF-1.7, NCML, RFC2616
CLU	JSON via REST API	http://devcloudnet.fmi.fi/api/	To be decided
ACCESS	To be decided	None	To be decided
ASC	CSW, geonetwork	http://catalogue2.sedoo.fr/geonetwork/srv (implementation on going)	ISO 19139
GRES	CSW, geonetwork	http://catalogue2.sedoo.fr/geonetwork/srv (implementation on going)	ISO 19139

Table 17: List of metadata standards and services implemented by July 2019

ACTRIS metadata should be registered or indexed in relevant metadata catalogues

Metadata catalogs	Description	ACTRIS DC unit indexed
GISC Offenbach		
NextGEOSS		
WIGOS	NaN	None
Copernicus	Defined by primary repository	None
re3data	To be defined	None
EOSC	To be decided	None

Table 18: ACTRIS metadata registered or indexed in relevant metadata catalogs.

4.3.3 Traceability of ACTRIS data

The term measurement traceability is used to refer to an unbroken chain of comparisons relating an instrument's measurements to a known standard, time, processing, software etc. Calibration to a traceable standard can be used to determine an instrument's bias, precision, and accuracy. The ability to trace a measurements back to its origin is important for several reasons; It increase the quality by facilitating back-out or reprocess bad data, and conversely, it allows reward and boost good data sources and processing techniques. This is also to ensure that proper attribution is given to data originators adequately reflecting their contributions through the data production chain.

ACTRIS works towards establishing traceability for all variables using persistent identifiers (PIDs). This work is in development, and need close interaction with the topical centres as well as National Facilities. Currently ACTRIS is using digital object identifiers (DOIs) for some level 3 datasets though the Data Cite Metadata Store API, and more will be implemented.

Currently, ARES unit assigns two different types of local persistent identifier (PID): * **Data Processing PIDs.** These PIDs identify unequivocally the characteristics of the instrument (including all its subparts) used to collect the Level 0 data. In particular, to each submitted Level 0 product it is assigned an alphanumeric ID which allows to retrieve all the details about the instrument configuration used to perform the measurement as well as the data processing configuration used to compute the corresponding Level 1 data products.
* **Dataset PIDs.** An internal PID generation system based on an alphanumerical "prefix"/"suffix" pattern identifies univocally each dataset downloaded through the ARES interfaces.

ACTRIS data will be assigned PIDs that are available through the metadata, the table below show the status by July 2019.

Data centre unit	PID service	Description	Standard
In Situ	To be decided		
ARES	Internal	Internal generation system of alphanumerical PIDs for data processing, Internal generation system of alphanumerical PIDs based on Handle System's pattern for datasets	RFCs 3650, RFCs 3651, RFCs 3652
CLU	To be decided	None	To be decided
ACCESS	Defined by primary repository	None	To be decided
ASC	To be decided	None	To be decided
GRES	To be decided	None	To be decided

Table 19: ACTRIS PID handlers

4.3.4: Version control of ACTRIS (meta)data

The ACTRIS DC aims at providing clear versioning of its data and metadata. Due to the decentralized nature of the Data Centre, this varies between the different data centre units, and implementation will be done on unit level.

As a guiding principle, all data submitted to ACTRIS passing quality assurance should be uniquely identified. In case of updates, a ID-number is generated, and previous data versions should be identifiable and kept available upon request while the latest version is served through the ACTRIS Data Centre.

A Versioning System has been implemented at ARES directly in the RDBMS by using DML (Data Manipulation Language) triggers. A new version of a file is produced when a user tries to modify data through a DML event. New versions will be centrally produced if new QC procedures and new processing features are released. Additionally new versions of the files will be allowed and centrally handled for fixing file bugs in particular for legacy data.

4.4 Accessible: Making data openly accessible [FAIR data]

The purpose of the data collection and generation of data products in ACTRIS is to provide open access to aerosol, cloud and trace gas in situ and remote sensing measurements of high quality (see section 1).

A guiding principle is that all ACTRIS data should be readable for both humans and machines using protocols that offer no limitations to access. ACTRIS Data Centre is organized as a distributed network of centralized repositories (see Figure 6). The ACTRIS data will be offered through the ACTRIS Data Centre portal, a web portal that allows the user to search, analyses and download data produced within ACTRIS (see Figure 7). Access to data and metadata will also be made possible by machine-to-machine interaction, enabling harvesting of metadata from the ACTRIS metadata catalog. Currently, machine-to-machine access of ACTRIS data varies between the different data units and their associated repositories.

There might also be data available through the ACTRIS Data Centre that is not directly ACTRIS data, but used in the production and interpretation of ACTRIS data.

4.4.1 ACTRIS data access and access protocols

General guidelines for access to ACTRIS data and services are available in the current ACTRIS access and service policy. Conditions of use should be indicated in section 3.4, and is covered by the attached license, unless stated otherwise.

The access protocol will be clearly described in the metadata. If direct access is limited due to size of data or sensitive data, contact information on institutional and/or personal level will be included. The data format and access protocol must be available as machine readable metadata.

Currently all data centre units maintain the access to the data, either directly through the unit specific repository or through the ACTRIS data portal.

The table shows the data access protocols.

DC unit	data format	Repository URL	Protocol	Authentication and authorization needed
In Situ	netCDF, NasaAmes, CSV, XML	http://ebas.nilu.no/	HTTP	No
ARES	netCDF	http://data.earlinet.org/	HTTP	Yes
CLU	netCDF	http://cloudnet.fmi.fi	HTTP	No
ACCESS (data portal)	Defined by primary repository	http://actris.nilu.no/	HTTP	For some data
ASC	netCDF (data conversion by 2020)	https://data.eurochamp.org/	HTTP	For some data
GRES	netCDF (data conversion by 2021)	https://gres.aeris-data.fr	FTP	No
ACCESS	varies	http://actris.nilu.no/Content/?pageid=226809f7a0ac49538914eeafb4448afa	FTP	No

Table 20: Data formats and access protocols

For In-Situ, CLU, GRES and and ASC unit, all data, metadata, tools and documentation are provided with free and fully open access to all users without authentication with username and password.

A Sign-On authentication system has been implemented at ARES unit. It is based on CAS (Central Authentication Service) project which implements natively multiple authentication protocols (CAS, SAML, OAuth, OpenID) and provides both authentication via username and password and via Google credentials. In order to gain access to ARES products (apart from Quicklooks, simple plots of Level 1 data) a user authentication (free and open to all users) is needed. Such authentication process has been implemented with the only purpose to allow feedback to the end user in case of software or data products updates.

In general, for all data that requires username and password, a Single-Sign-On service will be implemented, and used by all Data Centre units.

In all cases where access is restricted, information on how to access the data should be available through the metadata, in order to facilitate machine to machine interaction.

If specific software tools are need to access the data, documentation about the software and how to access it should be included, preferably in the metadata. Furthermore, ACTRIS digital tools (software etc.) will be available through open access repositories like GitHub. *A open source licence for software should be encouraged and applied when possible. All software related to ACTRIS data should aim at following the practice of open access if possible. For software related to access of level 2 data, the ACCESS unit is responsible together with the data centre units. To be discussed: For level 0 and 1, the topical centres and/or data centre unit are responsible for providing access to software related to ACTRIS level 0 and level 1.*

There are valuable and contributing networks to ACTRIS e.g. EMEP, GAW, EARLINET, and level 3 products covering bridge to external data bases and use of these data in combined products. The implementation and strategic and technical contributions of this is under development.

4.5 Interoperable: Making data interoperable [FAIR data]

As a guiding principle, ACTRIS should make sure that metadata and data use a formal, accessible, shared and broadly applicable language for knowledge representation in order to facilitate interoperability. Still, work remains to see if a common solution could be agreed upon. The intricate nature of the data and metadata might require the use of different solutions to suit the needs of different data centre units. As mention in section 4.3 metadata standard and vocabularies commonly used in the atmospheric domain should be applied, unless the common solutions do not address the specific need for the DC unit. Implementation of new standards for data and metadata used in the context of ACTRIS should be discussed by all the DC units. The aim should be to harmonize data and metadata as much as possible, both in terms of technical aspects related to implementation, but also making it easier for the end user to make use of the data.

By many of the DC units the Thredds Data Server (TDS) is used for serving data and metadata in an automated way as netCDF files through the OPeNDAP protocol (this apporach is implemented by In-Situ, ARES, ASC, GRES).

In addition to this, ARES provides a REST API for machine-to-machine interaction. The API serves metadata (info, provenance, versions, quality controls, etc.) in JSON format and data (specific files or datasets previously generated) in NetCDF format.

CLU is currently working on a RESTful API with similar services as ARES in development.

4.6 Reuseable: Increase data re-use [FAIR data]

The guiding principle is free and open access to ACTRIS data and ACTRIS data products, and the ACTRIS DC will facilitate data re-use by providing free and open access to ACTRIS data following the ACTRIS access and service policy and the open research data initiative of the European Commission.

As a result, the ACTRIS DC will implement one or multiple licenses for all ACTRIS level 2 data and NRT data that is available through the ACTRIS metadata catalog. Furthermore, the ACTRIS DC might also consider issuing a license on the use of metadata, in order to acknowledge ACTRIS when large amounts of metadata is harvested by third party application/services. ACTRIS aims to implement a license from the time ACTRIS becomes an ERIC (probably end of 2020 or early 2021). Until ACTRIS has decided upon and implemented one or more licenses, the current ACTRIS data policy will apply.

Several features have been implemented by In-Situ, ARES, CLU (or more) units to ensure reusability and traceability, in particular traceable data flow and version control of data products, see section 4.3.

In order to increase the reusability of data in ASC unit, these data are completed with rich metadata which are in open access from the website. These metadata provide detailed technical description of the chambers (size, volume, walls, irradiation system ...), experimental protocols used for the generation of the data, and an "auxiliary mechanism" which provides the chamber-dependent parameters affecting the observations. This last one is very useful for modelers who aim at simulating experiments performed in simulation chambers.

As regards ARES unit, all the characteristics of the LIDAR instrument configuration (laser, telescope, acquisition and detection system, processing configuration, ...) are reported as metadata in each ARES data product.

Availability of data can vary between the different data centre units. As an example, in situ data is typically submitted on an annual basis, and are therefore available the subsequent year, but other data centre units may provide NRT delivery of data; in addition, there may be campaign-based data. ACTRIS legacy data should be kept available for users, but may have a different data policy to the current ACTRIS data policy. If this is the case, this information should be available in the metadata.

Data centre unit	Data licence	Comment
In Situ	To be decided	
ARES	To be decided	
CLU	To be decided	
ACCESS	Will be defined by primary repository	
ASC	To be decided	
GRES	To be decided	

Table 21: Data licences

Responsible data centre unit	Software licence	Software link
In Situ	None	EBAS IO
ARES	None	Single Calculus Chain
CLU	MIT	CloudnetPy
ASC	None	None
GRES	None	None

Table 22: Software licences

5. Allocation of resources

ACTRIS Data Center is a distributed data center with scientific and data expert contributions as well as funding contributions from many institutions and sources. All host countries are contributing significantly to the operation and implementation through both national and international projects, in addition to considerable support from the institutions involved. Furthermore, there is large ongoing activity of making ACTRIS data FAIR, in particular this is the core of the work within the H2020 project ENVRI-FAIR. The ACTRIS DC budget in ENVRI-FAIR is ca 890 kEuro which leaves this project as one if the main funder of making ACTRIS data FAIR.

Details on costs of the various units is available upon request, and a part of the work within ACTRIS-PPP and ACTRIS-IMP (starting 1 January 2020).

6. Data security

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 in situ 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 the University of Lille, Villeneuve d'Ascq, France, and the cloud profile data in the Cloudnet DB at the Finnish Meteorological Institute in Helsinki, Finland.

6.1 Archiving and preservation of In-Situ 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 the levels 0-1-2 of data.

The complete data system is backed up regularly. This includes incremental back up of the data base 4 times per week, and two weekly back ups of the full data base to a server in a neighbour 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 a web application which checks files for syntactic and semantic validity before uploading. As an alternative submission method, especially for regular submission or submission of many files at once, ftp upload is possible.

A dedicated 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.

Ca 385 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. For more detail, see Table 2 and Table 3.

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. A long-term strategy for providing access to all ACTRIS data and other related services are in progress through the establishment of ACTRIS as a RI. ACTRIS is on the ESFRI (European Strategy Forum on Research Infrastructures) roadmap for Research Infrastructures, and a preparatory phase project is ongoing.

6.2 Archiving and preservation of ARES data

The ARES infrastructure is composed by seven virtual servers and two different SAN (Storage Area Network).

One server hosts a PostgreSQL database, a second and a third one are used to interface the data originators and endusers respectively. ARES data products are safely stored on a primary SAN. A full daily back up is made automatically and it is stored on a second backup SAN.

Another server is responsible for the provisioning of the whole database through THREDDS (Thematic Real-time Environmental Distributed Data Services). On the same server a CAS (Central Authentication Service) is configured to authenticate all ARES users centrally.

The current size of the PostgresSQL EARLINET database is about 1GB. The total amount of data submitted (NetCDF EARLINET files) is about 1.3 GB. An estimation of the growing rate of the database is 100-200MB/year. However a significant growth in number of files to be collected is expected because of: the use of the SCC (Single Calculus Chain) for the data submission, the inclusion of new products (pre-processed data, NRT optical properties, profiles, aerosol layers properties and multiwavelength datasets), increases of the number of aerosol remote sensing NF and increase of NF operating 24/7. We estimate that during the Implementation Phase the ACTRIS aerosol profile database could grow at a rate of about 300 GB per year.

The SCC is part of the ARES infrastructure and it is the standard EARLINET tool for the automatic analysis of lidar data. Three additional servers are needed to provide this further service: a calculus server where all the SCC calculus modules are installed and ran, a MySQL database where all the analysis metadata are stored in a fully traceable way and finally a web interface allowing the users to access to the SCC.

The ARES infrastructure 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.

6.3 Archiving and preservation of CLU data

The CLU database consists of a file archive connected to a relational metadata database, due to the nature of the typical use-case and data volume. The infrastructure comprises a webserver, an FTP server for incoming data streams, web and rsync server for outgoing data streams, processing servers, with data storage distributed across a series of virtual filesystems including incremental backups. Due to the data volume, most sites also hold an archive of their own Level 0 and Level 1 data, effectively acting as a second distributed database and additional backup.

The current size of the database is about 25 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 CLU infrastructure is maintained by FMI with long-term commitment for archiving and preservation. Publication of QA datasets will aid dataset preservation.

6.4 Archiving and preservation of GRES data

For the GRES unit, data are stored on disk on a server in Paris. As new data are provided once a year, a full backup is made yearly and stored on tape. We plan to have soon a second copy on tape in Palaiseau, France. The distance between both sites will be about 20km.

The GRES infrastructure is maintained by AERIS with long-term commitment for archiving and preservation.

6.5 Archiving and preservation of ASC data

Since the Eurochamp H2020 project, data from simulation chambers are managed by AERIS. It consists of a file archive connected to a MondoDB metadata database. Data files are stored on disk on a server located in Toulouse, France. A full daily backup is made automatically and stored on another site (in Tarbes, France). The distance between the database and the backup site is about 120km. We plan to have soon a copy on tape in Paris.

The ASC infrastructure is maintained by AERIS with long-term commitment for archiving and preservation.

6.6 Archiving and preservation of ACCESS data

The ACCESS unit is providing access to ACTRIS data through the ACTRIS data portal using the ASP.NET (Web Forms) Framework 4.5 and Internet Information Services (ISS) web-server. The metadata is harvested from each individual data center unit, currently In Situ (EBAS), ARES (EARLINET), CLU (CLOUDNET) and GRES (NDACC), using custom harvesting routines triggered by cronjobs on an Ubuntu server running custom scripts written in Perl/Python. The metadata itself is stored on a Oracle database server, version 11.2.0.4. Versioning and revision control is managed using subversion (SVN).

7. Ethical aspects

ACTRIS Ethical Guidelines describes the main principles of ethics to be applied within ACTRIS activities. These guidlines shall be acknowledged and followed by all persons affiliated to ACTRIS and should be supported by all participating institutions, including the Data Centre. These guidelines do not exclude other ethical issues (e.g. related to professional and scientific responsibility, governance, social and environmental responsibility and law abiding) brought up by the ACTRIS ERIC and its contractual ACTRIS partners, or by the Ethical Advisory Board of the ACTRIS ERIC. In general, everyone in ACTRIS should work in a socially ethical way keeping the integrity and fairness, and maintaining high level of trust and respect among the people working in ACTRIS and with the users and other stakeholders. One should always take into account that the mission of ACTRIS is to provide effective access for a wide user community to its resources and services, in order to facilitate high-quality Earth system research, to increase the excellence in Earth system research, and to provide information and knowledge on developing sustainable solutions to societal challenges.

8. Appendix

Appendix 1: List of ACTRIS variables from observational platforms and associated recommended methodology

List of ACTRIS variables and recommended methodology

Additional information: 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 utilizing these products will also be prepared, and standardized pre processed lidar data and NRT optical property profiles will be available.

Appendix 2: List of ACTRIS level 3 data products

List of ACTRIS level 3 data products

Appendix 3: ACTRIS In situ data centre unit (In-Situ) data life cycle

A3.1 Data Life Cycle Description

More tables to be added regarding the workflow, currently this is an example draft.

ACTRIS In situ data centre unit workflow diagram

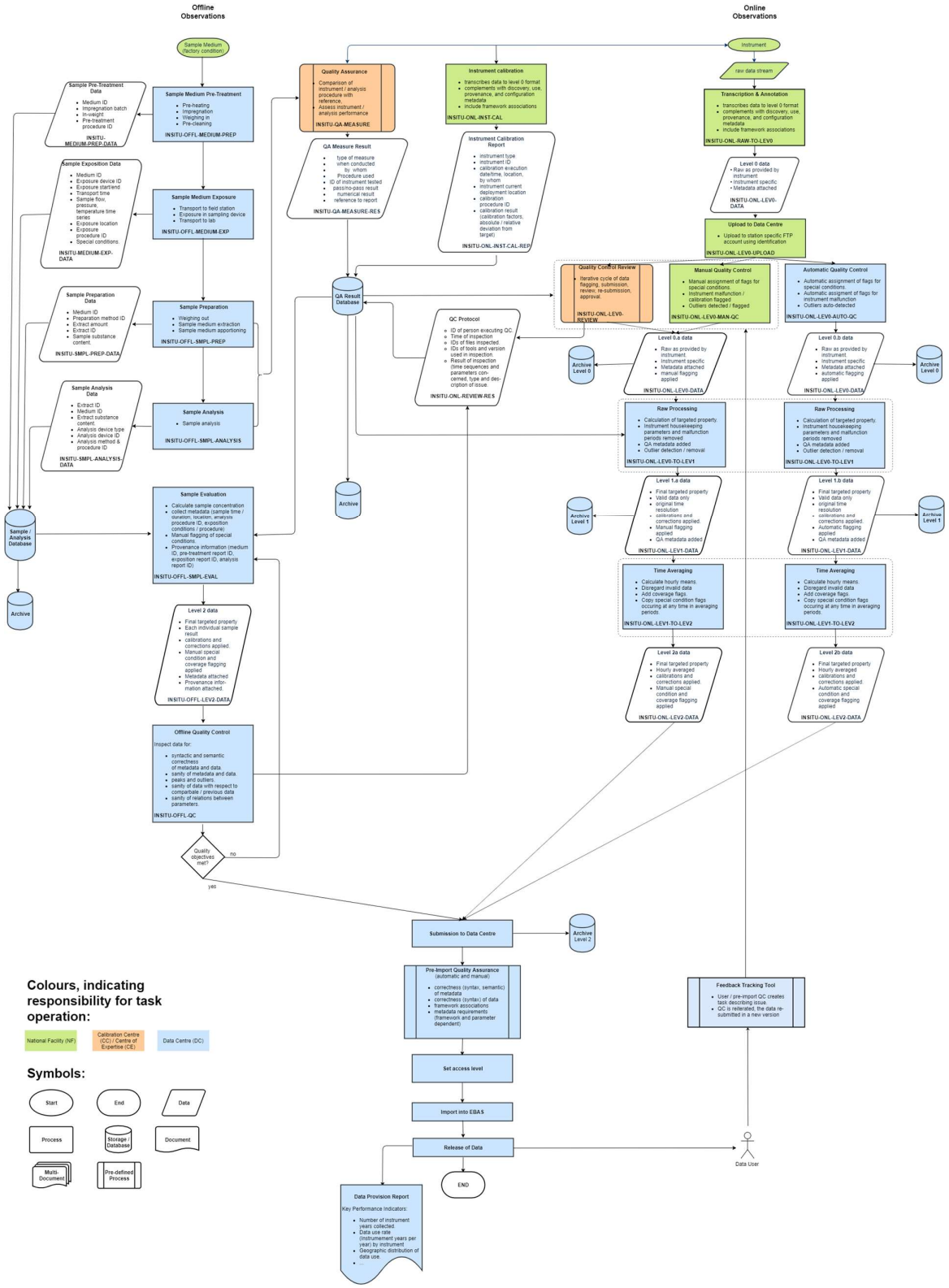
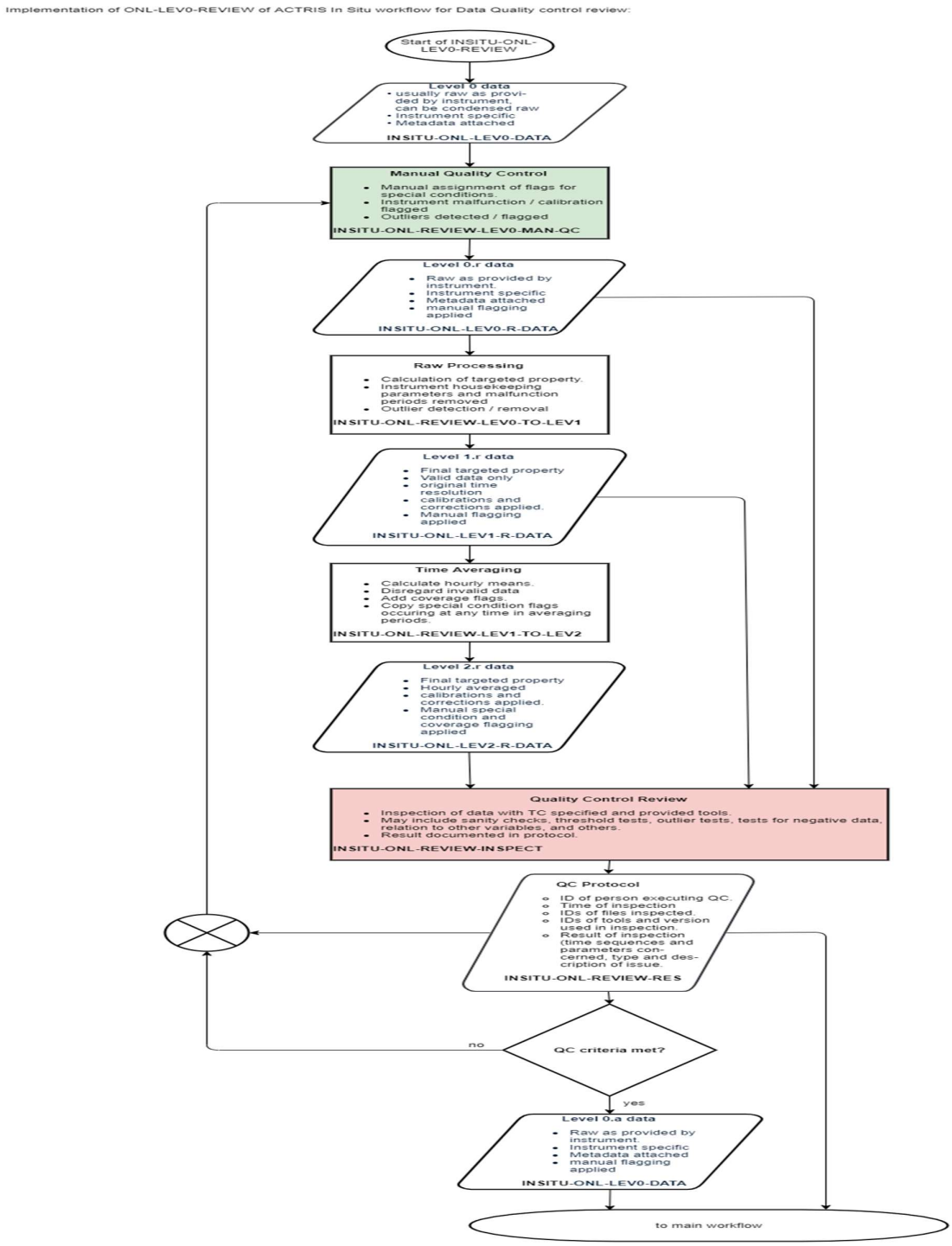


Figure 13: ACTRIS In Situ DC unit data workflow, describing the interaction between NFs, TCs, and DC In Situ in data production.



ACTRIS In situ data review workflow

Figure 14: ACTRIS In Situ DC unit data review workflow, a sub-workflow to the In Situ main data production workflow.

A3.2 Workflow Implementation Tables, by instrument type

In this version of the DMP, this Annex focuses on the distribution of responsibilities for workflow processing tasks, and a short specification of these. A specification of metadata and data items contained in data products and pre-products will follow in a later version.

For each workflow task, responsibilities include the following roles: * **Specification:** defining what is done in the tas. Includes step-by-step description, with formulas (algorithm description document (ADD), also called SOP, to be provided later). * **Implementation:** taking the ADD, and turning it into software. * **Operation:** running the software on a daily basis. Includes documentation of provenance while executing software. * **Application:** applying the software. Usually automatic, needs to be specified for manual tasks involving humans.

The task specifications and the distribution of roles between NFs, TCs, and DC are stated in tables linked below.

A3.2.1 Aerosol observations

A3.2.1.1 Integrating nephelometer

Nephelometer workflow implementation tables

A3.2.1.2 Filter Absorption Photometer

Filter Absorption Photometer workflow implementation tables

A3.2.1.3 Mobility Particle Size Spectrometer

Mobility Particle Size Spectrometer workflow implementation tables

A3.2.1.4 Condensation Particle Counter

Condensation Particle Counter workflow implementation tables

A3.2.1.5 Cloud Condensation Nucleus Counter

Cloud Condensation Nucleus Counter workflow implementation tables

A3.2.1.6 Aerodynamic / Optical Particle Size Spectrometer

Aerodynamic / Optical Particle Size Spectrometer workflow implementation tables

A3.2.1.7 Aerosol Chemical Speciation Monitor

Aerosol Chemical Speciation Monitor workflow implementation tables

A3.2.1.8 Proton-induced X-ray Emission

Proton-induced X-ray Emission workflow implementation tables

A3.2.1.9 Organic Tracers

Organic Tracers workflow implementation tables

A3.2.1.10 Organic Carbon / Elemental Carbon

Organic Carbon / Elemental Carbon workflow implementation tables

A3.2.1.11 Scanning Particle Size Magnifier / (Neutral) Air Ion Spectrometer / Nano Mobility Particle Size Spectrometer

Scanning Particle Size Magnifier / (Neutral) Air Ion Spectrometer / Nano Mobility Particle Size Spectrometer workflow implementation tables

A3.2.1.12 Particle Size Magnifier

Particle Size Magnifier workflow implementation tables

A3.2.2 Cloud observations

A3.2.2.1 Integrating Cloud Probe

Integrating Cloud Probe workflow implementation tables

A3.2.2.2 Ice Nucleus Counter

Ice Nucleus Counter workflow implementation tables

A3.2.2.3 Cloud Imaging Probe

Cloud Imaging Probe workflow implementation tables

A3.2.2.4 Cloud Droplet Probe

Cloud Droplet Probe workflow implementation tables

A3.2.2.5 Cloud Water Collector

Cloud Water Collector workflow implementation tables

A3.2.2.6 Cloud Aerosol Particle Sampler

Cloud Aerosol Particle Sampler workflow implementation tables

A3.2.3 Trace Gas Observations

A3.2.3.1 Volatile Organic Compounds

Volatile Organic Compounds workflow implementation tables

A3.2.3.2 Nitrogen Oxides

Nitrogen Oxides workflow implementation tables

A3.2.3.3 Condensable Vapours

Condensable Vapours workflow implementation tables

A3.2.3.4 Ozone

Filter Absorption Photometer workflow implementation tables to be added.

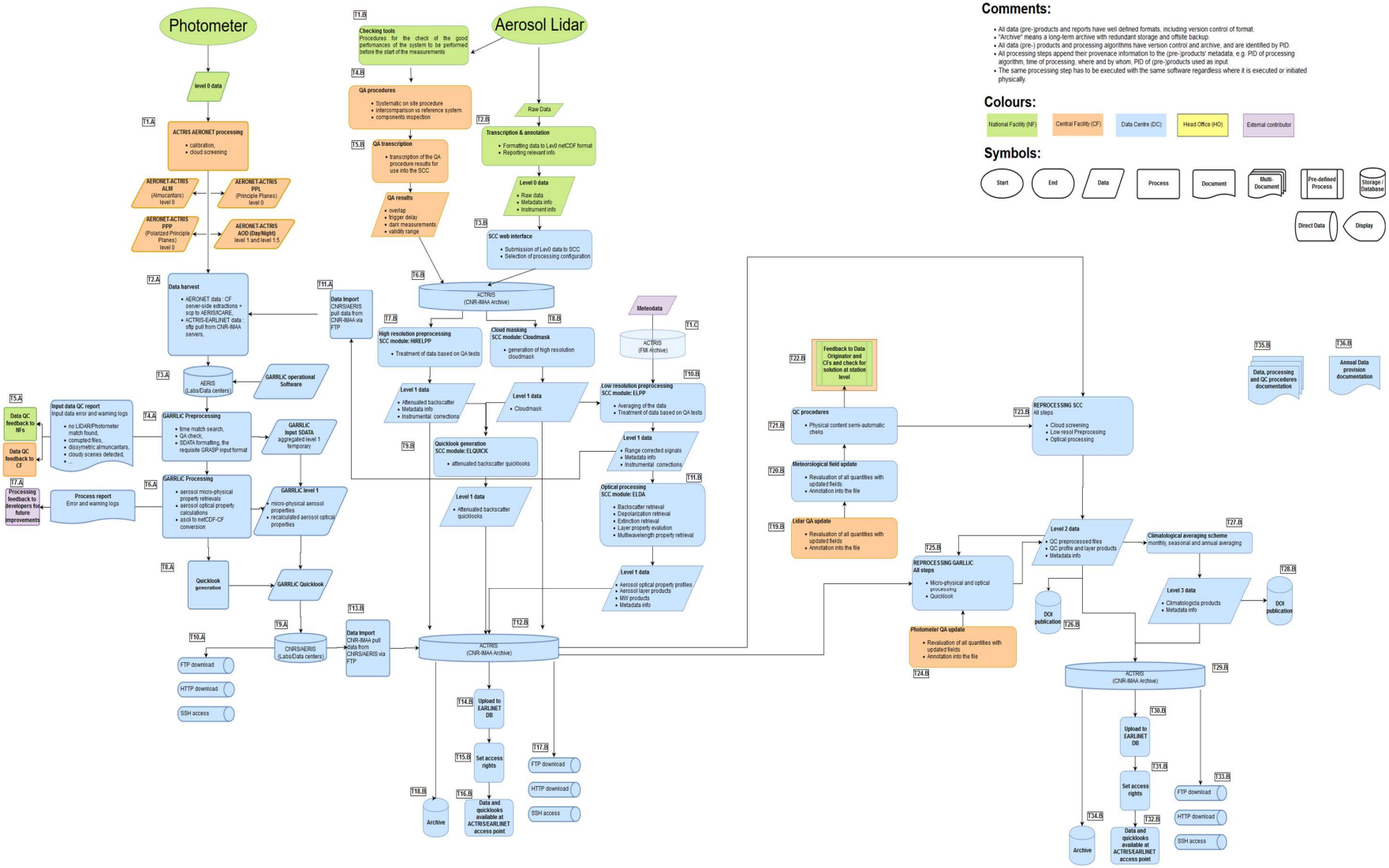
A3.2.3.5 Meteorological Base Parameters

Meteorological Base Parameters workflow implementation tables to be added.

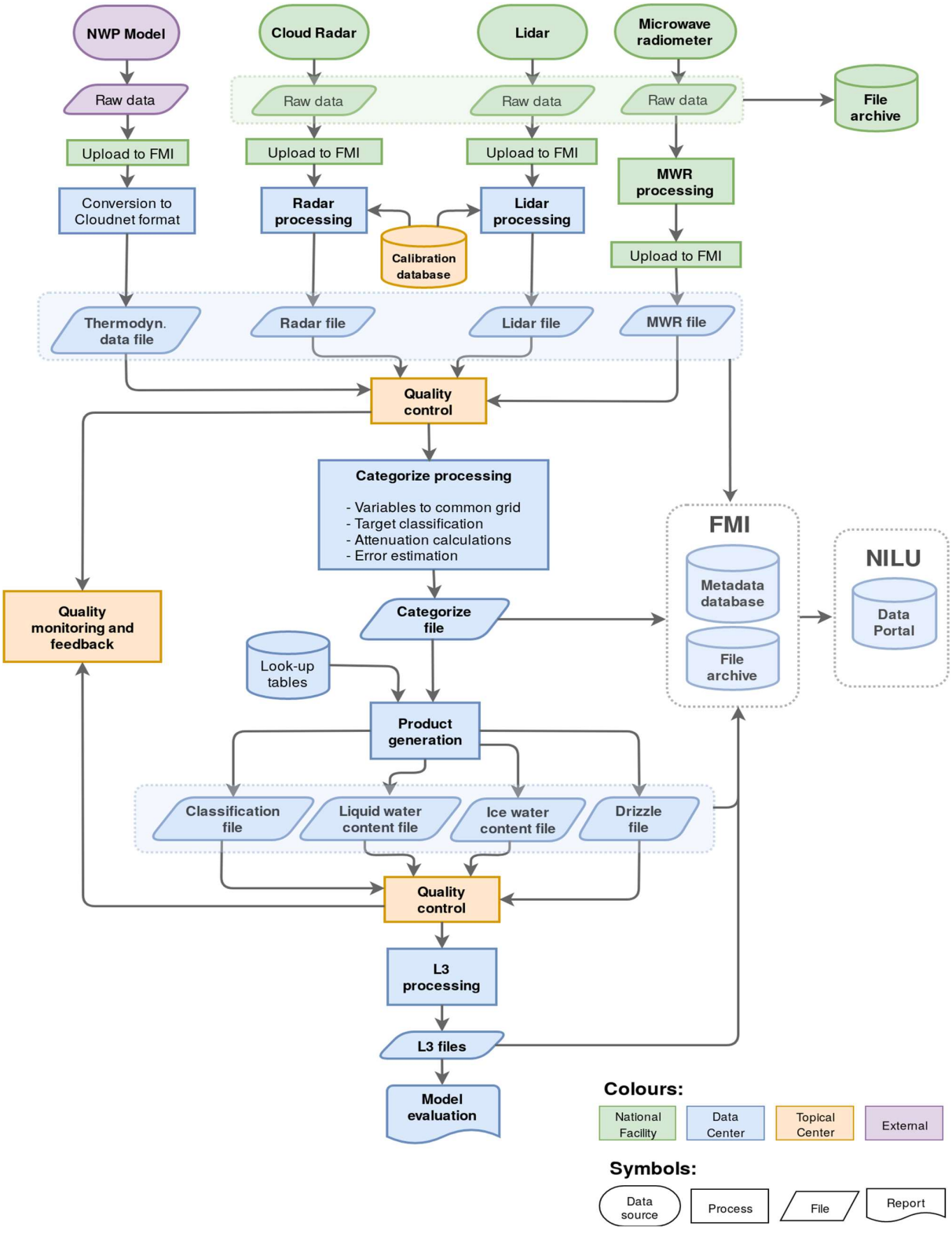
Appendix 4: ACTRIS Aerosol remote sensing data centre unit (ARES) data life cycle and workflow diagram

Link to seperate document describing the workflow in more detail.

To be added by ARES


Photometer
level 0 data
T1.A
ACTRIS AERONET processing
• calibration,
• cloud screening
AERONET-ACTRIS ALM (Almucantars) level 0
AERONET-ACTRIS PPL (Principle Planes) level 0
AERONET-ACTRIS PPP (Polarized Principle Planes) level 0
AERONET-ACTRIS AOD (Day/Night) level 1 and level 1.5
T2.A
Data harvest
• AERONET data : CF server-side extractions + scp to AERIS/ICARE;
• ACTRIS-EARLINET data : sftp pull from CNR-IMAA servers;
T3.A
AERIS (Labs/Data centers)
GARRLIC operational Software
T5.A
Data QC feedback to NFs
Data QC feedback to CF
Input data QC report
Input data error and warning logs
• no LIDAR/Photometer match found,
• corrupted files,
• dissymetric almucantars,
• cloudy scenes detected,
• ...
T4.A
GARRLIC Preprocessing
• time match search,
• QA check,
• SDATA formatting, the requisite GRASP input format
GARRLIC Input SDATA
aggregated level 1 temporary
T7.A
Processing feedback to developers for future improvements
Process report
Error and warning logs
T6.A
GARRLIC Processing
• aerosol micro-physical property retrievals
• aerosol optical property calculations
• ascii to netCDF-CF conversion
GARRLIC level 1
• micro-physical aerosol properties
• recalculated aerosol optical properties
T8.A
Quicklook generation
GARRLIC Quicklook
T9.A
T10.A
CNRS/AERIS (Labs/Data centers)
FTP download
HTTP download
SSH access
T1.B
Checking tools
Procedures for the check of the good performances of the system to be performed before the start of the measurements
Aerosol Lidar
T4.B
QA procedures
• Systematic on site procedure
• intercomparison vs reference system
• components inspection
T5.B
QA transcription
• transcription of the QA procedure results for use into the SCC
QA results
• overlap
• trigger delay
• dark measurements
• validity range
T2.B
Raw Data
Transcription & annotation
• Formatting data to Lev0 netCDF format
• Reporting relevant info
Level 0 data
• Raw data
• Metadata info
• Instrument info
T3.B
SCC web interface
• Submission of Lev0 data to SCC
• Selection of processing configuration
T6.B
ACTRIS (CNR-IMAA Archive)
T11.A
Data Import
CNRS/AERIS pull data from CNR-IMAA via FTP
T7.B
High resolution preprocessing SCC module: HIRELPP
• Treatment of data based on QA tests
T8.B
Cloud masking SCC module: Cloudmask
• generation of high resolution cloudmask
Meteodata
T1.C
ACTRIS (FMI Archive)
Level 1 data
• Attenuated backscatter
• Metadata info
• Instrumental corrections
Level 1 data
• Cloudmask
T10.B
Low resolution preprocessing SCC module: ELPP
• Averaging of the data
• Treatment of data based on QA tests
T9.B
Quicklook generation SCC module: ELQUICK
• attenuated backscatter quicklooks
Level 1 data
• Range corrected signals
• Metadata info
• Instrumental corrections
T11.B
Optical processing SCC module: ELDA
• Backscatter retrieval
• Depolarization retrieval
• Extinction retrieval
• Layer property evalution
• Multiwavelength property retrieval
Level 1 data
• Attenuated backscatter quicklooks
Level 1 data
• Aerosol optical property profiles
• Aerosol layer products
• MW products
• Metadata info
T13.B
Data Import
CNR-IMAA pull data from CNRS/HERIS via FTP
T12.B
ACTRIS (CNR-IMAA Archive)
T14.B
Upload to EARLINET DB
T15.B
Set access rights
T16.B
Data and quicklooks available at ACTRIS/EARLINET access point
T17.B
FTP download
HTTP download
SSH access
T18.B
Archive
T22.B
Feedback to Data Originator and CFs and check for solution at station level
T21.B
QC procedures
• Physical content semi-automatic cheks
T20.B
Meteorological field update
• Revaluation of all quantities with updated fields
• Annotation into the file
T19.B
Lidar QA update
• Revaluation of all quantities with updated fields
• Annotation into the file
T23.B
REPROCESSING SCC
All steps
• Cloud screening
• Low resol Preprocessing
• Optical processing
T35.B
Data, processing and QC procedures documentation
T36.B
Annual Data provision documentation
T25.B
REPROCESSING GARLLIC
All steps
• Micro-physical and optical processing
• Quicklook
T24.B
Photometer QA update
• Revaluation of all quantities with updated fields
• Annotation into the file
Level 2 data
• QC preprocessed files
• QC profile and layer products
• Metadata info
T27.B
Climatological averaging scheme
(monthly, seasonal and annual averaging
T26.B
DOI publication
Level 3 data
• Climatologicla products
• Metadata info
T28.B
DOI publication
T29.B
ACTRIS (CNR-IMAA Archive)
T30.B
Upload to EARLINET DB
T31.B
Set access rights
T32.B
Data and quicklooks available at ACTRIS/EARLINET access point
T33.B
FTP download
HTTP download
SSH access
T34.B
Archive
Comments:
• All data (pre-)products and reports have well defined formats, including version control of format.
• "Archive" means a long-term archive with redundant storage and offsite backup
• All data (pre-) products and processing algorithms have version control and archive, and are identified by PID.
• All processing steps append their provenance information to the (pre-)products' metadata, e.g. PID of processing algorithm, time of processing, where and by whom, PID of (pre-)products used as input.
• The same processing step has to be executed with the same software regardless where it is executed or initiated physically.
Colours:
National Facility (NF)
Central Facility (CF)
Data Centre (DC)
Head Office (HO)
External contributor
Symbols:
Start
End
Data
Process
Document
Multi-Document
Pre-defined Process
Storage / Database
Direct Data
Display

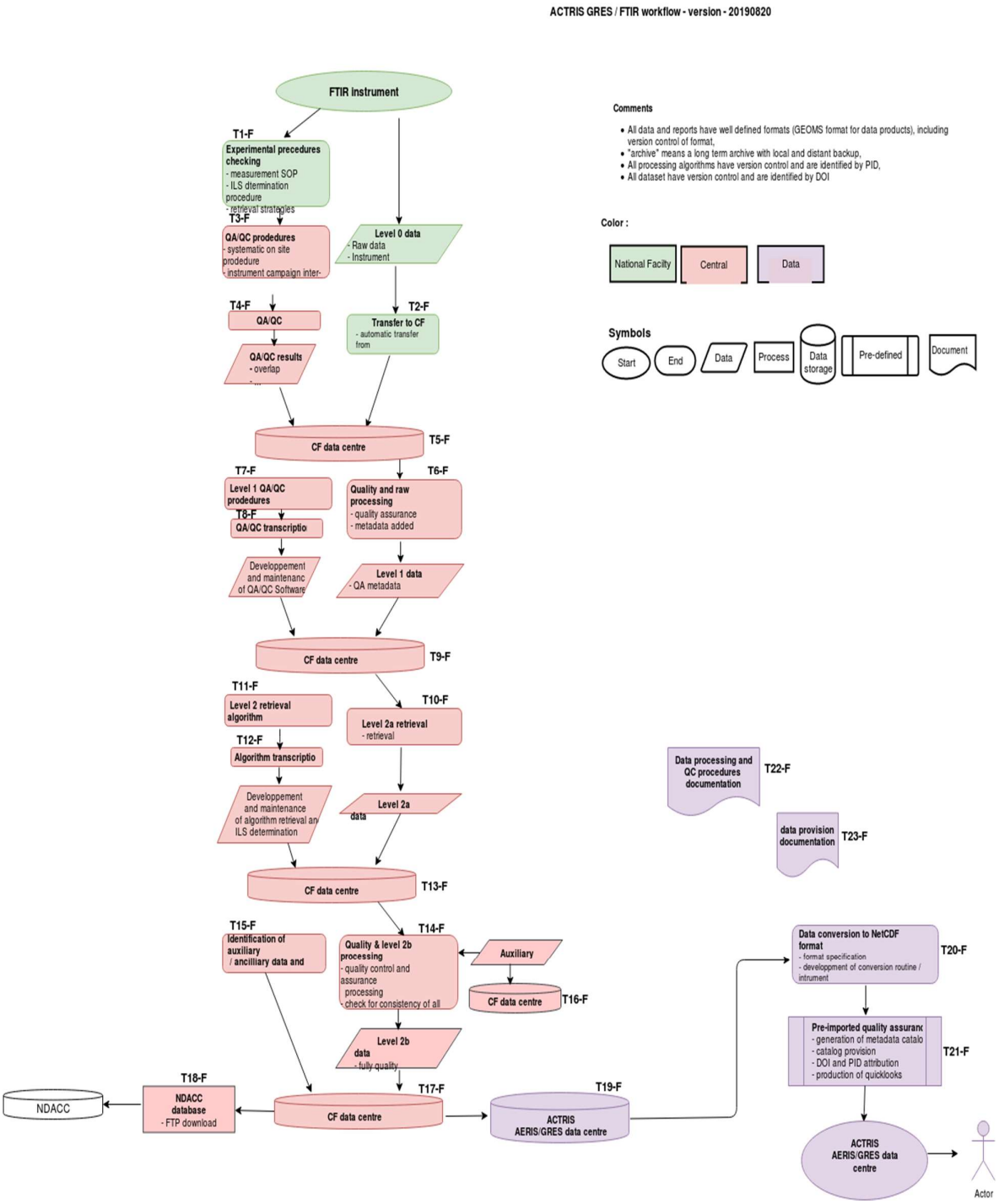
Appendix 5: ACTRIS Cloud remote sensing data centre unit (CLU) data life cycle and workflow diagram



ACTRIS Cloud remote sensing data centre unit workflow diagram

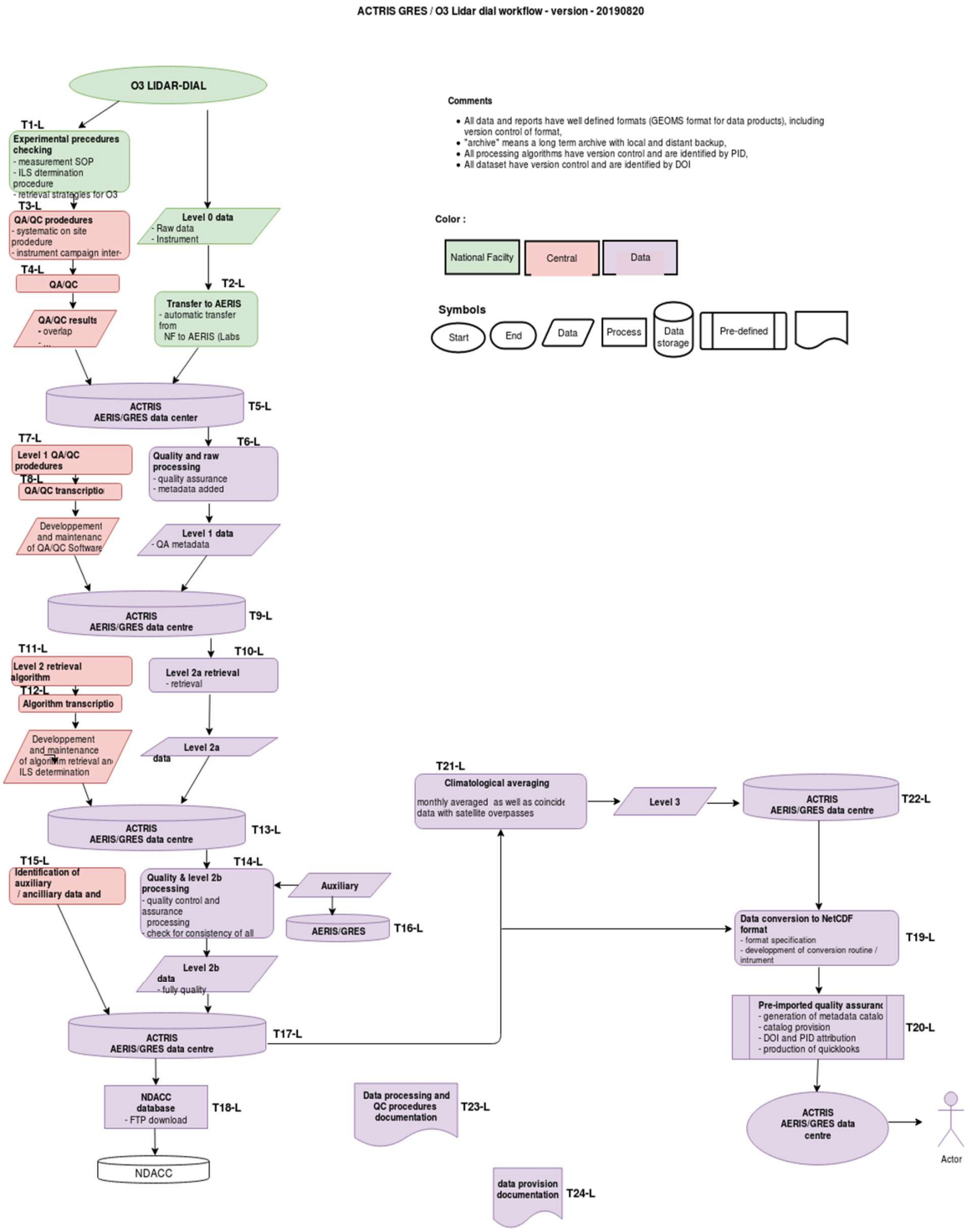
Appendix 6: ACTRIS trace gases remote sensing data centre unit (GRES) data life cycle and workflow diagram

ftir data



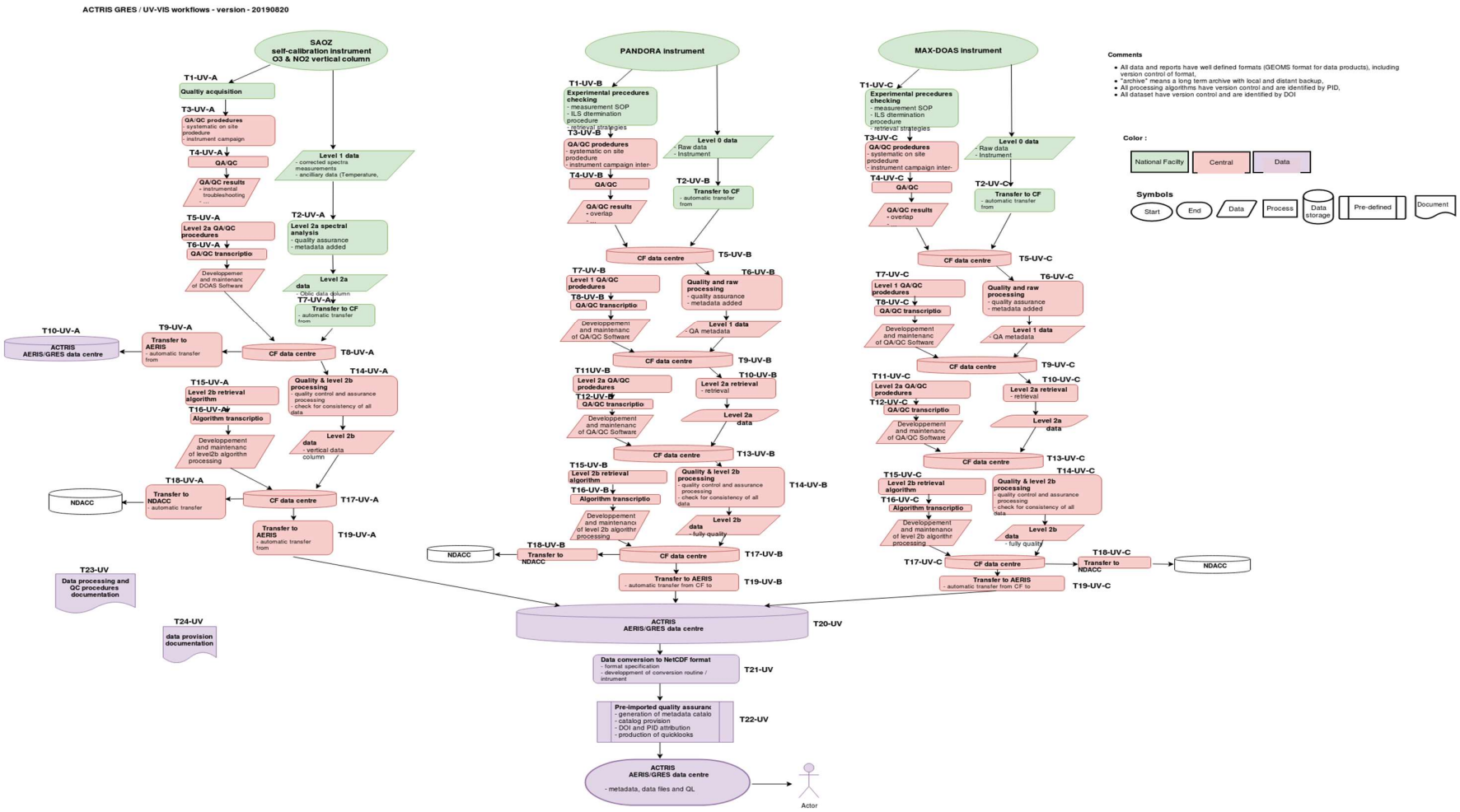
ACTRIS trace gases remote sensing data centre unit workflow diagram (ftir data)

lidar data



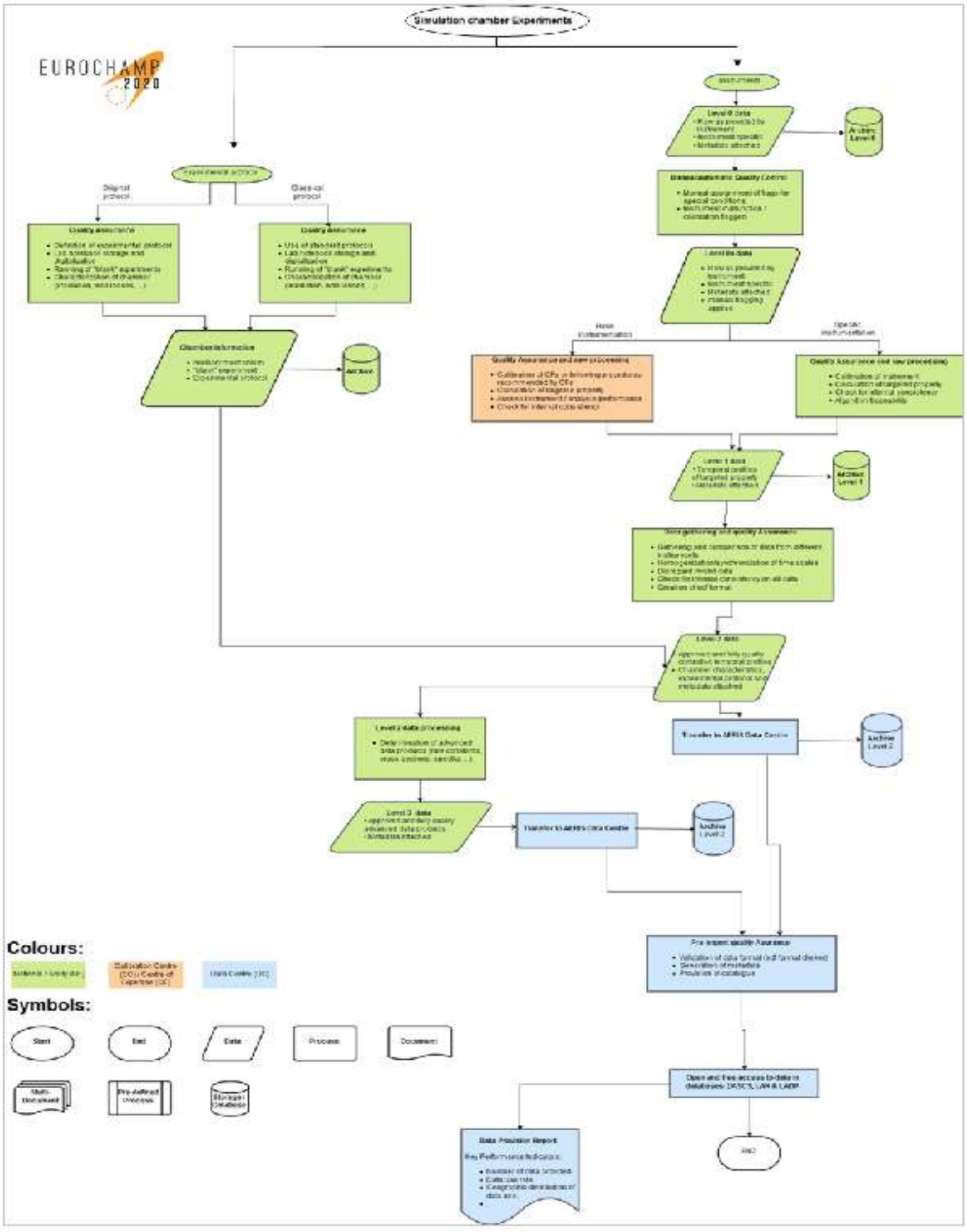
ACTRIS trace gases remote sensing data centre unit workflow diagram (lidar data)

uvvis data



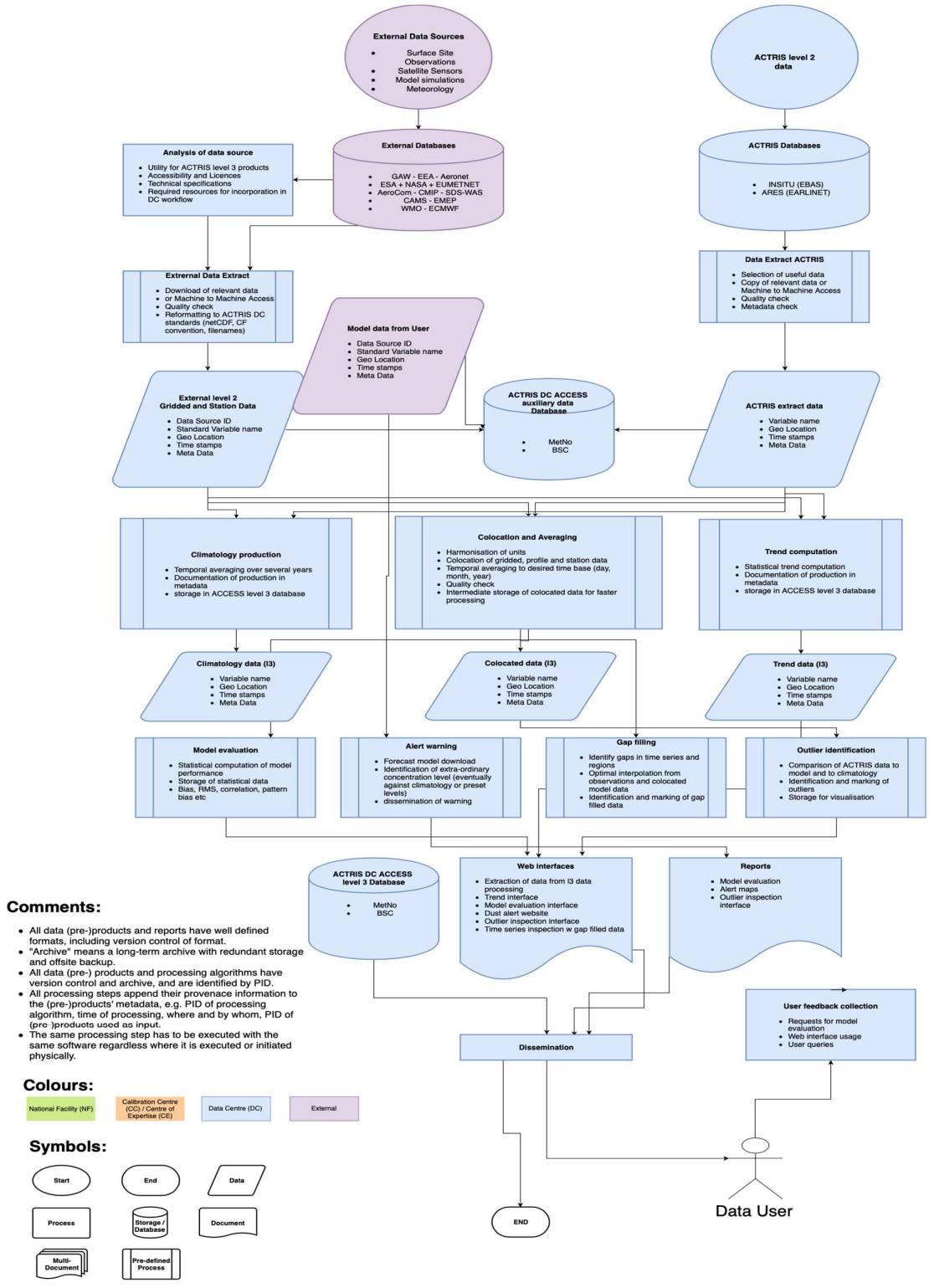
ACTRIS trace gases remote sensing data centre unit workflow diagram (uvvis data)

Appendix 7: ACTRIS Atmospheric simulation chamber data centre unit (ASC) data life cycle and workflow diagram



ACTRIS Atmospheric simulation chamber data centre unit workflow diagram

Appendix 8: Data lifecycle and workflow for ACCESS Data Centre Unit



ACTRIS ACCESS data centre unit workflow diagram

Appendix 9: Format and external data sources for level 3 variables

Below is a list of all lev3 variables that are listed in Annex II and the checkbox indicates whether they are included in the lists below or not:

- [ ] Column Water Vapor Content
- [ ] Climatology products for ACTRIS variables @ ACTRIS National Facilities across Europe
- [x] Collocation service of data from contributing networks
- [ ] PM retrieval @GAW sites
- [x] Single Scattering Albedo @ACTRIS National Facilities
- [ ] Integrated full-range particle number size distribution
- [ ] Source apportionment of submicron organic aerosols in Europe
- [ ] Volatile Organic Compounds (VOC) source attribution across Europe
- [ ] Cloud occurence at cloud in situ observational platforms
- [x] Direct Sun/Moon Extinction Aerosol Optical Depth (column)
- [x] Spectral Downward Sky Radiances
- [x] Aerosol columnar properties (GRASP-AOD)
- [x] ReOBS

- [x] Satellite data – combined with ground based ACTRIS data
- [x] Aerosol and Gas trend assessment
- [x] Data Interpretation and Outlier Identification Tool
- [x] Optimal interpolation and Gap filling tool
- [x] Model Evaluation Service
- [x] NWP Model Evaluation Service
- [x] Transport modelling products for assessment of source regions
- [x] Alert Service for National Facilities

Collected (other than ACTRIS L0-1-2)

Product	format	source	description
AERONET-NASA L1	csv	NASA/GSFC	https://aeronet.gsfc.nasa.gov
Terra+Aqua/MODIS	HDF4	AERIS	https://modis.gsfc.nasa.gov
CALIPSO	HDF4	AERIS	https://www-calipso.larc.nasa.gov
CLOUDSAT	HDF4	AERIS	http://www.cloudsat.cira.colostate.edu
PARASOL	HDF5	AERIS	http://www.icare.univ-lille1.fr/parasol
Aura/OMI	HDF4	AERIS	https://aura.gsfc.nasa.gov/omi
Terra/MISR	HDF4	AERIS	https://terra.nasa.gov/about/terra-instruments/misr
Metop/IASI	BUFR	AERIS	https://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Metop/MetopDesign/IASI/index.html
MSG/SEVIRI	NetCDF4	AERIS	https://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Meteosat/index.html
AeroCom	NetCDF4	METNO	https://aerocom.met.no/
NWP Model data	NetCDF4	NWP Centres	

Generated (systematic production)

Product	format	description
GRASP-AOD	NetCDF-CF	Aerosol size distribution retrieval from optical depth
ReOBS	NetCDF-CF	The ReOBS project proposes an advanced method to aggregate, quality-control and harmonize in one single NetCDF file as many available geophysical variables from a NF at hourly scale for the whole data record spanned by this ensemble of variables. This file allows to easily perform multiannual and multi-variable studies combining atmospheric dynamics and thermodynamics, radiation, clouds and aerosols, from ground-based observations associated to a NF.
Aerosol and Gas trend assessment	NetCDF-CF	Estimate of long term trends @ACTRIS sites, combining observations with models, interactive web visualization, automated assessment report
Data Interpretation and Outlier Identification Tool	NetCDF-CF	Quicklooks for time series data, compared to Copernicus Analysis and Reanalysis model products
? Optimal interpolation and Gap filling tool	NetCDF-CF	modal/data integration products which fill measurement gaps, eg in a time series, profile or field.
Alert Service for National Facilities	geoJSON **?**	Provide near real time update of special weather situations of interest for research activities at national facilities

Generated (on-demand services)

Some products will be generated through on-line services, and will generate datasets available for a limited time on a web server.

Product	format	description
Satellite data subsets	NetCDF-CF	Satellite data subsets, spatially and temporally colocated with ACTRIS ground-based measurements
Transport modelling products for assessment of source regions	NetCDF-CF	Backward transport modelling with FLEXPART to analyse air transirt and impact of various soucres. Develop tools to run FLEXPART operationally and automatically on a regular basis, e.g. monthly, for every site
Colocation service of data from contributing networks	NetCDF-CF	Benchmark data products including relevant EMEP and ACTRIS data: PM and/or sulphate with ACTRIS National Facilities compiled in one data product
Model Evaluation Service	NetCDF-CF	Automated model evaluation workflow, Evaluation reports of different complexity, NRT and reanalysis, climate models
NWP Model Evaluation Service	NetCDF-CF	Automated model evaluation workflow, evaluation reports of different complexity for NWP models, NRT and reanalysis, NWP models

Production of level 3 data solely based on data from ACTRIS observational platforms

List of ACTRIS level 3 data products

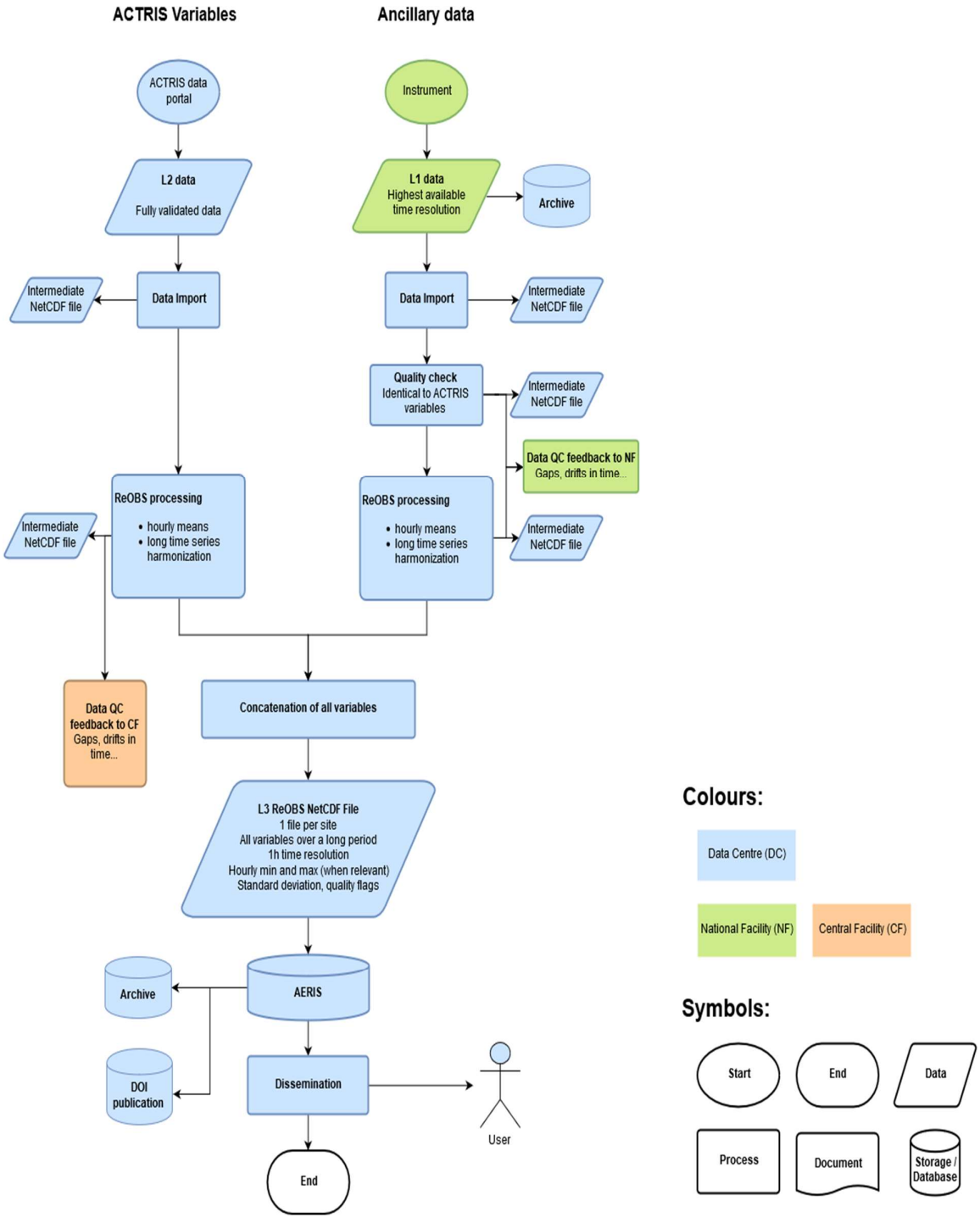
Production of ACTRIS level 3 data and tools through multi-source data integration services, employing external ground based measurement data

List of ACTRIS level 3 data products

Production of ACTRIS level 3 data products involving regional and global model data

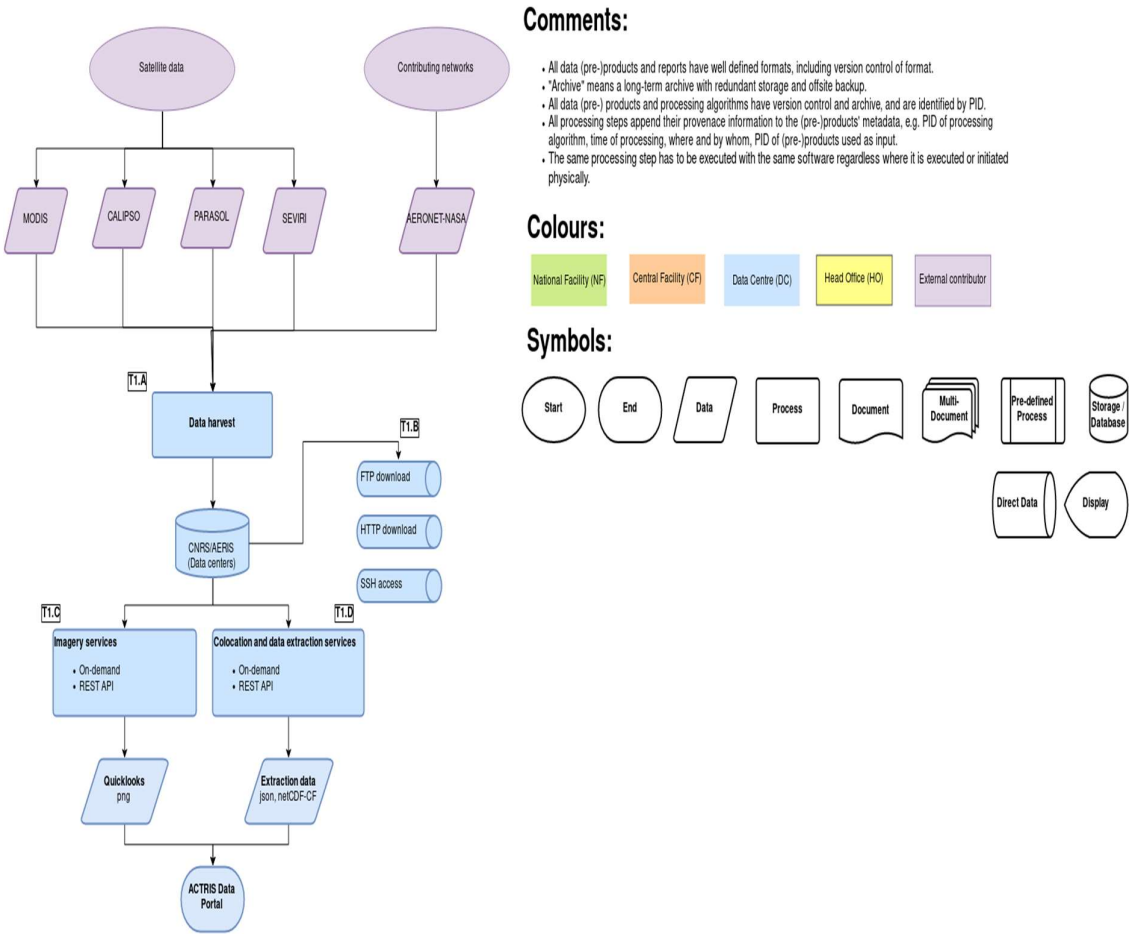
List of ACTRIS level 3 data products

Appendix 10: ReOBS workflow diagram



ReOBS workflow diagram

Appendix 11: Satellite data subsets workflow diagram



Satellite data subsets workflow

ACTRIS PPP (www.actris.eu) is supported by the European Commission under the Horizon 2020 – Research and Innovation Framework Programme, H2020-INFRADEV-2016-2017, Grant Agreement number: 739530