GECA: ESA'S NEXT GENERATION VALIDATION DATA CENTRE

Y.J. Meijer[(1)], T. Fehr[(1)], R.M. Koopman[(2)], A. Pellegrini[(1)], G. Busswell[(3)], I. Williams[(3)], M. De Mazière[(4)], S. Niemeijer[(5)], R. van Deelen[(5)]

[(1)]ESA/ESRIN, Via Galileo Galilei, 00044 Frascati (RM), Italy, Email: yasjka.meijer@esa.int
[(2)]GEO Secretariat, Case Postale 2300, CH-1211 Geneva 2, Switzerland
[(3)]Logica, Keats House, Springfield Drive, Leatherhead, KT22 7LP, United Kingdom
[(4)]BIRA-IASB, Ringlaan 3, B-1180, Brussels, Belgium
[(5)]S[&]T, Olof Palmestraat 18, 2616 LR Delft, The Netherlands

ABSTRACT

In the coming decade the availability of satellite data from Earth Observation (EO) platforms will exhibit a significant growth. The dataflow of the Sentinel 1-5 series will be much larger than the one of their preceding satellite missions. In addition, ESA develops a continuous series of Earth Explorer satellite missions. As geophysical validation of these EO data remains a high priority, ESA has initiated a project to develop a Generic Environment for Calibration/validation Analysis (GECA), which is considered to become the next generation validation data centre. The evolution part of GECA is in the interoperability between various validation data centres, offering several functionalities facilitating validation analysis with full traceability. One of these functions is the collocation engine which matches satellite data to correlative data and provides the option to download selected sub sets. It will also be possible to compare satellite and correlative data using 'best practice' analysis functions.

1. INTRODUCTION

This paper describes the "Generic Environment for Calibration/Validation Analysis" (GECA), which is a service and functionality currently under development in an ESA project. It describes the scope, context and objectives of the project. The project consists of two distinct, but connected activities:

1. evolution of ESA's validation data centre,
2. definition of quality information & action protocol.

1.1. Data Validation

Geophysical validation of satellite data is required to obtain confidence in the data quality. It involves comparison of satellite data products with independent measurements. Accurate characterisation of diurnal, geographical, seasonal and other dependencies are required. Fig. 1 gives an example for GOMOS.

Historically, ESA has placed major emphasis on the geophysical validation of Earth Observation (EO) products, in particular for atmospheric chemistry. To support the complex validation process, ESA has set up

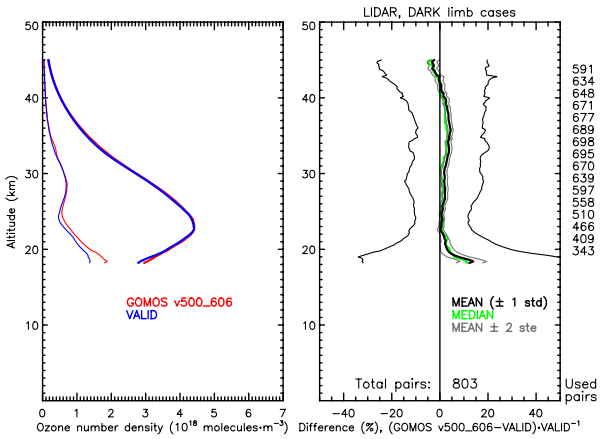

Figure 1. Example of GOMOS ozone profile validation results in an intercomparison with ground-based lidar data (picture credits: RIVM). Similar results can be found in [1].

validation data centres that host the correlative data acquired for calibration and validation of satellite sensors. Its 'first-generation' data centre implemented for the GOME-1 instrument had been derived from the infrastructure hosted by NILU for the Arctic science campaigns organised by the European Commission. Its 'second-generation' data centre has been implemented for the Envisat sensors AATSR, GOMOS, MERIS, MIPAS and SCIAMACHY, and contains specific functionality to further facilitate inter-comparisons and independent analyses during the validation process. Its metadata standard and format guidelines have been an evolution from the initial deliverable of the EU project COSE. The Envisat validation metadata and file structure guidelines [2] have become a de facto standard used by several compatible data centres from, among others, NASA (Aura Validation Data Centre), and the European Commission. It has also been endorsed by the Atmospheric Chemistry subgroup of the CEOS Working Group on Calibration and Validation. The current project consists of the definition and implementation of a further evolution, the 'third-generation' validation data centre, providing significantly more specialised functionality in support of the validation process.

Proc. 'Atmospheric Science Conference', Barcelona, Spain, 7–11 September 2009 (ESA SP-676, November 2009)

1.2. Using Satellite Data Quality Information

Measurements are only significant if their quality is specified. Traditionally, accuracy of ESA EO data products is assessed during validation campaigns, and recorded in product quality statements. Transient deviations from quality level are described in subsequent quality degradation reports and characterisation of complex quality dependencies is reported in cyclic, e.g., monthly or bi-monthly, quality assessment reports. Currently the sources of *a-posteriori* (i.e., formulated after product generation) information are not uniform, and are only available in (textual) documents. Users must interpret this external information (e.g., anomalies, bad datasets, and bad parameters) and adapt their analysis approach and tools to manually perform dataset correction. After corrective processing, the users will have to manually undo some of the earlier corrective actions. These operations are error prone and often not traceable. As a result, subjective interpretations enter the objective analysis, adversely affecting the outcome of scientific results.

It is the purpose of the GECA project to introduce a Quality Information and Action Protocol (QIAP) concept. This will allow the integration of available product quality information with data ingestion software resulting into uniform and automated action. The project includes an implementation for operational ESA missions. The definitions and architecture will be fully generic allowing further proliferation of the standard within the context of EO data harmonisation, e.g., in the frame of GMES.

2. GECA VALIDATION DATA CENTRE (GVDC)

The GECA Validation Data Centre (GVDC) will be a generic correlative data centre with significant evolution steps compared to traditional validation data centres. The evolution part is in the availability of specific functionalities facilitating the process of validation analysis. One of these functions is the collocation engine which matches satellite data to correlative data and provides the option to download selected sub sets. It will also be possible to compare satellite and correlative data using (best practice) analysis functions either via internet on the dedicated GECA server or locally at the user. More evolution is in the interoperability between various (validation) data centres through standardisation of (meta-) data, catalogue and data exchange offering a wider range of data.

2.1. Generic Correlative Data Centre

Instead of hosting correlative data related to a particular satellite mission, the GVDC will host data in a generic (meta-) data format (GDF) suitable to support any EO mission and covering multiple domains. In the first phase of the project, the database will be covering data related to ERS and ENVISAT including domains like atmospheric composition, land (cover, surface temperature (LST), vegetation, etc.), ocean (waves, Sea Surface Temperature (SST), etc.), land & sea ice and multi-spectral imaging (see Tab. 1).

Table 1. Domains covered by GECA with examples of current and future missions / instruments.

EO Domain	Satellite mission (instrument)
Atmospheric Chemistry	Envisat (GOMOS, MIPAS, SCIA), ERS-2 (GOME), Metop series (GOME-2), Sentinel-4, -5 & -5p
Atmospheric Dynamics	EarthCARE (ATLID, BBR, CPR, MSI), ADM-Aeolus (ALADIN)
Cryosphere (land- and sea-ice cover)	Cryosat-II (SIRAL), Envisat (RA-2), ERS-2 (RA-1)
Land (cover, LST, vegetation, etc.)	Envisat (AATSR, MERIS, RA-2, ASAR), ERS-2 (ATSR, RA-1, SAR), Sentinel-1 (SAR), Sentinel-3 (SLSTR, OLCI), SMOS
Multi-spectral imagery	Sentinel-2
Ocean and Coast (waves, colour, currents, SST, salinity, etc.)	Envisat (AATSR, MERIS, RA-2, ASAR), ERS-2 (ATSR, RA-1, SAR), Sentinel-1 (SAR), Sentinel-3 (SLSTR, OLCI), GOCE, SMOS (MIRAS)
Solid Earth (gravity, geoid, geomagnetic field)	Envisat (RA-2), ERS-2 (RA-1), GOCE, SWARM

GVDC compliant files can be generated with dedicated tools that will be available. The standard that is being developed, the GECA Data Format or short GDF, will be an evolution the Envisat metadata standard 4. Tools will be shared between compliant data centres. As some data providers are used to submit data to more than one data centre, a service for centralised submission is foreseen that, upon request, will automatically distribute files to various databases.

2.2. Data Centre Inter-Operability

Access to a wider range of correlative datasets is foreseen through interoperability. Currently data centre interoperability has started with the Aura Validation Data Centre (AVDC), ENVISAT Validation Data Centre (EVDC), European Aerosol Research Lidar Network (EARLINET), Global Atmospheric Watch Station In-formation System (GAWSIS), Global Earth Observation and Monitoring (GEOmon) and Network for the Detection of Atmospheric Composition Change (NDACC). In addition, there are plans to register all catalogue meta-data with the World Meteorological

Organisation Information System (WIS), which is in turn connected to the GEOSS clearinghouse. See Tab. 2 for the full list of data centres.

Exchange of catalogue metadata between the centres is currently operational as a prototype using the Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH v2, see [3]) (see Fig. 2). GVDC will host a metadata catalogue of the data available in the various peer data centres to which it is connected. In this way users can query datasets outside the GVDC. In the next phase of the interoperability, some data centres will agree to exchange actual data files (respecting the intellectual property rights). For the other data centres the user will have to access the relevant database directly using personal credentials to obtain the dataset of interest.

2.3. Campaign Management

The GVDC will support campaign managers and Cal/Val coordinators. It will empower the validation teams to review plans for correlative data acquisition and assess their adequacy for their requirements. During or after campaigns they can assess the campaign performance by comparing planned versus actual data acquisition.

2.4. Satellite Data Access

New is the direct access to ESA's multi-mission satellite data archive, which will allow validation scientists an easy way to obtain their required datasets. Users can query the product or satellite mission of their interest. The GVDC will perform a query in the relevant metadata using the Heterogeneous Missions Accessibility (HMA) standard [4]. The satellite data subset can then be selected for download and further analysis either via the GVDC server or (locally at) the user. Each validation step will allow the user to obtain the pre-processed satellite data in the GECA Data Format (GDF). The domains and satellite datasets covered by GECA are listed in Tab. 1.

2.5. Collocation Engine

A major evolution, compared to previous data centres, is the possibility to search for collocated pairs between satellite and correlative datasets. The GVDC user will be able to set certain basic, and if desired more advanced, criteria matching a satellite measurement under assessment with a correlative dataset. This other dataset can either be:

1. a measurement from another satellite,
2. a measurement from a non-satellite platform (i.e., ground-based, balloon-borne, air-borne, ship-borne or drifting), or
3. related to a fixed site.

Table 2. Data centres interoperating with GECA.

Data Centre	Main focus
AVDC (NASA)	Satellite validation
Earlinet (European)	Research and monitoring
EVDC (NILU/ESA)	Satellite validation
GAWSIS (Inter.nat.)	Station information
GeoMON (European)	Monitoring; data exchange/exploitation
GEOSS (Inter.nat.)	Data exchange/exploitation
NDACC (Inter.nat.)	Long-term monitoring/ Support to validation
WIS (WMO & Inter.nat.)	Data exchange/exploitation

The query output result will be presented together with collocation quality information related to the defined criteria. For example, time difference and measurement area overlap (or distance) between the collocated observations. Computer intensive, and hence long queries, can be scheduled or repeated at set intervals.

2.6. Analysis Environment

GVDC will allow users to directly perform intercomparisons based on 'best practice' analysis functions. This involves work currently performed within the Quality Assurance Framework for EO which, among others, establishes best practices for each community. A set of commonly used data handling and manipulation operations were selected.

Before actual intercomparison, a set of pre-processing functions will bring collocated data into a comparable state. This includes functionalities like conversion of paired data to a common format, inclusion of auxiliary data (e.g., meteorological parameters from ECMWF), parameter conversion (e.g., mixing ratio to concentration), regridding and smoothing of data. Note that data resulting from each stage on the server is also available for download and will be in the GECA Data Format (GDF). These functions will be implemented within GECA as generic building blocks. This allows the GVDC to be expanded in the future covering data from any new satellite mission and including new analysis functions.

The user has the choice to analyse the data either on the dedicated GECA server, which prevents the need of downloading the required datasets and exploits the available computing power, or the user can apply the GECA toolbox locally on downloaded data. The toolbox consists of the same building blocks as available on the server.

With the dedicated GECA server, users can avoid downloading large datasets to their computer and exploit available computing power. On the server the user can generate default or customised intercomparison reports (e.g., plots, statistics). Optionally the user can choose to retrieve ASCII files with the numerical data of each report component.

3. QUALITY INFORMATION & ACTION PROTOCOL (QIAP)

Currently many satellite data users do not take account of ESA's publicised quality reports. GECA will aim to resolve this by implementing a standard multi-mission protocol whereby *a-posteriori* quality issues are:

3. reported by satellite data users, and operational quality control teams;
4. investigated and managed by quality moderators;
5. publicised to satellite data users, with the possibility of applying corrections automatically (which remains the final user's choice), either through an end-user tool provided by GECA, or through external end-user tools compliant with the published protocol standard;
6. traceable at all times using a central database of quality issues.

QIAP will also allow computer systems to submit or query quality information over a standardised and automated interface.

Initially the QIAP implementation will target data from operational ESA missions, and not correlative validation data. Nevertheless, the concept will be sufficiently generic that it can also be applied to future ESA missions and possibly also non-satellite data.

4. CONCLUSION & OUTLOOK

GECA will offer a generic and comprehensive validation environment which includes much new functionality supporting the validation process. In addition, the QIAP concept will guarantee that quality information shall be properly taken into account in the use of the products.

The project has passed its system requirements review and has now started with the first phase of the implementation. An initial version of GECA will be operational early 2010, which will allow validation analysis of selected data sets from ERS and ENVISAT. By the end of 2010 the system will be fully operational for all missions (see Fig. 2) and accessible for external users.

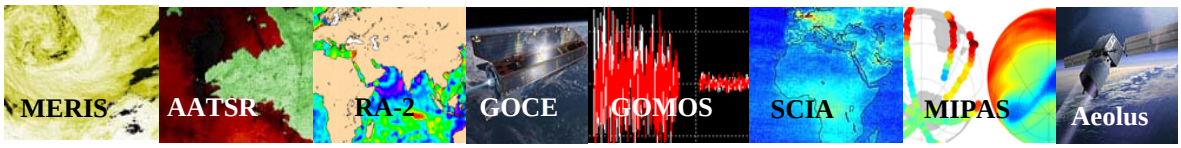

Figure 2. Impression of the generic character of GECA showing some of the covered satellite instruments.

5. REFERENCES

1. Van Gijsel, J. A. E., D. P. J. Swart, J. -L. Baray, H. Claude, T. Fehr, P. Von Der Gathen, S. Godin-Beekmann, G.H. Hansen, T. Leblanc, I.S. McDermid, Y.J. Meijer, H. Nakane, E.J. Quel, W. Steinbrecht, K.B. Strawbridge, B. Tatarov, E.A. Wolfram, 2009: Global validation of ENVISAT ozone profiles using lidar measurements, *International Journal of Remote Sensing*, 30, 15, pp. 3987—3994.
2. Bojkov, B.R., M. De Mazière, R.M. Koopman, 2002: Generic Metadata guidelines on atmospheric and oceanographic datasets for the Envisat Calibration and Validation Project, *ESA document*, pp 1-65.
3. OAI, 2002: The Open Archives Initiative Protocol for Metadata Harvesting, *http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm.*
4. OGC, 2006: OGC Catalogue Services Specification 2.0.0 – *EO Application Profile for CSW*, 2.0 version 0.1.4.

6. ACKNOWLEDGEMENTS

The authors are grateful for the contributions of the scientific partners in the consortium of the GECA project. These partners are from the following institutes: BIRA-IASB, CLS, CLS (BOOST), KNMI, RAL and University of Leicester.