

# MiMeS Discussion: Archival and Supplementary Data

David Bohlender NRC  
(wearing yet another hat...)

Science (.25%)

DAO (20%)

Admin (20%)

**CADC** (60%; GSA, FUSE, MOST, CVO, POOPSY, etc...)

# The Canadian Astronomy Data Centre (CADDC)

- Established in 1986 as one of three archiving centres for the HST
- CADDC staff designed, developed and operate the Gemini Science Archive (GSA), JCMT Science Archive (JSA), CFHT and many other facilities



<http://www.cadc.hia.nrc.gc.ca>

# GSA Example



The Canadian Astronomy Data Centre  
Herzberg Institute of Astrophysics

[CADCC Homepage](#)  
[Register](#)



## Gemini Science Archive

[Home](#)
[GSA Queries](#)
[GSA Help](#)
[Related Links](#)
[Credits](#)

New Resolver      Object      Right Ascension      Declination  
[Simbad\(simbad.u-strasbg.fr\)](#)      sigma Ori E      05 38 47.1      -02 35 40.5

[Download Datasets with Cal.](#)
[Download Datasets](#)
[Save Datasets](#)
[MarkAll](#)
[UnMarkAll](#)

Mark	Target Name	RA (J2000)	DEC (J2000)	Data Superset Name	Original File Name	Science Program	Observing Logs	Exposure Time	UT Date	Release Date	All Program Obs.	Processed Calibration Files	Nightly Raw Calibration Files	Instrument	Filter(s)
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-28-001</a>	S20040105S0015	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 12:59AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si3-9.7um</a>
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-31-001</a>	S20040105S0016	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 1:21AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si4-10.4um</a>
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-31-002</a>	S20040105S0017	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 1:44AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si4-10.4um</a>
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-31-003</a>	S20040105S0018	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 2:06AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si4-10.4um</a>
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-31-004</a>	S20040105S0025	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 2:44AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si4-10.4um</a>
<input type="checkbox"/>	S Ori 71	05 39 00.19	-02 37 05.9	<a href="#">GS-2003B-Q-33-31-005</a>	S20040105S0026	<a href="#">GS-2003B-Q-33</a>	<a href="#">Obs. Logs</a>	369.2	<a href="#">Jan 5 2004 3:07AM</a>	<a href="#">Jul 5 2005 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">TReCS</a>	<a href="#">Si4-10.4um</a>
N/A	Sigma Orionis IRS A	05 38 44.84	-02 35 57.1	<a href="#">GN-2008A-Q-111-1-001</a>	N20080327S0019	<a href="#">GN-2008A-Q-111</a>	Proprietary	30.0	<a href="#">Mar 27 2008 5:41AM</a>	<a href="#">Sep 27 2009 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">NIFS</a>	<a href="#">HK_G0600</a>
N/A	Sigma Orionis IRS A	05 38 44.84	-02 35 57.1	<a href="#">GN-2008A-Q-111-1-002</a>	N20080327S0020	<a href="#">GN-2008A-Q-111</a>	Proprietary	30.0	<a href="#">Mar 27 2008 5:42AM</a>	<a href="#">Sep 27 2009 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">NIFS</a>	<a href="#">HK_G0600</a>
N/A	Sigma Orionis IRS A	05 38 44.84	-02 35 57.1	<a href="#">GN-2008A-Q-111-1-003</a>	N20080327S0021	<a href="#">GN-2008A-Q-111</a>	Proprietary	30.0	<a href="#">Mar 27 2008 5:43AM</a>	<a href="#">Sep 27 2009 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">NIFS</a>	<a href="#">HK_G0600</a>
N/A	Sigma Orionis IRS A	05 38 44.84	-02 35 57.1	<a href="#">GN-2008A-Q-111-1-004</a>	N20080327S0022	<a href="#">GN-2008A-Q-111</a>	Proprietary	30.0	<a href="#">Mar 27 2008 5:44AM</a>	<a href="#">Sep 27 2009 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">NIFS</a>	<a href="#">HK_G0600</a>
N/A	Sigma Orionis IRS A	05 38 44.84	-02 35 57.1	<a href="#">GN-2008A-Q-111-1-005</a>	N20080327S0023	<a href="#">GN-2008A-Q-111</a>	Proprietary	30.0	<a href="#">Mar 27 2008 5:45AM</a>	<a href="#">Sep 27 2009 12:00AM</a>	<a href="#">List</a>		<a href="#">Cal. Files</a>	<a href="#">NIFS</a>	<a href="#">HK_G0600</a>

# CADC Infrastructure

- 12 Gen 5 processing nodes (and ~24 Gen 1 - 4)
  - two dual-core 3GHz processors; 16GB RAM
  - Gen 6 quad dual core processors with 64GB RAM on order
- 500TB of on-line storage on Gen 3 - 6 storage nodes
  - Gen 6 nodes have 60TB (60 x 1TB) each
  - Off site backup at UVic via 1Gb connection
- Sybase (+ hot standby) (16GB RAM/600GB storage) and 8 x DB2 database servers (4GB/100GB on each)
- Condor and “Grid-like” distributed computing available
- Programmatic access to *all stored files* (public or proprietary - with authorization): <http://www.cadc.hia.nrc.gc.ca/getData?...>  
i.e. you can use the CADC like a 500TB hard drive!

# The CFHT Archive

- Data from all CFHT instruments archived since about 1989
- Current interface very much geared towards imaging science
- Most recent development was for CFHTLS
- Currently only raw ESPaDOnS data archived, and collection not complete - mainly because of FITS header issues and lack of demand for raw data!

# ESPaDOnS Data Archiving

A number of issues for the CADC limit possible efforts:

- ★ Little motivation to ensure that the “raw” ESPaDOnS collection is complete
- ★ No access to processed ESPaDOnS data (which must be in FITS format for CADC ingestion)
- ★ No access to Libre-ESPRIT
- ★ MagIcS LEGACY database???

So little “official” CFHT archive development for MiMeS (or MaPP) is likely unless this changes. But...

# CADC Community Projects



- Stetson photometric standards
- Astronomy Meetings List
- ACS Virgo Cluster Survey
- SCUBA Legacy Catalogues
- **MiMeS metadata, products, catalogues....?**

# So, what would you like?

First, think big! **IF** all raw **AND** processed ESPaDOnS spectra were stored at the CADC:

- the processing Master (while relaxing in a restaurant with his wine) could execute a script that retrieves **all** ESPaDOnS I+V spectra, runs a regression analysis on every spectrum, plots B-field curves, stores results...

IF Libre-ESPRIT resided on CADC processing nodes:

- reprocess **all** raw spectra when a new version of the software is released

Thinking small, do some of this for MiMeS data at least?

# So, what would you like?

Thinking smaller:

➡ Use CADC database for catalogues,  
observing logs, etc...

➡ ...

➡ ...