Data Management Plan for

ACROSS - Atmospheric Chemistry of the Suburban Forest

A Data Management Plan created using DMP OPIDoR

Data website: https://www.across.aeris-data.fr ,

Data catalogue: https://across.aeris-data.fr/page-catalogue ,

Quicklooks (products): https://across.aeris-data.fr/products/

Aircraft data (SafirePlus): https://safireplus.aeris-data.fr/

DMP Creator: Christopher Cantrell

DMP Template: Science Europe - DMP template (english)

Last modified: 30-May-2022

Funders: ANR (ANR-20-CE01-0010, ANR-17-MPGA-0002), LEFE-CHAT, DIM-QI2, UPEC

Related Documents: NASA LaRC recommended standard variable names: https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf

CF Standard Name Table: https://cfconventions.org/Data/cf-standard-names/28/build/cf-standard-name-table.html

Project abstract:

ACROSS (Atmospheric ChemistRy Of the Suburban foreSt) is an integrative, innovative, multi-scale project within the “Make Our Planet Great Again” (MOPGA) initiative designed to advance understanding of the fate of the photochemical processing of urban and biogenic air mass mixtures in the Paris region. An ACROSS hypothesis is that the anthropogenic-biogenic air mass mixing leads to changes in the production of oxygenated volatile organic compounds (VOCs) whose properties alter their importance in incorporation into secondary organic aerosols (SOA) and their roles in production of ozone and other relevant secondary species. A likely important factor is NOx transport to suburban biogenic environments and the resulting modification of key chemical processes.

A key highlight of ACROSS is an intensive, multi-platform measurement campaign that will take place in the summer of 2022. The campaign includes a 40-meter tower and ground-based measurements in the Rambouilet suburban forest to the southwest of Paris, airborne regional observations across Paris and suburban forested areas, and several other multi-instrumented ground sites located in the urban, rural, and semi-rural Paris region. The data collected from this campaign will provide a unique snapshot of the properties and mixing of urban and biogenic air masses around one of the most populated and polluted European megacities. This new knowledge will contribute to the advancing of our understanding at the process level and lead to the ability to represent such complex systems in numerical models, ultimately resulting in improved capability to predict the impacts on air quality, regional climate, and global climate change.

Principal Investigators: Christopher Cantrell, Vincent Michoud

Project Principal Investigator Affiliation: LISA (Laboratoire Interuniversitaire des Systèmes Atmospheriques)

ORCID iD: 0000-0002-3844-1560 (Cantrell), 0000-0003-2779-9329 (Michoud)

Data Contact Person for overall project: Christopher Cantrell

ACROSS - Atmospheric Chemistry of the Suburban Forest

1. Data description and collection or re-use of existing data

1a. How will new data be collected or produced?

ACROSS is a multi-organization, multi-site, multi-platform atmospheric composition measurement project, whose key activity is collection of data by many instruments staged at the various project locations and platforms. Data will be collected from a variety of instruments by individual instrument principal investigators and formatted according to the project requirements. The data will be collected by instrument PIs as raw signals (voltages or other) that must be processed into atmospheric quantities (concentrations, mixing ratios, fluxes, or other). There can be preliminary processing (using generic calibration factors, for example) to achieve approximate levels, and then final processing (using field-collected calibration information) to achieve quantities with the highest possible quality possible with the given instrument and operating conditions. For ground-based data, the various data types have the following designations (aircraft data data types are slightly different, see later):

Level 0: raw signals collected from the instrument (i.e. voltages, currents, count rates or other), perhaps with some filtering to remove data corresponding to non-ambient operation (during calibrations or periods of instrument problems, for example).

Level 1a: field atmospheric quantities for informal examination in the field, using for example, a constant calibration factor. Not subject to rigorous quantity control or final calibration procedures. This applies to Quicklooks as well.

Level 1b: preliminary atmospheric quantities for examination in the initial period after the campaign. Could be of higher quality than field data (using for example, actual measured calibration factors), but not subject to full QA/QC procedures of final data.

Level 2: final atmospheric quantities processed using best calibration information available and carefully checked for errors in collection and processing, and for eventual submittal to data archives.

Level 3: quantities derived from level 2 data using combinations of data and/or numerical models. This includes, for example, quantities such as ozone production rates, hydroxyl radical reactivity, or aerosol growth rates. It also includes quantities related to comparison with measurements from other ACROSS instruments or those from past measurement campaigns (for example, MEGAPOLI).

For data collected aboard the ATR-42 aircraft, we adopt the SAFIRE recommendations for data levels (and file naming formats, see later). To distinguish from the data designations for ground-based observations, the term “Aircraft-Level” is used. The core data collected and processed by SAFIRE will also follow this convention.

Aircraft Level 0 (L0): digital or physical counts from the sensors and instruments operated during a campaign. No correction or processing is done. Same as ground-based Level 0.

Aircraft Level 1 (L1): measurements converted into physical parameters from level 0 data. Level 1 data serves a dual purpose: (1) To quickly judge the quality of the acquisition of each instrument and identify any malfunctions, and (2) to generate quick plots and “quicklooks” type images of observables for flight quality assessment, and information sharing. Level 1 data is not considered validated data. They are therefore not intended to be distributed and used for scientific interpretation.

Aircraft Level 2 (L2): geophysical quantities generated from level 0 data. These are data controlled and validated a posteriori by an expert or an expert code corresponding to the geophysical parameters measured during flight.

Aircraft Level 3 or 4 (L3 or L4): corresponding, respectively, to geophysical parameters averaged or synthesized from several level 2 observations, or from different origins (aircraft inter-comparisons, model data, satellites, etc.). If aircraft data are combined with non-aircraft observations, models, or satellites, use the conventions corresponding to the other data source (see Level 3 earlier, and file naming conventions later).

All data will be embargoed for access by the scientific team and selected others only for a one year from the deadline for final data submittal, which is planned for 6 months from the completion of the data acquisition (tentatively late July 2022). After the embargo period, the latest versions of the level 2 and 3 data will be released to the scientific community and the public.

Each individual dataset will be accompanied by a metadata document that describes in detail the instrument used to collect the data, the instrument principal investigator and team, the configuration of the instrument, the location of the instrument, the details of the conversion of raw signals to atmospheric concentrations and properties, and the coverage of the data. A template for the metadata will be provided to the scientific teams (see ACROSS metadata template.xlsx).

The data will be managed by AERIS (see description in section 6a).

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Data will be collected from the individual instruments with a variety of parameters measured, averaging times, raw and processed units and formats. In addition to the numeric data from instruments, there will also be images collected and archived to document the configuration of instruments during the campaign. The raw instrument data (level 0) will be processed by the instrument teams with various levels of scrutiny to produce more refined versions of the data (levels 1a, 1b, and 2).

In addition to conversion of level 0 data to atmospheric quantities, various types of derived data products will be produced according to the project requirements. These include time-series merges on various time bases, time averaging according to various needs, production of data of types and formats for use in constraining models, and statistical analysis of data to produce a variety of metrics (for example, species ratios or production rates). These derived data (level 3) will be archived within the ACROSS database and will be available to the scientific team and the public according to the schedule discussed in section 1a.

The primary data format for all of the ACROSS data will be NetCDF. This is prescribed by the data archiving organization, AERIS, and has become the standard format for archiving of geophysical data. Applications will be made available to easily convert between NetCDF and ascii formats (such as csv, txt, tab-delimited) for easy treatment by spreadsheet programs and user-written code (see ACROSS Metadata template.xlsx, tab NetCDF software). While Netcdf has become the de facto standard for data archival, there is a long history of data submitted to archives in ascii format (e.g. the NASA Langley format). The ability to convert among various formats allows easy use of the data by scientific communities and the public.

The amount of final data (level 2) collected during ACROSS is estimated to occupy about 2.3 Gb (97 instruments measuring at least one species, with 10 instruments measuring 20 species or more for a total of about 300 species; measurement times range from 50 Hz to 5 minutes, with an average of 1 minute; there will be 15 flights of 3.5 hour duration for the aircraft and 42 ground measurement days

nominally 24 hours per day (a few measurements will not measure during day or night); the aircraft instruments will produce about 4000 data points per species and the ground instruments about 60000; for each species there are at least two pieces of information: time and property (but could include uncertainties), assume 128 bits per pair). There will likely be 1-5 Gb of campaign-related photographs. Note that an instrument measuring a single species, operating at 50 Hz, 24 hours per day, for 42 days (of which there are a few) could generate 23 Gb of raw data. This means that if all data were stored at the maximum time resolution, the storage requirements could be much greater than 2.3 Gb. There could also be significant storage requirements for model output and level 3 data.

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Each dataset (collection of data stored together) will be connected to a metadata document according to AERIS requirements. The connection will be accomplished through standardized naming conventions (i.e. metadata file names and data file names will be similar, see ACROSS metadata template.xlsx). The specified AERIS format for metadata is compatible with ISO 19115, INSPIRE and other standards and extensible to adapt to the specific features of certain types of data (e.g. aircraft and ground platforms, remote sensing, and output of 0D and 3D models). They can be exported to standard formats or other portals on request of AERIS. Each dataset will also be assigned a Digital Object Identifier (DOI) by AERIS.

Data will be organized by AERIS according to the desires of the ACROSS science team. Each instrument will be assigned terminology to be used for file naming and version numbers. Data will be submitted to the archive according to the quality of the data (data levels) and associated deadlines. For example, field data (level 1a) will be submitted as soon as possible after collection so that it can be used for decision-making. Some of these data will be set up for “Quicklooks” on the ACROSS data website. Later, preliminary data (level 1b) will be submitted (soon after the completion of the campaign) that has been subjected to somewhat more rigorous examination than field data but is not yet in a final state. Then, final data (level 2) will be submitted that has been fully treated (as described in the metadata) in terms of identification and removal of spurious data, accurate application of calibration factors, and full compliance with the data submittal requirements. If errors are discovered as data are being used, revised data will be submitted in a timely manner to be included in the collection of data stored for each instrument, with version numbers incremented accordingly.

Each of these types of data will be clearly described, including processing procedures and methods for dealing with instrument problems and calibrations, in the metadata file linked to the instrument.

The metadata documentation will follow the recommendations and requirements of AERIS. This includes the categories of information listed in section 7a (also, see ACROSS metadata template.xlsx):

These apply most importantly to the scientific measurements and modelling for ACROSS but are also applicable to photographic and other documentation.

2b. What data quality control measures will be used?

The first step in ensuring high data quality will be asking the measurement team to carefully check their procedures for data collection and reduction. All data should be visually inspected for anomalous and outlier data, including various screening methods using well-tested codes and plots.

To the degree possible, measurements of the same quantity will be compared. These will be possible when similar instruments are deployed at the same site, but also during aircraft overflights of ground sites. Comparisons of ambient data will be performed at the highest temporal resolution possible that is consistent with the performance of the instruments involved in the comparisons, and also at lower temporal resolutions. Also, measurement PIs will be encouraged to exchange calibration standards when appropriate. The data for comparisons will be examined as soon as possible after the completion of the comparison period so that errors in instrument operation or data handling procedures can be identified and corrected.

The two types of comparison data possible: the comparison of selected ambient measurement periods. and the comparison of measurements of standards, and. The first type of comparison will be submitted to the archive within the ambient data files (levels 0-3) and noted in the corresponding metafile. The second type of comparison will be archived in separate files with slightly different formats and naming schemes.

Comparison data will be submitted to the data archive as level 3 or aircraft L4 data. Each comparison exercise will require a metadata sheet describing the origin of the data and the details of how the data were made compatible for comparison (converting to common time base, adjusting averaging time, etc.).

A data comparison manager will be identified who will be responsible for encouraging processing of data from comparison periods and submittal in a timely manner. The data comparison manager will perform the initial analysis of the comparisons and lead a team to report the results of the comparisons to the parties involved, the ACROSS team, and, if appropriate, to the wider scientific community through presentations and publications. The goal of instrument comparisons is not to criticize any measurements, but to improve the observations for the good of the project.

PIs should explain clearly in the metadata document (field 18) the philosophy of calibration, zeroing, and the use of this information in the reduction of the data.

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

All data will be stored in a raw format (level 0) by the individual instrument PIs and their teams. We ask that these teams regularly make backup copies of their data external to the data acquisition computers (e.g. thumb drive or cloud storage) and regularly back up these data copies. PIs will be encouraged to submit their raw data to the ACROSS data archive. Teams are encouraged to have their raw data duplicated in three to four locations, and to perform data backups daily (or after each aircraft flight).

During the data reduction process, teams are encouraged to regularly backup their data at various stages of the reduction process. They are also encouraged to make redundant copies of data submitted to the ACROSS archive.

The data located in common backup locations will be password protected with access only for the specific measurement teams.

3b. How will data security and protection of sensitive data be taken care during the research?

The LISA DT department has experts in computer hardware and software. Their expertise will be employed to ensure data is backed up regularly and safely. Use may also be made of IT consultants to supplement the DT expertise. In addition, AERIS provides data security and backup as part of their activities to provide data storage and access for national measurement activities.

In the event of data loss, the intent is to make use of the redundant copies of the data at various locations. If all copies are lost or the latest version of reduced data has not yet been archived, then use of professionals who are experts in data recovery will be engaged.

All level 0 data will be password protected and only available to the specific measurement teams. As data reduction moves forward, interim reduced data (level 1a and 1b) will be made available to the science team, but not the public. Access to these data will not involve revealing personal information or trade secrets. The measurement teams will be responsible for making sure that only proper data is shared.

Access to the final scientific data (levels 2 and 3) will follow the data protection policies of the ACROSS partner institutions. If there are issues that conflict between ACROSS and institution policies, the ACROSS management will contact the institutions involved to resolve the conflicts. If this is not possible, then the data collected by teams from the institutions in question may have to be archived separately.

4. Legal and ethical requirements, codes of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

There will be no personal data collected during ACROSS. The data are from scientific measurements of atmospheric composition and other related quantities.

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

All data collected during ACROSS are the property of the measurement teams and/or their institutions. These teams agree to share their data with the rest of the science team, and when the embargo period is over, with the wider scientific community and the public. Since these data were collected during a campaign supported by public funds, the data are ultimately owned by the public.

Data types and access:

1) Raw data (level 0) will only be accessible to the measurement team that collected the data. The teams are free to provide access to these data to whomever they wish.
2) Processed data (field including Quicklooks (level 1a), preliminary (level 1b), and final (level 2)) that is uploaded to the AERIS data archive will be available only to the ACROSS science team during the periods of (a) the campaign, (b) post-campaign data processing (6 months), and post-final data submittal embargo (1 year). This access availability includes measurement groups directly involved in ACROSS, measurement teams for the sTREEt and H2C campaigns, and the Airparif and Lig'Air air quality measurement networks (through data-sharing agreements). Other individuals can be given access to specific data by permission of that measurement PI. Access will be controlled by password protected websites and cloud-based datasets.
3) After the embargo period is complete (18 months after completion of the ACROSS campaign), the final processed data (revised as needed) stored on the AERIS data archive will be available to the public. Other versions (levels 0, 1a and 1b) of the data will be archived but will not be made publicly available.

In all cases, access to data does not imply freedom to use the data in whatever manner the user wishes. All use of the data must come with invitations of co-authorship of any scientific or non-scientific

publications and other conditions put forward by the measurement team. If the measurement team in question does not feel co-authorship is warranted, then a statement in the acknowledgements section of the paper should be considered.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

The data placed in the ACROSS archive consists of measurements collected as part of the ACROSS campaign by individual measurement teams. The assumption is that the data were collected by these teams in an ethical manner that does not involve cheating or theft of data from other teams, or the fabricating of data that are presented as the results of bona fide measurements.

If it is found that a team has behaved unethically in their participation in ACROSS, firm measures may be taken that could include, but not limited to, removal from the ACROSS science team and denial of access to the ACROSS data.

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

See section 4b for discussion of access to data by the science team and the public.

Final data (level 2) will be accessible to the scientific community and the public 18 months after the completion of the ACROSS measurement campaign. This will be managed by AERIS who will remove the password protected status of the final data at that time. During the embargo period, exclusive access to the data will be provided to the ACROSS measurement team to give them time to prepare scientific publications of the results of their measurements and the corresponding analysis.

The data will be archived in perpetuity by AERIS and will also be archived in other locations (including at LISA) to provide redundant backups of the data. AERIS routinely backs up data on their server, so this provides assurance that the data will always be available.

Should something happen to the AERIS organization and data archive in the future, data requests will be managed by LISA.

The ACROSS data website: https://www.across.aeris-data.fr includes information about ACROSS and the data (level 1a) processed during the campaign. These and higher levels of data (1b, 2 and 3) will be accessible through the AERIS catalogue: https://www.aeris-data.fr/catalogue. Aircraft data will be accessible at SafirePlus: https://safireplus.aeris-data.fr/ in addition to AERIS.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

The ACROSS final (level 2) data will be the most important data for long-term archival. The primary location of the data and its backup will be at the Data Service of Observatoire Midi-Pyrénées (called SEDOO) in Toulouse. The SEDOO is one of the Data and Services Centre of AERIS and will oversee the data management.

The mid-term preservation (10 years) is guaranteed by AERIS (see section 6a). The question of long-term preservation will be addressed soon in the framework of Data Terra RI. In case AERIS/SEDOO finds it cannot maintain this database, the responsibility could be transferred to another data centre within AERIS. AERIS is strongly supported at national level (CNRS, CNES, Meteo-France…). In the highly unlikely event that AERIS will have to close operations, we guarantee that we will migrate all content to other suitable repositories, and since all datasets will have DOIs (see 5d), all citations and links to datasets will not be affected.

Other ACROSS data will be archived by the individual measurement groups and will be included in the storage of data at LISA.

5c. What methods or software tools are needed to access and use data?

Data will be accessible and downloadable from a dedicated website created and managed by AERIS. Access to the data will not require intervention or handling of requests. Once the website is accessed, the data can be immediately downloaded. The data will be stored in NetCDF format that will be readily accessible into the foreseeable future. Access to tools for conversion of NetCDF to other formats will be linked to the data website. Should the NetCDF format become obsolete, the data will be converted into another format and archived.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

One of the services of AERIS (a trustworthy, long-term data repository and manger; see description in section 6a) is to provide DOI values for each dataset. These will be requested for each component of the ACROSS database and will be stored with the metadata viewable on the website.

AERIS (see section 6a) has an agreement to assign DOIs to datasets with the prefix 10.25326. A DOI will be requested for ACROSS datasets. It will be registered on Datacite (https://datacite.org/) using its metadata scheme. The landing page, provided by AERIS, will contain a sample citation, full metadata, and information on how to access the data.

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The cluster for Atmospheric data AERIS (https://www.aeris-data.fr), part of the French Data Terra Research Infrastructure (RI), has the objective to facilitate and enhance the use of atmospheric data,

whether from satellite, aircraft, balloon, or ground observations, or from laboratory experiments. It generates advanced products and provides services to facilitate data use, to prepare campaigns, and to interface with modelling activities. It consists of four Data and Service Centres (DSC) with more than 20 years' experience and strong expertise in data curation, storage, preservation, and dissemination by these entities: ICARE, ESPRI, SATMOS and SEDOO. Most of these data centres are involved in European initiatives and projects promoting the FAIR data principles and participating in the European Open Science Cloud (EOSC). AERIS is hosting the IAGOS data center and parts of the ACTRIS data center (trace gas remote sensing, atmospheric simulation chamber studies, and other types of data).

The AERIS organization will be ultimately responsible for the many data management/stewardship activities involved with the ACROSS data. For the short-term and intermediate term activities and decisions, PIs Christopher Cantrell and Vincent Michoud will manage the transition of the ACROSS data through its various stages.

Note that all the ACROSS data (ground-based and aircraft-based observations, model output, satellite observations, and derived quantities) will be archived at SEDOO and available via the ACROSS AERIS website (www.across.aeris-data.fr). The ACROSS aircraft-based observations will also be archived at the SAFIRE+ database archive and accessible via the SAFIRE+ website (www.safireplus.aeris-data.fr)

Dr. Cantrell created this Data Management Plan with the assistance of Dr. Michoud, AERIS representatives (Damien Boulanger and Guillaume Brissebrat) and the rest of the ACROSS Management Team. This DMP will be regularly reviewed and updated to ensure its content reflects the intent and needs of the ACROSS science team and the public interested in ACROSS scientific issues.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

AERIS will ensure that the ACROSS data will be FAIR. This is a component of their agreement to manage the ACROSS data. We do not expect any costs from the AERIS activities. The funds needed to create redundant backups discussed earlier, will come from the MOPGA ANR grant of Dr. Cantrell. The data are distributed by data centers, and so, by concept, the infrastructure and maintenance costs are mutualized. For 2 GB, we can estimate a yearly cost of 1000 €, covered by AERIS public funding.

7. File naming conventions

7a. Metadata file description document

Metadata files will be created for each data product (measurements from all platforms, model output, satellite observations, and derived quantities) for all the versions submitted (updates or higher levels of QA/QC) of data are created for a given product. The metadata document for previous versions is edited to include the new version number and date, which is logged with all the previous version numbers and dates in Field 15. The ACROSS Metadata file format is followed using the "ACROSS metadata template.xlsx". The metadata file naming conventions are:

ACROSS_Institution_platform_product_date-range.ext	(ground)
ACROSS_SAFIRE-ATR42_INSTRUMENT_date-range_flight-range.ext	(aircraft)
ACROSS_satellite_sensor_study-region_product_date-range.ext	(satellites)
ACROSS_insitution_mode-name_study-region_product_date-range.ext	(models)

Some example metadata file names:

ACROSS_NOx_LISA_RambForest_1-min_20220613-20220727.csv
ACROSS_PTRMS-VOCs_LISA_SAFIRE-ATR42_1-min_20220613-20220707.csv
ACROSS_CHIMERE_LISA_study-region_1-hr_20220613-20220727.csv

The metadata file format is (see file "ACROSS metadata template.xlsx":

Field No	Required	Parameter
1	*	Title
2	*	Abstract
3	*	Principal Investigator (name)
4	*	PI Institution
5	*	PI email address
6	*	Contact person for the data (name)
7	*	Data contact Institution
8	*	Data contact email address
9		Other persons involved in data (names)
10	*	Download links to data
11	*	DOI for dataset
12		Links to data format tools (e.g. format conversion)
13	*	Temporal coverage (start date, end date) UTC, indicate missing days
14	*	Spatial coverage (latitude, longitude, altitude)
15	*	Data version and date of creation
16	*	Platform name, location, and type or model name, type
17	*	Source: Instrument or model name
18		Principle of measurement or model
19	*	Parameters measured or output from model contained in data files. Refer to the NASA LaRC recommendations: (https://www-air.larc.nasa.gov/missions/etc/AtmosphericCompositionVariableStandardNames.pdf) or alternatively, the CF Standard Name Table (https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html)
20	*	Averaging time
21	*	Time format (midpoint or start-stop)
22	*	Missing data flags
23		Links to information on instrument, measurements, or model
24	*	Conditions on the use of these data
25	*	Other information about these data (detection limits, accuracy/uncertainty, propagation of errors, reasons for missing data, and other useful information)

7b. Product data files

Product data (both digital data and quicklooks images) file names are chosen to be consistent with the corresponding metadata file name. The parameters with examples for each type of filename are shown in the following table.

Parameter	Ground Platforms	Aircraft Platform	Satellites	Numerical Models
Project name	ACROSS	ACROSS	ACROSS	ACROSS
Institution	LISA, ICARE	NA	NA	LISA
Site or platform, satellite name	RambForest, PRG, SIRTA	SAFIRE-ATR42	Sentinel5P	CHIMERE
Satellite sensor	NA	NA	TROPOMI	NA
Study region	NA	NA	France	France
Product or instrument	NO, aero-scat, VOCs	O3	NO2, AOD	NO2, O3
Date	20220624	20220615	20220712	20220723
Time-resolution	1-min	1-Hz	daily	1-hour
Overpass time	NA	NA	1330	NA
Flight ID	NA	F12	NA	NA
Data level	NA	L2	NA	NA
Version: v level no. – version no. (vX-Y)	v2-3	NA	V2-1*	V2-2
Extension (netcdf [nc] or image [jpg, png, pdf])	.jpg	.nc	.nc	.nc

* Use satellite conventions for level numbers

The following terms are standardized for the sites and platforms.

Site or Platform	Standardized parameter name
Rambouillet Forest site, ground or tower	RambForest
Universite Paris Cité	PRG
SAFIRE ATR-42 aircraft	SAFIRE-ATR42
SIRTA	SIRTA

The quicklook or data file naming convention for ground-based observations is:

ACROSS_institution_site-or-platform_product_date_time-resolution_version.ext

The quicklook or data file naming convention for aircraft-based observations is:

ACROSS_SAFIRE-ATR42_INSTRUMENT_time-resolution_DATE_FLIGHTID_Data-level_version.ext

Where the DATE is in the format YYYYMMDD, the flight ID is “F” plus the flight number, the data level is L0, L1, L2, L3 or L4 (slightly different than the ACROSS definition for ground-based data in this document), and version is “v” plus a number that begins with 1 increasing as the data at a given level is revised. When the first data at a new level is generated, restart the numbering at 1. See the document “SAFIRE_Management_GUI_Guide_Management_Donees.pdf” for more details.

For satellite observations, the quicklook or data file naming convention is:

ACROSS_satellite_sensor_study-region_product_date_time-resolution_overpass-time_version.ext

For model output, the quicklook or data file naming convention is:

ACROSS_institution_model-name_study-region_product_date_time-resolution_version.ext

Some example product digital data file names (same file name format for quicklooks but with image (jpg, png, or pdf) extensions):

ACROSS_LISA_RambForest_NOx_20220613_1-min_v1b-2.nc
ACROSS_SAFIRE-ATR-42_PTRMS-VOCs_1-min_20220622_F7_L2_v1.nc

ACROSS_Sentinel5P_TROPOMI_France_NO2_20220622_daily_1330_v2-1.nc

ACROSS_LISA_CHIMERE_France_NO2_20220703_1-hr_v2-3.nc

Data files are created for each day or each flight. Time for digital data in UTC, and time for quicklooks in either CEST or UTC (clearly indicated on the image). The digital data will be in tabular format with first one or two columns giving the time in one of two formats, and the other columns giving quantities of the product corresponding to those times. The details of the product will come from the corresponding metadata file (see section 7a) and the product data file name. The first line of the data file will consist of the file name; the second line will contain the column headers. An example format of the data files for time format "midpoint" is:

ACROSS_LISA_RambForest_NOx_20220613_1-min_v1b-2.nc		
Time (midpoint)	NO, ug/m^3	NO2, ug/m^3
00:00:30	12.3	46.2
00:01:30	12.6	45.4
00:02:30	-99999.99	-99999.99
…	…	…

An example format of the data files for the time format "start-stop" is:

ACROSS_LISA_RambForest_NOx_20220613_1-min_v1b-2.nc			
Time (start)	Time (stop)	NO, ug/m^3	NO2, ug/m^3
00:00:00	00:00:58	12.3	46.2
00:01:00	00:01:57	12.6	45.4
00:02:00	00:02:55	-99999.99	-99999.99
…		…	…

Note that data files could contain columns with additional information, such as data uncertainty (if it varies on a point-by-point basis) and sample height (if it varies). Though not required, it is best to include a line for each time element, including the missing data flag as appropriate. If an entire day of data is not available (e.g. instrument failure), then it is not necessary to submit an empty file or one filled with missing data. Note days with missing data in the corresponding metafile, under parameter 13.

8. Quicklooks Data Presentations

The steps to making Quicklooks files involve the following steps. Some of these have nuances to a particular instrument or to the format of the raw data. For each day (24-hour period, from midnight on a day to just before midnight on that same day, or for each aircraft flight), using software developed by your team or that which came with the instrument, produce tables of data corresponding to the type of plot to be produced. The x-axis could be time, and the y-axis could be the mixing ratio of a

species. Or x-axis could be concentration, and the y-axis could be altitude. Suggest no more than two plots per file (one is better), although several data series could be shown in each plot (suggest no more than about 4).

Once the data table is produced, then a plot is produced with the software of your choice. To maximize readability and viewer understanding, consideration should be given to formatting the following features as described.

- For the time axis (usually the x-axis), suggest major tick marks every 3 hours, and minor tick marks every hour.
- On the plot, clearly indicate the time zone corresponding to the time shown. For Quicklooks, you can use local time (CEST) or universal time (UTC).
- Use visible colors for series or contours. Avoid difficult to see colors on the the screen (e.g. yellow).
- Use descriptive plot titles, including the species/parameter, measurement location, and date. Use description axis titles including measurement units and species/parameter. Include a legend if needed to understand the plot.
- Add light colored grids on both axes to allow easy comparison of data from one side of the plot to the other.