

Analytics Concepts Developing Intelligent Systems at ESA

CERN Data Analytics Workshop 20/2/2014

Salim Ansari

Earth Observation Fleet



Present Missions



Proba-V



CroSat



SMOS



Swarm



GOCE



Envisat



ERS



Proba-1

Future Missions



Sentinel-1



Sentinel-2



Sentinel-3



Sentinel-4



Sentinel-5



Sentinel-5P

Earth Observation Fleet



Present Missions

-  Proba-V
-  CroSat
-  SMOS
-  Swarm
-  GOCE
-  Envisat
-  ERS
-  Proba-1

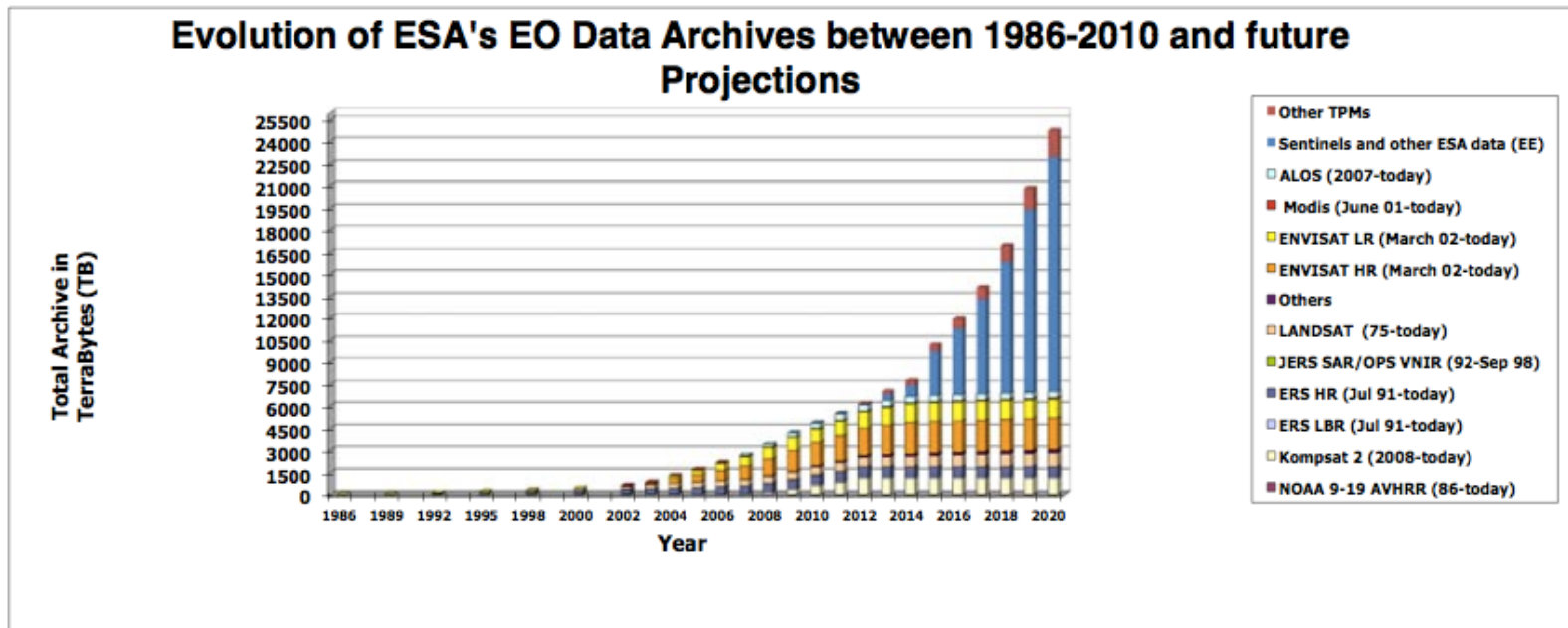
Challenges

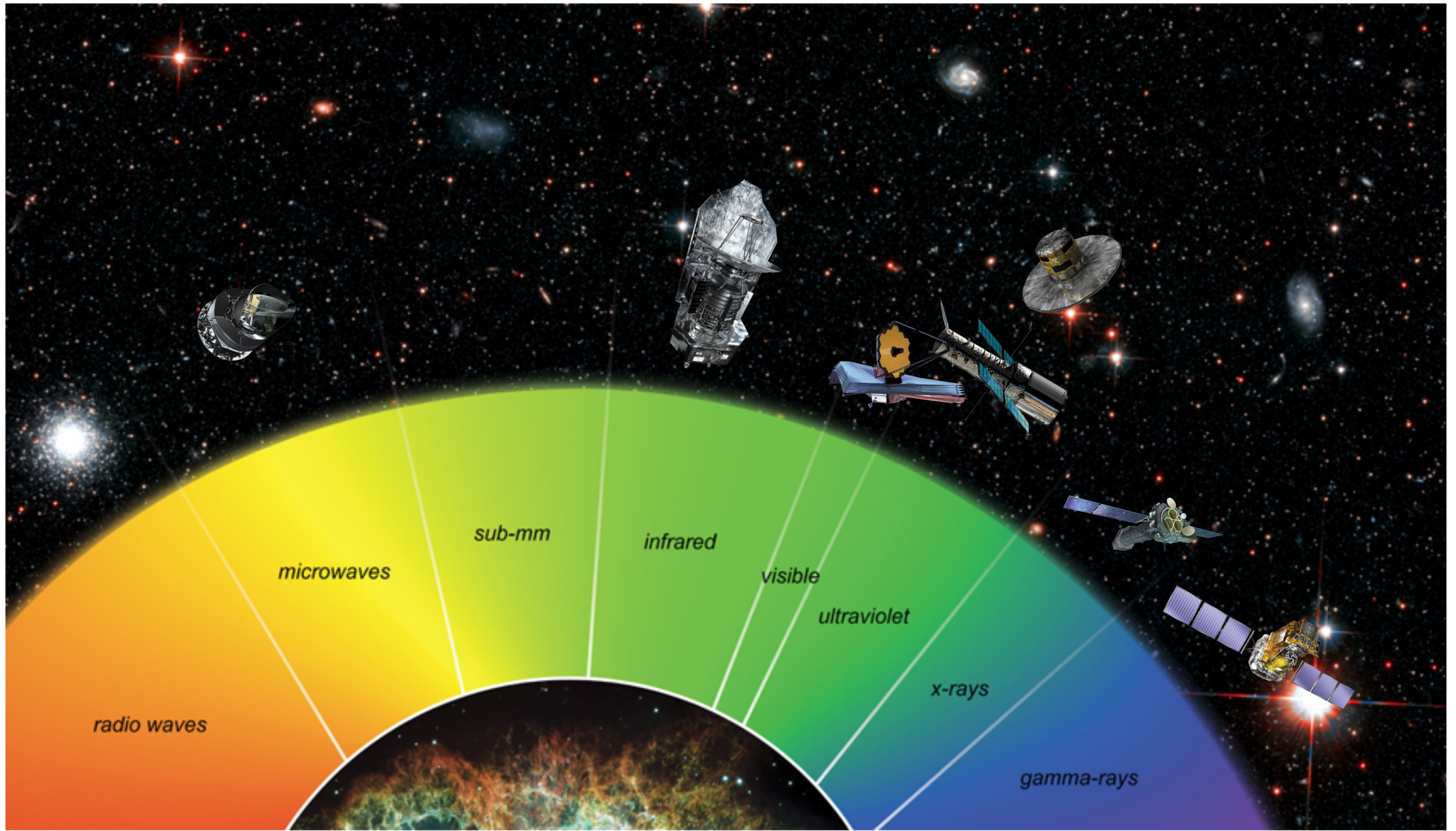
- Data
- Dissemination
- Data Combination
- Archiving

Future Missions

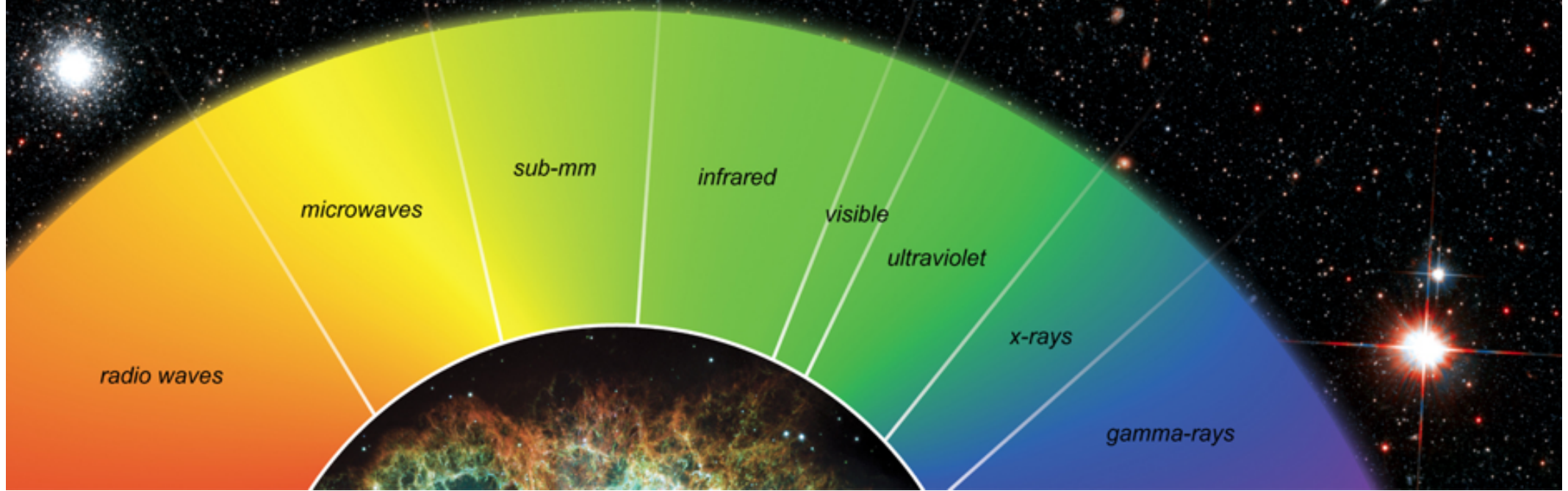
-  Sentinel-1
-  Sentinel-2
-  Sentinel-3
-  Sentinel-4
-  Sentinel-5
-  Sentine-5P

Earth Observation Data





Challenges
Data Processing
Data Dissemination
Multiple Domain Analysis
Archiving



1. Huge amounts of data are collected daily that go unnoticed.
2. Trends ranging from human perception in using information systems to research challenges could tremendously benefit from some structural design in analysing them
3. The confidence level of scientific results can increase exponentially if proper analytical tools are applied
4. The “live” approach to reporting on server/application performance can make a world’s difference, if we know right away where the problem lies
5. Optimizing our spacecraft operations by analysing orbital data can dramatically increase operational efficiency

I DON'T BELIEVE IN SURVEYS!!

I BELIEVE IN PREDICTIVE ANALYTICS

1. **ESA** has several directorates managing different programmes. From the Space Station to the Earth Observation fleet, to Astrophysics, there are different needs for different communities
2. The IT landscape may vary drastically from one programme to the next
3. SLA's are established between the IT Department and individual directorates
4. Require transparency and visibility of costs and costs tracking

The SLA Dashboard



iPad VPN
15:37
62%

Service Level Agreement

CAMERA

Directorate: **D/TEC**

Reporting date: **19/02/2014**

SLA Total Costs: **2,845,129**

SLA Year: **2014**

CITI

CIS

Service	Service Element	Costs
AMC V2		17
AMC PGMs		56
Application coordinator		162
ARTS for EOP-G		38
Based on 2012 actuals cost		221
D.F.		115
D.F.		189
DOORS for EOP-G		85
ECONOS DMS (Expenses)		225
EOP Network DMS		56
EOP Timesheet		40
Extension of PRISMA Maintenance and Support for D/TEC - CCN 25 - contact F. Fardouas		118
Galileo DMS		604
GMES DMS		73
IT Integrated support		162
NCE activities EOP part		162

Licenses

License ID	ESA Address Code	Server Name	LPU	Costs	Number of Copies per processor	Number of Processors
Total			264.3	17762.46		164
A001	EOP-GU	group-ferry.stg1.gps.esa.int	3.5	2352.70		1
A002	EOP-GU	mbmc-003	0.0	0.00		4
A003	EOP-GU	PDCO_BSP	3.5	2352.70		1
A004	EOP-GU	provida.pdf.pla.esm.esa.int	11.0	7394.20		1
0001	EOP-GU	REPORTEUR	3.5	2352.70		2
0002	EOP-GU	rgps0.pdf.pla.esm.esa.int	3.5	2352.70		1

Mobile Phone Costs

3417,74
7634,6
5695,87
6722,68
192,65
940,72
112,77
273,6
364,91
592,75
624,73
885,12
12579,58
16358,42

D/DG D/EIP D/EOP D/FCI D/HFM D/HSO D/IPL D/LAU D/NAV D/SRE D/TEC D/TIA DG

2013 2014 2015 2016

Other more challenging areas

1. Analytics can play a major role in simplifying scientific research!
2. Discovery tools can be automated thus freeing the scientist from having to carry out much of the search for knowledge to analytical tools and focus on the science
3. The data is there in huge volumes and in the Public domain!!
4. All it needs is some intelligent extraction protocol!



Examples of multiwavelength Astronomy



Today astronomers have to carry out very cumbersome searches through thousands of catalogues to find what they are looking for.

The screenshot shows the VizieR Catalog Selection Page on an iPad. The page title is "Catalog Selection Page" and it indicates "229 catalogs found having potential matches (142 really found)".

Search Criteria: Keywords: Crab

Preferences: max: 50, HTML Table, All columns, Compute, Mirrors: CDS, France

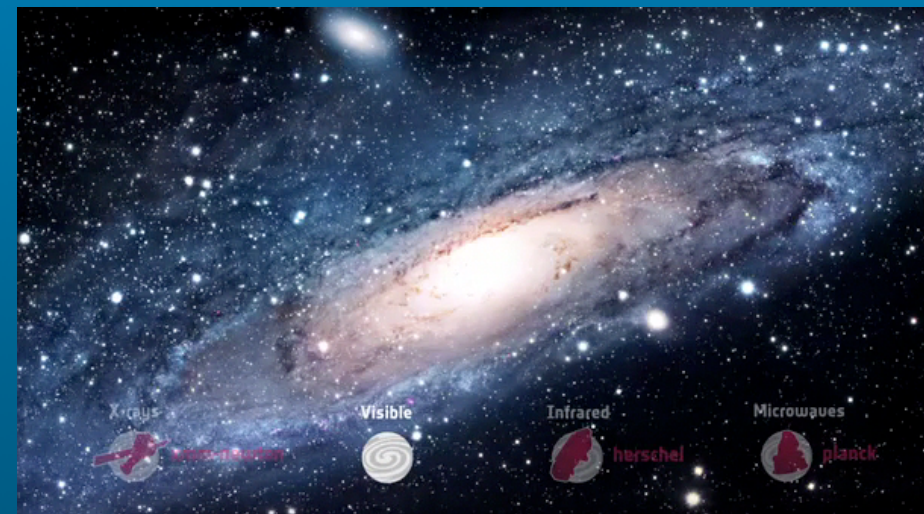
Actions: Reset All, Show table details, or Query selected Catalogs

Check	Code	Code	Code	Description	Similar Catalogs	ReadMe+ftp	Thumbnail
<input type="checkbox"/>	IR	I/319	(density 89) 140M	UKIDSS-DR9 LAS, GCS and DXS Surveys (Lawrence+ 2012)	2007MNRAS.379.1599L	ReadMe+ftp	
<input type="checkbox"/>	IR	I/314	(density 83) 122M	UKIDSS-DR8 LAS, GCS and DXS Surveys (Lawrence+ 2012)	2007MNRAS.379.1599L	ReadMe+ftp	
<input type="checkbox"/>	IR	I/297	(density 53) 1117M	NOMAD Catalog (Zacharias+ 2005)	2004AAS...205.4815Z	ReadMe+ftp	
<input type="checkbox"/>	IR	I/284	(density 51) 1045M	The USNO-B1.0 Catalog (Monet+ 2003)	2003AJ...125.984M	ReadMe+ftp	
<input type="checkbox"/>	IR	I/305	(density 49) 945M	The Guide Star Catalog, Version 2.3.2 (GSC2.3) (STScI, 2006)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/317	(density 48) 910M	The PPMXL Catalog (Roeser+ 2010)	2010AJ...139.2440R	ReadMe+ftp	
<input type="checkbox"/>	IR	I/311	(density 37) 563M	WISE All-Sky Data Release (Cutri+ 2012)	2012yCat.2311...0C	ReadMe+ftp	
<input type="checkbox"/>	IR	I/252	(density 36) 526M	The USNO-A2.0 Catalogue (Monet+ 1998)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/246	(density 34) 470M	2MASS All-Sky Catalog of Point Sources (Cutri+ 2003)		ReadMe+ftp	
<input type="checkbox"/>	IR	B/eso	(density 34) 6M	ESO Science Archive Catalog (ESO, 1991-2014) This archive log is limited to science observations (calibration and acquisition records are discarded)		ReadMe+ftp	
<input type="checkbox"/>	IR	B/iram	(density 28) 235k	IRAM Observation Logs (IRAM 1991-2014)		ReadMe+ftp	
<input type="checkbox"/>	IR	B/hst	(density 23) 466k	HST Archived Exposures Catalog (STScI, 2007)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/304	(density 18) 95M	Carlsberg Meridian Catalog 14 (CMC14) (CMC, 2006)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/322A	(density 17) 113M	UCAC4 Catalogue (Zacharias+, 2012)	2012yCat.1322...0Z	ReadMe+ftp	
<input type="checkbox"/>	IR	IX/28A	(density 15) 113k	ROSAT HRI Pointed Observations (IRXH) (ROSAT Team, 2000)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/225	(density 14) 539k	Catalog of Infrared Observations, Edition 5 (Gezari+ 1999)	1993scio.book.....G	ReadMe+ftp	
<input type="checkbox"/>	IR	VI/110	(density 13) 110k	Final Merged Log of IUE Observations (NASA-ESA, 2000)		ReadMe+ftp	
<input type="checkbox"/>	IR	B/cfht	(density 13) 631k	Log of CFHT Exposures (CADC, 1979-)		ReadMe+ftp	
<input type="checkbox"/>	IR	I/289	(density 12) 48M	UCAC2 Catalogue (Zacharias+ 2004) See also the more recent versions UCAC3 (I/315) and UCAC4 (I/322)	2004AJ...127.3043Z	ReadMe+ftp	
<input type="checkbox"/>	IR	IX/41	(density 12) 262k	XMM-Newton Serendipitous Source Catalogue 2XMMi-DR3 (XMM-SSC, 2010) Detailed description and explanations are available in the User Guide		ReadMe+ftp	

Examples of multiwavelength Astronomy



With the help of analytical methods, astronomers can choose the domains that can be automatically interrogated and results relevant to their research are regularly delivered



Dump 20,000 multiwavelength astronomical catalogues worth millions of data entries (astronomical objects) into a single Hadoop-like database.

1. Build multiwavelength queries of areas in the galaxy/universe that are of interest and extract the relevant records
2. Combine high resolution images from various multiwavelength sources to study the morphology and evolution of universal phenomena
3. Compare results to existing publications. Do they lead to a new discovery or are they simply serendipitous in nature?