Science Archives at the ESAC Science Data Centre

Christophe Arviset Head of ESDC, ESAC

SCIOPS 2015 Workshop, ESO 24/11/2015

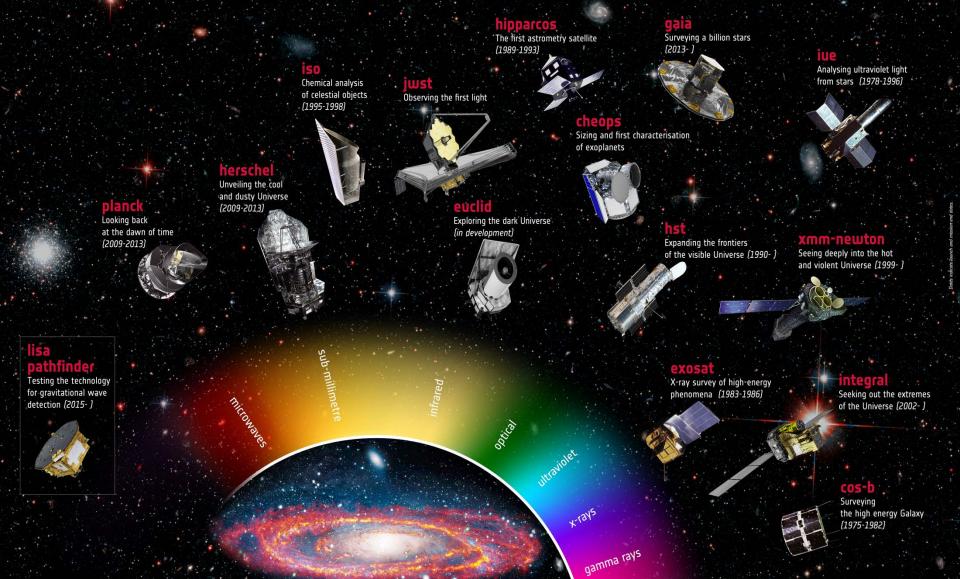
Issue/Revision: 1.0 Reference: ESAC Science Data Centre Status: Issued ESA UNCLASSIFIED - Releasable to the Public



Sa

→ ESA'S LEGACY OF SPACE OBSERVATORIES

ESA has a long history of space science missions that have studied the skies across the full spectrum of light, charted the Milky Way Galaxy and probed the fabric of the Universe. Even after completing observations, the legacy of these missions lives on in the form of vast databanks, and paves the way for the next generation of space observatories.







uice Studying Jupiter's icy moons (in development)



huygens Landing on Titan (15 January 2005)



solar orbiter The Sun up close (in development)

smart-I Exploring our Moon (2003-2006)

mars express Investigating the Red Planet (2003-)



ulysses Watching over the Sun's poles (1990-2009)

Measuring Earth's magnetic shield (2000-)



→ ESA'S LEGACY IN THE SOLAR SYSTEM

ESA has a long history of space science missions that have explored the Solar System, from the Sun, planets and moons to asteroids and comets, transforming our view of our planetary neighbourhood. Even after completing observations, the legacy of these missions lives on in the form of vast databanks, and paves the way for the next generation of spacecraft.

ESAC Science Data Centre *The Digital Library of the Universe*



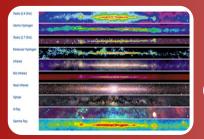
- Large set of science archives co-located at ESAC are a major research asset for community :
 - Astronomy, Planetary, Solar Heliospheric
- > Different types of data:
 - Raw data, calibrated processed data, high level data products, ...
 - All data public and available on-line after a short proprietary period
- Need to be kept readily available for future users and novel uses by various types of users:
 - Scientific Community (public access)
 - PI team and observers (controlled access)
 - Science Operations Team (privileged access)

Archive Strategy Plan for 5-20+ years

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 4 ESA UNCLASSIFIED - Releasable to the Public



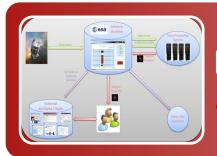
ESAC Science Archives Strategy



Enable maximum *scientific exploitation* of data sets



Enable efficient *long-term preservation* of data, software and knowledge, using modern technology

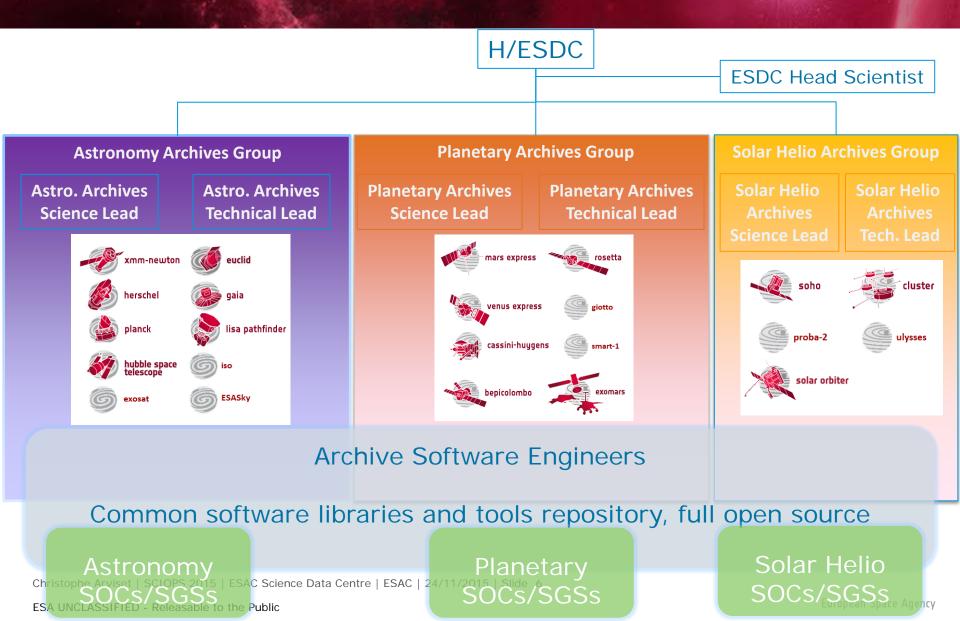


Enable cost-effective archive production by integration in, and across, projects

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 5

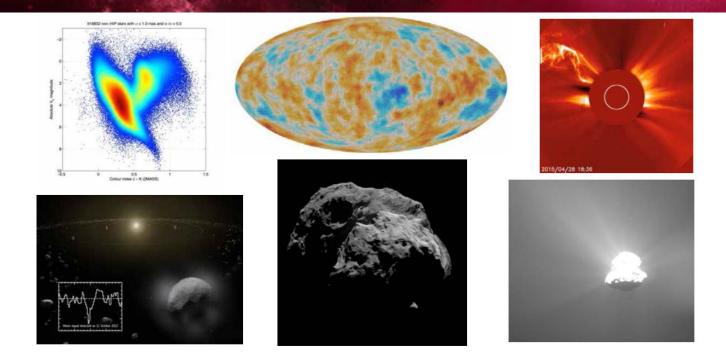
ESDC organizational structure





Enable Maximum Science Exploitation



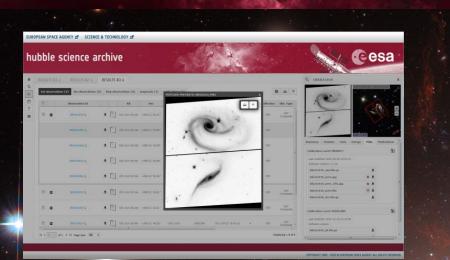


Scientists and Engineers working very closely together

Science driven Archives

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 7

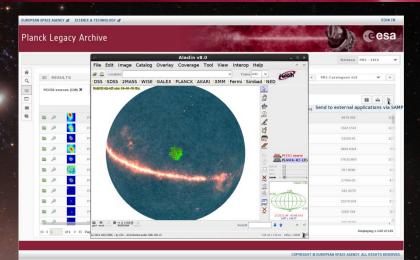
Individual Mission Science Exploitation



European HST Archive (Baines)

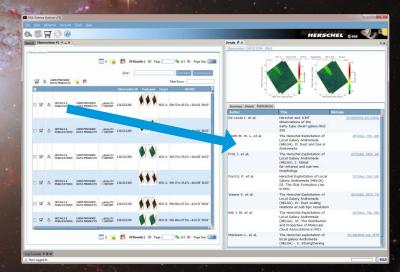
gaia archive					and the	and a	192			e	sa
ME SEARCH STATISTICS	HELP DO	CUMENTA	TION VOSPACE SHARE ADM	(IN)				97			
mple Fam ADQL Form Query F	Results										
	Job ra	me:							Que	гу езал	npres
Gala tables 0 public dual 0 public dual 0 public g10 fazzy_1000 0 public g10 fazzy_100000 0 public g10 fazzy_100000 0 public g10 fazzy_1000000 0 public g10 gal	num Antes	ELECT DIS ROM publi HERE 1=CO	TANCE (POINT 'IORS', alpha, delta : gog Cataloguesource unignis (Point TCNS', alpha, delt	0, POINT('1088',206.41683,-29. a),CIRCLE('1085',206.41883,-29	007013) AS dist. 00701, 0.08333	(33)) ORCE			Quer	y	
🗷 💮 public g10_imc	Status		Job	Creation date	Num. rows	Size	-			e	
B @ public.g10_mw	1		1444318635024D	08-Oct 2015, 17:37:15	4266	1 MB	8			• 🗆	
B () public g10_quasars B () public g10_smc	1		1444318523040D	08-Oct 2015, 17:37:03	4266	1 MB					
III @ mithe of Name: public	g10_mw		.8596137D	08-0:1-2015.173636	4266	1 MB					
B public gl Size (rows): 2143	\$10 Miky W 475885	sy sources t	.8513603D	08-041-2015, 17:25:13	4266	1.545					
B 🗍 public gr. Density map:	-		.84611160	08-Oct-2015, 17:34:21	4266	1.MB					
# @ publicity	10000	-	39073865D	30-Jun-2015, 12:11:14	4266	1 MB					
B publicig	-		15682813D	16-Mar 2015, 12:34:43	1313	665 KB					
II @ publicite			Lona Jgsl_Source	18-Mar-2015, 12:33:04	0	0.425					*
B 💮 publicity			.5487847D	09-Mar-2015, 16:38:08	1890	270 KB					
B () public.ucac4_original_valid	1		14259154565150	09-May-2015, 16:37:37	1313	665 KB					
User tables	1	0	whatch igst source cata	09-May 2015, 16:36:23	0	0.48			~ •	- 11	3+
III 🐻 🗋 user_jsalgado cata	1		1425318251124D	02 May 2015 18 44 11	1313	665 KB					
H 🔄 🗆 user pulgado early stars			and the second s	or our 2010, 10 went	1010	000 00		-		-	
	1 14 14	1-20 of 25					Select all jo				

Gaia Archive consortium release (Gonzalez)



esa

Planck Legacy Archive (Dupac, Lopez)



Herschel Science Archive (Verdugo, Teyssier)

Individual Mission Science Exploitation



SC STARCH IN THE	Check	Postcard	Tarret	Mission	Instrument	Instrument Host	Product Tile	Product Type	Instrument Made	Data Details
+ Target			TRANSFE	CARTH	PENSACON OFFICS F4.300MM	INC	ARUB-P.300,1960225,4745	PDS1/HODUCT.ORX/HATOMA	CANETA	and the second se
Mission			TREALEY	EARTH -	R01.0-2055-01504948-F4.3-52949	840	ARL/0-17,300,19960225,4745	P053,P100-07,08587641044	CANSIA	
ENDANIS 2016			IPPINALITY	DARTH	HADDELIKAD ZDIEL PLANAK /2 110MW	840	ARE/01-17,300,11060221,4041	PDS1,PRODUCT,OBSERVATIONAL	CANEDA	
INDEWCOWLKSE INDEWCOWLKSE INDEWCOWLKSE			THE REAL PROPERTY.	LAT:	UDITEMBEORIE-FLISTERIC CAMERA F4 762MM	840	ARLIN-0,300,19860225,4745	P051,P001.07,0858467044	CANEER	
			TRANSFER	6404	HASSELBLAD 2015-LOV SOMMA PL3 10594M	840	ARLIN-19,300,19980225,4245	PDS3/RODUCT.ORERVMIDIAL	CAVERA	Logovi D MULEC PHOTOCOVICE DATABLET
			10000201	EAITH	UDITEM COLETANT RELD CAMERA FA TREMM	1840	AR.8147,300,18862275,4745	PDS1_MODUCT_ORERIATONAL	CANETA	1985-V13CHXARL8H7,300,18800225,4146 Title
CASSING HEAREINS			TRIMALIEY	EARTH.	ROLE-2055-0V SOMWA FK3 129MM	BMO .	AR.8149.300,18862225,4744	POST, PRODUCT, ORSERVATIONAL	CAMERA	with the state of the second state of the seco
C HET										1.8 Product Class
MARS EXPRESS			(POALD)	LARDH	HADRED AD 2015 PLANAR (2-110M)	INO	ARUBHP_300_19802225,4746	POSI_PRODUCT_OBSERVATIONAL	CANERA	POSL/HODUCT_DESIGNATIONS. Start Time
SWILL MISSIONS FOR			SPHALIEY	DAILY .	PENDCON OFFICS-14 300MM	BMD .	ARUE-P.302,1980225,4746	P053/R00UCL085ERMIDINAL	CANENA	rate-op-zatas relotz Bolgi Time
PEAN SPACE ASENCY # SCIE		and Acres	101041-010	0.0074	Interfactive Pathon an Internal	81.17	ADD. REAL WAY ADDRESS 10. 11.17	0101 00124 01 10038-0174-01	/10108	MAG-OD 20108 20108

Ner and a second second

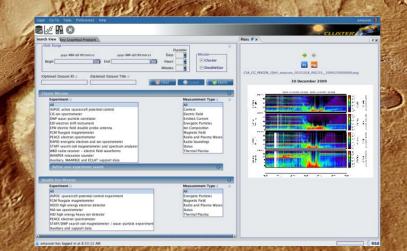
Planetary Science Archive (Lim, Grotheer) BepiColombo Archive (Martinez, Ortiz, Perez)

ine Medices Arbara	Jan 199					O In Sec. Sec.
5 15 H 🗣		0 0.0		和新聞	3 6 SOI	10
Dates Time Assessor	Observations #1 .A				Druh P x	**
revises -				_ 11	EX Clin Dynam (11647510)	
		[Mittents] + Page]	Se er ti de Bage Sier	an i li	to 1964751 Induced ET	
U Statument	250101	1000 Contraction (1000)	Non-Section Section	_	Observatory \$0H0	
Leturat	Detector	Observation Type	a Regis Date		Detector UT	
2 2 000	Floren	Live of sight collected ret	00000014-01/2007		Cleve Type: CME MATCH 285 Cleve Traf Son Put Diak	
					Churches CME WATCH185	
S & A ANA	UD	- 54	90-00-00 34/97/2987		Providented AZ file	
2 - 004	NA	10	100000034-0012007		Regim Date: 00:00:09 34:00:0007	T
					Ella Hama: ell'20070714.000000	
3 2 2 BOOK	HED	82	00-00-00 14/07/2007	•	Fig. County HTTS	
2 _ come	LETY	8.7	80-00-00 14-07-2007		Fix line 200040 bytes	
a de la constar	1000		and an an an an		Have Farger 123-Augstrone	2011/1 15 22:00
a la costar	UNION.	MP2	00000034/07/2007		Charlieren 105,3550C_AL_L008 Charlieren beckede	
a di citari					Too Postan 1424 more	Call Division
2 - coster	EPHIN.	808	06406814447/2007		The Arribe 1810 degrees	
a de come	UNIE I	82	00000014-01/2007		Rev See (2004) 2004) ersec	
				_	Top Court 1	
2 2 coster	EPHIN.	PU	85-00-08 14-01-2027		for Time 12.50 seconds	
Costa	(7+2)	MR2	00-00-00 14-07-2007		Dirgon Rachet Science	
					Ware Lief Red Augebors	
2 2 4 894	HED	14	06-00-08 14-07/2007	•	ANTITAL MART ACCULATION AND	
A COSTOR	LION -	M2	00000814/07/2007			and the second second
1 2 2 000	0.04		And the Designment		() () () () () () () () () ()	1
2 2 2 104	LED	82	00-00-00 3-4-97-2987			
100000				100	Concerned a	
3 & SAM	2-	Scat	00.00.09 \$4.07.2007			A CONTRACTOR OF A CONTRACT
2 - SHAN	2	Scan	00000014-01-2007	•		100 Carlos (1997)
2.0	677	CME WATCH 185	ID-00-08 14-07-2007	1000		

Soho Science Archive Including Proba-2 data



Archive Image Browser (Rosetta)

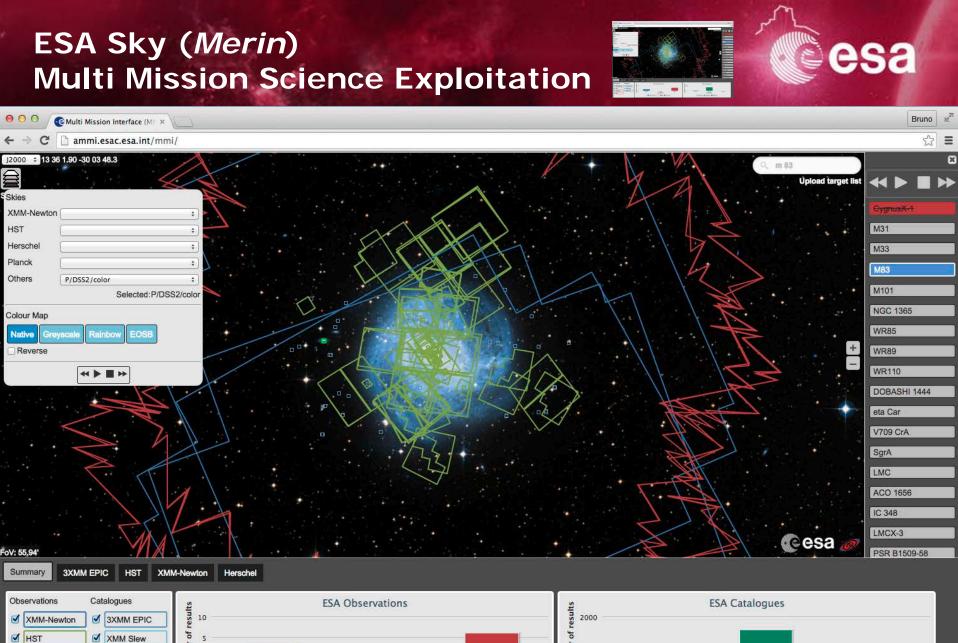


Cluster Science Archive, Including Double Star data

Need for multi mission science

- Large and wide set of ESA astronomy archives
- Science is more and more multi wavelength
- As missions enter into post operations and legacy phase, need to "hide" projects jargons and complexity, and offer simpler science data exploration tools
- Build on top of existing ESA archives
 - Re-use existing data infrastructure
 - Re-use of VO protocols (to make it open to other non ESA archives)
- ESASky for astronomy
- New PSA for planetary science
- Start thinking multi disciplinary (astronomy, planetary, solar heliospheric)

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 10

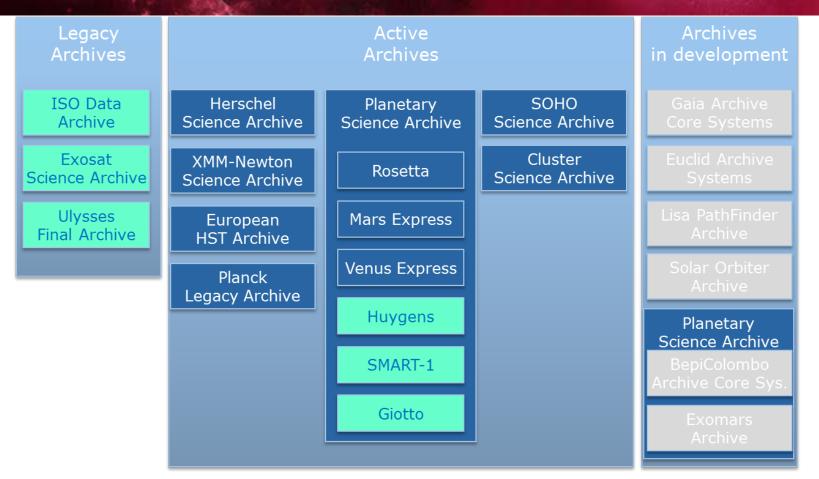






Long Term Preservation Strategy





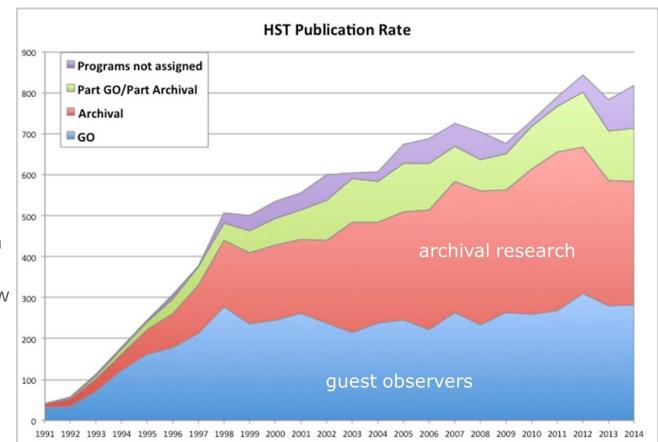
Preservation of data, software and knowledge

Regular Technology Evolution

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 12

Archival Science through Publications

- Number of Publications is a usual metrics to measure the success of a space science mission
- HST produced more than13.000 publications to date
- Number of publications coming from "archival" data (vs Guest Observers) has increased with time and now represents ~50% of HST annual publications
- Long Term Preservation of Data increases the science results from a mission !



Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 13

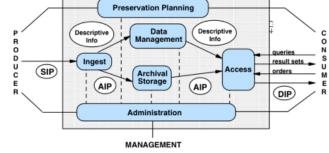
Long Term Preservation Strategy

- Consolidation of all ESA Space Science Archives at ESAC, with strong re-use across projects, ensuring easier and cheaper long term data preservation
 - Hardware infrastructure
 - Software architecture and code, including technology migration
 - Human technical and scientific expertise
 - Multi missions, multi instruments science exploitation
- Long term preservation of data processing capabilities (Navarro)
 - Preserve software coming from various places
 - Provide data processing capabilities as a "service"
 - Bring the "user software to the data" instead of the "data to the user"
- > Sharing and preservation of knowledge, including international cooperation
 - IVOA, IPDA, within ESA (with EO), CCSDS, PV2015, ...

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 14

ESAC Archives Technologies

- Flexible and modular system architecture
 - OAIS standard (Open Archival Information System)
 - Interfaces with data providers, external tools and archives through VO protocols
 - Enabling building added value services on top of our archives
- Need to support all phases
 - Development, operations, post operations, and legacy (LTDP – long term data preservation)
- Technology monitoring and update at regular intervals
 - Java Rich Client to Web Thin Layer
 - Database: Commercial to Open Source
 - Web Services and Interoperability (VO)
 - *"Big Data"* challenges (Gaia, Euclid)





Long Term Preservation Upcoming Challenges

- More and more ESA missions entering Legacy phases
 - Venus Express, Planck in 2016
 - Herschel, LisaPF in 2018
- Preserve and improved science exploitation facilities
- Archive technology migration for some old archives (ISO, ...)
- Setup Software preservation framework
- Define data persistent identifiers (ie DOI Digital Object Identifiers)
- ESA LTDP+ programme being defined for 2017-2022



Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 16

International Collaboration and Interoperability amongst Archives

IVOA - International Virtual Observatory Alliance



- Formed in 2002, 20 member projects
- Defines interoperability standards (VO framework) amongst astronomical (ground and space based) archives
- Working Groups and Interest Groups per technical domain (data access, data model, registry, applications, semantics, operations, ...)
 http://www.ivoa.net/

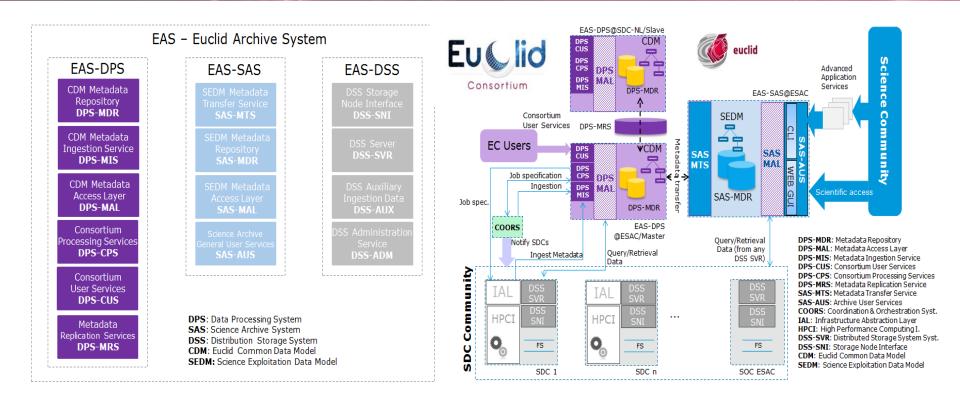
IPDA - International Planetary Data Alliance



- Formed in 2004, 12 space agencies
- Defines archiving guidelines for planetary data
- Defines interoperability standards amongst planetary archives
 http://planetarydata.org/

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 17

Archives Integration within Projects



Archive is fully part of Science Operations, from the start

Strong collaboration with SGS and Consortium

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 18

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 19

Archives full integrated in projects

- Archive is <u>not</u> an after thought at the end of the mission
- Archive is to be part of SciOps development from the beginning
- > Archive is more and more a collaborative process with many groups involved
 - Within Archives Groups
 - With SOC/SGS
 - With Consortium
 - With external Community
- Need for more collaborative methods
 - Scrum methodology for daily interactions with SOC/SGS
- Open source development for collaboration with Consortium

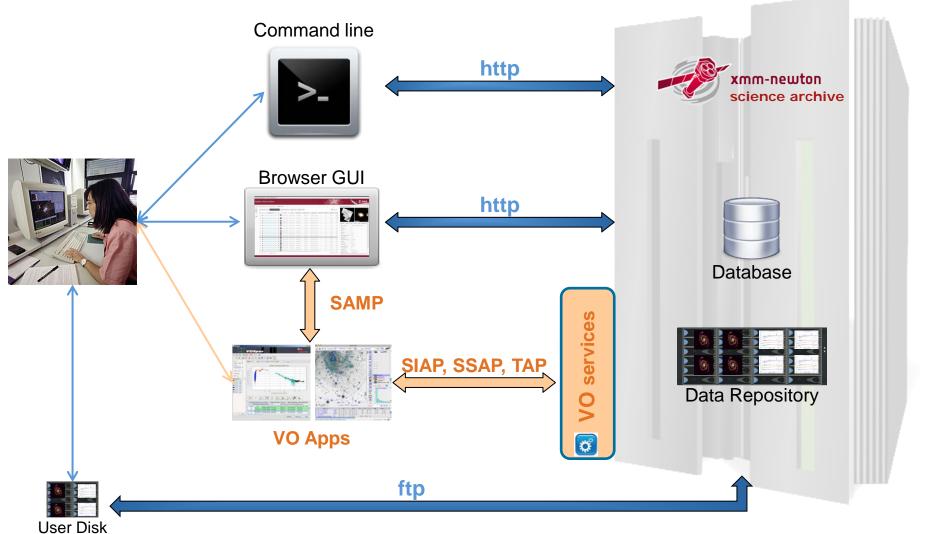






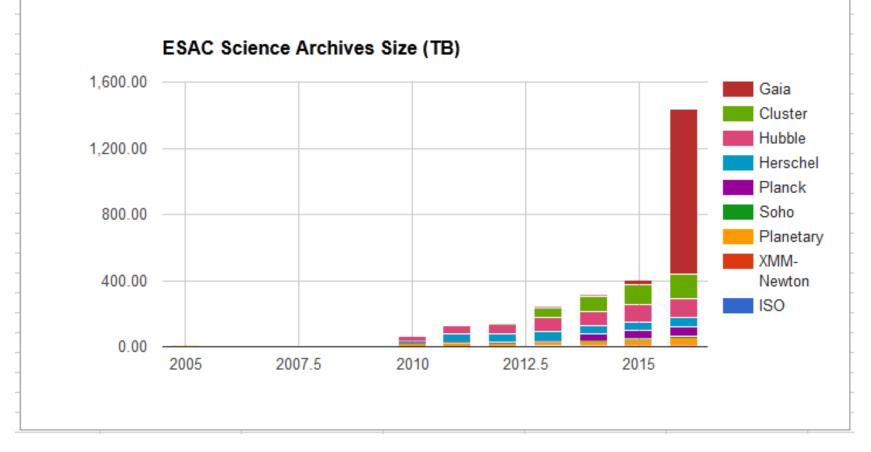
Standard Archives Architecture





Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 20

ESAC Archives Volume Evolution



All data stored on hard disks and distributed through Internet Euclid will add up to ~150 PBs by 2023

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 21

Need for new paradigm

New ways required to access the Gaia catalogue and associated data

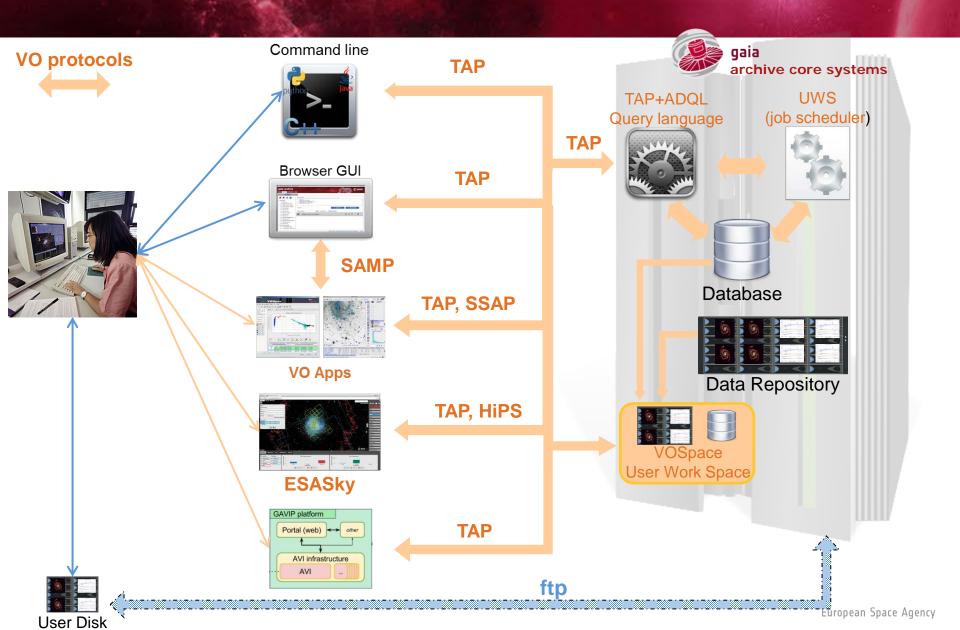
- Powerful query mechanism, asynchronicity of results
- One "query interface" for all archive services and VO services
- User can not download all catalogue and all data
 - Need to have user workspaces *IN* the Archive
 - User database space, user disk space
 - User workspace shareable amongst various users
- Bring user code to the data (OMullane)
 - Part of the user workspace in the archive
 - Share code with other users

The user works with the data WHERE the data is : Archive 2.0 concept

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 22

Gaia Archive Architecture





Mission and Archives metrics ?

- Mission success usually measured by number of scientific papers
- How to measure success of the archives ?
 - Number of active users / IP addresses
 - Volume of data being downloaded vs size of the archive
 - Ratio of data-out vs data-in ?
- Big differences between missions, between missions phases
- > Interested to hear how other data centres report on archives usage...



Science Archives

Fulfil various Science Operations functions

- Maximum science exploitation of data
- Ensure long term preservation of data, software and knowledge
- Support various functions of Science Ground Segments

Scientists and Engineers working together

- Archives must be science driven
- Archives require strong IT expertise

Towards Archives 2.0

• Open data, open source, open archives

Christophe Arviset | SCIOPS 2015 | ESAC Science Data Centre | ESAC | 24/11/2015 | Slide 25



Thanks!

http://archives.esac.esa.int Christophe.Arviset@esa.int

European Space Agency