

Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations



ESA Contract No. 4000118181/16/I-EF




Deliverable D7:

MAXDOAS Network Processing System Technical Requirements Document

Date: 06/11/2017

Version: 0.1

 BIRA-IASB	ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM ₄ DOAS)	Page 2 30 January 2018 Technical Requirements Document
--	---	--

Contributing authors:

Caroline Fayt (BIRA-IASB)

Francois Hendrick (BIRA-IASB)

Document change record

issue	date	item	comments
0.1.0	2017-11-06	–	Initial version

Table of Contents

Document change record	3
1 Introduction	6
1.1 Purpose	6
1.2 Scope	6
1.3 Definitions, acronyms and abbreviations	6
1.4 References	7
1.4.1 Applicable documents	7
1.4.2 Reference documents.....	7
1.4.3 Digital reference documents.....	7
1.5 Requirements convention.....	7
2 General Description	8
2.1 Relation to other projects.	8
2.2 Model description.	8
3 Specific Requirements.....	9
3.1 Functional requirements.	9
3.1.1 Level-1 data	9
3.1.2 Scientific modules package	9
3.1.3 Level-2 data products.....	10
3.1.4 Ancillary data.....	10
3.1.5 QA/QC.....	11
3.1.6 Configuration	11
3.1.7 Monitoring.....	11
3.1.8 Dashboards.....	12
3.2 Operational requirements.	12
3.3 Resources requirements	12
3.3.1 Compute servers	12
3.3.2 FTP server	12
3.3.3 Dynamic web pages.....	12
3.3.4 S/W requirements :	12
3.4 Security requirements.	13
3.4.1 FTP/HTTP	13
3.4.2 Backup of data and code.....	13

3.4.3	Dashboards web pages	13
3.5	Quality requirements	13
3.5.1	Modularity and flexibility.....	13
3.5.2	Performance	13
3.5.3	Scalability.....	13
3.5.4	Portability.....	14
3.5.5	Availability	14
3.5.6	Reliability requirements	14
3.5.7	Maintainability requirements.....	14

1 Introduction

1.1 Purpose


This document describes the technical requirements applicable to the FRM₄DOAS project entitled “Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations” (ESA contract 4000118181/16/I-EF).

1.2 Scope

The purpose of this project is to develop a centralised system providing harmonised ground-based reference data from a network of MAXDOAS instruments within a short latency period. The motivation is to provide a service that will help scientific research organisations having limited resources for operational processing of their data to deliver measurements according to standards required for the NRT validation of satellite data products to be generated as part of the Copernicus Sentinel programme.

1.3 Definitions, acronyms and abbreviations

AFGL	Air Force Geophysics Laboratory
API	Application Programming Interface
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
ECMWF	European Centre for Medium-Range Weather Forecasts
ESA	European Space Agency
FRM	Fiducial Reference Measurement
FTP	File Transfer Protocol
GEOMON	Global Earth Observation and MONitoring
GEOMS	Global Earth Observation Metadata Standard
HDF	Hierarchical Data Format
HPC	High Performance Computer
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
MADCAT	Multi Axis Doas - Comparison campaign for Aerosols and Trace gases
MAXDOAS	Multi-Axis Differential Optical Absorption Spectroscopy
NDACC	Network for the Detection of Atmospheric Composition Change
netCDF	Network Common Data Form
NORS	Demonstration Network Of ground-based Remote Sensing observations in support of the Copernicus Atmospheric Service
NRT	Near Real Time
OEM	Optimal Estimation Method
PhP	Hypertext Preprocessor, programming language for dynamic web pages
PI	Principal investigator
QA/QC	Quality Assurance/Quality Control
QA4ECV	Quality Assurance for Essential Climate Variables
S/W	Software
SAOZ	Système d’Analyse par Observations Zénithales

 BIRA-IASB	ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM ₄ DOAS)	Page 7 30 January 2018 Technical Requirements Document
--	---	--

SQL	Structured Query Language
SRD	Software Requirement Document
URL	Uniform Resource Locator
UV-Vis	Ultraviolet-visible
XML	Extensible Markup Language

1.4 References

1.4.1 Applicable documents

- [AD-1] ESA/ESRIN Statement of Work, ref. Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations, ENVI-SPPA-EOPG-SW-14-0003, July 2015
- [AD-2] FRM₄DOAS Technical proposal
- [AD-3] MAXDOAS Network Scientific Requirements (deliverable D2)
- [AD-3] MAXDOAS Instruments Review Document (deliverable D3)
- [AD-4] MAXDOAS Algorithm Round-Robin Definition and Results Document (deliverable D5)

1.4.2 Reference documents

1.4.3 Digital reference documents

- [DR-1] GEOMS UV/Vis Differential Optical Absorption Spectroscopy (DOAS) template description
<https://avdc.gsfc.nasa.gov/index.php?site=1876901039>
- [DR-2] PANDONIA Fiducial Reference Measurements for Atmospheric Composition
<http://www.pandonia.net/>
- [DR-3] NDACC recommendations for total O₃ and stratospheric NO₂ retrieval
<http://ndacc-uvvis-wg.aeronomie.be/tools.php>
- [DR-4] NORS project
http://nors.aeronomie.be/projectdir/PDF/NORS_D4.1_DFD.pdf

1.5 Requirements convention

The requirements specified in the following section are assigned a unique reference identifier, with the format :

RRR-x.y.z-nn

where :

- RRR** three characters for the type of requirements (FUN : functional, SEC : security, OPE : operational, RES : resources, QUA : quality)
- x.y.z** 3 digits number corresponding to the chapter number
- nn** 2 digits incremental number for each requirement of a given chapter.

2 General Description

2.1 Relation to other projects.

As a first step, a harmonised scheme for the determination of differential slant columns will be adopted. This will largely be based on results from parallel projects such as NORS, the MADCAT campaign and the ongoing QA4ECV project where standards for the spectral fitting of ozone, O₃, NO₂ and HCHO are being established, including recommendations on trace gas absorption cross-sections and corresponding spectroscopic database(s).

The project will rely on expertise gained in past activities and projects (e.g. the EU GeoMon and NORS projects) and will also be linked to parallel development on the fiducial air quality PANDORA network.

2.2 Model description.

Calibrated Level-1 data will be uploaded at regular intervals (that may differ from one station to another) on a FTP server at BIRA-IASB. The FTP incoming repository will be pooled several times per day (for example, every 15 minutes) and new files will be downloaded. After QA/QC test (in green in Figure 1), they will be submitted to the scientific chain that consist in slant columns retrieval with QDOAS S/W and stratospheric and/or tropospheric profiling algorithms.

The names of Level-1 files that passed the QA/QC tests will be catalogued in a first database with the corresponding metadata while a second data base will be constructed with the names of the Level-2 files. During the processing, information will be registered in logbooks. Monitoring tools in connection with the databases and logbooks will control the whole chain at each processing step and will generate alerts if an anomaly is detected.

Dash boards on web pages in relation with the Level-1, logbooks and Level-2 databases will allow data submitters to rapidly check if the process of their own files succeeded or not and to open status reports. The Level-2 products will be delivered in the GEOMS HDF format ([DR-1]) through HTTP pages.

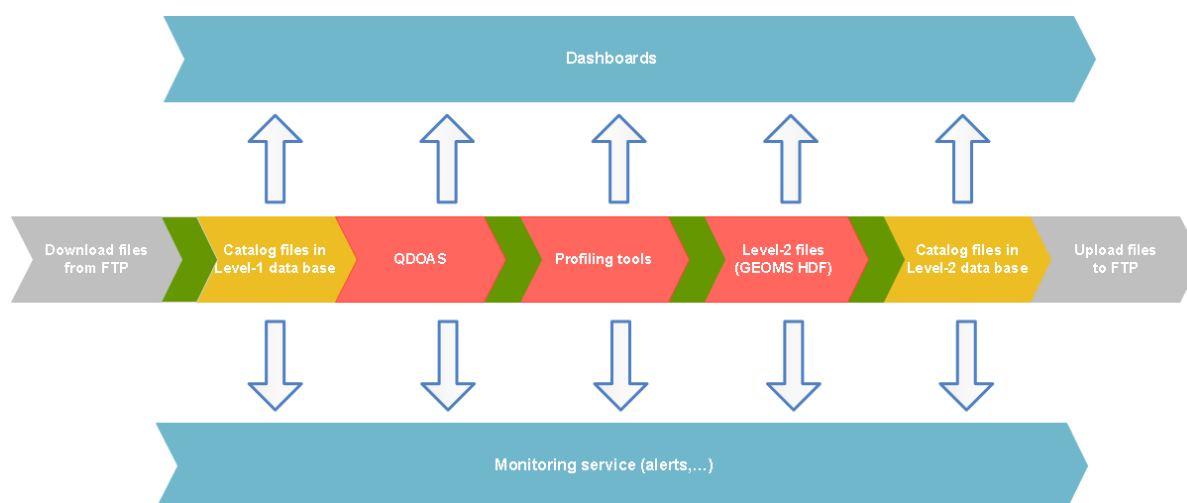



Figure 1 : The MAXDOAS processing chain

	<p align="center">ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM₄DOAS)</p>	<p align="right">Page 9 30 January 2018 Technical Requirements Document</p>
---	---	---

The MAXDOAS Network Processing System Architecture Design Document (Deliverable D-8) describes how the different modules of the processing chain will be linked.

3 Specific Requirements

3.1 Functional requirements.


3.1.1 Level-1 data

- [FUN-3.1.1-01] Before submitting Level-1 data, PIs or instrument operators shall register their instrument according to a procedure that will be defined by the consortium. Input configuration files for the spectral analysis for new instruments shall be set up jointly between BIRA and the instrument PI/data submitter.
- [FUN-3.1.1-01] Before submitting Level-1 data, PIs or instrument operators shall register their instrument according to a procedure that will be defined by the consortium. Input configuration files for the spectral analysis for new instruments shall be set up jointly between BIRA and the instrument PI/data submitter.
- [FUN-3.1.1-02] Data submitters have to make sure to submit only new or modified files on the FTP server.
- [FUN-3.1.1-03] Input data will consist in one day of calibrated ground-based DOAS measurements. The submitted spectra shall be dark-current and offset corrected. The expected input format is netCDF. A description of the file format will be provided to the users. For widely distributed systems such as MiniDOAS, EnviMes or SAOZ instruments, converters from the original format to the requested netCDF format will be distributed.
- [FUN-3.1.1-04] The file names shall satisfy a predefined format including information such as the instrument number, a channel number, the identification of the institute, the observation site, the measurement date, the times of the first and the last measurement... The information contained in the file names will be used to dispatch files in the correct directory for further processing.
- [FUN-3.1.1-05] The names of input/output files and metadata shall feed databases. Requests shall be possible from python scripts or dashboards through APIs (JSON).
- [FUN-3.1.1-06] Mechanisms shall be setup to follow the processing steps and to ensure the traceability of the processing chain. A process identification number will be assigned to intermediary files and Level-2 files generated by a new process and the different operations will be reported in logbooks.

3.1.2 Scientific modules package

Slant columns densities

- [FUN-3.1.2-01] QDOAS is the selected algorithm for slant columns retrieval. The program already uses netCDF format for output files. It shall be adapted to read netCDF Level-1 files. There shall be DOAS analysis configuration files per instrument and per product. The configuration shall be based on state-of-the-art recommendation settings.

	<p align="center">ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM₄DOAS)</p>	<p align="right">Page 10 30 January 2018 Technical Requirements Document</p>
---	---	--

[FUN-3.1.2-02] QDOAS generates output files in ASCII or netCDF formats. For this project, netCDF format will be preferred. Additional converters might be implemented/provided to feed the profiling algorithms with files in the format they use.

Tropospheric NO₂, HCHO and aerosols profiles

[FUN-3.1.2-02] The algorithm to retrieve tropospheric profiles from MAXDOAS measurements shall be the ones selected by the Round Robin exercise (see Deliverable D-5), i.e. the parameterized MAPA MC and Optimal-Estimation-Method-based MMF profiling algorithms.

Stratospheric NO₂ profile and total ozone retrieval

In addition to MAXDOAS tropospheric products, the system shall also deliver stratospheric profiles of NO₂ and total ozone from zenith-sky measurement.

[FUN-3.1.2-03] An algorithm developed at BIRA-IASB and based on the OEM (Hendrick et al., 2004) has been selected for stratospheric NO₂ profiles retrieval.

[FUN-3.1.2-04] The standard NDACC approach based on zenith-sky UV-vis observations at twilight will be used for total ozone retrieval.

3.1.3 Level-2 data products

[FUN-3.1.3-01] The granularity will be one file per instrument, per day and per supported product.

[FUN-3.1.3-02] In the frame of this project, the MAXDOAS processing system shall be able to deliver :

- Total ozone columns from twilight zenith-sky data (two values per file);
- Stratospheric NO₂ vertical profiles, from twilight zenith-sky data (two profiles per file);
- Tropospheric NO₂, HCHO and aerosols profiles from MAXDOAS scans
(one profile per scan).


[FUN-3.1.3-03] Delivered data should include the main measurement quantities, uncertainty information and quality flags as well as ancillary metadata which can be useful for the interpretation of comparisons between (MAX-)DOAS and satellite or model data, like averaging kernels, cloud conditions, location (latitude, longitude) of the effective air masses, etc. in addition to the trace gas or aerosol data.

[FUN-3.1.3-04] There will be Level-2 files per instrument, per day and per product. The Level-2 data that pass successfully QA/QC tests shall be automatically catalogued in Level-2 data base. The files shall satisfy GEOMS HDF format.

[FUN-3.3.3-05] The Level-2 files shall contain the information necessary to track back the process and identify the Level-1 sources.

3.1.4 Ancillary data

[FUN-3.1.4-01] The system will use ancillary data from sources designated by default (for example, AFGL for temperature/pressure profiles) but shall be able to extract other ones from other external sources (for example ECMWF) or ingest the ones from data providers.

 <p>BIRA-IASB</p>	<p>ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM4DOAS)</p>	<p>Page 11 30 January 2018 Technical Requirements Document</p>
--	---	--

3.1.5 QA/QC

- [FUN-3.1.5-01] Specific quality flags will be assigned to all records and will be updated if necessary throughout the processing chain. In the case a spectrum has been rejected, they shall indicate at which step of the process the issue happened.
- [FUN-3.1.5-02] The name and the content of Level-1 data files have to be validated in order to be catalogued in the Level-1 data base. These tests will consist not only in a verification of the format (for example, the presence and the validity of main fields) but also a check of the level of the signal.
- [FUN-3.1.5-03] Standardized and automated QA tests will be performed to validate the output at each step of the process. The instrument configuration files shall contain threshold values on a minimum number of variables for QA/QC tests. These variables should be selected in order to ensure robust QA/QC tests.
- [FUN-3.1.5-04] In case of the detection of anomalies and according to the type of issues (warning, fatal error), the system shall be able to continue the processing chain with eventually flagged data or stop it on fatal error.

3.1.6 Configuration

[FUN-3.1.6-01] The configuration of the system will consist in a collection of files specific to the modules and/or to the instruments. The format shall be selected in order to offer a high level of modularity, flexibility and to allow possible extension of options with needs in the future without revising the existing content.


3.1.7 Monitoring

Problems in the processing chain could come from different sources, for example :

- Warnings or errors returned by a scientific module (threshold(s) exceeded...)
- QA/QC procedure failed
- Missing data, bug in a module
- Technical problems (network issues, machine out of service...)
- ...

At all steps, different mechanisms shall be installed in order to detect and identify problems in the processing chain and to fix them as quickly and efficiently as possible.

- [FUN-3.1.7-01] Tools shall be developed to browse and cross-check the large amount of individual log files generated by the different modules, extract the minimum information helpful for tracking and eventually fixing an issue, deliver processing reports.
- [FUN-3.1.7-02] Alert messages shall be delivered to the operators and the data submitter in case of severe errors reported by one of the processing module or if a QA/QC procedure fails on a file.
- [FUN-3.1.7-03] The system shall be also able to detect stop of data stream.
- [FUN-3.1.7-04] Some indicators might be useful for further statistics (number of processed files per day, evolution of the data base, average processing time for a module, maximum number of

	<p align="center">ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM4DOAS)</p>	<p align="right">Page 12 30 January 2018 Technical Requirements Document</p>
---	--	--

simultaneous processes,...). As far as possible, the system shall anticipate the information to report in log files.

3.1.8 Dashboards

[FUN-3.1.8-01] The processing status of the submitted files will be available with dashboards available through web pages. Requests will be performed with web APIs. The format of exchanged data will be JSON. Regular reports (weekly or monthly) could be sent by mail to the data submitters.

3.2 Operational requirements.

[OPE-3.2-01] The MAXDOAS processing chain shall be fully automated but is not intended to be an operational system *stricto sensu*. According to the project requirements, it will be a demonstration system that will be maintained on a best-effort basis according to the availability of people.

3.3 Resources requirements

3.3.1 Compute servers

[RES-3.3.1-01] In order to increase the performances of the data processing, the system shall be designed in order to exploit as much as possible the execution in parallel of the different jobs.

3.3.2 FTP server

[RES-3.3.2-01] A dedicated FTP server will be available at BIRA-IASB with incoming for Level-1 radiance data. Rules shall be setup in order to control accesses.

[RES-3.3.2-02] Instead of a FTP repository, Level-2 files could be delivered through dedicated web pages where users would fill a form to select the data and would receive a mail with an URL to download a zip file with the requested files.


3.3.3 Dynamic web pages

[RES-3.3.3-01] To present dashboards with the status of submitted files, web pages shall be linked to the different databases and logbooks. Web pages will be developed with Joomla and will include web APIs for databases requests.

3.3.4 S/W requirements :

[RES-3.3.4-01] All the modules that compose the MAXDOAS processing system shall be open source and developed in languages that do not require licenses : C/C++, Fortran, Python, bash.

[RES-3.3.4-02] Scientific modules might be adapted/improved for the automatic processing. A collection of additional tools shall be developed to make the link between the scientific modules, for QA/QC procedures, to generate alerts.

 <p>BIRA-IASB</p>	<p>ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM₄DOAS)</p>	<p>Page 13 30 January 2018 Technical Requirements Document</p>
--	--	--

3.4 Security requirements.

3.4.1 FTP/HTTP

[SEC-3.4.1-01] A different password will be attributed per institute to upload Level-1 data on FTP.

[SEC-3.4.1-02] Users who want to download Level-2 data shall register first.

3.4.2 Backup of data and code

[SEC-3.4.2-01] Regular backup mechanisms shall be set up in order to avoid any loss of data (redundancy, magnetic bands) and to keep the history of the code.

3.4.3 Dashboards web pages

[SEC-3.4.3-01] Data submitters shall access to password-protected web pages related only to their instruments.

3.5 Quality requirements

3.5.1 Modularity and flexibility

The processing chains shall be modular and flexible enough.

[QUA-3.5.1-01] Changes in the code of one module shall not affect the operation of the processing chain

[QUA-3.5.1-02] The system shall be able to manage different versions of modules

[QUA-3.5.1-03] It shall be possible to restart an interrupted process from the step where the interruption occurred or a previous step.

[QUA-3.5.1-04] It shall be possible to reprocess the partial or total data base (selection of instruments or a measurement period).

[QUA-3.5.1-05] The system shall be able to test releases of existing programs or new algorithms in parallel with the automatic operation of existing ones

To reach these objectives, it is important to think about a good organization of the configuration files and to minimize dependencies of the modules.


3.5.2 Performance

[QUA-3.5.2-01] The automatized processing chain shall deliver Level-2 output files between 6 hours and 24 hours after the submission of spectra.

[QUA-3.5.2-02] If the process cannot successfully terminate, alerts shall sent to the appropriate people who will investigate and try to fix the issue in a reasonable delay w.r.t. their availability (workload, days off)

3.5.3 Scalability

[QUA-3.5.3-01] The demonstration of the processing system will be based on a selection of 11 sites available from project partners. The sites selected for the integration are listed in Table 6 of the Technical Proposal document.

 <p>BIRA-IASB</p>	<p>ESA Contract No. 4000118181/16/I-EF Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations (FRM4DOAS)</p>	<p>Page 14 30 January 2018 Technical Requirements Document</p>
--	---	--

[QUA-3.5.3-02] The system shall be designed to be able to ingest more instruments and to support other products in a transparent way in the future.

3.5.4 Portability

[QUA-3.5.4-01] The MAXDOAS processing system shall be designed for Linux machines only.

[QUA-3.5.4-02] The configuration files shall be also in a portable format (plain text, XML...)

3.5.5 Availability

The MAXDOAS processing system shall work continuously but for some reason it might need to be stopped and made not available for a selection of instruments or all instruments :

- Detection of a problem that affects seriously one or several modules or the data bases;
- Update of the operating systems that needs a restart of the machine
- Update of a module or a S/W used by the system

[QUA-3.5.5-01] It shall be possible to stop/restart the whole system with a general flag.

[QUA-3.5.5-02] It shall be possible to stop/restart just a module with a flag.

[QUA-3.5.5-03] It shall be possible to suspend the process at any step for one instrument.

[QUA-3.5.5-04] The system shall be able to detect the best server to execute a job in order to optimize the processing time and be as available as possible and not to saturate only one machine.

[QUA-3.5.5-05] The system shall be available for reprocessing of a given data set from and until the desired modules as far as input files have the required format and after providing the appropriate configuration files.

3.5.6 Reliability requirements.

[QUA-3.5.6-01] The integrity of the data base shall be kept in case of stop/restart of one or several modules or the reprocessing of a data set.

3.5.7 Maintainability requirements.

[QUA-3.5.7-01] The MAXDOAS processing system will be fully maintained and continuously improved until the end of the project. This will include retrieval algorithm maintenance and optimization (through interaction with algorithm designers), bug fixes, adjustments of the code, etc.

[QUA-3.5.7-02] A technical documentation of the MAXDOAS processing system will be provided at the end of the project. A more general description will be provided to people who want to join the network. The code is open source and will be documented as far as possible.