

## From ISO to Gaia : a 20 year journey through ESA Science Archives

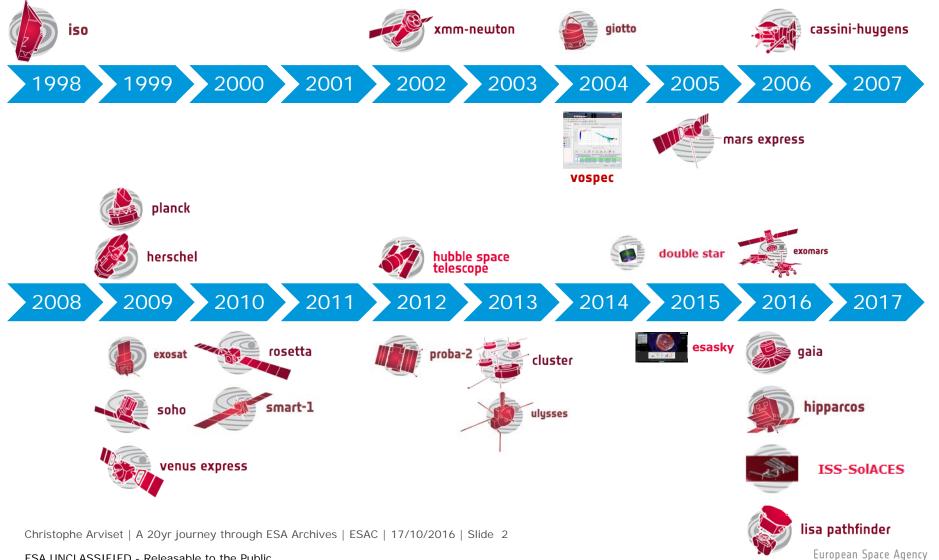
Christophe **Arviset**, Deborah **Baines**, Isa **Barbarisi**, Sebastien **Besse**, Guido **de Marchi**, Beatriz **Martinez**, Arnaud **Masson**, Bruno **Merin**, Jesus **Salgado** 

ESAC Science Data Centre ADASS 2016, Trieste, 17/10/2016

Issue/Revision: 1.0 Reference: A 20yr journey through ESA Archives Status: Draft ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

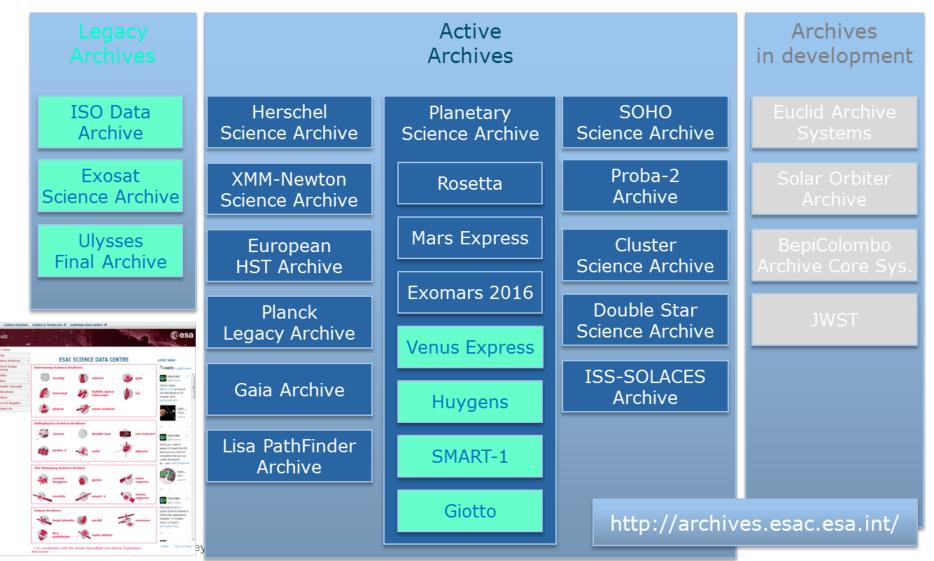
## **ESAC Archives : an ever growing family**



4.

## **Archives in different phases**

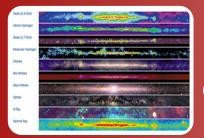




ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

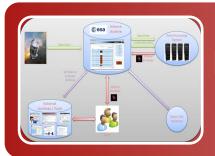
## **ESAC Science Archives Strategy**



# Enable maximum *scientific exploitation* of data sets



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

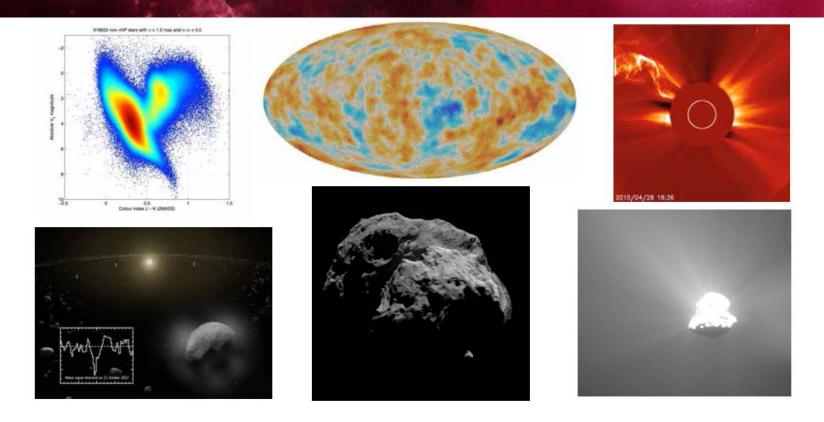


Enable cost-effective archive production by integration in, and across, projects

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 4

## **Enable Maximum Science Exploitation**





### Scientists and Engineers working very closely together

#### Science driven Archives

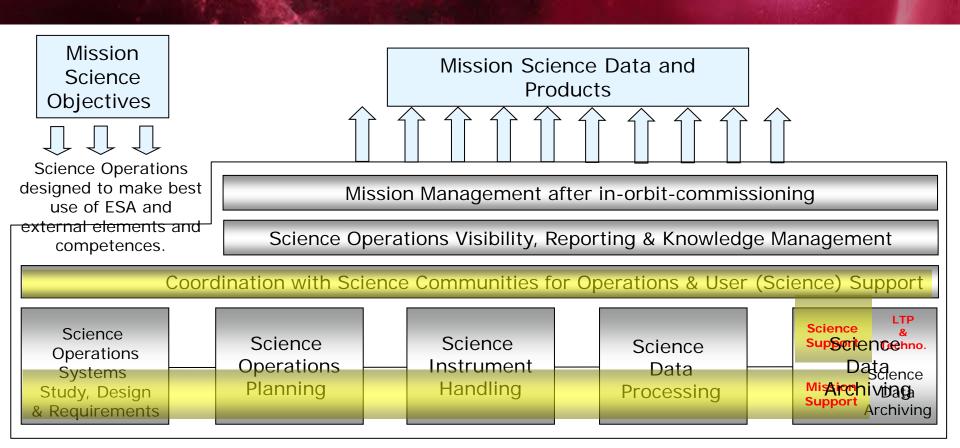
Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 5

## Long Term Preservation Strategy

- 1. 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 mission, multi instruments science exploitation
- 2. Long term preservation of data processing capabilities (on-going)
  - 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"
- 3. Sharing and preservation of knowledge, including international cooperation
  - IVOA, IPDA

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 6

# Archives as a Core Component of ESA Science Operations



Archive is NOT an afterthought at the end of the in-orbit operations !

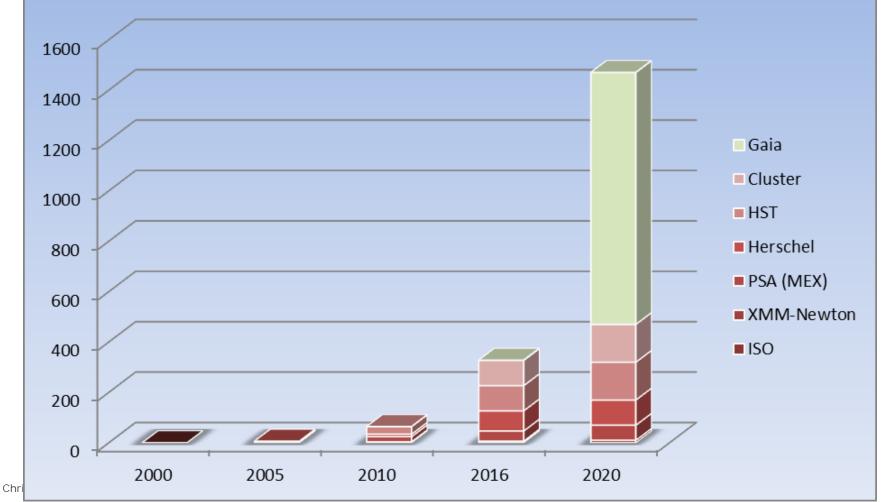
#### Archive development starts in the early phase of the project !

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 7

Archives Volume evolution (2000-2020)

## esa

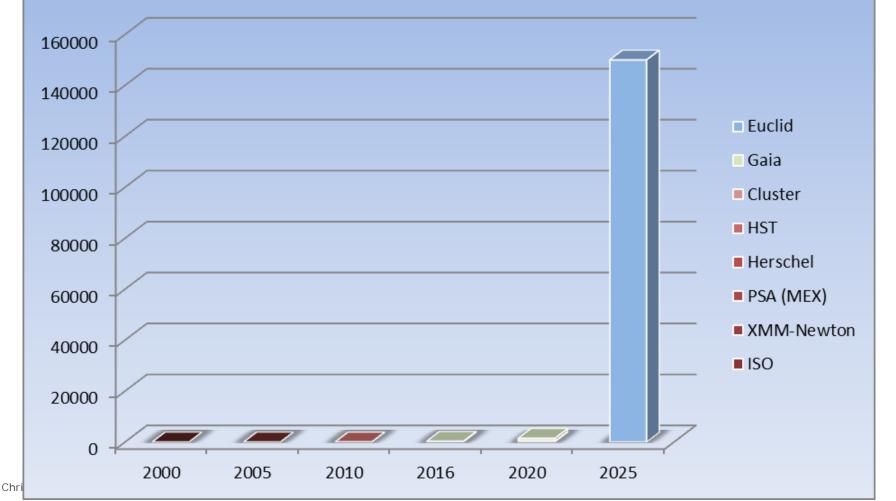
## **ESAC Archives Volume (in TB)**



Archives Volume evolution (2000-2025)

## esa

## **ESAC Archives Volume (in TB)**



## Software Design DOES Matter !

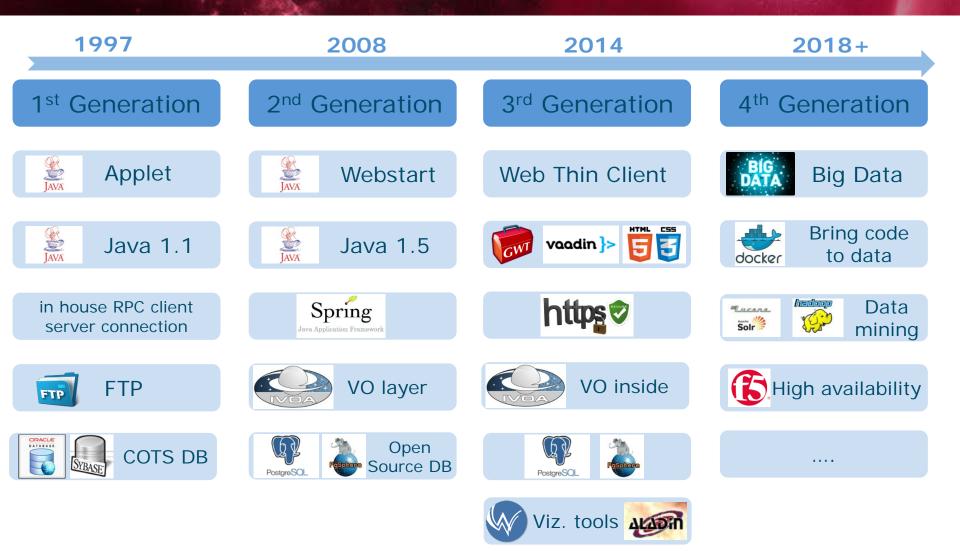


ISO Data Archive Top Level System Design - 1998 **ESAC Archives Common Architecture - 2009** esa BKRP Processed ISO data TOP LEVEL ARCHITECTURE. LEVEL 0. product Update (CDROM jukeboxes) Archive stored QUERY product INGEST MANAGER product BKRP 7 e-mail location product USER METADATA INTERFACE processing request request TDF reques Pipeline User Archive TDF Store User rocessing Engine nterface results processed REQUEST product EAS-DPS@SDC-NL/Slave FTP **Euclid Archive System** DPS CUS CDM product 2015 DPS Science Community í à à CDROM CPS Advanced DPS Application DPS MAL DATA Services MIS DPS-MDR EAS-SAS@ESAC Consortium SEDM User Services DPS-MRS STORAGE LAYER Metadat EC Users DPS CUS **VCDM** SAS SAS MAL MTS WEB DPS Job specification CPS DPS ⇒ Indestion DPS MAL Scientific access ൭ MIS SAS-MDR Job spec. DPS-MDR COORS EAS-DPS @ESAC/Master Query/Retrieval DPS-MDR: Metadata Repository Notify SDCs DPS-MAL: Metadata Access Layer Data (from any Ouerv/Retrieval Ingest Metadata DPS-MIS: Metadata Ingestion Service DSS SVR) Data DPS-CUS: Consortium User Services Community DPS-CPS: Consortium Processing Services DPS-MRS: Metadata Replication Service IAI IAI SAS-MTS: Metadata Transfer Service SAS-AUS: Archive User Services COORS: Coordination & Orchestration Syst. HPC] HPC1 . . . IAL: Infrastructure Abstraction Laver HPCI: High Performance Computing I. DSS-SVR: Distributed Storage System Syst. FS SDC FS FS DSS-SNI: Storage Node Interface CDM: Euclid Common Data Model SEDM: Science Exploitation Data Model Christophe Arviset | A SDC 1 SDC n SOC ESAC

ESA UNCLASSIFIED - FUILUAUDIO LO LITO I UDITO

## **Archive Technology Evolution**





Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 11

## 1<sup>st</sup> generation of Archives (1998...)

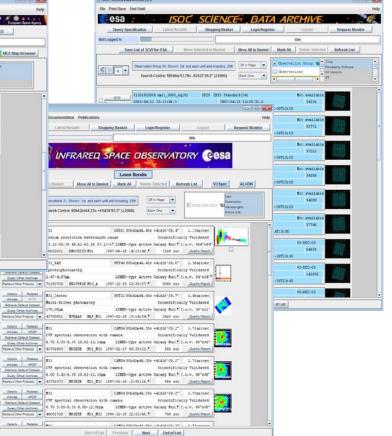
### Java Applets, html could not support all requirements at the time

12

#### 4- BOC Science Data Arc . . . . . . eesa Planetary Science Archive Query Specification Delivery Basket Login/Register Logout URDOT TMMA Menton Science Archow 707 1 Most I connect b File Print/Save Results Find Field Rediation Monitor Document lie'r. Cesa XMM-Newton Science Archiv Move All to Basket Mark All Delete Selected Refresh List MEX Map Browne Query Specification sitta Shopping Banket Login Reg < > · · Mie 23.9 hours 25 in Page 💌 Products 258. Showing results 1 to 25 Products 83 Move All to Basket Mark All Delete Selected ALADN 20060512 VCD R121A.TAB 2006-05-12T01:32:46.00 values 42 Shown, fat unli 25h 2006-05-12701:34:32.00 - 90W 0 2 1 25 in Page Expo 1001 Search Centre: 80h42m44.33s +41d16'87.5" (J2080) VEX-Y/V-SPICAV-3-SOIR-V1.0 VENUS EXPRESS SET/VENUS SPICAV 3 BOIR V1.0 SPICAV SOLAR OCCULTATION IN THE INFRARED SPICAV-SOLR con N.A. 20060512 WD TRT. TAB 2006-05-12T01: 32:46.00 2006-05-12701:34:32.00 A11157/681 #31 Com 0100 2000-06-25 05:46:12.0 2000-05-25 08:04:42.0 4709 Michael Wateon VEX-Y/V-SPICAV-3-SOIR-VI.0 VENUS EXFRESS SRY/VENUS SPICAV 3 SOIR VI.0 on N.A. SPICAV SOLAR OCCULTATION IN THE INFRAPED SPICAV-SOIR Icon N.A. RG81 3E5(1) R382 3E2(1) Galaxies and Scoups of Galaxies MUS-ACTIVE GALACT Quest Sheever Public Data Queity-Re-10060526 M01 C01 TET. 3 2006-05-26T01:39:15.00 2006-05-26T01:41:45.00 VEX-Y/V-SPICAV-3-SOIR-V1.0 VENUS EXPRESS SEV/VENUS SPICAV 3 SOIR V1.0 1-1 10h42a43.00s +41415 46.0" 26 axceed 0111570401 231 Care SPICAV SOLAR OCCULTATION IN THE INFRARED SPICAV-SOLR 0100 2000-06-25 08:12:41.0 2000-05-25 20:59:13.0 46012 Michael Watrom 20060526 M01 C01 167A.\* 2006-05-26T01:39:15.00 2006-05-26701:41:45.0 Demis Atlates EFF EFF(1) 2051 7F(1) 2052 7F(1) 9681 5E5(1) 9682 3E5(1) Duery Other Arshives Galaxies and Scoups of Galaxies - BOB-ACTIVE GALACY Quest Observer Public Data VEX-Y/V-SPICAV-3-SOIR-VI.0 VENUS EXPRESS SKY/VENUS SPICAV 3 SOIR VI.0 SFICAV SOLAR OCCULTATION IN THE INFRARED SPICAV-SOIR 0111570601 E31 Cage 10h42a43.00z +41415'45.0\* 20060526 M01 C01 166.7\* 2006-05-26T01:39:15.00 2006-05-26701:41:45 129 Sources 0191 2000-12-25 01:01:33.0 2000-12-28 03:43:23.0 13310 Michael Vetera EPH FT(1) M(31 FT(1) M(32 FT(1) R(3) SE3(1) R(52 SE3(1) AB B(5) (B FW)(1) Galexies and Scoups of Galexies - MCS-ACTIVE GALACY Guest Observer Public Data - Dualty-edium resolution wavelength range 0105270501 EDiCore 00h42a43.00s +41415'46.0" 26 arcse 0235 2061-06-29 03:16:05.0 2001-05-29 06:12:56.0 10611 Keith Mason Map: Any - Grid Zoom x1 + Icen W.A. -002001 KROCHIJU R31 Icon N.A. Details Aticles R501 SES(1) R502 SES(1) 08 UW01(2) NADIR 👻 White Dwarf Binaries, Nestron Star\* MOS-ACTIVE GALACY Guest Ob 11\_PAS Search Cancel Clear Search Print, Advanced Search trephotosetry .00"N 19250"E Query Other Archives 2. 47-4. 87pa 0105270101 #31Cor 10h42s43.00s +41415 46.0" 26 ercee 172 Sources 0205 2001-06-29 06-15-17 0 2001-06-29 27-10-20 0 52903 Eatrh Masor EPS F7(1) BUS1 F7(1) BUS2 FT(1) RUS1 SES(1) RUS2 SES(1) HB UV82(3) HB UV82(4) Cutaia Related M31\_Center Retrieve Default Densel | Halti-filter photometry 170.00LB Query Other Archives 0112570701 E31 Cage 00542a43.00s +41415'45.0" 26 axceed 0301 2302-01-06 15:30:40.9 2002-01-06 17:53:57.0 4517 Michael Watron Cutain Raisted 201 con N. A. Details Attained ROB1 SE3(1) R362 SE3(1) 68 0791(1) Galexies and Scoups of Galexies BOB-ACTIVE GALACY Quest Observer Public Data Qually CTT spe Start of List Previous Next End of List Cutala Raistet 832 Arioles HFDP Retrieve Default Dataset CTF spectral observation with casers **XMM-Newton Archive** Outsin Buister 811 Apples HPCP Reptars Default Dataset Chevy Chevy Australia Chevy Chevy Australia Chevy Chevy Australia Chevy Chevy Australia stats | Painted | w

#### **Planetary Science Archive**

#### **Integral SOC Archive**

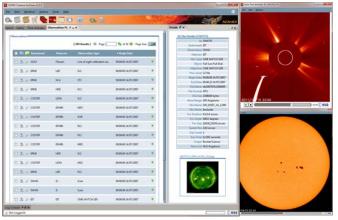


**ISO Data Archive** 

Christophe Arviset | A 20yr journey through E

## 2<sup>nd</sup> generation of Archives (~2005...)

#### Java web start, more user friendly GUI

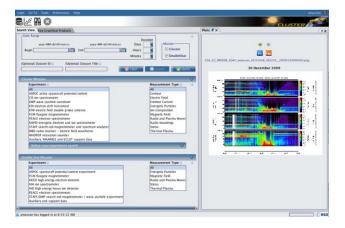


#### Soho Science Archive Including Proba-2 data

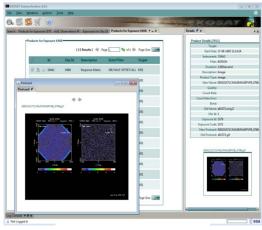


Christoph Herschel Science Archiver | ESAC | 17/10/2016 | Slide 13

ESA UNCLASSIFIED - Releasable to the Public



#### Cluster Science Archive, Including Double Star data



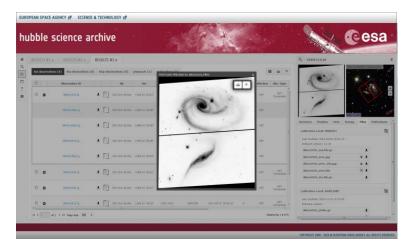
#### **Exosat Archive**

European Space Agency

4

## 3<sup>rd</sup> generation of Archives (~2013-...)

#### Thin layer web clients



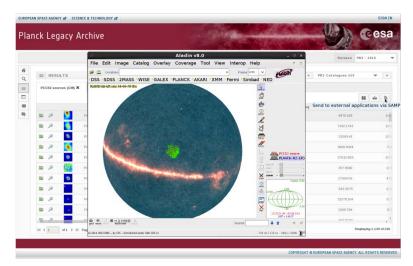
#### **European HST Archive**

gaia archive					Electro)	4/4	A.		e
HOME SEARCH STATISTICS	HELP DOCUMENTATION	VOSPACE SHARE ADM	IN		\$ . M	1	9		
Simple Form ADQL Form Query I	Results								
s ≈ * II <	Job name:							Quer	y esz
Gaint tables B    D    public class B    D    public class B    D    public class B    D    public cji0_lazzy_100000 B    D    public cji0_lazzy_1000000 B    D    public cji0_lazzy_1000000	FROM public, go	cataloguerourice IS(Point( ICRS , slpha, delt)	6, PODMT(1205), 266, 41683, -29- al, CIRCLE (13085, 266, 41683, -29-	Reset Form	33)) ORCER		SC 25mil	Query	1
B 💮 public.g10_imc	Status Jo	b	Creation date	Num. rows	Size				
B () publicigi0_mw B () publicigi0 quasars	✓ □ µ	44319635024D	08-Oct-2015, 17:37:15	4266	1 MB			•	
B Duble glo smc	✓ 0 14	44319523040D	08-Oct-2015, 17:37:03	4206	1.MB			-	
I Cathlent Name: publi	o g10_mw 510 Milky Way sources table	.8596137D	08-Oct-2015, 17:36:36	4206	1 MB			•	
III Public gl Size (rows): 2143	475885	#5136030	08-Oct-2015, 17:35:13	4265	1.665			•	
B Duble ge Density map		.84611160	08-Oct-2015. 17:34:21	4266	1 MB			•	
III (i) publicity III (i) publicity		99073865D	90-Jun 2015, 12:11:14	4266	1 MB			•	
III 🚍 publicig		35682813D	16-Mar 2015, 12:34 43	1313	665 KB			•	
# @ publicity		_cata_igst_source	16-Mar-2015, 12:33:04	0	0.68		8 1	0	
		54878470	09-Mar-2015, 16:38:08	1830	270 KB			•	
B 💮 publicity		259154565150	09-Mar-2015, 16:37:37	1313	666 KB				
🛙 🕘 public.ocac4_original_valid	🖌 🗆 14					-	-		
II () public.ucac4_original_valid II User tables		natch_igsl_source_cata	09-Mar-2015.16:36:23	0	0 #8	1.1	8 1		• III
🖲 💮 public.ucac4_original_valid	✓ □ ×	nakch_igsl_source_caka 25318251124D	09-May 2015, 16:36:23 02-May 2015, 18:44:11	0	0 KB 666 KB	8	8 1	•	

**Gaia Archive** 

ESA UNCLASSIFIED - R

| 17/10/2016 | Slide 14



esa

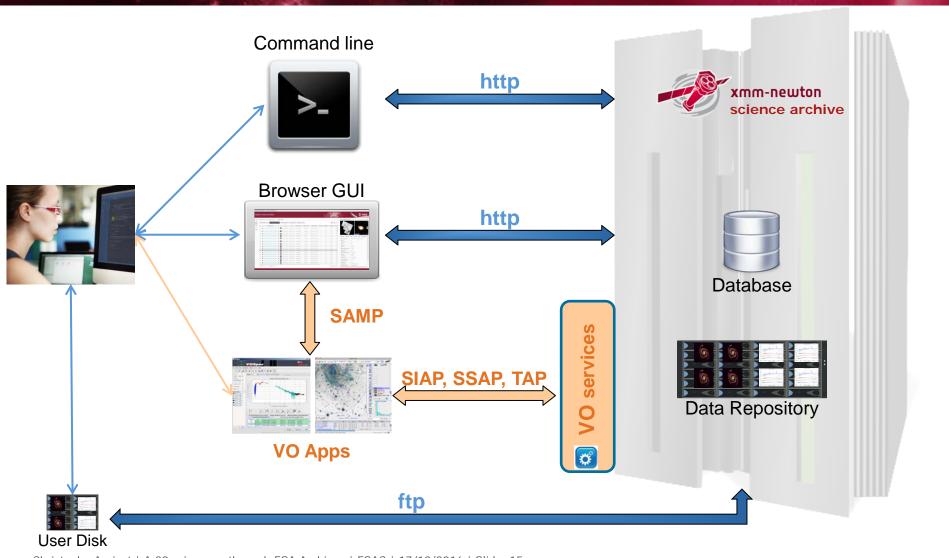
#### **Planck Legacy Archive**

(i) archives assome	salet/opsa_test/#ITab	le Viene							0 0 9	hadsop	+ •		+ 1	h 1
OPEAN SPACE AGENCY	of SCIENCE & TEN	DINOLOG	sy of cosm	os d'									_	STEN
50.80	science a	ana a	AT IN							X	in fact	e	es	
	A Hide All	5	<b>9</b> . (	2			QR	ŵ.		1 <b>1</b> 0	R05_CAM1_20160	097180627	TABL	x x
Rosetta		E	Pestcard	Product Identifier	Observation Start Tin +	Observation Stop Time	Target	Mission	Instrument	Processing Lev				
				BO5_CAM1_201608097231752	2016-08-09 23:17:52	2016-08-09 2517 52	679.64	Assets	NACAN	2.4		belos.		
			13	NOS_CANY_201408097231412	2016-08-09 23 14 12	2016-08-09 23 14:12	smca	Rosta	NUCAU	1		Sec.		
			3	ROS_CAM1_201608097230802	2016-08-09 23 08 12	2016-08-09 23 08 02	6795-5	Rosetta	NUCKI	- 2		1000		
				805_CAM1_201608087230422	2016-08-09 29 04:22	2016-08-09 23 04:22	17966	Rota	NUCRE	2	1	100		
				805,CAM1,201608097181617	2016-06-09 18:16:17	2016-08-09 18:16:17	6706.6	Roseta	NUCH	2		Surge		
epts .	~ 0		-	805,CAM1,201608091181238	2016-08-09 18 12:38	2016-06-00 1812:38	smca	Rosta	NUCAN	2.4	Summary Geor			
AUC	~ 0		E	R05,CAM1,201608091160627						1.0	Product Identifier	ROS, CAMIL,		3627
CONSERT				805_CAN1_201608097180247	2016-08-09 18:02-47	2016-08-09 18:02-47	679/C-G	Ameria	NGCAN		Start Time	2016-08-09 1		
COSAMA				805,CAM1,201608097131442	2016-08-09 13:14:42	2016-08-09 13 14-42	67916-6	Ronta	16030	2	Stop Time	2016-08-091	18:06:27	
MD45		۲	12		2016-06-09 13 11 62	2016-06-09 13 11 02	stric g	No.	NACAM		Target	Electra .		
NAICAM				805,CAM1,201608087130452	2016-08-09 15:54:52	2016-06-00 13:04:52	67/1C-G	Roma	NACAN	2	Instrument	NAVCAM		
CSAS			15	ROS_CANFL_201608097130112	2016-06-09 13-01-12	2016-06-09 13:01:12	6796-6	Roota	NHICAN		Processing Level	2		
utruments Types	~ 0			RD5_CAMT_20160839T081308	2016-06-09 08 13:08	2016-08-09 08 13:08	629,6,6	Rosette	NHOW	- 2	Instrument host	oteler		
			14	805_CAM1_201408097080928	2014-06-09 08:09 28	2016-06-01-08:01:28	8795-6	Roota.	1050200		Instrument type	CCD CAMER		
ine .	~ 0			805_CAM1_20%08097080318	2016-06-09 08:03:18	2016-06-09-08:08.18	smca	-toora	NACAM	2	Sub-instrument			
rocessing Level	~ •			RD5_CAM1_201608091015938	2016-08-09 07:59:38	2014-08-01 07:59:58	67866	Rents	NUCAH		Version	1.0		
ree Search			ž.,	905,CAM1,201608097211133	2016-06-09 03:11 33	2016-06-09 03 11:33	1795-5	Rosetta	NNCAN	2	Product Title	RDS_CAMP_2	0140809718	0627
9	-	1									See Label File			

#### Planetary Science Archive

## **Standard Archives Architecture**





## Need for new paradigm

- 1. 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
- 2. 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
- 3. Bring user code to the data
  - Part of the user workspace in the archive
  - Share code with other users

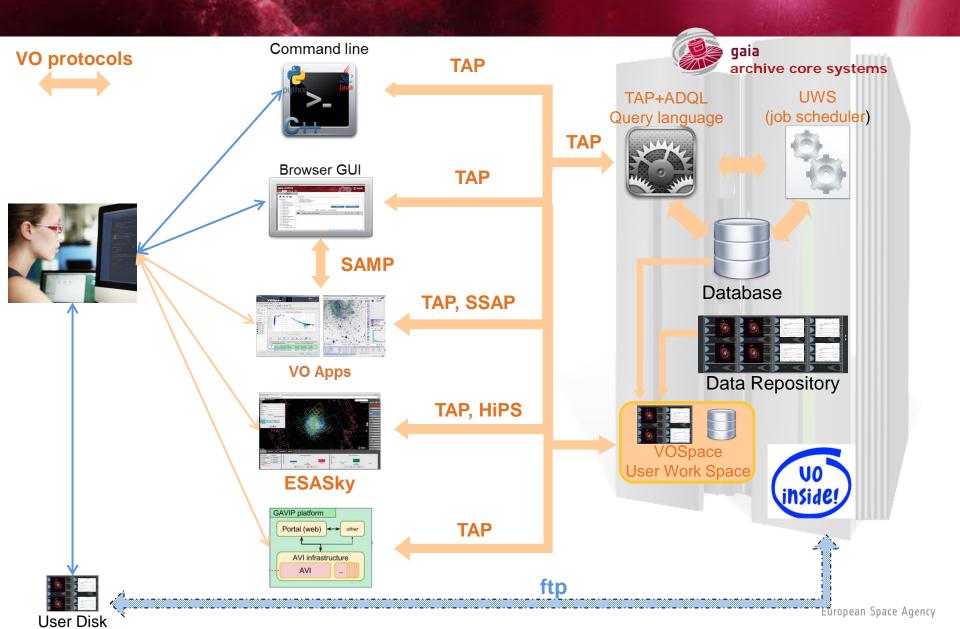


#### The user works with the data WHERE the data is : Archive 2.0 concept

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 16

## **Gaia Archive Architecture**

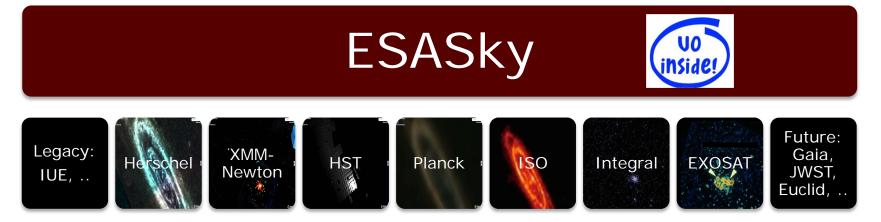




## Towards multi wavelengths Science data exploitation

- **Goal**: to facilitate data discovery and archival science for ALL users
  - Multi-wavelength
  - Project agnostic
  - Exploration
- Interface "on top of" all ESA astronomy archives

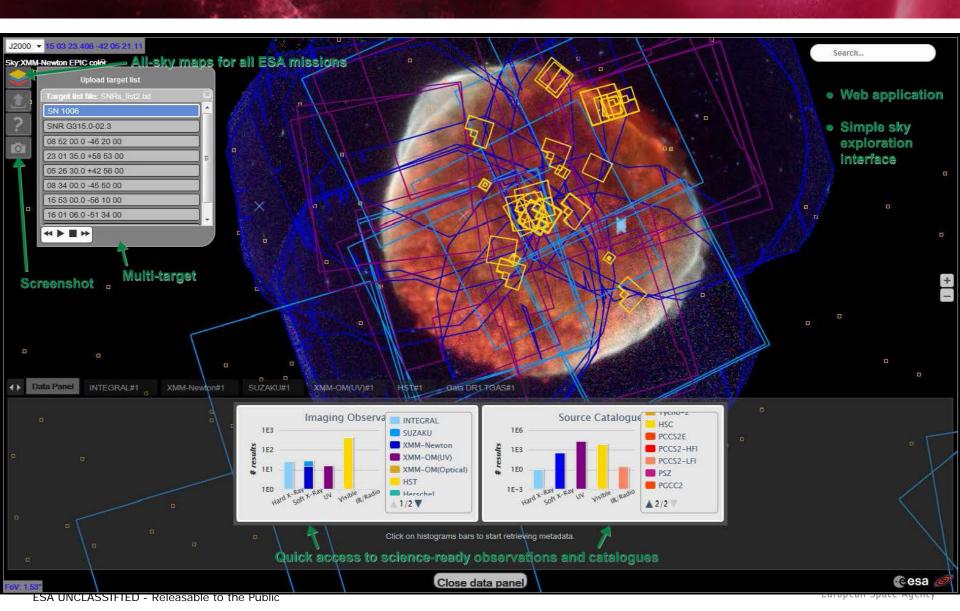




Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 18

### sky.esa.int





## Login or not Login : that is the question





## Semi Open access (Free access to query, need login to retrieve data)

- "Burden" for the user, especially for public data
- Required to access proprietary data
- Enables asynchronous services
  - (long queries, big volume download, user spaces)
- Easier usage stats, better knowledge of the users



#### Full Open access (no login)

- Instant access for the users
- No reason to login for public data access
- Usage stats based on IP addresses more complex and not as accurate

#### Conclusion : combine both ! Open access is default, offer login for extra services

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 20



#### 1. Mission success usually measured by number of scientific papers

ESA-led Missions	Launch	<2000	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	To
COS-B	1975	173	1																ľ
IUE	1978	3,217	147	130	76	70	89	31	21	30	25	26	39	30	15	19	15	14	3,9
Exosat	1983	706	8	3	4			2	1	2	1	2			1	2	3		7.
Hipparcos	1989	729	194	154	147	133	107	100	120	92	89	89	85	67	89	71	59	42	2,3
Ulysses	1990	898	83	138	39	89	57	56	36	42	52	38	59	34	43	34	34	19	1,
ISO	1995	490	180	138	146	129	123	107	67	35	33	33	28	10	5	5	3	2	1,
SOHO	1995	588	292	203	283	301	324	321	271	368	332	323	274	298	243	189	211	122	4,9
Huygens	1997		7	8	15	7	12	24	21	26	26	16	15	8	11	7	5		2
XMM-Newton	1999		23	91	100	231	329	298	327	334	278	314	332	352	298	330	306	220	4,1
Cluster	2000	54	1	38	22	65	127	168	145	123	183	176	192	232	187	171	155	124	2,1
INTEGRAL	2002					79	27	67	99	84	84	96	80	60	67	45	47	41	8
SMART-1	2003						2	8	10	5	10	12	6	5	7	7	7	1	8
Mars Express	2003					4	21	49	87	75	83	89	109	96	82	64	63	19	8
Rosetta	2004						22	16	36	58	29	17	39	25	54	29	41	54	4
Venus Express	2005								21	44	79	48	41	54	75	60	40	52	5
Herschel	2009												228	109	255	323	336	241	1,4
Planck	2009													48	69	118	328	225	7
PROBA-2	2009													7	7	22	13	6	5
Gaia	2013																32	25	5
Non ESA-led	-	-															-	-	
HST	1990	2,329	537	558	602	607	611	681	714	728	705	678	734	788	844	784	825	510	13,
Cassini	1997	22	1	16	18	12	49	108	150	157	173	199	234	184	177	135	127	38	1,
Double Star	2003						1	27	7	12	40	11	15	11	6	4	4	2	1
Suzaku	2005								20	50	78	138	92	109	101	94	78	2	7
AKARI	2006								2	29	22	21	47	50	61	37	29	3	3
Hinode	2006									62	105	135	133	134	129	100	97	63	9
IRIS	2013															3	22	33	5

pean Space Agency

			E	SA Science Arch	ives Usage S	Statistics				
SS	Mission	In-orbit operations	Period	Active users / month	Volume downloaded / month (TB)	% of total volume downloaded / month	Archive total size (TB)	Notes		
				Active	Archives					
			2015-H1	171	3.43	3%		CSA @ESAC data		
	Cluster	2000-	2015-H2	194	2.57	3%		distribution since Oct 2014		
	ordator	2000	2016-H1	171	2.34	2%		Since 2015-H2, archive size		
			2016-H2	124	2.85	3%		reflects only Science Data		
Mis			2015-H1	901	10.99	21%	52.67			
	Herschel	2009-2013	2015-H2	855	7.89	20%		7 Archive used as prime mean		
	Thersonier	2003-2013	2016-H1	944	14.74	23%		for data distribution		
			2016-H2	797	15.09	19%	78.35			
			2015-H1	204	1.01	1%	95.23			
	HST	1990-	2015-H2	181	0.19	0%		6 European HST Archive now at		
Ho		1000-	2016-H1	188	0.42	0%				
			2016-H2	231	0.13	0%	103.63			
			2015-H1	758	14.01	36%	39.40			
	Planck	2009-2013	2015-H2	615	12.16	23%		Public opening of the Planck		
	- ranon	2000 2010	2016-H1	609	1.23	2%		Archive on 21st March 2013		
			2016-H2	554	1.74	3%	52.00			
			2015-H1	395	5.74	20%	28.40	20 MEX, Rosetta and Huygens data also available via NASA 40 PDS 60		
	PSA (MEX, VEX, Ro		2015-H2	444	9.14	27%	34.20			
	Smart-1, Huygens a	nd Giotto)	2016-H1	403	5.52	15%	37.40			
			2016-H2	143	2.63	6%	41.60			
			2015-H1	168	0.34	5%	7.14			
	Soho @ ESAC	1995-	2015-H2	152	0.04	1%		3.00 ESA		
	00000 @ 20000		2016-H1	165	0.01	0%				
			2016-H2	152	0.01	0%	8.09			
			2015-H1	717	2.92		6.46			
	XMM-Newton	1999-	2015-H2	788	2.01	30%		Archive used as prime mean		
			2016-H1	930	2.07	30%		for data distribution		
			2016-H2	941	2.37	32%	7.39			
	Gaia	2013-	2016-H2	1,153	73.00	5214.29%	1.40	Gaia Archive at ESAC, released on 14/09/2016		
	ISS-SolACES	2008-	2016-H2	21	0. 11	8%	1.30	ISS-SolACES at ESAC released 31/08/2016 (values in GB)		
					y Archives					
			2015-H1	0	1.50	2%	83.90	D Legacy Archive		
	Exosat	1983-1985	2015-H2	1	0.93	1%				
			2016-H1	0	4.24	5%		(values in GB)		
			2016-H2	1	0.51	1%	83.90			
			2015-H1	88	0.49	0%	171.00			
	ISO	1995-1998	2015-H2	82	0.65			Legacy Archive since 2002		
		1000 1000	2016-H1	91	0.96	1%		(values in GB)		
			2016-H2	78	1.48		171.00			
			2015-H1	27	26.19		204.78	Legacy Archive		
	Ulysses	1990-2009	2015-H2	18	23.38		204.78	Legacy Archive Released in February 2013		
	019 33 63	1330-2009	2016-H1	33	18.88	9%	204.70	Avaluate in CD)		
he Ar			2016-H2	24	10.33	5%	204.78	(rados in ob)		
e Al										
AS	active users	means		d downloading data at leas						
_,			IP address downl	oading data at least once a	a month					

- 1. Mission success usually measured by number of scientific papers
- 2. 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 ?



. Miss				
	Mission	Data-in	Data-out	Ratio
. How	Cluster	30	29	1.0
	SOHO	4	5	1.2
	PSA	19	262	13.8
	Herschel	53	206	3.9
	HST	67	28	0.4
	Planck	42	168	4.0
	XMM	3	98	32.7
	Total (TB)	67	173	2.6

Christophe Arviset | A 20yr journey through ESA Archives | ESAC | 17/10/2016 | Slide 24

ESA UNCLASSIFIED - Releasable to the Public

- 1. Mission success usually measured by number of scientific papers
- 2. 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 ?
- 3. Still not perfectly unified metrics, still big differences
  - between disciplines,
  - between missions,
  - between missions phases
- 4. Would be useful to hear how others archive centre do?



## Looking back and looking forward

#### Archives serves various functions

- Maximum science exploitation of data
- Ensure long term preservation of data, software and knowledge
- Support Science Ground Segments from early phases

#### Scientists and Engineers working together

- Archives must be science driven
- Archives require strong IT expertise
- Regular technology evolution through time

#### Towards Archives 2.0

- Open data, open source, open archives
- Paradigm shift towards "bring your code to the data"

## Acknowledgements to all Archive stakeholders



And to people who played

a special role in the ESAC

#### Many thanks to all the ESAC Science Data Centre

Archives lifetime Pedro Osuna **ESDC ESDC ESDC** Inaki Ortiz • Leads Science Software Engineers Leads Ignacio Leon Science • Ground Instrument Segment Teams Jose Hernandez • Teams John Dowson Mission Mission Managers Consortium Archive Archive Science Scientists Users Project General Archives Scientists Public