NILU: OR 54/2002

NILU: OR 54/2002
REFERENCE: O-100064
DATE: DECEMBER 2002
ISBN: 82-425-1399-6

Operation of an Envisat Validation Data Centre

Estec Contract No 14419/00/NL/SF
Final Report

**Terje Krognes, Aasmund Fahre Vik,
Kjersti Karlsen Tørnkvist, Sam-Erik Walker,
Roland Paltiel, Trygve Bårde, Bjørn Gloslie,
Rita Larsen, Geir Ole Braathen, The Nguyen Thanh**

Preface

Ground based measurements are needed for evaluation of ENVISAT-1 measurements. NILU has created a storage facility for ground based measurement data with quality control of file formatting, and a web based file retrieval search engine.

This system has been active during the Commissioning phase under the current contract and will also be maintained during the Exploitation phase.

Contents

Summary

NILU has designed and implemented a system for organizing ground based measurement data, and for retrieval of the same data by scientists that perform comparisons with measurements from the ENVISAT satellite.

Data are stored in the HDF 4.x file format, and are accessed through a web-interface based on PHP server-side scripting, a MySQL database, an Apache server and Red Hat Linux running on a standard Intel processor PC.

This report explains the different system components and its related software packages, and describes how the data centre is operated.

Operation of an Envisat Validation Data Centre

1 Introduction

NILU has designed and implemented a system for organizing ground based measurement data, and for retrieval of the same data by scientists that perform comparisons with measurements from the ENVISAT satellite. The work has been performed in close co-operation with ESA and with representatives of the user community. The system is complex since it entails co-operation between wide spread scientific communities that have separate and different cultures and methods. In the ESA ENVISAT Calibration/Validation effort the measurements of stratosphere physicists, modellers and mathematicians, marine biologists, and space scientists need to be described within one common frame of reference.

2 System components

The main deliverables are summarized in the following paragraphs:

- Documentation (Data definitions and user instructions comprised in the Metadata Guidelines, + additional documents)
- Data file templates (mainly developed by the user community representatives)
- Software (tools for local generation and testing of data files, central file checking tools and display tools, web site, index database)
- Hardware (Intel PC with web server and database, and complete back-up machine, 200GB storage)

ESA selected the HDF 4.x file format for the file exchange, based on the established use of this format within ESA and some of the user groups. Main software tools have been developed in FORTRAN, IDL, PHP, SQL and UNIX shell-scripts. The system uses Red Hat Linux, Apache web-server with PHP server-side scripting, and a MySQL database.

Through extensive co-operation with the ESA project Official **Dr. Rob Koopman**, the system design has been extended and adapted according to the growing awareness of the user requirements. The final product has evolved considerably compared to what was envisioned when the Work Statement was created, and may be subject to further evolution.

3 The NILU Team

The NILU project leader from the start in the spring of 2000 until the spring of 2001 was **Dr. Bojan Bojkov**. He left NILU April 2001.

The Nguyen Thanh took over as the administrative project manager after Bojan Bojkov.

Dr. Aasmund Fahre Vik has joined the team to assist with the scientific part of the project management and takes part in IDL and FORTRAN programming.

Terje Krognes took part in working out data definitions, database structures and metadata guidelines, and has done a major part of the web development in PHP and SQL. He laid out the fundamentals of the system architecture, and interacted extensively in the detailed system design. He is also responsible for the user support, and handles updates of metadata in co-operation with the Metadata Board.

Sam-Erik Walker has programmed the main software tools in FORTRAN (ASC2HDF), and supported users with analysis of HDF file formatting problems.

Trygve Baarde has programmed the HDF file processing system in UNIX shell scripts, and assisted with hardware and OS installation.

Roland Paltiel has performed the the hardware and software installation.

Bjørn Gloslie has assisted with database configuration and administration.

Rita Larsen created the initial design of the Cal/Val web site, and has assisted with communication with the user group in administrative matters, such as data protocol administration.

Dr. Kjersti Karlsen Tørnkvist has joined the team as an IDL programmer.

Dr. Geir Ole Braathen has contributed as a scientific advisor.

4 Detailed system description

The system components are here described in a logical order when we follow a data file as it passes from the originator into the storage and forward to an end user. The DS (Data Supplier) needs to sign a data protocol and be registered in the system metadata. This allows the user access to the Cal/Val web site, and gives permission to upload data for one or more projects (AO's).

At the Cal/Val website (**http://nadir.nilu.no/calval/secure/**) the user will find Metadata Guidelines, file templates and other documents that help with formatting original data into an HDF file. A software tool named **ASC2HDF** is available for Windows, Linux, Solaris and HPUX users. This tool accepts data and metadata in two simple text files, and will generate an HDF file after extensively testing the input according to predefined rules and values in **table.dat**. **Table.dat** contains up-to-date information on all legal values in each metadata field. Whenever metadata are updated at the central site, a new version of **table.dat** is posted on the web.

When the HDF file has been successfully tested at the local site by **ASC2HDF**, it may be uploaded to the Cal/Val site (/**nadir/esa/incoming**) by ftp, or through a web upload page (**http://nadir.nilu.no/calval/secure/upload.php**). A set of

UNIX shell scripts is started every 5 minutes. These scripts check for new files in the /nadir/esa/incoming directory, and process each file by launching a FORTRAN program named **HDF2ASC**. The program verifies data and metadata according to the predefined rules in the **table.dat** – the same file that was used by **ASC2HDF**.

Even files that have been successfully tested by the originator, may be rejected at NILU, mostly due to inconsistencies in the file name (which reflects a subset of the metadata content), or due to duplicate file names or out-of-sequence version numbers. If the data supplier is not accredited for the AOID listed in the file, the file will also be rejected. An error report will automatically be emailed to the data supplier and the owner of the logon name that was used, and the file will be moved to a hidden directory.

If all checks out correctly, the received HDF file will be moved to a storage file tree starting
at **/nadir/esa/data**, and the file name, upload details and central metadata elements are stored in an **index database.** The system enforces consistent naming of variables and other metadata elements, and consistent spelling of names for people, organisations and sites.

The nadir index database contains the official list of allowed metadata values in the Cal/Val HDF data files, in addition to logs of uploaded/downloaded files, an overview of metadata contents, and the variable list of all accepted HDF data files. All this information is available to dynamic web pages at the Cal/Val web site. The main end user tool on this site is the "Search Data" page (**http://nadir.nilu.no/calval/secure/getdata4.php**), which allows the user to sort through the data files with advanced criteria selections. Filtering by data supplier, project, location, data source, data type, component and other metadata elements is supported. Data files may also be filtered by a "4-D box algorithm" (any file with data relevant for a given geographical location and time). Furthermore, files can be filtered by submission date and update status.

All data files that match the search criteria are listed in a new web page, with links to **HDF data file download**, to **comments**, and to a **variable list**. In the variable list page the user may select variables and generate an **on-line plot**. In the file list the user may also select multiple files for download as a tar-ball. The user may save the search criteria in the index database for convenient re-use at a later time.

Users that have an IDL license may download IDL scripts for HDF data file formatting (excluding the detailed error checking available in the FORTRAN version) and for plotting of data sets from HDF files (**http://nadir.nilu.no/calval/secure/idltools.php**).

Users that have signed an additional protocol for access to ECMWF data will find pre-computed **T106 ASCII extractions and plots** for the last 30 days, as well as **plots of isentropic and isobaric forecasts**. There is an on-line tool for extracting **T106 data into HDF files**. There is also an on-line facility for plotting **10-day back trajectories** (based on data calculated at DMI by Dr. B. Knudsen).

5 Operations

5.1 System development

The development process itself has been among the main operations through the entire project. The specifications have been continuously modified in co-operation with ESA. Any further development that causes changes to metadata definitions, database structure, or changes to system functionality, will be defined as separate projects.

5.2 User support

User support is the main purpose of the system, and has been the second main operation from the project start. Interaction with the users has been important for the evolvement of the system functionality. In addition to the documentation and information that users may download from the Cal/Val web site, the NILU team provide direct assistance to users that have problems with data file formatting (via email or telephone). This operation will continue into the Exploitation Phase as part of the "Operation of the Envisat Cal/Val Data Centre" project.

5.3 Metadata management

Routine metadata management has been provided since the first Rehearsal campaign, and will continue into the Exploitation Phase. This activity includes defining people as users, adding or changing metadata elements (names of persons, institutions or projects, component names, station names, etc.). Some of these activities require interaction with the Metadata Board (to maintain the scientific correctness of the metadata definitions). To ensure that old and new data files remain compatible, only additions to the metadata elements (not modifications or deletions) are encouraged when the project moves from the Commissioning Phase into the Exploitation Phase.

5.4 System management

The production server has (as of November 2002) been operational for 200 days without rebooting. System management is nevertheless a continuous task. A clone of the production server is kept ready for immediate use in case of system failure. Disk backups are made daily. Both the production server and the clone server need frequent operating system updates to maintain a reasonable security with regard to malicious crackers. The automated tasks are constantly monitored for stalled scripts or other malfunction. Some manual monitoring is also needed to manage the total system. These activities will continue into the Exploitation Phase.

5.5 Automated operations

The data file reception and testing is fully automated, including automatic email messages to the users for files that are not accepted. User requests for data file downloads, plots and ECMWF data extractions are automatically handled by web software. Pre-computed extractions and plots are generated and removed by automatic shell scripts. All these services will remain operational as long as the system itself is managed and kept on-line.

6 Compliance with the Work Statement of the contract with ESA

The statement of work defines 5 main tasks in section 3:

(1) Purchase and installation of database hardware and software,
(2) Interaction with users,
(3) Development of tools,
(4) Validation rehearsal and
(5) Data-centre services during and after the commissioning phase.

All tasks have been performed in accordance with the Work Statement except for task 3 where development of CORPORAL was removed in agreement with ESA. All deliverables (section 4) have been completed with the following exceptions: Data-comparison software (CORPORAL) will not be developed. Deliverable #5, CD-ROMS containing complete sets of data gathered by ACVT and MAVT members, will be delivered at the end of the project. The schedule (section 5) was delayed due to postponed launch of ENVISAT. NILU have attended meetings, workshops and reviews (section 6) in accordance with the Work Statement and will perform a final presentation of the project at ESRIN December 2002.

Norwegian Institute for Air Research (NILU)
P.O. Box 100, N-2027 Kjeller, Norway

REPORT SERIES	REPORT NO. OR 54/2002	ISBN 82-425-1399-6	
Scientific Report		ISSN 0807-7207	
DATE	SIGN.	NO. OF PAGES	PRICE
		8	NOK 150,-

TITLE	PROJECT LEADER
Operation of an Envisat Validation Data Centre	The Nguyen Thanh
Estec Contract No 14419/00/NL/SF Final Report	NILU PROJECT NO. O-100064
AUTHOR(S) Terje Krognes, Aasmund Fahre Vik, Kjersti Karlsen Tørnkvist, Sam-Erik Walker, Roland Paltiel, Trygve Bårde, Bjørn Gloslie, Rita Larsen, Geir Ole Braathen, The Nguyen Thanh	CLASSIFICATION * A
	CONTRACT REF. Estec Contract No. 14419/00/NL/SF

REPORT PREPARED FOR
EUROPEAN SPACE RESEARCH AND TECHNOLOGY CENTRE
Postbus 299, 2200 AG Noordwijk
Keplerlaan 1, 2201 AZ Noordwijk
The Netherlands

ABSTRACT

NILU has designed and implemented a system for organizing and presentation of ground based measurement data, and for retrieval of the same data by scientists that perform comparisons with measurements from the ENVISAT satellite.
Data are stored in the HDF 4.x file format, and are accessed through a web-interface based on PHP server-side scripting, a MySQL database, an Apache server and Red Hat Linux running on a standard Intel processor PC.
This report explains the different system components and its related software packages, and describes how the data centre is operated.

NORWEGIAN TITLE

Operation of an Envisat Validation Data Centre

KEYWORDS		
ESA	Envisat	Calibration and Validation

ABSTRACT (in Norwegian)

NILU har utviklet et system for lagring og presentasjon av bakkemålinger for sammenligning av måledata fra ENVISAT-satelitten.
Dataene er lagret i HDF 4 x format. Systemet aksesseres gjennom et web-grensesnitt med PHP, MySQL-database,. Apache server og Red Hat Linux på en Intel PC.
Denne rapporten viser hvordan systemet er bygd opp.

** Classification*
A Unclassified (can be ordered from NILU)
B Restricted distribution
C Classified (not to be distributed)