



Archives System Building Infrastructure: Re-engineering ESA's space based missions' archives

Pedro Osuna

Science Archives and Computer Support Engineering Unit (SRE-OE) Science Operations Department (SRE-O) Science and Robotic Exploration Directorate (SRE)





Introduction

3. Lessons learnt

- · Examples of working systems and the lessons learnt that can be derived from them
- · Lessons learnt from archive migration to new technologies
- Implications of new technologies for engineering processes, data storage, operations costs and system performance
- · Advantages and difficulties of building interoperables services
- · Common or re-usable systems for archives building





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

The EXOSAT Science Archive

Conclusion

Pedro Osuna





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

The EXOSAT Science Archive

Conclusion

Pedro Osuna





European Space Astronomy Centre (ESAC)

ESAC default location for ESA's:

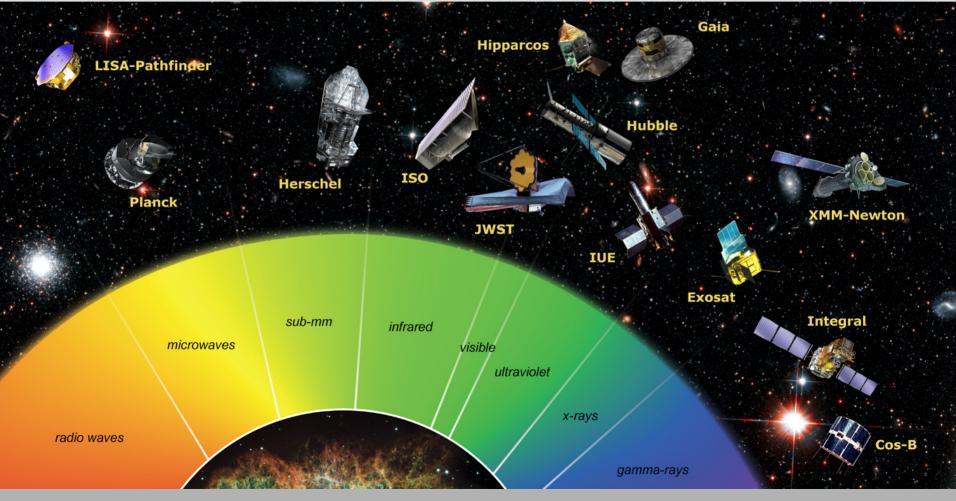
- Science operations:
 - long history with astronomical missions
 - expanding with solar system missions
- Science archives:
 - Astronomy
 - Planetary
- ESA VO activities:
 - ESAC is the European VO node for space-based astronomy







Space Based missions on Astronomy and Fundamental Physics



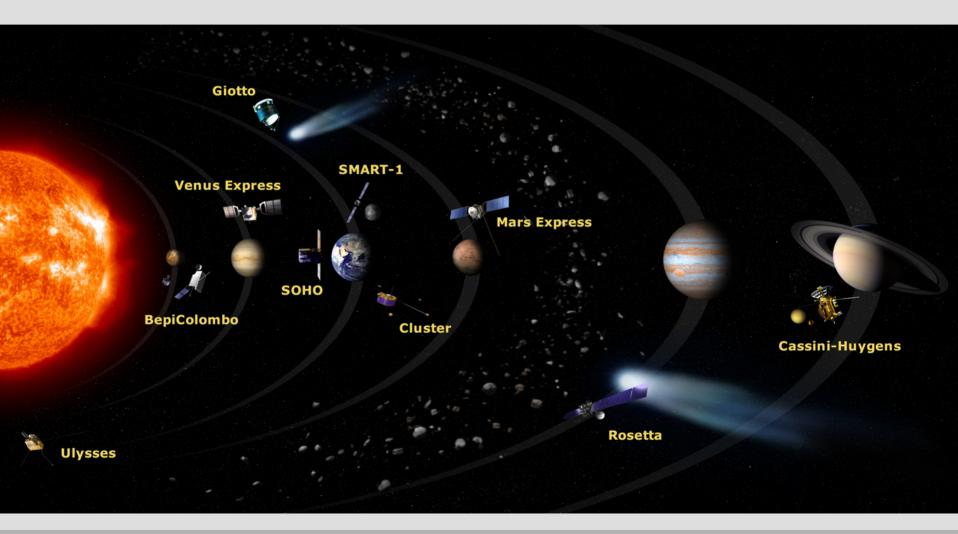
PV 2009 ESAC Madrid

Pedro Osuna





Solar System Space based missions



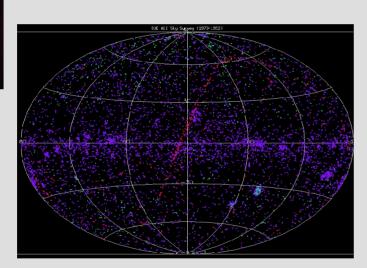
Pedro Osuna



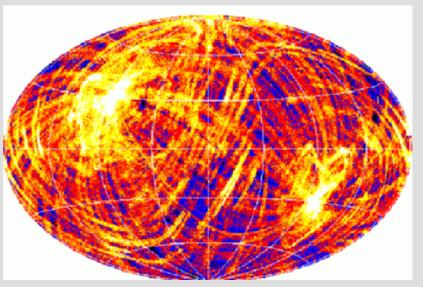


Before 1996: archives not at ESA

The International Ultraviolet Explorer (IUE) (1978-1996) The European Space Agency X-RAY Observatory (EXOSAT)











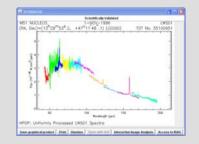
Starting a new era in archiving: The Infrared Space Observatory (ISO) Archive

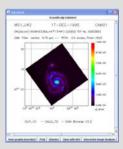


Print Tank First Field	Documentation Publications				
Query Specificate	Later Beauty	Sharenne Barthad	operflopping	time In	Respond Manifest
Logged In E					
	Dister APE		SERIVATOR	v Casa	
	STA DUTIONCE			Cood	
No. 1	industry Tanker Mees All	Laireit Revults	te Salariad Rafeeth L	Notional	ALADH
Colorest 1			21 11 7 101		un accenter
	Operative 5. Depet: 14 and		Rath Dre		anan an
Antern Anceres				-	
Belais Belaict	REL_PILL		47413138.9"		
Rattan Datest Latest	Secaral observation with o	Bertern 2 Salary	Scientifically Valid		
Street Other Reduced	BLINDLY CAR CAL FILLER		1356 and Burth		
and the second is	Carline	1 Past-11-10 241241411.	THE REAL PROPERTY.	MINI .	
Balata Balated	moc tite	PETTIC LINCOMIL, PLC -	-47411'50.2"		
Advist	Descenated and		Art. Briessifically Wal	Autod Inte	
Rational Delay8.2 sharet	101.0he	Bertert 2 Salary	T. c. v. 16	w827*	
atteis (Mai Podatt)	STADUES ATTAVES DOTTO	AL 1996-11-01 15:30:51.*	4286 sec	faget	
Estats Refere	Eli Batiens	1000113629x55.214 4			
Rolean Deput Dataset	Sedium corclution waveleng		Scimitizally Valid		
Liven Ottar Autorea	40.05-50,40 49.54-60.94 17		£. 1. 7. 94		
ferrore little Poduces	STADUST MINITE GALVIS	8 1996-11-01 10:01115.*	2514 sec	terest.	
Rates Kanped	BIC 1114 B 21	INTELININAL SEA	anarran an		
	Spectarghotaetay	HIGH LINESALLING	Bulentifically Valid		
Advist NP2P	2. C-L tha	Deptern 2 Galary	f.s.e. 24		
Aphane Detect Saland					
Roman Dariest Salant Down Office decisions					
Aphane Detect Saland	STATUSTS BARREL BROOM	at the transmission	3122 ANT Just	and the second	
Antoneo Deronit I olanet George Office Antoneo Remois Titler Products		Callo4 12620612.344		and the second se	
Anteres Derault Solant Deray (Disc Andreas Anterio Disc Andreas Anterio Anterio Anterio artist	1140013 HIEREL BROOM	CM0413529652.144	4741113.7	and the second	
Antonio Derault Dalant Deray Office Antonio Remote Dirac Products Antonio Antonio Orfice Factoria Office Factoria Darjust Dalant	1140015) BARKEL BRECH B 14 CVF CVF spectral abservation of	CARD413529652.34/ 4	HT411'55.7" Briestifically Talls		
Antores Descrit Default Descrit (Mar Rockson Remote Mar Rockson Antores Antores Satures Descrit (Mar Rockson Descrit (Mar Rockson Descrit (Mar Rockson	1140013 HIEREL BROOM	Campa 17629412.144 4 F176 campa - Sepfect 2 Galary	47411'13.7" Brimifically Talls f.s.v. 13	- 13	
Astrono Dereval Default Bores (Max Rotans Remoi Dire Potes) • • Potes dere Astrono Dereval Default Bores (Dire Rotans Bores (Dire Rotans Bores (Dire Rotans	1640653 BERREL BECOM 8.16 CVF CVF spectral observation of 5.15-15.49m 15200100 UNDROD CARDED	CANDA 13529652.344 4 0155 remets Septert 2 Unimry 8 1997-05-01 00:30:32.*	HTHLI'IL.T' Trimstiticily Tall f.s.v. 15 SIAE and	- 13	
Astrono Dereval Default Bores (Max Rotans Remoi Dire Potes) • • Potes dere Astrono Dereval Default Bores (Dire Rotans Bores (Dire Rotans Bores (Dire Rotans	1540045) BARKEL BYENG T 15 CYF CYF spectral abservation of 5.51-15.49m 15201261 LYTOPODY CARDYN Trad Belociae	CaRD4 15529612.144 4 F155 comes Septent 2 Galary 2 1997-05-51 05130132.* 19932 05554612.522 4	ATALL'EL.T" Repart firming Talla f.s.v. 13 SIAD and Joint COMD-DA.P"	- 1 %.	- 1
Antinen Freinel Felant Berry (Mar Bohna Ramon Dire Poten Adviss Adviss Gerry (Mer Rohna Gurry (Mer Rohna Ramon Dire Rohna Ramon Status)	1640653 BERREL BECOM 8.16 CVF CVF spectral observation of 5.15-15.49m 15200100 UNDROD CARDED	Camba 1 3529452.546 - (156 - cancia Septent 2 Galaxy 2 1997-05-01 01:39132.52 - LUMI2 010.548.52.525 - sctome	HTHLI'IL.T' Trimstiticily Tall f.s.v. 15 SIAE and		

- Available since December 1998:
 - http://iso.esac.esa.int/ida/
- Developed and released for the post mission phase
- Active development up to 2006, low maintenance now
- Content stable since 2002
- Around 400 GB of data (FITS) on hard disk of all levels of data (standard processing + high level data products)







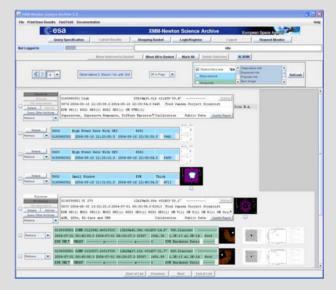
g

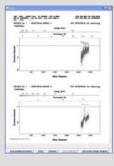


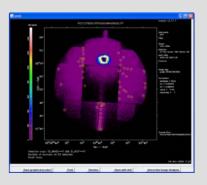


XMM-Newton Science Archive

- Available since April 2002:
 - http://xmm.esac.esa.int/xsa/
- Developed and released for the operational phase
- Still active development
- New data coming on daily basis
- 2 TB of data (FITS) on hard disk of raw and processed data, some catalogues
- Data processing done at Leicester
- On the fly reprocessing system available from the Archive (run at ESAC)











Archives at ESAC since ISO

- ESAC is the Centre where most of ESA Scientific Archives are developed, maintained and operated.
- ESAC Science Archives Team is giving support to various projects <u>http://www.rssd.esa.int/index.php?project=SAT</u>





ISO Data Archive Since December 1998

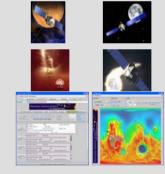


XMM-Newton Science Archive Since April 2002

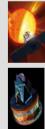




Integral SOC Herschel Science Data Archive Science Archive Since July 2005 Since 2009



Planetary Science Archive Giotto, Mars Express Rosetta, Venus Express Smart-1, Huygens Since March 2004





Soho, Exosat, Planck, GAIA, ... in the future





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

The EXOSAT Science Archive

Conclusion

Pedro Osuna





Investigation of possible improvements

- ISO Data Archive (IDA) was pioneer in using three tier architecture and Java technology, state-of-the-art technology at the time (~1995)
- Initially thought to serve only ISO data, the IDA archive technology was later applied to the XMM-Newton Science Archive (XSA)
- Other archives followed suite (see before...)
- This architecture has allowed not only the building of ESAC archives but also their integration within the VO, but...
- A lot of the technology used at the time is obsolete
- An example: some of the nowadays available-everywhere scroll-bars were implemented "by hand" in certain window interfaces





Setting the ground: listing possible issues

- Made an exercise of "self-auditory"....:
 - Communication Client-Server done through home made RPC (Java serialised objects)
 - Transport done through TCP-IP in compressed mode
 - Business layer mixing service, transport and logic in a single implementation
 - Business layer makes uses of port 80 and blocks its usage to any other application running on the same machine
 - load balancing, security and proxy redirection are not available in our "homemade" server

-

-





Investigation on frameworks and available technology

- Main points:
 - should be as open as possible, with a community big enough to ensure estability and permanence
 - should be as light as possible (both client and server)
 - should be modular and flexible
- Client layer:
 - should be light
- Server layer
 - should be robust and flexible
- DB and persistence layer
 - focus on Open source DB
 - find a proper persistence layer (!!)





The options and decisions

- Client layer:
 - Eclipse RPC versus InfoNode/JGoodies on Swing. Decided for lighter InfoNode/JGoodies/Swing
- Server Layer:
 - Application framework: Spring (used in all layers)
 - Server container: opted for the light way with Tomcat rather than heavier JBoss, Galshfish, Jonas....
- Persistence layer
 - Hibernate vs Ibatis. Both very good pros and cons. Decided finally for Hibernate, with better integration with overall project dependent Data Models





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

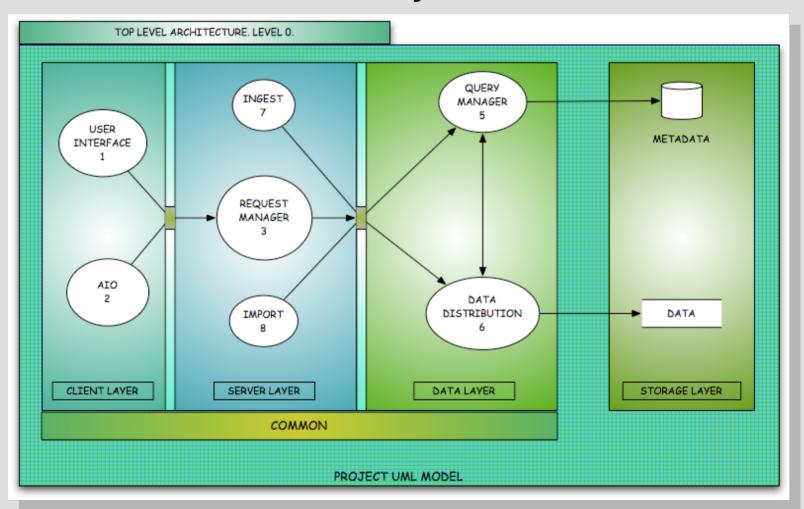
The EXOSAT Science Archive

Conclusion





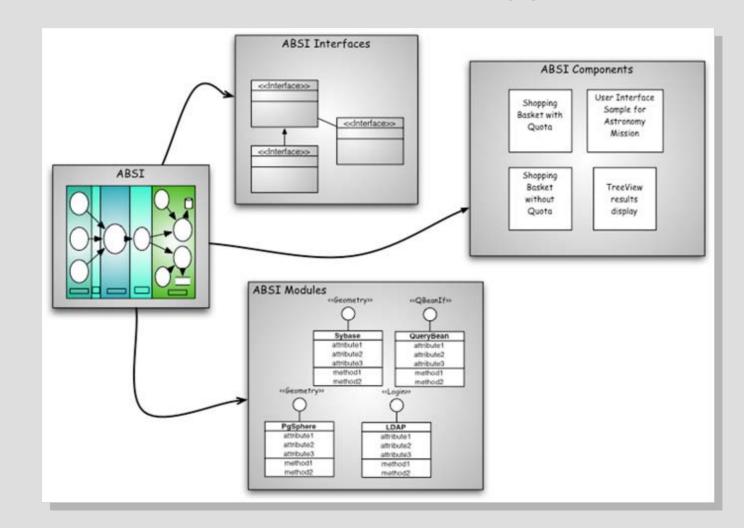
Standard three-layer architecture







ABSI Elements (II)







The problem of handling big amounts of metadata

- We are dealing with more than a million observations (granularity in SOHO different from Astronomical cases). Database Table indexing gets overly complicated and joins poorly performant
- □ To know how to apply the joins to the different attributes requested ("where" part of the query), we implement the **Dijkstra** algorithm (shortest path algorithm, graph theory).
- Dijkstra's algorithm, conceived by Dutch computer scientist Edsger Dijkstra in 1959, is a graph search algorithm that solves the single-source shortest path problem for a graph with non negative edge path costs, outputting a shortest path tree.
- □ This algorithm is often used in routing.
- We have applied it to our database tables and relationships
- On-line examples:

http://www.carto.net/papers/svg/dijkstra_shortest_path_demo/





with indiana in RA,

cuted

oen

Indexing spherical data in DB

- Indexing opherical data for search in a DD is traditional problem
 3.4. Circle
- Coord DFC.



- PgSp sourc
- Provi
 - inpl
 - con
 - vari
 - circ
 - sph
 - indexing or spherical data types

A spherical circle is an area around a 1 the radius of the circle. Usage cases at

· sites on earth having a maximum

• round cluster or nebula on sky spl

· a position with an undirected posi

A circle is specified using a spherical

Valid radius units are RAD, DEG, and I

Example 7. A circle around the Nor

sql> SELECT scircle '< (0d

< point , radius >

than zero.

- · several input and output formats
- Implemented in EXSA for the

3.9. Coordinates range



A spherical box is a coordinates range. Hence, you can select objects within a longitude range and latitude range. The box is represented using two spherical points: the southwest (pos_sw) and the northeast $(pos_n\Theta)$ edge of the box. The input syntax is:

```
( pos_sw, pos_ne )
```

```
or
```

```
pos_sw, pos_ne
```

Note:

- If the latitude of the southwest edge is larger than the latitude of the northeast edge, pgSphere swaps the edges.
- If the longitude of the southwest edge is equal to the longitude of the northeast edge, pgSphere assumes a full latitude range, except that the latitudes are equal, too.

Example 12. Input of a full latitude range

A full latitude range between $+20^{\circ}$ and $+23^{\circ}$.

sql> SELECT sbox '((0d,20d), (0d,23d))';

Example 13. A simple coordinates range

A coordinate range between -10° and $+10^{\circ}$ in latitude and 350° and 10° in longitude.

sql> SELECT sbox '((350d,-10d), (10d,+10d))';

PV 2009 ESAC Madrid

Pedro Osuna





New Archive creation Process

- Bottom-up approach: First build the general UML for the overall project. Then start building from there.
- UML → DB Design → Repository design → DAO (Data Access Objects) design → User Interface design
- Good UML design for project extremely important.
- Proper knowledge of the data by the SAT is crucial in order to build good Data repository and Data Distribution systems hundreds of mails interchanged between SOHO Archive Scientist and SAT Team on Data issues)





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

The EXOSAT Science Archive

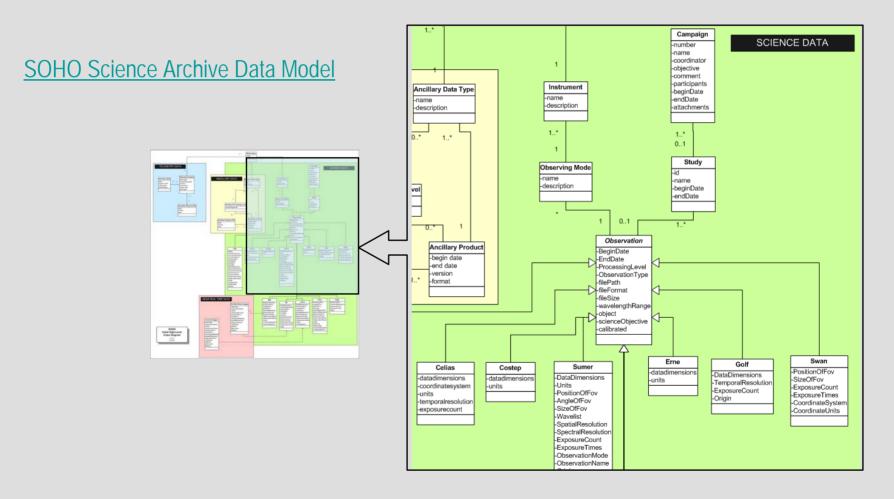
Conclusion

Pedro Osuna





SOHO Science Archive UML







Some numbers from SOHO

Data Type	Instrument	Data Format	Number of files
REALTIME	EIT	JPG	918402
REALTIME	LASCO	JPG	765356
REALTIME	MDI	JPG	23022
SCIENCE	CDS	FITS	295519
SCIENCE	CELIAS	CDF	80632
SCIENCE	COSTEP	ASCII	30921
SCIENCE	EIT	FITS	472144
SCIENCE	ERNE	ASCII	20596
SCIENCE	GOLF	FITS	4450
SCIENCE	LASCO	FITS	658419
SCIENCE	MDI	FITS	83494
SCIENCE	SUMER	FITS	118709
SCIENCE	SWAN	FITS	9122
SCIENCE	UVCS	FITS	87226
SCIENCE	VIRGO	FITS	23765





The SOHO Science Archive

http://soho.esac.esa.int/data/archive/index_ssa.html

General Usage (video)

Interoperability through IVOA protocols (video)

Time animator (video)

Pedro Osuna





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

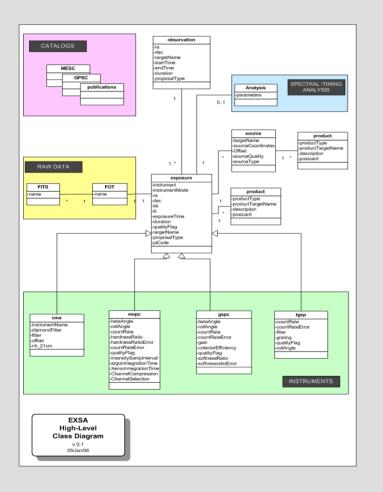
The EXOSAT Science Archive

Conclusion





The EXOSAT case



Different problem:

- not many instruments
- not a live project
- but facing standard issue in astronomy:
 - slow queries in spherical coordinates





The EXOSAT Science Archive

http://www.rssd.esa.int/index.php?project=EXOSAT&page=archive

General usage (video)





Adapting existing archives to new technology

	2 • • * # • * •
a \	
Results Display-	Sort Ortena 😟 🕢 Machinera 🐨
tain Query Panel	
Mars Express Earth Cotto	Clate Range Begin End End Myyy-MM-dd Hitmmiss Myyy-MM-dd Hitmmiss
Mission Rosetta Venus Express	Spacecraft Clock Settings Reset Start Reset Start Reset Stop Stop Count
Target Name	Inst Host
Dataset id	Reference Institution
AARS EXPRESS Mission	Mission Phase
C Target F	le © Coordinates
ASPERA HRSC MARS MARSIS ON	AECA \ SPICAM \ SPICE \
Product Id	St Data M
	Com 8 Cm
onsule + Ø x	

Re-engineered Planetary Science Archive prototype (**PSA**)

Products for Dataset 817 [98 Results] Page Product Id Details Image: Construction of the second of	irch	Data	sets #1	Products for Datas	at 817 * _ ×		
Product Id Details Product Id Details ATNM_P030602191 Paradic Details (ATNM_P030602191822_001) Mex Product Details (ATNM_P030602191822_001) Details Mex Product CK Max Resolution: Mex Product CK Id: ATNM_P030602191822_0013 Mex Product Ck Id: ATNM_P030602191822_0013 <t< th=""><th>Ē</th><th>Produc</th><th>ts for</th><th>Dataset 817</th><th></th><th></th><th></th></t<>	Ē	Produc	ts for	Dataset 817			
Product Id Details Product Id Details ATNM_P030602191 Paradic Details (ATNM_P030602191822_001) Mex Product Details (ATNM_P030602191822_001) Details Mex Product CK Max Resolution: Mex Product CK Id: ATNM_P030602191822_0013 Mex Product Ck Id: ATNM_P030602191822_0013 <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th></t<>							
Product Id Details (+ x m ATNM_P030602191 ATNM_P030602191822_001 ohrs ATNM_P05040100t Detector Id: Event Type: Kernel Type Id: SPICE ohrs ATNM_P06040100t Kernel Type Id: SPICE ohrs ATNM_RECONSTITU Scan Mode Id: St Data Product CK ohrs DE403_MASSES.TPC Id: ATNM_P030602191822_0013 ohrs DE403_MASSES.TPC Id: ATNM_P030602191822_0013 ohrs Scan Mode Id: St Data Product CK ohrs Spcraft Clock Start: 0 spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: ohrs Spcraft Reset Start: Spcraft Reset Start: ohrs Spcraft Reset Start: Start Time: 2003-06-03 00:55:27 ohrs Start Time: 2003-06-03 00:55:27 subPhase: EV ohrs SubPhase: EV Target Type: ohrs ATNHEXEDITRF93.TF N/A 1970-01-01 00:00:00					entre la constante de la const	ige Size:	25 🕶
Image: Construct Details (*** ** Image: Construct Details (ATNM_P030602191822_001) Details (*** Image: Construct Details (ATNM_P030602191822_001) Details (**** Image: Construct Details (ATNM_P030602191822_001) Details (**** Image: Construct Details (ATNM_P030602191822_001) Details (***** Image: Construct Details (ATNM_P030602191822_001) Details (************************************		THE		Product Id		hn	5
Mex Product Details (ATNM_P030602191822_001) Mex Product Details (ATNM_P030602191822_001) Detector Id: Detector Id: EarthfixeDiscort Mex Product Details (ATNM_P030602191822_001) Detector Id: EarthfixeDiau.TF Mex Product Details (ATNM_P030602191822_001) Details (ATNM_P030602191822_001) Details (ATNM_P030602191822_001) Mex Product Details (ATNM_P030602191822_001) Mex Product CK Belass: 1 Release: 1 Release: 1 Spcraft Clock Start: Spcraft Reset Start: SubPhase: EV Target Type: Target Type: EarthfixeDitrF93.TF N/A 1970-01-01 00:00:00 Othrs		0.2			Details 🖗 🕯 🗙	×	
Image: Second		Ŵ	2	ATNM_P03060219	Mex Product Details [ATNM_P030602191822_001	DOhrs	
Image: Second		11		ATNM P05040100		DOhrs	
ATINM_P060401000 ATINM_P060401000 Proc Level: Scan Mode Id: Scan Mode Id: Scan Product: CK Id: ATINM_P030602191822_0013 Orbit Number: Revision: 0 Spcraft Clock Start: Spcraft Clock Stop: Spcraft Reset Start: Spcraft Reset Start: SubPhase: EV Target Type: EARTHFIXEDITRF93.TF N/A 1970-01-01 00:00:00 .00hrs					Event Type:		
Proc Level: Scan Mode Id: St Data Product: CK Id: ATNM_RECONSTITU De403_MASSES.TPC De403_MASSES.TPC Release: 1 Release: 1 Revision: 0 Spcraft Clock Start: Spcraft Clock Start: Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Stop: Start Time: 2003-06-03 00:55:27 Start Time: 2005-07-14 01:51:37 SubPhase: EV Target Type: Start Time: N/A Target Type: Start Time: 2003-06-03 00:00 Other		W.	2	ATNM_P06040100		DOhrs	-
Scan Mode Id: 30hrs De403_MASSES.TPC St Data Product: CK Barbonic Id: ATTMM_P030602191822_0013 De403_MASSES.TPC Id: ATTMM_P030602191822_0013 De403_MASSES.TPC Release: 1 De403_MASSES.TPC Release: 1 De405.BSP Spcraft Clock Stop: Spcraft Clock Stop: Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: Start Time: 2003-06-03 00:55:27 Start Time: 2003-06-03 00:55:27 Target Name: N/A Target Type: Target Type: Deaction Target Type: Deaction N/A 1970-01-01 00:00:00 .00hrs							
St Data Product: CK Id: ATNM_P030602191822_0013 Orbit Number: Release: 1 Revision: 0 Spcraft Clock Start: Spcraft Clock Stop: Spcraft Reset Start: Spcraft Reset Start: SubPhase: EV Target Type: Revision: 0 Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: SubPhase: EV Target Type: Revision: 0 Spcraft Reset Start: Spcraft Reset Start: Spcraft Reset Start: SubPhase: EV Revision: 0 Spcraft Reset Start: Spcraft		頭	2	ATNM_RECONSTITU		DOhrs	-
Image: Construction of the second		1995					
Image: Second		1	2	DE403_MASSES.TP			- 🐢 🗧
Revision: 0 Spcraft Clock Start: Spcraft Clock Stop:							
Image: Second Clock Start: Spcraft Clock Start: Spcraft Clock Stop: Spcraft Clock Stop: Spcraft Reset Start: Spcraft Reset Start: SubPhase: EV Target Type: Target Type: Target Type: EARTHFIXEDITRF93.TF N/A 1970-01-01 00:00:00 .00hrs		34	2	DE403_MASSES.TP	Release: 1	9.0	
Spcraft Clock Stop: Spcraft Reset Start: Spcraft Reset Stop: Start Time: 2003-06-03 00:55:27 Start Time: 2003-06-03 00:55:27 Start Time: 2003-06-03 00:55:27 Start Time: 2005-07-14 01:51:37 SubPhase: EV Target Name: N/A Target Type: EARTHFIXEDITRF93.TF N/A 1970-01-01 00:00:00 .00 ♥					Revision: 0		
Image: Constraint of the section of		19	2.	DE405.BSP	Spcraft Clock Start:	.00	
Spcraft Reset Stop: Start Time: 2003-06-03 00:55:27 Start Time: 2005-07-14 01:51:37 SubPhase: EV Target Name: N/A Target Type: ARTHFIXEDITRF93.TF N/A 1970-01-01 00:00:00 .00hrs							
Image: Start Time: 2003-06-03 00:55:27			۵.	DE405.BSP		.00	*
Image: Start Higher Start Stop Time: 2005-07-14 01:51:37 Image: Start Stop Time: 2005-07-14 01:51:37							
Image: State of the state o		頭	2 4	EARTHFIXEDIAU.TF			1
Construction of the second secon							
Image: Type: Image: Type: Image: Type: Image: Type: Image: Type: Image: Type: Image: Type: Image: Type: Image: Type: Type: Image: Type: Type: Image: Type: T			2	EARTHFIXEDIAU.TF			
		-	-				
		15	- <u>-</u>	EARTHFIXEDITRF93			
🛒 🖭 🍌 EARTHSTNS_FX_050714.BSP N/A 1949-12-31 23:59:18 876599.00 🔮		Ħ	٤.	EARTHFIXEDITRF93	TF N/A 1970-01-01 00:00:00 .00H	nrs	٠
		Ħ	2	EARTHSTNS_FX_05	N/A 1949-12-31 23:59:18 876	599.00	





ESAC Archives History

ESAC Archives evolution

The "ABSI" concept

The SOHO Science Archive case

The EXOSAT Science Archive

Conclusion





Conclusions

- State of the art technology ever changing, need to adapt to new trend
- Many options available: time is needed to investigate different options
- Flexibility is paramount when making decisions: will have to re-assess technologies in some time
- ABSI concept proven useful: modularity is key issue
- First ABSI produced archives: SOHO and EXOSAT
- SOHO released to public community on 28 Sep 2009
- EXOSAT released to public community on 1 Dec 2009
- Will adapt existing ESA archives to new technology
- Archives can be accessed at:
 - <u>http://soho.esac.esa.int/data/archive/index_ssa.html</u>
 - <u>http://www.rssd.esa.int/index.php?project=EXOSAT&page=archive</u>
- <u>Science Archives Team pages</u>