

Appendix Q

ESATAP - Handling large thermal results data with HDF5

François Brunetti
(Dorea, France)














ESATAP
Handling Large Thermal Results Data with HDF5

ESA/ESTEC project managers:
Harrie ROOIJACKERS Hans Peter DE KONING

Authors:
alain.fagot@dorea.fr
francois.brunetti@dorea.fr

DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00










ESATAP Quick Presentation




- **ESATAP is a flexible post-processing tool of STEP-TAS/STEP-NRF data files.**
- **Main features:**
 - Handle the standard STEP-NRF thermal results data files.
 - Allow to implement own post processing scripts (components)
 - Provide a toolbox of around 40 basic components for statistical, thermal, mathematical calculation, heatflow, etc.
 - Provide interfaces with MS Word, MS Excel, OpenOffice, GnuPlot, etc.
 - Provide a graphical user interface for non-programmers allowing to easily implement/maintain/improve/enrich the Toolbox and/or own Local components or structured complex tasks.

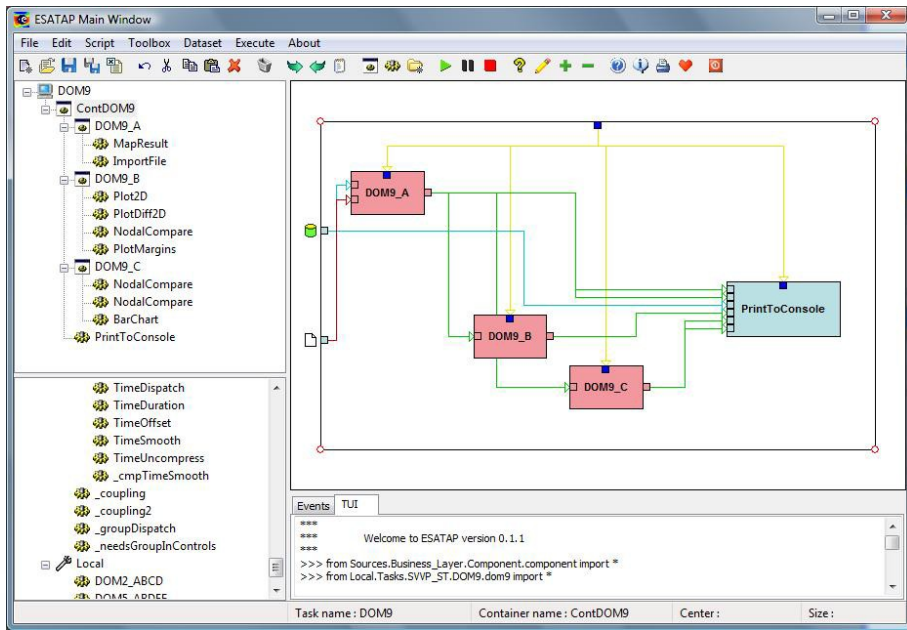
DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA



ESATAP Quick Presentation






21th European Workshop on Thermal & ECLS Software - DOREA

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00



Problematic (1)


- **Thermal analysis needs:**
 - Due to the improvement of processor capabilities, memory, disks and clusters: the size of thermal models is always increasing.
 - Thermal post processing analysis is software, hardware and data consuming: some model reduction techniques are implemented in order to reduce the large amount of data.
- **Classical model results data size evaluation**
 - Several thousands of nodes,
 - Several hundred thousands of conductors,
 - Many quantities (Temperatures, conductances, capacitances, areas, powers, etc.)
 - Several hundred of time steps (several orbit periods)



21th European Workshop on Thermal & ECLS Software - DOREA

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

Problematic (2)






- **Raw data size:**
 - Minimum results data array size order:
 - $100000 \times 100 \times 10 = 10$ Mega of doubles
 - > 80 Mbytes for conductors
 - Model data size:
 - $1000 \times 100 \times 10 = 8$ Mbytes
- **Data exchange format overhead:**
 - STEP-TAS part 21 (ASCII):
 - NRF datacube : 99% of the size, model 1 %
 - > 80 Mbytes of raw data \Leftrightarrow > 300 Mbytes
 - GFF : > 500 Mbytes
- **Objectives:**
 - An operational model issued from solvers generates a large amount of data, difficult to exchange. Improve the thermal data exchanges,
 - To reduce the time response as much as possible on results data access read to anticipate the needs.



21th European Workshop on Thermal & ECLS Software - DOREA

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

Studied solution



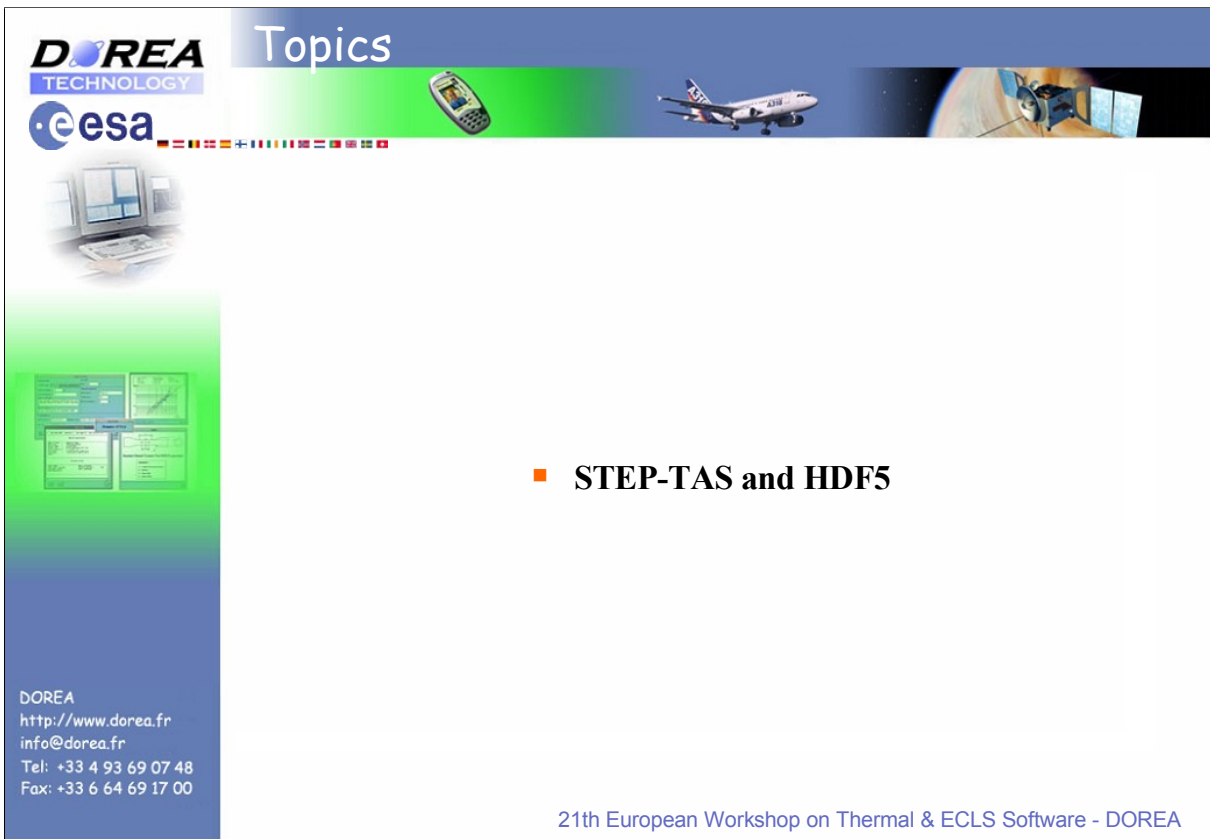



- **Solution:**
 - Use of a standard binary format, portable with indexing facilities,
 - I/O libraries allow to have the both archiving Part 21 ASCII format and HDF5 binary format for operational use.
 - Implement a generator able to provide the STEP-TAS (thermal) API for both part 21 and HDF5 data format.
 - Provide an efficient and optimized API able to quickly access slices of data within a large set of result values.

=> Adapt STEP-TAS SDK to HDF5

21th European Workshop on Thermal & ECLS Software - DOREA

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00



DOREA
TECHNOLOGY

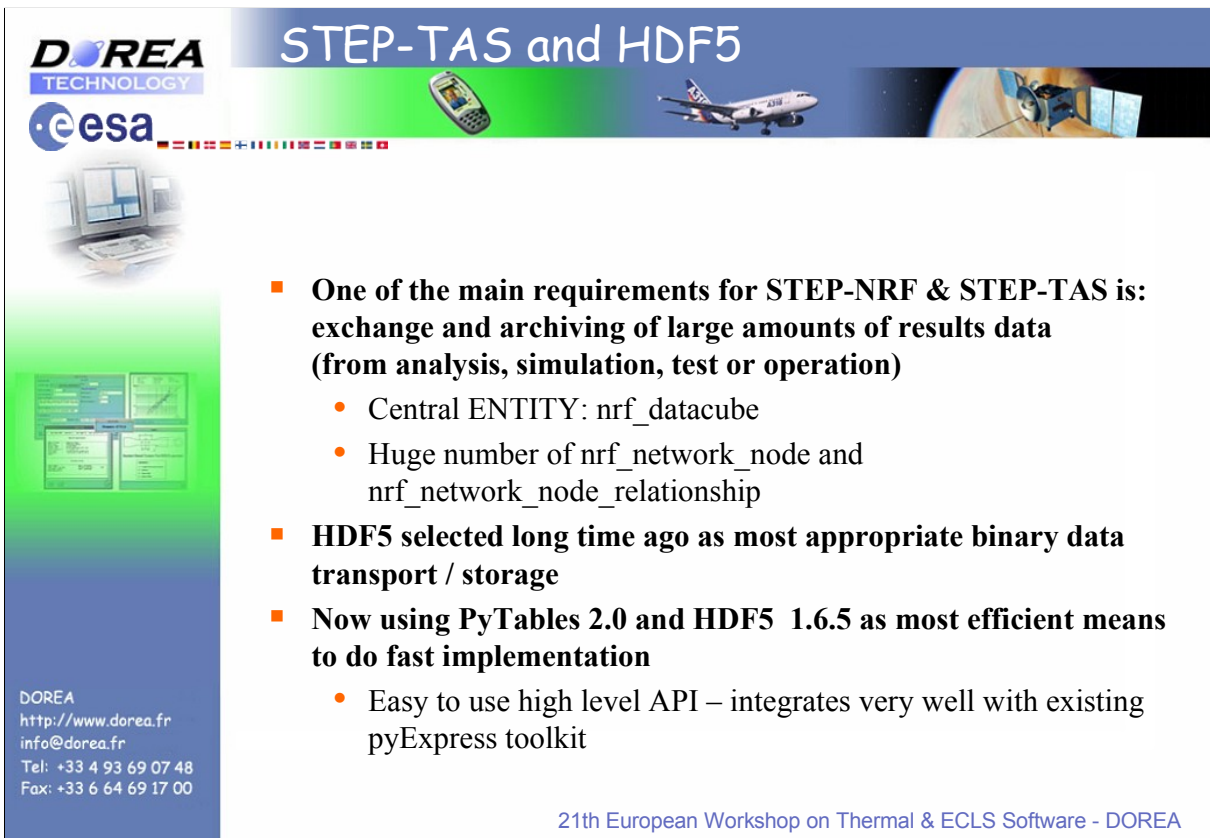
esa

Topics

- **STEP-TAS and HDF5**

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA



DOREA
TECHNOLOGY


esa

STEP-TAS and HDF5




- **One of the main requirements for STEP-NRF & STEP-TAS is: exchange and archiving of large amounts of results data (from analysis, simulation, test or operation)**
 - Central ENTITY: nrf_datacube
 - Huge number of nrf_network_node and nrf_network_node_relationship
- **HDF5 selected long time ago as most appropriate binary data transport / storage**
- **Now using PyTables 2.0 and HDF5 1.6.5 as most efficient means to do fast implementation**
 - Easy to use high level API – integrates very well with existing pyExpress toolkit

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA



Forseen Advantages of using HDF5



- **HDF5 retrieves data on demand from disk**
 - Contrary to Part 21 format where one must load the whole dataset in memory
 - HDF5 allows real time compression and chunking of Data
- **Forseen advantages on huge models**
 - Run time memory is maintained at reasonable level
 - HDF5 Dataset files are much smaller due to compression.
 - STEP-TAS HDF5 API allows to handle huge datasets where Part21 API failed.

DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA




Topics






- **STEP-TAS and HDF5**
- **STEP-TAS SDK**


DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00


21th European Workshop on Thermal & ECLS Software - DOREA



STEP-TAS HDF5 python SDK today

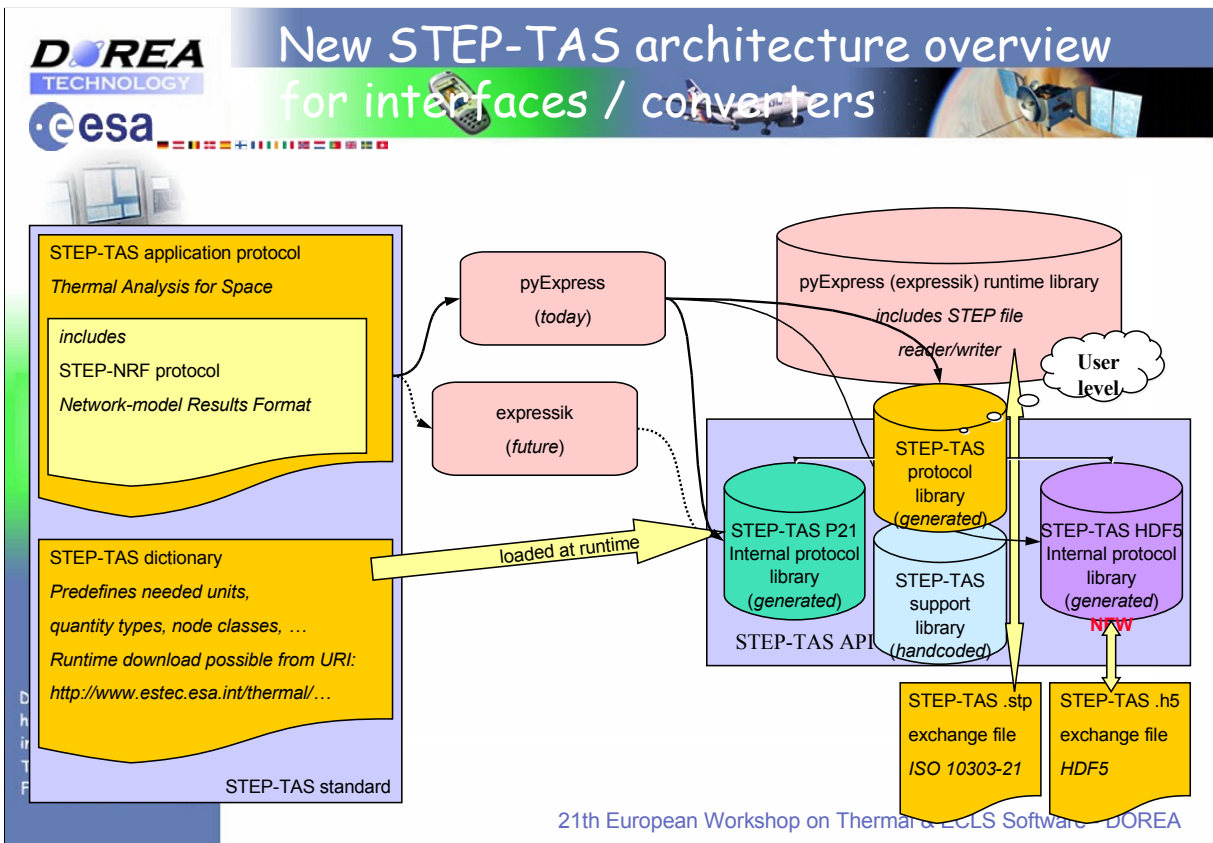





- **First release in Python language (Python 2.4.x)**
- **Based on:**
 - pyTables 2.0 <http://www.pytables.org>
 - HDF5 1.6.5 (NCSA) <http://hdf.ncsa.uiuc.edu/HDF5>
 - **Both Open Source**
- **SDK generated by PyExpress API generator**
- **STEP-TAS SDK keeps previous SDK's use and behavior while optimizing performance on large datasets.**
 - API remains the same at user's point of view,
 - The only difference is the selection of the repository format at Dataset creation.




DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA






STEP-TAS HDF5 python SDK - tuning





- **HDF5 offers a low memory consumption for large models**
- **But data access to HDF5 data can be time consuming**
 - HDF5 representation of an EXPRESS model consists in consequent number of tables (similar to Relational Databases Tables) and arrays.
 - Reading data of an Express instance means reading data in several HDF5 tables/arrays
- **As to provide best execution time/Memory management, STEP-TAS generated SDK implements a fine handling of data between run-time memory and HDF5 datasets.**

DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA





Topics




- **STEP-TAS and HDF5**
- **STEP-TAS API**
- **Some Benchmarks**


DOREA
http://www.dorea.fr
info@dorea.fr
Tel: +33 4 93 69 07 48
Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA





Benchmark 1







- **Model contents**
 - 1112 thermal nodes,
 - Datacube 1: 1111 Observable item x 14 quantities x 576 states
=> **895 104** values
 - Datacube 2: 1 Observable item x 14 quantities x 576 states => 8 064 values
- **size of dataset in P21 file : 93,6 Mb**
- **size of dataset in HDF5 file : 30,3 Mb**
- **Objective of the bench**
 - The benchmark is executed twice with P21 and with HDF5 files
 - Opening dataset
 - Display lists of dataset content (Models, Nodes, conductors, quantities, results,)
 - Execute 4 « ReadObservableItemData » on first datacube (retrieve values for defined quantities, nodes, times)
 - Make a copy of input dataset





DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA

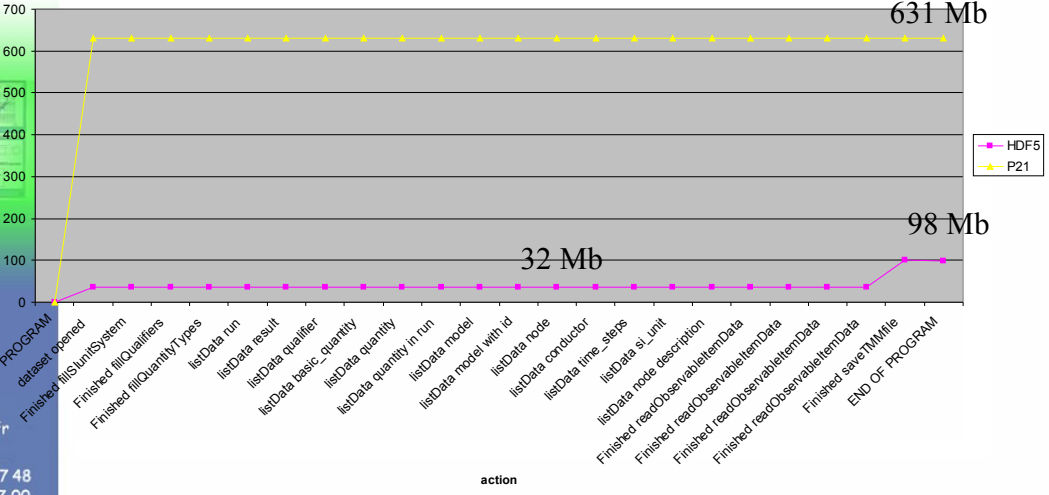



Benchmark 1 - Memory usage





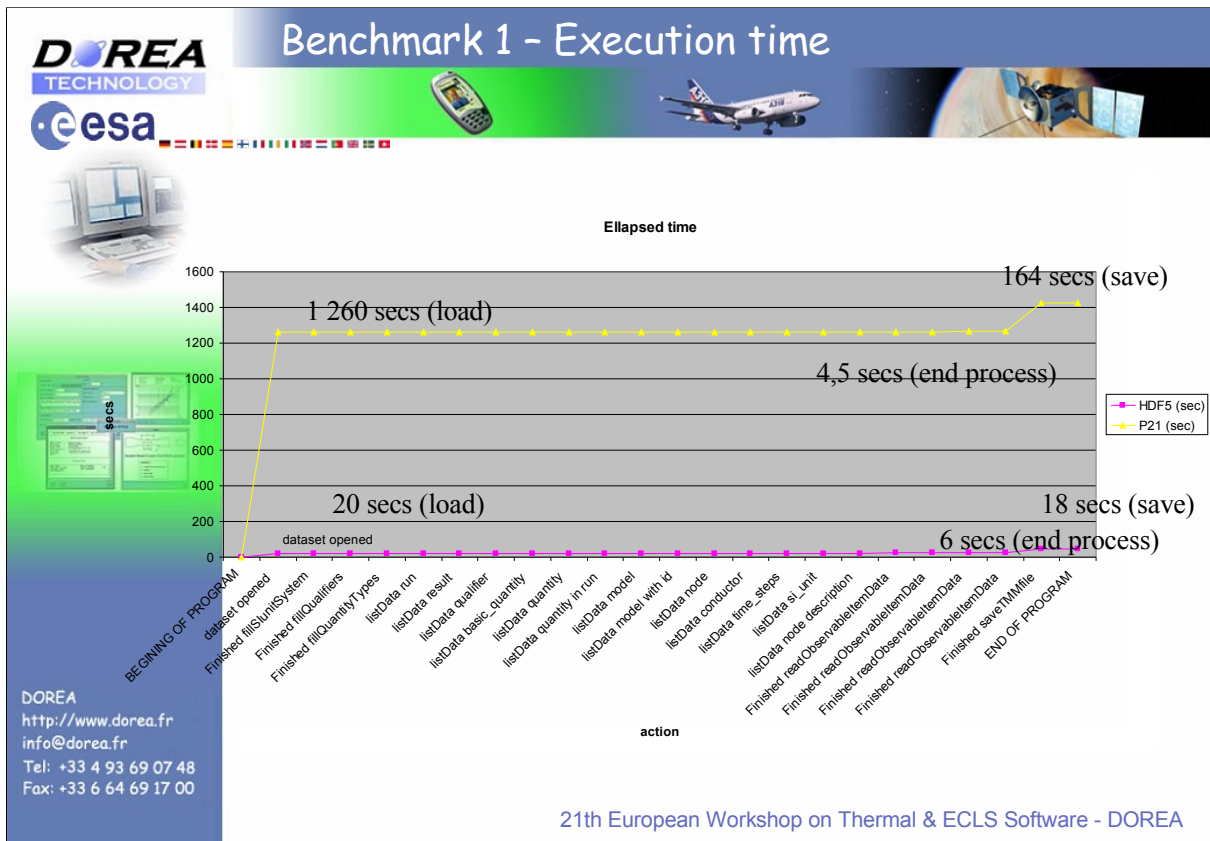
Memory used



Action	HDF5 (Mb)	P21 (Mb)
BEGINNING OF PROGRAM	0	0
dataset opened	~50	631
Finished fillUnitsSystem	~50	631
Finished fillQualifiers	~50	631
Finished fillQuantityTypes	~50	631
listData run	~50	631
listData result	~50	631
listData qualifier	~50	631
listData basic_quantity	~50	631
listData quantity	~50	631
listData quantity /n run	~50	631
listData model	~50	631
listData model with id	~50	631
listData node	~50	631
listData conductor	~50	631
listData time_steps	~50	631
listData el_unit	~50	631
listData node description	~50	631
Finished readObservableItemData	~50	631
Finished readObservableItemData	~50	631
Finished readObservableItemData	~50	631
Finished readObservableItemData	~50	631
Finished saveTMMfile	98	631
END OF PROGRAM	98	631

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA

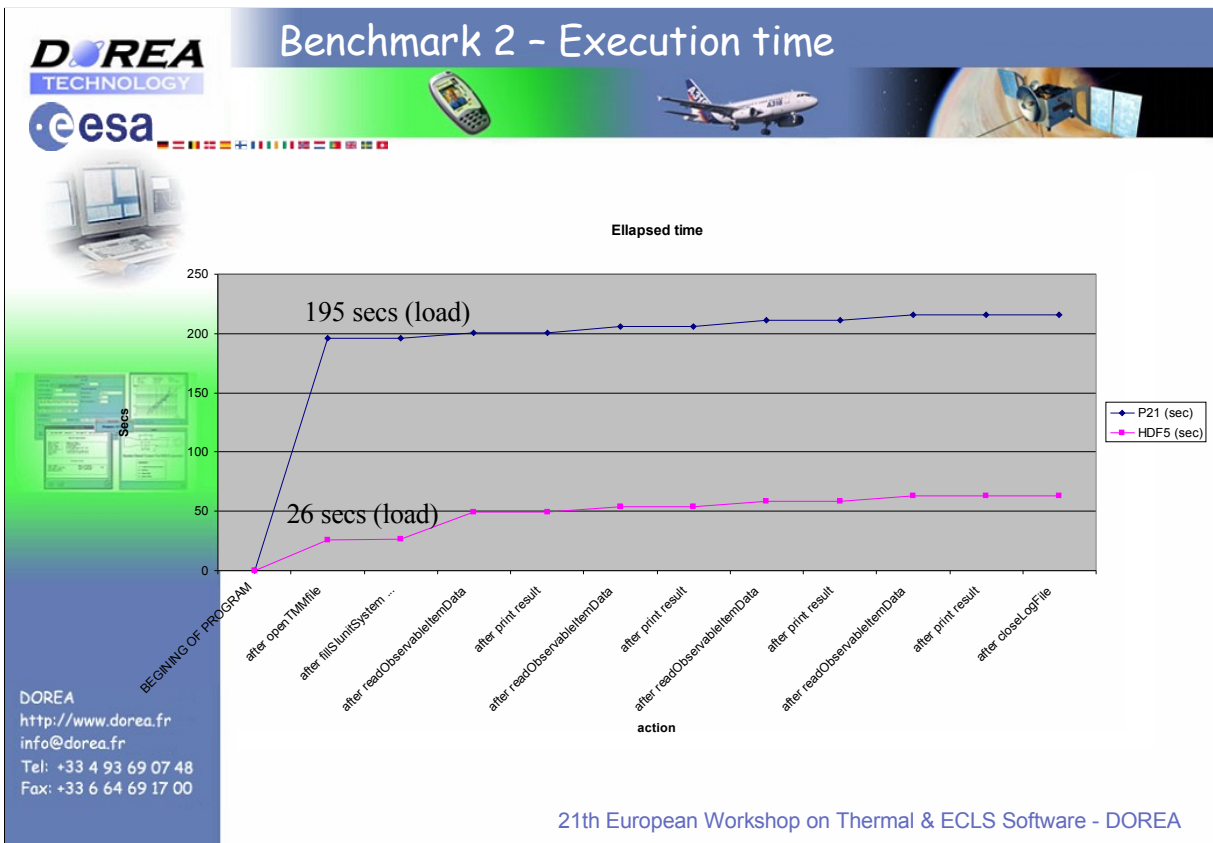
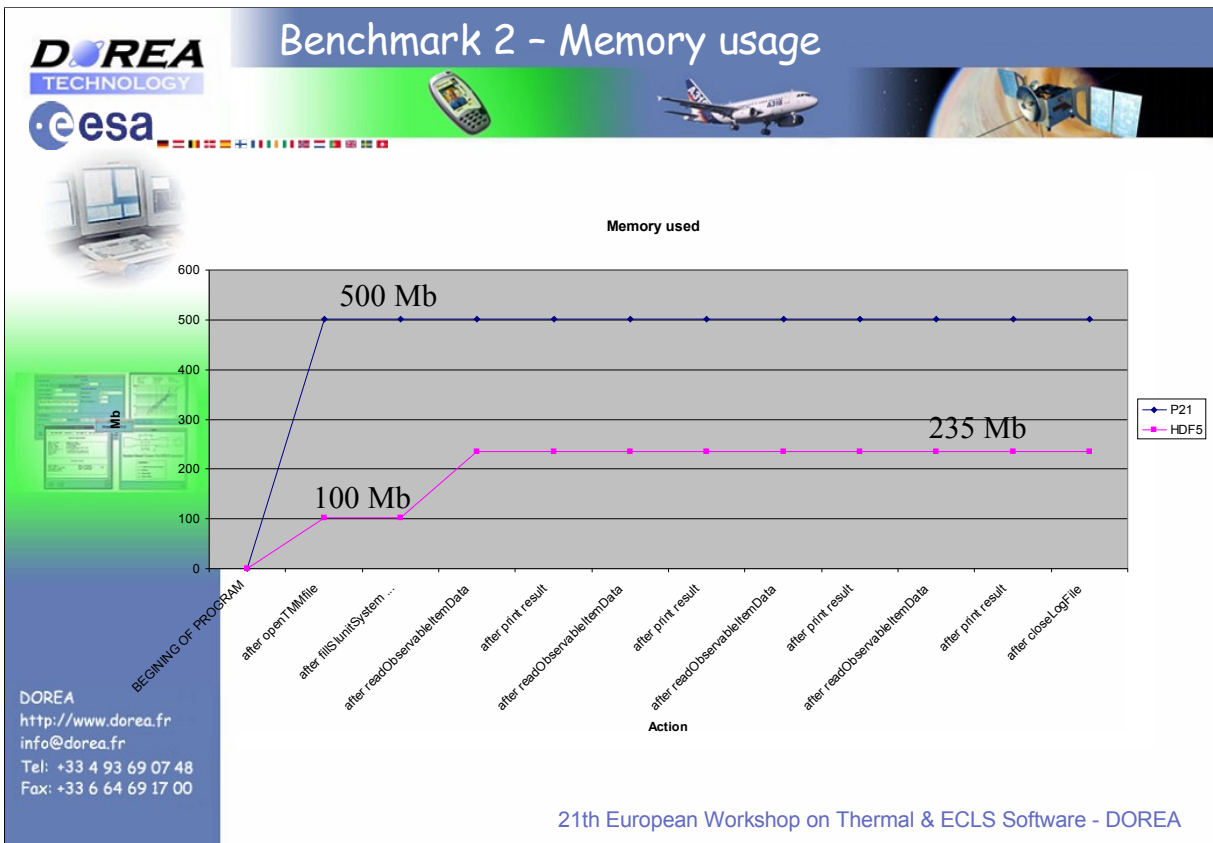




Benchmark 2

- size of dataset in P21 file : 12 Mb
- size of dataset in HDF5 file : 5,9 Mb
- Model contents
 - 1112 thermal nodes, 106 959 conductors
 - Datacubes with 15554 values to 210016 values
- Objective of the bench
 - The benchmark is executed twice with P21 and with HDF5 files
 - Opening dataset
 - Execute 4 « ReadObservableItemData » on first datacube (retrieve values for defined quantities, nodes, times)




DOREA
<http://www.dorea.fr>
 info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00



21th European Workshop on Thermal & ECLS Software - DOREA



Demo










- **Data size:**
 - Extracting and handling 2 periods of temperatures from a group of 30 nodes within 1000 nodes, 20 quantities.
- **Extracting data:**
 - Part 21 or other ASCII results formats:
 - Around 1h30 (with 1h28 of data loading)
 - HDF5:
 - < 30 s (20 s of data table structure preload when opening)
- **Simple post processing calculation**
 - Part 21 or other ASCII results formats:
 - Same order than for extraction.
 - HDF5:
 - < 60 s (20 s of data table structure preload when opening)



DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA

Conclusion

- **STEP-TAS 5.2 SDK for HDF5:**
 - Actually under system validation
 - Availability : December 2007
- **ESATAP HDF5:**
 - Distribution process is being finalized
 - Availability : January 2008
 - Platforms: Windows XP/Vista, Linux
- **ESATAN (DUMPTAS HDF5):**
 - Availability : December 2007

DOREA
<http://www.dorea.fr>
info@dorea.fr
 Tel: +33 4 93 69 07 48
 Fax: +33 6 64 69 17 00

21th European Workshop on Thermal & ECLS Software - DOREA

