#	Title	Authors	Abstract
001	The Gemini Science Archive: Current Status and Future Prospects	Colin Aspin, Gemini Observatory David Bohlender, CADC, HIA, Victoria	We will present details of the design, development and operation of the Gemini Science Archive, the GSA. The full basic archive was released to the public in July 2004 and has many inovative features. We will additional discuss the Phase III development of the GSA including many advanced capabilities which will allow the GSA to publish extensive VO-enabled data products.
002	VOTFilter, Bridging VO Services to Industrial Desktop Application	Chenzhou CUI, National Astronomical Observatory of China Kui WU, Beijing Institute of Astronautical System Engineering	VOTable is an XML format defined for the exchange of tabular data in the context of the Virtual Observatory. It is the first Proposed Recommendation defined by International Virtual Observatory Alliance, and has obtained widely support from both VO communities and many astrophysics projects. Taking advantage of XML file format of OpenOffice, and XML filter VOTFilter is developed by the Chinese Virtual Observatory project. Using the VOTFilter, one can read and write files in VOTable format, edit and analyze astronomical data using OpenOffice Calc. In the paper, we will describe the importance of the toolkit, introduce its usage and discuss possible extension for the current version. Some technical problems met during its development will be also discussed.
003	GaiaGrid : Its Implication and Implementation	S. G. Ansari, Directorate of Science, European Space Agency M.A.C. Perryman, Directorate of Science, European Space Agency M. ter Linden, DutchSpace	Gaia is an ESA space mission to determine positions of 1 billion objects in the Galaxy at micro-arcsecond precision. The data analysis and processing requirements of the mission involves about 20 institutes across Europe, each providing specific algorithms for specific tasks, which range from relativistic effects on positional determination, classification, astrometric binary star detection, photometric analysis, spectroscopic analysis etc. In

			an initial phase, a study has been ongoing over the past 3 years to determine the complexity of Gaia's data processing. Two processing categories have materialised: core and shell. While core deals with routine data processing, shell tasks are algorithms to carry out data analysis, which involves the Gaia Community at large. For this latter category, we are currently experimenting with use of Grid paradigms to allow access to the core data and to augment processing power to simulate and analyse the data in preparation for the actual mission. We present preliminary results and discuss the sociological impact of distributing the tasks amongst the community.
005	Hardware Acceleration for Astronomical Data Analysis	Eric Sessoms, NRAO	The floating-point performance of commercial Graphics Processing Units (GPUs) currently exceeds that of more costly CPUs by a factor of 3 and the parallel architecture of GPUs is so well suited to many physical problems that researchers have reported practical performance gains of 6- to 10-times with some algorithms. As GPU performance is on an exponential growth curve doubling every 6-months, instead of doubling every 18-months as main line CPUs, GPUs are an increasingly cost-effective computational resource. We present an overview of GPU architecture and discuss how that architecture makes these performance gains possible. We then explore the programming model for GPUs in greater detail, introduce programming frameworks that have been developed to take advantage of GPU performance, and identify a number of data processing packages already available for GPUs.
012	Global covariance analysis of the SIM	Global covariance analysis of the SIM astrometric grid	The all-sky, high accuracy astrometric grid of ~1300 red giants will provide the basis for astrometric analysis of target stars and extragalactic objects observed with the Space Interferometry Mission (SIM)

	astrometric grid		facility. Due to the differential character of interferometric measurements with SIM, the grid data will be correlated globally, the more strongly the closer the stars on the sky. Those correlations will have far-reaching significance for astrophysical projects using ensembles of stars, e.g., determination of the distance to the LMC, and the rotation curve of the Galaxy. The ongoing grid simulations with the newly developed
014	SAADA: Astronomical Databases Made Easy	L. Michel, H. Nguyen Ngoc & C. Motch	Many astronomers wish to share datasets with their community but have not enough manpower to develop databases having the functionalities required for high-level scientific applications. The SAADA*) project aims at automatizing the creation and deployment process of such databases. A generic but scientifically relevant data model has been designed which allows to build databases by only providing a limited number of product mapping rules. Databases created by SAADA rely on a relational database supporting JDBC and covered by a Java layer including a lot of generated code. Such databases can simultaneously host spectra, images, source lists and plots. Data are grouped in user defined collections whose content can be seen as one unique set per data type even if their formats differ. Data sets can be correlated one with each other using qualified links. These links help, for example, to handle the nature of a cross-identification (e.g., a distance or a likelihood) or to describe their scientific content (e.g., by associating a spectrum to a catalogue entry). The SAADA query engine is based on a language well suited to the data model which can handle constraints on linked data, in addition to classical astronomical queries. These constraints can be applied on the linked objects (number, class and attributes) and/or on the link qualifier values. Databases created by

			SAADA are accessed through a rich WEB interface or a Java API. We are currently developing an inter-operability module implanting VO protocols. *) French acronym for Auto-configurable Archiving System for Astronomical Databases.
015	Software for detection of Optical Transients in observations with Rapid Wide-Field Camera	Anton Biryukov, SAI, SAO. Gregory Beskin, SAO. Sergey Bondar, RIPI. Kevin Hurley, SSL. Eugeny Ivanov, RIPI. Sergey Karpov, SAO. Elena Katkova, RIPI. Alexei Pozanenko, IKI. Ivan Zolotukhin, SAI, SAO	A software complex for detection and investigation of optical transients (OTs) in observations with rapid wide-field camera has been developed. Data stream is a sample of digitized TV-CCD frames with size of 1360x1024 pixels, exposure time of 0.13 sec and frame frequency of 7.5 Hz. Software performs the following tasks: real time data transfer to the LAN; accumulation of initial data (total volume up to 0.5 Tb per night); real time data reduction - detection and classification of OTs (stationary and moving), determination of their equatorial coordinates, magnitudes and parameters of trajectory, their possible identification with known objects and transmission of information about OTs (Alerts) to the local and global networks. It takes 0.4 seconds (3 successive frames) to conclude about detection of OT and to calculate of detected object parameters. OT detection algorithm is based on the pixel-to-pixel comparison of current frame with frame averaged over 10-100 previous ones. All observed objects initially considered as motional. It allow to separate really motion and stationary transients in the field. For each frame the Awaiting Area (AA) is found for every observed Object on basis of its trajectory parameters. In this area algorithm suppose to detect appropriate object on current frame. The Object is considered as stationary (real) OT, if after 2 frames, its position of first detection lies inside current AA.

			The faintest detectable object has 11.5 magnitude (close to V-band). During 150 nigths of observations from autumn 2003 till summer 2004 in average 100 satellities and 200 meteors were registered per night.
016	MIGALE: milstones and roadmap	Igor Chilingarian, Sternberg Astronomical Inst. of Moscow State U. Philippe Prugniel, CRAL Observatoire de Lyon Hector Flores, GEPI Observatoire de Paris Jean Guibert, GEPI Observatoire de Paris Regis Haigron, GEPI Observatoire de Paris Isabelle Jegouzo, GEPI Observatoire de Paris Frederic Royer, GEPI Observatoire de Paris Francoise Tajahmady, GEPI Observatoire de Paris Gilles Theureau, GEPI Observatoire de Paris Jacques Vetois, GEPI Observatoire de Paris	MIGALE (Multiparametric Virtual Instrument to Study the Evolution of Galaxies) is a collaboration between two laboratories in France: CRAL (Lyon) and GEPI (Paris), Sofia observatory (Bulgaria) and Moscow State University (Russia). The project is now supported by Programme National Galaxies (France) and two national virtual observatories, VO-France and RVO. The observational projects on two large telescopes (VLT and Russian 6-m) proposed by MIGALE are being carried out. The best known resource provided and maintained by MIGALE is HyperLeda - the database containing homogenized information for millions of galaxies. MIGALE distributes the Pleinpot open source software platform that can be used to easily build astronomical databases and online tools and has vocation to conform to the VO standards in the nearest future. PostgreSQL is used as DBMS. In 2004 three large Pleinpot powered databases related to the studies of the evolution of galaxies were opened to the public: *GIRAFFE archive, http://dbgiraf.obspm.fr/ArchiveGiraffe/ *HiGi, http://klun.obs-nancay.fr/ *ASPID, http://alcor.sao.ru/db/aspid/ We built the prototypes of the tools that would constitute the proposed virtual instrumentation in 2005-2007: *DisGal3D to make deconvolution of 3D

			spectral data *PEGASE.HR to build high resolution synthetic spectra of galaxies *SPIKeR to determine the internal kinematics and stellar content of galaxies from integrated light spectra We plan to release three databases supposed to become major sources of scientific data for galaxies: *DisGal: 3D data for distant galaxies *Fabry-Perot: 3D data from Fabry-Perot interferometers all over the world *ASPID-GE: reduced 3D data from MPFS (6m telescope)
017	Optical Camera with high temporal resolution to search for transients in the wide field	Ivan Zolotukhin, Sternberg Astronomical Institute, Moscow, Russia Gregory Beskin, Special Astrophysical Observatory, Nizhnij Arkhyz, Karachai-Cherkessia, Russia Anton Biryukov, Sternberg Astronomical Institute, Moscow, Russia Sergey Bondar, Research Institute for Precision Instruments, Moscow, Russia Kevin Hurley, Space Science Laboratory, University of California, Berkeley Evgeny Ivanov, Research Institute for Precision Instruments, Moscow, Russia Sergey Karpov, Special Astrophysical Observatory, Nizhnij Arkhyz, Karachai-Cherkessia, Russia Elena Katkova, Research	One of the poorly understood aspects of gamma-ray bursts (GRBs) is prompt optical emission. For its successful registration one have to observe independently from space borne gamma-ray telescopes and use optical instruments with wide field of view. To realize this approach wide field optical camera with high temporal resolution has been developed. Main objective (15 cm diameter, F/1.2) of the camera projects 21x16 degrees area onto image intensifier photocathode (gain - 150, scaling factor - 0.22). Special optics transfers image from intesifier's output to the TV-CCD camera with frame frequency of 7.5 Hz (0.13 sec exposure time). Observational data (17 Mb/sec transfer rate) is transmitted to the local PC which broadcasts it through the LAN to the storage computer equipped with RAID and to the PC for real-time processing. The chosen network configuration allows to keep raw data obtained during 8 hours observing set with 0.13 s resolution (volume up to 0.5 Tb) and detect on the fly both stationary and moving OTs up to V < 11.5. System provides classification of detected objects, magnitude

		Institute for Precision Instruments, Moscow, Russia Alexey Pozanenko, Space Research Institute, Moscow, Russia	and coordinates with 35 accuracy in 0.4 sec (3 frames). The camera operated in test mode from May to November, 2003; since November 2003 the camera monitors HETE-2 WXM field of view. Expected numbers of GRBs hitting in field of view of the camera which will be registered simultaneously with WXM (HETE-2) and BAT (SWIFT) are 1-2 and 4 per year, correspondingly.
020	Gbit/s I/O at the European VLBI dataprocessor - clustered silicon and long fiber	Harro Verkouter, JIVE Huib Jan van Langevelde, JIVE Friso Olnon, JIVE Arpad Szomoru, JIVE Steve Parsley, JIVE Mike Garrett, JIVE	In the European VLBI Network, a number of high-data-rate radio-astronomy R&D projects are taking shape. We deal with Gbits/s on both on the input-side and the output side of the EVN MkIV VLBI correlator at JIVE. The EVN MkIV correlator is capable of handling 16Gbit/s input datarate and a maximum output datarate of 1.28Gbit/s (160MByte/s). At the inputside of the correlator various developments are well underway. Hard disk recorded VLBI observations are supported. Another project deals with enabling real-time imaging VLBI, which was demonstrated to work on April 28th 2004. Work is in progress to enable capturing the full output rate of the MkIV correlator. This will yield high time/frequency resolution data, which will enable more scientific output from VLBI datasets as the field-of-view will be enhanced by a factor of 100 or more. In order to match the expected datarate and volume, a combined storage/compute cluster was installed. The hardware is tuned for processing large volumes in a streaming fashion: InfiniBand is used as cluster interconnect. Research is going on in the field of distributed

			processing. New algorithms will be developed to enable parallel processing of VLBI data. This effort is done in conjunction with other radio-astronomical institutes in an EU funded effort. An archive with access methods like the Astrophysical Virtual Observatory is developed when we are able to generate fine-tuned (in time/frequency resolution and/or desired field-of-view) datasets from the very large raw datasets. Raw datasets typically will be on the order of 4TB for a canonical VLBI experiment.
021	Harmonic development of an arbitrary function of the Moon/Sun/planets coordinates to Poisson series	Sergey Kudryavtsev, Sternberg Astronomical Institute of Moscow State University	A new algorithm of spectral analysis of an arbitrary function of the Moon/Sun/planets coordinates tabulated over a long period of time is proposed. Expansion of the function to a Poisson series is directly made where the amplitudes and arguments of the series' terms are high-degree time polynomials (as opposed to the classical Fourier analysis where the terms' amplitudes are constants and the arguments are linear functions of time). This approach leads to an essential improvement in accuracy of the harmonic development of the tabulated function over a long-term interval and reduction of the approximating series' lenght. To test the algorithm, we calculated the Earth-Moon distance on every day within [1000BC-5000AD] by using the ELP2000-85 analytical theory of lunar motion and then made the spectral analysis of the tabulated values with help of the new algorithm. All coefficients of the Poisson series composing the ELP2000-85 theory have been successfully found; the maximum deviation of the lunar distance values calculated by our series from those given by the exact ELP2000-85 model does not exceed 1.5 centimeters over the whole interval of six thousand years.

			The work is supported in part by grant 02-02-16887 from the Russian Foundation for Basic Research.
022	An $\mbox{\cal O}(N \log M)$ Algorithm for Catalogue Matching	Drew Devereux David J. Abel Robert A. Power Peter R. Lamb	A basic problem in data fusion in Virtual Observatories is that of how to determine which source records from different catalogues actually refer to the same source, on the basis of spatial co-location. This is the {\it catalogue matching} problem, and it is inherently costly to solve. Our algorithm applies filter-refine and plane sweep techniques. Pre-processing consists of a sort by declination, and the active list is a queue indexed by a binary tree. The algorithm is $\mbox{\cal O}(N \log M)$ in both I/O and processor costs, with only moderate memory requirements. Empirical assessment on catalogues of up to a billion records suggests that the algorithm performs at least an order of magnitude better than the techniques in current use.
023	Distributed Data Mining for Astrophysical Data	Sabine McConnell David Skillicorn	We present a data mining technique in which the datasets are partitioned by attributes such that datasets contain subsets of information for all available objects . Our approach is evaluated on a variety of astrophysical datasets and is suitable for a distributed setting, where the introduction of real time data analysis imposes an upper bound on the size of the data that can be processed. Using a decision tree technique, we show that this ensemble technique can achieve the same or better classification accuracy compared to an approach where the complete dataset is contained in a single location. At the same time, this substantially increases the amount of data that can be processed.
024	Hires: Super-resolution	Charles Backus, JPL Thangasamy Velusamy, JPL Tim Thompson, JPL	We present a fast deconvolution program based upon the R-L algorithm and its generalization to the case of redundant sky coverage addressed by Aumann, Fowler,

	for the Spitzer Infrared Telescope	John Arballo, JPL	and Melnyk. The program makes extensive use of FFT's, resamples using a drizzle method, and accounts for projection and distortion effects that are important in Spitzer data. We address some of the tactical issues involved, such as spatial frequency response, background removal and ringing near sharp transitions, and the importance of accurate psf images. We conclude with a few examples of processed Spitzer observations.
025	he EVLA Software System Overall Design	Tom Morgan, NRAO, Socorro, NM Kevin Ryan, NRAO, Socorro, NM Ken Sowinski, NRAO, Socorro, NM Boyd Waters, NRAO, Socorro, NM	We review the overall design of the software system being constructed to operate the Expanded Very Large Array (EVLA). The scope of the system spans proposal preparation and submission to the NRAO through archiving of final output data products in forms suitable for export to data reduction packages and for data mining. Many VLA functions currently being performed by hand will be replaced by automated subsystems. Other software-based procedures as well as monitor and control systems and data archives are being redesigned to handle enhanced correlator configurations and higher output data volumes, and high bandwidth receiver, sampler and data transfer technologies being incorporated into the antennas. Each of the major software subsystems will be addressed with respect to inputs, outputs, functionality and connectivity to data bases and other system components.
028	Fast Algorithms for Massive-Scale Classification Problems: Toward 1	Alexander Gray, Carnegie Mellon University Gordon Richards, Princeton University Robert C. Nichol, University of Portsmouth Robert Brunner, NCSA/University of Illinois Urbana-Champagne	Quasar detection from SDSS data is an example of a classification problem which must simultaneously achieve as high accuracy as possible while somehow retaining computational tractability in the face of datasets with up to millions of objects in three or more measurements. We present a new algorithm for the highly-accurate Nonparametric Bayes Classifier (the Bayes-optimal classifier for arbitrary

	Million Quasars		distributions), which makes it computable for such massive datasets. We have used it to achieve the largest quasar catalog to date, and believe this fundamental statistical/computational technology can enable similar large-scale astronomical data mining projects.
029	VOSpec: A tool for handling Virtual Observatory compliant Spectra	Pedro OSUNA ESAC/ESA Christophe ARVISET ESAC/ESA Isa BARBARISI ESAC/ESA Jesus SALGADO ESAC/ESA	The European Space Agency (ESA) has been heavily involved in the development of the Simple Spectral Access Protocol (SSAP) of the International Virtual Observatory Alliance (IVOA) by providing access to the Infrared Space Observatory (ISO) mission spectral data. The European Space Astronomy Centre (ESAC) at Villafranca del Castillo near Madrid, Spain, will be the centre for Space Based Virtual Observatory activities within the European Space Agency. The VOSpec tool is part of the already started VO activities of ESA. There are lots of tools for spectra analysis and display in the astronomical community. However, there is a need for a "VO-enabled" tool that can superimpose spectra coming from different projects within the VO and in VO-format. One of the main reasons for the current absence of such a tool is the fact that astronomical spectra are not as well defined as, e.g., images. The IVOA has made a significant effort in trying to define a Simple Spectral Access Protocol by which part of this problem is solved, giving the rules to be able to handle spectral data in the VO context. By defining an algorithm based on dimensional analysis for the superimposition of spectra coming from VO resources, ESAC has been able to create a tool that

			can superimpose VO spectra that declare their metadata in a specific way. To do this, VO SSAP-compatible resources need only give three new extra fields in their VOTable SSAP. At the time of writing this abstract, ESAC is planning to propose these three extra metadata fields for acceptance at the IVOA as optional inputs in the definition of the Simple Spectral Access Protocol.
031	Reliable, Automatic Transfer and Processing of Large Scale Astronomy Datasets	Tevfik Kosar, University of Wisconsin-Madison George Kola, University of Wisconsin-Madison Robert Brunner, NCSA Miron Livny, University of Wisconsin-Madison Michael Remijan, NCSA	Astronomers are increasingly obtaining larger datasets, particularly in the optical and near-infrared. Unfortunately, the technologies to process large amounts of image data and share the data and the results with collaborators spread around the globe, have not kept pace with the data flow. In the past, this type of software has also required significant human involvement to deal with failures. We have designed and implemented a fault-tolerant system that can process large amounts of astronomy images using idle CPUs on desktops, commodity clusters and grid resources. It reliably replicates data and results to collaborating sites and performs on-the-fly optimization to improve throughput. It is highly resilient to failures and can recover automatically from network, storage server, software and hardware failures. To demonstrate the capabilities of this framework, we have successfully processed three terabytes of DPOSS images using idle grid resources spread across three organizations. This system is freely available to others groups who wish to significantly decrease the time and effort required to perform large scale data transfer and processing.
032	Linearization of Spitzer Infrared	John Fowler, California Institute of Technology	The Spitzer Infrared Spectrometer data are taken via read-without-reset measurements to obtain multiple samples forming a photometric "ramp" for each pixel in an echellogram. Each ramp is linearized via a

	Spectrometer Data Via Minimization of Chi-Square With Correlated Errors		quadratic model. After linearization, a quality-assurance test is performed to determine how linear each pixel's ramp has become. This is accomplished by fitting a straight line to the ramp via chi-square minimization. The goodness of fit is of primary importance, since this determines whether the inevitable deviations from linearity are statistically significant given the estimated photometric noise. Because the latter is dominated by photon noise which is summed up the ramp, the chi-square parameter used to measure goodness of fit must include the effects of correlated errors. This paper describes the construction of the full error covariance matrix and its use in the chi-square minimization.
034	VWhere: An Extensible Filtering and Data Exploration Guilet	Michael S. Noble, Massachusetts Institute of Technology	In this paper we describe VWhere, a S-Lang/Gtk guilet which combines graphical ease of use with powerful data exploration and filtering. Analysis in VWhere amounts to manipulating sets of region filters applied to one or more plots. Input vectors originate from either files or in-memory arrays, with plots and regions created via mouse. A text window is also provided, in which additional vectors may be fabricated from arbitrary mathematical expressions. A list of points passing all filters is output, which can be applied to other datasets, create output files, etcetera. In contrast with the (requested update of this truncated abstract)
036	A Java Thick Client User Interface for Grid Processing	Ted Hesselroth, Spitzer Science Center	A user interface (CAPRI) which is configurable at runtime has been developed which allows application features to be maintained and upgraded on a central server, available to users without the need for reinstalling software. The user interface is specified by an XML file

	.		accessed through a URL and parsed by the open-source SWIX library, which returns a completely laid-out container with the application’s controls. A set of generic model-view-controller-actions classes are also instantiated based on parsing of the input XML file. Hierarchical relationships present in the XML file are reflected in membership relationships among the classes. An event-driven architecture with a central event handler allows for convenient extensibility. Client/server software is based on the Java Web Services package. SAAJ message passing based on SOAP is the transaction medium between the client and server, allowing input and output files to be exchanged. The server has access to data and computing resources and brokers the requested computation. Secure shell is supported as an option for transfers between the server and computing grid. Sun Grid Engine software is used to manage the cluster of processing nodes, which can be diskless workstations booted via PXE from the grid engine manager system. This application has been deployed at the Spitzer Science Center to allow rapid interactive processing of science data.
039	VISTA Data Flow System: Pipeline Processing for WFCAM and VISTA	James R. Lewis, Cambridge Astronomy Survey Unit M.J. Irwin, Cambridge Astronomy Survey Unit S.T. Hodgkin, Cambridge Astronomy Survey Unit Peter S. Bunclark, Cambridge Astronomy Survey Unit	The UKIRT Wide Field Camera (WFCAM) on Mauna Kea and the VISTA IR mosaic camera at ESO, Paranal, with respectively 4 Rockwell 2k x 2k and 16 Raytheon 2k x 2k IR arrays on 4m-class telescopes, represent an enormous leap in deep IR survey capability. With combined nightly data-rates of typically 1TB, automated pipeline processing and data management requirements are paramount. Pipeline processing of IR

		D.W. Evans, Cambridge Astronomy Survey Unit R.G. McMahon, Cambridge Astronomy Survey Unit	data is far more technically challenging than for optical data. IR detectors are inherently more unstable, while the sky emission is over 100 times brighter than most objects of interest, and varies in a complex spatial and temporal manner. In this presentation we describe the pipeline architecture being developed to deal with the IR imaging data from WFCAM and VISTA, and discuss the primary issues involved in an end-to-end system capable of: robustly removing instrument and night sky signatures; monitoring data quality and system integrity; providing astrometric and photometric calibration; and generating photon noise-limited images and astronomical catalogues.
040	Christophe.Arviset@esa.int	Christophe ARVISET, ESA / ESAC Isa BARBARISI, ESA / ESAC John DOWSON, ESA / ESAC Jose HERNANDEZ, ESA / ESAC Iñaki ORTIZ, ESA / ESAC Pedro OSUNA, ESA / ESAC Jesús SALGADO, ESA / ESAC Guillermo SAN MIGUEL, ESA / ESAC Aurèle VENET, ESA / ESAC	In June 2004, ESA decided to rename the VILSPA centre into ESAC for European Space Astronomy Centre. All astronomy missions (ISO, XMM-Newton, Integral, Herschel, Planck, Gaia, etc..) will be located at ESAC. ESAC will also host most of the ESA astronomy and planetary Scientific Archives, becoming the main space science data provider in Europe. ESAC Archive Group had already been involved in some VO activities. The ISO Data Archive and the XMM-Newton Science Archive have already been VO-enabled via the SIAP (Simple Image Access Protocol). Furthermore, the IDA has been the first archive to be SSA (Simple Spectra Access) compliant. All these features were demonstrated at the AVO first science demonstration in January 2004. Thanks to these successes, ESA has decided to have its own ESA-VO project to become the European Space based astronomy VO actor. Dedicated manpower and computer resources have been allocated to the ESA-VO project to participate more actively in all IVOA initiatives.

			ESA-VO main axes of development are the following: - Keep on making all ESA astronomy archives VO compliant (SIAP, SSA, VOQL. …) - Build an ESA-VO Portal including a multi-missions search interface based on VO protocols - Inter-operability from the existing ESAC archives to remote archives via VO protocols - Development of specific VO applications such as VOSpec, a tool for retrieval and overlay display of Spectra from SSAP VO-enabled archives, including various unit conversion as per metadata specified in the SSA - Building a ESA VO registry - Building a ESA VO Grid
042	A Framework for Parallel Data Analysis on a Distributed Grid	Jeffrey P. Gardner, Pittsburgh Supercomputing Center Andrew Connolly, University of Pittsburgh	Virtual observatories will give astronomers easy access to an unprecedented amount of data. In many cases, mining these data will require the power of parallel computers. These machines may be a small Beowulf cluster, a large massively parallel platform, or a collection of parallel machines distributed across a Grid. Harnessing the power of these machines can be difficult, since parallel programs often demand a great deal of time and expertise to develop. However, nearly all analysis of large astronomical datasets use a limited number of data structures: trees and grids. Our framework will provide a flexible, extensible, and easy-to-use way of using these data structures for data analysis on parallel and grid-distributed machines. By minimizing development time for these platforms, we will enable wide-scale knowledge discovery on massive datasets.

045	XMM-Newton: Approaching 5 years of succesful science operations	Carlos GABRIEL Matteo GUAINAZZI Leo METCALFE	The major ESA X-ray observatory, XMM-Newton, is approaching 5 years of operations in flight. The design concepts driving the activities of the Science Operations Centre at the European Space Astronomy Centre in Spain have shown their validity for this astrophysics long-term mission. Scientific exploitation of data by the astronomical community is facilitated especially through a system resting on 4 pillars: - scientific data dissemination: from raw telemetry up to processed and calibrated high-level science products, such as images, spectra, source lists, etc; - dedicated science analysis software: development and distribution of mission specific software, as well as of continuously updated instrument calibration; - scientific archive: access to data but also to high level information on data contents through state-of-the-art, in-house developed archival facilities; - documentation: continuously updated documentation on all aspects of spacecraft and instrument operations, data reduction and analysis, with particular focus on the most important scientific results obtained by XMM-Newton, reachable through a comprehensive set of project web pages. We intend to review all these and related aspects forming the basis of a modern astronomical observatory. Our aim is to show how the innovative but solid elements in the approach to each the mentioned points contribute to forefront science in the light of some of the outstanding scientific results achieved by XMM-Newton.
046	The ALMA	Jeff Kern, NRAO	The Atacama Large Millimeter Array (ALMA) requires

	Real Time Monitor and Control Bus	Rodrigo Amestica, NRAO	high precision real time control in a distributed system. To meet the unique requirements of ALMA a special purpose monitor and control bus, based on the industry standard CAN bus, has been developed. We will discuss the design and performance of the ALMA Monitor & Control Bus as well as it's software implementation in a Linux based real-time operating system.
047	W projection : a new algorithm for wide field imaging with radio synthesis arrays	Tim Cornwell, NRAO Kumar Golap, NRAO Sanjay Bhatnagar, NRAO	Wide field imaging with low frequency synthesis arrays is limited by a number of troublesome effects. First amongst these is the "non-coplanar baselines" distortion whereby the integral relationship between sky brightness and measured visibility function is not a simple Fourier transform. A piece wise approximation to the integrals can be used and forms the basis of the facet approaches used for the last 15 years. These approaches are difficult to program and perform relatively poorly. We have developed a novel, high performance algorithm based upon convolution of the visibility samples with a Fresnel kernel. We interpret the Fresnel kernel as being required to propagate the electric field to a common reference plane. The role of Fresnel diffraction in radio inteferometry seems to have been unrecognized previously.
049	Overview of the Gemini Science Archive	Severin Gaudet, Canadian Astronomy Data Centre Adrian Damian, CADC Norm Hill, CADC David Bohlender, CADC Sharon Goliath, CADC Geoffrey Melnychuk, CADC Colin Aspin, GEMIN	Gemini Science Archive (GSA) is a new science archive developed by the Canadian Astronomy Data Center (CADC) to provide the scientific community with tools for effective on-line access to data collected by the Gemini telescopes. The first release of the GSA basic archive (aka Phase 1) is scheduled for early fall 2004, with other subsequent releases containing advanced capabilities to follow after. Benefiting from the knowledge and experience of other telescope archives built and operated by CADC, GSA was

			also designed to be one of the first scientifically effective archive of ground-based operations. As such, from the very beginning GSA was considered an integrated part of the planning, observation, calibration, data reduction and data distribution process that occur at Gemini. This fundamental shift in data archiving, allowed for the implementation of some novel approaches to the data handling and processing: 1. support for data produced directly by telescopes' instruments as well as other related data (weather, logs etc) 2. database schema to separately store meta-data and catalogs 3. easy to edit data dictionary to specify data formats, ranges, transformations etc 4. rapid availability of data due to electronic transfer of data from the Gemini telescopes 5. generation of derived data for easy access by the user or for compatibility with other telescopes The paper addresses all these aspects of GSA design and also lays down new features planed for future releases.
051	Creating Data that Never Die: Building a Spectrograph Data Pipeline in the Virtual Observatory Era	D.J. Mink, SAO W.F. Wyatt, SAO J.B. Roll, SAO S.P. Tokarz, SAO M.A. Conroy, SAO N. Caldwell, SAO, M.J. Kurtz, SAO, M.J. Geller, SAO	Data pipelines for modern complex astronomical instruments do not begin when the data is taken and end when it is delivered to the user. Information must flow between the observatory and the observer from the time a project is conceived and between the observatory and the world well past the time when the original observers have extracted all the information they want from the data. For the 300-fiber Hectospec low dispersion spectrograph on the MMT, the SAO Telescope Data Center is constructing a data pipeline which provides assistance from preparing and

			submitting observing proposals through observation, reduction, and analysis to publication and an afterlife in the Virtual Observatory. We will describe our semi-automatic pipeline and how it has evolved over the first nine months of operation.
052	Deploying the AstroGrid: Science Use Ready	Nicholas A Walton, Institute of Astronomy, University of Cambridge, UK Paul Harrison, Jodrell Bank Observatory, Victoria University of Manchester, UK Martin Hill, Institute for Astronomy, University of Edinburgh, UK Anita Richards, Jodrell Bank Observatory, Victoria University of Manchester, UK	AstroGrid (see http://www.astrogrid.org), a UK eScience project with collaborating groups drawn from the major UK data archive centres, is creating the UK's first virtual observatory. We describe the current state of the AstroGrid testbed deployment system. Related presentations at this meeting cover AstroGrid and the international context and technical details of the software components of AstroGrid. In order to ensure rapid user feedback AstroGrid is fully deploying each iteration release, connected to relevant data and application products, in such a fashion as to allow scientific use of that release. The early users are primarily the science advisory group and now the beta tester communities of the project. The scientific functionality of the current deployment is highlighted. This includes access to a sophisticated workflow capability. Its use in allowing image extraction from multiple image datasets, input photometry file creation, redshift determination, and visualisation for the outputs to allow discovery of high redshift objects is described. Further examples show the use of the latest Astrophysical Data Query language standard and how it is being used to enable large data queries of remote databases in searching for low mass objects in the Pleiades with results returned to a virtual 'MySpace' user storage area, where further visualisation and processing can be performed.

			We note how the deployed system is being tested by the science community, and how that comment and feedback is vital in informing the project as to future releases including the December 2004 'Release 1'. The latest AstroGrid deployment release is available from http://www.astrogrid.org/release.
055	Simulation of the Future LSST Data Pipeline	Ghaleb Abdulla, David Liu, Jim Garlick, Sergei Nikolaev, Marcus Miller, Michael Franklin, Kem Cook, Jim Brase	In this paper we describe our approach to build a pipeline simulator for the future Large –scale Synoptic Survey Telescope (LSST). The simulated pipeline will be used to research and evaluate software architectures that are efficient and flexible. It will also be used to define the real-time software and hardware requirements needed to support the anticipated LSST data rates. The LSST data pipeline requirements are still being defined, however, previous surveys can provide a good source for data requirements. Our approach is to use the SuperMacho data pipeline as a prototyping tool to identify a framework for building Modular Data-Centric Pipeline (MDCP) architectures. The prototyping is done in a hierarchical fashion to help capture and define the general data attributes (schema) first. We then model other necessary components based on science and performance requirements. We use identified schemas or data attributes as a way to define a data model for LSST. !
056	Integrating Legacy Code Into Virtual Observatories - A	Joerg M. Colberg, University of Pittsburgh Ryan Scranton, University of Pittsburgh Andrew J. Connolly, University of Pittsburgh	As virtual observatories are being built, the integration of existing astronomical analysis codes is becoming increasingly important. This is particularly true for standard applications that are widely and often used. In these cases, creating what amounts to on-site data-mining facilities is clearly beneficial

	Test Case		for a wide variety of users, including, but not limited to, users with very little current knowledge of VO technologies or users interested in using the machinery for educational purposes. Given the nature of most legacy codes, there is a number of problems that have to solved. Here, we present an application that makes a very fast and efficient n-point correlation function code interact with the Sloan Digital Sky Survey. We discuss the basic ideas and setup, technical problems encountered, and possible extensions.
057	Mosaicking with MOPEX	David Makovoz, Caltech Iffat Khan, Caltech	We present MOPEX - a software package for image mosaicking and point source extraction. MOPEX has been developed for the SPITZER Space Telescope. This presentation concentrates on the mosaicking aspects of the package. MOPEX features the use of several interpolation techniques, coaddition schemes, and robust and flexible outlier detection based on spatial and temporal filtering. A number of original algorithms have been designed and implemented in MOPEX. Among them is direct plane-to-plane coordinate transformation, which allows at least an order of magnitude speed up in performing coordinate transformation by bypassing the sky coordinates. The dual outlier detection makes possible outlier detection in the areas of even minimal redundancy. Image segmentation based on adaptive thresholding is used for object detection, which is part of outlier as well as point source detection. Efficient use of computer memory allows mosaicking of data sets of very deep coverage of thousands of images per pointing, as well as areas of sky covering many square degrees. Although designed for SPITZER data, MOPEX does not require any SPITZER-specific fits header keywords to run , and can be applied to other data, that have standard header information on the image

			geometry and pointing. The package is available for distribution at http://ssc.spitzer.caltech.edu/postbcd/.
068	The robotic Liverpool Telescope data flow model	Robert. J. Smith, Liverpool John Moores University on behalf of the LT Operations Group, Liverpool John Moores University	The Liverpool telescope is a 2m fully robotic telescope own by Liverpool John Moores University (UK) and operating at Observatorio Roque de los Muchachos in the Canary Islands. It is entirely autonomous and is obtaining data for a wide variety of science programmes without any nightly supervision, dynamically scheduling the observations according to both weather conditions and certain `fairness' criteria. In order to take advantage of the telescope's capabilities; flexibility, efficiency and very rapid response to targets of opportunity; an integrated procedure of data handling had to be developed. For example, imaging data of photometric fields are reduced whilst the telescope acquires the next field, providing almost real-time data quality assessments, which can be fed back into the scheduling decisions. We will describe the full data flow model from phase II observation specification, though data acquisition, automated pipeline reduction and distribution to the appropriate astronomer using secure transfer protocols.
076	he Design of the W. M. Keck Observatory Archive	G. Bruce Berriman, Michelson Science Center and Infrared Processing and Analysis Center Thomas Bida, Lowell Observatory David Ciardi, Michelson Science Center Albert Conrad, W. M. Keck Observatory Anastasia Laity, Infrared Processing and Analysis Center	The Michelson Science Center and the W. M. Keck Observatory are building an archive that will serve data obtained at the Keck Observatory. The archive has begun operations and is ingesting level 0 (uncalibrated) observations made with the recently upgraded High Resolution Echelle Spectrometer (HIRES); these observations will be publicly accessible after expiration of a proprietary period.Observatory staff have begun using the archived data to determine the long-term performance of the HIRES instrument. The

		Jeffrey Mader, W.M. Keck Observatory Naveed Tahir-Kheli, Infrared Processing and Analysis Center Hien Tranh, W. M. Keck Observatory	archive is housed at the MichelsonScience Center (MSC)and employs a modular design with the following components:1. Data Evaluation and Preparation: images from the telescope are evaluated and native FITS headers are converted to metadata that will support archiving; 2. Trans Pacific Data Transfer: metadata are sent daily by email and ingested into the archive in a highly fault tolerant fashion, and FITS images are written to DVD's andsentto IPACeach week; 3. Science Information System:inherited from the NASA/IPAC Infrared Science Archive, it provides all the functionality needed to support database queries and processing of requests; and a web-based 4. User Interface, a thin layer above the information system that accepts user requests and returns results. The design offers two major cost saving benefits:it overcomes the geographical separation between the telescope and the archive and enables development at Keck and at MSC to proceed independently; and it permits direct inheritance of the IRSA architecture.
077	Recent developments of the ISO Data Archive	Alberto Salama, ESA ISO Data Centre Team, ESA	ESA's Infrared Space Observatory (ISO) performed 30,000 scientific observations, covering all areas of astronomy, as the world's first true infrared observatory. Two spectrometers, a camera and an imaging photo-polarimeter jointly covered wavelengths from 2.5 to 240 microns. Launched in 1995, ISO was operational until May 1998. All data had been re-processed with the end-of-mission calibration to populate the first homogeneous ISO Data Archive, which opened to the community in December 1998. Through the ensuing four years of the Post-operations Phase, ESA's ISO Data Centre developed and refined the ISO Data Archive to offer the ISO data to the worldwide astronomical community, and together with the several National Data Centres, worked to fill the archive with the best systematically processed and

			calibrated data products which could be achieved for the huge ISO database. These products allow users to select from the archive data sets of interest for deeper study with interactive analysis tools. During ISO's Active Archive Phase, which runs until December 2006, the ISO Data Centre continues to work with active National Data Centres to bring the archive into its final shape. In this talk we will review the major upgrades of the ISO Data Archive accomplished in this phase: (i) the incorporation of Highly Processed Data Products, the result of dedicated projects focused on cleaning the pipeline products from residual instrumental artefacts, (ii) the detailed characterisation of each observation via a Data Quality Report and (iii) the integration within the Virtual Observatories.
091	On the use of IDL for instrument control.	Steve Mazuk, The Aerospace Corporation Catherine Venturini, The Aerospace Corporation	The Aerospace Corporation has developed a near-infrared and visible spectrograph that is used for astronomical observations at Lick observatory’s 3 meter telescope.This paper describes the instrument control and data handling system, which employs the Interactive Data Language (IDL) for both the user interface and instrument control. The system employs IDL in a client-server design to control all aspects of data acquisition, and has been operational for several years. The use of IDL has simplified the system design and allowed for extensive modifications.
093	Interoperability in action: the Aladin Experience	F. Ochsenbein P. Fernique F. Bonnarel M. Allen T. Boch F. Genova A. Schaaff	Aladin was born 8 years ago as a small applet named "Aladin-lite", devoted to a visual correlation of sky images with the astronomical objects described in catalogues and databases. Since then, Aladin has been continually developed to provide good visualisation of catalog data overlaid on the observed pixels, and to be easily applicable to the holdings of existing image

			and data servers. Recently, Aladin was used as the main component of the Astrophysical Virtual Observatory prototype, demonstrating the use VO interoperability standards for enabling real science. But interoperability is not just a set of standards used to describe accurately the data -- it is also a means by which programs and tools can interact and cooperate. Several ways of communicating with and from Aladin have been developed, including a command language (scripting), java interfaces, and plug-ins. This presentation will focus on the lessons learned during the Aladin experience in several aspects: the strengths and weaknesses of the programming language (Java); the importance of the standards and protocols, their surprises and their challenges; the difficulties of actual interoperation; the necessity of keeping proper references to the data producers and maintainers -- and maintaining compatibility with the users' material.
097	MaxBCG: The Importance of Database Systems for the Virtual Observatory and the Grid	María A. Nieto-Santisteban, Johns Hopkins University Alexander S. Szalay, Johns Hopkins University Jim Gray, Microsoft Research Aniruddha R. Thakar, Johns Hopkins University William J. O’Mullane, Johns Hopkins University James Annis, Fermi National Accelerator Laboratory	The Maximum-likelihood Brightest Cluster Galaxy (MaxBCG) application searches for galaxy clusters that span a large dynamic range of cluster masses and provides good redshift estimates using the cluster red sequence galaxies. MaxBCG was developed originally in tcl using the SDSS Astrotools package and ran on the Terabyte Analysis Machine (TAM), a 10-CPU cluster specially tuned to solve this type of problem. The same application code was grid-enabled and used to test the Chimera Virtual Data System created by the GriPhyN project. As is common in astronomical file-based applications, the TAM and Chimera implementations used hundreds of thousands of files for the computations, files served from the SDSS Data Archive Server. The data required to run the MaxBCG is also available in the SkyServer database. At Johns Hopkins University, we have implemented a version using SQL on a SQL Server 2000 cluster that processes the same sky area an order of magnitude faster than

			TA! M. We are also working on an efficient grid-enabled system that does not require moving large volumes of data across the network. When the user submits the application, upon authentication and authorization, the SQL code is deployed and executed on the available data-grid nodes hosting the SkyServer database system. In this paper we describe why and how the Grid and the Virtual Observatory can take advantage of Database Management Systems.
099	The Architecture of the NASA/IPAC Infrared Science Archive (IRSA)	John C. Good, Infrared Processing and Analysis Center, Caltech Anastasia Alexov, Infrared Processing and Analysis Center, Caltech G. Bruce Berriman, Infrared Processing and Analysis Center, Caltech Nian-Ming Chiu, Infrared Processing and Analysis Center, Caltech Thomas H. Jarrett, Infrared Processing and Analysis Center, Caltech Mih-seh Kong, Infrared Processing and Analysis Center, Caltech Anastasia Clower Laity, Infrared Processing and Analysis Center, Caltech Serge Michel Monkewitz, Infrared Processing and Analysis Center, Caltech Naveed D. Tahir-Kheli, Infrared Processing and Analysis Center,	IRSA employs a component-based architecture based on a set of portable, standalone data access and processing programs.New archive interfaces are built by calling these modules through a standard command-control protocol.The inherent re-use of this approach makes the system highly extensible and dramatically simplifies supporting a wide range of projects. For instance, the IRSA tools are also being used for the WM Keck Observatory archive, interfaces to the Spitzer Space Observatory archive and support for the COSMOS Hubble Treasury Program.The architecture seamlessly supports National Virtual Observatory protocols via output mode switches in the data access modules. All this requires a remarkably small number of components.The external interfaces are:Gator, for general catalog queries; Atlas, for spatial queries of mixed collections of tables and image sets; and an interface currently under development to support complex multi-table database queries.These in turn utilize the data access modules described above including:ISISQL; for DBMS access;and Montage, for general image reprojection and mosaicking.

		Caltech Saille Warner-Norton, Infrared Processing and Analysis Center, Caltech Angela Zhang, Infrared Processing and Analysis Center, Caltech	Two signed JAVA applets are used for presentation and data fusion across IRSA and and external archives (another form of interoperabily):Oasis, for display of images and tables; and QtPlot(an extension of Berkeley's PtPlot) for general XY plotting. The architecture is rounded out by two integration tools:RADAR, which provides general region statistics and can organize the execution of all the above services for a complete view of a region of the sky;and ROME, a general request management system developed as part of the NVO.
100	Modern Statistical Methods for GLAST Event Analsis	Robin D Morris, USRA-RIACS Johann Cohen-Tanugi, Stanford Linear Accelerator Center	We describe the ongoing development of an event analysis algorithm for the Large Area Telescope (LAT) instrument on the Gamma-ray Large Area Space Telescope (GLAST).We show how it is possible to construct an algorithm that incorporates accurately the physics of the detector, both in terms of the processes that produce secondary particles and photons and most importantly, in terms of the multiple scattering processes of the charged particles.The posterior distribution is modeled as a mixture, where the discrete modes of the mixture correspond to different physical processes that can cause the same detector response.We use sequential importance sampling ("particle filters") to determine the distribution over the components of the mixture, that is, the possible configurations in terms of secondary processes and particles, and Markov chain Monte Carlo to explore the details of the distribution in each configuration.This allows us to estimate the mixture distribution (over configurations), where each mixture component is non-gaussian due to the multiple scattering and the detector geometry.The result is an algorithm that estimates the full distribution over the energy, azimuth and elevation of the incident photons, and aims to extract as much information as

			possible from the detector response.
101	Detection of Rare Objects in Massive Astrophysical Datasets Using Innovative Knowledge Discovery Technology	Alvaro Soto, Pontificia Universidad Catolica de Chile Antonio Cansado, Pontificia Universidad Catolica de Chile Felipe Zavala, Pontificia Universidad Catolica de Chile	This work presents new knowledge discovery in databases (KDD) technology that aids the scientist in the detection of rare types of astronomical objects. Our main idea is that while computers have the power to search huge amounts of data, the expert has the domain knowledge to efficiently lead this search. Our system builds upon two main components: a probabilistic model able to scale to large datasets and a set of modules to interact with the scientist. The probabilistic model represents the joint uncertainty of objects attributes in an astrophysical catalog. The model consists of a combination of Bayesian networks, and Gaussian mixture models, trained through an accelerated version of the expectation maximization algorithm (EM). The model provides the system with three main capabilities: detection of rare objects, through the identification of points with low probability, explanation of the sources of the anomalies, through the identification of the unusual attributes of an object, and scalability to large datasets, through the use of efficient data structures and algorithms. The modules to interact with the scientist incorporate innovative active learning techniques that use the expert feedback to progressively improve the performance of the probabilistic model. Using a modified version of the Distance-Weighted K-nearest neighbor algorithm, the system performs an interactive exploration of the space of potential rare objects. Areas where the system receives a positive feedback from the scientist, meaning that a detected object is indeed a rare object, are fully explored. Our system is currently being tested using data from the release 1 of the Sloan Digital Sky Survey.

107	AXedrizzle - Spectral 2D Resampling using Drizzle	Martin Kuemmel, Space Telescope - European Coordinating Facility Jeremy Walsh, Space Telescope - European Coordinating Facility Soeren Larsen, Space Telescope - European Coordinating Facility Richard Hook, Space Telescope - European Coordinating Facility	The aXe spectral extraction software was designed to extract spectra from grism images such as those taken with the Advanced Camera for Surveys on HST. Aided by an object catalogue and a configuration file, aXe extracts and calibrates the spectra in an unsupervised way. The usual way to combine spectra from several dithered grism images is to coadd the individual 1D spectra extracted for each object to create a deep 1D spectrum. This procedure rebins the data twice, and a complex weighting scheme is required to propagate pixel defects and cosmic ray affected pixels through the reduction process. In order to mitigate these drawbacks we have developed aXedrizzle, in which a deep 2D spectrum of each object is formed by co-adding all the 2D spectra in the individual images. For the 2D resampling of the individual 2D spectra we employ the stsdas.drizzle software. The transformation coefficients for drizzle are computed such that the combined drizzled image resembles an ideal long slit spectrum with the dispersion direction parallel to the x-axis and cross-dispersion direction parallel to the y-axis. The final 1D extraction of the spectra can be done with standard aXe tasks, or any other longslit extraction program. In this contribution the new reduction scheme is introduced, and its advantages are discussed extensively. Results from Hubble Ultra Deep Field ACS/HRC Parallels program reduced with aXedrizzle are presented to demonstrate the feasibility of this first implementation of the drizzle code to combine spectral data. Possible applications to ground-based multi-object spectroscopy are discussed.
112	Features of the AstroGrid	Tony Linde, University of Leicester Andrew Lawrence, University of	AstroGrid is a UK e-Science project. Its aim is to develop a complete Virtual Observatory infrastructure and facilitate its deployment at key astronomical data

	approach to Virtual Observatory architecture	Edinburgh Nicholas Walton, University of Cambridge	centres throughout the UK. The project began in 2001 and s now nearing the end of its first phase, with a complete (V1.0) release of all software components planned for Dec 2004. A second phase has been approved and funded and will run to end 2007. This presentation will describe the key conceptual features of the AstroGrid approach to VObs architecture, including: # service-oriented architecture # browser-based access to all services # compilation of astronomical tasks into complex workflows # access to distributed file storage (MySpace) # configurable approach to dataset access and application execution # full management of users and groups and their access authorities We will show how the AstroGrid suite of services fully incorporate current and pending recommendations from the standards working groups of the International Virtual Observatory Alliance (IVOA), including: # common web service interface # both 'local' and 'full' registries for resource discovery # data exchange using the VOTable standard We will describe the advanced features of AstroGrid which allow developers to: # make complex registry queries using XQuery # read and write directly to MySpace # deploy existing command line tools

			In conclusion, we will describe developments planned for AstroGrid-2, including: # client-side "astronomer's workbench" # command-line interface and common scripting language # semantically-rich resource discovery # advanced data mining and visualisation services # full rollout across the UK
118	Japanese Virtual Observatory (JVO) prototype 2	Masahiro Tanaka, NAOJ Yuji Shirasaki, NAOJ Satoshi Honda, NAOJ Yoshihiko Mizumoto, NAOJ Masatoshi Ohishi, NAOJ Naoki Yasuda, U. Tokyo Yoshifumi Masunaga, Ochanomizu U. Yasuhide Ishihara, Fujitsu Ltd. Katsumi Abe, Fujitsu Ltd. Jumpei Tsutsumi, Fujitsu Ltd. Hiroyuki Nakamoto, SEC Ltd. Yuusuke Kobayashi, SEC Ltd. Tokuo Yoshida, SEC Ltd. Yasuhiro Morita, SEC Ltd.	We describe the architecture of the Japanese Virtual Observatory (JVO) prototype system version 2.JVO aims at seamless access to astronomical data archives stored in distributed data servers as well as data analysis environment.For this purpose, it is important to build a framework for access to remote servers, including remote procedure calls (RPCs) and data transfer.A data request for distributed database is written in the JVO Query Language.The JVO system parses the query language, decomposes it into individual remote procedures such as retrieval of catalog, image and spectrum and cross matching, and generate a work flow.Based on this work flow, remote procedures are called.For RPCs of JVO prototype system 1, we employed Globus toolkit 2 (GTK2).However, latency time of GTK2 RPCs was too long for successive short-time jobs.Therefore, we employ Globus toolkit 3 (GTK3) for JVO prototype system 2.As the result, we find that Grid Service in GTK3 imp! roves performance of RPC.In addition to Grid Service, Reliable File Transfer (RFT) is used for efficient data transfer.Astronomical data stored in distributed servers are discovered through a registry server which provides metadata discussed in the IVOA registry working group and is built using a XML database.
121	Grid-related activity	Claudio Vuerli, INAF-O.A.Trieste Leopoldo Benacchio, INAF-	INAF, the Italian National Institute of Astrophysics, is currently involved in several Grid-related projects, at both the national and international

	in progress at INAF	O.A.Padova Fabio Pasian, INAF-O.A.Trieste Juan Alcalà, INAF-O.A.Capodimonte Andrea Baruffolo, INAF-O.A.Padova Ugo Becciani, INAF-O.A.Catania Enrico Cascone, INAF-O.A.Capodimonte Adriano Fontana, INAF-O.A.Roma Giuseppe Longo, Università di Napoli	levels. Grid.it is a multi-disciplinary national project with the purpose of setting up and testing a national Grid infrastructure for the Italian research community. Astrophysics contributes with two applications. Their goal is to access astrophysical archives and catalogues (namely the pilot archive of the Telescopio Nazionale Galileo and the GSC-II) and to gridify the pipelines working on retrieved data (e.g. ASTRO-WISE). Work has been done to make the Grid middleware able to access and publish DBMS entities on the Grid. DRACO is a national project extending Grid.it. Five INAF institutes plus two Universities are involved in this project, aimed at using Grid technology for processing and mining huge amounts of data. Test-bed applications are contributed, while some nodes are validating dedicated technological solutions. DRACO is also the project allowing the Italian community to participate in the Virtual Observatory; several DRACO applications are VO-enabled already. Several initiatives are currently in progress to gridify pipelines for the Planck project. A grid-aware pipeline working on simulated Planck data will be officially proposed to the EGEE project. In the meantime Planck pipelines are being tested on Gilda, a test-bed Grid infrastructure setup to host applications to be later ported on EGEE. A joint collaboration with ESA has also been started for this purpose. Test pipelines using the Planck simulation software have been successfully tested on the Grid infrastructure in use for Grid.it and Gilda.
124	Reducing VIMOS IFU data with	Ralf Palsa, European Southern Observatory Carlo Izzo, European Southern	To cope with the ever increasing number of operational instruments, and with the quantity and complexity of the produced data, the European Southern Observatory

	Gasgano and CPL-based pipeline recipes	Observatory Nick Kornweibel, European Southern Observatory Derek J.McKay, European Southern Observatory and Rutherford Appleton Laboratory Michele Peron, European Southern Observatory	(ESO) has developed the Common Pipeline Library (CPL) and the front-end application Gasgano. The CPL is a library that standardises and streamlines the implementation of pipelines, while Gasgano permits the interactive use of high-level pipeline recipes. The context of IntegralField Spectroscopy offers an ideal test-case where the advantages of using the CPL and Gasgano can be evaluated. The pipeline algorithmsfor the processing of the VIMOS Integral Field Unit (IFU) data were developed at ESO in only six staff-months, taking full advantage of the ready-to-use functionality available in the CPL. The data reduction strategy applied by the VIMOS IFU pipeline recipes is based on the MOS paradigm: the scientific spectra are extracted, wavelength calibrated and corrected for the transmission differences between fibres, on the basis of the extraction mask and flux measurements obtained from an associated set of flat-field and arc-lamp exposures. The use of the Gasgano interface greatly simplifies the task of classification and correct association of the large number of calibrations and intermediate products involved in the process.
125	Monte Carlo Image Analysis in Radio Interferometry	Urvashi R.V., NRAO/NMT T.J. Cornwell, NRAO	Image analysis, such as component fitting of radio interferometric images has traditionally been based on likelihood techniques applied to the deconvolved images. The analysis usually ignores the uncertainty arising from the process of deconvolution. Ideally one would estimate the properties of components representing the entire emission present in the raw, dirty image. In practice, this is not feasible given the large dimensionality of the parameter space. We present an intermediate approach in which a Bayesian image analysis is performed to fit components to sub-regions of the dirty image, taking full account of the point spread function. Prior to the fitting of emission inside a given region, the emission outside

			of the region is removed from influence by subtracting a previously deconvolved image. Our method produces samples of the posterior distributions for the number and parameters of elliptical gaussian components within the region of interest. We compare the performance of this approach to the standard methods.
129	Astronomical database related applications in the Grid.it project	Alessandra Volpato, INAF-O.A.Padova Giuliano Taffoni, INAF-O.A.Trieste Serena Pastore, INAF-O.A.Padova Claudio Vuerli, INAF-O.A.Trieste Andrea Baruffolo, INAF-O.A.Padova Riccardo Smareglia, INAF-O.A.Trieste Giuliano Castelli, INAF-O.A.Trieste Fabio Pasian, INAF-O.A.Trieste Leopoldo Benacchio, INAF-O.A.Padova	The Astronomical Observatories of Trieste and Padova are involved in the Grid.it project, a multidisciplinary research project, funded by the Italian Ministry for Education, University and Research, aiming at 'Enabling platforms for high-performance computational grids oriented to scalable virtual organizations'. The workpackage of this project named Grid Applications for Astrophysics focuses on exploring the use of grid technologies for the development of astrophysical applications. Within this workpackage the main goal of the OAPd group is to study the portability to the Grid of an existing system for the consultation of large astronomical catalogues, currently serving on the net the Second Guide Star Catalog (GSC-II). In the same framework, the Astronomical Observatory of Trieste (OAT) Technology Group is dealing with the similar problem of integrating in the Grid the archive of observational data from the Italian Galileo National Telescope (TNG) and, as a further step, to provide processing for the data retrieved through grid-enabled pipelines. Since the grid infrastructure provided by the Italian National Institute for Nuclear Physics (INFN) is based on GT-2, an adequate model for DBMSes is not available. For this reason, INFN and INAF institutes are collaborating to design and implement a suitable architecture to integrate DBMSes in the existing grid infrastructure. In the meanwhile INAF OATs and OAPd

			have developed temporary solutions to circumvent the problem, by adopting two complementary approaches: one based on a client-server paradigm and the other on Web-Services architecture. In this paper we report about these activities and the preliminary results obtained so far.
137	Mosaicing with interferometric radio telescopes : Efficient algorithm for imaging and image plane corrections.	S. Bhatnagar, NRAO Socorro K. Golap, NRAO Socorro T.J. Cornwell, NRAO Socorro	High resolution imaging of objects larger than the primary beam of the antennas in an interferometric radio telescope require the mosaic imaging mode.Moscing is expected to be widely used with upcoming new telescopes like the ALMA and the EVLA.Development of algorithms to solve for the errors which limit the dynamic range with this mode of imaging is scientifically important. Since the emission spans many primary beams, primary beam and the pointing offsets related image plane errors adversely affect, and currently limit the dynamic range of the mosaiced images. Efficiency of algorithms to solve for such antenna based errors using the measured data itself, depends on the accuracy of the image reconstruction algorithm. Such algorithms as well as the image reconstruction algorithms are both computationally expensive. The dominant cost of such an algorithm comes from the computation of the derivative of the objective function with respect to the pointing offsets. Residual visibility computation also require the image model. This in turn requires re-sampling (gridding) of the observed visibilities from all the mosaic pointing observations on a regular grid for Fourier inversion.The dominant cost of imaging, particularly for large number of pointings, strongly depends on the cost of gridding the visibilities on a regular grid. In this paper, we present a relatively efficient visibility gridding algorithm and the results from an

			algorithm for solving for the antenna based pointing offsets.Implications on the imaging dynamic range and computational costs for mosaicing will also be discussed.
144	OpenSkyQuery & OpenSkyNode - the VO Framework to Federate Astronomy Archives	William O'Mullane, JHU Tamas Budavari, JHU Vivek Haridas, JHU Nolan Li, JHU Tannu Malik, JHU Masatoshi Oishi, NAOJ Alexander Szalay, JHU Aniruddha Thakar,JHU Ramon Williamson,NCSA	OpenSkyNode and ADQL are the major new steps in the Data Access layer of the Virtual observatory. Presented as a poster last year OpenSkyQuery(OSQ) now has a functioning portal which allows cross matches between catalogs on thirteen nodes. The portal also allows a user to upload their own list of source to be cross matched with the registered SkyNodes. This is one of the first systems utilizing of the IVOA's nascent standard Astronomical Data Query Language(ADQL). We shall demonstrate the functioning of the portal as it communicates with the nodes and the NVO searchable registry. We shall present the OSQ architecture and explain its use of WebServices as an open standard, Java and C#implementations of OpenSkyNode demonstrates that the open standard does indeed work.We shall also discuss the newest ADQL0.8 specification which will allow ADQL to be used to query registries in the VO indeed the the NVO searchable registry is already accepting ADQL0.7.4. and implem! ents OpenSkyNode0.7.4.
145	Extending Sherpa with S-Lang	Douglas Burke, Harvard-Smithsonian Center for Astrophysics Aneta Siemiginowska, Harvard-Smithsonian Center for Astrophysics Stephen Doe, Harvard-Smithsonian Center for Astrophysics	The modeling and fitting program Sherpa is one of the cornerstones of the CIAO X-ray astronomy analysis software package. It allows users to fit their one- and two-dimensional data sets - such as spectra, radial profiles, or images - in an interactive environment. The addition of the S-Lang interpreted language (http://www.s-lang.org/) in the CIAO 3.0 release offers the user the capability to not only control Sherpa from a scripting language, but also to customize and extend Sherpa's capabilities by using the S-Lang interfaces to CIAO's infrastructure

			(libraries). In this talk we will describe a number of extensions to Sherpa written in S-Lang, including: plot customizations; alternative user interfaces to Sherpa's commands; extra plots for comparing fits and image data; and interactive data manipulation. These extensions require no compilation of code by the user and are available for download from the CIAO web pages (http://cxc.harvard.edu/ciao/).
147	development of SAOImage DS9: Lessons learned from a small but successful software project	William Joye, Smithsonian Astrophysical Observatory Eric Mandel, Smithsonian Astrophysical Observatory	We will present a discussion of lessons learned during the development of SAOImage DS9. Starting with general observations on scientific software development, we will discuss our design and implementation cycle, allocation of effort and resources, keys to our success, overall strategies that have worked well (and those that have not), and future issues and challenges. We hope our experience will be of use to other small software development projects.
148	Linking and tagging initiatives at the Astrophysical Journal	Greg Schwarz, AAS Journals Staff Scientist	At the start of 2005 the Astrophysical Journal will be offering three new linking and tagging options for the electronic edition.The linking projects are in conjunction with partnerships with ADS, CDS, and NED. Authors may use new AASTeX mark up to tag data sets of participating data centers.A common name resolver hosted at ADS verifies the data set tag during copy editing and forwards the reader who follows the links to the appropriate data center.A demonstration of this capability can be seen in the September Spitzer ApJ Supplement.In a similar vein authors we will support object linking to SIMBAD and NED.For this project we will also query authors to supply the most important objects of their paper during peer review.These object lists will be supplied to the data centers to help

			them integrate the information into their databases and potentially check for errors.The SIMBAD and NED object web pages that are now created weeks and months after publication at CDS will be also be dynamically added to the navigation bar once they are made available.The final endeavor is a controlled set of major astronomical facility keywords to help organizations track the effectiveness of their telescopes.The ultimate goal of each project is to allow authors make their papers more useful and to help researchers seamlessly navigate between the journal and the data archives.
151	Grist: Grid Data Mining for Astronomy	Joseph C. Jacob, Jet Propulsion Laboratory, California Institute of Technology Roy Williams, California Institute of Technology Jogesh Babu, The Pennsylvania State University S. George Djorgovski, California Institute of Technology Matthew J. Graham, California Institute of Technology Daniel S. Katz, Jet Propulsion Laboratory, California Institute of Technology Ashish Mahabal, California Institute of Technology Craig D. Miller, Jet Propulsion Laboratory, California Institute of Technology Robert Nichol, ICG, University of Portsmouth, PO1 2EG, UK Daniel E. Vanden Berk, The	The Grist project is developing a novel grid-technology based system as a prototype research environment for astronomy with massive and complex datasets.When complete, this knowledge extraction system will consist of a library of distributed grid services controlled by a workflow system, compliant with standards emerging from the grid computing, web services, and virtual observatory communities.The science drivers for Grist include finding high redshift quasars, studying peculiar variable objects, search for transients in real-time, and the fitting of SDSS QSO spectra to measure black hole masses.Grist services will also be part of a compelling vehicle for outreach as a component of the “hyperatlas” project to serve high-resolution multi-wavelength imagery over the internet.In support of these science objectives, the Grist framework will provide the enabling fabric to tie together distributed grid services in the areas of data access, federation, mining, so! urce extraction, image mosaicking, catalog federation, data subsetting, statistics (histograms, kernel density estimation, and R language utilities exposed by VOStatistics services), and visualization.Interactive deployment and control of

		Pennsylvania State University Harshpreet Walia, Jet Propulsion Laboratory, California Institute of Technology	these distributed services will be provided from an intuitive, graphical desktop workflow manager.The new grid services paradigm explored in Grist will pave the way for a new era of distributed astronomy, with tremendous flexibility that allows software components to be deployed as services that are: (i) controlled and maintained by the authors; (ii) close to the data source for efficiency; or (iii) controlled by the end users so they have control over policies and level of service.
152	Data Mining in the Era of Virtual Observatorie	Andrew Connolly, University of Pittsburgh Joerg Colberg, University of Pittsburgh Simon Krughoff, University of Pittsburgh Ryan Scranton, University of Pittsburgh Andrew Moore, Carnegie Mellon University	With the development of the Virtual Observatories astronomers now have access to massive, distributed data sets covering many decades of the electromagnetic spectrum. To fully realize the scientific potential of the VO we require analysis tools that can scale to the size and dimensionality of these collections of data and that are interfaced directly onto the Virtual Observatories (i.e. to make the analysis of VO data as seamless and transparent as our access to the data themselves). In this talk we present the initial implementation of a toolbox for data mining in the era of the Virtual Observatory. We will discuss a range of tools designed to analyze the large, multidimensional data sets that are currently available through the VO. We will specifically address analyses that cover a broad range of questions that astronomers might want to ask of the VO data: how to measure the clustering of galaxies from large photometric and spectroscopic surveys using a single and parallel computing environment, how to identify anomalous or unusual sources in large multidimensional data sets, how to track moving sources in time domain data and how to integrate the multi-frequency data available through the VO with external data sets. We will discuss how the algorithms for addressing these questions are designed and

			implemented and how these tools are integrated into a webservice environment together with the associated challenges for undertaking statistical astronomy on large heterogeneous data sets.
164	Mining the Astronomical Archives for the "Tadpole" Galaxy	T.H. Jarrett, IPAC/Caltech Ismael Perez Fournon, IAC Gordon Stacey, Cornell U Donovan Domingue, IPAC/Caltech	We present results from an investigation of the peculiar"Tadpole Galaxy" (UGC 10214).This work is motivated by the Spitzer Wide-area Infrared Extragalactic Survey (SWIRE)observations of the ELAIS N1 region, which includes UGC10214and its local environment. The Spitzer imaging data comprisethe four mid-infrared channels of IRAC and the 24-micron MIPSobservations.To supplement these mid-infrared data, we have (1) carried out a ground-based program to image the Tadpole athigh resolution in the optical and near-infrared windows, and (2) systematically searched through publicly available astronomicalmission archives.Our presentation includes the data miningprocess used to collect and organize the complex and disparatedata sets that come from mission/observatory archives, notablythe IRSA (infrared), MAST (HST), HEASARC (high energy), NED &LEDA (extragalactic) and CDS/VizieR (general) archives.Wespotlight the use of three virtual observatory tools: IRSA'sinventory and data exploration services RADAR and ATLAS, andthe NVO data inventory service.
166	VO-enabling a major astronomical analysis and reduction software system	David Giaretta, Starlink/RAL Malcolm Currie, Starlink/RAL Stephen Currie, Starlink/RAL Mark Taylor, Starlink/U of Bristol Norman Gray, Starlink/U of Glasgow Peter Draper, Starlink/U of Durham David Berry, Starlink/U C lancs	This talk will describe the Starlink experience in its process of VO-enabling its existing comprehensive collection of astronomical analysis and reduction software packages. Starlink has been involved from the early stages of the VO in the UK and internationally. It has been working actively in the IVOA process, for example the contributing to the VOTable standard development as well as creating the most complete VOTable library, as part of the STIL. While not tracking every twist and turn, Starlink has

		Alasdair Allan, Starlink/U of Exeter	incorporated, in pilot studies, several of the favoured Grid techniques of the day. Each time the aim has been to explore options for making minimum modifications while Grid/VO enabling the bulk of the software. As the GRID/VO standards (appear) to be stabilising and providing greater functionality, VO-enabled Starlink applications are being released, having adequate functionality, robustness, level of documentation and quality. This talk will describe some of the design choices we have made, the VO capabilities we are incorporating and contributing and the lessons learned.
172	Time Domain Explorations With Digital Sky Surveys	A. Mahabal (Caltech) M. J. Graham (Caltech) S. G. Djorgovski (Caltech) R. Williams (Caltech) P. Kollipara (Caltech) E. Krause (Caltech) B. Granett (Caltech) C. Baltay (Yale University) D. Rabinowitz (Yale University) A. Rengstorf (NCSA/UIUC) R. Brunner (NCSA/UIUC) M. Bogosavljevic (Caltech) A. Bauer (Yale University) P. Andrews (Yale University) N. Ellman (Yale University) S. Duffau (Yale University) J. Snyder (Yale University) N. Morgan (Yale University) J. Musser (Indiana University) S. Mufson (Indiana University) M. Gebhard (Indiana University)	One of the new frontiers of astronomical research is the exploration of time variability on the sky at different wavelengths and flux levels. We have carried out a pilot project using DPOSS data to study strong variables and transients, and are now extending it to the new Palomar-QUEST synoptic sky survey. We report on our early findings and outline the methodolgy to be implemented in preparation for a real-time transient detection pipeline. In addition to large numbers of known types of highly variable sources (e.g., SNe, CVs, OVV QSOs, etc.), we expect to find numerous transients whose nature may be established by a rapid follow-up.Whereas we will make all detected variables publicly available through the web, we anticipate that email alerts would be issued in the real time for a subset of events deemed to be the most interesting.This real-time process entails many challenges, in an effort to maintain a high completeness while keeping the contamination low.We will utilize distributed Grid services developed by the GRIST project, and implement a variety of advanced statistical and machine learning techniques.
177	The BUCS	Rychard Bouwens, UC Santa Cruz	Interpreting today's high-resolution multiwavelength

	Library for Cloning Complete Galaxy Samples	Garth Illingworth, UC Santa Cruz Dan Magee, UC Santa Cruz	data on galaxies can be a challenging task.Real galaxies have complex morphologies, which depend upon wavelength, and can be found on images with a wide variety of beam smearing and signal-to-noise.To properly measure the evolution of galaxy structure and morphology across cosmic time, these properties need to be treated.It is to this end that we have developed the BUCS software suite. With this package, it is possible to select complete samples of galaxies off an image set and then project these objects to higher redshift accounting for pixel k-corrections, cosmic surface brightness dimming, and image specific noise and PSF.The end result of these simulations are no-evolution image sets that can be directly compared against the observations.This suite is now available to do science and as an illustration of its capabilities we describe its use in interpreting z~6-8 galaxies (i and z-dropouts) from the Hubble Ultra Deep Field. Work is also under way to port these capabilities to the web.The purpose of this talk is to introduce this effort to the astronomical community as a whole.
178	New Features for VO-Enabled Data Intensive Science with the SDSS Data Release 3	Ani Thakar, JHU Alex Szalay, JHU Jim Gray, Microsoft Research William O'Mullane, JHU Tamas Budavari, JHU Maria Nieto-Santisteban, JHU George Fekete, JHU Nolan Li, JHU Robert Lupton, Princeton	With the third data release (DR3), the SDSS Catalog Archive Server (CAS) serves up 3 Terabytes of catalog data for nearly 200 million celestial objects via the SkyServer web pages at skyserver.sdss.org.Several changes and new features have been added that facilitate data-intensive science with the data and advance VO standards and technologies. 1. The CasJobs batch query system unveiled at ADASS XIII now allows stored procedures and functions in the user's MyDB. This enables users to bring their program to the data rather than the other way around, which is far more efficient for data-intensive applications.We replicate user account details between installations to allow

			distributed CasJobs configurations.This form of replication raises issues that are crucial to the VO community. We are also attempting to secure the system using X509 certificates through WS-Security. This ties in with the distributed storage effort within the VO. 2. New color images in the Visual Tools based on the ASINH stretch that are deeper and more feature-rich.We briefly discuss the algorithm. 3. The traffic logging system is now quite detailed and extensive, allowing many useful and interesting statistics to be gathered.The Support Services specification proposed for IVOA includes a logging section modeled on this apporach. 4. A much faster HTM (spatial index) library with an upgraded interface. 5. An enhanced object crossid facility allowing users to run their own SQL query on the matching objects. We briefly describe these features and list future enhancements anticipated with SQLServer Yukon and SDSS-DR4.
181	Registries and Publishing in the Virtual Observatory	Raymond Plante, NCSA	A registry, as a repository containing descriptions of available data and services, is where data discovery and data publishing first meet in the Virtual Observatory (VO).In the last year, the IVOA framework for registries has advanced from a working prototype system to a comprehensive set of standard interfaces.I review the emerging standards for registries and the metadata they contain and discuss how registries around the world can work together to create an up-to-date picture of available VO resources.I illustrate how the standards--in particular, the metadata standards--allow the

			registries to be used in a variety of ways.I highlight some of the current challenges being examined.One such issue is that of registry curation: ensuring sufficient accuracy, consistancy, and completenes to make the descriptions useable.Finally, I focus on the NVO approach to registries and how it fits into the larger process of publishing in the VO.
183	Astronomical Computing within Data Archives	Adrian Pope Alex Szalay Jim Gray	As astronomical datasets continue to grow in size we begin to face serious issues with transporting data from archives to computing resources used to do calculations with the data.We have begun to attack this problem by performing as many calculations as possible within the archive.One example is the use of a SQL Server archive with Hierarchical Triangular Mesh (HTM) spatial indexing software to pixelize the low redshift Sloan Digital Sky Survey (SDSS) Main Galaxy Sample (MGS) before performing Large Scale Structure (LSS) analyses.The resulting counts-in-cells data is an order of magnitude smaller than the input catalog, dramatically reducing the amount of data transported to external programs.We also present a system for analytically calculating the spatial relationships between spherical polygons within an archive that has been used to analyze the complex geometry of SDSS survey footprint and masks, a necessity for preparing samples of data suitable for detailed LSS studies.The software was written (by Alex Szalay and Jim Gray) in the high level Transact-SQL language, with the use of the HTM library.Finally we describe the creation and utilization of a Monte Carlo realization of the SDSS for studying the effects of calibration errors and masking on LSS studies.The realization contains hundreds of millions of random points and was created using a cluster of SQL Server machines.
187	SALSA:an	omas Quinn, University of	Both astrophysical simulations and large astronomical

	interactive tool for parallel analysis and visualization of Astronomical data sets	Washington, Graeme Lufkin, University of Washington, Greg Stinson, University of Washington, Filippo Gioachin, University of Illinois, Orion Lawler, University of Illinois, Laxmikant Kale, University of Illionois	surveys are producing catalogs containing of order one billion objects.Analyzing such data sets is difficult with a serial package, and custom coding of parallel software is tedious. To meet these challenges, we have developed Salsa, a parallel, interactive analysis tool for point-like data such as particles in an N-body simulation or object catalogs from a sky survey. The user runs a graphical client application on their desktop, which communicates with the server running on a parallel machine. The client controls the display of the simulation, which is rendered by the server. The client can enter high level code, like Python, that is executed on the server, implementing new functionality and providing programmatic control. The client is currently written in Java and the server is written in the CHARM++ parallel language for maximum portability.The client/server pair of Salsa is currently used for active research in N-body simulations, and its use on Sloan Digital Sky Survey data will be demonstrated. The server can read particle or catalog data, render different types of visualizations, create and manipulate groups of objects, and generate statistics on these groups.Information about the object data and control over groups is exposed to the embedded Python interface. The client manipulates the simulation view, defines groups, and sends user-written code to the server.