

TIDE, the TropOMI Integrated Development Environment

C.M. Plevier, H. de Wolf, P.G. Beerthuizen

*Dutch Space B.V.
P.O.Box 32070
2303 DB Leiden, the Netherlands
Email: info@dutchspace.nl*

ABSTRACT

Dutch Space's experience and technology development for processing of earth observation (EO) remote sensing data provided the basis for a new approach of using a distributed architecture for end-to-end system simulation in the definition of the Tropospheric Ozone Monitoring Instrument (TropOMI) and related processing chain. All major stakeholders in this process can combine their own simulators (even when protected by IPR, and/or running behind company firewalls) in a network which is based on a computing and data grid paradigm. The GridAssist toolset thereby takes care of all workflow and access handling management, providing a distributed environment to define and optimise the instrument and processing chain. This system (TropOMI Integrated Development Environment, TIDE) prototype is currently in daily operational use for several months.

Complementary to TIDE and other grid-based developments a prototype of a Virtual Data Centre (VDC) for semantic web based access of EO and other large data sets of various natures and build was recently demonstrated. The VDC keeps the peculiarities of the many database aspects hidden for the user, providing unified access methods that can also be made part of an automated processing chain, through the Dutch Space developed Yggdrasill portal tool set.

The paper describes in more detail the design aspects and considerations used for the creation of the prototype system which is currently in daily use by the TropOMI definition team.

DISTRIBUTED END-TO-END SIMULATION FOR INSTRUMENT DEFINITION

TIDE (TropOMI Integrated Development Environment) is a design simulator running distributed over the internet to assist in the development of the TropOMI instrument. TIDE aims for early discovery of instrument level 2 requirements by involving end-users in phase A instrument design simulations. This improves the overall quality of specifications and design of both the instrument and related processing chain with respect to the intended use of the instrument. The TropOMI Integrated Development Environment was build by Dutch Space during a study project with KNMI, SRON, ARGOSS and TNO.

Heritage in Remote Sensing Data Processing

The Netherlands has a strong heritage in the design, building, calibrating of Earth atmosphere observing instruments, such as the GOME instrument on ESA's ERS-2 satellite (launched in 1995), SCIAMACHY[1] on ESA's ENVISAT (launched in 2002), OMI[2] on NASA's EOS-Aura satellite (launched in 2004), and GOME-2 on EUMETSAT's MetOp (launched in 2006).

This is complemented by expertise that was gained in the development of related software:

- For the OMI instrument on ENVISAT, the Ground Data Processing Software processed raw instrument data (level 0) to calibrated measurement data (level 1b), using attitude and ephemeris data for the spacecraft and ancillary data.
- In order to validate ENVISAT Product Data Sets early in the development cycle, the Data Product Analysis Tool was developed for ESA. It validated the format of the data, checked the internal consistency of the file structure, compared the instrument data against the meta-data in the file header and provided an overview of the contents of the file, and optionally extracted data for detailed inspection.
- Another tool working on the ENVISAT PDS converted these files into the Hierarchical Data Format (HDF), a multi-object file format for (graphical and numerical) scientific data. This has been incorporated in the Netherlands SCIAMACHY Data Centre as developed under co-ordination of KNMI.

- In order to support end-to-end processing chains early in the development cycle, the End-to-end system Analysis and Service Environment (EASE) provided a framework for the integration and control of various software elements (simulations, or incorporating final applications) which make up an entire data acquisition and processing chain. Based on CORBA technology, this framework enabled distributed concurrent engineering over the Internet. It was used in the Ozone Application Simulation and Explorer (OASE) project to support the geographically distributed processing of atmospheric ozone data, using the ERS-2/GOME simulator preceded by the MODTRAN radiative transfer model.
- This EASE concept was combined with the upcoming Grid technology into Grid-aware End-to-end system Analysis and Service Environment (GREASE), which provided benefits of Grid computing in a Computational Grid environment using applications that themselves were not Grid-aware. The work flow management tool of GREASE allowed users to construct scenarios consisting of chains of applications, which were dispatched to Grid processing resources for execution. It has been used for applications related to the Ozone Monitoring Instrument (OMI), operational since 2004 when it was launched on the NASA AURA satellite.
- Further development (including the simpler sGrid alternative to Globus toolkit) resulted into the GridAssist software suite, which has also been or will be deployed supporting users in the following environments:
 - GaiaGrid, in which GridAssist controlled a simulation of the entire GAIA mission (covering five years of observations of 1000 astrometric binary stars) in 4 days, by executing 183 jobs on 23 nodes distributed in 8 institutes of 5 countries.
 - “THEmatic Vertical Organisations and Implementation of Collaborative Environments” (THEVOICE) project, in which GridAssist provided the basis for a grid-based infrastructure and webservices in an ESA study that aimed to investigate the understanding of requirements in selected science and application communities and to demonstrate and validation of innovative networked collaborative working modes
 - SarGrid, a precursor of a scalable infrastructure that can support large scale SAR processing
 - Dutch Space internally supporting Earth Observation instrument studies.
 - The Tropospheric Monitoring Instrument Integrated Development Environment (TIDE), and its successor Downstream Evaluator for Scenes of TROPOMI Operational Products (DEScTOP).

The expertise from the above developments was contributed to the EU FP6 framework project “Dissemination and Exploitation of GRIDs in Earth science” (DEGREE), which ran from 2004 to 2006, and aimed to increase the awareness and uptake of GRID technology and infrastructure by EU Earth Science Industry and Research communities.

The expertise and toolset for easy-to-deploy frameworks for Grid-based workflow processing that supports distributed e-collaboration provided the starting point for the development of ‘Yggdrasil’, a toolkit for a distributed virtual data centre. It provides the means to make data accessible, through a web portal, from geographically distributed data providers. A prototype of the Yggdrasil virtual data centre has been deployed for the national ‘Climate Changes Spatial Planning’ programme, and for LOFAR (Low Frequency Array), a project to build a ‘software radio telescope’ from an interferometry array of 25000 dipole antennas distributed across the Netherlands and further.

TropOMI and related processing chain

Tropospheric Monitoring Instrument (TropOMI) is a nadir-viewing grating-based imaging spectrograph in the line of OMI and SCIAMACHY. TROPOMI is planned for ESA’s upcoming Sentinel-5 precursor satellite.

To facilitate early design trade-offs, end-users of the instrument work together in a study called TIDE (TropOMI Integrated Development Environment). In this study, a simulation of the instrument to be built is coupled with a chain of various processing algorithms in order to obtain data products that are evaluated to improve the future usability of the instrument under development.

The TIDE processing chain, as shown in Fig. 1, is built around the instrument simulator that produces level 0 data as it is sent to ground by telemetry. This simulator is immediately followed by a first ground processing step that is in essence an inverse instrument simulator. This step produces observation data, called level 1, from the raw level 0 instrument data converted to science parameters and corrected with calibration information. This part is the domain of the instrument developers (Dutch Space and TNO).

The instrument simulation is fed with high resolution spectra computed by radiative transfer models from predefined trace gas columns specifically selected for the required simulation. The level 1 data products, produced by the instrument and ground station simulator, are further processed to obtain the following Level 2 data products: NO₂, CO, CH₄. This part is the domain of KNMI and SRON.

Series of trace gas columns can be placed in a geographical scene which when run through the instrument simulator chain results in geographically annotated level 2 data. This information is used by value-adders, represented in TIDE by Argoss, to provide end-users with nice recognizable maps showing air quality information.

Network of Simulators

In systems with many stakeholders and disciplines, simulators consist of many models of subsystems and models of external interfacing systems. These models often have a different origin and underlying specialism than others. Thus, building a complete system simulator is of the same complexity as the system itself. This complexity is often avoided by restricting the scope of the simulations or by using inaccurate and simplified models of subsystems. An additional reason for this approach is to limit the required processing power and execution time.

By including (early) models from suppliers, the specialised experience and heritage of the supplier is re-used and wasteful duplicate modelling work is avoided. When timing constraints are not (yet) of relevance, the subsystem simulation can be provided as an internet service. One advantage of a simulation service is that the simulator and its environment can be maintained and developed by the supplier himself. Mind that quality aspects, such as configuration control remain important.

With subsystem simulators being provided as services, different stakeholders can compose the system simulation that is relevant for their part in the development using the same components as others, therefore assuring consistency. In a distributed simulation, components do not compete for processing power. Similar to having simulator instances for several stakeholders, a simulation service can be backed by multiple processors. These resources are then available for all stakeholders, introducing both resource redundancy that improves availability and improved usage that reduces the amount of resources needed.

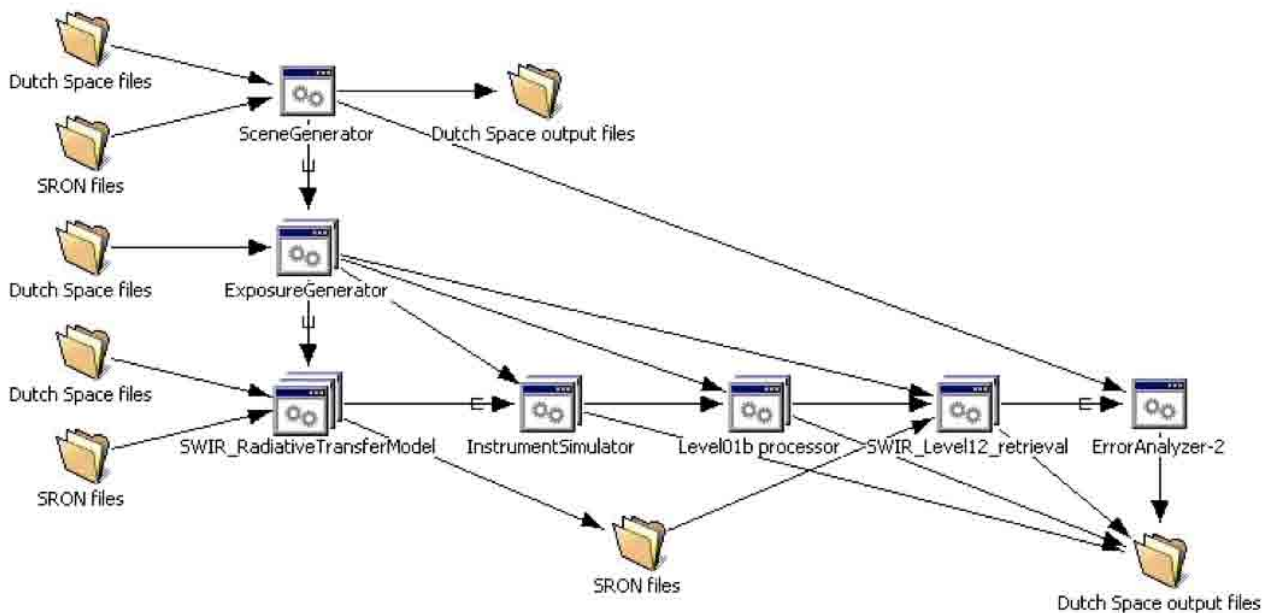


Fig. 1. TropOMI processing chain/workflow

Workflow and Access Handling

TIDE makes use of the GridAssist software framework, which in turn utilizes the Grid computing system. GridAssist is developed and maintained by Dutch Space. Grid computing is a form of distributed computing where a virtual computer is created, which consists of a cluster of networked computers that act together to perform large computing tasks. The clustering of the network is usually fairly loose. GridAssist is a software framework that provides benefits of computing in a Grid environment to applications that are not inherently 'Grid-aware', for users who are not experts on Grid technology. It provides a portal for access to applications, resources and data using high-speed networks, a scenario builder and a controller for scheduling jobs.

Workflow tool

A key component of GridAssist is its graphical user interface. Hidden behind the GUI is a lot of complex technology that the user need not be aware of. An example of the GUI created by GridAssist is shown in Fig. 2. The window shows three tabs:

Workflow: The GridAssist Workflow Pane allows the user to construct workflows in an easy way. The tool provides a graphical user interface where the user can create workflows using the drag-and-drop principle.

Monitor: The GridAssist Monitoring Pane allows the user to track the progress of the jobs that he/she created in the GridAssist Workflow Pane. It also allows the user to terminate a job.

Registry: The GridAssist Registry Pane can be used by administrators and application integrators to manage the information about applications and resources available for the user. This information is stored in the GridAssist Registry.

The large window on the right contains the main graphical display of the workflow. Three main items can be used:

Applications: Applications are the executable programs.

Data Storage: These are the directories where the input and output data are stored. The files that are transferred can be viewed by moving the mouse over the connection that links to the Data Storage.

Data connections: Between applications themselves and between applications and data storage, data connections need to be established. The workflow tool will check the correctness of the connections between the services, by checking that the output produced by one application matches the type of input required or expected by the destination application.

The GridAssist workflow tool greatly assists in easily setting up a workflow. Here we describe the main steps that need taken to build such a workflow. First the user starts the Workflow Tool on his own client machine. GridAssist needs

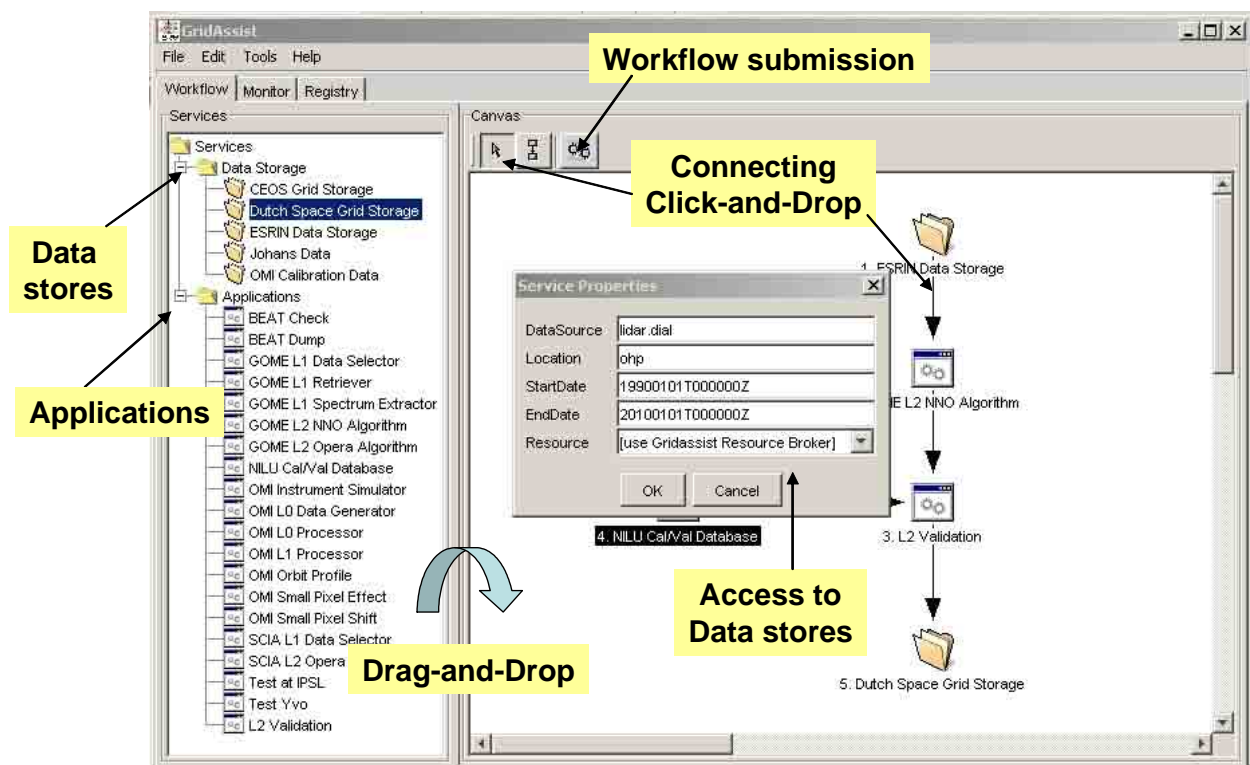


Fig. 2. Workflow tool

Java RE for it to run. The next step is that the tool queries the Registry in order to obtain information about available services. The workflow tool presents these services in a kind of palette in the left panel. The items in that left panel can now be dragged-and-dropped in the right panel. In the right panel, they will appear as icons. Now, the most important step is to connect the selected services. This is done by first clicking on the “link” button at the top of the right panel and then by drawing a line between the two services. The contents of the connection can be inspected by moving the mouse over the line (e.g. the file to be written is shown), and edited by double clicking on the line.

When the user has completed defining the workflow, it can be saved as a local file for later use. The next step is to click the submit button, which submits the workflow to the Workflow Controller. After the workflow has been submitted, it will be processed automatically, without requiring user intervention.

Grid Processing ‘Under the Hood’

After the workflow has been submitted, the workflow is parsed by a workflow engine. The workflow engine arranges file transports and decides which application needs to run when and how often. A request to run an application is submitted to the dispatcher, which finds a suitable processor for the application and transports the files as specified.

For the provision of applications as services, a simple toolkit was created in previous project called sGrid, as the existing grid toolkits were found to be complex and difficult to install. Using sGrid it is easy to add an application processor to the grid system from behind a firewall, requiring standard http access to the internet only (like a browser).

Each application needs to be administered as a grid service (deployment). To fit in with other applications, interface control documents can be used to agree on the details. This should contain definition of input and output files, specification of commands, required architecture and properties that specify whether an input file is optional or whether one-or-more files are expected. This might not seem very complex, but based on experience it is advised to involve an information technologist to support this task.

Virtual Data Centre

Based on techniques used for distributed computing, a Grid-based e-collaboration extension has been made with the Yggdrasill Virtual Data Centre. This toolkit makes it possible to access data, through a web portal, from geographically distributed data providers.

Yggdrasill is based on the sGrid technology, which is used to:

- Send announcements about new data products from the data providers to the central catalogue, which can be queried by users through the web portal
- Fulfil orders placed by users, by sending the order from the portal to the remote data providers, and retrieving the related data product from the remote data provider to the portal, where it can be downloaded by the user.

The system supports customized data products, where users can specify settings for customization parameters in the web portal in order to obtain custom-made data. The customization process may include data format conversion, visualization or even complete model runs to create the desired output.

A prototype of the Yggdrasill virtual data centre has been deployed for the national ‘Climate Changes Spatial Planning’

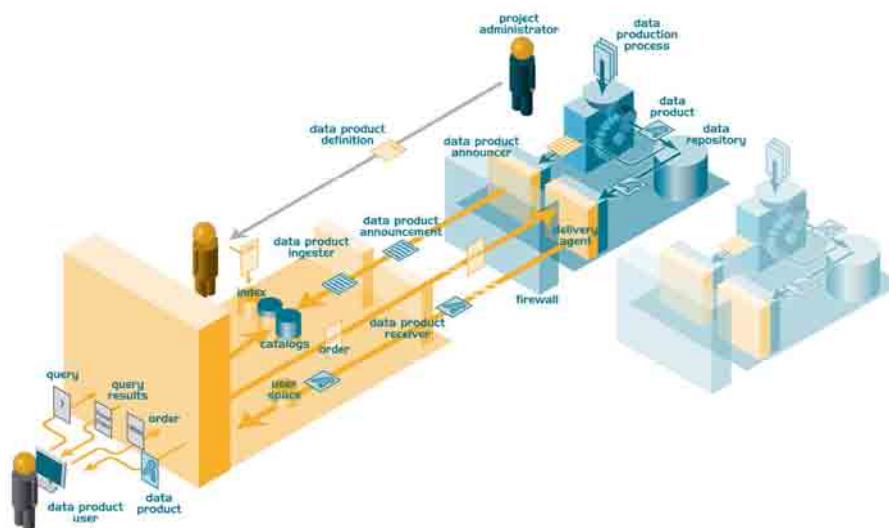


Fig. 3. Yggdrasill virtual data centre

programme[3], which focuses to enhance joint-learning between those to communities and people in practice within spatial planning in order to get climate change and climate variability one of the guiding principles for spatial planning in the Netherlands.

Yggdrasil has also been deployed to support the LOFAR[4] (Low Frequency Array), a project to build a ‘software radio telescope’ from an interferometry array of 25000 dipole antennas distributed across the Netherlands and Germany, extending into Great Britain and France, and possibly in other European countries such as Poland, Sweden and Ukraine.

USING TIDE

For instrument and processing characterisation, a homogeneous input scene is used to set up the computing chain. However, the system can also be used using more realistic atmospheric input scenes for the Netherlands. The left panel of Fig. 4 shows the results for one of the trace gases, CH₄, as retrieved from the simulated observations of the instrument. The same simulation run was repeated with a 2 degree temperature difference applied to the instrument optics. The final result in terms of the relative difference with the first run is shown in the right panel of Fig. 4.

REFERENCES

- [1] H. Bovensmann, J.P. Burrows, M. Buchwitz, J. Frerick, S. Noel, V.V. Rozanov, K.V. Chance, A.P.H. Goede, “SCIAMACHY: mission objectives and measurement modes”, J. Atmos. Sci. 56, 127-155 (1999).
- [2] P.F. Levelt, G.H.J. van den Oord, M.R. Dobber, A. Malkki, H. Visser, J. de Vries, P. Stammes, J.O.V. Lundell, H. Saari, “The Ozone Monitoring Instrument”, IEEE Trans. Geosci. Remote Sens. 44 (5), 1093-1101 (2006).
- [3] <http://www.klimaatvoorruijnte.nl>
- [4] <http://www.lofar.org>

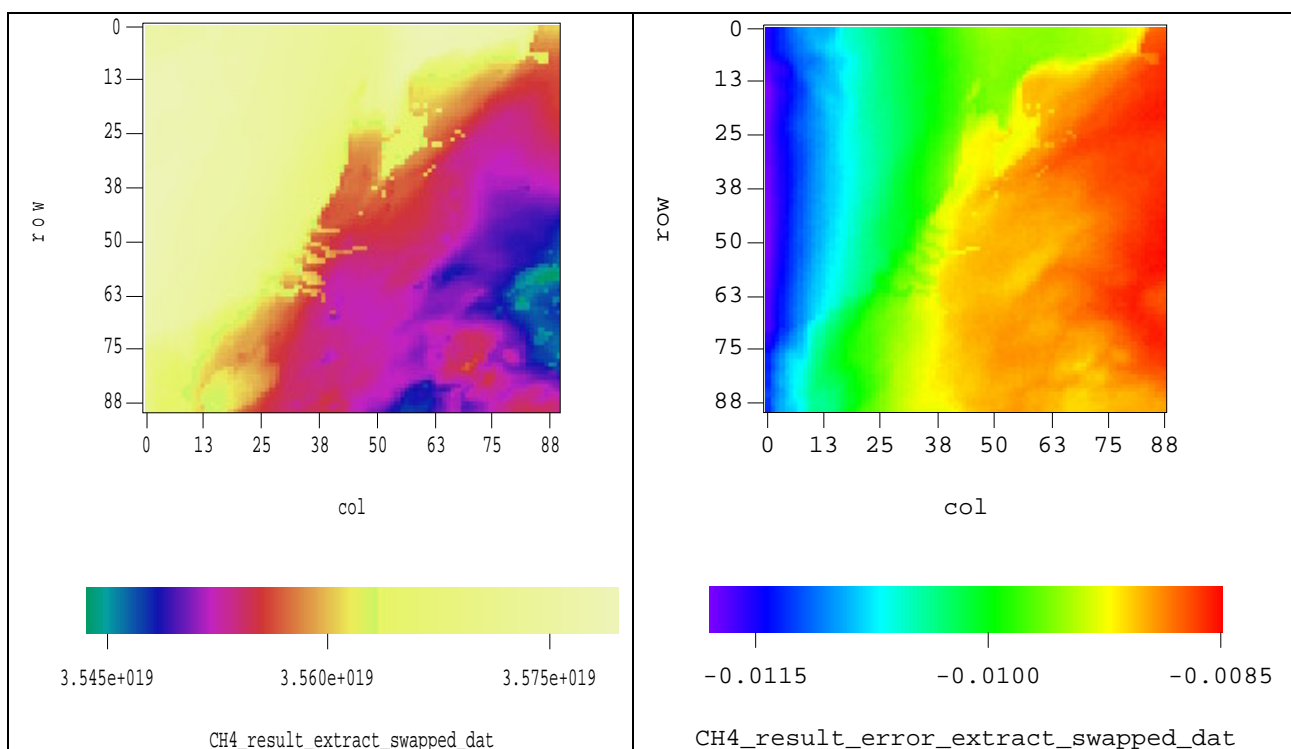


Fig. 4. retrieved CH₄ field (left) and the effect of 2 degrees temperature difference (right).