The ESA Cluster Active Archive

C.H. Perry, T. Eriksson, P. Escoubet, S. Esson, H. Laakso, S. McCaffrey, S. Sanderson, H. Bowen, A. Allen and C. Harvey

Cluster 5th anniversary symposium, 19-23 Sept 2005
Introduction

• Overview
• Standards – data formats and metadata
• Infrastructure – hardware and network
• Products – raw data, processed data
• Tools and information systems
• Web based data access
• Availability for beta testing
• Summary
Overview

**Aims of the CAA:-**

- The CAA should contain (all) the Cluster high resolution data
- The data should be of the best (achievable) quality
- The data should be suitable for detailed science investigations
- Data should be publicly accessible (no data rights)
- Need to start now while expertise still available

ESA providing support to instrument teams to ensure this

In addition the CAA will hold ancillary products and support information including

- Auxiliary data (such as orbit and attitude)
- Survey data and plots
- CSDS data including JSOC parameters
- Documentation
Implementation Review – June 2005

CAA Overview

Reviews

Core team

System dev.

Standards

Reporting

Product dev.

Cluster 5th anniversary symposium
19-23 September 2005
Standards

The CAA Metadata working group has developed the CAA Metadata Dictionary (CAA-CDPP-TN-0002)

- Metadata provides a standard machine readable way of storing descriptions of the data

- The metadata is used to aid the consistent location and handling of all data products and in particular the digital parameter (science) data

- There has been much iteration between CAA and instrument teams. A minor update (v2.02) of the MDD is due out shortly

- Interoperability issues being addressed in conjunction with SPASE (Space Physics Search and Exchange)
Standards – Digital Parameter Data

The CAA Data Formats working group has developed the Cluster Exchange Format (DS-QMW-TN0010)

- ASCII format suitable for long term archival
- Simple tabular, comma separated format
- Self describing via CAA metadata header
- Standard gzip compression used for efficient storage

CAA also plan to allow receipt and delivery of data in binary NASA CDF format

- CAA will translate CDF data to CEF for long term storage

Conceptually the files making up a dataset are treated as a single time series.
Standards – Digital Parameter Data

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Code</th>
<th>Type</th>
<th>Value (°C)</th>
<th>Additional Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-10-01T00:00:00Z</td>
<td>20:15:19.068Z</td>
<td>DZ</td>
<td>I</td>
<td>0.814000</td>
<td>-33.7000, 21.9000, -60781.2, 51415.6, -37972.4</td>
</tr>
<tr>
<td>2000-10-01T00:01:00Z</td>
<td>20:16:05.077Z</td>
<td>MZ</td>
<td>I</td>
<td>0.814000</td>
<td>-33.8000, 21.9000, -60717.5, 51351.9, -37972.4</td>
</tr>
<tr>
<td>2000-10-01T04:03:11Z</td>
<td>20:03:11.067Z</td>
<td>CY</td>
<td>I</td>
<td>0.880000</td>
<td>-54.4000, 22.4000, -33767.4, 39310.3, -43642.7</td>
</tr>
<tr>
<td>2000-10-01T07:52:56Z</td>
<td>20:32:56.084Z</td>
<td>TL</td>
<td>I</td>
<td>0.947000</td>
<td>-83.0000, 2.60000, 1146.80, 16947.4, -37144.1</td>
</tr>
<tr>
<td>2000-10-01T09:07:05Z</td>
<td>20:07:05.067Z</td>
<td>CZ</td>
<td>I</td>
<td>0.970000</td>
<td>-58.8000, 8.80000, 12806.1, 6180.10, -28224.4</td>
</tr>
<tr>
<td>2000-10-01T09:35:00Z</td>
<td>20:35:00.081Z</td>
<td>QL</td>
<td>I</td>
<td>0.978000</td>
<td>-45.3000, 9.40000, 16565.1, 1720.20, -23254.9</td>
</tr>
<tr>
<td>2000-10-01T10:05:00Z</td>
<td>20:05:00.086Z</td>
<td>VL</td>
<td>I</td>
<td>0.986000</td>
<td>-27.6000, 9.80000, 19750.7, -3185.60, -16628.8</td>
</tr>
</tbody>
</table>
CAA Hardware/Software Status

Room at ESTEC has been refurbished (air con etc)

System hardware was delivered during August 2004

Basic configuration (operating system and most third party software, CVS, web server etc) complete

System is running Debian version of Linux

MySQL Database installed

High speed network connection via ESAGrid (1 Gbit/s)

Data ingestion, cataloguing and basic web access (developed by Sinead McCaffrey) capability in place (limited to data fetch)
CAA Infrastructure – System/Network

System

Network

<table>
<thead>
<tr>
<th>Data</th>
<th>Address</th>
<th>Average Speed KB/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDI</td>
<td>mpe.mpg.de</td>
<td>750</td>
</tr>
<tr>
<td>ASPOC</td>
<td>rhea.iwf.oeaw.ac.at</td>
<td>250</td>
</tr>
<tr>
<td>CIS</td>
<td>cis.cesr.fr</td>
<td>95</td>
</tr>
<tr>
<td>WBD</td>
<td>cassini.physics.uiowa.edu</td>
<td>120</td>
</tr>
<tr>
<td>EFW</td>
<td><a href="http://www.cluster.irfu.se">www.cluster.irfu.se</a></td>
<td>200</td>
</tr>
<tr>
<td>RAPID</td>
<td>sun2.mpaei.wdg.de</td>
<td>850</td>
</tr>
<tr>
<td>PEACE</td>
<td>msslx.mssl.ucl.ac.uk</td>
<td>200</td>
</tr>
<tr>
<td>WHISPER</td>
<td><a href="http://www.lpce.cnrs.fr">www.lpce.cnrs.fr</a></td>
<td>770</td>
</tr>
<tr>
<td>RAL</td>
<td><a href="http://www.cluster.rl.ac.uk">www.cluster.rl.ac.uk</a></td>
<td>650</td>
</tr>
<tr>
<td>ESOC</td>
<td>cddsa.esoc.ops.esa.int</td>
<td>90</td>
</tr>
</tbody>
</table>
Overview of instrument deliveries

<table>
<thead>
<tr>
<th>Task Name</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEF Delivery/Ingestion</td>
<td>Tue 02/01/01</td>
<td>Tue 31/12/02</td>
</tr>
<tr>
<td>ASPC</td>
<td>Tue 02/01/01</td>
<td>Sat 30/06/01</td>
</tr>
<tr>
<td>CIS</td>
<td>Tue 02/01/01</td>
<td>Sun 30/06/02</td>
</tr>
<tr>
<td>DWP</td>
<td>Thu 01/02/01</td>
<td>Thu 01/02/01</td>
</tr>
<tr>
<td>EDI</td>
<td>Tue 02/01/01</td>
<td>Tue 31/12/02</td>
</tr>
<tr>
<td>EFW</td>
<td>Fri 02/02/01</td>
<td>Mon 31/12/01</td>
</tr>
<tr>
<td>FGM</td>
<td>Sat 03/02/01</td>
<td>Tue 01/05/01</td>
</tr>
<tr>
<td>PEACE</td>
<td>Tue 02/01/01</td>
<td>Mon 31/12/01</td>
</tr>
<tr>
<td>RAPID</td>
<td>Tue 02/01/01</td>
<td>Sat 30/06/01</td>
</tr>
<tr>
<td>SAPPF</td>
<td>Thu 04/01/01</td>
<td>Mon 31/12/01</td>
</tr>
<tr>
<td>WHISPER</td>
<td>Fri 02/02/01</td>
<td>Fri 31/08/01</td>
</tr>
<tr>
<td>WBD</td>
<td>Sat 03/02/01</td>
<td>Mon 31/12/01</td>
</tr>
</tbody>
</table>

- So far for 2001:
  - About 75 datasets per spacecraft
  - More than 1 TB of CEF data in 200,000 files
  - Several more TB of raw, binary Level-1 and graphical products still to be ingested.
Cluster Science Data System Prime and Summary Parameter Data.

- The CSDS PP/SP will also be ingested into the CAA
- Data will be translated from CDF to CEF
- More importantly metadata will be translated to the CAA MDD
- Availability will track provision of high resolution products

- Original CDF versions of the PP/SP will continue to be available from the national data centres
- PP intervals ingested into the CAA are likely also to be available as public data from the national data centres
- Updates to PP/SP will also be added to CAA but there will be a lag compared to availability via CSDS.
For the Cluster extended mission raw data delivery on CD-ROM from ESOC will cease.

- CAA will provide RDM data accessible over the network.
- DDS access will continue unchanged.
- A web based system for data retrieval has been put in place and has been undergoing testing over the last month.
- Requests can be made for just the portions of the RDM directory structure that are required, thus saving network bandwidth.
- Support for automated transfers is provided via the standard wget utility.
- Details of the interface and examples are available on the CAA technical forum.
- The operational system is expected to be available for testing next week.
Tools

A number of tools already available to handle CEF data

• CAAtools (developed by Tobias Eriksson) provides a set of low level tools including verification of CEF syntax (CEFpass), generating XML metadata descriptions (CEF2XML) and for combining, splitting and extracting intervals from one or more files (CEFcombine)

• QSAS complete data visualisation and analysis package developed by Imperial College and Queen Mary University of London

• A set of generic IDL functions for reading CEF data are under production

• The CIS team are updating their CI visualisation package to support CEF data products

• A working group is being set up to coordinate these efforts
CAA – QSAS support for CEF-2

Cluster, Magnetic Field Magnitude, full resolution

C4

C2

C3

C1

1 minute

Cluster 5th anniversary symposium

19-23 September 2005
Cluster Active Archive

WELCOME TO THE CLUSTER ACTIVE ARCHIVE

The CAA is a database which will contain all the Cluster high resolution and other science data, together with auxiliary (orbit, attitude, command history etc.) data, survey data, CSDS data and plots selected JSOC data (predicted and identified events, geometric and magnetic positions).

This temporary site is currently under construction and will be released in May 2005. If you would like to register to view the pre-release version of the site (for testing and feedback purposes) please register using the link on the side menu to participate in the pre-release trials.
CAA – Web Access

Cluster Active Archive

Error - Login Required.

You are not logged in. Your session may have expired, or you have logged out. Please log in to search the Cluster Active Archive database.
Cluster Active Archive

Cluster 5th anniversary symposium

19-23 September 2005
CAA – Web Access

Select by time range, spacecraft, experiment, type or any combination

CLUSTER ACTIVE ARCHIVE SEARCH PAGE
Select search parameters below...

Start Time
(Year:Month:Day:Hour:Minute:Second)

Stop Time
(Year:Month:Day:Hour:Minute:Second)

Observatory
- Cluster-1
- Cluster-2
- Cluster-3
- Cluster-4

Experiment
- ASPOC
- CIB
- DWIP
- EDI
- EFW
- FOM
- PEACE
- RAPID
- STAFF
- VGD
- WHISPER

Measurement Type
- Activity_Index
- Electric_Field
- Energetic_Particles
- Instrument_Status
- Ion_Composition
- Magnetic_Field
- Neutral_Atom_Images
- Neutral_Loss
- Particle_Correlator
- Radio_and_Plasma_Waves
- Radio_Soundings
- Spacecraft_Status
- Thermal_Plasma

Instrument Type
- Antenna
- Channeltron
- Data_Processing_Unit
- Double_Sphere
- Electron_Drift
- Electrostatic_Analyser
- Faraday_Cup
- Fan Feedback
- HF_Radar
- Langmuir_Probe
- Long_Wire
- Magnetometer
- Mass_Spectrometer
- Micro-channel_Plate
- Monopole
- Quadripolar_Analyser
- Resonance_Sounder
- Search_Coll
- Spacecraft_Potential_Control
- Spectral_Power_Receiver
- Waveform_Receiver

Cluster 5th anniversary symposium

19-23 September 2005
Dataset selection

Datasets

Expand/Collapse

- **Cluster 1**
  - CIS-CODIF1
  - **CIS-HIM**
    - C1_CP_CIS-HA_H8_MAGIONS_CS (3D ion distributions for CIS-HA, High-Sensitivity, Magnetosheath Mode)
    - C1_CP_CIS-HA_H8_MAGIONS_PEF (3D ion distributions for CIS-HA, High-Sensitivity, Magnetosheath Mode)
    - C1_CP_CIS-HA_H8_MAGIONS_FF (3D ion distributions for CIS-HA, High-Sensitivity, Magnetosheath Mode)
    - C1_CP_CIS-HA_H8_MAGIONS_RC (3D ion distributions for CIS-HA, High-Sensitivity, Magnetosheath Mode)
  - **ED1**
  - **EF1**
    - C1_CP_EFW_L1_P1 (Probe 1 to spacecraft potential)
    - C1_CP_EFW_L1_P12 (Potential difference measured between probes 1 and 2)
    - C1_CP_EFW_L1_P2 (Probe 2 to spacecraft potential)
    - C1_CP_EFW_L1_P3 (Probe 3 to spacecraft potential)
    - C1_CP_EFW_L1_P34 (Potential difference measured between probes 3 and 4)
    - C1_CP_EFW_L1_P4 (Probe 4 to spacecraft potential)
    - C1_CP_EFW_L2_EF (Electric field (high-pass filtered))
    - C1_CP_EFW_L2_P (Spacecraft potential)
    - C1_CP_EFW_L3_P (Spacecraft potential [spin resolution])
  - **FOM**
    - C1_CP_FOM_5VPS (Magnetic field, 5 vectors/second resolution)

**CAA – Web Access**

http://caa.estec.esa.int/...
Select products to download

SEARCH RESULTS

File Set
- C1_CPEF0L1_P1 (No. files: 10, Total size: 17109.75 KB)

Download

CLUSTER ACTIVE ARCHIVE SEARCH PAGE
Select search parameters below...

You have attempted to download size of [32556584.96kB] The maximum allowed is [100000kB]

DOWNLOAD RESULTS
To download just click on the link below
Search Results ready for download

Files returned in Zip file, one file per selected dataset
CAA – Beta Testing

The CAA web interface will be available for registration and beta testing from Monday 26th September and is expected to last for several months.

http://caa.estec.esa.int/

General caveats during the beta testing activity:-

- The system will be undergoing ongoing development and enhancement that may result in it being offline at short notice.

- There are known issues with some data products. All data retrieved should be treated with caution and instrument teams should be consulted before these data are used for serious scientific investigations.

- An inter-calibration technical working group (led by Harri Laakso) is being set up to look at areas for improvement in data quality.

- Please provide feedback either via the technical forum or e-mail the CAA team at caateam@rssd.esa.int
Further information from http://caa.estec.esa.int

Technical Forum: http://www.cluster.rl.ac.uk/caa-bin/yabb/YaBB.pl
LiveLink: http://www.rssd.esa.int/llink/livelink (requires login)
Summary

• High resolution data from most of the Cluster instruments is starting to become available from the CAA.

• Ongoing data delivery and ingestion activities will extend the coverage and number of products available.

• Ongoing development of the user interface will concentrate on provision of access to graphical products and on demanding plotting.

• The CAA will open for beta testing next week – anyone wishing to participate in this activity is welcome to do so.

http://caa.estec.esa.int/

For further information contact: caateam@rssd.esa.int