THE TECHNOLOGY AND ATMOSPHERIC MISSION PLATFORM (TAMP)

Stefano Natali[(1)], Simone Mantovani[(1)], Gerhard Triebnig[(2)], Daniel Santillan[(2)], Marcus Hirtl[(3)], Thorsten Fehr[(4)]

[(1)] SISTEMA GmbH, Vienna, Austria
[(2)] EOX IT Services, Vienna, Austria
[(3)] Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
[(4)] ESA ESTEC, Noordwijk, The Netherlands

ABSTRACT

The scientific and industrial communities are being confronted with a strong increase of satellite missions and related data. This is in particular the case for the Atmospheric Sciences communities, with the upcoming Copernicus Sentinel-5 Precursor, Sentinel-4, -5 and -3, and ESA's Earth Explorers scientific satellites ADM-Aeolus and EarthCARE. The challenge is not only to manage the large volume of data generated by each mission / sensor, but to process and analyze the data streams. Creating synergies among the different datasets will be key to exploit the full potential of the available information. Integrating Earth-Observation data with ground based observations and numerical models, is the basis for a new data exploitation paradigm which opens new research and commercial opportunities.

As a preparation activity supporting scientific data exploitation for Earth Explorer and Sentinel atmospheric missions, ESA is funding the technology study and prototype implementation of the "Technology and Atmospheric Mission Platform" (TAMP). The services and tools are developed along use cases defined with users from different scientific and operational fields and implemented according to their requirements to ensure acceptance of TAMP platform by the atmospheric community.

The current work provides summary information about the collected requirements, the TAMP system design and implementation.

Index Terms— Atmospheric Applications, Big Data, Collaborative e-Infrastructure, Data Visualization, Virtual Research Environment

1. INTRODUCTION

New generations of satellite-borne sensors (e.g. Atmospheric Sentinels, Earth Explorer missions) will provide an unprecedented amount and variety of data to be used by the Atmospheric Sciences community and to complement / to be assimilated in numerical weather predictions. As in many other fields, Atmospheric Sciences are experiencing / facing a new challenge: moving from the observational science, to the theoretical science, then to the simulation science, reaching the intensive / massive data exploitation science (.e.g see the Fourth Paradigm, [1]).

Atmospheric Sciences will get a huge benefit from the combined and effective use of a large variety and volume of data that completely fit with the concept of big data (see the 4Vs of Big Data, [2]). As example, Hirtl et al. demonstrate the improved performance in forecasting particulate matter concentration assimilating satellite-based Aerosol Optical Thickness (AOT) maps into numerical models [3].

In order to be ready to exploit upcoming satellite datasets, there is a need to set up and demonstrate the feasibility of a new approach, enriching as much as possible the Atmospheric Sciences scenario.

The role of the potential system users is crucial in the development of a new platform: users can define needs and potential limitation of the system, and its usefulness in their daily work. Furthermore there is a need to establish new collaboration modalities and tools among users: the concept of virtual laboratory shall be implemented, promoting large centralized infrastructure investments than local small ones,

Technological challenges to be faced for the implementation of the TAMP platform span from the big data management related to multi-source and multi-dimensional datasets (see e.g. [8], [9], [10], [11], [12]), to multi-dimensional data visualization needs (see, as example, the V-MANIP project portal [7]), to remote and massive data processing, e.g. managed via message brokers and task dispatched technologies (see, as example, [12], [13]).

In the current work, the development of the TAMP platform is described, highlighting on the one hand the fundamental role of users in the definition of the platform requirements and in its validation, on the other the strong technological effort of the project partners in using cutting-edge technologies to implement the demonstrator.

2. REQUIREMENTS DEFINITON

The definition of the system requirements represents the fundament on which the platform is built. Requirements have been defined following user, data, and technological needs and constraints. A set of high level European scientific groups (hereafter referred as Scientific and Technical Forum, STF) have been selected to define the user requirements to be successively translated into system requirements. The STF members belong the European Centre for Medium-range Weather Forecast (ECMWF, UK), the German Aerospace Center (DLR, Germany), the Belgian Institute for Space Aeronomy (BIRA, Belgium), the Royal Netherlands Meteorological Institute (KNMI, The Netherlands), the National Institute of Research and Development for Optoelectronics (INOE, Romania), the Austrian Meteorological Service (ZAMG, Austria).

The STF members have been requested to describe a potential use of the TAMP platform defining:

- Which data to use (satellite-based data, numerical models, validation datasets
- How these products shall be prepared to be used for their specific scopes (pre-processing activities such as subsetting, remapping, …)
- How these products can be used together (e.g. comparison between two products providing the same field, combined processing, …)
- How these products / processing results could be effectively visualized and manipulated via advanced multi-dimensional graphic user interfaces (GUI)
- How the results of the analysis / processing shall be exported or directly published onto specific media.

As result of the consultation, a set of 272 unique user requirements have been collected, to be translated into system requirements. The full traceability of the technical (system and modules) requirements has been achieved by means of cross reference matrices to ensure that any change in user or technical requirements can be correctly addressed. Finally eight use cases have been selected to be supported by the implemented software to validate the system performance.

3. SYSTEM DESIGN

The implemented TAMP platform will be deployed on a remote infrastructure providing access via web browser as well as via direct machine connection (e.g. secure shell connection). The platform shall store a large amount of data and shall make available computational resources to perform massive data processing.

Three main logical layers are defined to represent the TAMP platform:

- The Graphic User Interface GUI) layer, containing:
    - The Public Information Portal (PIP): it shows the project main features, providing a social-like page where posts from the system administrator (e.g. availability of new datasets) and users (e.g. specific results achieved via the platform) are available
    - A user page, from which each user can configure preference, post news and access to the Data Analysis and Visualization Environment (DAVE)
    - A data ingestion page, from which a user can upload owned data to be used within the platform
    - The Data Analysis and Visualization Environment (DAVE), from which the user can access to the data, visualize, select, run data assessment tools and save the processing results.
- The server-side layer, that contains five main modules:
    - The processing libraries, a set of data access and processing libraries to support processing modules
    - The data ingestion module, to pre-process and store system retrieved data as well as user provided data
    - The Web Coverage Service (WCS) data access module, to allow internal and external modules to retrieve the stored and processed data via standardized interfaces
    - Data processing modules, where data assessment and processing modules are deployed
    - User virtual machines: these are computational resources available for the users, from which the users can access to all system-available data, deploy owned processors, run the processors on system data and visualize the results via the DAVE GUI.
- The data storage layer, that contains both the system collected or user provided data and the processing results, in two separated areas. Both storage areas can be accessed via the WCS module.

4. SYSTEM IMPLEMENTATION

To pursue the virtual laboratory approach, a dedicated infrastructure to host the TAMP platform has been made available by the Austrian Meteorological Service: 64 processors, 128 GB RAM and 5TB storage are shared amongst the system modules and user virtual machines.

Data to be collected are summarized as follows:

- Satellite-based Ozone products (total column, profile) from SCIAMACHY, MIPAS, GOMOS, GOME, SCISAT, ODIN
- Satellite-based Aerosol products (total column, profile) from SCIAMACHY, MIPAS, GOMOS, GOME, OMI, MODIS, CALIPSO
- Satellite-based NO2 products (total column, profile) from SCIAMACHY, MIPAS, GOMOS, GOME, OMI
- Satellite-based CH4 products from MIPAS

- Satellite-based SO2 products from OMI
- Numerical weather prediction models data coupled with chemistry (MACC model data, BASCOE Reanalysis, WRF/CHEM simulations) low resolution/ global coverage and high resolution/ regional coverage
- Correlative data from AERONET, EARLINET, CLOUDNET, PANDONIA, EMEP emission inventory, TNO emission inventory, ESA EVDC, NASA AVDC, ACTRIS (for SO2)

The platform makes available the following functionalities for data (pre-)processing and data assessment:

- Pre-processing and processing tools
  - Remap data onto a pre-defined geographic grid
  - Additions / subtractions of data
  - Units Conversion
  - Vertical integration
  - Combined spatial-temporal means
  - Extraction/ use of quality flags
  - Filtering of level 2 data according to quality flag
  - Chemical speciation monitor
  - Level 3 product generations including averaging kernel (by users)
- Data Assessment
  - Compute basic statistical tests (subset of the DELTA A&P benchmarking tool)

The users are allowed to update and run software packages and scripts in the following languages:

- Fortran (90)
- C / C++
- Python
- Shell scripts

Data can be extracted from the platform in netCDF(-CF) format; plots can be exported in png / ps / pdf format.

The TAMP platform is, at this stage, a demonstration platform freely accessible by STF members and associated scientists of the same institutions. Nevertheless its design and implementation technologies allow a quick system upscale and transfer to operation to accommodate requests from a wider community. The future evolution of the platform will be defined based on the STF members' feedback and ESA follow-on strategies.

5. RESULTS AND CONCLUSIONS

The upcoming scenario of data availability is driving the development of new concepts of research environments, where large data storage and processing capabilities are available to a wide range of users that can jointly develop and validate new algorithms. The TAMP project aims at demonstrating the approach for Atmospheric Sciences users, where a strong increase of satellite-based products is expected in the near future, and where multi-sensor, multi-temporal and multi-dimensional data can be used in synergy to improve current and future products, paving the way for a future data assessment operational platform.

6. AKNOWLEDGEMENTS

The current study is funded by the European Space Agency under the contract: Technology and Atmospheric Mission Platform (C.n. 4000112832/14/I-CS).

7. REFERENCES

[1] Tony Hey, Stewart Tansley, and Kristin Tolle (Eds.). The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, Redmond, Washington, (2009).

[2] http://en.wikipedia.org/wiki/Big_data Big Data Wikipedia Page (visited on 06.11.2015).

[3] Hirtl, M., Mantovani, S., Krüger, B. C., Triebnig, G., Flandorfer, C., Bottoni, M., & Cavicchi, M.: Improvement of air quality forecasts with satellite and ground based particulate matter observations. Atmospheric Environment, 84, 20-27 (2014)

[4] Sentinel-3 Team GMES-S3OP-EOPG-TN-13-0001 Sentinel-3 User Handbook Issue: 1.0'Date: 2 Sept. 2013

[5] Paul Ingmann, B. Veihelmann, J. Langen, D. Lamarre, H. Stark, G. BazalgetteCourrèges-Lacoste: Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p. Remote Sensing of Envi-ronment 120 (2012) 58–69.

[6] Various Authors: ADM-Aeolus Science Report. European Space Agency publication, SP-1311, 'ADM-Aeolus' April 2008, ISBN 978-92-9221-404-3, ISSN 0379-6566

[7] http://demo.v-manip.eox.at/ Interactive geospatial explorations (V-MANIP Demo Applications) (visited on 12.02.2016).

[8] http://www.esa.int/esaLP/LPearthcare.html EarthCARE Web Portal (visited on 12.02.2016).

[9] http://www.esa-aerosol-cci.org/ ESA CCI Aerosol web site (visited on 12.02.2016).

[10] http://www.esa-cloud-cci.org/ ESA CCI Cloud web site (visited on 12.02.2016)

[11] http://www.esa-ozone-cci.org/ ESA CCI ozone web site (visited on 12.02.2016).

[12] http://www.esa-ghg-cci.org/ ESA CCI greenhouse gases web site (visited on 12.02.2016).

[13] http://www.celeryproject.org/ Celery home page (visited on 12.02.2016).

[14] http://www.rabbitmq.com/ RabbitMQ home page (visited on 12.02.2016)