Atmospheric Data Access for the Geospatial User Community

ADAGUC Data Products Standard

http://adaguc.knmi.nl

Document Version : 1.0
Data Format Version : 1.0

Summary:
This document describes the standard that is used for all data products that are created, used and disseminated within the ADAGUC project. It includes a filename convention, required metadata and usable data structures for raster and vector data. The metadata follow the NetCDF Climate and Forecast Metadata Conventions. All data products are stored in HDF5 format created by the netCDF-4 API.

Delivery Slip

	Name	Institute/Company	Date	Signature
Prepared	S. de Witte	SRON	03/10/08	
Approved	B. A. Domenico S. Nativi O. V. Wilhelmi	Unidata IMAA NCAR	03/10/08	
Accepted	J. van de Vegte	KNMI	03/10/08	

Authors

Name	Institute/Company	Contact
S. de Witte	SRON	S.de.Witte@sron.nl
R. van Hees	SRON	R.M.van.Hees@sron.nl
M. Plieger	KNMI	Maarten.Plieger@knmi.nl
R. Sluiter	KNMI	Raymond.Sluiter@knmi.nl
W. Som de Cerff	KNMI	sdecerff@knmi.nl
J. van de Vegte	KNMI	vegtevd@knmi.nl
E. de Vreede	KNMI	Ernst.de.Vreede@knmi.nl
H. Manders	KNMI	Hans.Manders@knmi.nl
R. de Jeu	VU	richard.de.jeu@falw.vu.nl
N. de Reus	VU	nils.de.reus@ivm.vu.nl

Acknowledgement
The authors of this document like to thank B.A. Domenico (Unidata), S. Nativi (IMAA) and O.V. Wilhelmi (NCAR) for their contributions.

Document Log

Issue	Date	Comment	Author/Partner
0.1	13/02/08	Initial Release	S. de Witte
0.2	29/02/08	Input from RvH, MP, RdJ added	S. de Witte et al.
0.3	05/03/08	Input from KNMI added	
0.3.1	12/03/08	Edits from SRON after KNMI input	S. de Witte et al.
0.4	21/03/08	Updates after ADAGUC meeting (13/03/08)	S. de Witte et al.
0.5	31/03/08	Updates to Chapter 1 to 4	S. de Witte et al.
0.6	10/04/08	Updates after ADAGUC meeting (02/04/08)	S. de Witte et al.
0.7	20/06/08	Updates after ADAGUC meetings @ EGU (14/04/08) and SRON (22/05/08)	S. de Witte et al.
0.8	15/08/08	Updates after input from B.A. Domenico, S. Nativi and O.V. Wilhelmi and further discussions during the ADAGUC meeting at KNMI (06/08/08)	S. de Witte et al.
1.0	02/10/08	Final updates for the first official release	S. de Witte et al.

Document Change Record

Issue	Item	Reason for Change
0.2		
0.3		Additions and structure change
0.3.1		See Document Log
0.4		See Document Log
0.5		See Document Log
0.6		See Document Log
0.7		See Document Log
0.8		See document Log
1.0		First official release

Copyright notice and disclaimers

Copyright

1. All information in this document is copyrighted and may not be used except as provided in this copyright notice or other proprietary notice provided with the relevant materials.

2. The information contained in this document may be downloaded, printed or copied provided that ALL copies retain the copyright and any other proprietary notices contained on the materials. No material may be modified, edited or taken out of context such that its use creates a false or misleading statement or impression as to the positions, statements or actions of the ADAGUC Consortium (consisting of KNMI [Royal Nederlands Meteorological Institute], SRON [Netherlands Institute for Space Research], VU-FALW [Vrije Universiteit Amsterdam, Faculty of Earth and Life Sciences] and WUR [Wageningen University and Research Centre]).

3. The names of copyright holders may NOT be used in advertising or publicity pertaining to this document, its content, specifications, or software without specific, written prior permission of the copyright holder. Title to copyright in such document will at all times remain with copyright-holders.

4. The ADAGUC Consortium is not responsible for the content of any references to this document. Please understand that any referring document is independent from the ADAGUC Consortium, and the ADAGUC Consortium has no control over the content of such document. In addition, when a document refers to this document, this does not mean that the ADAGUC Consortium endorses or accepts any responsibility for the content, or the use, of such document.

5. The information the ADAGUC Consortium published in this document may contain references or cross references to the ADAGUC Consortium specifications, projects, programs and services that are not announced or available in your country.

6. The information in this document may contain technical inaccuracies or typographical errors. Information may be changed or updated without notice. The ADAGUC Consortium may make improvements and/or changes in the materials contained in or described in this document at any time without notice. The ADAGUC Consortium may also make changes in this copyright notice without notice. User is bound by such revisions and should therefore periodically check for said revisions to review the then current copyright notice.

Limitation on Warranties

ALL INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS." THE ADAGUC CONSORTIUM MAKES NO REPRESENTATIONS OR WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE OR NON-INFRINGEMENT. AS TO INFORMATION PUBLISHED IN THIS DOCUMENT, THE ADAGUC CONSORTIUM MAKES NO REPRESENTATION OR WARRANTY THAT THE CONTENTS OF THIS DOCUMENT IS FREE FROM ERROR OR SUITABLE FOR ANY PURPOSE; NOR THAT IMPLEMENTATION OF SUCH INFORMATION WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

Please note that some jurisdictions may not allow the exclusion of implied warranties, so some of the above exclusions may not apply to you.

Limitation on Liability

IN NO EVENT WILL THE ADAGUC CONSORTIUM BE LIABLE TO ANY PARTY FOR ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES FOR ANY USE OF THIS DOCUMENT, INCLUDING, WITHOUT LIMITATION, ANY LOST PROFITS, BUSINESS INTERRUPTION, LOSS OF PROGRAMS OR OTHER DATA ON YOUR INFORMATION HANDLING SYSTEM OR OTHERWISE, EVEN IF THE ADAGUC CONSORTIUM IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Contents

1 Introduction

1.1 Background

Many tools and data formats exist for atmospherical data. To disseminate this wealth of information to the geospatial communities is still very difficult. It is complicated to easily share data among scientists representing the geospatial communities without performing some cumbersome conversions. ADAGUC aims to reduce the need for scientists to invent their own converter tools. Selected space borne atmospheric and land datasets will be made accessible to a Geographic Information System (GIS) for data comparison, resampling, selection, manipulation and visualization. The user community is intensively involved in the project for the definition of the use cases.

The deliverables of the ADAGUC project are: Open Source conversion tools, selected atmospheric datasets in GIS format and web services to provide the datasets and to demonstrate the usability of the methodology to the geospatial and atmospheric community. Dissemination of results is pursued by publications, workshops and (inter)national cooperation.

1.2 Overview

This document builds further on the User Requirements Document (URD) [A05] and the Software Specification Document (SSD) [A06].

The content of this document:

- Chapter 1, introduction.
- Chapter 2, relevant user requirements derived from the URD and relevant boundary conditions derived from the SSD.
- Chapter 3, filename conventions.
- Chapter 4, file format standard of the HDF5/netCDF-4 format adopted in the ADAGUC project.
- Chapter 5, import and export conversions.
- Chapter 6, units used within the ADAGUC project.
- Chapter 7, coordinate systems.
- Chapter 8, limitations, open issues and future challenges.

1.3 Purpose and scope

This document describes the standards that are applicable to the ADAGUC data products and relevant for the atmospheric and geospatial user communities. In particular, it describes the file name conventions and (meta)data standards that will be used within the ADAGUC project. Furthermore, it describes the file formats of netCDF-4 data products, extended and implemented on top of the HDF5 data format, adopted in the ADAGUC project. And it describes units and coordinate systems used in the ADAGUC project.

1.4 Standards applicability

The conventions and standards described in this document must be followed for all data product files within the ADAGUC systems and should be considered for data product files originating from the ADAGUC systems. The application of these standards will be strictly enforced for all systems developed under ADAGUC

responsibility, and will be encouraged as far as possible for other systems, not under ADAGUC responsibility.

1.5 Applicable documents

A01 International Organization for Standardization (ISO), *http://www.iso.org*
A02 OGC, Open Geospatial Consortium, Inc.®, *http://www.opengeospatial.org*
A03 Dutch Geoservices Profiles, *http://www.geonovum.nl/standaarden.html*
A04 INSPIRE, Infrastructure for Spatial Information in the European Community, *http://www.ec-gis.org/inspire*
A05 ADAGUC User Requirements Document (URD), V1.3, *http://adaguc.knmi.nl*
A06 ADAGUC Software Specification Document (SSD), V0.1, *http://adaguc.knmi.nl*
A07 Dutch metadata profile for geographical data, version 1.1, 2 juni 2006, *http://www.geonovum.nl/informatiemodellen/nen3610*
A08 Dutch metadata profile for services, *http://www.geonovum.nl/services/nieuws---services*
A09 International Standard ISO 19115:2003, *http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020*
A10 International Standard ISO 8601:2004, *http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40874*
A11 International Standard ISO 639-2:1998, *http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=4767*

1.6 Reference documents

R01 HDF5: File Format Specification, Version 2.x, *http://www.hdfgroup.org/HDF5/doc/H5.format.html*
R02 WMO, World Meteorological Organization, *http://www.wmo.int*
R03 GALEON project, *http://www.unidata.ucar.edu/projects/THREDDS/GALEON/Home.html*
R04 UMN Mapserver, *http://mapserver.gis.umn.edu*
R05 GDAL, *http://www.gdal.org*
R06 NetCDF, *http://www.unidata.ucar.edu/software/netcdf*
R07 NetCDF Climate and Forecast (CF) Metadata Conventions, version 1.2 (May 2008), *http://cf-pcmdi.llnl.gov*
R08 PyTables - Hierarchical Datasets in Python, *http://www.pytables.org*
R09 KNMI HDF5 Data Format Specification, v3.5., *http://www.knmi.nl/publications/showAbstract.php?id=368*
R10 HDF5 Dimension Scale Specification and Design Notes, version 1.0, *http://www.hdfgroup.uiuc.edu/HDF5/doc_1.8pre/supplements/H5DimScales*
R11 Geonovum, *http://www.geonovum.nl*
R12 Framework of standards for the Dutch GII, version 1.1, 14 juni 2006, *http://www.rgi.nl/downloads/files/framework%20of%20standards%20v1.1.pdf*
R13 World Meteorological Organization (WMO) keyword list, *http://www.wmo.int/pages/prog/www/metadata/WMO-keywords.html*
R14 SCIAMACHY, *http://envisat.esa.int/instruments/SCIAMACHY*
R15 PROJ.4 - Cartographic Projections Library, *http://trac.osgeo.org/proj*
R16 Netherlands SCIAMACHY Data Center (NL-SCIA-DC), *http://neonet.knmi.nl/neoaf*
R17 THREDDS project, *http://www.unidata.ucar.edu/projects/THREDDS*
R18 NASA-EOS data product levels, *http://observer.gsfc.nasa.gov/sec3/ProductLevels.html*
R19 HDF5: API Specification Reference Manual, *http://hdfgroup.org/HDF5/doc/RM/RM_H5Front.html*
R20 UDUNITS: Unidata units library, *http://www.unidata.ucar.edu/software/udunits/udunits-1/index.html*
R21 CF Standard Name Table, version 9, 10 June 2008, *http://cf-pcmdi.llnl.gov/documents/cf-standard-names*
R22 Understanding Map Projections, ArcInfo™ 8, Melita Kennedy, Environmental Systems Research Institute, Inc., 2000, *http://www.kartografie.nl/geometrics/Map%20projections/Understanding%20Map%20Projections.pdf*
R23 UMN Mapserver Filter Encoding (FE), *http://mapserver.gis.umn.edu/docs/howto/filterencoding*

1.7 Abbreviations and acronyms

ADAGUC	Atmospheric Data Access for the Geospatial User Community
AMSR	Advanced Microwave Scanning Radiometer
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
BIL	Band Interleaved by Line
BIP	Band Interleaved by Pixel
BSQ	Band SeQuential
CF	Climate and Forecast
ECMWF	European Centre for Medium-Range Weather Forecasts
EOS	Earth Observing System
ESA	European Space Agency
FRESCO	Fast Retrieval Scheme for Cloud Observables
GALEON	Geo-interface to Atmosphere, Land, Earth, Ocean, NetCDF
GDAL	Geospatial Data Abstraction Library
GIS	Geographical Information System
GOME	Global Ozone Monitoring Experiment
GRD	ESRI Grid format
GRIB	GRIdded Binary
HDF	Hierarchical Data Format
IMAA	Institute of Methodologies for Environmental Analysis
IMAU	Institute for Marine and Atmospheric Research, Utrecht
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
IUP	Institute of Environmental Physics, University of Heidelberg
JPG	Joint Photographic Experts Group Interchange Format
KML	Keyhole Markup Language
KNMI	Royal Nederlands Meteorological Institute
LatLon	Latitude, Longitude
LPRM	Land Parameter Retrieval Model
MERIS	Medium-spectral Resolution Imaging Spectrometer
NCAR	National Center for Atmospheric Research, Boulder CO
NetCDF	Network Common Data Form
NRT	Near Real Time
OGC	Open Geospatial Consortium
OMI	Ozone Monitoring Instrument
PNG	Portable Network Graphics
QoS	Quality of Service
RD	Rijksdriehoeksmeting (Dutch Grid)
RGI	Ruimte voor Geo-Informatie
RIVM	Rijksinstituut voor Volksgezondheid en Milieu
RT	Real Time
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY
SHP	ESRI Shapefile
SSD	Software Specification Document
SRON	SRON Netherlands Institute for Space Research
TBD	To Be Determined
TBW	To Be Written
THREDDS	Thematic Realtime Environmental Distributed Data Services
TIFF	Tagged Image File Format

UCAR	University Corporation for Atmospheric Research
UMN	University of Minnesota
URD	User Requirements Document
UTC	Coordinated Universal Time
UTM	Universal Transverse Mercator
VU-FALW	Vrije Universiteit Amsterdam, Faculty of Earth and Life Sciences
WCS	Web Coverage Service
WFS	Web Feature Service
WMO	World Meteorological Organization
WMS	Web Map Service
WUR	Wageningen University and Research Centre
XML	eXtensible Markup Language

1.8 Terminology

Dataset (ISO)	A collection of one product kind (same source/type). E.g. the SCIAMACHY [R14] daily average ozone. In this document when referring to the ISO term [A01], 'Dataset (ISO) ' will be used, otherwise the HDF5 definition for dataset is meant.
Dataset (HDF5)	In the HDF5 data model 'dataset' is used for a multi-dimensional array of data elements, together with supporting metadata [R01].
Product	One file from an ISO dataset containing data
Metadata	Description on ISO dataset level according to the ISO19115 standard [A09]. There will also be metadata information on measurement level, mostly based by the netCDF-CF standard [R07].
Regular grid	A grid with square pixels with fixed X and Y dimensions
Vector data	A polygon described by 4 points (tile)
NetCDF	A set of interfaces for array-oriented data access and a freely-distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data.
PyTables File	HDF5 file which includes at least mandatory PyTables attributes
PyTables Leaf	HDF5 dataset

2 User requirements and boundary conditions

2.1 Introduction

The user requirements of the ADAGUC system are described in [A05], the relevant user requirements for this document are given in paragraph 2.2, and additional boundary conditions derived from the SSD are given in paragraph 2.3.

The ADAGUC project has both a national and international focus. To minimize repeating tasks the (inter)national developments are monitored closely and form the basis of this document. In the Netherlands geo-standards are coordinated by Geonovum [R11], especially the Dutch metadata profiles for geographical data [A07] and services [A08] are important. In international context the developments by the GALEON project [R03] which focuses on netCDF [R06] and CF Conventions [R07], and the developments by OGC [A02], WMO [R02] and ISO [A01] are closely followed.

2.2 User requirements

In this section the user requirements relevant for this document are summarized.

UR-SOW-1	ADAGUC services will be built using open standards
UR-SOW-2	ADAGUC services will use open source software solutions where practically possible.
UR-SOW-3	ADAGUC will make maximum use of existing NL-SCIA-DC solutions
UR-SOW-4	ADAGUC services shall be compliant with international standards (ISO [A01], OGC [A02])
UR-SOW-5	ADAGUC services shall be “INSPIRE ready”
UR-SOW-6	ADAGUC services shall be compliant with the Dutch Geoservices profiles [A03]
UR-SOP-2	ADAGUC will provide OGC compliant web service access to datasets
UR-FUDA-8	All created software shall comply with the RGI guidelines (OGC, Geonovum).
UR-FUDA-16	ADAGUC shall deliver data in a geographic coordinate system using Latitude-Longitude with WGS84 datum
UR-FUDA-17	ADAGUC should be able to deliver data in the ERTS89/UTM coordinate system (INSPIRE)
UR-FUDA-18	ADAGUC should be able to deliver data in the RD (Rijksdriehoek) coordinate system.
UR-FUDA-26	ADAGUC shall import data in the ASCII format
UR-FUDA-27	ADAGUC shall import data in the HDF-4 format
UR-FUDA-28	ADAGUC shall import data in the HDF-5 format
UR-FUDA-29	ADAGUC shall import data in the netCDF format
UR-FUDA-30	ADAGUC shall import data in the BIL format
UR-FUDA-31	ADAGUC shall import data in the GeoTIFF format
UR-FUDA-32	ADAGUC shall import data in the GRD format
UR-FUDA-33	ADAGUC shall import data in the GRIB format
UR-FUDA-34	ADAGUC shall export data in the GRD format
UR-FUDA-35	ADAGUC shall export data in the netCDF format
UR-FUDA-36	ADAGUC shall export data in the HDF-5 format
UR-FUDA-38	ADAGUC shall export data in the GeoTIFF format
UR-FUDA-39	ADAGUC shall export data in the SHP (Shapefile) format

UR-FUDA-40	ADAGUC shall export data in KML (Google Earth) format
UR-FUDA-41	ADAGUC shall use the HDF-5 format as the common format for storage
UR-FUDA-42	Metadata shall be provided conform the INSPIRE guidelines
UR-FUDA-43	Metadata shall be provided conform the Dutch Geoservices profiles
UR-FUDA-44	ISO19115 [A09] shall be used to document product metadata
UR-INTF-1	A uniform filename convention for all ADAGUC datasets shall be used
UR-INTF-2	The ADAGUC filename convention shall be based on existing international initiatives (ISO, OGC).
UR-INTF-3	All internet services from ADAGUC shall be OGC compliant.
UR-INTF-4	All datasets shall use the appropriate WMO [R02]/OGC/CF [A02] standard for variable names
UR-INTF-5	All datasets shall use the ISO standard for metrics
UR-INTF-6	The ADAGUC web services shall be interoperable with international initiatives, like GALEON
UR-INTF-7	The ADAGUC system shall be interoperable with the NL-SCIA-DC system
UR-INTF-8	Dataset metadata shall be exchangeable between different instances of ADAGUC (NL-SCIA-DC and VU-FALW)
UR-QUA-1	Quality of datasets shall be documented and described using the appropriate metadata fields.
UR-QUA-2	Dataset quality information shall be provided by the dataset provider

2.3 Boundary conditions

From the SSD [A06] some extra boundary conditions can be derived. These boundary conditions may impose extra requirements for filename conventions and the formats/description of metadata.

GDAL:
ADAGUC will use the GDAL library [R05] to enable read and write access of non-native data formats. GDAL is widely accepted as the standard library for geo-based file conversions and used in OGC compliant services like UMN Mapserver [R04].

HDF5:
The HDF5 format is used to store all the ADAGUC products. The data products will follow the HDF5 1.8.x specification [R01], which includes support for dimension scales that provide an elegant solution for data presentation.

NetCDF:
Many users of ADAGUC use netCDF. Therefore it should be possible to use the netCDF API to generate and read ADAGUC products with netCDF. The netCDF version 4 API will be used within the ADAGUC project, because it is implemented on top of the HDF5 library, making it possible to create HDF5 and NetCDF files with the same API.

PyTables:
Many users of ADAGUC use Python. PyTables improves the accessibility from Python. PyTables imposes some additional metadata fields [R08].

Regular grids:
The support of grids is limited to grids with square pixels with fixed X and Y dimensions.

Vector data:
Vector data is defined as a polygon described by 4 points (a tile). The ADAGUC Data Products Standard will not deal with other line or vector structures.

Time Zone:
All date/time fields, as used in the filenames or stored as fields/atributes in de files itself, will always be stated in the Coordinated Universal Time (UTC) timezone.

Open Source:
To be able to freely distribute the ADAGUC tools, any program used and/or developed within the ADAGUC project shall be Open Source.

3 Filename convention

3.1 Introduction

The ADAGUC project will serve many different data products from different sources. Data products from different sources will be converted to the ADAGUC file format and stored in a central location. The filename convention should provide the information to make a clear distinction between the data products. The general directive of this filename convention is to keep the filename as short as possible, while still be able to uniquely identify the different products. Both internally stored products, as well as exported products from the webservices will follow the filename convention described in paragraph 3.2.

3.2 Data product filename format

The data product files shall be named using a fixed set of elements, each of fixed size, separated by underscores "_". These elements constitute a smaller set of information which ensure that each filename is unique, within the context of the ADAGUC project. See Table 3-1 for an overview. If this set of information is smaller than the elements fixed length it will be appended with underscores "_" to make sure that all filenames are of fixed length. An exception to this latter rule is the extension (.XXX). This element does not need to be appended with underscores "_" if the extension is smaller than three characters.

The general format that applies to all ADAGUC data product files is:

```
MMMMM_CCCC_T_F_PPPPPPP_LLL_------------instance-id------------.XXX
```

where

- `MMMMM` is a five-letter mission, instrument or model identifier (see paragraph 3.3),
- `CCCC` is the file class (OPER, CONS, TEST, etc.),
- `T` is the file type (V for vector data, R for raster data, I for in situ data and O for other),
- `F` is a flag for exported products (via the ADAGUC webservices) indicating whether the original internally stored ADAGUC product has been customized (resampled, resized or changed in any other way); if so, use a capital letter C, otherwise an underscore "_",
- `PPPPPPP` is the product acronym (see paragraph 3.3),
- `LLL` is the level of the data in the product starting with a capital letter L (for instance L1B, L2 or L3)
- `------------instance-id------------` is the instance ID that uniquely identifies one instance of a data product; the last four characters (digits only) of the instance ID always indicate the version number of the product (`vvvv`), starting at "`0001`"; the version is separated from the rest of the instance ID by an underscore "_",
- `.XXX` is the extension (for instance .nc, .h5, .tif).

The instance ID has the following format:

`yyyymmddThhmmss_YYYYMMDDTHHMMSS_vvvv`

where

- `yyyymmddThhmmss` is the validity start date/time in UTC or the creation date/time in UTC. If the first and second date/time in the instance ID (`yyyymmddThhmmss` and `YYYYMMDDTHHMMSS`) differ, `yyyymmddThhmmss` is the validity start date/time. If they are equal, it is the creation date/time. In case the validity period instance ID format is used, it can have the special value `00000000T000000` for beginning of mission, or if a validity period is not applicable (conform ISO8601 [A10]),
- `YYYYMMDDTHHMMSS` is the validity stop time in UTC or the creation date/time in UTC. If the first and second date/time in the instance ID (`yyyymmddThhmmss` and `YYYYMMDDTHHMMSS`) differ, `YYYYMMDDTHHMMSS` is the validity stop date/time. If they are equal, it is the creation date/time. In case the validity period instance ID format is used, it can have the special value `99999999T999999` for end of mission, or if a validity period is not applicable,
- `vvvv` is the version number as defined above.

Filename element		**Element number**	**Regular Expression** (without separating underscores)	**Length** (without separating underscores)
Mission/Instrument/Model		0	(?=^.{5}$)(^[A-Z0-9]{1,5}_{0,4}$)	5
File class		1	^[A-Z0-9]{4}$	4
Product type		2	^[RVIO]$	1
Custom flag		3	^[C_]$	1
Product acronym		4	(?=^.{7}$)(^[A-Z0-9]{1,7}_{0,6}$)	7
Product level		5	^L[0-4][A-I_]$	3
Instance ID	Creation Date/Time, Validity Period	6a	^[0-9]{4}[019][0-9][0-39][0-9]T[0-29][0-9][0-69][0-9]{3}_[0-9]{4}[019][0-9][0-39][0-9]T[0-29][0-9][0-69][0-9]{3}$	31
	Version	6b	^[0-9]{4}$	4
Total (with separating underscores, without extension)				63

Table 3-1 Filename elements overview

Below are some examples of internally stored filenames:

```
SCIA__OPER_R___TMTNO2__L3__20060101T000000_20060131T235959_0001.nc
AMSR__OPER_R___LPRMSMD_L3__20070704T133000_20070704T133000_0003.nc
```

And examples of exported filenames (via the ADAGUC webservices), in case resampling, resizing or any other change of the data (compared to the internally stored product) took place:

```
SCIA__OPER_R_C_TMTNO2__L3__20060101T000000_20060131T235959_0001.nc
AMSR__OPER_R_C_LPRMSMD_L3__20070704T133000_20070704T133000_0003.nc
```

Note that, apart from the custom flag, only the validity period and extension for exported products could be different from the internally stored product. All other filename elements (including the version) should remain the same.

3.3 Missions, Instruments, Models, Datasets and Products

The following tables describe the situation at the time of writing this document. The latest list of possible missions, instruments, models, datasets and products can be found at the ADAGUC website (http://adaguc.knmi.nl).

Table 3-2 lists possible missions, instruments and models and the abbreviations / acronyms which should be used in the filename (filename element 0). Table 3-3 describes possible datasets and the abbreviations / acronyms which should be used as variable names within the data product and, in case of a product with a single dataset, in the filename as well (filename element 4). Table 3-4 lists possible combinations of datasets which can be combined in one data product. In that case, the abbreviations / acronyms from Table 3-3 are used as variable names within the data product, while the abbreviations / acronyms from Table 3-4 are used in the filename (filename element 4).

Mission/Instrument/Model	**Abbreviation / Acronym** (filename element 0)
SCIAMACHY	SCIA
ECMWF	ECMWF
MERIS	MERIS
AMSR	AMSR
OMI	OMI
LPRM	LPRM

Table 3-2 Abbreviations / Acronyms of possible missions, instruments and models

Dataset	**Abbreviation / Acronym** (variable name and, in case of a single dataset product, filename element 4)
Total Daily Tropospheric NO2	TDTNO2
Total Daily Column NO2	TDCNO2
Total Monthly Tropospheric NO2	TMTNO2
Total Monthly Column NO2	TMCNO2

Relative Humidity	RELHUM
Land Parameter Retrieval Model Soil Moisture (from the Descending swaths)	LPRMSMD
Land Parameter Retrieval Model Soil Moisture (from the Ascending swaths)	LPRMSMA
Daily cloud pressure (FRESCO)	CLOUDPR
Daily cloud fraction (MERIS: SciaMAX, Sciamachy: FRESCO)	CLOUDFR
Volumetric soil water layer 1	SWVL1
Soil surface temperature layer 1	STL1
Large Scale precipitation (accumulated)	LSP
Convective precipitation (accumulated)	CP
Boundary layer height	BLH
10 metres wind U-component	10U
10 metres wind V-component	10V
2 metres air temperature	2T
Total precipitation (accumulated)	TP
Averaged monthly column CH_4	AMCCH4
Total Ozone with SCIAMACHY algorithm	TOSOMI
Total Ozone with GOME algorithm	TOGOMI

Table 3-3 Abbreviations / Acronyms of possible datasets

Combined Datasets	**Abbreviation / Acronym** (filename element 4)
TDTNO2, TDCNO2	TDNO2
TMTNO2, TMCNO2	TMNO2
CLOUDPR, CLOUDFR	FRESCO
LPRMSMD, LPRMSMA	LPRMSM

Table 3-4 Abbreviations / Acronyms of possible combinations of datasets, which can be stored in a single data product (i.e. in one data file). See Table 3-3 for descriptions of the individual datasets.

4 Internal format standard

4.1 Justification

Based on user requirement UR-FUDA-41, the Hierarchical Data Format Version 5 (HDF5) [R01] is the common format for storage within the ADAGUC project. HDF implements a model for managing and storing data. The model includes an abstract data model and an abstract storage model (the data format). In addition it has libraries to implement the abstract model and to map the storage model to different storage mechanisms. The HDF5 library [R19] provides a programming interface to a concrete implementation of the abstract models. Furthermore, the library implements a model of data transfer, i.e. efficient movement of data from one stored representation to another stored representation.

There are several good reasons to use the HDF5 data format:

- The HDF5 API provides platform independent data access
- The HDF5 API provides uniform and quick data access, also to very large datasets.
- Many scientific applications recognize HDF5 data products

Disadvantages for novice users are the large number of HDF5 modules involved to read/write data, and complexity of the HDF5 API. For this reason, we have developed the ADAGUC internal file structure (see paragraph 4.2 through 4.4) such that writing or reading an ADAGUC compliant product can be achieved by only using the HDF5 high level API (v1.8) [R19].

Many users from the geospatial community specified HDF5 as the preferred format for scientific data exchange, users from the atmospherical community specified NetCDF as the preferred format for scientific data exchange. Therefore a convenient API is the netCDF-4 API [R06], as NetCDF-4 is extended and implemented on top of the HDF5 data format. With NetCDF-4 one can serve both worlds simultaneously. Therefore the use of the netCDF-4 API [R06] is recommended.

4.2 Internal format description

4.2.1 Introduction

All data products within the ADAGUC project are stored in HDF5 format and can, if needed, be exported in several formats described in paragraph 5.1. In this chapter we provide the current version of the ADAGUC general metadata specification and the ADAGUC data format specification for vector, raster and model based data.

4.2.2 HDF5 File Organization

An HDF5 file is a container file for storing a variety of scientific data and is composed of two primary types of objects: groups and datasets. An HDF5 group is a grouping structure containing HDF5 objects. An HDF5 dataset, or variable, is an object consisting of (a multidimensional array of) data elements. Any HDF5 dataset may have an associated attribute list. This list is a user-defined HDF5 structure that provides extra information about an HDF5 object (metadata). The attributes are described in paragraph 4.3 and 4.4. A general attribute named “Conventions” is

required at the top-level of the product to indicate to which version of the NetCDF Climate and Forecast (CF) Metadata Conventions [R07] the product complies (for instance, "CF-1.2").

4.2.3 ADAGUC HDF5 internal file structure

The ADAGUC internal file structure contains variables which can be subdivided in three types: variables to store the metadata, variables to store the dimension scales and variables to store the actual data. A schematic overview of the ADAGUC file structure for the different types is given in Figure 4-1. The three types of variables (metadata, dimension scales and data) are indicated with A, B and C respectively.

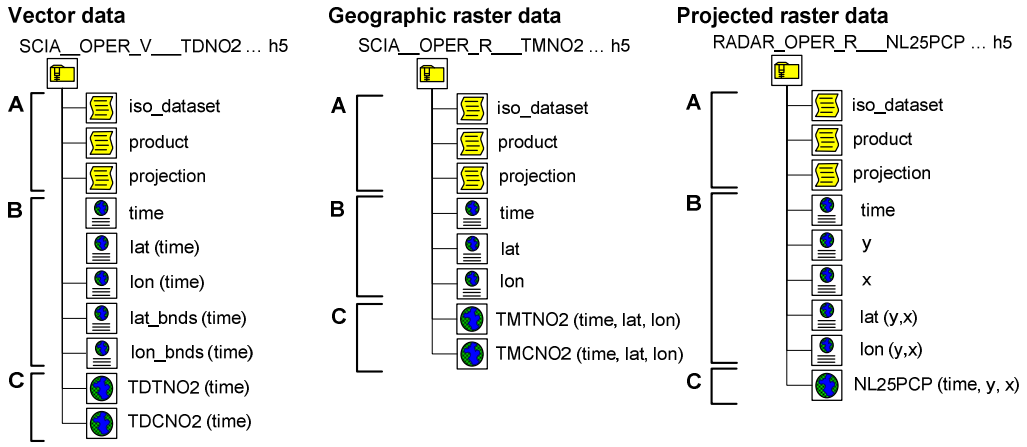
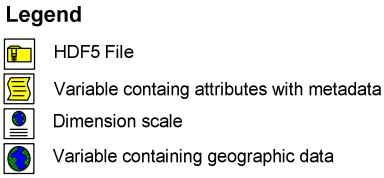


Figure 4-1 Overview of the file structure used in the ADAGUC project. Left: file with vector data. Middle: file with geographic raster data. Right: file with projected raster data. A: the metadata part; B: the dimension scale part; C: the data.



(A) Metadata

Metadata is stored in attributes which are assigned to a variable. The variables 'product', 'iso_dataset' and 'projection' are used to store the metadata. The iso_dataset variable contains global attributes which are valid for all files in the ISO dataset, while the product variable contains specific metadata about the variables within the file. Specific attributes are assigned to all variables. Detailed information about the metadata for each of the variables can be found in paragraph 4.3.

(B) Dimension scales

Dimension scales are used as coordinate variables and time intervals. They provide information about geolocation and time to the data variables. The dimension scales follow the NetCDF Climate and Forecast (CF) Metadata Conventions General Metadata format specification [R07]. Dimension scales are lists containing coordinates or time intervals. Figure 4-2 shows geographic raster data with the dimensions time, lat and lon. Its size is, in this example, 4x6x12. This means that there are 4 indexes for time, 6 for latitude and 12 for longitude. According to the CF conventions the dimensions should have the order time, lat, lon.

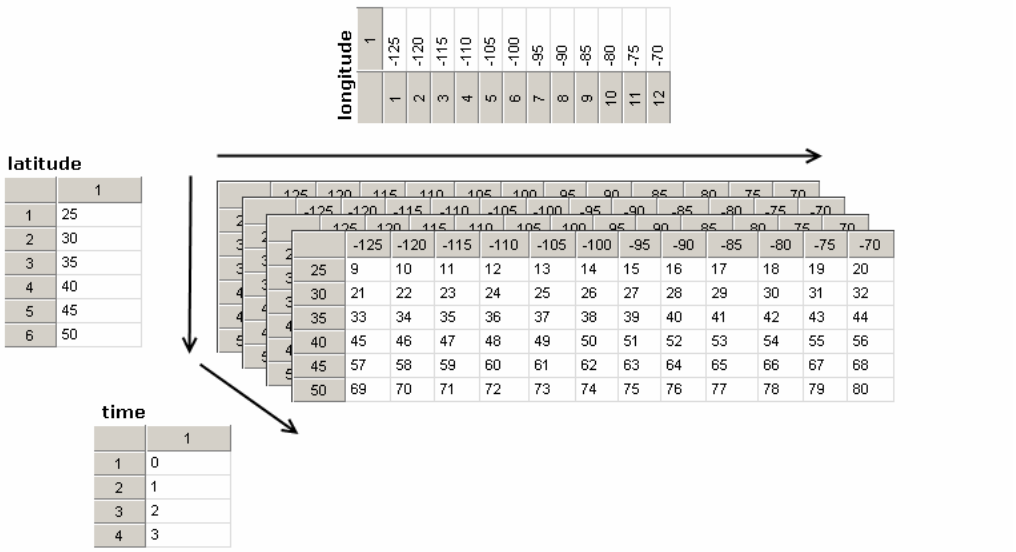


Figure 4-2 Geographic raster data containing three dimensions (time, latitude and longitude) with a size of 4x6x12. The dimension scales are drawn along the axes of the data field.

Dimensions scales for vector data

For vector data the dimension scales are lat, lon, lat_bnds, lon_bnds and time. The lat and lon dimensions represent the center of the tile, while lat_bnds and lon_bnds represent the corners of the tile. These dimensions scales are a function of time, just like the variables TDTNO2 and TDCNO2 in Figure 4-1 (left). The geolocation of a tile and its corresponding value can be read by obtaining its corner coordinates from lat_bnds and lon_bnds and its value from the data variables (for example TDTNO2). The time of the corner coordinates and the time of the data value must be the same.

Dimensions scales for raster data using a geographic coordinate system

In case the raster data uses a geographic coordinate system, the required dimensions are time, lat and lon which represent the center of the pixel. Figure 4-2 shows an example of raster data using a geographic coordinate system.

Dimensions scales for raster data using a projected coordinate system

A projected coordinate system is defined on a flat, two-dimensional surface. When the raster data is projected the dimension scales x and y are included which represent the projected coordinates. The dimension scales lat and lon become a function of y and x (see Figure 4-1, right), providing the latitude and longitude at the location of y and x, respectively. In this case y and x contain the projected coordinates, while lat and lon represent the geographic latitudes and longitudes in degrees at location y, x. For detailed information see the NetCDF Climate and Forecast (CF) Metadata Conventions [R07].

(C) Data

Part "C" contains the data in the file structure. The data is stored in variables which names are defined in Table 3-3. The variables are functions of the dimension scales. In case of vector data the variable is a function of time. For geographic raster data the variable is a function of time, lat and lon. For projected raster data the variable is a function of time, y and x (see Figure 4-1).

4.3 Metadata format specification

4.3.1 Overview of required variables and attributes

The three metadata variables (iso_dataset, product and projection) act as a container for the metadata attributes. The iso_dataset variable itself is a string (see paragraph 4.3.3), the other two variables (product and projection) are of arbitrary type since they do not contain data.

The ADAGUC ISO metadata complies with the ISO19115 standard [A09]. This standard is a content standard, i.e. it defines which elements should be described.
The attributes of the ISO dataset variable form the ISO dataset metadata (Table 4-1). ISO19115 [A09] describes many mandatory and optional fields. To decide which fields should be applied, the "Nederlandse Kernset Metadata" (Dutch ISO metadata profile, [A07]) is defined and used here.

ID	Variable	Attribute Name	Standard
M0	iso_dataset		ADAGUC
M0-0		language	ISO-19115:2003
M0-1		organisationName_metadata	ISO-19115:2003
M0-2		linkage_metadata	ISO-19115:2003
M0-3		role_metadata	ISO-19115:2003
M0-4		datestamp	ISO-19115:2003
M0-5		metadataStandardName	ISO-19115:2003
M0-6		metadataStandardNameVersion	ISO-19115:2003
M0-7		code	ISO-19115:2003
M0-8		codeSpace	ISO-19115:2003
M0-9		title	ISO-19115:2003
M0-10		date	ISO-19115:2003
M0-11		dateType	ISO-19115:2003
M0-12		abstract	ISO-19115:2003
M0-13		status	ISO-19115:2003
M0-14		organisationName_dataset	ISO-19115:2003
M0-15		linkage_dataset	ISO-19115:2003
M0-16		role_dataset	ISO-19115:2003
M0-17		keyword	ISO-19115:2003
M0-18		useLimitation	ISO-19115:2003
M0-19		accessConstraints	ISO-19115:2003
M0-20		spatialRepresentationType	ISO-19115:2003
M0-21		denominator	ISO-19115:2003
M0-22		language_dataset	ISO-19115:2003
M0-23		organisationName_distributor	ISO-19115:2003
M0-24		linkage_distributor	ISO-19115:2003
M0-25		role_distributor	ISO-19115:2003
M0-26		statement	ISO-19115:2003

Table 4-1 Required attributes of the ISO dataset variable, valid for the whole ISO dataset

Table 4-2 gives an overview of the general required metadata for all products in the form of attributes of the product variable and indicates to which standard the attributes comply.

ID	Variable	Attribute Name	Standard
M1	product		ADAGUC
M1-0		title	CF
M1-1		institution	CF
M1-2		source	CF
M1-3		history	CF
M1-4		references	CF
M1-5		comment	CF
M1-6		ref_doc	ADAGUC
M1-7		ref_doc_version	ADAGUC
M1-8		format_version	ADAGUC
M1-9		input_products	ADAGUC
M1-10		originator	ADAGUC
M1-11		creation_date	ADAGUC
M1-12		software_version	ADAGUC
M1-13		file_class	ADAGUC
M1-14		type	ADAGUC
M1-15		acronym	ADAGUC
M1-16		level	ADAGUC
M1-17		variables	ADAGUC
M1-18		validity_start	ADAGUC
M1-19		validity_stop	ADAGUC
M1-20		style	ADAGUC

Table 4-2 Required product metadata, valid for one product

Table 4-3 gives an overview of the required metadata for all products in the form of attributes of the projection variable and indicates to which standard the attributes comply.

ID	Variable	Attribute Name	Standard
M2	projection		ADAGUC
M2-0		projection_name	ADAGUC
M2-1		EPSG_code	ADAGUC
M2-2		proj4_params	ADAGUC
M2-3		grid_mapping_name	ADAGUC

Table 4-3 Required projection metadata (note that an EPSG code does not necessarily have to exist for certain combinations of proj4 parameters, see paragraph 4.3.4 for details).

4.3.2 Additional attributes

Paragraph 4.3.1 describes the mandatory attributes for all ADAGUC products. The following additional attributes are allowed in an ADAGUC HDF5 data product.

Scale and offset:
The preferred way to indicate a scale and an offset is to use the Scale+Offset Filter in HDF5 (from version 1.8 onwards). This filter compresses scalar (integer and floating-point) datatypes which stay within a range [R01].

Alternatively, one could use scale and offset attributes. The scale and offset attributes are used to convert integer values into meaningful physical quantities. Within the netCDF Climate and Forecast Conventions [R07, paragraph 8.1] this is referred to as "Packed Data". The attributes scale_factor and add_offset are used to define the scale and offset factors. To retrieve a physical quantity from an integer value the input value is first multiplied with the scale factor and second the offset factor is added:
Physical Quantity = (byte value * scale_factor) + add_offset
The input value is called the packed data, while the resulting value is called the unpacked data. The packed data type must be of type byte, short or int. The resulting unpacked data type must have the same data type as the scale_factor and add_offset attributes. These attributes are usually of data type float or double.

Projection metadata:
Projection parameters can be added in case of a projected coordinate system, for example a polar stereographic projection. These projection attributes are derived from appendix F of the netCDF Climate and Forecast (CF) Metadata Conventions [R07]. Example attributes are listed in Table 4-7 for a polar stereographic projection.

Any other attributes can only be stored in the data product attached to a variable called "custom", or, alternatively custom attributes can be added anywhere in the data product as long as the attribute prefix is "custom_" (i.e. the attribute name starts with "custom_").

4.3.3 ISO Dataset metadata

The ISO dataset variable holds the general metadata valid for the whole ISO dataset. It should be equal for all files belonging to a particular ISO dataset. The ISO dataset variable itself holds the version of the ISO dataset (for instance "`ISO-19115:2003`").

Attribute Name	Description	Type	Comment
`language`	Language used to describe the metadata	String	ISO639-2 [A11] should be used here: English = 'eng', Dutch = 'dut'
`organisationName_metadata`	Complete name of the organization responsible for the metadata (responsible party).	String	Use the complete name of the organization. Abbreviation of the name may be added. Example: Koninklijk Nederlands Meteorologish Insitituut (KNMI)
`linkage_metadata`	URL of the organization	String	Example: http://adaguc.knmi.nl
`role_metadata`	Function performed by the responsible party	Code list[1]	Example: pointOfContact. Code list (B5.5) is available in appendix of [A07]
`datestamp`	Date the metadata was created or altered	Date	Date coded according to ISO8601 [A10]: YYYY-MM-DD. Example: 2006-01-23
`metadataStandardName`	Name of the metadata standard name (including profile name) used	String	Default value is: ISO19115
`metadataStandardNameVersion`	Version (profile) of the metadata standard used	String	Default value: Nederlandse metadatastandaard voor geografie 1.1

[1] A Code list consists of a enumeration (or list) of possible entries from which a value has to be chosen. The chosen value will have type 'String'.

code	Alphanumeric value identifying the reference system of the dataset	Code list	EPSG codes should be used, e.g. 28992 (RD), 7408 (RD/NAP), 4937 (ETRS89), 4326 (WGS84)
codeSpace	Name or identifier for the organisation or person responsible for the code reference system	String	Default value: EPSG EPSG codes should be used for better interoperability.
title	Name by which the cited resource is known	String	Name of the dataset
date	Reference date for cited resource	Date	Date coded according to ISO8601 [A10]: YYYY-MM-DD. Example: 2006-01-23. The type of event the date is valid for is in dateType.
dateType	Event used for the reference date	Code list	Code list (B5.2) can be found in [A07].
abstract	Brief narrative summary of the content of the resource	String	Short description of the dataset content
status	Status of the resource	Code list	Code list (B5.23) can be found in [A07]. Examples: Complete, onGoing
organisationName_dataset	Complete name of the organization responsible for the dataset (responsible party).	String	Use the complete name of the organization. Abbreviation of the name may be added. Example: Koninklijk Nederlands Meteorologish Insitituut (KNMI) Note: _dataset is added because of usage within an HDF group
linkage_dataset	URL of the organization responsible for the dataset	String	Example: http://www.knmi.nl Note: _dataset is added because of usage within an HDF group
role_dataset	Function performed by the responsible party	Code list	Example: pointOfContact. Code list (B5.5) is available in appendix of [A07] Note: _dataset is added because of usage within an HDF group
keyword	Commonly used word(s)or formalized word(s) or phrases(s) used to describe the subject	String	The use of keyword lists is encouraged (e.g. from netCDF-CF [R07], WMO keywords list [R14])
useLimitation	Limitation affecting the fitness for use of the resource or metadata.	String	Examples: 'Not to be used for navigation', 'Do not use the dataset for scales larger than 1:50,000'
accessConstraints	Access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource or metadata.	String	Mandatory field. If no restrictions, use 'N/A' or 'none' as value
spatialRepresentationType	Method used to spatially represent geographic information	Code list	Example: vector. Code list is available in appendix of [A07].
denominator	The number below the line in a vulgar fraction (1:)	Integer > 0	Scale on which the dataset is validly represented. Example: 1000
language_dataset	Language(s) used within the dataset	String	ISO639-2 [A11] should be used here: English = 'eng', Dutch = 'dut' Note: _dataset is added because of usage within an HDF group

organisationName_distributor	Complete name of the organization responsible for the distribution of the dataset	String	Use the complete name of the organization. Abbreviation of the name may be added. Example: Koninklijk Nederlands Meteorologish Insitituut (KNMI) Note: _distributor is added because of usage within an HDF group
linkage_distributor	URL of the organization responsible for the distribution of the dataset	String	Example: http://www.knmi.nl Note: _distributor is added because of usage within an HDF group
role_distributor	Function performed by the responsible party. In this case always 'Distibutor'	Code list	Note: _distributor is added because of usage within a HDF group
statement	General explanation of the data producers' knowledge about he lineage of a dataset	String	General description or remarks regarding the history of the dataset

Table 4-4 Detailed ISO dataset metadata

4.3.4 Product metadata

The product variable holds the general metadata valid for the whole product. The time unit used within the ADAGUC HDF5 file format specification is seconds. The time zone is always UTC.

Attribute Name	Description	Type	Comment
title	Short description of the file contents	string	Short description of the contents of the product
institution	Institution or organization where the original data was produced	string	
source	Method of production of the original data	string	Mission, Instrument or Modelname (1st element of the filenmame, see Table 3-1)
history	List of the applications that have modified the original data		
references	References that describe the data or methods used to produce it	string	Reference to a document or webpage describing the product
comment	Miscellaneous information about the data or methods used to produce it	string	Extra notes (if needed, otherwise leave empty)
ref_doc	Name of the document describing the dataset (ISO)	string	E.g. the Ozone product description
ref_doc_version	Version of the document describing the dataset (ISO)	string	E.g. "1.0"
format_version	Version of the ADAGUC data format being used for the product	string	This is the ADAGUC data format version mentioned on the front page of this document.
input_products	List of input products	string	
originator	Creator of the product	string	
creation_date	Creation date/time in UTC	date/time	
software_version	The version of the software which created the file	string	E.g. "V1.0.1"
file_class	File Class (2nd element of the filenmame, see Table 3-1)	string	E.g. "OPER", "CONS", "TEST", etc.
type	Product Type (3rd element of the filenmame, see Table 3-1)	string	Can be either "R", "V", "I" or "O"
acronym	Product Acronym (5th element of the filenmame, see Table 3-1 and Table 3-3)	string	E.g. "TRPNO2"

level	Product Level (6th element of the filenmame, see Table 3-1)	string	E.g. "L3"
variables	List of all variables containing the actual data of the product	string	Variables in section C of Figure 4-1
validity_start	Start date/time of the validity period of the product (1st part of the 7th element of the filenmame, see Table 3-1)	date/time	E.g. "20060101T000000"
validity_stop	End date/time of the validity period of the product (2nd part of the 7th element of the filenmame, see Table 3-1)	date/time	E.g. "20061231T235959"
style	Style / form of the attributes	string	Can be either "lowercase", "UPPERCASE", "camelCase", "separating_character" or "mixed"

Table 4-5 Detailed product metadata

4.3.5 Projection metadata

The projection variable holds the projection metadata for this product. Projection metadata is required to define the type of coordinate system, type of projection and datum for each dataset. The projection metadata attributes follow the Climate and Forecast Conventions [R07], with the addition of three attributes from the ADAGUC data products standard. These additional attributes are `projection_name`, `EPSG_code` and `proj4_params`. The netCDF Climate and Forecast grid mapping attributes can be found in Appendix F of the netCDF Climate and Forecast (CF) Metadata Conventions [R07]. The attributes depend on the type of coordinate system and type of projection.
Examples of projection metadata are given for two datasets (TMNO2 and NL25PCP) in the last column of Table 4-6 and Table 4-7, respectively. The first dataset uses a geographic coordinate system (Table 4-6). For this type of coordinate system the attribute `grid_mapping_name` is set to "latitude_longitude". This represents a geographic coordinate system with units in degrees and datum WGS84. This type of coordinate system is used for mapping the TMNO2 dataset, see Figure 4-1 (middle).
The second dataset uses a projected coordinate system with a polar stereographic projection (Table 4-7). This type of projection requires several extra parameters: `latitude_of_projection_origin`, `straight_vertical_longitude_from_pole`, `scale_factor_at_projection_origin`, `false_easting` and `false_northing`. Optionally, the ellipsoidal figure associated with the geodetic datum to approximate the shape of the earth can be defined using the `semi_major_axis`, `semi_minor_axis`, `longitude_of_prime_meridian` and `inverse_flattening` parameters. This type of coordinate system is used for mapping the NL25PCP dataset, see Figure 4-1 (right).

Attribute Name	Description	Type	Comment
projection_name	Full (human readable) name of the projection being used	string	E.g. for TMNO2: "Latitude Longitude"
EPSG_code	The EPSG code of the projection being used (if available)	string	E.g. for TMNO2: "EPSG:4326" If an EPSG code is not defined for the proj4_params use "UNDEFINED", if the EPSG code is unknown use "UNKNOWN".

proj4_params	List of space separated tag=value pairs to be used with the PROJ.4 projection library [R15]. Each tag is preceded with a +	string	E.g. for TMNO2: "+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs"
grid_mapping_name	Name of the grid mapping	string	E.g. for TMNO2: "latitude_longitude"

Table 4-6 Projection parameters for a standard geographic coordinate system.

Attribute Name	Description	Type	Comment
projection_name	Full (human readable) name of the projection being used	string	E.g. for NL25PCP: "Polar Stereographic"
EPSG_code	The EPSG code of the projection being used (if available)	string	E.g. for NL25PCP: "UNKNOWN" If an EPSG code is not defined for the proj4_params use "UNDEFINED", if the EPSG code is unknown use "UNKNOWN".
proj4_params	List of space separated tag=value pairs to be used with the PROJ.4 projection library [R15]. Each tag is preceded with a +	string	E.g. for NL25PCP: "+proj=stere +lat_0=90 +lat_ts=60 +lon_0=0 +k=1 +x_0=0 +y_0=0 +a=6378.14 +b=6356.75 +units=m +no_defs"
grid_mapping_name	Name of the grid mapping	string	E.g. for NL25PCP: "polar_stereographic"
latitude_of_projection_origin	The latitude chosen as the origin of rectangular coordinates for a map projection.	double	E.g. for NL25PCP: 60
straight_vertical_longitude_from_pole	The longitude to be oriented straight up from the North or South Pole.	double	E.g. for NL25PCP: 0
scale_factor_at_projection_origin	A multiplier for reducing a distance obtained from a map by computation or scaling to the actual distance at the projection origin.	double	E.g. for NL25PCP: 1
false_easting	The value added to all abscissa values in the rectangular coordinates for a map projection.	double	E.g. for NL25PCP: 0
false_northing	The value added to all ordinate values in the rectangular coordinates for a map projection.	double	E.g. for NL25PCP: 0
inverse_flattening	Used to specify the inverse flattening (1/f) of the ellipsoidal figure associated with the geodetic datum and used to approximate the shape of the Earth (*optional*).	double	E.g. for NL25PCP: 298.183263207102
semi_major_axis	Specifies the length, in meters, of the semi-major axis of the ellipsoidal figure associated with the geodetic datum and used to approximate the shape of the Earth (*optional*).	double	E.g. for NL25PCP: 6378140
semi_minor_axis	Specifies the length, in meters, of the semi-minor axis of the ellipsoidal figure associated with the geodetic datum and used to approximate the shape of the Earth (*optional*).	double	E.g. for NL25PCP: 6356750

longitude_of_prime_meridian	Specifies the longitude, with respect to Greenwich, of the prime meridian associated with the geodetic datum (*optional*).	double	E.g. for NL25PCP: 0

Table 4-7 Projection parameters for a projected coordinate system with a polar stereographic projection. Descriptions are derived from Appendix F of the netCDF Climate and Forecast (CF) Metadata Conventions [R07].

4.4 Data format specification

4.4.1 General format

The ADAGUC data products shall follow the HDF5 file format specification [R01]. The HDF5 library provides a programming interface to a concrete implementation of the HDF storage format. The ADAGUC file format requires the usage of "Dimension Scales" [R10]. Therefore, it is advised to use the HDF5 dimension scale API (introduced in HDF5 v1.8) or the netCDF-4 API [R06]. The usage of the netCDF-4 API is preferred (see paragraph 4.1), where the creation mode flag should include "NC_NETCDF4" to cause the netCDF library to create a HDF5/netCDF-4 file.

The datasets included in the ADAGUC HDF5 products need to be uniform in time and space. They should be related and share the same metadata. See Table 3-4 for possible combinations of datasets within one data product (i.e. within one file).

Optionally variables and datasets can have PyTable attributes [R08]. Datasets need to have attributes that define geolocation and time. However, geolocation and time can be supplied in different datasets, preferably by using the HDF5 Dimension Scale API [R01, R07, R10].

Each variable or dimension scale in section B of Figure 4-1 will have, at least, the attributes described in Table 4-8. These attributes follow the NetCDF Climate and Forecast (CF) Metadata Conventions [R07].

Attribute Name	Description	Type	Comment
long_name	Full name of the variable	string	E.g. "longitude" for variable "lon", "latitude" for variable "lat"
units	Unit of the variable	string	E.g. "degrees_east"

Table 4-8 Required dimension scale attributes

Each dataset in section C of Figure 4-1 will have, at least, the attributes described in Table 4-9. These attributes follow the NetCDF Climate and Forecast (CF) Metadata Conventions [R07].

Attribute Name	Description	Type	Comment
long_name	Full name of the variable	string	E.g. "air_pressure_at_cloud_top" for variable "cloudTopPress"
units	Unit of the variable	string	E.g. "hPa"
standard_name	Standard name for the variable	string	E.g. "precipitation_flux"
coordinates	True latitude and longitude coordinates	string	E.g. "lon lat"

`grid_mapping`	Mapping between the given coordinate variables and the true latitude and longitude coordinates	string	In general, this is the name of a variable that provides the description of the mapping with a set of attributes. For ADAGUC products this is always "projection", i.e. pointing to the projection variable described in Table 4-6

Table 4-9 Required dataset attributes

4.4.2 Additional attributes

Paragraph 4.4.1 describes the mandatory attributes for all ADAGUC products. The following frequently used attributes can be added to variables of an ADAGUC HDF5 data product.

coordinates:
The latitude and longitude coordinates of a horizontal grid that was not defined as a Cartesian product of latitude and longitude axes, can be represented using two-dimensional coordinate variables. These variables are identified as coordinates by use of the 'coordinates' attribute: "`coordinates = lon lat`".

_FillValue:
This attribute defines a value used to represent missing or undefined data. It is recognized by the netCDF library as the value to pre-fill disk space allocated to the variable it belongs to. Its data type should be the same as the data type of the variable.

4.4.3 Raster based data

TBW (examples/dump of a raster based HDF5 file could be specified here)

4.4.4 Vector based data

TBW (example/dump of a vector based HDF5 file could be specified here)

4.4.5 Model based data (additional)

TBW (example/dump of a HDF5 file with model data could be specified here)

5 Conversion format standard

5.1 Imported data formats

The ADAGUC system shall import data in the formats described in the next sections.

5.1.1 ASCII

TBW

5.1.2 HDF4

TBW

5.1.3 HDF5

TBW

5.1.4 NetCDF

TBW

5.1.5 BIL

TBW

5.1.6 GeoTIFF

TBW

5.1.7 GRD

TBW

5.1.8 GRIB

TBW

5.2 Exported data formats

The ADAGUC system shall export data in the formats described in the next sections. The ADAGUC system primarily depends on GDAL/OGR [R05] for export of data. When GDAL/OGR support is extended to new data formats those formats may be added to the list of supported formats (if applicable). As the formats are already implemented in GDAL/OGR no additional specifications are needed in most cases.
GDAL/OGR consists of two libraries. The Geospatial Data Abstraction Library (GDAL) is used for raster data access, the OGR Simple features library is used for vector data access.

5.2.1 GRD

No additional specifications needed.

5.2.2 NetCDF

Raster netCDF-CF compliant files will be created by the GDAL ADAGUC netCDF raster driver. The driver will be in netCDF3 mode to make interoperability with most netCDF readers possible. The exported files comply with version 1.2 of the netCDF Climate and Forecast Conventions (CF) [R07].

Because the netCDF driver for OGR currently lacks a write function, *vector* netCDF files can not be exported to netCDF. However, it is possible to download vector netCDF4-CF files directly from the ADAGUC system (in the internally stored format, without any modifications).

5.2.3 HDF5

The exported HDF5 format will be the ADAGUC HDF5 format described in paragraph 4.2, 4.3 and 4.4. The file will be exported to HDF5 using the GDAL netCDF driver, in netCDF4 mode. Because netCDF4 internally uses the HDF5 file format, the driver will export files in the HDF5 file format. Optionally, to be able to build a memory map of the contents of the HDF5 file efficiently with PyTables, one can choose to have a set of attributes included that provide basic hinting about the kind of data and layout (for datasets only) in each group and PyTabes leaf (= HDF5 dataset) of the file [R08]. Listed below are the mandatory PyTables attributes:

Attribute Name (File/Group)	Type	Comment
CLASS	string	'GROUP'
TITLE	string	same as title
VERSION	string	'1.0'
PYTABLES_FORMAT_VERSION	string	'2.0' [only for root-group]

Table 5-1 PyTables attributes at the top-level of the product

Attribute Name (Table)	Type	Comment
CLASS	string	'TABLE'
TITLE	string	description of table
VERSION	string	'2.6'
FIELD_X_NAME	string	name of column 'X'
FIELD_X_FILL	string	default value of field 'X'
NROWS	integer	number of rows

Table 5-2 PyTables attributes for tables

Attribute Name (EArray)	Type	Comment
CLASS	string	'EARRAY'
TITLE	string	description of dataset
VERSION	string	'1.3'
EXTDIM	string	number of extendable dimension

Table 5-3 PyTables attributes for EArrays

5.2.4 GeoTIFF

No additional specifications needed.

5.2.5 SHP (Shapefile)

No additional specifications needed.

5.2.6 KML (Google Earth)

KML-support in GDAL is under development.

6 Units

The units and standard names used in the ADAGUC data products standard follow the netCDF Climate and Forecast Conventions (CF) [R07] where possible. The units attribute is required for all variables except for the `iso_dataset`, `product` and custom variables. The value of the units attributes is a string that can be recognized by Unidata's Udunits package [R20].

Standard names are used to precisely describe the physical quantities being represented. The standard name is applied to the attribute `standard_name`, which takes a string as value. The set of available standard names are listed in the standard name table [R21].

7 Coordinate systems

7.1 Introduction

Basically there are two types of coordinate systems, geographic and projected coordinate systems. A geographic coordinate system uses latitude and longitude to reference a point on the earth's surface. The earth's surface is represented as a three-dimensional spherical surface. The sphere is approximated by an ellipsoid. The ellipsoid can be described by two radii: the longer radius is called the semi major axis, and the shorter radius is called the semi minor axis. The datum of a geographic coordinate system defines the position of the spheroid relative to the center of the earth, e.g. the origin and orientation of latitude and longitude lines [R22].
A projected coordinate system is a flat, two-dimensional representation of the earth. The projected coordinate system uses linear units of measure. The projected coordinate system is based on a spherical earth model, and its coordinates are related to geographic coordinates by a projection transformation. The projection transformation is called the map projection. It is used to convert the latitude and longitude coordinates to x, y coordinates on the flat projection. Examples of map projections are conic, cylindrical and planar surfaces [R22].

7.2 Coordinate systems within ADAGUC

The coordinate systems used in the ADAGUC *raster* data products are the coordinate systems defined in Appendix F of the NetCDF Climate and Forecast conventions [R07]. For *vector* data only the geographic coordinate system with name "latitude_longitude" is supported. This means that coordinates of vector data are always represented by latitude and longitude on a spherical earth.
The approximation of the ellipsoid of the earth's surface for all coordinate systems can be defined by the `inverse_flattening`, `semi_major_axis` and `semi_minor_axis` attributes. The inverse flattening is related to semi major axis and semi minor axis with the formula:

```
inverse_flattening = semi_major / (semi_major - semi_minor)
```

The Proj4 library [R15] in combination with GDAL is used to translate between coordinate systems. Chapter 4.3.5 lists examples for a geographic and a projected coordinate system.

8 Limitations and open issues

8.1 Introduction

This chapter discusses the current limitations and open issues, including: the granularity level of the published datasets, the need for a catalog service, etc.. It is important to explain to the community that the ADAGUC project is aware of these topics and is fully aware of the present shortcomings and (future) challenges. Naturally, these topics might be addressed in future projects.

8.2 Web Coverage Service

The current Web Coverage Service (WCS) is based on UMN MapServer [R04] and GDAL [R05]. GDAL is the data translator tool which is used by UMN MapServer to convert between various geographical formats. When a WCS GetCoverage request is done to UMN MapServer, one raster product is returned. The data is converted to the needs of the user. For example the geographical region, projection, resolution and file format can be adjusted.

8.2.1 Datasets

The current limitation of the GDAL/UMN MapServer combination [R05, R04] is that only one raster product for each request can be returned. For example, a user wants to retrieve Tropospheric NO2 and he or she is interested in the measurement error for each pixel as well. Within the ADAGUC system it is currently not possible to return a raster dataset with its accompanying error field in one request. Two requests need to be done, one request to retrieve the data and one request to retrieve the error fields.

8.2.2 Dimensions

The current implementation of the ADAGUC raster driver in GDAL limits the maximum number of dimensions not to exceed three. Only a 2D or 3D grid is supported. GDAL supports multiple bands within a raster. This is used as the third dimension in the ADAGUC raster driver.

8.3 Web Feature Service

Time requests to the OGC [A02] services are used to request data for a certain time domain. The parameter to make these requests is named TIME. In contrast to the Web Map Service (WMS) and the Web Coverage Service (WCS), the Web Feature Service (WFS) does not support time requests. It is however possible to make selections in time by using WFS Filter Encoding (FE) [R23] which makes it possible to query the data. Filter Encoding is supported by UMN MapServer, but date comparison is not supported by UMN MapServer. Therefore the ADAGUC system will currently not support WFS Filter Encoding or WFS Time requests.

8.4 OGR Library

Currently there is no support in OGR to create ADAGUC netCDF files from other vector formats. One of the problems is that the ADAGUC vector netCDF file format supports four vertices for each polygon. With OGR it is possible to use point, polygons and multi polygons with an unlimited amount of vertices. Because of this constraint, write support for the OGR ADAGUC driver has not been included.

8.5 Catalog services

At the moment, ADAGUC does not provide catalog services, due to present lack of a well recognized and adapted standard and the consequent limitations in discovery clauses. Therefore, implementation was not feasible within the timeframe of the project. A follow up should investigate the different existing standards and try to select one in line with INSPIRE and GMES.

8.6 Granularity

TBW – This section will explain a) the adopted approach to select the granularity level of published data, b) the consequential limitations for accessing the other possible levels of data granularities and c) the impact on the metadata ingestion/harvesting process.