ADASS XXVIII Abstract Book

Brought to you by a Makefile, calling python code that operated on a Excel spreadsheet.

Generated 2018-10-21T01:18:28.084278

P8.1: A GPU implementation of the harmonic sum algorithm

Theme: Time Domain Astronomy
Author(s):
Karel Adamek (University of Oxford, Department of Engineering Science, Oxford e-Research Centre)
Wesley Armour (University of Oxford, Department of Engineering Science, Oxford e-Research Centre)

Contact: *karel.adamek@oerc.ox.ac.uk*

Detecting pulsars in time-domain radio astronomy using Fourier transform based techniques is a convenient and computationally efficient way to extract the faint single pulses from the noise in which they sit. However this technique, called periodicity searching, has some pitfalls. One of these is that the power contained in pulsar signal is spread into multiple harmonics in calculated power spectra. The incoherent harmonic sum algorithm is one way to rectify this. The algorithm sums the power that is spread across multiple harmonics back into a single fourier bin. This increases signal-to-noise ratio of detected pulsars and allows us to detect weaker pulsars as a result. Harmonic summing also forms part of search techniques used for detecting accelerated pulsars where a two-dimensional harmonic sum is required. Or in more complex cases such as for pulsars with elliptical orbits where a three-dimensional harmonic sum might be necessary. However porting the harmonic sum to many-core architectures like GPUs is not straightforward. The main problem that must be overcome is the very unfavourable memory access pattern, which gets worse as the dimensionality of the harmonic sum increases. We present a set of algorithms for calculating the harmonic sum that are more suited to many-core architectures such as GPUs. We present an evaluation of the sensitivity of these different approaches, the performance and discuss GPU portability.

P7.1: STELLAR ACTIVITY EFFECTS ON MOIST HABITABLE TERRESTRIAL ATMOSPHERES AROUND M DWARFS

Theme: Software for Solar Systems Astronomy
Author(s):
Mahmuda Afrin Badhan (University of Maryland College Park)
Eric T. Wolf (University of Colorado Boulder)
Ravi Kumar Kopparapu (NASA Goddard Space Flight Center)
Eliza M.-R. Kempton (University of Maryland College Park)
Giada Arney (NASA Goddard Space Flight Center)
Shawn Domagal-Goldman (NASA Goddard Space Flight Center), Drake Deming (University of Maryland College Park)

Contact: *afrin20m@astro.umd.edu*

The first habitable zone (HZ) exoplanets to have their atmospheres characterized will likely be tidally-locked planets orbiting nearby M dwarfs. Future transit spectroscopy of such planets is part of the communitys plan in assessing their habitability. 3D climate modeling has shown tidally-locked HZ terrestrial planets, at the inner HZ of M dwarfs, may possess significantly enhanced water vapor content in the lower atmosphere compared to the 24 hour analog. For model M dwarfs with T 3000K in particular, such inner HZ planets have been shown to retain the moist atmosphere for low Earth-like instellation levels. This is promising for both habitability as well as our ability to spectrally detect this with the upcoming James Webb Space Telescope. However, while strong vertical mixing is expected to loft the water vapor high enough into the atmosphere, M dwarfs are typically high XUV environment. To assess whether the water vapor destruction continuously driven by such stellar UV activity levels would affect detectability, we run 1-D photochemical models of these atmospheres under varying stellar UV activity. To remain in the moist greenhouse regime as established in Kopparapu et al. 2017, we take the 3D model simulated abundances and temperature profiles for a N_2-H_2O dominated planet around a 3300K M dwarf. We also explore additional chemical complexity by introducing new species to our atmosphere. We find that as long as the atmosphere is well-mixed at 1 mbar and higher pressures, UV activity has no impact on our H_2O detectability. We also find that even the highest UV scenario does not produce spectrally significant O_3 for JWST.

P4.1: Performance analysis of the SO/PHI software framework for on-board data reduction

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
K. Albert (Max Planck Institute for Solar System Research (MPS))
J. Hizberger (Max Planck Institute for Solar System Research (MPS))
D. Busse (Max Planck Institute for Solar System Research (MPS))
J. Blanco Rodrguez (Universidad de Valencia)
J. P. Cobos Carrascosa (Instituto de Astrofisica de Andaluca (IAA - CSIC))
B. Fiethe, Institute of Computer and Network Engineering at the TU Braunschweig (IDA), A. Gandorfer, MPS, Y. Guan, IDA, M. Kolleck, MPS, T. Lange, IDA, H. Michalik, IDA, S. K. Solanki, MPS, J. C. del Toro Iniesta, IAA - CSIC, and J. Woch (MPS)

Contact: *albert@mps.mpg.de*

The Polarimetric and Helioseismic Imager (PHI) is the first solar spectropolarimeter that goes to deep space. It will be launched on-board the Solar Orbiter (SO) spacecraft, to orbit the Sun in highly elliptical orbits. SO/PHI has stringent requirements on its science data accuracy, while it is subjected to highly dynamic environments by the spacecraft. The orbits, however, severely limit the amount of telemetry, making the download of numerous full datasets and additional calibration data unfeasible. In addition, they also increase the command-response times of the instrument. To overcome these limitations, SO/PHI implements autonomous on-board instrument calibration, and autonomous on-board science data reduction. The algorithms are implemented within a system of software framework and dedicated hardware implementations in reconfigurable FPGA-s. The system is designed to overcome the resource and computational power limitations of on-board computing, do secure metadata logging and warn ground support of possible errors in the data processing flow. In this contribution we do an in-depth performance analysis of the on-board data processing system. We test multiple data pipelines on realistic data in a step by step analysis, and tune the pipeline parameters to achieve optimal results. We finally analyse the accuracy of the resulting scientific data products calculated by the instrument.

P3.1: Exoplanets data visualization in multidimensional plots using virtual reality in DACE

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Fabien ALESINA (University of Geneva)
Nicolas BUCHSCHACHER (University of Geneva)
Julien BURNIER (University of Geneva)

Contact: *fabien.alesina@unige.ch*

The Data Analysis Center for Exoplanets (DACE) is a web platform based at the University of Geneva (CH) dedicated to exoplanets data visualization, exchange and analysis.

This platform is based on web technologies using common programming languages like HTML and JavaScript. During the past 3 years, the plotting tools has been improved in order to display large datasets on the platform, dealing with browsers performances constraints.

The next challenge is to display the exoplanets data in multidimensional plots. The web virtual reality technology has been added on DACE, and allows the user to display the data in virtual reality devices like cardboards and headsets.

The virtual reality is used for displaying 3D plots of synthetic planetary populations, discovered exoplanets from different archives, and 3D planetary systems with a star and its orbiting planets.

The used technologies are webVR, external GPUs called eGPUs in order to increase laptop performances, HTC vive pro headset and google cardboards.

O2.1: Hit the Ground Running: Data Management for JWST

Theme: Management of Large Science Project
Author(s):
Anastasia Alexov (Space Telescope Science Institute)
Mark Kyprianou (Space Telescope Science Institute)
Catherine Kaleida (Space Telescope Science Institute)

Time: Thu 11.45 - 12.00

Contact: *alexov@stsci.edu*

As the launch of James Webb Space Telescope (JWST) approaches a team of engineers and scientist is hard at work developing the Data Management Subsystem (DMS) for JWST with its cadre of complex imaging and spectral instruments. DMS will perform receipt of science and engineering telemetry data; will perform reformatting, quality checking, calibration, data processing; will archive the data; will have tools for retrieving the data; will have the capacities for reprocessing the data; will have external/public calibration tools; will provide user notification, search, and access tools for JWST science and engineering data; will distribute data to the end user; provide extensive user analysis/visualization tools; and, will provide support for contributed data products from the community. We will give an overview of the software components, the hardware they run on, the programming languages/systems used, the complexity of the tested end to end science data flow, the current functionality of the system and what's to come for the JWST Data Management Subsystem in preparation for launch.

O13.1: Receiving Credit for Research Software

Theme: Other
Author(s):
Alice Allen (Astrophysics Source Code Library)

Time: Wed 10.00 - 10.15

Contact: *aallen@ascl.net*

Though computational methods are widely used in many disciplines, those who author these methods have not always received credit for their work. This presentation will cover recent changes in astronomy, and indeed, in many other disciplines, that include new journals, policy changes for existing journals, community resources, changes to infrastructure, and availability of new workflows that make recognizing the contributions of software authors easier. This talk will include steps coders can take to increase the probability of having their software cited correctly and steps researchers can take to improve their articles by including citations for the computational methods that enabled their research.

P4.2: Optimization of Aperture Photometry in the Chandra Source Catalog

Theme: Data Science: Workflows Hardware Software Humanware

Author(s):
Christopher Allen (Smithsonian Astrophysical Observatory)
Francis A. Primini (Smithsonian Astrophysical Observatory)
Joseph B. Miller (Smithsonian Astrophysical Observatory)

Contact: *ceallen@cfa.harvard.edu*

As part of the creation of the Chandra Source Catalog, the Aperture Photometry systems characterize source photon counts, count rates, energy flux, and photon flux for over 300,000 sources. These sources are characterized per observation and across blockings of similar observations, leading to approximately three million unique characterizations. As development progressed and our test sets grew both in both complexity and size, runtime issues were identified - scenarios where the system was too resource-intensive or had projected processing times far outside of the project timeframe. These issues lead to a continual reevaluation of our implementation of the tool, all the while keeping the underlying scientific algorithms intact. Herein we will discuss the algorithm, the challenges we encountered scaling it up for production, and the optimizations we added as a result.

These design decisions cover a broad scope, from the mundane of restructuring nested loops to minimize file I/O, to the intricate of modifying our statistical sampling methods and replacing our hypermesh-based fitting with a Markov chain Monte Carlo sampling that was then converted into a probability distribution.

Analysis tools used to identify code hotspots will be reviewed, as well as the coding tools we used to refactor some computationally expensive sections of the tool. In exploring these decisions in detail, the pros and cons of both the chosen solution and alternative options considered but not chosen will be discussed.

The tools high-level design will also be analyzed, elaborating on how we were able in a single tool to perform analysis on the broad variety of sources the catalog encounters - from point sources to broad extended emissions, and from the faintest detections to brightest.

P2.1: Toward common solutions for data access, discovery and interoperability

Theme: Management of Large Science Project

Author(s):
Mark Allen (CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-6700 Strasbourg)

Contact: *mark.allen@astro.unistra.fr*

The ASTERICS project has supported a program of work to make enablethe data from large astronomy infrastructures (ESFRI) and pathfinder projects to become available for discovery and usage by the whole astronomical community, interoperable in the Virtual Observatory framework, and accessible with a set of common tools. We highlight the successes and lessons learned.

O1.1: Automatic Classification of TESS Planet Candidates using Deep Learning

Theme: Machine Learning in Astronomy
Author(s):
Megan Ansdell (UC Berkeley)
Yani Ioannou (Cambridge University)
Michele Sasdelli (Cortexica Vision Systems)
Hugh Osborn (Lab of Astrophysics of Marseille)
Jeff Smith (NASA Ames / SETI Institute)
Massimo Mascaro, Google; Adam Lesnikowski, Nvidia; Daniel Angerhausen, Center for Space and Habitability; Jon Jenkins, NASA Ames

Time: Tue 14.30 - 14.45

Contact: *ansdell@berkeley.edu*

We present results from a NASA Frontier Development Lab (FDL) project to automatically classify exoplanet candidates identified by the Transiting Exoplanet Survey Satellite (TESS) using artificial intelligence (AI) techniques applied with compute resources provided by the Google Cloud Platform. NASA FDL is an applied AI research accelerator aimed at implementing cutting-edge AI technologies to challenges in the space sciences. NASA's TESS mission, launched in April 2018, is conducting an all-sky transit survey to discover thousands of exoplanets around bright stars that are amenable to follow-up observations needed to characterize planet sizes, masses, and even atmospheres. TESS observes sectors of the sky for 27 days, after which the TESS pipeline generates flux time series for each stellar target, identifies any transit-like signatures, and calculates diagnostic metrics for each event. Time is of the utmost importance given the short ground-based observing window following each TESS observing sector. Thus prior to the first TESS science data release, we developed a deep neural network classification system to rapidly identify planet candidates or flag false positives. We generated raw TESS data using a physics-based instrument and sky simulator, Lilith, which models the full range of instrument behavior (e.g., camera optics, readout electronics, spacecraft jigger) and includes realistic instances of stellar astrophysics (e.g., transiting planets, eclipsing binaries, stellar variability). This simulated data was then processed through the TESS pipeline to produce complete archivable data products, which we used as ground-truth data to train our deep neutral net classification system and evaluate its performance characteristics. We also show transfer learning results from a neural network trained on 150,000 Kepler light curves, then fine-tuned on simulated TESS data, and evaluate the domain gap for our model between these two distinct data sources.

P1.1: Cherenkov Shower Detection Combining Probability Distributions from Convolutional Neural Networks.

Theme: Machine Learning in Astronomy
Author(s):
Mauricio Araya (UTFSM)
Francisco Casas (UTFSM)
Rodrigo Caceres (UTFSM)
Edson Carqun (UTFSM)

Contact: *mauricio.araya@usm.cl*

Ground-based gamma-ray observatories such as the Cherenkov Telescope Array presents new challenges for astronomical data analysis. The dynamics of the atmosphere and the complexity of Cherenkov shower are two uncertainty sources that needed to be embraced rather than corrected.

As each telescope only has access to a separated patch, the partial information of each one has to be combined. For instance, when blots can be identified on the images, the application of Hillas parameters allows to identify the approximate direction to the projection's center. This information can be combined for several telescopes using stereoscopic reconstruction to converge on a single point. The limitation of this technique however is that it performs regressions to a predefined blot shapes, not using all the information contained in the images. Thus, deep learning techniques based on Convolutional Neural Networks have been applied with promising results. However, they rely on very large networks that process all the telescope images at once, which might not scale properly when dealing with large arrays.

We propose to run several separate instances of an smaller network for each telescope, but that are able to retrieve a probability distribution instead of approximate coordinates for the sought point. This probability distribution can be arranged by the network so it can express certainty about the direction or the distance to the center of the projection separately. The distributions retrieved by all the telescopes can be combined to get a final probability distribution. Preliminary results shows the viability of this approach to identify the center and assign a confidence value to the result.

P10.1: ESA Science Archives and ESASky

Author(s):
Christophe Arviset (ESA-ESAC)
Deborah Baines (QUASAR for ESA)
Guido de Marchi (ESA)
Beatriz Martinez (RHEA for ESA)
Bruno Mern (ESA)
Juan Gonzalez, SERCO for ESA Elena Racero, SERCO for ESA Jesus Salgado, QUASAR for ESA

Contact: *Christophe.Arviset@esa.int*

The ESAC Science Data Centre (ESDC), located at ESAC, Madrid, Spain, hosts all science archives from more than 20 ESA space science missions, and various more are in preparation.

Data across the full electromagnetic spectrum can be obtained through individual astronomy archives for Gaia, XMM-Newton, Herschel, HST, Planck, ISO, EXOSAT and more recently from Lisa PathFinder as well. Furthermore, ESASky is an open science web application providing full access to the entire sky as observed by all ESA and partners astronomy missions.

The Planetary Science Archive offers a one-stop source for data from all ESA planetary missions (Rosetta, Mars Express, Exomars TGO, Venus Express, Huygens and Giotto). In addition, the heliophysics science archives provide access to data from Soho, Cluster, Ulysses, Proba-2, Double Start and ISS-SolACES.

P4.3: Utilizing Conda for Fermi Data Analysis Software Releases

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):

Joseph Asercion (Fermi Science Support Center/NASA GSFC)

Contact: *joseph.a.asercion@nasa.gov*

The Fermi Gamma-Ray Space Telescope mission provides, via the Fermi Science Support Center (FSSC), a suite of data analysis tools to assist the high energy astrophysics community in working with Fermi data. For many years these tools were distributed via both precompiled binaries and source tarball downloads on the FSSCs website. Due to the complexity of the tools and restrictions on development the downloads carried with them a large complement of third-party software which often caused package conflicts on users machines and bloated the size of the complete analysis package. To alleviate these problems the Fermi development team has decided to update the distribution pipeline to utilize the Conda package management system. This has allowed the development team to greatly reduce the software package size, eliminate a large category of bugs which once were prevalent, and target a decrease in software update turnaround/release time. In this talk, I will outline the process the development team took to convert our legacy codebase into a Conda compatible form and outline the lessons learned throughout this process.

P10.2: Baseline-dependent dimensional reduction techniques for radio interferometric big data compression

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Marcellin Atemkeng (Department of Physics and Electronics, Rhodes University)
Elhadji Moustapha SECK (African Institute for Mathematical Sciences, Cameroon)
Oleg Smirnov (SKA/Department of Physics and Electronics, Rhodes University)
Sphesihle Makhathini (SKA/Department of Physics and Electronics, Rhodes University)

Contact: *m.atemkeng@gmail.com*

Modern radio interferometers like MeerKAT, ASKAP, and LOFAR produce large amounts of data. The high data rates are driven by the fine time and frequency sampling of these instruments, as well as high angular resolution. For MeerKAT this amount of data is around anywhere from 64GB to 0.5 TB of row visibilities per second. The SKA will generate orders of magnitudes higher than this. Therefore, data compression is required to save memory and improve processing time. Offringa (2016) showed that lossy compression can compress LOFAR data by a factor that is greater than 5. A natural method to compress the data is through averaging, either baseline-dependent averaging or windowing as described in Atemkeng et al. (2018). Kartik et al. (2017) showed that the data can be compressed using a Fourier dimensionality reduction model (FDRM). This is in the gridded visibilities and not in the continuous visibilities. The gridded visibilities lie on a regular grid where data for all baselines have been interpolated together making it difficult to gauge an acceptable variance thresholding of the non-zero singular values (see Kartik et al. (2017)). Since each baseline sees the sky differently, decorrelation is baseline dependent and the noise variance is different per-visibility: these effects cannot be taken into account in the FDRM. Some applications (e.g. transient search, or to build up a wide-FoV interferometric archive) would require storing the raw data from all the baselines and not the gridded data. This work will studies the different algorithms in literature for big data dimensional reduction and apply them to visibilities data. The reduction should be baseline-dependent.

P5.1: DEAVI: Dynamic Evolution Added Value Interface

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Deborah Baines (Quasar Science Resources)
Ignacio de la Calle (Quasar Science Resources)
Jos Mara Herrera-Fernndez (Quasar Science Resources)
Aitor Ibarra (Quasar Science Resources)
Jess Salgado (Quasar Science Resources)
Luis Valero-Martin, Quasar Science Resources

Contact: *dbaines@sciops.esa.int*

We present DEAVI, an Added Value Interface (AVI) to manage and exploit data from the ESA missions Gaia and Herschel. AVIs are software packages that provide scientists with the mechanisms to submit their own code to be executed close to the ESA mission archives. GAIA AVIs are deployed at the Gaia Added Value Interface Platform (GAVIP), a Python-based platform designed and developed by ESA and hosted at the European Space Astronomy Centre (ESAC). The proposed AVI is part of the software package being developed by Quasar Science Resources for the StarFormMapper (SFM): A Gaia and Herschel Study of the Density Distribution and Evolution of Young Massive Star Clusters project, funded by the European Union under the Horizon 2020 programme.

P5.2: New Python developments to access CDS services

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Matthieu Baumann (CDS, Observatoire de Strasbourg)
Thomas Boch (CDS, Observatoire de Strasbourg)
Pierre Fernique (CDS, Observatoire de Strasbourg)

Contact: *matthieu.baumann@astro.unistra.fr*

We will present recent developments made in the frame of the ASTERICS project and aimed at providing Python interface to CDS services and Virtual Observatory standards. Special care has been taken to integrate these developments into the existing astropy/astroquery environment. A new astroquery.cds module allows one to retrieve image or catalogue datasets available in a given region of the sky described by a MOC (Multi Order Coverage map) object. Datasets can also be filtered through additional constraints on their metadata. The MOCPy library has been upgraded: performance has been greatly improved, unit tests and continuous integration have been added, and the integration of the core code into the astropy.regions module is under way. We have also added an experimental support for creation and manipulation of T-MOCs which describe the temporal coverage of a data collection.

P3.2: VisIVO Visual Analytics Tool an EOSC Science Demonstrator for data discovery

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Ugo Becciani (INAF - Istituto Nazionale di Astrofisica)
Fabio Vitello (INAF - Istituto Nazionale di Astrofisica)
Eva Sciacca (INAF - Istituto Nazionale di Astrofisica)
Alessandro Costa (INAF - Istituto Nazionale di Astrofisica)
Simone Riggi (INAF - Istituto Nazionale di Astrofisica)
Sergio Molinari, INAF - Istituto Nazionale di Astrofisica

Contact: *ugo.becciani@inaf.it*

VisIVO is an integrated suite of tools and services for data discovery that include collaborative portals, mobile applications, visual analytics tool and a number of key components such as workflow applications, analysis and data mining functionalities. Space missions and ground-based facilities produce massive volumes of data and the ability to collect and store them is increasing at a higher pace than the ability to analyze them. This gap leads to new challenges in the analysis pipeline to discover information contained in the data. VisIVO Visual analytics tool for star formation regions focuses on handling these massive and heterogeneous volumes of information accessing the data previously processed by data mining algorithms and advanced analysis techniques with highly interactive visual interfaces offering scientists the opportunity for in-depth understanding of massive, noisy, and high-dimensional data. The aforementioned challenges demands an increasing archiving and computing resources as well as a federated and interoperable virtual environment enabling collaboration and re-use of data and knowledge. Thus, the connection with the European Open Science Cloud is being investigated exploiting the EGI services such as the Check-in (for federated authentication and authorization) and the Federated Cloud (for analysis and archiving services). Recently the VisIVO development has been exploited for the experimentation of cutting-edge interactive visualization technologies for the improvement of teaching and scientific dissemination. This work is being carried out as a knowledge transfer from astrophysical sciences to geological sciences in the context of an international collaboration to innovate teaching, learning and dissemination of earth sciences, using virtual reality.

P5.3: Breathing New Life Into An Old Pipeline: Precision Radial Velocity Spectra of TESS Exoplanet Candidates

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
G. B. Berriman (Caltech/IPAC-NexScI)
D. Ciardi (Caltech/IPAC-NexScI)
B. J. Fulton (Caltech/IPAC-NexScI)
J. C. Good (Caltech/IPAC-NexScI)
M. Kong (Caltech/IPAC-NexScI)
H. Isaacson (U.C. Berkeley) J. Walawender (W. M. Keck Observatory)

Contact: *gbb@ipac.caltech.edu*

The High Resolution Echelle Spectrograph (HIRES) at the W.M. Keck Observatory (WMKO) is one of the most effective Precision Radial Velocity (PRV) machines available to U.S. astronomers, and will play a major role in radial-velocity follow-up observations of the tens of thousands of exoplanets expected to be discovered by the Transiting Exoplanet Sky Survey (TESS) mission. To support this community effort, the California Planet Search (CPS) team (Andrew Howard, PI) has made available a PRV reduction pipeline that will be available to all U.S. astronomers from February 2019 onwards. Operation of the pipeline has strict requirements on the manner in which observations are acquired, and these will be fully documented for users at the telescope.

The pipeline is written in IDL, and was developed over time for internal use by the CPS team in their local processing environment. Development of a modern version of this pipeline in Python is outside the scope of our resources, but it has been updated to support processing in a generic operations environment (e.g. changes to support multiple simultaneous users) We have developed a modern, Python interface to this updated pipeline, which will be accessible as a remote service hosted behind a firewall at the NASA Exoplanet Science Institute (NExScI). Users will be able to use Python clients to access data for input to the pipeline through the Keck Observatory Archive (KOA). The pipeline which will create calibrated and extracted 1D spectra and publication-ready time series, which can be visualized and analyzed on the client side using tools already available in Python. The Python client functions interface with the pipeline through a series of server-side web services. Users will have access to a workspace that will store reduced data and will remain active for the lifetime of the project. This design supports both reduction of data from a single night or long-term orbital monitoring campaigns.

The service is on-schedule for deployment in December 2018.

This project is a collaboration between NExScI, WMKO, KOA and CPS.

P6.1: Fundamentals of effective cloud management for the new NASA Astrophysics Data System

Theme: DevOps Practices in Astronomy Software

Author(s):
Sergi Blanco-Cuaresma (Harvard-Smithsonian Center for Astrophysics)
Alberto Accomazzi (Harvard-Smithsonian Center for Astrophysics)
Michael J. Kurtz (Harvard-Smithsonian Center for Astrophysics)
Edwin Henneken (Harvard-Smithsonian Center for Astrophysics)
Carolyn S. Grant (Harvard-Smithsonian Center for Astrophysics)
Donna M. Thompson, Roman Chyla, Steven McDonald, Golnaz Shapurian, Timothy W. Hostetler, Matthew R. Templeton, Kelly E. Lockhart.
    Affiliation: Harvard-Smithsonian Center for Astrophysics

Contact: *sblancocuaresma@cfa.harvard.edu*

The new NASA Astrophysics Data System (ADS) is designed as a service-oriented architecture (SOA) that consists of multiple customized Apache Solr search engine instances plus a collection of microservices, containerized using Docker, and deployed in Amazon Web Services (AWS). For complex systems, like the ADS, the loosely coupled architecture can lead to a more scalable, reliable and resilient system if some fundamental questions are addressed. After having experimented with different AWS environments and deployment methods, we decided in December 2017 to go with Kubernetes for our container orchestration. Defining the best strategy to properly set-up Kubernetes has shown to be challenging: automatic scaling services and load balancing traffic can lead to errors whose origin is difficult to identify, monitoring and logging the activity that happens across multiple layers for a single request needs to be carefully addressed, and the best workflow for a Continuous Integration and Delivery (CI/CD) system is not self-evident. This raises several fundamental questions: how do you update your service when there is a new release of a microservice? How do you troubleshoot issues, when requests follow a complex path through the architecture? We have been using Kubernetes for almost a year now, both in our development and production environments. This poster highlights some of our findings.

P10.3: Creating and managing very large HiPS: the PanSTARRS case

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Thomas Boch (CDS, Observatoire de Strasbourg)

Contact: *thomas.boch@astro.unistra.fr*

HiPS (Hierarchical Progressive Surveys) is a proven Virtual Observatory standard which enables an efficient way to deliver easily potentially huge images collection and allows for fast visualisation, exploration and science applications. CDS has recently published the HiPS for PanSTARRS g and z-bands images, covering three quarter of the sky at a resolution of 250mas per pixel. We will describe in our poster the challenges we faced and the lessons learnt in generating and distributing these HiPS made of 47 million FITS tiles, amounting to 10 trillion pixels and more than 20TB per band. In particular, we will detail the methods we developed to optimize the generation, the storage and the transfer of the HiPS. In addition, a color HiPS, based on the two already available HiPS, has been made available and can be visualised from HiPS clients, like Aladin Desktop or Aladin Lite.

P6.2: Executable user documentation for in-development software

Theme: DevOps Practices in Astronomy Software

Author(s):
Catherine Boisson (Observatoire de Paris)
Jose Enrique Ruiz (IAC)
Christoph Deil (MPIK)
Axel Donath (MPIK)
Bruno Khelifi (APC)
for the Gammapy team

Contact: *catherine.boisson@obspm.fr*

One key aspect of software development is feedback from users and beta-testers. This community is not always aware of the developments undertaken in the base-code, neither use the tools and practices followed by the developers to deal with issues related with a non-stable software in continuous evolution. The open-source Python package for gamma-ray astronomy, Gammapy, provides its beta-tester user community with versioned reproducible environments and executable documentation, based on Jupyter notebooks and virtual environment technologies from conda and Docker. We believe this set-up greatly improves the user experience for a software in prototyping phase, as well as communication with the user community. We will present such a setup where code and tutorial are versioned coupled.

P2.2: Towards a National Center for Optical and Infrared Astronomy: Opportunities and Challenges in Science, Software, and Data

Theme: Management of Large Science Project
Author(s):
Adam S. Bolton (NOAO/NCOA)

Contact: *bolton@noao.edu*

Following guidance from the US National Science Foundation (NSF), the present-day operating programs of the US National Optical Astronomy Observatory (NOAO) and the Gemini Observatory are being restructured along with the future operations of the Large Synoptic Survey Telescope into a single organization, currently known as the NSFs National Center for Optical and Infrared Astronomy (NCOA). NCOA will be the focal point for federal investment in open-access ground-based optical and infrared astronomy for the foreseeable future. I will give an overview of NCOA planning and implementation, with a focus on the new opportunities in science delivery, software, and data systems that this new organization will provide.

O11.3: ProvTAP: A TAP service for providing IVOA provenance metadata

Theme: Quality Assurance of Science Data

Author(s):
Franois Bonnarel (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR7550, F-67000 Strasbourg)
Mireille Louys (Universit de Strasbourg, CNRS, ICube, UMR 7357, F-67000 Strasbourg)
Mathieu Servillat (LUT, Observatoire de Paris, PSL Research University, CNRS, F-92190 Meudon)
Markus Nullmeier (Zentrum fr Astronomie der Universitt Heidelberg, Astronomisches Rechen-Institut D-69120 Heidelberg)
Michle Sanguillon (Laboratoire Univers et Particules de Montpellier, Universit de Montpellier, CNRS/IN2P3, France)
Anastasia Gaskin, Leibniz-Institut fr Astrophysik Potsdam (AIP), Germany Kristin Riebe, Leibniz-Institut fr Astrophysik Potsdam (AIP), Germany Ole Streicher, Leibniz-Institut fr Astrophysik Potsdam (AIP), Germany

Time: Mon 16.30 - 16.45

Contact: *francois.bonnarel@astro.unistra.fr*

In the astronomical Virtual Observatory, provenance metadata provide information on the processing history of the data. This is important to assert quality and truthfulness of the data, and to be potentially able to replay some of the processing steps.

The ProvTAP specification is a recently proposed IVOA Working draft defining how to serve IVOA provenance metadata via TAP, the Table Acces Protocol, which allows to query table and catalog services via the Astronomical Data Query Language (ADQL). ProvTAP services should allow finding out all activities, entities, or agents that fulfil certain conditions.

Several implementations and developments will be presented. The CDS ProvTAP service describes provenance metadata for HiPS generation. The CTA ProvTAP service will provide access to metadata describing the processing of CTA event lists. GAVO prototyped specialised query functions that could facilitate accomplishing the goals of ProvTAP users.

I4.1: Massive Data Exploration in Astronomy: What Does Cognitive Have To Do With It?

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Kirk Borne (Booz Allen Hamilton)

Time: Wed 14.00 - 14.30

Contact: *borne_kirk@bah.com*

There has been a tendency for astronomers to avoid unsupervised data exploration, due to the characterization of this approach as a non-scientific fishing expedition. But, a cognitive approach to massive data exploration has the potential to amplify hypothesis formulation and question generation for greater astronomical discovery. The incorporation of contextual data from other wavelengths and other surveys provides the basis for seeing interestingness in the multi-dimensional properties of sources that might otherwise appear uninteresting in a single survey database. Some suggested methods for cognitive exploration will be presented, including computer vision algorithms that are used in robotics to see patterns in the the world, but these can be used to see emergent patterns in the multi-dimensional parameter space of astronomical data.

I12.1: An Overview of the LSST Image Processing Pipelines

Theme: Algorithms
Author(s):
Jim Bosch (Princeton University / LSST)

Time: Mon 11.15 - 11.45

Contact: *jbosch@astro.princeton.edu*

In this talk, I'll walk through LSST's Data Release and Alert Production pipelines, highlighting how we produce various important datasets. I'll also call attention to the algorithms where LSST will need to push the state of the art or operate qualitatively differently from previous surveys.

O5.4: aflak: Pluggable Visual Programming Environment with Quick Feedback Loop Tuned for Multi-Spectral Astrophysical Observations

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Malik Olivier Boussejra (Keio University)
Shunya Takekawa (Nobeyama Radio Observatory)
Rikuo Uchiki (Keio University)
Kazuya Matsubayashi (Okayama Astrophysical Observatory)
Yuriko Takeshima (Tokyo University of Technology)
Makoto Uemura, Hiroshima University Issei Fujishiro, Keio University

Time: Thu 10.00 - 10.15

Contact: *malik@boussejra.com*

In the age of big data and data science, some may think that artificial intelligence would bring analytical solution to every problem. However, we argue that there is still ample room left for human insight and exploration thanks to visualization technologies. New discoveries are not made by AI (yet!). This is true in all scientific domains, including astrophysics. With the improvements of telescopes and proliferation of sky surveys there is always more data to analyze, but not so many astronomers. We present aflak, a visualization environment to open astronomical datasets and analyze them. This papers contribution lies in that we leverage visual programming techniques to conduct fine-grained, astronomical transformations, filtering and visual analyses on multi-spectral datasets with the possibility for the astronomers to interactively fine-tune all the interacting parameters. By visualizing the computed results in real time as the visual program is designed, aflak puts the astronomer in the loop, while managing data provenance at the same time.

O10.1: AstroCut: A cutout service for TESS full-frame image sets

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
C. E. Brasseur (Space Telescope Science Institute)
Rick White (Space Telescope Science Institute)
Jonathan Hargis (Space Telescope Science Institute)
Susan Mullally (Space Telescope Science Institute)
Scott Fleming (Space Telescope Science Institute)
Mike Fox - Space Telescope Science Institute Arfon Smith - Space Telescope Science Institute

Time: Tue 09.30 - 09.45

Contact: *cbrasseur@stsci.edu*

The Transiting Exoplanet Survey Satellite (TESS) launched this past March and will have its first data release near the end of this year. Like that of the Kepler mission, the TESS data pipeline will return a variety of data products, from light curves and target pixel files (TPFs) to large full frame images (FFIs). Unlike Kepler, which took FFIs relatively infrequently, TESS will be taking FFIs every half hour, making them a large and incredibly valuable scientific dataset. As part of the Mikulski Archive for Space Telescope's (MAST) mission to provide high quality access to astronomical datasets, MAST is building an image cutout service for TESS FFI images. Users can request image cutouts in the form of TESS pipeline compatible TPFs without needing to download the entire set of images (750 GB). For users who wish to have more direct control or who want to cutout every single star in the sky, the cutout software (python package) is publicly available and installable for local use.

In this talk we will present the use and design of this software, in particular how we were able to optimize the cutout step. The main barrier in writing performant TESS FFI cutout software is the number of files that must be opened and read from. To streamline the cutout process we performed a certain amount of one-time work up front, which allows individual cutouts to proceed much more efficiently. The one-time data manipulation work takes an entire sector of FFIs and builds one large (45 GB) cube file for each camera chip, so that the cutout software need not access several thousand FFIs individually. Additionally we transpose the image cube, putting time on the short axis, thus minimizing the number of seeks per cutout. By creating these data cubes up front we achieved a significant increase in performance. We will show examples of this tool using the currently available simulated TESS data, and discuss use cases for the first data release. We will finish by discussing future directions for this software, such as generalizing it beyond the TESS mission.

P4.4: Streamlining Pipeline Workflows: Using Python with an Object-Oriented Approach to Consolidate Aggregate Pipeline Processes

Theme: Data Science: Workflows Hardware Software Humanware

Author(s):
Matthew Brown (W. M. Keck Observatory)
Jeff Mader (W. M. Keck Observatory)
Josh Riley (W. M. Keck Observatory)
Luca Rizzi (W. M. Keck Observatory)
Bruce Berriman (NASA Exoplanet Science Institute)
Chris Gelino, Anastasia Laity, Melanie Swain, Mihseh Kong - NASA Exoplanet Science Institute

Contact: *mbrown@keck.hawaii.edu*

The Keck Observatory Archive (KOA), a collaboration between the NASA Exoplanet Science Institute and the W. M. Keck Observatory, serves science and calibration data for all current and retired instruments from the twin Keck Telescopes. In addition to the raw data, we publicly serve quick-look, reduced data products for four instruments (HIRES, LWS, NIRC2 and OSIRIS), so that KOA users can easily assess the quality and scientific content of the data. In this paper we present the modernization of the Data Evaluation and Processing (DEP) Pipeline, our quality assurance tool to ensure science data is ready for archival. Since there was no common infrastructure for data headers, the DEP pipeline had to evolve to accommodate new instruments through additional control paths each time an instrument or instrument upgrade was added. Over time, new modules to assist with the processing were added in a variety of languages including IDL, C, CSH, PHP, and Python. The calls to multiple interpreters caused a lot of overhead. This project was an initiative to consolidate the DEP pipeline into a common language, Python, using an object-oriented approach. The object-oriented approach allows us to abstract out the differences and use common variables in place of instrument-specific values. As a result, for new instruments one only needs to modify a new subclass with the differing values in order to work with the pipeline. By consolidating everything to Python, we have seen an increase in efficiency, ease of operation, and ease of maintenance.

O10.3: No-SQL databases: An efficient way to store and query heterogeneous astronomical data in DACE.

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Nicolas Buchschacher (University of Geneva)
Fabien Alesina (University of Geneva)
Julien Burnier (University of Geneva)

Time: Tue 10.00 - 10.15

Contact: *nicolas.buchschacher@unige.ch*

Data production is growing every day in all domains. Astronomy is particularly concerned with the recent instruments. While SQL databases have proven their performances for decades and still performs in many cases, it is sometimes difficult to store, analyse and combine data produced by different instruments which do not necessarily use the same data model. This is where No-SQL databases can help to solve our requirements: how to efficiently store heterogenous data in a common infrastructure ? SQL database management systems can do a lot of powerful operations like filtering, relation between tables, sub-queries etc. The storage is vertically scalable by adding more rows in the tables but the schema has to be very well defined. In the opposite, No-SQL databases are not restrictive. The scalability is horizontal by adding more shards (nodes) and the different storage engines have been designed to easily modify the structure. This is why it is well suited in the big data era. DACE (Data and Analysis Center for Exoplanets) is a web platform which facilitates data analysis and visualisation for the exoplanet research domain. We are collecting a lot of data from different instruments and we regularly need to adapt our database to accept new data sets with different models. We recently decided to do a major change in our infrastructure after using PostgreSQL to use CASSANDRA for the storage and Apache Solr as an indexer to do sophisticated queries among a huge number of parameters. This recent change accelerated our queries and we are now ready to accept new data sets from futur instruments and combine them with older data to do better science. DACE is funded by the Swiss National Centre of Competence in Research (NCCR) PlanetS, federating the Swiss expertise in exoplanet research.

P6.3: Development, tests and deployment of web application in DACE

Theme: DevOps Practices in Astronomy Software

Author(s):
Julien Burnier (Universit de Genve)
Nicolas Buschacher (Universit de Genve)
Fabien Alesina (Universit de Genve)

Contact: *julien.burnier@unige.ch*

The Data and Analysis Center for Exoplanets (DACE) is a web platform based at the University of Geneva (CH) dedicated to extrasolar planets data visualisation, exchange and analysis. This platform is based on web technologies using common programming languages like HTML and Javascript for the front-end and a Java REST API for the back-end. Over the last 6 months, the process to maintain, develop, test and deploy the applications has been dramatically improved to facilitate the maintenance and the integration of new features. The goal of such automation is to let more time to focus on development and reduce the duplicated work. To achieve this result, we migrated our Java application to the Maven software project management and added unit tests. We implemented a pipeline on GitLab which consists of executing the tests and deploy the application in a dev environment at every commit. The front-end side is then tested using the Selenium web browser automation to simulate the user - website interactions and compare the new results with the old ones. Once all the tests are validated, a manual action on the GitLab interface can be done to deploy the application on the official web site and we ensure the compatibility of the new features with the production version. We are currently working to have a very complete set of tests on both back and front end in order to remove the manual part of production deployment and to have a fully automated integration of our applications.

P12.1: The JWST Data Calibration Pipeline

Theme: Algorithms

Author(s):
Howard Bushouse (Space Telescope Science Institute)
Jonathan Eisenhamer (Space Telescope Science Institute)
James Davies (Space Telescope Science Institute)

Contact: *bushouse@stsci.edu*

STScI is developing the software systems that will provide routine calibration of the science data received from the James Webb Space Telescope (JWST). The processing uses an environment provided by a Python module called stpipe that provides many common services to each calibration step, relieving step developers from having to implement such functionality. The stpipe module provides common configuration handling, parameter validation and persistence, and I/O management. Individual steps are written as Python classes that can be invoked individually from within Python or from the stpipe command line. Any set of step classes can be configured into a pipeline, with stpipe handling the flow of data between steps. The stpipe environment includes the use of standard data models. The data models, defined using yaml schema, provide a means of validating the correct format of the data files presented to the pipeline, as well as presenting an abstract interface to isolate the calibration steps from details of how the data are stored on disk.

P4.5: Full Stack Data Science: Using Python to download, clean, analyze and visualize Gaia data

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Juan Luis Cano (Satellogic)
Roberto Izquierdo
Mariona Badenas

Contact: *juanlu001@gmail.com*

Python has become of the most important programming languages in the software industry and in Astronomy, specially thanks to its popularity in the Data Science field. However, with the explosion of the Big Data movement, the tooling became quite fragmented and one single language turned out to be insufficient: efficient data access has to be done through databases using SQL, all the Hadoop stack is written in Java or Scala, and modern visualization is usually developed in JavaScript. Even if some current solutions, such as PySpark, enable the use of Python for Big Data, the deployment is often complex and the workflow fragile. This complexity is an obstacle for scientists without formal training in Software Engineering and also for the general public, particularly in Astronomy.

In recent times, the Python ecosystem has evolved to catch up with the Big Data world, and new tools have appeared to process, analyze and visualize bigger than RAM datasets. On the one hand, libraries like numba allow Python to scale up by compiling a subset of the language to assembly code, and on the other hand projects like Dask are bringing distributed and out-of-core capabilities to the usual small data tools of the Python in Data Science stack, allowing easy reuse of existing codebases and easy scaling to tens to thousands of nodes.

In this talk we will showcase the power of these modern Python packages to perform Full Stack Data Science on the Gaia Data Release 2. We will demonstrate how to easily download Gaia data, how to process big amounts of data on a modest laptop, and how to scale that to a distributed cloud computing cluster to increase performance with minimal cost and code changes. We will also feature special Python libraries for Big Data visualization that avoid common visualization pitfalls. Lastly, we will justify how these tools simplify the access to Astronomical data in general, discuss the limitations of the approach and talk about future developments.

P11.1: Rectification and wavelength calibration of EMIR spectroscopic data with Python

Theme: Quality Assurance of Science Data
Author(s):
Nicols Cardiel (Universidad Complutense de Madrid, Spain)
Sergio Pascual (Universidad Complutense de Madrid, Spain)
Jess Gallego (Universidad Complutense de Madrid, Spain)
Cristina Cabello (Universidad Complutense de Madrid, Spain)
Francisco Garzn (Instituto de Astrofsica de Canarias, Spain)
Marc Balcells, Nieves Castro-Rodrguez, Lilian Domnguez-Palmero, Peter Hammersley, Nicolas Laporte, Lee R. Patrick, Roser Pell, Mercedes Prieto, Alina Streblyanska

Contact: *cardiel@ucm.es*

EMIR, the near-infrared camera and multi-object spectrograph operating in the spectral region from 0.9 to 2.5 microns, has been commissioned at the Nasmyth focus of the Gran Telescopio Canarias. One of the most outstanding capabilities of EMIR is its multi-object spectroscopic mode which, with the help of a robotic reconfigurable slit system, allows to take around 53 spectra simultaneously. This poster describes how important reduction steps, concerning image rectification and wavelength calibration, are performed with the help of PyEmir, the python code developed as part of the contribution of the Universidad Complutense de Madrid in this instrument.

P12.2: Jitter and readout sampling frequency impact on the Athena/X-IFU performance

Theme: Algorithms
Author(s):
M.Teresa Ceballos (Instituto de Fsica de Cantabria (CSIC-UC))
B. Cobo (Instituto de Fsica de Cantabria (CSIC-UC))
P.Peille (CNES)

Contact: *ceballos@ifca.unican.es*

The X-ray Observatory Athena is the mission selected by ESA to implement the science theme The Hot and Energetic Universe for L2 (the second Large-class mission in ESAs Cosmic Vision science programme).

One of the two X-ray detectors designed to be onboard Athena is X-IFU (X-ray Integral Field Unit), a cryogenic microcalorimeter based on Transition Edge Sensor (TES) technology that will provide spatially resolved high-resolution spectroscopy. X-IFU will be developed by an international consortium led by IRAP (PI), SRON (co-PI) and IAPS/INAF (co-PI) and involving ESA Member States, Japan and the United States.

X-ray photons absorbed by X-IFU detector generate intensity pulses that must be detected and reconstructed on-board to recover their energy, position and arrival time. The software prototype package (SIRENA) in development at IFCA/Spain contains a set of processing algorithms under study to get the best compromise between performance and availability of onboard computing resources.

Recently, the baseline for the processing algorithms has been defined selecting the optimal filtering for energy reconstruction and the Single Threshold Crossing as the triggering mechanism. This combined selection provides the best compromise results for the mission requirements, based on the analysis of the simulated data.

However, the non-perfect sampling of the pulses rising-edges could result in significant errors of the reconstructed energies with the standard optimal filtering algorithm. If not corrected properly, these errors could induce a prohibitive broadening of the energy resolution.

We present here the analysis of the magnitude of this effect and propose a correction. In addition, we evaluate the impact of a reduced readout sampling frequency in the energy resolution, once the jitter correction has been applied.

P9.7: An automatic data collection and analysis software for GRB studies and its result

Theme: Multi-Messenger Astronomy

Author(s):
Wei Chen (Institute of High Energy Physics, Chinese Academy of Sciences)
Weihua Lei (Huazhong University of Science and Technology)

Contact: *chenwei@ihep.ac.cn*

Gamma-ray bursts (GRBs), transient gamma rays from the sky, are currently the most luminous explosions in the universe. After its discovery in 1973, GRB has attracted more and more attentions due to its special properties, such as high energy, high redshift, and rich multi-band data. In this work, I mainly work on the research of central engine of GRB with its large collections of GRB multi-band data. I try to constrain the central engine models based on the statistical analysis of large GRB sample. As the first step, I develop Capella, an automatic data collection and statistical analysis code dedicated for GRB studies. We then collect large amounts of GRB data with Capella from all available public data archive. One of the outputs of the software is the signature of a newborn Black Hole (BH) from the collapse of a supra-massive millisecond magnetar. We put special focus on the central engine of magnetar. If magnetars are the central engines of some GRBs, then several evolutionary results of magnetars are expected: (1) the immediate collapse into a BH; (2) the collapse of a supra-massive magnetar into a BH after it spins down; and (3) a stable magnetar. From GRB samples, we find two candidates exhibiting evidence of a magnetar central engine at early afterglow and a newborn starving BH at late time. Our results suggest that some magnetars may indeed collapse into a BH, which provides direct support for the magnetar and BH central engine models. Our results suggest that some magnetars may indeed collapse into a BH, which provides direct support for the magnetar and BH central engine models.

P13.1: Binospec@MMT: a database-driven model of operations, from planning of observations to data reduction and archiving

Theme: Other
Author(s):
Igor Chilingarian (SAO)
Sean Moran (SAO)
Martin Paegert (SAO)
Daniel Fabricant (SAO)
Jan Kansky (SAO)
Warren Brown, SAO

Contact: *igor.chilingarian@cfa.harvard.edu*

Binospec is a new optical multi-object spectrograph operated at the f/5 focus of the 6.5-m converted MMT telescope at Mt.Hopkins, Arizona commissioned in Nov/2017. Here we describe a software system for the Binospec spectrograph driven by the PostgreSQL relational database that covers all stages of the instrument operations from the slit mask design to the data reduction pipeline and archiving and distribution of raw and reduced datasets. We use an interactive web based front end to design slit masks, which submits the configurations directly into the database, which then are sent to the mask cutting facility. When slit masks are installed into the spectrograph, the process is logged in the same database, which is also connected to the MMT telescope scheduler and Binospec instrument control software. All exposures collected with the instrument are recorded into the database in real time, which allows us to perform automated pipeline data reduction. Finally, the metadata of pipeline reduced products are ingested into the database and the products (along with raw data) are distributed to the PIs. We demonstrate that using currently available technologies and very limited manpower it is possible to build a complete database-driven software system similar to those deployed by major space missions like Chandra, XMM, and HST.

O3.2: An HDF5 Schema for SKA Scale Image Cube Visualization

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Angus Comrie (University of Cape Town)
Adrianna Pinska (University of Cape Town)
Rob Simmonds (University of Cape Town)
Russ Taylor (University of Cape Town)

Time: Mon 09.45 - 10.00

Contact: *accomrie@gmail.com*

In this paper, we describe work that has been performed to create an HDF5 schema to support the efficient visualization of image data cubes that will result from SKA Phase 1 and precursor observations. The schema has been developed in parallel to a prototype client-server visualization system, intended to serve as a testbed for ideas that will be implemented in replacements for the existing CyberSKA and CASA viewers.

Most astronomy image files are currently packaged using the FITS standard, however this has a number of shortcomings for very large images. The HDF5 technology suite provides a data model, file format, API, library, and tools, which are are all open and distributed without charge. This enables structured schemas to be created for different applications. We will show how these can be beneficial to packaging radio astronomy (RA) data. In particular, our interest is in supporting fast interactive visualization of data cubes that will be produced by the SKA telescope. Existing HDF5 schemas developed for RA data were unable to meet our requirements. The LOFAR HDF5 schema did not meet performance requirements, due to the approach of storing each 2D image plane in a separate group. The HDFITS schema serves as a starting point for an HDF5 schema that maintains round-trip compatibility with the FITS format, but lacks the additional structures required for pre-calculated and cached datasets. Therefore, we have created a new schema designed to suite our application, though this may be advantageous for other processing and analysis applications.

The schema is similar to that of HDFITS, but extensions have been added to support a number of features required for efficient visualization of large data sets. We will discuss these extensions and provide details on performance improvements with commonly used access patterns. We will also describe real-world performance when used with our prototype visualization system.

P6.4: GDL - GNU Data Language 1.0

Theme: DevOps Practices in Astronomy Software
Author(s):
Coulais, Alain (LERMA, Observatoire de Paris et CNRS)

The name and order is still in discussion in the team. (including Duvert, Arabas, Gastaud, Schellens, Flinois, Si Lounis)

Some may be in acknowledgements.

Contact: *alain.coulais@obspm.fr*

On behalf of the GDL team, we will be glad to use the occasion of ADASS 2018 to announce the 1.0 release of GDL - GNU Data Language. GDL is the free and open-source drop-in replacement for IDL - Interactive Data Language. The 1.0 release, coming out 15 years after Marc Schellens has first made GDL public, marks several milestones attesting to the maturity of the project. Among these - as compared with versions presented at the previous editions of ADASS - there are updates in the plotting functionalities, major advancements in the Windows version of GDL, rewritten support for IDL save files, multi-threaded performance improvements and additions targeting compatibility with newer versions of IDL (including new data types).

GDL 1.0 is also the first major release following the move of GDL development from SourceForge (thank you for the last 15 years of hospitality!) to Github. This relocation was an occasion for a major effort targeting developer-workflow-oriented automation which helps us reduce entry barriers for new contributions, ensure increasing code coverage with automatic tests, facilitate maintenance and user support as well as ensure result reproducibility and traceability. Several of the lessons learn, insights gained and remaining challenges related with the development of the continuous integration process for the project will be reported.

Finally, the presentation will summarize current availability of packaged versions of GDL, compatibility with popular IDL-written astronomy-related software, potential directions for the upcoming releases, and the key areas where GDL could benefit from further user feedback and new contributions.

P2.3: Triumphs and Challenges of the Astropy Project: Open Development of a Python Library for Astronomy

Theme: Management of Large Science Project

Author(s):
Steven M. Crawford (Space Telescope Science Institute)

on behalf of the Astropy Collaboration

Contact: *scrawford@stsci.edu*

The Astropy Project is a community composed of individuals and institutes that are interested in building an open source framework in Python for astronomy. Using collaborative development tools and following an open development philosophy, the project provides an ecosystem of interoperable packages. We describe the project infrastructure designed to facilitate collaboration on the development of open source software and provide an update of the most recent releases. The project has had a number of successes and challenges while being one of the largest collaboratively developed software package in astronomy. We discuss the aspects of the project that contributed to its success while also highlighting some of the challenges to the project and future developments.

P4.6: Astrocook: your spectral analysis recipe book, now with a GUI

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Guido Cupani (INAFOATs)
Giorgio Calderone (INAFOATs)
Stefano Cristiani (INAFOATs)
Giuliano Taffoni (INAFOATs)
Valentina D'Odorico (INAFOATs)
Paolo Di Marcantonio, INAFOATs

Contact: *guido.cupani@inaf.it*

Astrocook (http://github.com/DAS-OATs/astrocook) is a budding Python package for the analysis of quasar spectra. The project was first presented last year at ADASS and has since grown in scope and perspective with the development of a dedicated graphical user interface to control its multiple recipes for line detection, continuum determination and fitting of the absorption systems. As we will show, the Astrocook GUI is not only a user-friendly tool to access the packages functionalities, but it also provides a framework to solve the automation-vs-customization conundrum of spectral analysis. Through the GUI recipe editor, the users can now design their own workflows, tweak their parameters on test data, and launch them over extended datasets with no need for scripting. Such approach has been proved effective in fields where the human analysis fails (e.g. the assessment of completeness in the detection of the absorption systems) and is meant to set a standard for the next-generation spectrograph data analysis tools.

O1.4: A method to detect radio frequency interference based on convolutional neural networks

Theme: Machine Learning in Astronomy

Author(s):
Dai Cong (College of Information and Science Technology, Beijng Normal University)
Zuo Shifan (National Astronomical Observatories, Chinese Academy of Sciences)
Liu Wei (College of Information and Science Technology, Beijng Normal University)
Yu Xianchuan (College of Information and Science Technology, Beijng Normal University)
Chen Xuelei (National Astronomical Observatories, Chinese Academy of Sciences)
Zhu Meng College of Information and Science Technology, Beijng Normal University

Time: Tue 16.15 - 16.30

Contact: *daicong2014@163.com*

Along with the rapid development of telecommunication, radio frequency interference (RFI) generated from diverse human produced sources like electronic equipment, cell phones, GPS and so on can contaminate the weak radio band data. Therefore, RFI is an important challenge for radio astronomy. RFI detection can be regarded a special task of image segmentation. As for RFI signals, they appears in the form of point, vertical or horizontal lines. However, most existing convolution neural networks (CNNs) perform classification tasks, where the output is the single classification label of an image. The U-Net enables classification of each pixel within the image, which is suitable and competitive for image segmentation. Thus, in this paper, we implement the U-Net of 14 layers with framework of Keras to detect RFI signals. The U-Net can perform the classification task of clean signal and RFI. Also, the U-Net is a kind of extended CNN with symmetric architecture, which consists of a contracting path to capture context information and extract features and an expanding path to get precise localization. It extracts the features of RFI for learning RFI distribution pattern and then calculates the probability value of RFI for each pixel. Then we set a threshold to get the results flagged by RFI. We train the parameter of the U-Net with Tianlai data(A radio telescope-array, the observing time is from 20:15:45 to 24:18:45 on 27th of September 2016, the frequency is from 744MHz to 756MHz and the number of baseline is 18528). The experimental results show that, compared with the traditional RFI flagging method, this approach performs better with satisfying accuracy and takes into account the relationship between different baselines, which contributes to correctly and effectively flag RFI.

P7.2: CCSDSPy - Convenient Decoding of Binary Spacecraft Telemetry

Theme: Software for Solar Systems Astronomy

Author(s):
Daniel da Silva (NASA/GSFC, Trident Vantage Systems, Johns Hokpins University)

Contact: *dasilva@jhu.edu*

CCSDS is a low-level packet format used by many NASA and ESA missions for low-level telemetry, and often contains tightly packed bits to reduce downlink requirements. CCSDSPy is a python library to ease decoding of CCSDS data for satellite ground systems. The driving need of this library is to ease the management of difficult to decode telemetry containing fields that are not byte-aligned, odd-length, or otherwise annoying.

Poster preview: https://docs.google.com/presentation/d/1Ab3DaM_UwP4rSL_lESZCZdHsQmwp-WEPFoek-wFdCW4/edit?usp=sharing

P5.4: Gaia Photometric Science Alerts Data Flow

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
A. Delgado (Institute of Astronomy, University of Cambridge, UK)
S. Hodgkin (Institute of Astronomy, University of Cambridge, UK)
D.W. Evans (Institute of Astronomy, University of Cambridge, UK)
D.L. Harrison (Institute of Astronomy, University of Cambridge, UK)
G. Rixon (Institute of Astronomy, University of Cambridge, UK)
F. van Leeuwen Institute of Astronomy, University of Cambridge, UK; G. van Leeuwen Institute of Astronomy, University of Cambridge, UK; A. Yoldas Institute of Astronomy, University of Cambridge, UK;

Contact: *mdelgado@ast.cam.ac.uk*

Since the Gaia Photometric Science Alerts project started operations in 2014 with the goal of highlighting transient events detected by the Gaia mission, more than 6000 transients have been discovered or their existence confirmed. Gaia data is transmitted daily from the spacecraft, and received and processed by the Gaia Photometric Science Alerts System at the Institute of Astronomy (Cambridge, UK) to extract the transient events candidates. After data processing and filtering, the candidates are eyeballed using a web application where related information for every alert is displayed to assist in deciding whether the candidate should be announced and made publicly available in different formats and methods to the wider astronomical community. For supporting astronomers to improve their interactions with each other, and to prioritise and optimise the follow-up of the alerts, the Gaia Marshall web application provides contextual data to the transients. The Marshall acts as a meeting point to exchange information, and improve classifications. In this poster we present an outline of the Gaia alerts data flow, from the alert candidates extraction to the follow-up.

P10.4: Serving large scale survey data for ASKAP with SIA2 and SODA

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
James Dempsey (CSIRO)
Minh Huynh (CSIRO)

Contact: *james.dempsey@csiro.au*

The Australian SKA Pathfinder (ASKAP) telescope is a survey radio telescope that will produce image cubes of over 1 TB each in size. A key problem is how astronomers will work with these huge files. The CSIRO ASKAP Science Data Archive (CASDA) has taken the approach of allowing astronomers and their scripts to work with parts of the files as well as entire files. Both sub cubes and spectra can be extracted in-place for arbitrary subsets of the cubes. The Server-side Operations for Data Access (SODA) and Simple Image Access v2 (SIA2) virtual observatory protocols provide the foundations of these capabilities. I will review how we implemented these protocols and our planned future work in this area.

O12.3: GWCS - A General Approach to Astronomical World Coordinates

Theme: Algorithms
Author(s):

Nadia Dencheva (Space Telescope Science Institute)
Perry Greenfield (Space Telescope Science Institute)

Time: Mon 14.00 - 14.15

Contact: *dencheva@stsci.edu*

GWCS is a package for managing the World Coordinate System (WCS) of astronomical data. It takes a general approach to the problem of expressing arbitrary transformations by supporting a data model which includes the entire transformation pipeline from input coordinates (detector by default) to world coordinates (standard celestial coordinates or physical quantities). Transformations from the detector to a standard coordinate system are combined in a way which allows for easy manipulation of individual components. The framework handles discontinuous models (e.g. IFU data) and allows quantities that affect transforms to be treated as input coordinates (e.g. spectral order). It provides flexibility by allowing access to intermediate coordinate frames. The WCS object is serialized to a file using the language independent Advanced Scientific Data Format (ASDF).Alternatively the ASDF object can be encapsulated in a FITS extension. The package is written in python and is based on astropy. It is easy to extend by adding new models and coordinate systems.

P4.7: GMRT Archive Processing Project

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Shubhankar Deshpande (National Center for Radio Astrophysics (TIFR))
Rathin Desai (National Center for Radio Astrophysics (TIFR))
Archit Sakhadeo (National Center for Radio Astrophysics (TIFR))
Shadab Shaikh (National Center for Radio Astrophysics (TIFR))
Yogesh Wadadekar (National Center for Radio Astrophysics (TIFR))
C.H. Ishwara Chandra - National Center for Radio Astrophysics (TIFR) Huib Intema - Leiden University

Contact: *shubhand@cs.cmu.edu*

The GMRT Online Archive houses over 80 terabytes of interferometric observations obtained with the GMRT since the observatory began operating as a facility in 2002. The utility of this vast data archive, likely the largest of any Indian telescope, can be significantly enhanced if first look (and where possible, science ready) processed images can be made available to the user community. We have initiated a project to pipeline process GMRT images in the 150, 240, 325 and 610 MHz bands. The SPAM pipeline is being used for this purpose. A prototyping run has been successfully completed and the results are encouraging. The thousands of processed continuum images that we will produce will prove useful in studies of distant galaxy clusters, radio AGN, as well as nearby galaxies and star-forming regions. Besides the scientific returns, a uniform data processing pipeline run on a large volume of data can be used in interesting ways. For example, we will be able to measure various performance characteristics of the GMRT telescope and their dependence on waveband, time of day, RFI environment, backend, galactic latitude etc. in a systematic way. Since the SPAM pipeline also carries out direction dependent modeling of ionospheric phase errors, we will also be able to measure differential ionospheric phase delays over thousands of sightlines over the entire solar cycle to better understand the properties of the earth's ionosphere. A variety of data products such as calibrated UVFITS data, sky images, Hierarchical Progressive Survey (HiPS) images, PyBDSF catalogs, AIPS processing logs will be delivered to users via the GMRT online archive. Data products will be compatible with standard Virtual Observatory protocols.

O11.1: Adding Science Validation to the JWST Calibration Pipeline.

Theme: Quality Assurance of Science Data
Author(s):
Rosa Diaz (Space Telescope Science Institute)
Macarena Garcia Marin (ESA)

Time: Mon 16.00 - 16.15

Contact: *rdiaz@stsci.edu*

The JWST Calibration Pipeline is a set of steps separated into three main stages; looking to provide the best calibration for all JWST instruments, observing modes, and a wide range of science cases. Careful scientific validation and verification are necessary to determine consistency and quality of the data produced by the calibration pipeline. With this goal in mind, the scientist at STScI have supported validation testing for most of the major builds. Our experience with HST and the realization of the effort it would take to consistently and reliably test after each build, even after launch, motivated us to think about streamlining the process. We started building unit tests to verify that the calibration pipeline produced the expected results. However, the need for a more in-depth scientific validation of the wide range of science cases that will be observed by JWST requires a different strategy; one that not only validates the accuracy of the data but that also provides with reliable metrics for all science cases. We are working on defining a more complete set of science validation tests cases and simulated data that can be integrated within an automated building and testing framework; allowing full science verification and validation of the calibration pipeline in short time scales as well as quality assurance of the calibration products.

Archiving this goal has been an arduous task, not only limited by the state of development of the software and the availability of accurate data for testing, but also by the diversity of ideas coming from a large group of scientist from different teams, resources, and conflicting schedules. In this talk, I will present the integration of the science validation testing framework within the build process. I will also discuss the challenges we faced to make this possible, the steps we took, and how this work will help us support the development of the JWST Calibration Pipeline after launch.

O5.2: Astropy and the Virtual Observatory

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):

Tom Donaldson (Space Telescope Science Institute)

Time: Wed 16.45 - 17.00

Contact: *tdonaldson@stsci.edu*

The International Virtual Observatory Alliance (IVOA) has been defining standards for interoperable astronomical data exchange since 2002. Many of these standards are being used successfully and extensively by archives and end user tools to enable data discovery and access. Nevertheless a skepticism persists in parts of the community about the utility and even relevance of these standards, as well as the processes by which they were written. By contrast, the Astropy Project, with its very different processes (and somewhat different goals), has been widely embraced by the community for the usefulness and usability of its interoperable Python packages. In this talk I will discuss what these projects might learn from each other, and how more collaboration might benefit both projects and the community in general.

P6.5: Automating Multimission Access: rolling out a flexible Virtual Observatory-based infrastructure

Theme: DevOps Practices in Astronomy Software

Author(s):
Theresa Dower (STScI)
Bernie Shiao (STScI)

Contact: *dower@stsci.edu*

As a part of The Mikulski Archive for Space Telescopes' (MAST) mission to develop and support user-friendly and scientifically useful search tools and foster inter-archive communication and interoperable standards, we integrate International Virtual Observatory (IVOA) standards in our infrastructure. These standard practices allow us to provide layers of simple and complex interfaces supporting both direct scientific research and inter-archive pipeline coordination. Here we discuss an architecture based on Table Access Protocol (TAP) services for MAST holdings, how we are automating the deployment and maintenance of these services to provide more data coverage with modern research software including Astropy, and how we can update these pieces individually to continue meeting unique technological challenges.

P10.5: Archive-2.0: Metadata and Data Synchronisation between MAST, CADC, and ESAC

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Patrick Dowler (National Research Council Canada)
Mara Arvalo (European Space Astronomy Centre)
Javier Durn (European Space Astronomy Centre)
Daniel Durand (National Research Council Canada)
Sverin Gaudet (National Research Council Canada)
Jonathan Hargis, Brian McLean, Oliver Oberdorf, David Rodriguez (Mikulski Archives for Space Telescopes) Brian Major (National Research Council Canada)

Contact: *pdowler.cadc@gmail.com*

The Canadian Astronomy Data Centre (CADC) and the European Space Astronomy Centre (ESAC) maintain mirrors of and provide user access to the HST data collection. A new mirroring approach was needed to improve consistency and support future missions like JWST. The Common Archive Observation Model (CAOM) is used as the core model for all data holdings at the CADC and the Mikulski Archive for Space Telescopes (MAST) and was extended to support a metadata and data synchronization system that allows the partners to maintain a complete copy of the entire HST collection using metadata and files generated at MAST.

The metadata synchronization process relies on a simple RESTful web service operated by the metadata source (MAST) and a metadata harvesting tool run by the mirror centres (CADC and ESAC). The harvesting tool normally operates in incremental mode (recent changes) to maintain an up-to-date copy of the metadata. Consistency of the metadata is insured through the use of a robust metadata checksum algorithm; a full validation mode can be used to check and fix cases where incremental harvest events or deleted observation events were missed (rare) or source and destination metadata checksums do not match.

We have also developed a new file synchronization tool that leverages CAOM metadata to discover and retrieve files from the source (MAST) to the mirror sites. Through the use a backend plugins, the CADC and ESAC have extended the file synchronization to interface with their respective site-specific storage systems.. Like the metadata harvesting tool, file synchronization normally operates in incremental mode: it uses local CAOM metadata to discover new or modified files and schedule downloads. A separate mode performs downloads. A validation mode performs a full comparison of files referenced in CAOM with those in the local storage system and (optionally) schedules downloads to fix any discrepancies.

Apart from the common data collection, services and tools described above, partners are allowed to extend CAOM metadata with more information specifically intended to provide added value features. ESAC, for instance, introduced information about publications in their instance of the Archive.

P1.2: Deep-Learning Continuous Gravitational Waves

Theme: Machine Learning in Astronomy
Author(s):
Christoph Dreissigacker (Albert Einstein Institute Hannover)
Rahul Sharma (Albert Einstein Institute Hannover)
Reinhard Prix (Albert Einstein Institute Hannover)
Christopher Messenger (SUPA, School of Physics & Astronomy, University of Glasgow)

Contact: *christoph.dreissigacker@aei.mpg.de*

The search for continuous gravitational waves from unknown spinning neutron stars presents an open computational challenge: optimal fully-coherent matched filtering is computationally impossible, and empirical semi-coherent methods are the best current alternative known. There has been promising progress recently in applying Deep Convolutional Neural Networks as a detection method for binary black-hole coalescence signals (George&Huerta, Gabbard et al, Gebhard et al (2017)). Here we present results of our study on the feasibility and potential of using similar networks to search for continuous gravitational waves directly in the detector strain data.

P9.8: CALET Gamma-ray Burst Monitor web-analysis system

Theme: Multi-Messenger Astronomy
Author(s):
Ken Ebisawa (ISAS/JAXA)
Satoshi Nakahira (RIKEN)
Takanori Sakamoto (Aoyama Gakuin University)
Atsumasa Yoshida (Aoyama Gakuin University)

Contact: *ebisawa@isas.jaxa.jp*

CALET (CALorimetric Electron Telescope) has been installed on the Japanese Experiment Module Exposed Facility of the International Space Station (ISS) since August 2015. CALET carries two observational instruments, the CALorimeter (CAL), whose main target is cosmic-rays, and the CALET Gamma-ray Burst Monitor (CGBM). CGBM is composed of a single Soft Gamma-ray Monitor (SGM) and two Hard X-ray Monitors (HXM). Their energy bands are 100 keV – 20 MeV (SGM) or 7 keV – 1 MeV (HXM), and the fields of view are one steradian (HXM) or the whole sky beside earth (SGM). The CGBM data are released from DARTS (http://darts.isas.jaxa/astro/calet) in the standard FITS format soon after the pipe-line processing. So that global users can easily analyze CGBM data without prior knowledge, we are developing a user-friendly CGBM web-tool (http://darts.isas.jaxa.jp/astro/calet/cgbmweb/LCViewer.html). Using this tool, users can display light-curves from different instruments in different energy bands, visualize the location of CALET (ISS) at the time of observation, and the fields of view on the sky. Users can interactively change scales of the light-curves, choose time-intervals to extract the burst and the background energy spectra. These spectral files can be downloaded as well as the instrument responses, so that users can carry out spectral model fitting analysis using XSPEC.

P8.2: Prototype Implementation of a Web-Based Gravitational Wave Signal Analyzer: SNEGRAF

Theme: Time Domain Astronomy
Author(s):
Satoshi Eguchi (Fukuoka University)
Shota Shibagaki (Fukuoka University)
Kazuhiro Hayama (Fukuoka University)
Kei Kotake (Fukuoka University)

Contact: *satoshieguchi@fukuoka-u.ac.jp*

A direct detection of gravitational waves is one of the most exciting frontiers for modern astronomy and astrophysics. Gravitational wave signals combined with classical electro-magnetic observations, known as multi-messenger astronomy, promise newer and deeper insights about the cosmic evolution of astrophysical objects such as neutron starts and black holes. To this end, we have been developing an original data processing pipeline for KAGRA, a Japanese gravitational wave telescope, for optimal detections of supernova events. As a part of our project, we have just released a web application named SuperNova Event Gravitational-wave-display in Fukuoka (SNEGRAF) in this autumn. SNEGRAF accepts the users' theoretical waveforms as a plain text file consisting of a time series of h+ and hx (the plus and cross mode of gravitational waves, respectively), then displays the input, a corresponding spectrogram, and power spectrum together with KAGRA sensitivity curve and the signal-to-noise ratio; we adopt Google Visualization API for the interactive visualization of the input waveforms. However, it is a time-consuming task to draw more than 10^5 data points directly with JavaScript, although the number can be typical for a supernova hunt by assuming a typical duration of the event and sampling rate of the detectors; a combination of recursive decimations of the original in the server-side program and an appropriate selection of them depending on the time duration requested by the user in a web browser achieves an acceptable latency. In this poster, we present the current design, implementation and optimization algorithms of SNEGRAF, and its future perspectives.

P13.2: JWST Association Generation: Piecing It All Together

Theme: Other
Author(s):
Jonathan Eisenhamer (AURA/STScI)

Contact: *eisenhamer@stsci.edu*

The basic unit of JWST data products is an exposure for each science instrument detector. Relationships between multiple exposures are captured in an association. An association is a means of identifying a set of exposures that belong together and may be dependent upon one another. JWST associations are created by the Association Generator. Using a set of association rules, the Generator filters through an Association Pool, basically a list of exposures and related meta-data, to produce Associations. These associations are then used as input to the appropriate calibration pipelines.

P5.5: CASA, the Common Astronomy Software Applications for Radio Astronomy

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):

Bjorn Emonts (National Radio Astronomy Observatory)

Contact: *bjornemonts@gmail.com*

CASA, the Common Astronomy Software Applications package, is the primary data-reduction and -analysis software for the Atacama Large Millimeter/submillimeter Array (ALMA) and the Karl G. Jansky Very Large Array (VLA), and is frequently used also for other radio telescopes. The package can process both interferometric and single dish data, and one of its core functionalities is to support the ALMA and VLA pipelines. The CASA infrastructure consists of a set of C++ tools bundled together under an iPython interface as data reduction and analysis tasks, which are easy to use by the astronomical community. The National Radio Astronomy Observatory (NRAO) guides an international team of developers and scientists from various institutions, who maintain and continuously improve the CASA software. This poster will give you an overview of the current status of the CASA software, and upcoming changes that the CASA team plans for the near future.

P12.3: A simple survey for cross-matching method

Theme: Algorithms
Author(s):
Dongwei Fan (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *fandongwei@nao.cas.cn*

In order to find a practicable method to build an online cross-matching service, we test several index and search methods. Indexing methods includes the directly matching, HEALPix based matching and zones algorithm. Sorted list and several search trees are also be inspected, e.g., Binary Search Tree, Red-Black Tree, B-Tree. From this survey, we can see that HEALPix based Binary Search on sorted array is the fastest and simple way to cross-match in memory, and the environment is easy to be restored from hardisk. If there are two unsorted catalog on disk, a red-black balance tree with HEALPix indices would be a good choice. But when the catalog is too big to cache in memory, the memory-hardisk-swapping much slows down the efficiency. The key is keeping more points in the memory and do the Binary Search. Not only the speed to cross-match, how to efficiently exporting the rest data columns in catalogs are also be considered. A cross-match web service build on these method is released.

P12.4: Extragalactic stellar photometry and the blending problem

Theme: Algorithms
Author(s):
Feinstein, Carlos (IALP/Conicet- FCAG/UNLP)
Baume, Gustavo (IALP/Conicet- FCAG/UNLP)
Mara Jimena, Rodriguez (IALP/Conicet- FCAG/UNLP)
Mara Marcela Vergne (IALP/Conicet- FCAG/UNLP)

Contact: *cfeinstein@fcaglp.unlp.edu.ar*

The images provided by the Advanced Camera for Surveys at the Hubble Space Telescope (ACS/HST) has the amazing spacial resolution of 0.05/pixel. Therefore, it could resolve individual stars in nearby galaxies and in particular young blue stars in associations and open clusters of the recent starburts. These data are useful for studies of the extragalactic young population using color magnitude diagrams (CMD) of the stellar groups. However, even with the excellent indicated spatial resolution, the blending of several stars in crowded fields can change the shape of the CMDs. Some of the blending could be handled in the cases they produce particular features on the stellar PSF profile (e.g. abnormal sharpness, roundness, etc). But in some cases, the blend could be difficult to detect, this is the case were a pair or several stars are in the same line of sight (e.g. observed in the same pixel). In this work, we investigated the importance of the blending effect in several crowded regions, using both numerical simulations and real ACS/HST data. In particular, we evaluated the influence of this effect over the CMDs, luminosity functions (LFs) and reddening estimations obtained from the observations.

P5.6: Data Analysis Tools for JWST and Beyond

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Henry Ferguson (Space Telescope Science Institute)
on behalf of the JWST Data Analysis Development Forum

Contact: *ferguson@stsci.edu*

Data Analysis tools are essential for transforming data into knowledge. They are distinct from pipeline tools in that they usually require interaction and tailoring to specific scientific needs, but parts of the analysis process can often be turned into a pipeline after an initial exploratory phase. Data analysis tools for JWST are being added to the open-source Python+Astropy, either as complete packages or as contributions to new packages. The tools include libraries to manipulate and transport complex geometric transformations (gWCS and ASDF), libraries for image analysis (imexam and photutils) and tools for analyzing spectroscopy (Specviz, MOSviz and Cubeviz). This talk will provide an overview of the tools, highlighting areas that are ripe for collaboration, and a brief summary of the agile development process.

P8.3: Time in Aladin

Theme: Time Domain Astronomy

Author(s):
Pierre Fernique (Observatoire Astronomique de Strasbourg - CDS)
Daniel Durand (National Research Council Canada - CADC)
Ada Nebot (Observatoire Astronomique de Strasbourg - CDS)

Contact: *Pierre.Fernique@astro.unistra.fr*

We present recent Aladins developments designed to handle and display the astronomical time dimension. Aladin was originally dedicated to visualise astronomical data in terms of spatial coordinates. Based on the same technology, we have incorporated a new dimension in Aladin: the time. This new Aladin prototype, based on the core of Version 10, incorporates two new components: a time view window and a time coverage capability. The time view window is a simple Aladin extension to its regular window originally designed to handle longitude VS latitude graphics. This new graphic method is dedicated to draw scatter plots where the primary axis is time and the secondary axis can be magnitude, flux, radial velocity, etc. The original spatial view and the new time view are fully interoperable allowing the users to select objects in either views to see them selected in all views simultaneously. The time coverage capability is based on the technology supporting the Multi-Ordered Coverage (MOCs), replacing the HEALPix space discretisation with a time scale instead. Thus the way the user manipulates time coverage is similar to space coverage manipulation, for instance performing fast coverage intersections or unions, generating a coverage from a list of sources, etc. These new capabilities are already available in the Aladin Beta version available on the Aladin CDS Web site.

O5.1: The NOAO Data Lab: Design, Capabilities and Community Development

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Michael Fitzpatrick (NOAO)
Knut Olsen (NOAO)

Glenn Eychaner (NOAO), Leah Fulmer (NOAO), Lijuan Huang (NOAO), Stephanie Juneau (NOAO), David Nidever (NOAO), Robert Nikutta (NOAO), Adam Scott (NOAO)

Time: Wed 16.30 - 16.45

Contact: *fitz@noao.edu*

We describe the NOAO Data Lab, a new science platform to efficiently utilize catalog, image and spectral data from large surveys in the era of LSST. Data Lab provides access (through multiple interfaces) to many current NOAO, public survey and external datasets to efficiently combine traditional telescope image/spectral data with external archives, share results and workflows with collaborators, experiment with analysis toolkits and publish science-ready results for community use. The architecture, science use-case approach to designing the system, its current capabilities and plans for community-based development of analysis tools and services are presented. Lessons learned in building and operating a science platform, challenges to interoperability with emerging platforms, and scalability issues for Big Data science are also discussed.

P13.3: The COSPAR Capacity Building Initiative: entering a new phase

Theme: Other
Author(s):
Carlos Gabriel (ESA-ESAC)

Contact: *carlos.gabriel@sciops.esa.int*

The Capacity Building Programme is considered today one of the flagships of COSPAR (COmmittee for SPace Research) activities. It started in 2001 as a tentative project to widening expertise in space science and promoting the use of data archives from space missions, particularly in astronomy, in developing countries. In the last 17 years we have held 15 workshops in astronomy, teaching advanced students and young researchers in developing countries how to analyse data from diverse space missions like XMM-Newton, Chandra, Swift, Hubble, Fuse, Gallex, Rosetta, Mars Express, Fermi, Suzaku, Herschel, Spitzer, NuStar, as well as related science. A first period of settling down the Programme went on for 8-9 years, in which the initial concepts were refined, the usefulness of the workshops confirmed and an associated Fellowship created. A second period, marked by a continuous expansion followed, this time with a Panel heading the Programme, in which the different main space science disciplines were represented. This year we have had a period of reflection, concluding in the COSPAR General Assembly in July, about what can be done better, what is still missing in the Programme, and how we could use the impending Panel reorganisation to renew and expand our objectives, initiating a new era of the CBP. In this paper I will discuss the main decisions taken about: a) ways to better evaluate the impact of the Programme; b) a larger interaction between the diverse space disciplines represented in it; and c) an extension of the Programme with the aim of motivating younger students to move in the direction of space sciences in developing countries.

P6.6: Agile meets Requirements

Theme: DevOps Practices in Astronomy Software
Author(s):
Julio Gallegos (Serco for ESA)
Fernando Perez (RHEA for ESA)
Mauro Casale (ESA)

Contact: *julio.gallegos@sciops.esa.int*

Initially, the Science Ground Segment development for the BepiColombo mission; a joined venture to explore Mercury with two separate spacecraft ESAs Mercury Planetary Orbiter and JAXAs Mercury Magnetospheric Orbiter; followed a requirements flow-down approach, i.e. a V-model; from the Science Implementation Requirements Document (SIRD) to System to Sub-system requirements, split into Launch and Mercury systems; each system with iterative versions every six months. After two years into the development, it was decided to introduce Agile methodology for the development of the iterative versions. The decided approach was to have 6 sprints of one month each with a final formal release at the end of the six-sprint period. This transition required the implementation of user-stories to convey the requirements to the development team; the use of a continuous integration/continuous deployment environment and the automation of several tasks. This paper describes the drivers for the transition to a Scrum development, its advantages and drawbacks; and the full traceability from the (SIRD) to the validation of the product, through system and subsystem requirements, user-stories and test-cases; to maintain information consistency. We consider this a good example of the applicability and effectiveness of an Agile-Scrum development methodology to a long-life development cycle program.

P13.4: MIRISim: the JWST-MIRI simulator

Theme: Other
Author(s):
Vincent Geers (UK Astronomy Technology Centre - STFC - UKRI, Edinburgh, United Kingdom)
MIRI European Consortium (jwst-miri.roe.ac.uk)

Contact: *vincent.geers@stfc.ac.uk*

MIRISim is the simulator package for the JWST Mid-Infrared Instrument (MIRI), created by the MIRI European Consortium. MIRISim is designed to simulate photon propagation through MIRI, and delivers detector images consistent with the expected on-orbit performance. The simulated data have the same uncalibrated data format that will be made available to JWST observers, and include all metadata required for processing with the JWST calibration pipeline (under development by STScI). MIRISim is written in Python 3 and released as part of an Anaconda environment that is publicly available at www.miricle.org. We present an overview of MIRISim together with example simulations.

P13.5: Preparing for JWST: a detailed simulation of a MOS deep field with NIRSpec

Theme: Other
Author(s):
Giovanna Giardino (ATG for European Space Agency (ESA), ESTEC, Nethrelands)
Pierre Ferruit (European Space Agency (ESA), ESTEC, Nethrelands)
Jacopo Chevallard (Institut dAstrophysique de Paris, France)
Emma Curtis-Lake (Institut dAstrophysique de Paris, France)

Nina Bonaventura, Aurelien Jarno , Peter Jakobsen, Arlette Pecontal, Laure Piqeras, and the JADES collaboration

Contact: *Giovanna.Giardino@esa.int*

JWST/NIRSpec will be the first multi-object spectrograph (MOS) to fly in space and it will enable the simultaneous measurement of up to 200 spectra over the wavelength range 0.6-5.0 micron, allowing us to study the rest-frame optical properties of large samples of galaxies out to z 9, and the rest-frame UV out to z10. This powerful instrument mode, however, requires careful planning of the observations and good understanding of the processing steps necessary to go from the detectors' count-rate images to background subtracted, calibrated spectra.

To support the community in preparing NIRSpec MOS programs and getting ready to analyze the data, we present here a set of simulations closely mimicking the deep spectroscopic observations that will be performed as part of the JADES survey, a joint effort of the NIRCam and NIRSpec GTO teams. The simulations are made possible by the NIRSpec Instrument Performance Simulator software. This tool consists of two main components: a Fourier Optics wave propagation module coupled with a detailed model of the instruments optical geometry and radiometric response and a detector module reproducing the noise properties and response of NIRSpec's two H2RG sensors. The targets for the simulations were selected from the JWST Extragalactic Mock Catalog, JAGUAR (Williams et al. 2018).

The simulation data package delivered here include more than 60 count-rate images corresponding to the exposures break-down of the low and medium resolution part of one of the two NIRSpec deep-field spectroscopic programs of the JADES survey. The simulated data consists of three dither pointings, for 4 different instrument configurations (low and medium resolution over the entire NIRSpec wavelength range), plus the extracted, background subtracted, spectral traces for each of the 370 targets and corresponding 2D-rectified spectra and calibrated 1D spectra, as well as the mock astronomical data used as the simulation input.

O1.3: Importance of Feature Selection in ML models

Theme: Machine Learning in Astronomy
Author(s):

Sankalp Gilda (University of Florida)

Time: Tue 16.00 - 16.15

Contact: *s.gilda@ufl.edu*

Importance of Feature Selection in ML models

An ever looming threat to astronomical applications of ML, and especially DL, is the danger of overfitting data. In particular, we refer to the problem of stellar parameterization from low-mid resolution spectra. The preferred method to deal with this issue is to develop and use spectral indices - this requires careful measurements of equivalent widths of blended spectral lines. This is prone to use error, and does not often result in very accurate results wrt the output parameters. In this work, we tackle this problem using an iterative ML algorithm to sequentially prune redundant features (wavelength points) to arrive at an optimal set of features with the strongest correlation with each of the output variables (stellar parameters) - T_{eff}, $\log(g)$ and [Fe/H]. We find that even at high resolution with tens of thousands of pixels (wavelength values), most of them are not only redundant, but actually decrease the mean absolute errors (MAEs) of the model output wrt the true values of the parameters. Our results are particularly significant in this era of exploding astronomical observational capabilities, when we will undoubtedly be faced with the 'curse of dimensionality'. We illustrate the importance of feature selection to reduce noise, improve model predictions, and best utilize limited computational and hardware resources on various downsampled and degraded synthetic PHOENIX spectra, by convolving the raw high res (500,000) sources to low and mid res (2,000 - 15,000).

O10.6: Driving Gaia Science from the ESA Archive: DR2 to DR3

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
J. Gonzlez-Nez (European Space Astronomy Centre - ESAC - ETSET Universidade de Vigo)
J. Salgado (European Space Astronomy Centre - ESAC)
R. Gutirrez-Snchez (European Space Astronomy Centre - ESAC)
J.C. Segovia (European Space Astronomy Centre - ESAC)
J. Durn (European Space Astronomy Centre - ESAC)
E. Racero (ESAC), J. Osinde (ESAC), P. De Teodoro (ESAC), A. Mora (ESAC), J. Bakker (ESAC), B. Mern (ESAC), C. Arviset (ESAC), F. Aguado-Agelet (U. Vigo)

Time: Wed 09.30 - 09.45

Contact: *jgonzale@sciops.esa.int*

Released 25th April, Gaia DR2 hosted in the ESA Gaia archive is leading a paradigm shift in the way astronomers access and process astronomical data in ESA archives.

An unprecedented active community of thousands of scientists is making use of the latest IVOA protocols and services (TAP, DataLink) in this archive, benefitting of remote execution and persistent, authenticated, server side services to speed up data exploration and analysis. The availability of a dedicated Python library for this purpose is connecting the archive data to new data processing workflows.

The infrastructure serving this data has been upgraded from DR1, now making use of replication, clustering, high performance hardware and scalable data distribution systems in new ways for ESA astronomical archives. VO orientation of the archive has been strengthened by the provision of Time Series in DR2 through use of a VO aware format and protocol.

In order to cover the overwhelming data volume of DR3, new services will be offered to the general astronomical community. Remote execution of code, with notebook services and access to data mining infrastructure as a service are the topics under development.

In this talk, it will be described how the current archive does enable to analyse Gaia data more effectively linked to how this is changing data analysis workflows. The infrastructure created for this purpose will be described, and the architecture and plans under implementation for DR3.

F4: Image Processing in Python With Montage

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
John Good (Caltech/IPAC-NExScI)
G. Bruce Berriman (Caltech/IPAC-NExScI)

Time: Tue 15.30 - 16.00

Contact: *jcg@ipac.caltech.edu*

The Montage image mosaic engine (http://montage.ipac.caltech.edu; https://github.com/Caltech-IPAC/Montage) has found wide applicability in astronomy research, integration into processing environments, and is an examplar application for the development of advanced cyber-infrastructure. It is written in C to provide performance and portability. Linking C/C++ libraries to the Python kernel at run time as binary extensions allows them to run under Python at compiled speeds and enables users to take advantage of all the functionality in Python. We have built Python binary extensions of the 59 ANSI-C modules that make up version 5 of the Montage toolkit. This has involved a turning the code into a C library, with driver code fully separated to reproduce the calling sequence of the command-line tools; and then adding Python and C linkage code with the Cython library, which acts as a bridge between general C libraries and the Python interface.

We will demonstrate how to use these Python binary extensions to perform image processing, including reprojecting and resampling images, rectifying background emission to a common level, creation of image mosaics that preserve the calibration and astrometric fidelity of the input images, creating visualizations with an adaptive stretch algorithm, processing HEALPix images, and analyzing and managing image metadata.

The material presented here will be made freely available as a set of Jupyter notebooks posted on the Montage GitHub page.

P5.7: The Present State of XSPEC and CFITSIO, Astronomical Analysis Packages Maintained by NASA's HEASARC.

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Craig A. Gordon (Innovim, NASA/Goddard Space Flight Center)
Keith A. Arnaud (CRESST/UMD/GSFC)
William Pence (NASA/Goddard Space Flight Center)

Contact: *craig.a.gordon@nasa.gov*

We discuss the present state of XSPEC and CFITSIO, two of the more widely used packages in HEASOFT, the software suite maintained and distributed by the HEASARC at NASA/Goddard Space Flight Center. XSPEC is an interactive X-ray spectral fitting program that has been used in almost 10,000 refereed publications over the last three and a half decades, and underwent a full redesign in the 2000s to object-oriented C++. We examine its modular design which allows for the expansion of such capabilities as new fitting statistics, plotting packages, and models (including those supplied by the end-user). We also show how originally unanticipated features were incorporated into the design, such as parallel processing and an extension module providing full access to XSPEC from Python (PyXspec). CFITSIO is a C and Fortran library that is used in much astronomical software for performing fast operations on FITS files. It has recently undergone a transition in support from its original author to the HEASARC programming staff, and we discuss challenges this entails and also some of its recent enhancements.

P12.5: Astrophysical codes migration into Exascale Era

Theme: Algorithms
Author(s):
David Goz (INAF - OATs)
Sara Bertocco (INAF - OATs)
Luca Tornatore (INAF - OATs)
Giuliano Taffoni (INAF - OATs)

Contact: *david.goz@inaf.it*

The ExaNeSt H2020 EU-funded project aims to design and develop an exascale ready prototype based on low-energy-consumption, ARM64 cores and FPGA accelerators. We participate to the application-driven design of the platform and to the validation of the prototype. To carry on this work we are using three widely used astrophysical codes: a direct N-body code, called Hy-Nbody and derived from HiGPUs (Capuzzo-Dolcetta+13), to simulate stellar cluster dynamics; PINOCCHIO (Monaco+02), a code aimed to generate catalogues of dark matter halos and their merger histories; GADGET-3, evolution of the public code GADGET-2 (Springel+05), a state-of-the-art N-body and hydrodynamical code for large-scale, high-resolution numerical simulations of cosmic structure formation and evolution. Core algorithms of the aforementioned applications have been improved in such a way to increasingly fit to the exa-scale target platform.

Waiting for the ExaNest prototype release, we are performing some tests and code tuning operations on ARM SoC facility. We deployed a testbed HPC cluster based on 64-bit ARMv8 Cortex-A72/Cortex-A53 core design, powered by the Mali-T860 embedded GPU, and SLURM Resource Manager. In parallel we are porting a kernel of Hy-Nbody on FPGA aiming to obtain high performance-per-watt solutions for our algorithms.

In this poster we describe how we re-engineered the applications and we show first test results on ARM SoC.

P2.4: Gaia DPAC Project Office: Coordinating the production of the largest star catalogue.

Theme: Management of Large Science Project
Author(s):
G. Gracia-Abril (Gaia DPAC Project Office, ESAC; Astronomisches Rechen-Institut, Zentrum fr Astronomie, Uni. Heidelb)
D. Teyssier (Gaia DPAC Project Office, ESAC; Telespazio Vega UK Ltd for ESA / ESAC)
J. Portell (Gaia DPAC Project Office; Institut de Cincies del Cosmos, Universitat de Barcelona (IEEC- UB))
A.G.A. Brown (Leiden Observatory, Leiden University)
A. Vallenari (INAF - Osservatorio astronomico di Padova)
- F. Jansen, Mission Operations Division, Operations Department, Directorate of Science, European Space Research and Technology Centre - U. Lammers, European Space Astronomy Centre (ESA/ESAC)

Contact: *ggracia@sciops.esa.int*

The ESA Gaia satellite is creating the most accurate map ever of the Milky Way. The second release of the Gaia archive was made public in April 2018. The impact of this release in the scientific community can be quantified with the number of papers submitted since its publication, in average around 2 papers per day based on the released data since it was published.

The Data Processing and Analysis Consortium (DPAC) is in charge of processing the Gaia data. More than 400 scientists and engineers distributed over 80 institutions in 20 countries work in DPAC. The challenge is to process the, today's figures, more than 1 trillion CCD observations, of around 2 billion sources, in 6 processing centers with more than 10 processing pipelines. The data produced is not only astrometric data but also high quality photometry, radial velocities and stellar parameters for many objects types. All these data are interdependent, output of some pipelines are required as input by other systems and, in many cases, the dependencies are cyclic.

The DPAC Project Office, together with the sub group leaders in DPAC, is responsible to coordinate the consortium activities, managing inter-dependencies between the different groups and pipelines and schedule and monitor the data deliveries across the processing centers. We will present DPAC management structure and activities, the lessons learnt in the preparation of DR2 and the challenges for future releases.

P5.8: The Dutch contribution to the ESRC

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Yan Grange (ASTRON)
Zheng Meyer - Zhao (ASTRON)
Michael Wise (ASTRON)

Contact: *grange@astron.nl*

In current-day radio astronomy, large interferometric arrays play an important role. The data rates coming out such instruments makes it a challenge to distribute data to and process data by end-users, especially those who are not closely involved in the telescope array itself. With the advent of the SKA, processing on local compute systems is becoming impractical and the allocation of compute resources, as well as the support for users to process their data need to be a common effort. To do this, the processing of the data will be done at local Science Data Centres, the SKA Regional Centres (SRC). In our contribution, we present the work that is being done by ASTRON, as a main Dutch partner in the federated European Science Data Centre (ESRC). ASTRON operates two SKA pathfinder instruments (LOFAR and APERTIF), bringing in experience with large-scale processing of current-day telescopes.

O10.7: Harnessing the power of archival data to increase scientific output: the JCMT experience.

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Sarah Graves (East Asian Observatory)

Time: Wed 09.45 - 10.00

Contact: *s.graves@eaobservatory.org*

EAO, operators of the James Clerk Maxwell Telescope (JCMT) recently reanalysed more than five years of publications using our data, identifying which of our TAC-approved projects have gone onto be published, and identifying which types of publication lead to large numbers of follow-on publications. This has also allowed us to identify which observations have not been published either by their PIs or by teams using our archive.

We have also examined the data products and interfaces from existing, highly productive data collections (both those from our own observatory and from other telescopes), identifying the features that we believe have helped encourage uptake and.

We are currently engaged in using this information in several ways, While this includes both 'match-making' between exciting orphaned observations and investigators, we are also reevaluating our choice of observatory produced data products and considering additional, simple interfaces to encourage more use of these products.

This paper will discuss the observatory software, databases and workflow required to do this work, as well as discussing the findings from our work that are most relevant when running or planning observatory archives.

P3.3: Open-source web tools for spectroscopic and imaging data visualization for the VOXAstro initiative

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality
Author(s):
Kirill Grishin (Department of Physics, Moscow State University; Sternberg Astronomical Institute)
Igor Chilingarian (Smithsonian Astrophysical Observatory; Sternberg Astronomical Institute)
Ivan Katkov (Sternberg Astronomical Institute)
Ivan Zolotukhin (Sternberg Astronomical Institute; Universit e de Toulouse; UPS-OMP, IRAP)

Evgeniy V. Rubtsov, Department of Physics, Moscow State University;Sternberg Astronomical Institute Svyatoslav Borisov, Department of Physics, Moscow State University;Sternberg Astronomical Institute

Contact: *kirillg6@gmail.com*

VOXAstro stands for Virtual Observatory tools for eXtragalactic Astrophysics. This initiative includes several projects such as Reference Catalog of Spectral Energy Distribution (RCSED; http://rcsed.sai.msu.ru/) and Las-Campanas Stellar Library. Here we present a set of flexible open-source tools for visualization of spectral and imaging data. Using web-visualisation libraries FlotJS and Dash we developed interactive viewers for displaying low- and high-resolution spectra of stars and galaxies, which allow one to view spectra having resolution up to R=80000 without putting a significant load on server and client sides, which is achieved by choosing the adaptive spectral binning window and dynamically preloading the datasets. We implemented a number of additional features like multiple spectra display, output of header info (e.g. stellar atmospheric parameters or stellar population properties of galaxies), display of emission lines decomposition parameters (fluxes, widths etc.) The spectral viewers can be easily embedded into any archive or database web-site. We also present a cutout service that extracts data on the fly from the UKIDSS near-infrared imaging survey and generates colour composite RGB stamps, which we use, e.g. in the RCSED web-site as an embedded service. The service is built using Astropy python library and uses IVOA SIAP to access images, which it then cutouts on the fly. In the coming years we plan to expand the capabilities of our spectroscopic and imaging visualization services and use them in future projects within VOXAstro.

P1.20: Analyzing WIND data using machine learning

Theme: Machine Learning in Astronomy
Author(s):
Luiz Fernando Guedes dos Santos (CUA/GSFC)
Teresa Nieves-Chinchilla (CUA / GSFC)
Barbara J Thompson (NASA/GSFC)
Michael S Kirk (NASA/GSFC)

Contact: *51guedesdossantos@cua.edu*

Coronal mass ejections (CMEs) are large-scale explosions of magnetic field and plasma from the Sun's corona and the primary drivers of terrestrial space weather. The fastest CMEs can reach Earth in 1-5 days expanding in size as they travel due to their strong entrained magnetic fields. Multiple viewpoint and observations require many assumptions to model the 3D CME dynamic and kinematic evolution. Although in-situ the measurements provide us great advance in how CMEs are and evolve, it still needs many assumptions to model it. Using data from Earth-directed ICME events and developing methodologies using Machine Learning techniques we expect to improve the analysis of all this data already available and search for relations not yet observed. New methods can improve categorization, relationships between quantities and evolution of the CMEs itself.

I6.1: DevOps: the perfect ally for Science Operations for a large and distributed astronomy project.

Theme: DevOps Practices in Astronomy Software
Author(s):
Rocio Guerra Noguero (European Space Agency)

Time: Thu 09.00 - 09.30

Contact: *rguerra@sciops.esa.int*

The Gaia Science Operations Centre (SOC) is an integral part of a large consortium responsible for Gaia data processing. Serving terabytes of processed data on a daily basis to other Processing Centres across Europe makes unique demands on the processes, procedures, as well as the team itself. In this talk I will show how we have embraced the DevOps principles to achieve our goals on performance, reliability and teamwork.

P4.8: Computational Astrophysics with Go

Theme: Data Science: Workflows Hardware Software Humanware

Author(s):
Pramod Gupta (Department of Astronomy, University of Washington)

Contact: *psgupta@uw.edu*

Go is a relatively new open-source language from Google. It is a compiled language and so it is quite fast compared to interpreted languages. Moreover, the creators of the language focused on minimizing complexity. Hence, even though the language was not designed for scientific computing, its speed and simplicity make it quite attractive for scientific computing.

In this paper, I discuss the suitability of Go for Computational Astrophysics based on using Go for Monte Carlo Radiative Transfer. The Go compiler has fast compile times and gives helpful error messages. Go has a standard code formatting tool gofmt so everyone's code looks the same. This makes it easier to read code written by others. The language has excellent documentation and a large number of built-in libraries. The run time performance is within a factor of two of an equivalent C, C++ or Fortran program. Go does memory management and run time checks such as array bounds checking which C, C++ and Fortran do not do. This increases the running time but it also increases the reliability of the code.

Due to Go's newness and since it is not targeted at scientific computing, there are a limited number of scientific computing libraries. Hence, Go may not be a feasible choice for projects which depend on specialized libraries. However, for projects which are not dependent on such libraries, Go is an excellent language for Computational Astrophysics.

P13.6: MegaPipe 2.0: 10000 square degrees of CFHT MegaCam imaging

Theme: Other
Author(s):

Stephen Gwyn (Canadian Astronomy Data Centre)

Contact: *gwyn.cadc@gmail.com*

MegaPipe, the MegaCam data processing pipeline at the CADC, has been upgraded to version 2.0 and has processed 10000 square degrees of the sky. MegaPipe has been operating since 2008. It was originally intended to increase the usage of archival MegaCam data by calibrating and stacking the images as they became public. That focus moved to include processing data from the CFHT Large Programs such as the NGVS, OSSOS, VESTIGE and CFIS. MegaPipe 2.0 represents several improvements. The advent of GAIA means that the astrometric calibration is considerably more accurate. The public release of Pan-STARRS allows photometric calibration of images that even if they were taken under non-photometric conditions, by using the PS1 stars as in-field standards. Together this means that almost every MegaCam image can be astrometrically/photometrically calibrated to sufficient accuracy to allow stacking (30 mas and 0.01 magnitudes respectively). The other change going to MegaPipe 2.0 is how the images are stacked. Previously, MegaPipe only stacked images that were taken on more or less the same part of the sky. This limited the number of images that could stacked. MegaPipe 2.0 instead stacks on set of 10000x10000 pixel tiles, each half a degree square evenly covering the whole sky. The result that twice much sky area can stacked. There are now 10000 square degrees of imaging in the ugriz bands available for download at: http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/access/graph.html

P3.4: BaSC - A Bayesian path to improved source finding in radio astronomy

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Peter Hague (University of Cambridge)
Haoyang Ye (University of Cambridge)
Bojan Nikolic (University of Cambridge)
Steve Gull (University of Cambridge)

Contact: *prh44@cam.ac.uk*

Widely used source detection and characterisation methods in radio interferometry borrow from those developed for optical astronomy. This approach does not account for the fundamental differences in the data, and rely on algorithms such as CLEAN to preprocess interferometric maps in such a way that information may be lost.

We present BaSC (Bayesian Source Characterisation) - a Python library that calculates likelihoods for sources based on visibilities rather than CLEANed maps, and applies Bayesian methods to generate evidence values for different models of the part of the sky being observed. BaSC is extensible, and can for example provide the ability to account for data from other wavelengths in its calculations.

We will discuss the limitations of existing methods in cases such as the confusion of two sources, and demonstrate the performance of BaSC on both simulated and real observations.

BaSC is freely available at https://www.github.com/petehague/BASC

P6.7: The conceptual design of amateur public observatory software framework

Theme: DevOps Practices in Astronomy Software

Author(s):
Jun Han (National Astronomical Observatories, Chinese Academy of Sciences)
Dongwei Fan (National Astronomical Observatories, Chinese Academy of Sciences)
Chenzhou Cui (National Astronomical Observatories, Chinese Academy of Sciences)
Yunfei Xu (National Astronomical Observatories, Chinese Academy of Sciences)
Changhua Li (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *hanjun@nao.cas.cn*

As the great demand of public observatory, the amateur has built lots of observatories. They usually carry out astronomy observation in remote field stations after traveling great distances. It is not only wasting lots of time, but also not safe. As a result, we designed a hardware integration system last year to support remote observation, and the printed circuit board platform is in making. Based on this platform, we propose a conceptual design of amateur public observatory software framework as the observation control system. It could not only satisfy multiple observation modes' requirement, but also is easy to operate and more intelligent. It will be a huge boost to astronomy education and citizen science.

P12.6: Pixel mask Filtering of CXC Datamodel

Theme: Algorithms

Author(s):
Helen He (Smithsonian Astronomical Observatory)
Mark CresitelloDittmar (Smithsonian Astronomical Observatory)
Kenny Glotfelty (Smithsonian Astronomical Observatory)

Contact: *hhe@cfa.harvard.edu*

We present pixel mask filtering algorithm, a new attribute to Datamodel library. Datamodel, a Chandra X-Ray Center data analysis software (CIAO) package, facilitates a wide range of on-the-fly data manipulation capabilities such as copy, differ, binning and filtering. The pixel mask filtering of a 2-D binary image is integrated into the CIAO region filtering syntax and logic, and thus complements the conventional analytic shapes filtering of circle, rectangle, polynomial. The integration allows for the standard operations on combining masks and analytic shapes such as include, exclude, intersect, and union. The mask is stored in a data block, referenced by the region filter string in the resulting file's data subspace. We will highlight the features and usage of pixel mask filtering in the CIAO software, and discuss some divergent effects between mask filtering and region shape filtering.

I7.1: The Minor Planet Center Data Processing System

Theme: Software for Solar Systems Astronomy
Author(s):
Matthew J. Holman (Harvard-Smithsonian Center for Astrophysics)
Matthew J. Payne (Harvard-Smithsonian Center for Astrophysics)
James M. Bauer (University of Maryland)

Time: Tue 11.15 - 11.45

Contact: *mholman@cfa.harvard.edu*

The Minor Planet Center (MPC) is the international clearing house for all ground-based and space-based astrometric and radar observations of asteroids, comets, trans-Neptunian objects, and outer satellites of the giant planets. The MPC assigns designations, provides up-to-date ephemerides, and coordinates follow-up observations for these objects. To meet the needs of the community, the MPC currently receives and processes over two million observations per month and maintains a catalog of orbits of more than 700K objects.

Although the MPC processes observations of all minor solar system bodies, its focus is near-Earth objects (NEOs). All MPC operations are organized around this central function. The MPC is required to warn of NEOs approaching within 6 Earth Radii within the coming 6 months. Thus, the main components of the MPC's data processing system enable real-time identification of candidate NEOs, with possible impact trajectories, within a much larger volume of asteroids and other solar system objects. A few such alerts are issued each year, including that for ZLAF9B2/2018 LA. In addition, The MPC facilitates follow up observations and the coordination of observing assets for efficient recovery searches for NEOs.

We anticipate that the data volumne will increase to a factor of 10 to 100 over the next decade as surveys such as LSST and NEOCam come online, augmenting the already-large volume from programs such as Pan-STARRS, the Catalina Sky Survey, NEOWISE, and ZTF. Thus, we are in the process of building and testing a new MPC data processing system. The goals are to maximize accuracy, data accessibility, automation, and uptime while minimizing latency and maintaining dependable archives of all data received.

In this talk I will highlight the challenges faced by the MPC, demonstrate the key components of our data processing system, and describe a number of algorithmic advances that support a much more efficient and reliable system.

The MPC operates at the Smithsonian Astrophysical Observatory, part of the Harvard-Smithsonian Center for Astrophysics (CfA), under the auspices of the International Astronomical Union (IAU). The MPC is 100% funded by NASA as a functional sub-node of the Small Bodies Node (SBN) of the NASA Planetary Data System at U. Maryland.

P5.9: HEASOFT: A FITS Data Processing and Analysis Software Suite

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):

Bryan Irby (NASA Goddard Space Flight Center)

Contact: *bryan.k.irby@nasa.gov*

For over 20 years, HEASOFT has provided an ever-growing suite of utilities for processing and analyzing data in the FITS (Flexible Image Transport System) format, dedicated to missions supported by the HEASARC (High Energy Astrophysics Science Archive Research Center) at NASA's Goddard Space Flight Center, but supplying numerous generic processing and analysis utilities to the scientific community as well. HEASoft currently contains nearly 4M lines of code in over 1000 individual tasks, supports 15 high energy missions, and is used by thousands of individuals or institutions. With the number of users and partner missions increasing every year, we continue to focus on portability, modularity, and exploring new modes of distribution.

P7.3: The SunPy Ecosystem

Theme: Software for Solar Systems Astronomy
Author(s):
Jack Ireland (NASA GSFC)
Stuart Mumford (University of Sheffield, UK)
Nabil Freij (Universitat de les Illes Balears)
Daniel Ryan (USRA/NASA GSFC)
Steven Christe (NASA GSFC)
Albert Shih, NASA GSFC SunPy Community

Contact: *jack.ireland@nasa.gov*

We discuss the development of SunPy and the ecosystem of related Python packages that provide specialized functionality to the heliophysics community. SunPy is a community-developed, free and open-source software package for solar physics. It aims to provide a comprehensive data-analysis environment that allows researchers within the field of solar physics to perform common tasks simply.

SunPy has released 9 beta versions (0.x) and work is now ongoing on SunPy 1.0 whose purpose will be to provide a stable API. SunPy 1.0 will provide the ability to search and download data from multiple locations including the Virtual Solar Observatory (VSO), Joint Science Operations Center (JSOC) and the Heliophysics Event Knowledgebase (HEK); offers TimeSeries and Map objects that enable the manipulation of time series and two-dimensional image data respectively; and enables transformations between solar and astronomical coordinate systems. SunPy 1.0 will support Python 3.6+ only, and the SunPy 0.9.x series will provide bugfix support to existing functionality for Python 2.7.x until the scheduled end of Python 2 support at the end of 2019. SunPy is built using functionality from Astropy, pandas, matplotlib and other scientific packages commonly used in astronomy.

SunPy is supported by the development of affiliated packages that provide additional functionality built on top of the core library. For example, IRISPy is an SunPy-affiliated package designed for reading, manipulating and visualizing data taken with the Interface Region Imaging Spectrometer (IRIS). The package ndcube is a SunPy-affiliated package designed for handling n-dimensional datacubes described by a World Coordinate System translation, permitting powerful manipulation of multi-dimensional data. The package drms provides access to Michelson Doppler Imager (MDI), and Solar Dynamics Observatory (SDO) data. Other packages in the SunPy ecosystem are fiasco, a Python interface to the Chianti atomic database and solarbextrapolation, a library for extrapolating 3D magnetic fields from line-of-sight magnetograms. We discuss how SunPy can support an expanding heliophysical data-analysis environment.

P1.3: A new implementation of deep neural network to spatio-spectral analysis in X-ray astronomy

Theme: Machine Learning in Astronomy
Author(s):
Hiroyoshi Iwasaki (Rikkyo University)
Yuto Ichinohe (Rikkyo University)
Yasunobu Uchiyama (Rikkyo University)
Hiroya Yamaguchi (NASA GSFC/University of Maryland)

Contact: *h.iwasaki@rikkyo.ac.jp*

Recent rapid developments in Deep Learning, which can implicitly capture structures in high-dimensional data, will lead to the opening of a new chapter of astronomical data analysis. As a new implementation of Deep Learning techniques in the fields of astronomy, we here report our application of Variational Auto-Encoder (VAE) using deep neural network to spatio-spectral analysis of the data from the Chandra X-ray Observatory, in the particular case of Tychos supernova remnant. Previous applications of Machine Learning techniques to the analysis of SNRs have been limited to principal component analysis (Warren et al. 2005; Sato & Hughes 2017) and clustering without dimensional reduction (Burkey et al. 2013). We have implemented an unsupervised learning method combining VAE and Gaussian Mixture Model (GMM), where the reduction of dimensions of the observed data is performed by VAE and clustering in the feature space is by GMM. We have found that some characteristic features such as the iron knots in the southeastern region can be automatically recognized through this method. Our implementation exploits a new potential of Deep Learning in astronomical research.

P13.7: Abstracting the storage and retrieval of image data at the LSST

Theme: Other

Author(s):
Tim Jenness (LSST/AURA)
Jim Bosch (Princeton)
Pim Schellart (Princeton)
K-T Lim (SLAC)
Andy Salnikov (SLAC)
Michelle Gower (NCSA)

Contact: *tjenness@lsst.org*

Writing generic data processing pipelines requires that the algorithmic code does not ever have to know about data formats of files, or the locations of those files. At LSST we have a software system known as the Butler, that abstracts these details from the software developer. Scientists can specify the dataset they want in terms they understand, such as filter, observation id, date of observation, and instrument name, and the butler translates that to one or more files which are read and returned to them as a single Python object. Conversely, once they have finished processing the dataset they can give it back to the butler, with a label describing its new status, and the butler can write it in whatever format it has been configured to use.

The butler system is not LSST-specific and is entirely driven by external configuration to suit a specific use case. In this poster we describe the core features of the butler and the associated architecture.

P1.4: Variable Star Classification Using Multi-View Metric Learning

Theme: Machine Learning in Astronomy

Author(s):
K.B. Johnston (Florida Institute of Technology)
S.M. Caballero-Nieves (Florida Institute of Technology)
A.M. Peter (Florida Institute of Technology)
V. Petit (University of Delaware)
R. Haber (Florida Institute of Technology)

Contact: *kyjohnst2000@my.fit.edu*

Comprehensive observations of variable stars can include time domain photometry in a multitude of filters, spectroscopy, estimates of color (e.g. U-B), etc. When it is considered that the time domain data can be further transformed via digital signal processing methodologies, the potential representations of the observed target star are limitless. If the goal is classification of observed variable stars, using this multitude of representations/views can become a challenge as many of the modern pattern classification algorithms in industry are limited to single input and single output. Presented here is an initial review of multi-view classification as applied to variable star classification, to address this challenge. The variable star UCR Starlight dataset and LINEAR dataset are used to generate a baseline performance estimate. The LM3L algorithm is applied to a set of generic features, and the performance with regard to a generic feature space is evaluated. A matrix-variate implementation of the LM3L algorithm is designed and presented specifically for this task as well; the matrix-variate implementations are novel developments.

F3: Visualization in IRSA Services using Firefly

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Emmanuel Joliet (IPAC/Caltech)
Xiuqin Wu (LSST/IPAC)

Time: Tue 10.45 - 11.15

Contact: *ejoliet@ipac.caltech.edu*

NASA/IPAC Infrared Science Archive (IRSA) curates the science products of NASA's infrared and submillimeter missions, including many large-area and all-sky surveys. IRSA offers access to digital archives through powerful query engines (including VO-compliant interfaces) and offers unique data analysis and visualization tools. IRSA exploits a re-useable architecture to deploy cost-effective archives, including 2MASS, Spitzer, WISE, Planck, and a large number of highly-used contributed data products from a diverse set of astrophysics projects.

Firefly is IPAC's Advanced Astronomy WEB UI Framework. It was open sourced in 2015, hosted at GitHub. Firefly is designed for building a web-based front end to access science archives with advanced data visualization capabilities.The visualization provide user with an integrated experience with brushing and linking capabilities among images, catalogs, and plots. Firefly has been used in many IPAC IRSA applications, in LSST Science Platform Portal, and in NEDs newly released interface.

In this focus demo, we will show case many data access interfaces and services provided by IRSA based on Firefly. It will demonstrate the reusability of Firefly in query, data display, and its visualization capabilities, including the newly released features of HiPS images display, MOC overlay, and the interactions between all those visualization components.

P1.5: Multiscale spatial analysis of young stars complex using the dbscan clustering algorithm recursivel

Theme: Machine Learning in Astronomy
Author(s):
Joncour Isabelle (UGA/UMD)
Morayx, Estelle (UGA)
Duchne, Gaspard (UGA/Brekeley)
Mundy, Lee (UMD)

Contact: *isabelle.joncour@univ-grenoble-alpes.fr*

Clustering and spatial substructures sudies of the distribution of young stellar objects (YSOs) in star forming regions may be used as a key tracers of (1) their birth sites and (2) their dynamical evolution with time. This work aims to provide a framework to identify and analyze the multilevel topological substructures of star forming regions using recursively the dbscan density based clustering algorithm.

Sweeping the full range of local length scale from the scale of the whole region down to the wide pair regime, we apply the dbscan algorithm on young star spatial distribution, to identify the density connected components at each spatial scale with a high level of confidence from random expectations. We constrain the free parameters of the algorithm using correlation functions and nearest neighbor statistics. From that analysis, we derive (1) the density components spectrum defined as the number of detected density components as a function of the local scale and (2) the clusterTree object (dendrogram analogue) associated to the full multilevel topology of the density components. To quantify the complexity of the star forming region spatial structure, we then further introduce the Strahler order as an indicator of the level and depth of the substructures. We use this framework to analyze three different density profile types of star forming region: fractal, Plummer and random distribution showing that we indeed (1) recover the global density profile from the spatial point distribution analysis and (2) derive characteristic structural signatures for each type of region. These intrinsic characterizations of the topological properties of the star clusters are then used to study the topological structure of Taurus star forming complex.

O8.2: The ZTF Alert Stream: Lessons from the first six months of operating an LSST precursor

Theme: Time Domain Astronomy

Author(s):
Mario Juric (University of Washington)
Maria Patterson (University of Washington)
Eric Bellm (University of Washington)

Time: Wed 12.05 - 12.20

Contact: *mjuric@astro.washington.edu*

The Zwicky Transient Facility (ZTF) is an optical time-domain survey that is currently generating about one million alerts each night for transient, variable, and moving objects. The ZTF Alert Distribution System (ZADS; Patterson et al.) packages these alerts, distributes them to the ZTF Partnership members and community brokers, and allows for filtering of the alerts to objects of interest, all in near-real time. This system builds on industry-standard real-time stream processing tools: the Apache Avro binary serialization format and the Apache Kafka distributed streaming platform. It leverages concepts and tools being developed for LSST (Python client libraries), with the source code publicly available on GitHub.

This talk will give an overview of the ZTF alert distribution system. We will examine lessons learned from six months of operating an LSST precursor alert stream (both from the operator and end-user perspective), discuss opportunities for standardization, and implications for the LSST.

P4.9: JWST Data Management Subsystem Operations: Preparing to Receive, Process, and Archive JWST Data

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Catherine Kaleida (The Space Telescope Science Institute)
Faith Abney (The Space Telescope Science Institute)
Matthew Burger (The Space Telescope Science Institute)
John Scott (The Space Telescope Science Institute)

Contact: *ckaleida@stsci.edu*

The James Webb Space Telescope (JWST) is a cornerstone in NASA's strategic plan, serving as the premier tool for studying the earliest stars and galaxies and for understanding the origins and future of the universe and the galaxies and solar systems within it. The Data Management Subsystem (DMS) is an integral part of the systems JWST needs to achieve these goals, as it serves as the interface between JWST and the astronomers who use it. We outline the JWST DMS Operations and detail the systems and tools that will be used to ensure that the unprecedented JWST data products are of the highest quality possible and available in the archive as quickly as possible. We also describe the rehearsals that are taking place, in order to ensure the operations systems, personnel, and procedures are ready well in advance of the spacecraft launch.

P12.7: The algorithms behind the HPF and NEID pipeline

Theme: Algorithms
Author(s):
Kyle F. Kaplan (University of Arizona)

Contact: *kfkaplan@email.arizona.edu*

Abstract: HPF and NEID are new high-resolution stabilized echelle spectrometers at the forefront of using radial velocity techniques to search for terrestrial mass exoplanets. Nightly data taken at the telescopes with large format detectors must be automatically processed into radial velocities in less than 24 hours. This requires a large investment in computer power and memory along with an automated pipeline that can check the quality of the data and handle issues without the need for human intervention. I will present an overview of our pipeline and discuss the, sometimes novel, algorithms and techniques we use to turn the unprocessed 2D echellograms into optimally extracted 1D spectra. These algorithms include the use of polygon clipping to rectify the curvature found in the beams on the detectors, the ability to fully account for aliasing in under-sampled data on the detector using flat lamp spectra, and the use of pixel count histograms to automatically match similar exposures and check the quality of the data. I will also discuss how our pipeline is built up from many independent modules, making it robust against failure and allowing it to be easily modifiable.

P12.8: Alpha-X: An Alpha Shape-based Hierarchical Clustering Algorithm

Theme: Algorithms

Author(s):

Ramsey L. Karim (University of Maryland, College Park)
Isabelle Joncour (Univ. Grenoble Alps)
Lee Mundy (University of Maryland, College Park)

Contact: *rkarim@astro.umd.edu*

This project is an ongoing exploration into the utility of alpha shapes in describing hierarchical clustering. Alpha shapes, a concept introduced by Edelsbrunner and Mucke 1994 that relates to a generalization of convex hulls known as alpha hulls, describe boundaries of regions around point clusters that are associated (point to point) at distances less than some characteristic length scale α. Algorithms for finding alpha shapes are based on the Delaunay triangulation of the point set, which leads to the notion of both positive space, in the inclusion of these triangles into the shapes, as well as negative space, in their exclusion. This well-defined negative space can indicate gaps in point clouds at a given α scale; hence the alpha shape approach can define both over- and under- densities of points. The concept of alpha shapes is a discrete approach and can thus be applied to sets of positions of stars to evaluate stellar clustering and associated voids.

One feature we are developing is the representation of point-cloud substructures as α values in tree representations which capture the hierarchical lineage of structure at different values of α. With this approach, alpha shapes could be used in an alternate hierarchical cluster detection method that characterizes clusters and their complementary gaps over a range of length scales and naturally yields defined boundaries for these cluster and gap shapes. Similar approaches using this method have been successfully implemented outside of astronomy, in fields such as molecular biology, pattern recognition, and digital shape sampling (Varshney, Brooks, and Wright 1994; Edelsbrunner 2010). Our project will continue to develop and validate this methodology, comparing it to extant methods used within our field to verify that it offers novel and significant analysis products before moving into specific scientific applications.

P4.10: CIAO: A Look Under the Hood of Chandra's X-Ray Imaging and Analysis Software Configuration Management- Past, Present, and Future.

Author(s):
Zeke Kaufman (Smithsonian Astrophysical Observatory)
Mark CresitelloDittmar (Smithsonian Astrophysical Observatory)
Janet Evans (Smithsonian Astrophysical Observatory)
Omar Laurino (Smithsonian Astrophysical Observatory)
Warren McLaughlin (Smithsonian Astrophysical Observatory)
Joe Miller, Smithsonian Astrophysical Observatory

Contact: *zkaufman@cfa.harvard.edu*

Abstract. The CIAO (Chandra Interactive Analysis of Observations) software suite is approaching two decades of service**1 and CIAO remains the primary analysis package from the Chandra X-Ray Observatory. Despite the package's maturity, CIAO continues to undergo active development from a diverse group of developers using a multiple programming languages and build architectures. Keeping up with the ever evolving capabilities in hardware, software, version control systems and paradigm shifts in software development methodologies presents a challenge to both developers and configuration management teams. This paper provides an overview of how the CIAO software suite has evolved over the years with a particular emphasis on configuration management of the system. This paper examines CIAO's integration with various off the shelf software as well as CIAO's packaging and distribution system and how these have evolved. We will conclude with an outlook on the future direction of CIAO infrastructure including possible integration with modern package management systems such as Conda, plans for continuous integration, and use of tools such as pip for extending capabilities in CIAO.
    **1 CIAO version 1.0 was released October 1999

P3.5: Vissage: viewing polarisation data from ALMA

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Wataru Kawasaki (National Astronomical Observatory of Japan)
Yuji Shirasaki (National Astronomical Observatory of Japan)
Christopher Zapart (National Astronomical Observatory of Japan)
Akira Yoshino (National Astronomical Observatory of Japan)
Eisuke Morita (National Astronomical Observatory of Japan)
Tsuyoshi Kobayashi, George Kosugi, Masatoshi Ohishi, and Yoshihiko Mizumoto (National Astronomical Observatory of Japan)

Contact: *wataru.kawasaki@nao.ac.jp*

Vissage (VISualisation Software for Astronomical Gigantic data cubEs) is a standalone FITS browser, primarily aiming to offer easy visualisation of huge, multi-dimensional FITS data from ALMA. We report its new features including basic capabilities of easily viewing polarisation data from ALMA.

P7.4: ZChecker: Finding Cometary Outbursts with the Zwicky Transient Facility

Theme: Software for Solar Systems Astronomy
Author(s):
Michael S. P. Kelley (University of Maryland)
Dennis Bodewits (Auburn University)
Quan-Zhi Ye (Caltech/IPAC)

Contact: *msk@astro.umd.edu*

Cometary science benefits from wide-field time-domain optical surveys. Aside from the discovery of new comets, such surveys can provide a better description of known objects. For example, we can quantify intrinsic brightness variation with heliocentric distance and true anomaly (i.e., season); potentially estimate dust-to-gas ratio and its variation with time (when relevant filters are used); and identify cometary outbursts or other interesting phenomenon for follow-up. We describe ZChecker, automated software for finding and visualizing known comets in Zwicky Transient Facility (ZTF) survey data. ZChecker uses on-line ephemeris generation and individual data product meta-data (observation time, image corners, and the world coordinate system) to efficiently identify and download images of targets of interest in the ZTF archive. Photometry of each target is measured, and the images rotated to place the comet-Sun vector at a constant position angle. To help identify comets for follow-up investigations, ZChecker then scales each image of each target to a common heliocentric and geocentric distance, then combines the data into nightly and bi-weekly averages. The difference between the two shows variations in brightness and morphology that potentially indicate, e.g., an outburst of activity or the motion of a precessing jet. Example ZChecker output is presented, including outbursts of 29P/Schwassmann-Wachmann 1 and 64P/Swift-Gehrels, and a curved dust feature at comet C/2017 M4 (ATLAS).

I3.2: 3D Data Visualization in Astrophysics

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality
Author(s):
Brian R. Kent (NRAO)

Time: Mon 09.00 - 09.30

Contact: *bkent@nrao.edu*

We present unique methods for rendering astronomical data - 3D galaxy catalogs, planetary maps, data cubes, and simulations. Using tools and languages including Blender, Python, and Google Spatial Media, a user can render their own science results, allowing for further analysis of their data phase space. We aim to put these tools and methods in the hands of students and researchers so that they can bring their own data visualizations to life on different computing platforms.

O12.1: Auto-multithresh: A General Purpose Automated Masking Algorithm for Clean

Theme: Algorithms

Author(s):
Amanda Kepley (National Radio Astronomy Observatory)
Takahiro Tsutsumi (National Radio Astronomy Observatory)
Crystal Brogan (National Radio Astronomy Observatory)
Remy Indebetouw (National Radio Astronomy Observatory)
Ilsang Yoon (National Radio Astronomy Observatory)
Brian Mason, National Radio Astronomy Observatory

Time: Mon 11.45 - 12.00

Contact: *akepley@nrao.edu*

Generating images from radio interferometer data requires deconvolving the point spread function of the array from the initial image. This process is commonly done via the clean algorithm, which iteratively models the observed emission. Because this algorithm has many degrees of freedom, producing an optimal science image typically requires the scientist to manually mask regions of real emission while cleaning. This process is a major hurdle for the creation of the automated imaging pipelines necessary to process the high data rates produced by current and future interferometers like ALMA, the JVLA, and the ngVLA. In this talk, we present a general purpose masking algorithm called auto-multithresh that automatically masks emission during the cleaning process. This algorithm was initially implemented within the tclean task in CASA 5.1. The tclean implementation significant performance improvements in CASA 5.3. The auto-multithresh algorithm is in production as part of the ALMA Cycle 5 and 6 imaging pipelines. It has also been shown to work with data from telescopes like the VLA and ATCA. We describe how this algorithm works, provide a variety of examples demonstrating a success of the algorithm, and discuss the performance of the algorithm. Finally, we close with some future directions for producing science ready data products that build on this algorithm.

P4.11: DALiuGE/CASA based processing for the extragalactic HI observations with FAST.

Author(s):
V. Kitaeff (The International Centre for Radio Astronomy Research, The University of Western Australia)
M. Zhu (National Astronomical Observatories, Chinese Academy of Sciences)
L. Staveley-Smith (The International Centre for Radio Astronomy Research, The University of Western Australia)
R. Tobar (The International Centre for Radio Astronomy Research, The University of Western Australia)
K. Vinsen (The International Centre for Radio Astronomy Research, The University of Western Australia)
A. Wicenec,The International Centre for Radio Astronomy Research, The University of Western Australia C. Wu, The International Centre for Radio Astronomy Research, The University of Western Australia

Contact: *slava.kitaeff@uwa.edu.au*

We present a prototype for the spectral-line data reduction pipeline based on the graph-based execution framework DALiuGE, and the CASA single-dish spectral-line package. The pipeline has been designed for the drift-scan mode of FAST multi-beam telescope targeting extra-galactic HI observations.

P1.6: The construction of a new stellar classification template library for the LAMOST 1D Pipeline based on LAMOST DR5

Theme: Machine Learning in Astronomy
Author(s):
Xiao, Kong (National Astronomical Observatories, Chinese Academy of Sciences)
A-li, Luo (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *kongx@nao.cas.cn*

With the ability of capturing four thousands spectra in one exposure, the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) is a special quasi-meridian reflecting Schmidt telescope located in Xinglong Station of national Astronomical Observatory, China. It released the fifth spectral data this year, containing 8,171,443 star, 153,090 galaxy, 51,133 QSO and 642,178 unknown type, all of which are classified by the LAMOST 1D pipeline. This pipeline is used for spectral analysis, aiming to determine the spectral type and redshifts of the spectra observed by LAMOST by matching with spectral templates. Generally, the performance of the stellar classification greatly depends on the quality of templates. In this paper, we construct a new stellar template library, which is supposed to increase the number of types and the accuracy of the classification, for LAMOST based on the data from LAMOST DR5. All the 9 million spectra from LAMOST DR5 are participated in this construction experiment and they are gathered in 500 groups using k-means clustering method. Those group centers corresponding to spectra less than 1000 are abandoned at first. Then the weighted average spectrum (group center) is served as the template spectrum in each group. Initially, 417 centers are obtained. We visually inspect all template spectra and discard 181 centers due to low spectral quality or the similarity between different group centers. Furthermore, the types of the remained clustering centers are assigned by the subclass of spectra from LAMOST DR5. Meanwhile, 19 templates whose subclass are difficult to determine are also abandoned. Afterwards, we obtain a new template library containing 197 LAMOST template spectra with 82 different MK classes. Finally, the feasibility and accuracy of using this template for classification has been verified by comparing and analyzing the classification results of several control groups of data.

P4.12: Data Processing Challenges for CTA

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Karl Kosack (CEA Paris-Saclay and CTA Observatory)

for the CTA Observatory

Contact: *karl.kosack@cea.fr*

The Cherenkov Telescope Array (CTA) Observatory will generate hundreds of petabytes of raw and simulated data during its operation—orders of magnitude more than any current gamma-ray instrument. We present the current status of the design and prototyping efforts for a system to handle the large volume of raw data, to produce a nearly equal volume of simulated data needed to characterize the instrumental response, and to process it all to provide much more compact science data to users. This includes our efforts at a comprehensive architectural model, as well as the use of modern technologies for distributed data processing and storage, and for a python-driven analysis.

P12.9: Qualification of Sparse Modeling Technique for radio interferometric imaging of ALMA

Theme: Algorithms

Author(s):
George Kosugi (National Astronomical Observatory of Japan)
Takeshi Nakazato (National Astronomical Observatory of Japan)
Shiro Ikeda (The Institute of Statistical Mathematics)

Contact: *george.kosugi@nao.ac.jp*

Sparse modeling is widely used in image processing, signal processing, and machine learning recently. Thanks to the research and progress in statistical mathematics along with the evolution of computational power, the technique is supposed to be applicable to the radio interferometric imaging for the data obtained by ALMA (Atacama Large Millimeter-submillimeter Array). We've developed a new imaging tool based on the sparse modeling approach and experimentally implemented on CASA (Common Astronomy Software Application) which is an official reduction software for the ALMA data. The poster presentation gives supplemental information to the oral talk in session 11 by Nakazato et al. The new imaging technique with sparse modeling is a computationally intense process even for the latest CPUs. In the poster, several ideas and practices to reduce the calculation time will be presented. The comparison to CLEAN imaging by using artificial data (simulated data) will also be presented.

I9.1: Coordinating observations among ground and space-based telescopes in the multi-messenger era

Theme: Multi-Messenger Astronomy
Author(s):
Erik Kuulkers (European Space Agency)

Time: Tue 16.30 - 17.00

Contact: *erik.kuulkers@esa.int*

The emergence of time-domain multi-messenger (astro)physics asks for new and more efficient ways of interchanging information, as well as collaboration. Many space- and ground-based observatories have web pages dedicated to showing information about the complete observations and planned observation schedule. The aim would be to standardise the exchange of information about observational schedules and set-ups between facilities and in addition, to standardise the automation of visibility checking for multiple facilities. To reach this, we propose to use the VO protocols (ObsTAP-like) to write services to expose these data to potential client applications and to develop cross facilities visibility servers.

P2.5: Lessons Learned from the behemoth JWST Data Management effort

Theme: Management of Large Science Project
Author(s):
Mark Kyprianou (STScI)
Anastasia Alexov (STScI)
Katherine Kaleida (STScI)

Contact: *kyp@stsci.edu*

As with any large project, there are always lessons to be learned for future endeavors that can be applied and/or considered for similar endeavors. The development of the Data Management Subsystem (DMS) for the James Webb Space Telescope (JWST) provides such an opportunity to take advantage of experiences and events that can be applied to other missions. The DMS software and hardware tool set is extensive; it involves more than 100 staff including developers, scientists and systems engineers. It receives science and engineering telemetry data from the spacecraft, performs quality checking, calibration, data (re-)processing; archives the data; provides user notification, search, and access tools for JWST science and engineering data; distributes data to the end user; provides extensive user analysis/visualization tools. We will discuss the evolution of DMS and its interaction within the Science and Operations Center (S&OC) covering keys to our success and also those that did not work out so well.

P11.2: DRAGONS - Data Reduction for Astronomy from Gemini Observatory North and South

Theme: Quality Assurance of Science Data
Author(s):
Kathleen Labrie (Gemini Observatory)
Kenneth Anderson (Gemini Observatory)
Ricardo Cardenes (Gemini Observatory)
Chris Simpson (Gemini Observatory)
James Turner (Gemini Observatory)

Contact: *klabrie@gemini.edu*

DRAGONS, Data Reduction for Astronomy from Gemini Observatory North and South, is Gemini's new Python-based data reduction platform. DRAGONS offers an automation system that allows for hands-off pipeline reduction of Gemini data, or of any other astronomical data once configured. The platform also allows researchers to control input parameters and in some cases will offer to interactively optimize some data reduction steps, e.g. change the order of fit and visualize the new solution.

The project makes good use of other open source projects. The data interface, Astrodata, uses at its core astropys NDData and io.fits. The input parameters configuration system uses a slightly modified version of LSST's pex.config. The project is also working with the astropy community to define the tools needed for building spectroscopic data reduction packages.

DRAGONS is used at the observatory for nighttime quality assessment. The same software will be used for quicklook reduction of target of opportunity and LSST follow-up observations, and as the tool the researchers can use to prepare their Gemini data for analysis.

P2.6: The VLA Sky Survey - operations, data processing and archiving

Theme: Management of Large Science Project

Author(s):
Mark Lacy (NRAO)
Claire Chandler (NRAO)
Amy Kimball (NRAO)
Steve Myers (NRAO)
Stephan Witz (NRAO)

Contact: *mlacy@nrao.edu*

The VLA Sky Survey (VLASS) is an ambitious project to image the entire sky visible to the VLA in three epochs. The Survey isl being carried out at a frequency of 2-4GHz in full linear polarization (Stokes I,Q,U) and a resolution of 2.5, making it the highest angular resolution all-sky radio survey ever attempted. VLASS will collect 0.5PB of raw data, and 0.3PB of Basic Data Products (calibration tables, images and catalogs) will be created and archived at NRAO. There are also opportunities for external groups to create higher level Extended Data Products. In this poster, we present a summary of the scheduling, operations, data processing and archiving of the survey, as well as some early results from the first set of observations, covering half the sky.

P5.10: Optimization Strategies for running Legacy Codes

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Jason Lammers (University of Maryland College Park)
Dr. Peter Teuben (University of Maryland College Park)

Contact: *jlammers0@gmail.com*

Legacy codes can be both incredibly useful as well as cumbersome. Although they produce robust results from well-defined problems, they plague the user with complex installations, complicated user interfaces, and inflexibilities. Short of rewriting such codes, scripts which emulate the effects of legacy codes can lead to tremendous overhead and more headache for the user. Nonetheless, these legacy codes can provide invaluable insight to a skilled user who takes the time to understand its inner workings.

In this paper, we will showcase examples of the NEMO Stellar Dynamics Toolkit and how analysis simulation codes can be optimized using modern scripting languages. Not only is it possible, it is relatively easy to speed up computation time and increase user-friendliness by integrating modern scripting languages into codes which would otherwise overwhelm an inexperienced user.

O12.4: Data-Driven Pixelisation with Voronoi Tessellation

Theme: Algorithms
Author(s):
Marco C Lam (Astrophysics Research Institute, Liverpool John Moores University)

Time: Mon 14.15 - 14.30

Contact: *c.y.lam@ljmu.ac.uk*

In modern Astrophysics, Voronoi Tessellation is a rarely used as a pixelisation scheme. While it exists, it is almost exclusively used in signal enhancements and simulations. In Observational Astronomy, with Gaia, ZTF, DES etc. data becoming available, LSST and Euclid coming online in the next decade, this branch of science is becoming more and more data-driven. HEALPix, HTM and Q3C offer excellent ways to pixelise the celestial sphere, the implementations completely separate the background information from the signal. There are excellent use cases to have them independent from each other, but there are also cases when this becomes a burden in computation when we have to process more pixels than necessary or require post-hoc calculations to group pixels at different resolution levels to form larger segments. With Voronoi Tessellation, it can generate a one-to-one mapping of data points to Voronoi cells where anywhere inside the cell is the closest to the governing data point. We illustrate the application of Voronoi Tessellation in a set of magnitude and proper motion-limited data how it can simplify the survey properties of the 3pi Steradian Survey from the Pan-STARRS 1, where the footprint area is imaged by the 60 CCDs at $10\hat{5}$ pointings over 3.5 years.

O11.2: Quality assurance in the ingestion of data into the CDS VizieR catalogue and data services

Theme: Quality Assurance of Science Data
Author(s):
Gilles Landais (CDS)
Pierre Ocvirk (CDS)
Mark Allen (CDS)

VizieR team (T.Boch, M.Brouty, E.Perret, FX. Pineau, T.Pouvreau, P.Vannier)

Time: Mon 16.15 - 16.30

Contact: *gilles.landais@astro.unistra.fr*

VizieR is a reference service provided by the CDS for astronomical catalogues and tables published in academic journals, and also for associated data. Quality assurance is a key factor that guides the operations, development and maintenance of the data ingestion procedures. The catalogue ingestion pipeline involves a number of validation steps, which must be implemented with high efficiency to process the 1200 catalogues per year from the major astronomy journals. These processes involve an integrated teams of software engineers, specialised data librarians (documentalists) and astronomers, and various levels of interaction with the original authors and data providers. Procedures for the ingestion of associated data have recently been improved with semi-automatic mapping of metadata into the IVOA ObsCore standard, with an interactive tool to help authors submit their data (images, spectra, time series etc.). We present an overview of the quality assurance procedures in place for the operation of the VizieR pipelines, and identify the future challenges of increasing volumes and complexity of data. We highlight the lessons learned from implementing the FITS metadata mapping tools for authors and data providers. We show how the quality assurance is an essential part of making the VizieR data comply with FAIR (Findable, Accessible, Interoperable and Re-useable) principles, and the necessity of quality assurance in for the operational aspects of supporting more than 300,000 VizieR queries per day through multiple interactive and programmatic interfaces.

P6.8: Application of Google Cloud Platform in Astrophysics

Theme: DevOps Practices in Astronomy Software

Author(s):
MARCO LANDONI (INAF - Osservatorio Astronomico di Brera)
RICCARDO SMAREGLIA (INAF - OSS. ASTRONOMICO TRIESTE)
GIULIANO TAFFONI (INAF - OSS ASTRONOMICO TRIESTE)
ANDREA BIGNAMINI (INAF - OSSERVATORIO ASTRONOMICO TRIESTE)

Contact: *marco.landoni@inaf.it*

The availability of new Cloud Platform offered by Google motivated us to propose nine Proof of Concepts (PoC) aiming to demonstrated and test the capabilities of the platform in the context of scientifically-driven tasks and requirements. In this poster, we review the status of our initiative by illustrating 3 out of 9 successfully closed PoC that we implemented on Google Cloud Platform. In particular, we illustrate a cloud architecture for deployment of scientific software as microservice coupling Google Compute Engine with Docker and Pub/Sub to dispatch heavily parallel simulations. We detail also en experiment for HPC based simulation and workflow executions of data reduction pipelines (for the TNG-GIANO-B spectrograph) deployed on GCP. We compare and contrast our experience with on-site facilities comparing advantages and disadvantages both in terms of total cost of ownership and reached performances.

P10.6: Mapping Data Models to VOTable

Author(s):
Omar Laurino (Smithsonian Astrophysical Observatory)
Gerard Lemson (Johns Hopkins University)
Mark Cresitello-Dittmar (Smithsonian Astrophysical Observatory)
Tom Donaldson (Space Telescope Science Institute)
Laurent Michel (Observatoire de Strasbourg)

Contact: *olaurino@cfa.harvard.edu*

Data providers and curators provide a great deal of metadata with their data file. This metadata is invaluable for users and for Virtual Observatory software developers. In order to be interoperable, the metadata must refer to common data models. A new specification is being developed by the IVOA Data Modeling Working Group to define a scheme for annotating VOTable instances in a standard, consistent, interoperable fashion, so that each piece of metadata can unambiguously refer to the correct data model element it expresses. With this specification data providers can unambiguously and completely represent metadata in the VOTable format as data model instances, and clients can build faithful representations of such instances. The mapping can be extended to formats other than VOTable.

P10.7: FileDB, a pattern for querying cosmological simulations

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Gerard Lemson (JHU)

Contact: *glemson1@jhu.edu*

Cosmological simulations these days consist of 100s of billions of particles and 100s of snapshots. Summary data products derived from these data sets such as cluster catalogues are now frequently stored in large, publically accessible relational databases. But there is an increasing demand for access to the raw simulation data, which however are generally too large and less suited for a purely relational treatment. I will present FileDB, a system for querying such simulations from within a relational database, without requiring the ingestion of the raw data into the database itself. FileDB uses a special organization of the simulation snapshots along space filling curves to achieve efficient support for the main, spatial query patterns. We have created for this purpose a query library inside SQL Server that calculates the overlap between 3D query volumes and these curves to derive efficient retrieval patterns of the required data. This library is callable in SQL though user defined functions, and has been ported to Java for use in Oracle for example. We have implemented FileDB on the Millennium database and other simulations accessible in the online SciServer/CasJobs query service. As example I will show SQL queries joining the raw data to the halo catalogues stored in the database to determine density profiles. The fact that FileDB leaves the files in their original form means they can also be accessed directly on the file system. This is particularly useful for global analysis of the files, for which databases could be a bottle neck. I will show examples how SciServer/Compute provides access to these same simulations from within Jupyter notebooks running server side in Docker containers that have the data volumes directly mounted. Finally, the FileDB pattern easily translates to other domains such ocean circulation models, allowing the creation of multi-petabyte database in a cheap and efficient manner.

P1.7: Design of KNN Star-QSO Classification Algorithm Based on Cloud Computing

Theme: Machine Learning in Astronomy
Author(s):
Changhua Li (National Astronomical Observatories, Chinese Academy of Sciences)
Chenzhou Cui (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *lich@nao.cas.cn*

KNN classification algorithm is a mature and simple machine learning algorithm, which is widely used in astronomical spectral classification. However, with the increasing sample size, the traditional KNN algorithm cant meet the requirements of performance. Based on the cloud computing environment of China Virtual Observatory, the KNN algorithm is optimized based on SPARK + HDFS architecture. Through Map/Reduce, the large samples are segmented and distributed on different SPARK computing nodes, and the obvious acceleration effect is achieved. By comparing with the python-ML general algorithm, KNN's computation time can be reduced by nearly 20 times in a large scale samples

I1.1: Deep learning of astronomical features with big data.

Theme: Machine Learning in Astronomy
Author(s):
Maggie Lieu (ESA)

Time: Tue 14.00 - 14.30

Contact: *maggie.lieu@sciops.esa.int*

In Astronomy, there is a tendency to build machine learning codes for very specific object detection in images. The classification of asteroids and non-asteroids should be no different than the classification of asteroids, stars, galaxies, cosmic rays, ghosts or any other artefact found in astronomical data. In computer science, it is not uncommon for machine learning to train on hundreds of thousands of object categories, so why are we not there yet? I will talk about image classification with deep learning and how we can make use of existing tools such as the ESA science archive, ESAsky and citizen science to help realise the full potential of object detection and image classification in Astronomy.

P11.3: stginga: Ginga Plugins for Data Analysis and Quality Assurance of HST and JWST Science Data

Theme: Quality Assurance of Science Data

Author(s):
Pey Lian Lim (STScI)
Eric Jeschke (NAOJ)

Contact: *lim@stsci.edu*

stginga is an image visualization package to assist in data analysis and quality assurance of science data from Hubble Space Telescope (HST) and James Webb Space Telescope (JWST). It is based on the Ginga toolkit for building scientific viewers. In this poster, we will describe the main plugins developed for data analysis and quality assurance tasks with stginga. We also discuss the basic outline of writing a Ginga plugin, with pointers to documentation and examples.

P11.4: The quality assurance of Chinese solar physics historical observation data archives

Theme: Quality Assurance of Science Data

Author(s):
Ganghua Lin,NAOC
Suo Liu,NAOC

Contact: *lgh@bao.ac.cn*

The archives of Chinese solar physics history observation data mainly came from the 5 observation stations, they are based on 17 kinds of observation data, among them the longest observation data has been for more then 90 years. So it should be a needed taken seriously by each data worker that how to make so many kinds of data are processed better, so that a high quality archives can be constructed, and be provided conveniently to users for using. This paper introduces the every steps of processing of the archives, for example, from the digitalization of the original data archive, archiving, data processing, to share, in the each step of processing includes the quality assurance.

O6.1: Agile and DevOps from the trenches at ASTRON

Theme: DevOps Practices in Astronomy Software
Author(s):
Marcel Loose (ASTRON)

Time: Thu 09.30 - 09.45

Contact: *loose@astron.nl*

A few years ago the Software development teams at ASTRON decided to adopt the Agile/Scrum software development method. We are building instruments and software that push technological boundaries. Requirements often lack sufficient detail and are subject to constant change, whilst the first data from a new instrument or early prototype become available. The unknown unknowns largely outnumber the known unknowns. Agile/Scrum has proven to be successful in situations like these.

We stumbled and fell, but gained a lot of experience in how Agile development techniques can be used in the scientific arena. We learned what works, and what does not work. We became more and more convinced that Agile/Scrum can be very effective in the area of Scientific Software development. In this presentation I would like to take you by the hand and revisit the journey we have made, in the hope that you will learn from the mistakes that we have made, and the lessons that we have learned.

P11.5: A TripleStore implementation of the IVOA Provenance Data model

Theme: Quality Assurance of Science Data

Author(s):
Mireille Louys (Icube and CDS, Observatoire astronomique , University of Strasbourg)
Franois-Xavier Pineau (CDS, Observatoire astronomique de Strasbourg, University of Strasbourg)
Franois Bonnarel (CDS, Observatoire astronomique de Strasbourg, University of Strasbourg)
Lucas Holzman (Ensiie , Evry)

Contact: *mireille.louys@unistra.fr*

The IVOA has proposed a standard for capturing the provenance metadata in the production and distribution of astronomical data. We present an implementation in a tripple store for the provenance information recorded for a collection of astronomical images. The ontology applied is derived from PROV-O from the W3C and from the IVOA Provenance data model. SPARQL queries based on the Data model concepts allow to select datasets based on a wide range of provenance properties and have proven to be efficient in the tripple store representation. The data model of the SIMBAD CDS data base has also been tested, and has proven to scale very efficiently in the tripplestore strategy as well.

P9.9: Searching for Optical Counterparts to Gravitational Wave Events with the Catalina Sky Survey

Theme: Multi-Messenger Astronomy
Author(s):
Michael Lundquist (University of Arizona/Steward Observatory)
David Sand (University of Arizona/Steward Observatory)
Eric Christensen (University of Arizona/Lunar & Planetary Laboratory)

Contact: *mlundquist@email.arizona.edu*

On 17 August 2017, the era of multi-messenger, gravitational wave astronomy began with the discovery of the optical counterpart to the gravitational wave event GW170817. In this poster, we outline our software and strategy for discovering optical counterparts to future gravitational wave events using data from the Catalina Sky Survey.

P13.8: Getting ready for the fourth Asterics DADI virtual observatory school

Theme: Other
Author(s):
Katharina Lutz (CDS Observatoire astronomique de Strasbourg)
Mark Allen (CDS Observatoire astronomique de Strasbourg)
Ada Nebot (CDS Observatoire astronomique de Strasbourg)

Contact: *katharina.lutz@astro.unistra.fr*

We present the current status of the EURO-VO tutorials in the light of the upcoming fourth Asterics DADI virtual observatory school.

P12.10: Modeling Narrow Rings with a Single Chain of Gravitating Particles

Theme: Algorithms
Author(s):
Yuxi Lu (University of Maryland College Park)
Douglas Hamilton (University of Maryland College Park)

Contact: *lucylulu12311@gmail.com*

Narrow rings have been discovered around Saturn, Uranus and the other gas giants. These structures are hundreds of thousands of km across yet only a few km wide. The most well-known radial confinement theory involves Shepherd moons in which each ring is bracketed by a pair of small satellites. Saturns F ring and the Uranian Epsilon ring have exactly such moonlet pairs. In other cases, two body resonances seem to be responsible for sculpting ring edges. However, for the vast majority of rings, neither of these machinists is active. We are investigating the possibility of three-body resonances inducing stability in narrow rings using a N-body code, hnbody. We modified the code to simulate narrow rings using a chain of particles by including the gravity from only the nearest neighboring particles. We study the strengths and limitations of this code by measuring the precession rate for rings modeled with different numbers of particles and/or nearest neighbors. We have developed graphic tools in Python to directly visualize the chain of particles using polar. The FFT is particularly useful for finding weak resonances caused by a distance satellite or the more exotic three-body resonances.

P5.11: Arcade: A User Focussed, Visual Compute Environment in CANFAR

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):

Brian Major (National Research Council Canada)
JJ Kavelaars (National Research Council Canada)
Sbastien Fabbro (National Research Council Canada)
Daniel Durand (National Research Council Canada)

Contact: *major.brian@gmail.com*

For a number of years CANFAR (Canadian Advanced Network for Astronomy Research) has offered virtual machines (VMs) as a way to do both interactive computing and batch processing in the cloud. A VM is a general and flexible base offering that can suit nearly any given astronomy compute project, but demands of users a thorough understanding of the intricacies of software installation and maintenance, and requires significant effort to achieve initial benefits. Arcade is an effort by the Canadian Astronomy Data Centre (CADC) to offer the CANFAR community an astronomy-focussed, easier-to-use, and more intuitive science platform.

Arcade is composed of a number of pre-built, specialized application bundles ('game consoles') that appear in a graphical desktop environment. They run independently of each other in the cloud, allowing Arcade to optimize console execution to their particular resource requirements. Consoles have user-specific access to a shared file system and other CANFAR cloud services, outfitting users with a variety of computing, analysis and storage tools.

At the CADC we see a lot of potential in Arcade and will be planning its evolution based on a number of key questions. How can scalability be best achieved? How can we reduce the burden of software and infrastructure maintenance for users and operators? Can we allow users to customize their Arcade experience? How can consoles be launched in batch processing? How can we best leverage open source technology and development from other projects? We shall discuss the core concepts of Arcade and explore its potential in respect to these questions and from feedback from the astronomy computing community.

P5.12: Galaxy Cataloguing Expert System (GCES): Use Cases

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Maria Jose Marquez (EUMETSAT)
Tams Budvari (JHU)
Luis Manuel Sarro (UNED)

Contact: *mariajose.marquez@eumetsat.int*

Title: Galaxy Cataloguing Expert System (GCES): Use Cases Lead author: Mara Jos Mrquez The observational nature of Science, especially relevant in domains such as Astronomy and Astrophysics sets important challenges in the data analysis domain. Nowadays astronomical surveys yield challenging images with crowded populations, overlapping objects, and faint detections. Here we present a Galaxy Cataloguing Expert System (GCES) to better exploit the knowledge that can be extracted from the surveys. This expert system involves the combination of appropriate AI methodologies, i.e. Bayesian inference, Voronoi tessellation and computer vision all this integrated in an automatic and configurable processing pipeline and requiring intense and intelligent computational capabilities. The GCES system addresses a multiple problem, i.e. i) label sources in an astronomical image according to the probability that each one is blended or contaminated with surrounding sources, using Voronoi tessellation and rule-based system; ii) extract photometric measurements of extended objects using contour-based approaches; and iii) carry out a probabilistic cross-match of sources from different images and bands, using all astro-photometric information available. We demonstrate here the use of the GCES in full with the most representative use cases. When the GCES system is used with the COSMOS catalogue, apparent false detections are re-discovered as real sources. This demonstrates the benefits of using artificial intelligence methodologies in the exploitation of Big Data and the consequent knowledge discovery. The paper is structured as follows: section 1 presents an introduction of the multi-axis problem and its context; section 2 describes the expert data processing and analysis pipeline (GCSE) and its specific use for the problem indicated in section 1; finally section 3 presents relevant results obtained from the different use cases with real COSMOS data sets along with a critical assessment.

O10.4: Data-driven Space Science at ESA Science Data Centre

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Beatriz Martinez (Rhea System S.A. for ESA)
Isa Barbarisi (Serco for ESA)
Juan Gonzalez (Serco for ESA)
Bruno Merin (ESDC / ESA)
Sara Nieto (Rhea System S.A. for ESA)
Hector Perez Rhea System S.A. for ESA, Jesus Salgado QUASAR for ESA, Pilar de Teodoro
Aurora for ESA

Time: Wed 09.00 - 09.15

Contact: *beatriz.martinez@esa.int*

For many scientists nowadays, the first step in doing science is exploring the data computationally. New approaches to data-driven science are needed due to the big increase of space science missions data in volume, heterogeneity, velocity and complexity. This applies to ESA space science missions, whose archives are hosted at the ESA Science Data Centre (ESDC). Some examples are the Gaia archive -whose size is estimated to grow up to 1PB and 6000 billion of objects-, the Solar Orbiter archive -which is expected to handle several time series with more than 500 millions of records- and the Euclid archive, which shall be able to handle up to 10PB of data. The ESDC aims, as a major objective, to maximize the scientific exploitation of the archived data. Challenges are not limited to manage the large volume of data, but also to allow collaboration between scientists, to provide tools for exploring and mining the data, to integrate data (the value of data explodes when it can be linked with other data), or to manage data in context (track provenance, handle uncertainty and error). ESDC is exploring solutions for handling those challenges in different areas. Specifically: storage of big catalogues through distributed databases (ex. Greenplum, Postgres-XL,); storage of long time series in high resolution via time series oriented databases (TimeScaleDB); fulfil data analysis requirements via Elasticsearch or Spark/Hadoop; and enabling scientific collaboration and closer access to data via JupyterLab, Python client libraries and integration with pipelines using containers. In this presentation we are going to take a tour of these approaches.

P1.8: Putting more intelligence into the reduction and analysis of SITELLE data.

Theme: Machine Learning in Astronomy
Author(s):
Thomas Martin (Universit Laval)
Laurent Drissen (Universit Laval)

Contact: *thomas.martin.1@ulaval.ca*

SITELLE is the imaging Fourier transform spectrometer of the Canada-France-Hawaii Telescope (CFHT). It delivers cubes of 4 million spectra at moderate resolution (R 3000 - 10000) on a 11'x11' FOV with a spatial resolution of 0.32.

Each data cube contains up to 34 Go of data and a handful of different astrophysical objects which cannot all be detected easily (especially at low signal-to-noise ratio) : high-redshift galaxies, planetary nebulae, emission-line stars, compact hii regions, etc. Object detection and classification is a challenge. We have developed tools based on convolutional neural networks that greatly help reasearcher to quickly find the most hidden objects.

Data reduction have also been enhanced with deep neural networks for cosmic ray detection and stars detection.

Finally, spectral model fitting can also take advantage of the python library Tensor Flow which provides GPU optimized minimization algorithm.

We will also see more of the marvelous data obtained recently.

P8.4: Insight-HXMT Timing Analysis of New Black Hole Candidate MAXI J1820+070

Theme: Time Domain Astronomy
Author(s):

Ma Xiang (max@ihep.ac.cn)
Qu Jinlu (qujl@ihep.ac.cn)

Contact: *max@ihep.ac.cn*

We present the Insight-HXMT timing analysis of the new black hole candidate MAXI J1820+070 discovered in March 2018.We find that the source exhibits state transitions from LHS to HIMS and eventually to SIMS. Low frequency quasi-periodic oscillations(QPO) is found in the hard state and intermediate states and its centroid frequency increase with time while Q and RMS do not show evolution. With the large effective area of Insight-HXMT at high energy band, we are able to present the energy dependence of the QPO amplitude and centroid frequency up to 200 keV which supplies more constraints on the region of QPO.

P1.9: Saving endangered animals with Astro-Ecology

Theme: Machine Learning in Astronomy
Author(s):
Paul Ross McWhirter (Liverpool John Moores University)
Josh Veitch-Michaelis (Liverpool John Moores University)
Claire Burke (Liverpool John Moores University)
Steve Longmore (Liverpool John Moores University)
Owen McAree (Liverpool John Moores University)
Serge Wich Liverpool John Moores University

Contact: *P.R.McWhirter@ljmu.ac.uk*

Conservation science is experiencing an unprecedented challenge in identifying and protecting endangered species across the world. The large stretches of land and sea require innovative solutions for the monitoring of endangered populations. Drones equipped with high resolution cameras with supporting data from satellites have helped to mitigate these challenges. Unfortunately, it is difficult to detect animals from optical images when they might only be a matter of a few pixels across from the perspective of an overhead drone. The Astro-Ecology research group at Liverpool John Moores University is working on deploying thermal infrared cameras on drones to detect animals from their body heat and therefore at night as well as during the day. In the thermal infrared band, animals appear as bright sources on a dark, colder background. Using these techniques, multiple populations have been classified and studied.

As the project scaled up, the initial manual classifications of animal populations became unfeasible given the quantity of the collected thermal images. The nature of these images attracted astrophysicists to join the project as the automated detection and classification of bright sources on a dark background has been of particular interest to observational astronomy for decades. The application of readily available software such as Sextractor was surprisingly successful identifying the bright sources in the data. This success was short-lived as the thermal background could change strongly depending on the time of day and the local climate. Using the Moderate Resolution Imaging Spectroradiometer satellites (MODIS), a model of land surface temperature variation across daily and yearly timescales was produced. This allowed for future data gathering flights to be optimised for the local conditions prior to the expedition. The atmosphere also results in an absorption effect on the observed sources in the thermal infrared primarily due to water content as a function of the distance from the camera to the source and the air temperature.

These sources of error must be corrected for the successful identification of an animals surface temperature. As the primary goal of this project is the detection and classification of animals, it is important that they be distinguishable from the background and have thermal profiles indicative of their species. This requires the development of an image pre-processing pipeline, similar to the calibrations applied to astronomical instrumentation, so that the detection and classification algorithms are applied to data with the expected thermal profiles where the temperature recorded by the instrument matches that of the target source. Individually modelling the thermal profile of every animal is quite difficult yet a large quantity of data has been collected containing these thermal profiles (once pre-processed).

We are developing a machine learning pipeline based heavily on the approaches in modern computer vision which are simultaneously being employed within current astronomical image-based approaches. Our approach to this task requires a large set of labelled training data of multiple species. Whilst we have collected a large quantity of data, it would take a long time to manually identify the sources of interest. To address this problem we are using citizen science through the Zooniverse web portal. Using a detection method based on thresholding the thermal data based on the land surface temperature conditions, we identify sources with higher than expected temperature. The Zooniverse site requests users to perform a set of tasks designed to remove false positives, identify missed sources and classify the detected animals. Upon the completion of the

citizen science workloads, the training data can then be applied to a selection of computer vision machine learning algorithms to produce models for the detection and classification of animals from their thermal profiles.

F1: Building LOFAR As A Service: Processing Petabytes with just a click

Theme: Management of Large Science Project
Author(s):
Alexandar Mechev (Leiden Observatory)
Huib Intema (Leiden Observatory)
Aske Plaat (Leiden Institute of Advanced Computer Science)

Time: Mon 10.45 - 11.15

Contact: *apmechev@strw.leidenuniv.nl*

The LOFAR Radio Telescope produces PetaBytes of data each year. Processing such volumes is not possible at clusters provided by academic institutions, and thus needs to be launched and managed at a High Throughput Cluster. With increasing complexity of LOFAR workflows, building and maintaining new scientific workflows on a distributed architecture becomes prohibitively time consuming. To make pipeline development and deployment easy and data processing fast, we integrate cluster middleware and LOFAR software with a leading workflow orchestration software, Airflow. The result is a flexible application that can launch and manage LOFAR processing. With Airflow, we can easily create a service for LOFAR users to process their data transparently.

P3.6: Realtime telescope and data visualization using web technologies

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Pablo Mellado (IRAM, Institut de Radioastronomie Millimetrique)

Contact: *mellado@iram.es*

When performing onsite or remote observations in a telescope, it is very critical to have a good feedback about the current status of your ongoing observation. Nowadays the web technology have evolved to allow us to get data in realtime with the advantage of using just a web browser.

In our telescope we found the need of updating our previous outdated monitoring system which is currently showing information about the status of the telescope, the last scans plots, a view of the current weather conditions, etc

This poster shows how the data visualization has improved by using newer technologies like microservices, websockets and messaging, as well as the structure developed to integrate them succesfully in a reliable and more attractive way.

P13.9: ALiX: An advanced search interface for AladinLite

Theme: Other

Author(s):

Laurent Michel (Observatoire Astronomique de Strasbourg)
Thomas Boch (CDS Strasbourg)
Wang Jie (Universit Technologique de Belfort Montbliard)

Contact: *laurent.michel@astro.unistra.fr*

ALiX is a flexible catalog portal based on Aladin Lite. It it designed to use an interactive sky view as a primary selection tool. The ALiX view is constantly updated with data queried from the host database. It offers advanced functionalities allowing to mix local data with VO data. Users can draw areas of interest by hand and bookmark views. ALiX has no dependency with any specific data source; it can be integrated in any existing portal.

P10.8: The new science portal and the programmatic and VO interfaces of the ESO science archive

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Alberto Micol (ESO)
Nausicaa Delmotte (ESO)
Vincenzo Forch (ESO)
Nathalie Fourniol (ESO)
Olivier Hainaut (ESO)
Uwe Lange, ESO; Khan Ahmed Mubashir, ESO; Joerg Retzlaff, ESO; Martino Romaniello, ESO; Devendra Sisodia, Pactum Gmbh; Malgorzata Stellert,Tekom Gmbh; Felix Stoehr, ESO; Ignacio Vera, ESO; Stefano Zampieri, ESO;

Contact: *amicol@eso.org*

In 2018 new powerful ESO science archive interfaces have become available to the astronomical community. The main functionalities were already presented at the ADASS 2017, but one year has gone by not without challanges, and with some useful additions. The new archive interfaces include two main components.

The ESO Archive Science Portal: Interactive access via web pages to browse and explore the archive with interactive, iterative queries. The results are presented in real time in various tabular and/or graphic forms, including interactive previews, allowing an evaluation of the usefulness of the data which can then be selected for retrieval.

The direct database and Virtual Observatory access: The inherent limitation in the intuitive way that the web interface enables archive content to be discovered is that it is unsuited to more complex queries, such as those that include sequences with logical statements like and, or and not, or queries that join different sources of information. This restriction can be overcome by providing direct access to the ESO database tables. Extensive documentation is provided in terms of practical examples, which are intended to provide templates for users to customise and adapt to their specific needs.

In this first release, processed data from the LPO are supported. Future plans include expanding the support to ALMA processed data and raw data from the LPO. It is planned that these new access points will gradually replace the current ones for La Silla Paranal data, while ALMA will keep maintaining a dedicated, separate access.

P8.5: Evaluating Methods for Flare Detection in GALEX Light Curves

Theme: Time Domain Astronomy
Author(s):

Chase Million (Million Concepts)
Scott Fleming (Space Telescope Science Institute)
Parke Loyd (Arizona State University)

Contact: *chase@millionconcepts.com*

The gPhoton flare catalog (GFCat) project will produce light curves and measurement statistics for thousands of stellar flares serendipitously observed by the Galaxy Evolution Explorer (GALEX) mission, with energies between log10(27) and log10(32) ergs in two ultraviolet bandpasses. This will require an automated search for flares in millions of light curves with time resolution 1 minute. A variety of methodologies have been used for the detection and measurement of stellar flares, many requiring substantial human intervention, and there is no clear consensus approach. We evaluate the performance of a variety of autonomous flare detection and measurement algorithms using injection and recovery testing from simulated GALEX light curves.

P5.13: Finding Your Place in the Cosmos with WCSTools

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Jessica Mink (Smithsonian Astrophysical Observatory)

Contact: *jmink@cfa.harvard.edu*

The WCSTools package has been updated to include access to arbitrary-length keywords in FITS headers and the UCAC5 and GAIA catalogs. Examples of the use of those and older little -known features will be presented.

P13.10: Starting up a Data Model for Exoplanetary Data

Theme: Other
Author(s):
Marco Molinaro (INAF - OATs)
Eleonora Alei (INAF - OAPd)
Serena Benatti (INAF - OAPd)
Andrea Bignamini (INAF-OATs)
Franois Bonnarel (CDS, Observatoire de Strasbourg)
Mario Damasso / INAF - OATo Mireille Louys / CDS, Observatoire de Strasbourg Michele Maris
/ INAF - OATs Valerio Nascimbeni / INAF - OAPd

Contact: *marco.molinaro@inaf.it*

The search for, study and characterisation of extrasolar planets and planetary systems is a grow-
ing and improving field of astrophysical research. Alongside the growing knowledge on the field
the data resources are also growing, both from observations and numerical simulations. To tackle
interoperability of these data an effort is starting (under the EU H2020 ASTERICS project) to
delineate a data model to allow a common sharing of the datasets and collections of exoplanetary
data. The data model will pick up model components from the IVOA specifications, either existing
or under investigation, and attach new ones where needed. This contribution presents the first
results in drafting the exoplanetary systems dedicated data model. It shows relationships with
existing and proposed IVOA models and presents the new key components not yet available in the
interoperable scenario. The results here reported are those provided by a dedicated face-to-face
meeting attended by the authors. They cover a first set of requirements and considerations and
take into account aspects like the observations of exoplanetary systems, the usage of existing ex-
oplanets catalogues, the investigation of atmospheres of confirmed exoplanets and the simulation
of exoplanet's atmospheres devoted to characterize exoplanets habitability. Use cases and require-
ments are also presented to form the basis of this modelling effort.

I5.1: Hubble in the Cloud: A Prototype of a Science Platform at STScI

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
Ivelina Momcheva (STScI)

Time: Wed 16.00 - 16.30

Contact: *imomcheva@stsci.edu*

The availability of high-quality, highly-usable data analysis tools is of critical importance to all astronomers as is easy access to data from our archives. In this talk I will describe the approach to developing the prototype of a new cloud-based data management environment for astronomical data reduction and analysis at STScI. I will examine the decisions we made, I will demonstrate the prototype and I will discuss what new areas of scientific exploration and discovery are opened by this platform.

P12.11: Image reconstruction method for an X-ray telescope with an angular resolution booster

Theme: Algorithms

Author(s):
Mikio Morii (The Institute of Statistical Mathematics)
Shiro Ikeda (The Institute of Statistical Mathematics)
Yoshitomo Maeda (Institute of Space and Astronautical Science)

Contact: *morii@ism.ac.jp*

We propose an image reconstruction method for an X-ray telescope with an angular resolution booster proposed by Y. Maeda et al. at ISAS/JAXA. The booster consists of double coded masks in front of an X-ray mirror. In order to have a better sky image from an off-focus image, a proper image reconstruction process must be applied. The new image reconstruction method is based on the Bayesian statistics, where the traditional Richardson-Lucy algorithm is extended with a prior of sparseness and smoothness. Such a prior is desirable for astronomical imaging because astronomical objects have variety in shape from point sources, diffuse sources (supernova remnants, clusters of galaxies, and pulsar wind nebula) to mixtures of them (point sources in Galactic planes). As a result, the image resolution would be improved from a few arcmin to 10 arcsec. The performance of the X-ray telescope is demonstrated with simulated data: point sources and diffused X-ray sources such as Cas A and Crab Nebula. Through the demonstration, the angular resolution booster with the image reconstruction method is shown to be feasible.

O4.2: New Synthesis Imaging Tool for ALMA based on the Sparse Modeling

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Takeshi Nakazato (National Astronomical Observatory of Japan)
Shiro Ikeda (The Institute of Statistical Mathematics)
Kazunori Akiyama (National Radio Astronomy Observatory, MIT, Harvard University, NAOJ)
George Kosugi (National Astronomical Observatory of Japan)
Masayuki Yamaguchi (University of Tokyo, National Astronomical Observatory of Japan)
Mareki Honma (National Astronomical Observatory of Japan)

Time: Wed 14.45 - 15.00

Contact: *takeshi.nakazato@nao.ac.jp*

A new imaging tool for radio interferometry has been developed based on the sparse modeling approach. It has been implemented as a Python module operating on Common Astronomy Software Applications (CASA) so that the tool is able to process the data taken by Atacama Large Millimeter/submillimeter Array (ALMA). In order to handle large data of ALMA, the Fast Fourier Transform has been implemented with gridding process. The concept of the sparse modeling for the image reconstruction has been realized with two regularization terms: L1 norm term for the sparsity and Total Squared Variation (TSV) term for the smoothness of the resulting image. Since it is important to adjust the size of the regularization terms appropriately, the cross-validation routine, which is a standard method in statistics, has been implemented. This imaging tool runs even on a standard laptop PC and processes ALMA data within a reasonable time. The interface of the tool is comprehensible to CASA users and the usage is so simple that it consists of mainly three steps to obtain the result: an initialization, a configuration, and a processing. Remarkable feature of the tool is that it produces the solution without human intervention. Furthermore, the solution is robust in the sense that it is less affected by the processing parameters. For the verification of the imaging tool, we have tested it with two extreme examples from ALMA Science Verification Data: the protoplanetary disk, HL Tau as a typical smooth and filled image, and the lensed galaxy, SDP.81 as a sparse image. In our presentation, these results will be presented with some performance information. The comparison between our results and those of traditional CLEAN method will also be provided. Finally, our future improvement and enhancement plan to make our tool competitive with CLEAN will be shown.

P5.14: ESAC Science Exploitation and Preservation Platform Reference Architecture

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Vicente Navarro (European Space Agency)
Ruben Alvarez (European Space Agency)
Fernando Perez (Rhea for European Space Agency)
Christophe Arviset (European Space Agency)
Javier Ventura Traveset (European Space Agency)
Roberto Prieto - European Space Agency Angel Martin Furones - Politechnical University of Valencia

Contact: *vicente.navarro@esa.int*

At ESA, active science missions like Gaia, Planck and XMM-Newton have developed precursor systems enabling the provision of advanced applications for the execution and instantiation of data analysis pipelines. Simultaneously, developments in missions like Euclid, as well as programmes like Galileo and Copernicus are tackling the creation of cyberinfrastructures capable of acquiring, processing, distributing and analysing massive amounts of data in an effective way. These initiatives have led to the implementation of solutions, commonly known as Thematic Exploitation Platforms.

ESAC Science Exploitation and Preservation Platform (SEPP) project drives the consolidation of past experiences and future needs into a reference framework to foster research through the provision of space science data, products and services. SEPP aims at integrating information and processing assets into a single environment to deliver advanced analysis and collaboration services.

This work presents SEPPs multi-mission reference architecture, which leverages on mainstream big data, cloud, virtualisation and container technologies to create a software as a service (SaaS) computing environment. This environment pivots around the paradigm shift characterised by the move of processing components to the data, rather than the move of data to the users.

SEPP puts the focus on the science community, to promote their contributions and involvement in the form of data and processing extensions. It provides storage space for users to bring their data and processing code close to the archives, encouraging execution of user-customised pipelines. Along the same lines, the integration of VO standards, tools like JupyterLab and on-demand instantiation of applications from a web based Science App Store, maximise interoperability and collaboration across the community.

I8.1: Data challenges of the VO in Time Domain Astronomy

Theme: Time Domain Astronomy
Author(s):
Ada Nebot (CDS, Observatoire Astronomique de Strasbourg)

Time: Wed 11.20 - 11.50

Contact: *ada.nebot@astro.unistra.fr*

Surveys specifically designed to monitor the transient sky have opened the window for discovery and exploration through time domain. Source classification and transmission of the alerts for further follow-up as well as analysing possible periodicity in variable sources poses a challenge with the huge amounts of data synoptic missions are providing. We will review some of the challenges of Time Domain data and we will share some of the tools and services that are being built within the Virtual Observatory to discover, access, visualise and analyse Time Domain data, focusing in particular on Time Series.

P10.9: HXMT Archive and Data Process System

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Jianyin, Nie (IHEP, CAS)
Mingyu, Ge (IHEP, CAS)
Xiaobo, Li (IHEP, CAS)
Chengkui, Li (IHEP, CAS)
HaiSheng, Zhao (IHEP, CAS)

Contact: *niejy@ihep.ac.cn*

Hard X-ray Modulation Telescope (HXMT) is the first Chinese X-ray Space Observatory. It covered detection limit from 5keV 150keV. HXMT Scientific Data Center (HSDC) , who producing and charging Data Product is a branch of HXMT Groud SegmentAfter over one year operation HSDC has accumulated over 300TB Data, include about 10TB raw data and 290TB high level production. HSDC also developed a set of software to process, manager, monitor and analyse data production base on Computing cluster and Storage cluster. Next, We will Use more advanced algorithms blossoming in Big-Data field and AI technology. The data processing software set will also be extended to other telescope.

P10.10: SCIENCE EXPLOITATION IN A BIG DATA ARCHIVE: THE EUCLID SCIENTIFIC ARCHIVE SYSTEM

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Sara Nieto (ESAC)
Bruno Altieri (ESAC)
Damien Noriet (ESAC)
Pilar de Teodoro (ESAC)
Jesus Salgado (ESAC)
Fabrizio Giordano (ESAC) Elena Racero (ESAC) Juan Gonzalez (ESAC) Monica Fernandez (ESAC) Bruno Merin (ESAC)

Contact: *sara.nieto@sciops.esa.int*

Euclid is an ESA mission and a milestone to explore the dark Universe. Euclid will map the 3D distribution of up to two billion galaxies and dark matter associated with them. It will hence measure the large-scale structures of the Universe across 10 billion light years, revealing the history of its expansion and the growth of structures during the last three-quarters of its history. In total Euclid will produce up to 26 PB per year of observations.

The Euclid Archive System is a joint development between ESA and the Euclid Consortium and is led by the Science Data Centres of the Netherlands and the ESDC (ESA Science Data Centre). The EAS is composed by three different subsystems: Data Processing System (DPS), Distributed Storage System (DSS) and Science Archive System (SAS). The SAS is being built at the ESDC and is intended to provide access to the most valuable scientific data, which is currently estimated in 10 PB of images, catalogue and spectra, after 6 years mission.

Big-data technologies, driven by nature and data volume, are transforming the way of doing scientific research towards collaborative platforms whose first goal is to enable access and process large data sets in ways that could not be done downloading the data. Distributed shared-nothing architectures for databases and data processing allowing scaling-out, joint with interactive analysis tools are currently the main technologies explored as part of the Euclid scientific archive. Some examples are: JupyterLab, Apache Spark, GreenPlum and PostgresXL. We will describe how Euclid in the context of the ESDC and in collaboration with the Gaia archive, envisages such a challenge to reach the scientific goals of the mission.

O1.2: Acceleration of Non-Linear Minimisation with PyTorch

Theme: Machine Learning in Astronomy
Author(s):
Bojan Nikolic (University of Cambridge)

Time: Tue 14.45 - 15.00

Contact: *b.nikolic@mrao.cam.ac.uk*

Minimisation (or, equivalently, maximisation) of non-linear functions is a widespread tool in astronomy, e.g., maximum likelihood or maximum a-posteriori estimates of model parameters. Training of machine learning models can also be expressed as a minimisation problem (although with some idiosyncrasies). This similarity opens the possibility of re-purposing machine learning software for general minimisation problems in science.

I show that PyTorch, a software framework intended primarily for training of neural networks, can easily be applied to general function minimisation in science. I demonstrate this with an example inverse problem, the Out-of-Focus Holography technique for measuring telescope surfaces, where a improvement in time-to-solution of around 300 times is achieved with respect to a conventional NumPy implementation. The software engineering effort needed to achieve this speed is modest, and readability and maintainability are largely unaffected.

P13.11: Subaru Telescope Network 5 or STN5 - The new computer and network system at Subaru Telescope

Theme: Other
Author(s):
Junichi Noumaru (Subaru Telescope, National Astronomical Observatory of Japan)
Kiaina Schubert (Subaru Telescope, National Astronomical Observatory of Japan)
Thomas Winegar (Subaru Telescope, National Astronomical Observatory of Japan)
Eiji Kyono (Subaru Telescope, National Astronomical Observatory of Japan)
Hitomi Yamanoi (Subaru Telescope Mitaka Office, National Astronomical Observatory of Japan)

Contact: *noumaru@naoj.org*

Subaru Telescope has recently completed the procurement and installation of the 5th contract of the computing environment called Subaru Telescope Network 5 or STN5. The original commitment to incorporate telescope sub-systems not managed by contract vendor was a high priority with STN5. We were successful with the procurement of sufficient computing resources supporting the Hilo base and the summit facilities, to move analysis machines, virtual machines and instrument control machines called OBCPs that have been re-developed to virtual machines.

As in past contracts the core network environment has been migrated to a 10 Gbps Cisco core switch. This guarantees the highest available bandwidth to critical observation systems, such as Gen2, PFS, SCExAO, Hyper Suprime-Cam, and Summit Virtual Environment. This also allows future upgrades of upstream network switches throughout the dome structure supporting instruments.

The analysis environment and has been enhanced and the virtual machine environment has been increased at both the Hilo base and the summit facilities. The latter allows for organizational VMs developed by other divisions to be migrated to managed environment.

In 2018, Subaru Telescope Archive System or STARS, which is a part of STN5, has been developing a FITS correction system, at the request of HSC Team, for modification of FITS keywords calculated by post-observation reduction clusters. Example keywords are for seeing and transparency. Users can download either FITS originals with ASCII corrections or the updated FITS corrected file.

The future goal of STN5 is to allow for Computer & Data Management Division to be relieved of system and network administration during the 1st year of the contract. Gather necessary resources, manpower, and knowledge, to administer the current contract system.

P13.12: The impact of Virtual Observatory Tools in Astronomy Research in Nigeria

Theme: Other
Author(s):
Ikechukwu Anthony Obi (Centre for Basic Space Science)
Okere Bonaventure (Centre for Basic Space Science)
Offor Peter (Centrd forBasic Space Science)

Ikechukwu Anthony

Contact: *tonykassidy_z@yahoo.com*

Astronomy development in Nigeria has, over the years, been hampered by lack of adequate astronomical facilities and access to data due to poor internet connectivity. In this paper, we discuss how recent introduction of the Virtual Obervatory tools in schools and universities has brought a fast-growing interest in astronomy amongst students as well as boosting the level of astronomy research being carried out in arious universities.

P7.5: The Analysis of Periodic Variation of Sunspot Groups and the X-ray Flare Classes

Theme: Software for Solar Systems Astronomy

Author(s):
Oloketuyi Jacob (Yunnan Observatories, Chinese Academy of Sciences, Kunming and UCAS)
Liu Yu (Yunnan Observatories, Chinese Academy of Sciences, Kunming)
Zhao Mingyu (Yunnan Observatories, Chinese Academy of Sciences, Kunming)

Contact: *jacob.oloketuyi@ynao.ac.cn*

The present solar cycle 24 has been thought to be unusually quiet in many respects. Both sunspot group numbers and flare numbers show a very low level of occurrence compared to the previous cycles. This study investigated the variations and relationship exhibited between the sunspot group numbers and the solar x-ray flares classes (B, C, M and X) in the solar cycles 23 and 24. The Observations of sunspot groups and the solar x-ray flares are archived in the National Oceanic and Atmospheric Administration (NOAA) website. The study was conducted using the continuous wavelet transforms, and the cross-correlation methods. Firstly, we found that the total daily solar flare produced is asynchronous and in antiphase with the sunspot group numbers while it is positively correlated with the sunspot group numbers on cross-correlation analysis. Periodicities of short and intermediate variations were also observed with sunspot group numbers having a significant periodicity around 400 days. The phase asynchrony between the sunspot group numbers and the total flares which is a result of the drop in B and C flares during solar maximum could be attributed to overwhelming of M and X flares produced by magnetic flux system from the active region/sunspot groups.

O10.5: Bringing together the Australian sky - coordination and interoperability challenges of the All-Sky Virtual Observatory

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Katrina Sealey (AAO-MQ)
Simon O'Toole (AAO-MQ)
Lloyd Harischandra (AAO-MQ)
Elizabeth Mannering (AAO-MQ)
Minh Vuong (AAO-MQ)

Time: Wed 09.15 - 09.30

Contact: *katrina.sealey@aao.gov.au*

The Australian All-Sky Virtual Observatory (ASVO) consists currently of 5 nodes. There are 2 nodes with optical astronomical data; Data Central (MQ) and Skymapper (ANU). There are 2 nodes with radio data; Murchison Wide Field Array (MWA, Curtin) and CSIRO ASKAP Science Data Archive (CASDA, CSIRO). The last node is the Theoretical Astrophysical Observatory (TAO, Swin). These 5 nodes work together under the unified ASVO.

The Australian astronomical user community is driving multi-node and multi-wavelength use cases, for example, querying Data Central spectroscopic data with Skymapper imaging data. Meeting the user requirements of the community comes with complexities and challenges. Some of the challenges we are facing include a single sign-on (unified authorisation/authentication) and the querying and representation of very different remote data, such as, overlaying GaLactic and Extragalactic All-sky MWA Survey (GLEAM) data stored in Western Australia with imaging data stored in Eastern Australian states. This presentation will discuss the challenges and successes in both co-ordinating the Australian ASVO and providing interoperability across the 5 nodes.

P1.10: Identifying contaminants in astronomical images using convolutional neural networks

Theme: Machine Learning in Astronomy
Author(s):
Maxime Paillassa (Laboratoire dastrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, alle Geoffrey Saint-Hilaire, 3)
Emmanuel Bertin (Sorbonne Universit, CNRS, UMR 7095, Institut dAstrophysique de Paris, 98 bis bd Arago, 75014 Paris)
Herv Bouy (Laboratoire dastrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, alle Geoffrey Saint-Hilaire, 3)

Contact: *maxime.paillassa@u-bordeaux.fr*

In this work, we propose to use convolutional neural networks to detect contaminants in astronomical images. Each contaminant is treated in a one vs all fashion. Once trained, our networks are able to detect various contaminants such as cosmic rays, hot and bad pixels defaults, saturated pixels, diffraction spikes, nebulosities, persistence effects, satellite trails, residual fringe patterns, or tracking errors in images, encompassing a broad range of ambient conditions, PSF sampling, detectors, optics and stellar density. The CNN is performing semantic segmentation: it can output a probability map, assigning to each pixel its probability to belong to the contaminant or the background class, except for tracking errors where another convolutional neural network can assign to a whole focal plane the probability that it is affected by tracking error. Training and testing data have been gathered from real data originating from various modern CCD and near-IR cameras or simulated.

P10.11: WAN OPTIMIZATION FOR ALMA DATA

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Jose Parra (ALMA Observatory)
Alejandro Barrientos (ALMA Observatory)
Maxs Simmonds (ALMA Observatory)

Contact: *jose.parra@alma.cl*

Proof of Concept

In the era of large observatories and huge data transmission through the networks, we did a proof of concept using Riverbed technology to address the ALMA data distribution challenge: limited bandwidth, far away locations and large amounts of data. In the ALMA case from the Atacama desert in Chile to our regional centers in the northern hemisphere. The PoC shows the percentage of reduction in network traffic, after applying de-duplication, tcp window optimization and virtual latency improvement on scienctific data.

P4.13: Runing GTC data reduction pipelines in Jupyter

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):

Sergio Pascual (Universidad Complutense de Madrid)
Nicolas Cardiel (Universidad Complutense de Madrid)

Contact: *sergiopr@fis.ucm.es*

The data reduction pipelines for the GTC (Gran Telescopio Canarias) instruments EMIR and MEGARA are based on the same framework (numina, https://zenodo.org/record/1341361). The pipeline can be run either automatically at the telescope or using a command line interface. We have added support to run the pipeline inside a Jupyter notebook, with the python interpreter. The new classes in numina provide persistent storage of reductions and querying capabilities, to retrieve older reductions

P13.13: Anticipated Data Products from the Europa Imaging System (EIS) on Europa Clipper

Theme: Other
Author(s):
Wes Patterson (Johns Hopkins University Applied Physics Laboratory)
Elizabeth Turtle (Johns Hopkins University Applied Physics Laboratory)
Alfred McEwen (University of Arizona)
Randolph Kirk (United States Geological Survey)
Carolyn Ernst (Johns Hopkins University Applied Physics Laboratory)

Contact: *Wes.patterson@jhuapl.edu*

EIS consists of wide-angle (WAC) and narrow-angle (NAC) cameras, and each can operate either in pushbroom mode with up to 6 colors (plus clear) or in framing mode in the clear bandpass. The NAC has a 2-axis gimbal that can point +/- 30 degrees from nadir. The Europa Clipper mission will include over 40 close flybys of Europa with varying geometries, leading to a variety of image products. Standard (pipeline) data products will include raw data, calibrated and map-projected products, a suite of mosaics, and Digital Terrain Models (DTMs). Map-projected products will have up to 7 co-registered images in clear and 6 color bandpasses. Special challenges include removal of geometric distortions due to spacecraft pointing jitter, photometric normalization, and stray light corrections to best search for and study potential cryo-volcanic plumes. Advanced data products will include a global mosaic at 50 m/pixel, global color mosaic at 300 m/pixel, regional mosaics at 10-50 m/pixel, a global map of bolometric hemispheric albedo for thermal studies, geodesy datasets to measure Europas orbital libration, limb fits for global shape to constrain ice shell properties, change detection sequences to discover current activity, and co-registered products with other Europa Clipper remote sensing data (UV, near-IR, thermal-IR, radar altimetry and sounding). Both cameras can acquire stereo images and we will produce DTMs of selected areas, including along each nadir illuminated ground-track with the WAC, to characterize clutter in radar sounding profiles. The highest-resolution images will be greater than or equal to 0.5 m/pixel scale, sufficient to evaluate candidate landing site for a future landed mission. Public outreach products will include full-disk color images, stereo anaglyphs, and color and high-resolution flyover movies.

P6.9: Transforming Science Code into Maintainable Software, Insights into the G-CLEF Exposure Time Calculator

Theme: DevOps Practices in Astronomy Software
Author(s):
Charles Paxson (Smithsonian Astrophysical Observatory)

Contact: *cpaxson@cfa.harvard.edu*

We explore a common workflow in research institutions where science code is transformed into robust, maintainable, and expandable code. The case study presented is the Exposure Time Calculator (ETC) for the Giant Magellan Telescope Consortium Large Earth Finder (G-CLEF) spectrometer. We describe the process we took to develop requirements documentation and a web application from science code. The ETC provides a rich set of features to help the scientists estimate the performance of the instrument including: the computation of exposure time, SNR, and precision radial velocity, GUI text results, downloadable FITS standard compliant summary of results, and graphical displays. We highlight the importance of a requirements document for information exchange between scientist and engineer, where principles and assumptions can be collaboratively understood and solidified. As the document matures, scientists may use it to specify new requirements. We discuss the importance of making physical interpretations of the code, of understanding and ultimately cleaning of science code magic numbers, and of comprehending the overall flow. This detailed analysis is important since requirements morph as the project progresses. Therefore, a modular design, especially segmenting calibration tables from source code, allow hardware engineering data upgrades without affecting software code. Science code need not be efficient in processing speed, and illuminating black boxes allow potential speed improvements. The software implementation is verified by comparison with hand calculations by scientists and standard data sets (in our case, flux comparisons to stellar standards of known type), and implementing best practices such as unit tests. We analyze this process for our own purposes as well as sharing this process to assist future software engineers.

P4.14: Reprocessing all the XMM-Newton scientific data: a challenge for the Pipeline Processing System

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Jose Vicente Perea-Calderon (RHEA for ESA/ESAC)

Contact: *jose.perea@sciops.esa.int*

2019 will mark the 20-year anniversary of the XMM-Newton Mission. So far, the mission has successfully completed a total of around 14.000 pointing observations, and it is expected to continue for many more years, producing a huge number of high-quality science data products.

Data processing of those observations is carried out by the XMM-Newton Pipeline Processing System (PPS) and the products are delivered to the XMM-Newton Science Archive (XSA). During the two decades many changes have been implemented in the data processing software, partly following improvements to the calibration of the science instruments. Several re-processing campaigns have been undertaken along the mission in order to have an up-to-date and uniformly processed set of high-level science data products in the archive. A new re-processing exercise is to be carried out, as it has been more than six years since the last re-processing campaign in 2011.

Unlike the daily mission operations where a a limited number of observations have to be processed by PPS, a whole mission re-processing is a real challenge. An individual XMM-Newton Pipeline job (of one observation) can take up to five hours of computer processing time, some of them even longer. To achieve the processing of thousands of observations in a reasonable period of time requires a special preparation including a deep analysis of the computing resources. An extreme optimization of the resources sharing becomes essential in our case.

Besides the optimization of the computing infrastructure usage, a set of software tools had to be developed in order to cope with the management and monitoring of this enormous number of individual Pipeline jobs.

P6.10: Centralisation and management of science operations procedures and test cases using SOCCI

Theme: DevOps Practices in Astronomy Software
Author(s):
Fernando Perez Lopez (ESAC)
Vicente Navarro (ESAC)
Kaarel Hanson (CGI Finland)
Kaarel Lumi (CGI, Finland)
Henri Liiva (CGI, Finland)

Contact: *fernando.perez@esa.int*

The Science Operations Configuration Control Infrastructure (SOCCI) is a single, highly customizable platform for system engineering, providing tools and guidelines for: requirement management, problem and change management, test management, project and document management, source version control and continuous integration. This infrastructure provides support the software development and maintenance processes of SCI-O Science Operations units at the European Space Astronomy Centre (ESAC). SOCCI reduces effort and knowledge to setup & maintain the Systems Engineering Environment (SEE) and supports the users by providing guidelines and good practices learnt from previous experiences.

The development of SOCCI started in 2014 and it is being operationally used from June 2017. However, the range of functionalities already covered by SOCCI have been recently extended through SOCCI Evolution and SOCCI Test Framework projects. This paper describes the design, implementation and use of SOCCI for two major new functionalities: the operational procedures management and documentation including their scheduling and automatic execution; and the testing automation including the importing and exporting of test results from external test tools.

P13.14: The CDS HEALPix library

Theme: Other
Author(s):
Pineau (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)
Fernique (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)
Boch (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)

Contact: *francois-xavier.pineau@astro.unistra.fr*

The CDS is releasing a new HEALPix library implemented in Java. Before possibly porting it in C, we are experimenting with the Rust programming language. It allows the library to be compiled into WebAssembly or native code, and thus to easily plug it into web browsers, Python codes or PostgreSQL, to name but a few. The library is distributed under the 3-clause BSD licence. It focuses on our own needs, on performances and accuracy.

We are investigating the potential usage of the WebAssembly version into Aladin Lite. The objective is twofold: supporting deeper orders (up to 24, the present limit being 13), and changing the current GPL license to a less restrictive one. Aladin desktop has already started to resort to the Java version.

Unlike the "official" library, coming from the cosmology community, our Java version do not currently supports Spherical Fourier Transformations and do not support the RING scheme (also it is able to transform cells numbers from the nested scheme to the ring scheme and vice-versa). In return, it do support additional features like: distinguishing between cells partially and fully overlapped by a cone; exact cells-overlapped-by-cone solution; very-fast approximate cells-overlapped-by-cone function dedicated to cross-matches; supporting self-intersecting polygons of any size; fast ordered list of small cells surrounding a larger cell; MOC support in cells-overlapped-by-cone queries; BMOC support (MOC storing an additional flag for each cell); etc.

I11.1: The BagIt Packaging Standard for Interoperability and Preservation

Theme: Quality Assurance of Science Data
Author(s):

Raymond Plante (National Institute of Standards and Technology)

Time: Mon 14.30 - 15.00

Contact: *raymond.plante@nist.gov*

BagIt is a simple, self-describing format for packaging related data files together which is gaining traction across many research data systems and research fields. One of its great advantages is in how it allows a community to define and document a profile on that standard–that is, additional requirements on top of the BagIt standard that speaks to the needs of that community. In this presentation, I will summarize the key features of the standard, highlight some important profiles that have been defined by communities, and talk about how this standard is being used as part of the NIST Public Data Repository. I will compare and contrast the use of BagIt for enabling interoperability (e.g. for transfering data between two systems) and its use for preseravation. I will then give an overview of the NIST BagIt profile for preservation as well as introduce a general-purpose MultiBag profile which addresses issues of evolving data and scaling to large datasets.

P10.12: The VLITE Database Pipeline

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Emil Polisensky (NRL)
Emily E. Richards (ATA, LLC)
Tracy Clarke (NRL)
Namir E. Kassim (NRL)
Wendy Peters (NRL)

Contact: *emil.polisensky@nrl.navy.mil*

A post-processing pipeline to adaptively extract and catalog point sources has being developed to enhance the scientific value and accessibility of data products generated by the VLA Low-band Ionosphere and Transient Experiment (VLITE; http://vlite.nrao.edu/) on the Karl G. Jansky Very Large Array (VLA). In contrast to other radio sky surveys, the commensal observing mode of VLITE results in varying depths, sensitivities, and spatial resolutions across the sky based on the configuration of the VLA, location on the sky, and time on source specified by the primary observer for their independent science objectives. Previously developed tools and methods for generating source catalogs and survey statistics proved inadequate for VLITE's diverse and growing set of data. A raw catalog of point sources extracted from VLITE images is created from source fit parameters stored in a queryable database. Point sources are measured using the Python Blob Detector and Source Finder software (PyBDSF; Mohan & Rafferty 2015). Sources in the raw catalog are associated with previous VLITE detections in a resolution- and sensitivity-dependent manner, and cross-matched to other radio sky surveys to aid in the detection of transient and variable sources. Final data products include separate, tiered point source catalogs grouped by sensitivity limit and spatial resolution.

P12.12: Efficient FFT-based $\mathcal{F}$-statistic implementation for continuous-gravitational-wave searches

Theme: Algorithms
Author(s):

Reinhard Prix (Albert-Einstein-Institute Hannover)
Chris Messenger (School of Physics & Astronomy, University of Glasgow)
Maximillian Bensch (Albert-Einstein-Institute Hannover)
Christophe Choquet

Contact: *reinhard.prix@aei.mpg.de*

We present a highly efficient implementation of the coherent $\mathcal{F}$-statistic for continuous gravitational waves, which leverages the power of the FFT for substantial speed improvements compared to the ("Demodulation") algorithm used by Einstein@Home previously. The basic idea for this goes back to Jaranowski, Krolak, Schutz (1998). Our implementation features a number of performance and usability improvements compared to previous implementations, among other things the ability to chose an arbitrary frequency resolution, internal enforcement of efficient power-of-two FFTs, and the use of optimal time-domain (windowed-)sinc-interpolation for barycentering. This code has been ported to GPUs (using openCL), resulting in further efficiency gains by 1-2 orders of magnitude compared to the CPU version.

O7.1: ESASky: A New Window for Solar System Data Exploration

Theme: Software for Solar Systems Astronomy

Author(s):
Elena Racero (ESAC)
Fabrizio Giordano (ESAC)
Benoit Carry (Observatoire de la Cte d'Azur)
Jerome Berthier (IMCCE-Observatoire de Paris)
Juan Gonzalez (ESAC)
Deborah Baines - ESAC Bruno Merin - ESAC Jesus Salgado - ESAC H.Norman - ESAC M.Lpez-Caniego - ESAC Pilar de Teodoro - ESAC Christophe Arviset - ESAC

Time: Tue 11.45 - 12.00

Contact: *eracero@sciops.esa.int*

Allowing the solar system community fast and easy access to the astronomical data archives is a long-standing issue. Moreover, the everyday increasing amount of archival data coming from a variety of facilities, both from ground-based telescopes and space missions, leads to the need for single points of entry for exploration purposes. Efforts to tackle this issue are already in place, such as the Solar System Object Image Search by the Canadian Astronomy Data Centre (CADC), plus a number of ephemeris services, such as Horizons (NASA-JPL), Miriade (IMCCE) or the Minor Planet & Comet Ephemeris Service (MPC).

Within this context, the ESAC Science Data Centre (ESDC), located at the European Space Astronomy Centre (ESAC) has developed ESASky (http://sky.esa.int), a science driven discovery portal to explore the multi-wavelength sky providing a fast and intuitive access to all ESA astronomy archive holdings. Released in May 2016, ESASky is a new web application that sits on top of ESAC hosted archives, with the goal of serving as an interface to all high-level science products generated by ESA astronomy missions. The data spans from radio to x-ray and gamma-ray regimes, with Planck, Herschel, ISO, HST, XMM-Newton and Integral missions.

We present here the first integration of the search mechanism for solar system objects through ESASky. Based on IMCCE Eproc software for ephemeris precomputation, it allows fast discovery of photometry observations from ESA missions that potentially contain those objects within their Field Of View. In this first integration, the user is able to input a target name and retrieve on-the-fly the results for all the observations from the above-mentioned missions that match the input provided, that is, that contains within the exposure time frame the ephemerides of such objects.

Finally, we will also discuss current developments and future plans in strong collaboration with some of the main actors in the field.

F2: SciServer: Collabroative data-driven science

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles
Author(s):
M. Jordan Raddick (Johns Hopkins University)

Time: Mon 15.30 - 16.00

Contact: *jordan.raddick@gmail.com*

The SciServer team is pleased to announce its final production system of SciServer, offered free to the scientific community for collaborative scientific research and education.

SciServer is an online environment for working with scientific big data, and specifically datasets hosted within our ecosystem. Researchers, educators, students, and citizen scientists can create a free account to get 10 GB of file storage space and access to a virtual machine computing environment. Users can run Python, R, or Matlab scripts in that environment through Jupyter notebooks. Scripts can be run either in Interactive mode, which displays results within the notebook, or in Batch mode, which writes results to the users personal database and/or filesystem.

SciServer hosts a number of datasets from various science domains; within astronomy, it features all data releases of the Sloan Digital Sky Survey (SDSS), as well as other datasets from GALEX, Gaia, and other projects. The SciServer system also incorporates the popular SkyServer, CasJobs, SciDrive, and SkyQuery astronomy research websites, meaning that SciServer Compute offers APIs to read and write from these resources. All these features ensure that computation stays close to data, resulting in faster computation therefore faster science. SciServer also allows users to create and manage scientific collaborations around data and analysis resources. You can create groups and invite collaborators from around the world. You and your collaborators can use these groups as workspaces to share datasets, scripts, and plots, leading to more efficient collaboration.

We will present a highly interactive demo of SciServer, highlighting the latest features and with an emphasis on science use cases. Please bring your questions let us know what you have not yet been able to do with SciServer, and we will help you do it. We are actively looking for new collaborations, feature requests, and scientific use cases. Please let us know how we can help you do your science!

O3.3: Analysis of Astronomical Data using VR: the Gaia catalogue in 3D

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Emanuel Ramirez (Quasar Science Resources)
Juan Gonzlez Nez (ESA-ESAC)
Jos Hernandez (ESA-ESAC)
Jess Salgado (ESA-ESAC)
Alcione Mora (ESA-ESAC)
Uwe Lammers (ESA-ESAC), Bruno Mern (ESA-ESAC), Deborah Baines (ESA-ESAC), Guido de Marchi (ESA-ESTEC), Christophe Arviset(ESA-ESAC)

Time: Mon 10.00 - 10.15

Contact: *eramirez@quasarsr.com*

Since 2016, the ESAC Science Data Centre have been working on a number of Virtual Reality projects to visualise Gaia data in 3D. The Gaia mission is providing unprecedented astrometric measurements of more than 1 billion stars. Using these measurements, we can estimate the distance to these stars and therefore project their 3D positions in the Galaxy. A new application to analyse Gaia DR2 data will be publicly released for Virtual Reality devices during 2018. In this presentation we will give a demo of the latest version of the Oculus Rift application and will show specific use cases to analyse Gaia DR2 data as well as a demonstration on how can Virtual Reality be integrated into a data analysis workflow. We will also show how can new input techniques such as hand-tracking can bring new levels of freedom in how we interact with data.

O7.2: The PDS Approach to Science Data Quality Assurance

Theme: Software for Solar Systems Astronomy

Author(s):
Anne C Raugh (University of Maryland)

Time: Tue 12.00 - 12.15

Contact: *araugh@umd.edu*

The Planetary Data System (PDS) has been mandated by NASA not merely to preserve the bytes returned by its planetary spacecraft, but to ensure those data are usable through generations - 50-100 years into the future. When PDS accepts data for archiving, it must be complete, thoroughly documented, and as far as possible autonomous within the archive (that is, everything needed to understand and use the data must be in the archive as well). Two pillars support the PDS mission: The PDS4 Information Model, and the mandatory External Peer Review.

The PDS4 Information Model codifies metadata not just for structure, but for provenance, interpretation, and analysis as well. The XML document structures defined for the current implementation of the model and its various constituent namespaces define minimum requirements and present best practices for describing all these aspects of the archival data. The schematic enforcement of these requirements provides a simple, automated approach to ensuring the metadata are present and well-formed.

The PDS External Peer Review is required for all data prior to acceptance for archiving. Equivalent to the refereeing process for journal articles, The PDS External Peer Review presents the candidate data to discipline experts unaffiliated with the creation of the data. These reviewers exercise the data in its archival form by reproducing published results, doing comparative analysis between the candidate data and similar or correlated results, and so on, using only the archival resources. These reviewers then determine if the data are of archival quality, and where needed, formulate a list of corrections and additions required prior to archiving.

Together, the Information Model guides data preparers to producing well-formatted, well-documented data products while the External Peer Review ensures the archive submission is complete, usable, and of sufficient quality to merit permit preservation - and support - as part of the Planetary Data System archives.

P4.15: MeerKAT: Operational Workflow and Data Analysis

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Rosly Renil

Contact: *rosly@ska.ac.za*

MeerKAT, the next generation radio telescope is already operational for commissioning and analysing large datasets for science community. This poster explains the operational processes in place for engineers, scientists, commissioners, etc to analyse the data-sets for further data mining that improves various processes. It is also worth to note the software processes and hardware involved, different analysis tools used for telescopic operations in producing good science data.

P6.11: Automating the Fermi Software Development Process with Continuous Integration Practices

Theme: DevOps Practices in Astronomy Software
Author(s):
Alexander Reustle (NASA / GSFC)

Contact: *alexander.w.reustle@nasa.gov*

In 2017 The Fermi LAT collaboration embarked on a plan to modernize its legacy system for software development. The goal was to employ a set of techniques known collectively as Continuous Integration to improve reliability and speed of release. This brief talk will describe the strategies and tools employed, as well as the experiences gained while migrating to this new system.

P3.7: New visualization Features in Firefly

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Trey Roby (IPAC, Caltech)
Xiuqin Wu (IPAC, Caltech)

Contact: *roby@ipac.caltech.edu*

Firefly is IPAC's Advanced Astronomy WEB UI Framework. It was open sourced in 2015, hosted at GitHub. Firefly is designed for building a web-based front end to access science archives with data visualization capabilities. Firefly provides integrated interactive data visualization capabilities with images, catalogs, and plots. It has been used in many IPAC IRSA applications, in LSST Science Platform Portal, and in NEDs newly released interface.

The team has been working on adding new features to Firefly visualization for last year. The most recent addition is HiPS image display, released to public in IRSA Viewer. This talk gives an overview of Firefly's new HiPS support and how we make the smooth transition between HiPS and FITS image displays.

P4.16: Euclidizing external tools: an example from SDC-IT on how to handle software and humanware

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Erik Romelli (INAF - OATs)
Marco Frailis (INAF - OATs)
Samuele Galeotta (INAF - OATs)
Daniele Tavagnacco (INAF - OATs)
Davide Maino (University of Milano, INFN)
Claudio Vuerli, INAF-OATs Gianmarco Maggio, INAF-OATs Giuliano Taffoni, INAF-OATs

Contact: *erik.romelli@inaf.it*

Euclid is an upcoming space mission aimed at studying the dark Universe and understanding the nature of the so called Dark Matter and Dark Energy. The launch is scheduled at the moment for the 2021. A huge amount of data, up to more than 70 PB, will be produced by the two on-board instruments (VIS and NISP) and the data processing will be a crucial aspect of the mission, especially dealing with the performance of the code involved in the scientific analysis.

Due to the expected amount of data and estimated number of cores a distributed computing system on several Science Data Centers (SDCs) has been implemented. To ensure a uniform environment in all SDCs, any software designed for Euclid must comply with a set of common rules and must be implemented in a predefined framework.

Not all the code is designed and implemented ex novo for Euclid purposes; usually data analysis pipelines inherit already existing software tools, designed outside the Euclid Consortium. SDCs are in charge of the integration of external code within the official Euclid software environment.

We will present an overview of how that was done at the Italian SDC in Trieste and present some practical examples related to the integration of an external tool into the Euclid environent. The topic will be discussed taking into account the technicality and focused on the crucial, but usually ignored, aspect of human interfaces.

P5.15: Stellar atmospheric parameters from full spectrum fitting of intermediate and high-resolution spectra against PHOENIX/BT-Settl synthetic stellar atmospheres

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Evgenii Rubtsov (Sternberg Astronomical Institute MSU; M.V. Lomonosov Moscow State University, Faculty of Physics)
Igor Chilingarian (Smithsonian Astrophysical Observatory; Sternberg Astronomical Institute MSU)

Contact: *ev.rubtcov@physics.msu.ru*

We present a new technique implemented in IDL for determination of the parameters of stellar atmospheres using PHOENIX and BT-Settl synthetic stellar spectra. The synthetic spectra provide good coverage in the Teff, logg, [Fe/H], [/H] parameter space over a wide wavelength range and allow to fit observed spectra of a vast majority of stars. Our procedure also determines radial velocities and stellar rotation, and it takes into account flux calibration imperfections by fitting a polynomial continuum. Thanks to using pixel fitting, we can exclude certain spectral features, which are not present in the models, such as emission lines (chromospheric emission in late-type stars or discs around Be stars). We perform a non-linear chi2 minimization with the Levenberg-Marquardt method that is applied to the entire spectrum, with the exception of areas with peculiarities: emission lines, model shortcomings (incompleteness of the spectral line lists used for the atmospheric model calculation). We take into account systematic errors of the surface gravity estimates introduced by synthetic atmospheres by applying a correction computed from the comparison of our results with those obtained using asteroseismology. We present the comparative statistical analysis of optical spectral libraries ELODIE, INDO-US, MILES, UVES-POP, and a new near-infrared Las Campanas Stellar Library and discuss prospective applications of our technique.

P13.15: Science Data Pipeline of NICER

Theme: Other
Author(s):
Kristin Rutkowski

Contact: *kristin.l.rutkowski@nasa.gov*

NICER (Neutron star Interior Composition ExploreR Mission) data flow from the mission aboard ISS, through the science processing pipeline, and to the High Energy Astrophysics Archive. This poster will describe the different steps in the pipeline, detailing how the data are converted to required formats, the different flows of data, and how the status of the data is tracked throughout the entire cycle. We use a combination of Python, Perl, and shell scripts, along with a relational database, to move the data through the pipeline. The HEASOFT software package is used to process the data. The data are moved to long term storage at the HEASARC, where interested parties may retrieve them for further study.

P13.16: Schroedingers code: Source code availability and transparency in astrophysics

Theme: Other
Author(s):
P. Wesley Ryan (Astrophysics Source Code Library)
Alice Allen (University of Maryland)
Peter Teuben (University of Maryland)

Contact: *wes@ascl.net*

Astronomers use software for their research, but how many of the codes they use are available as source code? We examined a sample of 166 papers from 2015 for clearly identified software use, then searched for source code for the software packages mentioned in these research papers. We categorized the software to indicate whether source code is available for download and whether there are restrictions to accessing it, and if source code was not available, whether some other form of the software, such as a binary, was. Over 40% of the source code for the software used in our sample was not available for download.

As URLs have often been used as proxy citations for software and data, we also extracted URLs from one journals 2015 research articles, removed those from certain long-term reliable domains, and tested the remainder to determine what percentage of these URLs were accessible in September and October, 2017. We repeated this test a year later to determine what percentage of these links were still accessible. This poster will present what we learned about software availability and URL accessibility in astronomy.

P10.13: Gaia DR2 and the Virtual Observatory: VO in operations new era

Author(s):
Jesus Salgado (ESDC/ESAC)
Juan Gonzalez (ESDC/ESAC)
Raul Gutierrez (ESDC/ESAC)
Juan Carlos Segovia (ESDC/ESAC)
Alcione Mora (Gaia SOC/ESAC)
Thomas Boch CDS Mark Allen CDS Nigel C. Hambly ROE Stelios Voutsinas ROE Markus Demleitner ARI Gregory Mantelet ARI, CDS Javier Duran ESDC/ESAC Elena Racero ESDC/ESAC Pilar de Teodoro ESDC/ESAC Bruno Merin ESDC/ESAC Christophe Arviset ESDC/ESAC

Contact: *jesusjuansalgado@gmail.com*

During the last decade, the IVOA (International Virtual Observatory Alliance) has been tasked with the difficult task of defining standards to interchange astronomical data. These efforts have been supported by many IVOA partners in general and by the ESAC Science Data Centre (ESDC) in particular, that have been collaborating in the definition of standards and in the development of astronomical VO-inside archives. New ESDC archives, like Gaia, ESASky and the ones in development like Euclid, makes use of VO standards not only as a way to expose the data but, also, as the architectural design of the system.

The Gaia Data Release 2 archive, has been a global effort done not only concentrated into the central archive at ESAC, but also as a collaboration with partner data centers like CDS, ARI, ROE and other members of DPAC. All Gaia partners make use of VO protocols to expose Gaia data as a principle. With this release, the level of dissemination and community endorsement of the VO protocols have entered into a new phase. The percentage of the Gaia expert community that makes use of VO standards has been increased to an unprecedented level of use; 34,000 users accessing the ESA Gaia Archive interface; over 5,000 advanced users sending more than 1,5 million data analysis queries during the only the first week and just counting the ESA Gaia Archive.

These standards are offered to the community in a transparent way (like SAMP or VOSpace to interchange data), as VO procotols extensions like the Tabular Access Procol extension (TAP+) or as direct use, like the Astronomical Data Access Language (ADQL) that the users learn in order to implement data mining scientific use cases that were almost impossible in the past.

We will describe how VO protocols simplify the work of design and implementation of the astronomical archives and the current level of endorsement by the scientific community.

P10.14: An overview of the OVGSO data centre

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Michele Sanguillon (LUPM)
Charlotte Vastel (IRAP)
Jean-Michel Glorian (IRAP)

Contact: *Michele.Sanguillon@umontpellier.fr*

The OVGSO (Observatoire Virtuel du Grand Sud-Ouest: https://ov-gso.irap.omp.eu/) is one of the six French Astrophysical Data Center recognized by INSU (Institut National des Sciences de lU-nivers) since 2013. It aims at providing a common environment for all the Astrophysical researches, based on observed/theoretical/modeled data access and tools interoperability. The OVGSO gathers five different themes: Sun-Earth (STORMS: https://stormsweb.irap.omp.eu/, CLIMSO-DB: http://climso.irap.omp.eu/), planetary plasmas (CDPP: http://cdpp.irap.omp.eu/), inter-stellar medium (CASSIS: http://cassis.irap.omp.eu/, CADE: http://cade.irap.omp.eu/, KIDA: http://kida.obs.ubordeaux1.fr/), stellar spectra (PolarBase: http://polarbase.irap.omp.eu/, Pol-lux: http://pollux.oreme.org/), high energy astrophysics (SCC-XMM: http://xmmssc.irap.omp.eu/). We will present in the poster the different tools that have been developed and the OV standards and protocols that have been used within this Data Center. For example, spectroscopic observa-tions can be retrieved using SSAP (Single Spectral Access Protocol) to access the IVOA services, to reach some observed data and link them to theoretical atomic and molecular databases such as VAMDC (Virtual Atomic and Molecular Data Centre: http://www.vamdc.eu/) using TAP (Table Access Protocol). SAMP (Simple Application Messaging Protocol) is also used to send spectra on the tools developed by OVGSO. The goal of the OVGSO is to share common tools for different thematics in astrophysics for a better use of the astrophysical data. This need has revealed to be crucial with the wealth of high-spatial/ spectral resolution spectra for the past 15 years, with new generation of ground/space-based observatory. Dealing with a large quantity of data leads to optimized tools and a better and common access to the many databases.

P1.11: A hybrid neural network approach to estimate galaxy redshifts from multi-band photometric survey.

Theme: Machine Learning in Astronomy
Author(s):

Felipe Souza (Instituto Nacional de Pesquisas Espaciais)
Amita Muralikrishna (Instituto Nacional de Pesquisas Espaciais)
Walter dos Santos Jr (Instituto Nacional de Pesquisas Espaciais)
Rafael Santos (Instituto Nacional de Pesquisas Espaciais)

Contact: *rafael.santos@inpe.br*

Machine learning methods have been used in cosmological studies to estimate variables that would be hard or costly to measure precisely, like, for example, estimating redshifts from photometric data. Previous work showed good results for estimating photometric redshifts using nonlinear regression based on an artificial neural network (Multilayer Perceptron). One of the problems identified on the previous work is that a single network with multiple neurons on the hidden layer for the regression may not be able to yield the best results. In this work we explore a hybrid neural network approach that uses a Self-Organizing Map to separate the original data into different groups, then applying the Multilayer Perceptron to each neuron on the Self-Organizing Map to obtain different regression models for each group. Preliminary results indicate that in some cases better results can be achieved, although the computational cost may be increased.

P1.12: Chatting with the astronomical data services.

Theme: Machine Learning in Astronomy
Author(s):
Andr Schaaff (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)
Alexis Guyot (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)
Thomas Boch (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)
Sbastien Derriere (Universit de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbour)

Contact: *andre.schaaff@astro.unistra.fr*

In our everyday life, we use increasingly the voice to interact with assistants for heterogeneous requests (weather, booking, shopping, etc.). We present our experiments to apply the Natural Language Processing to the querying of astronomical data services. It is of course easy to proto-type something. But is it realistic to propose it as a new way of interaction in a near future, as an alternative to the traditional forms exposing parameter fields, check boxes, etc.? To answer to this question, it is necessary to answer before to the most fundamental question: is it possible to satisfy professional astronomers needs through this way? We have not started from scratch as we have useful tools and resources (name resolver, authors in Simbad, missions and wavelengths in VizieR, etc.) and the Virtual Observatory (VO) brings us standards like TAP, UCDs, ..., imple-mented in the CDS services. The interoperability, enabled by the VO, is a mandatory backbone. We explain how it helps us to query our services in Natural Language. And how it will be possible in a further step to query the whole VO through this way. We present our pragmatic approach based on a chatbot interface (involving Machine Learning) to reduce the gap between good and imprecisely/ambiguous queries. Comments (necessarily enthusiastic) are welcome. Collaborations too.

P11.6: The IVOA Provenance Data Model

Theme: Quality Assurance of Science Data

Author(s):
Mathieu Servillat (LUTH, Observatoire de Paris - CNRS - PSL, France)
Kristin Riebe (AIP, Potsdam, Germany)
Franois Bonnarel (CDS, Strasbourg, France)
Anastasia Galkin (AIP, Potsdam, Germany)
Mireille Louys (CDS, Icube, Strasbourg, France)
Markus Nullmeier, ARI, Heidelberg, Germany Michle Sanguillon, LUPM, Montpellier, France Ole
Streicher, AIP, Potsdam, Germany

Contact: *mathieu.servillat@obspm.fr*

I will present the Provenance data model proposed as a standard to the IVOA. The objective
is to describe how provenance information can be modeled, stored and exchanged within the as-
tronomical community in a standardized way. We follow the definition of provenance as proposed
by the W3C, i.e. that "provenance is information about entities, activities, and people involved
in producing a piece of data or thing, which can be used to form assessments about its quality,
reliability or trustworthiness." Such provenance information in astronomy is important to enable
any scientist to trace back the origin of a dataset (e.g. an image, spectrum, catalog or single
points in a spectral energy distribution diagram or a light curve), a document (e.g. an article, a
technical note) or a device (e.g. a camera, a telescope), learn about the people and organizations
involved in a project and assess the quality as well as the usefulness of the dataset, document or
device for her own scientific work.

P5.16: ASVO MWA project: lower technical barrier to access MWA data

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Robert Shen (Astronomy Australia Limited)
Randall Wayth (ICRAR/Curtin University)
Andreas Wicenec (ICRAR/University of Western Australia)

Contact: *robert.shen@astronomyaustralia.org.au*

The Murchison Widefield Array (MWA) is a low-frequency radio telescope operating between 80 and 300 MHz. The MWA project aims to perform large surveys of the entire southern sky hemisphere and acquire deep observations on targeted regions. For the last five years, the MWA has generated more than 20Pb of raw data, but there has been no easy way for the astronomical community to access this data or for MWA researchers to share data among their collaborators. The ASVO MWA project builds on the previous pilot project and aims to address these issues by reducing the technical barriers for the community to discover, download and use both public and proprietary MWA data. Specifically, the ASVO MWA project has four service components: (1) Calibration component, which allows users to apply a system determined calibration solution to their selected data products using an IVOA compliant UWS based pipeline; (2) Authentication component, which provides users with a single sign-on via EduGAIN by collaborating with Spherical Cow Group (SCG) on their existing Identify Management Project; (3) VOSpace component, which provides an IVOA compatible VOSpace service and functionality to enable users to interact with their data in an IVOA compatible environment; (4) MWA web development component, which provides a modern web user experience by enhancing the existing ASVO MWA pilot GUI and system management interfaces. The ASVO MWA project will lower the barrier for astronomers to enter the world of big data analysis, processing and computational modelling. Further steps in the development of the ASVO MWA over the next few years will enhance ASVO MWA functionality and act as an early pilot towards the forthcoming SKA regional centre. In this talk, we will present details of the implementation and look-and-feel of the ASVO MWA project.

P8.6: Applications of the in-memory database Redis in processing transient event alerts

Theme: Time Domain Astronomy
Author(s):
Min-Su Shin (Korea Astronomy and Space Science Institute)

Contact: *msshin@kasi.re.kr*

We present results of using the in-memory database Redis in processing transient event alerts. The Redis works in two different ways for processing alerts. First, the publication-subscription model in the Redis allows us to adopt it as a message delivery system for multiple local alert clients. Second, we use the features of indexing and storing geolocation information in the Redis to enable low-latency matching of transient locations with custom catalogs. The current system collects event alerts by using VOEvent streams and detecting changes in web pages/feeds. We also introduce our efforts of migrating the system from the Redis message delivery environment to the NATS-based message processing configuration as well as application of Uber's H3 spatial indexing model instead of the Redis geolocation support.

P13.17: Pipeline Processing of Stratospheric Terahertz Observatory (STO-2) Galactic Plane Survey

Theme: Other
Author(s):
R.F. Shipman (SRON)
Y.M. Seo (JPL Caltech)
V. Tolls (Harvard Smithsonian Center for Astrophysics)
P.F. Goldsmith (JPL Caltech)
C. Kulesa (University of Arizona)

Contact: *russ@sron.nl*

The Stratospheric Terahertz Observatory is a balloon born experiment which observed 2.5 deg2 of the Galactic Plane in the [CII] 158 micron transition. We describe the STO-2 data processing and data products as well as the challenges present with these data.

P13.18: VO service in Japan : Registry service based on Apache Solr and SIA v2 service for Japanese Facilities

Theme: Other
Author(s):
Yuji Shirasaki (National Astronomical Observatory of Japan)
Christopher Zapart (National Astronomical Observatory of Japan)
Masatoshi Ohishi (National Astronomical Observatory of Japan)
Yoshihiko Mizumoto (National Astronomical Observatory of Japan)

Contact: *yuji.shirasaki@nao.ac.jp*

Currently more than 20 thousands of VO services are registered in the VO registry database. Keyword Search is the most popular way to find a resource. There can be a lot of ways to implement this capability, and the performance depends on how to index the document describing the resource metadata.

Apache Solr is an open source search platform and uses the Lucene Java search library for full-text indexing. The document is indexed after the process that removes stop words (not adequate for a keyword e.g. a, the, and, is, are ...) and stems a word to a root word (e.g. clustering to cluster). Thus for the query of a keyword cluster, the Solr search system will return documents that contain a word cluster, clustering, or clustered. Solr also can handle single- or multi-token synonyms and abbreviation. For an example, SN, supernovae, and SNe are translated to supernova in the process of indexing. So it increases the probability for a given keyword to hit a desired resource metadata. In order to incorporate this feature, we upgraded the registry service behind the JVO portal to the one based on Solr from XML based DB.

The data from Nobeyama Legacy project were released from the JVO portal on 1st June in 2018. The data are now distributed also through the most recent VO standard interface called SIA-v2. The data of ALMA and Subaru telescope are also accessible through the SIA-v2. We are now working on distributing the data of Hyper Suprime-Cam SSP/DR1 and also the data of Hitomi satellite under the collaboration with C-SODA/JAXA.

We present those our recent development on Japanese Virtual Observatory system.

P1.13: MeerKAT Radio Frequency Interefence characterization using Machine Learning

Theme: Machine Learning in Astronomy
Author(s):
Isaac sihlangu (University of Cape Town)
Bruce Bassett (University of Cape Town, Square Kilometer Array Project)
Amit Mishra (University of Cape Town)

Contact: *isihlangu@ska.ac.za*

The Square Kilometer Array (SKA) project is an international effort to build the worlds largest radio telescope, with a square kilometre (one million square metres) of collecting area spread over sites in Southern Africa and Australia. It will be the largest and the most sensitive radio telescope ever built and will observe the sky with unprecedented depth and detail. Science results from the SKA telescope are expected to change our understanding of the Universe.

Detection and mitigation of Radio Frequency Interference (RFI) is one of the biggest challenges when analysing the observational data collected from the modern radio telescopes. Due to industrial and technological developments, radio astronomy observations are being threatened by increasing levels of RFI. Also, an important byproduct of the increased collecting area of the these telescopes is their increased sensitivity and as the sensitivity of the these instruments increases, so does the sensitivity to unwanted signals (i.e satellites, mobile cellphones, airplanes, etc). Furthermore the dawn of big data has begun with next generation telescopes such as the SKA that will generate orders of magnitude more data than the current generation. The combination of these issues means that old techniques used to detect and mitigate RFI signals are no longer sufficient. More sophisticated and automated techniques will be required.

My research is focused on finding novel techniques and methods that can provide additional insight into various types and sources of RFI phenomena detected by radio telescopes, specifically via the techniques of machine learning and deep learning. My work aims to characterise and classify RFI signals detected by a brute-force algorithm. I will study and catalogue the different kinds of RFI signals that are detected. Attributes found to be statistically discriminative will be extracted from RFI signals and will be used as feature sets. The attributes extracted will be used to train machine learning algorithms such as Random Forests, Decision Trees and Neural Networks that will classify different types of RFI categories. Thus, this analysis will help in separating outlier signals into different kinds of RFI or signals of astronomical importance. We will also be able to predict when certain kinds of RFI are likely to occur which will aid in proper scheduling of observations.

O5.3: Lilith: A Versatile Instrument and All-Sky Simulator for use with Space-Based Astrophysics Observatories

Theme: Science Platforms: Tools for Data Discovery and Analysis from Different Angles

Author(s):
Jeffrey C. Smith (SETI Institute)
Peter Tenenbaum (SETI Institute)
Jon M. Jenkins (NASA Ames Research Center)
Joseph D. Twicken (SETI Institute)

Time: Thu 09.45 - 10.00

Contact: *jeffrey.smith@nasa.gov*

To help facilitate the development of the Transiting Exoplanet Survey Satalite (TESS) data analysis pipeline, it was necessary to produce simulated flight data with sufficient fidelity and volume to exercise all the capabilities of the pipeline in an integrated way. As a generator of simulated flight data, Lilith, was developed for this purpose. We describe the capabilities of the Lilith software package, with particular attention to the interaction between the implemented features and the pipeline capabilities that it exercises. Using a physics-based TESS instrument and sky model, Lilith creates a set of raw TESS data which includes models for the CCDs, readout electronics, camera optics, behavior of the attitude control system (ACS), spacecraft orbit, spacecraft jitter and the sky, including zodiacal light, and the TESS Input Catalog. The model also incorporates realistic instances of stellar astrophysics, including stellar variability, eclipsing binaries, background eclipsing binaries, transiting planets and diffuse light. This simulated data is then processed through the TESS pipeline generating full archivable data products. Full instrumental and astrophysics ground truth is available and can be used as a training set for TESS data analysis software, such as when training a machine learning classifier for planet candidates. Our intention is to continue to tune Lilith as real TESS flight data becomes available, allowing for an up-to-date simulated set of data products to complement the mission flight data products, thereby aiding researchers as they continue to adapt their tools to the TESS data streams. We discuss the execution performance of the resulting package, and offer some suggestions for improvements for instrument and sky simulators to be developed for other missions.

P1.14: Mock Datasets and Galaxy Merger Statistics from Cosmological Hydro Simulations

Theme: Machine Learning in Astronomy
Author(s):
Greg Snyder (STScI)

Contact: *gsnyder@stsci.edu*

I will describe efforts to blend cosmological simulations with surveys of distant galaxies. In particular, I will discuss our work to create and interpret millions of synthetic images derived from the Illustris project, a recent large hydrodynamic simulation effort. Recently, we showed that because galaxies assembled so rapidly, distant mergers are more common than the simplest arguments imply. Further, we improved image-based merger diagnostics by training many-dimensional ensemble learning classifiers using the simulated images and known merger events. By applying these results to data from the CANDELS multi-cycle treasury program, we measured a high galaxy merger rate in the early universe in broad agreement with theory, an important test of our cosmological understanding.

P13.19: Azimuthal variation of oxygen abundance in galactic stellar discs of the EAGLE simulations

Theme: Other
Author(s):
Martin Solar (Universidad Andres Bello (UNAB))
Patricia Tissera (Universidad Andres Bello (UNAB))

Contact: *msolar95@gmail.com*

Context: The spatial distribution of oxygen abundance in galactic stellar discs is decisive to understand how efficiently metals that are synthesized in massive stars can be redistribute across the galaxy and this provide information about disc assembly, star formation processes and chemical evolution.

Aims: We study the azimuthal variations of the metallicity gradients (slopes and zero points) and confronted with the gradients obtained by performing azimuthal averages.

Methods: To get dispersions of metallicity gradients in the galactic stellar discs, there is necessary to divide the galactic plane in 6 sub-regions. To these 6 sub-regions the metallicity gradient is obtained, which is important to get the median and standard deviation of the slope and the zero point, this will be the transcendental variables. The data analysis es based on the EAGLE simulations.

Results: We get 3 groups of galaxies, separated by his standard deviation of the slope of the metallicity gradient; those with low, medium and high dispersion. From low to high standard deviation of the slope, the galaxies show a tendence to be more massive, with more number of particles and a higher value of the radius at half mass (all for young stellar populations).

Conclusions: We find that there is a dispersion in the metallicity profiles depending on the direction chose to estimate them. The preliminary analysis of the dispersions suggest that there is a dependen on morphology: well-defined discs have less dispersion. They also have larger sSFR, hence, thery are more actively forming stars. There is no clear correlation with the total stellar mass (including the bulge). Next steps will be focused on improving this analysis and comparing with available observations.

P12.13: Tensor Clusters for extracting and summarizing components in spectral cubes

Theme: Algorithms
Author(s):
Mauricio Solar (UTFSM, Chile)
Humberto Farias (UTFSM, Chile)

Contact: *mauricio.solar@usm.cl*

As it is already known, modern observatories like the Atacama Large Millimeter/ sub- millimeter Array (ALMA) and the Very Long Baseline Array (VLBA) generate large- scale data. This problem of data size that has been extensively studied in recent times but there is another equally relevant problem, the data cubes also exhibit unprecedented complexity and dimensionality. These complex sources can be modeled using tensor algebraic methods. In this representation the cube has two physical axes (Right ascension(Ra)/Declination(Dec)), and a third axis (Wavelet, Time, Intensity), a representation in a tensor of order 3. This representation leads to a concise description of cubes of astronomical data. Over this representation was applied a tensor-cluster approach to find the morphology of components expected in these complex data cubes. To achieve this goal the TensorFit library with strong GPU support to handle spectral cubes in a tensor mode was used.

P4.17: Development on Data Analysis Software on GECAM

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Dr.Xinying Song (Institute of High Energy Physics, Chinese Academy of Science)

Contact: *songxy@ihep.ac.cn*

Introduction on Development on Data Analysis Software on GECAM.

P2.7: Contracts Tracking for Astronomical Infrastructure Projects in Jira

Theme: Management of Large Science Project
Author(s):
Travis Stenborg (Astronomy Australia Limited)

Contact: *mqastro@acm.org*

A custom Jira workflow has been developed at Astronomy Australia Limited (AAL) for contract status management. The Jira cloud-hosted system provides quick determination of contracts' statuses for AAL's program of astronomical infrastructure projects. Passive, overview-style contracts' statuses are available from the standard, browser-accessible Jira dashboard and more detailed information accessible via views. Additionally, a configurable level of active contract status updates are provided via integration of workflow transitions with user-specified watchers.

P13.20: Running the Fermi Science Tools on Windows

Theme: Other
Author(s):
Thomas E. Stephens (GSFC/Innovim)

Contact: *thomas.stephens@nasa.gov*

The Fermi Science Tools, publicly released software for performing analysis on science data from the Fermi Gamma-ray Space Telescope, have been available and supported on Linux and Mac since launch. Running the tools on a Windows based host has always required a virtual machine running Linux. New technologies, such as Docker and the Windows Subsystem for Linux has made it possible to use these tools in a more native like environment.

In this poster we look at three different ways to run the Fermi Science Tools on Windows: via a VM, a Docker container, and using the Windows Subsystem for Linux. We present the steps necessary to install the tools, any issues or problems that exist, and benchmark the various installations. While not yet officially supported by the Fermi Science Support Center, these Windows installations are checked by staff when new releases are made.

I10.1: Astronomical archives: Serving up the Universe

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era
Author(s):
Felix Stoehr (ESO/ALMA)

Time: Tue 09.00 - 09.30

Contact: *felix.stoehr@gmx.de*

This talk first briefly reviews some of the current context of storing and making astronomical data discoverable and available. We then discuss the challenges ahead and look at the future data-landscape when the next generation of large telescopes will be online, at the next frontier in science archives where also the content of the observations will be described, at the role machine-learning can play as well as at some general aspects of the user-experience for astronomers.

P11.7: DOI in Astronomy

Theme: Quality Assurance of Science Data
Author(s):
Streicher, Ole (Leibniz Institute for Astrophysics Potsdam)
Enke, Harry (Leibniz Institute for Astrophysics Potsdam)

Contact: *ole@aip.de*

The use of DOI for astronomical data sets is becoming part of good scientific practice. Although a fairly small set of metadata, almost all aspects of findability are covered. The advantage is clearly, that this is no astronomy special but a widely accepted means catering to the quest of interoperability across disciplines, albeit on a basic level. DOI also move the onus of responsibility for proper data sets from the individual scientist to the institutions. Even the IVOA starts to look for ways to incorporate DOI for datasets into their registry and other services. Still much has to be done: the agreement on curation procedures and what would be considered as a data set to stam DOI on. At AIP we provide DOI for a range of datasets, and th poster will provide some insight in the curation procedures and decisions about what we consider a data set.

P1.15: A Machine Learning Approach to Detect Dark Matter Particles Under Extreme Class Imbalance and Large Datasets

Theme: Machine Learning in Astronomy

Author(s):
R. Sutrisno (University of Houston)
R. Vilalta (University of Houston)
A. Renshaw (University of Houston)

Contact: *rasutrisno@uh.edu*

Experiments attempting to directly detect dark matter have reached a high level of sensitivity, where hunting a single dark matter particle interacting within a detector demands filtering out millions -or even billions- of events originating from background sources. The Darkside-50 collaboration is an international experiment conducted at the Laboratori Nazionali del Gran Sasso in Italy, where low-radioactivity liquid argon is utilized within a dual-phase time projection chamber to detect weakly interacting massive particles (WIMPS), one of the leading candidates for dark matter. The Darkside-50 experiment faces two main data-analysis challenges: extreme class imbalance and large datasets. In this paper we show how machine learning techniques can be employed, even under the presence of samples exhibiting extreme class-imbalance (i.e., extreme signal-to-noise ratio). In our data-analysis study, the ratio of negative or background events to positive or signal events is highly imbalanced by a factor of 10^7. This poses a serious challenge when the objective is to identify a signal that can be easily misclassified as background. We develop and employ a modification of the Synthetic Minority Over-Sampling (SMOTE) technique to alleviate the class-imbalance problem, by artificially generating new signal samples using a k-nearest neighbor approach; we enhance the mechanism behind SMOTE to restrict the pool of k-nearest neighbors to minimize the overlap between signal and background events. To expedite the analysis of large quantities of data, we undersample background events. The result is a new sample distribution that facilitates building accurate predictive models. Experimental results on real data obtained from the Darkside-50 experiment show the benefits of our proposed approach.

P10.15: The TESS Science Data Archive

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Daryl Swade (Space Telescope Science Institute)
Scott Fleming (Space Telescope Science Institute)
Jon M. Jenkins (NASA Ames Research Center)
David W. Latham (Harvard-Smithsonian Center for Astrophysics)
Edward Morgan (Massachusetts Institute of Technology)
Susan E. Mullally, Space Telescope Science Institute Roland Vanderspek, Massachusetts Institute of Technology

Contact: *swade@stsci.edu*

The Transiting Exoplanet Survey Satellite (TESS) is a survey mission designed to discover exoplanets around the nearest and brightest stars. TESS is a NASA Astrophysics Explorer mission that was launched on April 18, 2018. The Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute serves as the archive for TESS science data.

TESS will conduct large area surveys of bright stars and known M dwarfs within about 60 parsecs. TESS will observe a single 24 degree by 96-degree sector of the sky for approximately 27 days before pointing to a new sector, surveying almost the entire sky over the two-year prime mission.

Science data are captured in two observation types: Target Data and Full Frame Images. Target Data result from a subarray of pixels read out with a two-minute cadence from approximately 15,000 stars per sector. Target stars change every sector, but targets near the ecliptic poles will be observed in multiple sectors. The entire field of view is captured in Full Frame Images taken at thirty-minute cadence. Full Frame Images provide a rich source of space-quality continuous light curves for many astronomical investigations.

Archive data products originate from multiple elements within the TESS ground segment. The Payload Operations Center provides the archive with raw data from the spacecraft, operational files, and focal plane characterization models. The Science Processing Operations Center pipeline generates FITS files for the Target Data and Full Frame Images. Light Curves and Centroids are extracted from the Target Data. Threshold Crossing Events are identified from the Light Curves in the Transiting Planet Search pipeline and Data Validation Reports are generated. MAST will stage catalog data generated by the TESS Science Office. The TESS Input Catalog at present contains approximately half a billion persistent luminous objects over the entire sky that are potential two-minute targets or are needed to document nearby fainter stars that contaminate the target photometry. The TESS Objects of Interest Catalog lists planetary candidates identified as Threshold Crossing Events. In addition, the TESS Science Office will manage the TESS Follow-up Observing Program in which ground and space based telescopes will be used for further observations of TESS Objects of Interest. Follow-up Observing Program participants will be responsible for submission of follow-up data to MAST.

The services provided by MAST for the TESS mission are to store science data and provide an Archive User Interface for data documentation, search, and retrieval. Target Data and Full Frame Images are expected to be available in MAST within two months after conclusion of a sectors data collection. Current estimates predict 30 Terabytes of TESS data available through MAST for each year of the mission.

The MAST architecture is designed to support multiple missions. MAST currently serves data from the Hubble Space Telescope, Kepler, and approximately ten other missions, as well as numerous additional high-level science data products. MAST will be the JWST data archive. The TESS mission takes advantage of this multi-mission architecture to provide a cost effective archive that allows integration of TESS data with data from other missions.

P1.16: Spectral Classification of Galaxies using Deep Neural Networks and Self-taught Learning

Theme: Machine Learning in Astronomy
Author(s):
Yihan Tao (National Astronomical Observatories, Chinese Academy of Sciences)
Yanxia Zhang (National Astronomical Observatories, Chinese Academy of Sciences)
Chenzhou Cui (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *y.tao@nao.cas.cn*

Spectral classification using machine learning methods is highly in need these days for large sky survey (e.g. LAMOST and SDSS) . These survey produce large amount of spectra dataautomated methods can save human researchers from the painful task of classifying and checking the spectra by eye. Recent years, many research has found it promising by using deep neural networks. However, it has the problem of difficult to understand the model, and the trained model is hardly reusable for a new data set.

In previous work, we carried out experiments using supervised machine learning methods (Logistic Regression, Random Forest and Linear SVM) to select active galactic nucleus (AGNs) from a dataset of galaxy spectra (including AGNs, starburst galaxies, starforming galaxies, and others) provided by SDSS DR14. The results show that the accuracy of these traditional machine learning method is acceptable as the classification accuracy reaches 91%. However, it can be limited because we used linear PCA of 20 components to present spectra features.

In this paper, we investigate the accuracy of applying deep neural networks (DNNs) to learn the feature representation from the galaxy spectra. We trained a DNN of three hidden layer and its test accuracy achieved higher than 95%. To make the model reusable and more interpretable, we also employed self-taught learning (unsupervised feature learning) to let the spectra data speak for itself. The advantage of self-taught learning and unsupervised feature learning is that we can get the algorithms to learn a feature representation from large amount of unlabeled data. This is especially useful for some kind of galaxies (such as LINERs, BL Lacs) which we have relatively small amount of labeled spectra. We trained a 7-layer auto-encoder (2600-1024-256-20-256-768-1024-2600) and then feed into the classifiers. In addition, we present and visualize the features learnt from our auto-encoder and try to explain which part of the spectra is more important for classification.

O12.2: Performance-related aspects in the Big Data Astronomy Era: architects in software optimization

Theme: Algorithms
Author(s):
Daniele Tavagnacco (INAF-OATs)
Marco Frailis (INAF-OATs)
Samuele Galeotta (INAF-OATs)
Davide Maino (INFN)
Erik Romelli (INAF-OATs)
Claudio Vuerli - INAF-OATs, Giuliano Taffoni - INAF-OATs, Gianmarco Maggio - INAF-OATs

Time: Mon 12.00 - 12.15

Contact: *daniele.tavagnacco@inaf.it*

In the last decades the amount of data collected by Astronomical Instruments and the evolution of computational demands have grown exponentially. Today it is not possible to obtain scientific results without prodigious amounts of computation. For this reason, the software performance plays a key role in modern Astronomy data analysis. Scientists tend to write code with the only goal of implementing the algorithm in order to achieve a solution. Code modifications to gain better performance always come later. However, as computing architectures evolve to match the performance that is demanded, the coding task has to encompass the exploitation of the architecture design, the single-processor performance and parallelization. To facilitate this task, programming languages are progressing and introducing new features to fully make use of the hardware architecture. Designing a software that meets performance, memory efficiency, maintainability, and scalability requirements is a complex task that should be addressed by the software architect. The complexity stems from the existence of multiple alternative solutions for the same requirements, which make tradeoffs inevitable. In this contribution we will present part of the activity done at the Italian Science Data Center for the ESAs cosmological space mission Euclid which regards the software performance optimization. In particular, considering the programming languages selected for the development of the Euclid scientific pipelines, we will present some C++ and Python examples focusing on the main aspects of human contribution in the optimization of the code from the performance, memory efficiency and maintainability point of view.

P3.8: TOPCAT and Gaia

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality
Author(s):
Mark Taylor (University of Bristol)

Contact: *m.b.taylor@bristol.ac.uk*

TOPCAT, and its command line counterpart STILTS, are powerful tools for working with large source catalogues. ESA's Gaia mission, most recently with its second data release, is producing source catalogues of unprecedented quality for more than a billion sources. We present here some examples of how TOPCAT and STILTS can be used for analysis of Gaia data.

P13.21: QAC: Quick Array Combinations with CASA

Theme: Other
Author(s):
Peter Teuben (University of Maryland)

Contact: *teuben@astro.umd.edu*

A simple python layer in CASA was developed to aid in writing scripts for array (single dish and interferometric) combination. Although initially developed for TP2VIS, running simulations and comparing with other array combination methods, and adding regressions became cumbersome, and QAC was developed. Both ALMA, ngVLA and CARMA simulations are already supported, but extending to more generic array are planned.

P12.14: Robust Registration of Astronomy Catalog

Theme: Algorithms
Author(s):
Fan Tian (Johns Hopkins University)
Tamas Budavari (Johns Hopkins University)
Amitabh Basu (Johns Hopkins University)

Contact: *ftian4@jhu.edu*

Due to a small number of reference sources, the astrometric calibration of images with a small field of view is often inferior to the internal accuracy. An important experiment with such challenges is the Hubble Space Telescope. A possible solution is to cross-calibrate overlapping fields instead of just relying on standard stars. Following Budavari and Lubow (2012), we use infinitesimal 3D rotations for fine-tuning the calibration but re-formalize the objective to be robust to large number of false candidates in the initial set of associations. Using Bayesian statistics, we accommodate bad data by explicitly modeling the quality which yields a formalism essentially identical to M-estimation in robust statistics. Our preliminary results show great potentials for these methods on simulated catalogs where the ground truth is known.

O4.1: Data Science =! Software Engineering. Exploring a workflow for ALMA operations.

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Ignacio Toledo (Joint ALMA Observatory)
Jorge Ibsen (Joint ALMA Observatory)

Time: Wed 14.30 - 14.45

Contact: *ignacio.toledo@alma.cl*

In the last few years Data science has emerged as a discipline of its own to address problems where data is usually heterogeneous, complex and abundant. In a nutshell, data science allows to provide answers to situations where a hypothesis can be formulated and later can be either confirmed or rejected following standard scientific methodology using data as raw material. Data science has been called differently depending of the domain (business intelligence, operational management, astroinformatics) and it has been recently in the center of a hype related to artificial intelligence and machine learning. It has been quickly adopted by the digital industry as the tool to distill information of massive operational data sets.

Among the many tools data science requires (mathematics, statistics, domain knowledge of the data sets,), IT infrastructure and software is by far the most visible and there is at present a whole ecosystem available as open source projects. The downside of this is data science is commonly confused with IT and software development, which creates conflicts between engineering- and scientific- mindsets, and leads to wrongly applying software development methodologies to it neglecting the experimental nature of the problem. In summary, creating the data lab becomes more important than answering questions with it.

In the domain of ALMA operations, there are many instances that can be identified and described as data science cases or projects ranging from monitoring array elements to understand performances and predict faults for engineering operations to routine monitoring of calibrators for science operations purposes. We have identified already around 30 different initial questions (or data science cases) and found that several of them have been addressed through individual efforts.

In parallel, several enabling platforms or frameworks have appear in the ecosystem that provides data scientists with both the laboratory equipment to conduct their experiments as well as enabling tools for collaboration, versioning control, and deploying results in production with a quick turnaround.

This talk aims to summarize the results of our exploration to apply data science workflows to resolve ALMA operations issues, identify suitable platforms that are already in use by the industry, share our experience in addressing specific ALMA operations data cases, and discuss the technical and sociological challenges we encountered along the way.

O4.5: Towards new solutions for scientific computing: the case of Julia

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):

Maurizio Tomasi (Universit degli Studi di Milano)
Mos Giordano (Universit del Salento)

Time: Thu 12.00 - 12.15

Contact: *maurizio.tomasi@unimi.it*

This year marks the consolidation of Julia (https://julialang.org/), a programming language designed for scientific computing, as the last version before 1.0 has just been released (0.7). Among its main features, expressiveness and high execution speeds are the most prominent: the performance of Julia code is similar to statically compiled languages, yet Julia provides a nice interactive shell and fully supports Jupyter; moreover, it can transparently call external codes written in C, Fortran, and even Python without the need of wrappers. The usage of Julia in the astronomical community is growing, and a GitHub organization named JuliaAstro takes care of coordinating the development of packages. In this talk, we will provide an overview of Julia and JuliaAstro. We will also provide a real-life example by discussing the implementation of a Julia-only simulation pipeline for a large-scale CMB experiment.

P4.18: An on-site data reduction pipeline for the Goodman Spectrograph

Theme: Data Science: Workflows Hardware Software Humanware

Author(s):

Simon Torres-Robledo (SOAR Telescope)
Cesar Brinceno (SOAR Telescope)

Contact: *storres@ctio.noao.edu*

The Goodman Spectroscopic Pipeline (GSP) is reaching some maturity and behaving in a stable manner. Though its development continues, we have started a parallel effort to develop a real-time version of the GSP, which aims at obtaining fully reduced data within seconds after the spectrum has been obtained at the telescope. Most of the required structure, algorithms and processes already exist with the offline version of the GSP. The real-time or online version differs in its requirements for flow control, calibration files, image combination, reprocessing, observing logging assistance, etc. Here we present results obtained with the offline version of GSP with various instrument setups, and outline the route for implementation of the real time, online version.

P12.15: Development of auto-multithresh: an automated masking algorithm for deconvolution in CASA

Theme: Algorithms
Author(s):
Takahiro Tsutsumi (NRAO)
Amanda Kepley (NRAO)
IIsang Yoon (NRAO)
Urvashi Rau (NRAO)

Contact: *ttsutsum@nrao.edu*

A general purpose automated masking algorithm for deconvolution was developed in order to support automated data processing in ever-increasing data volumes of the current and future radio interferometers as described by Kepley et al. in this meeting. In this presentation, we describe some technical details of the implementation of the automated masking algorithm named auto-multithresh which was integrated into the refactored imaging task (tclean) in CASA. We also discuss our approach that we took for the development, which loosely follows the iterative model, so that the implementation is refined progressively for its functionality and performance based on testing and updated requirements throughout prototyping in Python to the final production in C++.

F5: Workflows Management using Pegasus

Theme: Data Science: Workflows Hardware Software Humanware
Author(s):
Karan Vahi (USC Information Sciences Institute)
Mats Rynge (USC Information Sciences Institute)

Time: Wed 10.45 - 11.15

Contact: *vahi@isi.edu*

Workflows are a key technology for enabling complex scientific applications. They capture the interdependencies between processing steps in data analysis and simulation pipelines, as well as the mechanisms to execute those steps reliably and efficiently in a distributed computing environment. They also enable scientists to capture complex processes to promote method sharing and reuse and provide provenance information necessary for the verification of scientific results and scientific reproducibility. Application containers such as Docker and Singularity are increasingly becoming a preferred way for bundling user application code with complex dependencies, to be used during workflow execution. The use of application containers ensures the user scientific code is executed in a homogenous environment tailored for application, even when executing on nodes with widely varying architecture, operation systems and system libraries. This demo will focus on how to model scientific analysis as a workflow and execute them on distributed resources using the Pegasus Workflow Management System (http://pegasus.isi.edu).

Pegasus is being used in a number of scientific domains doing production grade science. In 2016 the LIGO gravitational wave experiment used Pegasus to analyze instrumental data and confirm the first ever detection of a gravitational wave. The Southern California Earthquake Center (SCEC) based at USC, uses a Pegasus managed workflow infrastructure called Cybershake to generate hazard maps for the Southern California region. In March 2017, SCEC conducted a CyberShake study on DOE systems ORNL Titan and NCSA BlueWaters to generate the latest maps for the Southern California region. Overall, the study required 450,000 node-hours of computation across the two systems. Pegasus is also being used in astronomy, bioinformatics, civil engineering, climate modeling, earthquake science, molecular dynamics and other complex analyses.

Pegasus allows users to design workflows at a high-level of abstraction, that is independent of the resources available to execute them and the location of data and executables. It compiles these abstract workflows to executable workflows that can be deployed onto distributed and high-performance computing resources such as DOE LCFs like NERSC, XSEDE, local clusters, and clouds. During the compilation process, Pegasus WMS does data discovery, locating input data files and executables. Data transfer tasks are automatically added to the executable workflow. They are responsible for staging in the input files to the cluster, and for transferring the generated output files back to a user-specified location. In addition to the data transfers tasks, data cleanup (cleanup data that is no longer required) and data registration tasks (catalog the output files) are be added to the pipeline. For managing users data, Pegasus interfaces with a wide variety of backend storage systems (with different protocols). It also has variety of reliability mechanisms in-built ranging from automatic job retries, workflow-checkpointing to data reuse. Pegasus also performs performance optimization as needed. Pegasus provides both a suite of command line tools and a web-based dashboard for users to monitor and debug their computations. Over the years, Pegasus has also been integrated into higher level domain specific and workflow composition tools such as Portals, HUBzero and Wings. We also recently have added support for Jupyter notebooks, that allows users to compose and monitor workflows in a Jupyter notebook.

P3.9: VisIVO Visual Analytics Tool an EOSC Science Demonstrator for data discovery

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Fabio Vitello (INAF-Astrophysical Observatory of Catania)
Eva Sciacca (INAF-Astrophysical Observatory of Catania)
Ugo Becciani (INAF-Astrophysical Observatory of Catania)
Alessandro Costa (INAF-Astrophysical Observatory of Catania)
Simone Riggi (INAF-Astrophysical Observatory of Catania)
Sergio Molinari, INAF-Institute for Astrophysics and Space Planetology

Contact: *fabio.vitello@inaf.it*

VisIVO is an integrated suite of tools and services for data discovery that include collaborative portals, mobile applications, visual analytics tool and a number of key components such as workflow applications, analysis and data mining functionalities.

Space missions and ground-based facilities produce massive volumes of data and the ability to collect and store them is increasing at a higher pace than the ability to analyze them. This gap leads to new challenges in the analysis pipeline to discover information contained in the data.

VisIVO Visual analytics tool for star formation regions focuses on handling these massive and heterogeneous volumes of information accessing the data previously processed by data mining algorithms and advanced analysis techniques with highly interactive visual interfaces offering scientists the opportunity for in-depth understanding of massive, noisy, and high-dimensional data.

The aforementioned challenges demands an increasing archiving and computing resources as well as a federated and interoperable virtual environment enabling collaboration and re-use of data and knowledge. Thus, the connection with the European Open Science Cloud is being investigated exploiting the EGI services such as the Check-in (for federated authentication and authorization) and the Federated Cloud (for analysis and archiving services).

Recently the VisIVO development has been exploited for the experimentation of cutting-edge interactive visualization technologies for the improvement of teaching and scientific dissemination. This work is being carried out as a knowledge transfer from astrophysical sciences to geological sciences in the context of an international collaboration to innovate teaching, learning and dissemination of earth sciences, using virtual reality.

P1.17: Analysis of Stellar Spectra from LAMOST DR5 with Generative Spectrum Networks

Theme: Machine Learning in Astronomy

Author(s):
Rui Wang (National Astronomical Observatories,Chinese Academy of Sciences, Beijing, China)
Luo, Ali (National Astronomical Observatories,Chinese Academy of Sciences, Beijing,China)

Contact: *wangrui@nao.cas.cn*

We derived the fundamental stellar atmospheric parameters (Teff, log g, [Fe/H] and [/Fe]) of low-resolution spectroscopy from LAMOST DR5 with Generative Spectrum Networks(GSN), which follows the same scheme as a normal ANN with stellar parameters as the inputs and spectrum as outputs. After training on PHOENIX theoretical spectra, the GSN model performed effectively on producing synthetic spectra. Combining with Bayes framework, application in analysis of LAMOST observed spectra become efficient on the Spark platform. Also, we examined and validated the results by comparing with reference parameters of high-resolution surveys and asteroseismic results. Our method is credible with a precision of 130K for Teff, 0.15 dex for log g, 0.13 dex for [Fe/H] and 0.10 dex for [/Fe].

P10.16: The Murchison Widefield Array's VO compliant archive

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Andreas Wicenec (International Centre for Radio Astronomy Research, University of Western Australia)
Dave Pallot (International Centre for Radio Astronomy Research, University of Western Australia)
Greg Sleap (Murchison Widefield Array, Curtin Institute of Radio Astronomy)
Christopher Jordan (Murchison Widefield Array, Curtin Institute of Radio Astronomy)
Marcin Sokolowski (Murchison Widefield Array, Curtin Institute of Radio Astronomy)
Hasan Rayan, Cipher Tel Ltd Pty

Contact: *andreas.wicenec@uwa.edu.au*

The Murchison Widefield Array is a SKA pre-cursor operated on the site of the planned SKA1-LOW. It has been operational since more than 4 years and during that time it has collected around 25 PB of data. During 2018 we have been working on a separately funded project to implement a completely new access system to the MWA archive, which also includes an on-the-fly calibration pipeline and is based on IVOA standards. As part of the implementation we have also implemented a new VOSpace server system in Python, which has been released as an open source contribution to the community. In this paper we are presenting an overview of the new MWA archive system, including the web interface, authentication/authorisation, data processing, staging and delivery.

I2.1: Establishing the SKA Regional Centre Network: Mesh Management and Culture Change

Theme: Management of Large Science Project
Author(s):
Michael W. Wise (ASTRON Netherlands Institute for Radio Astronomy)

Time: Thu 11.15 - 11.45

Contact: *wise@astron.nl*

The Square Kilometre Array (SKA) is an ambitious project to construct the worlds most powerful radio telescope and enable transformational scientific discoveries across a wide range of topics in physics and astronomy. With two telescopes sites located in the deserts of South African and West Australia, an operational headquarters based in the UK, and 12 different member countries contributing to the design and construction, the SKA is truly a global endeavor. Once operational, the SKA is expected to produce an archive of science data products with an impressive growth rate on the order of 700 petabytes per year. Hosting the resulting SKA archive and subsequent science extraction by users will require a global research infrastructure providing additional capacity in networking, storage, computing, and support.

This research infrastructure is currently foreseen to take the form of a federated, global network of SKA Regional Centres (SRCs). These SRCs will be the primary interface for researchers in extracting scientific results from SKA data and, as such, are essential to the ultimate success of the telescope. The unprecedented scale of the expected SKA data stream, however, requires a fundamental change in the way radio astronomers approach extracting their science. Efforts are already underway in various countries around the world to define and deploy the seeds of what will grow into a community-provided research infrastructure that can deliver SKA science. In this talk, I will give an update on these initial efforts as well as the various technological, management, and sociological challenges associated with establishing the SKA Regional Centre network.

P13.22: Software and data policies for Nature journals

Theme: Other
Author(s):
Paul Woods (Nature Research)

Contact: *paul.woods@nature.com*

I briefly summarise the current code and data availability policies that apply to Nature, Nature Astronomy, Nature Geoscience and other Nature journals. Full policies can be found at: https://www.nature.com/authors/policies/availability.html

P1.18: Research on Automatic Recognition of Radio Frequency Interference Based on Deep Learning

Theme: Machine Learning in Astronomy

Author(s):
Jian Xiao (Tianjin University)
Zhicheng Yang (Tianjin University)
Ce Yu (Tianjin University)
Bo Zhang (National Astronomical Observatories, Chinese Academy of Sciences)
Min Long (College of Engineering, Boise State University)

Contact: *xiaojian@tju.edu.cn*

Radio frequency interference (RFI) mitigation is a key phase in data pipeline of radio telescopes. Classical RFI mitigation methods depend on the RFI physical characteristics, often fail to recognize some complicated patterns or result in misrecognition. We developed a novel approach of RFI recognition and automatic flagging with an improved convolution neural network. The network was constructed based on U-net with much deeper network structure for more complicated patterns and more components to reduce recognition error caused by over-fitting. Compared to the most widely used classical method – SumThreshold, our method has better performance on both precision and recall rate, it also outperformed the U-net, a traditional deep learning model, and KNN,a typical machine learning model.

P12.16: An algorithm of selection of meteor candidates in GWAC system

Theme: Algorithms
Author(s):

YangXu (National Astronomical Observatories, Chinese Academy of Sciences)

Contact: *yxu@nao.cas.cn*

With its large field of view, GWAC can record hundreds of meteors every day. These meteors disturb the detection process of optical transients in the GWAC system, and may be a valuable treasure for some meteor research groups. It is therefore very important to accurately find all of these meteors. To address the challenge of precisely distinguishing meteors from other elongated objects in a GWAC-like sky survey system, we design and implement a meteor candidate recognition algorithm, including the recognizing and morphology analysis of the light curves of the meteor candidates. Although the algorithm may filter out some real meteors, it can provide a sample of meteor with high confidence. After processing the images of Mini-GWAC taken in two months, we detect 109,000 elongated objects in which more than 90 percent of objects are not meteor. Among the elongated objects, about 5.9% objects are identified as meteors with high confidence, after the filters based upon an existence in a single frame, a single peak in the light curves, and a slow variation of the light curves.

P13.23: Super-resolution Imaging of the Protoplanetary Disk HD 142527 using Sparse Modeling

Theme: Other

Author(s):
Masayuki Yamaguchi (The University of Tokyo, National Astronomical Observatory of Japan)
Kazunori Akiyama (National Radio Astronomy Observatory, MIT Haystack Observatory)
Akimasa Kataoka (National Astronomical Observatory of Japan)
Misato Fukagawa (National Astronomical Observatory of Japan)
Takashi Tsukagoshi (National Astronomical Observatory of Japan)
Takayuki Muto at Kogakuin University, Shiro Ikeda at the Institute of Statistical Mathematics, Hiroshi Nagai at NAOJ, Mareki Honma at NAOJ, Ryohei Kawabe at NAOJ and the University of Tokyo

Contact: *masayuki.yamaguchi@nao.ac.jp*

High-resolution observations of protoplanetary disks with radio interferometers are crucial for understanding the planet formation process. Recent observations using Atacama Large Millimeter/submillimeter Array (ALMA) have revealed various small-scale structures in disks. In interferometric observations, the observed data are an incomplete set of Fourier components of the radio source image. The image reconstruction is therefore essential in obtaining the images in real space. The CLEAN technique has been widely used, but recently, a new technique using the sparse modeling approach is suggested. This technique directly solves a set of undetermined equations and has been shown to behave better than the CLEAN technique based on mock observations with VLBI (Very Long Baseline Interferometry). However, it has never been applied to ALMA-like connected interferometers nor real observational data. In this work, for the first time, the sparse modeling technique is applied to observational data sets taken by ALMA. We evaluate the performance of the technique by comparing the resulting images with those derived by the CLEAN technique. We use two sets of ALMA archival data at Band 7 (350GHz) for the protoplanetary disk around HD 142527. One is taken in the intermediate-baseline array configuration, and the other is in the longer-baseline array configuration. The image resolutions reconstructed from these data sets are different by a factor of 3. We compare images reconstructed using sparse modeling and CLEAN. We find that the sparse modeling technique can successfully reconstruct the overall disk emission. The previously known disk structures appear on both images made by the sparse modeling and CLEAN at its nominal resolutions. Remarkably, the image reconstructed from intermediate-baseline data using the sparse-modeling technique matches very well with that obtained from longer-baseline data using the CLEAN technique with the accuracy of 90 % on the image domain.

P9.10: aLigo/Virgo Gravitational Wave Electromagnetic Counterparts Host Galaxy Candidates Filtering Pipeline

Theme: Multi-Messenger Astronomy
Author(s):
Xu Yunfei (NAOC)
Fan Dongwei (NAOC)
Xu Dong (NAOC)
Cui Chenzhou (NAOC)

Contact: *xuyf@nao.cas.cn*

An ongoing effort of the emergent field of multi-messenger astronomy is to discover electromagnetic (EM) counterparts of transient gravitational wave (GW) with EM follow-up observation. After each detection of the GW event, aLIGO -Virgo will distribute the GW source sky confidence distribution data to collaborators, with which astronomers can carry out follow-up observations at different wavelengths. However, based on the current observation accuracy, the 90% credible localization area of GW sources cover hundreds of square degrees. This makes the identification of EM counterpart challenging even with wide-field telescopes. Due to the expected gravitational wave event source reside in or near a galaxy, so restricting follow-up observations to galaxies can significantly decrease the required number of pointing, the total integration time and the number of false positives found. Based on this idea, we set up a pipeline to query the GLADE catalogue based on the position error range and distance range of GW events published by aLIGO / Virgo, and sort the results according to their confidence of the locations, b band brightness, (metering) redshift to quickly screen out the galactic host candidates. The system realizes a series of automatic operations, such as analyzing GW alarm, extracting GW position data, querying catalog, filtering and sorting of host galaxy candidates, and notifying, etc. The whole process does not need manual intervention. By building the index of the GLADE catalog based on the HEALPix celestial sphere partitioning method, combining with the MOC query method based on the GW position data, the fast search of the host galaxy candidates was realized, one query and sort tasks time consumption is less than 10s. The overall pipeline is highly efficient and can perform a series of operations for data visualization, host galaxy candidates filtering, and observation list report within 5 minutes after receiving a GW event alert.

O3.1: An introduction to FITSWebQL

Theme: Astrophysical Data Visualization from Line Plots to Augmented and Virtual Reality

Author(s):
Christopher Zapart (NAOJ)
Y. Shirasaki (NAOJ)
M. Ohishi (NAOJ)
Y. Mizumoto (NAOJ)
W. Kawasaki (NAOJ)
T. Kobayashi (NAOJ), G. Kosugi (NAOJ), E. Morita (NAOJ), A. Yoshino (NAOJ), S. Eguchi (Fukuoka Univ.)

Time: Mon 09.30 - 09.45

Contact: *chris.zapart@nao.ac.jp*

The JVO ALMA WebQL web service - available through the JVO ALMA FITS archive - has been upgraded to include legacy data from other telescopes, for example Nobeyama NRO45M in Japan. The updated server software has been renamed FITSWebQL. In addition, a standalone desktop version supporting Linux, macOS and Windows 10 Linux Subsystem (Bash on Windows) is also available for download from http://jvo.nao.ac.jp/ chris/ .

The FITSWebQL server enables viewing of even 100GB-large FITS files in a web browser running on a PC with a limited amount of RAM. Users can interactively zoom-in to selected areas of interest with the corresponding frequency spectrum being calculated on the server in near real-time. The client (a browser) is a JavaScript application built on AJAX, WebSockets, HTML5, WebGL and SVG.

There are many challenges when providing a web browser-based real-time FITS data cube preview service over high-latency low-bandwidth network connections. The upgraded version tries to overcome the latency issue by predicting user mouse movements with a Kalman Filter in order to speculatively deliver the real-time spectrum data at a point where the user is likely to be looking at. The new version also allows one to view multiple FITS files simultaneously in an RGB composite mode (NRO45M FUGIN only), where each dataset is assigned one RGB channel to form a colour image. Spectra from multiple FITS cubes are shown together too.

The talk gives a brief tour of the FITSWebQL main features. We also touch on some of the recent developments, such as an experimental switch from C/C++ to Rust (see https://www.rust-lang.org/) for improved stability, better memory management and fearless concurrency, or attempts to display FITS data cubes in the form of interactive on-demand video streams in a web browser.

O10.2: AXS: Making end-user petascale analyses possible, scalable, and usable

Theme: Databases and Archives: Challenges and Solutions in the Big Data Era

Author(s):
Petar Zecevic (University of Washington DIRAC Institute, University of Zagreb)
Colin T. Slater (University of Washington)
Mario Juric (University of Washington)

Time: Tue 09.45 - 10.00

Contact: *Petar.Zecevic@fer.hr*

We introduce AXS (Astronomy eXtensions for Spark), a scalable open-source astronomical data analysis framework built on Apache Spark, a state-of-the-art industry-standard engine for big data processing.

In the age when the most challenging questions of the day demand repeated, complex processing of large information-rich tabular datasets, scalable and stable tools that are easy to use by domain practitioners are crucial. Building on capabilities present in Spark, AXS enables querying and analyzing almost arbitrarily large astronomical catalogs using familiar Python/AstroPy concepts, DataFrame APIs, and SQL statements. AXS supports complex analysis workflows with astronomy-specific operations such as spatial selection or on-line cross-matching. Special attention has been given to usability, from conda packaging to enabling ready-to-use cloud deployments.

AXS is regularly used within the University of Washington's DIRAC Institute, enabling the analysis of ZTF (Zwicky Transient Facility) and other datasets. As an example, AXS is able to cross-match Gaia DR2 (1.8 billion rows) and SDSS (800 million rows) in 2 minutes, with the data of interest (photometry) being passed to Python routines for further processing. Here, we will present current AXS capabilities, give an overview of future plans, and discuss some implications to analysis of LSST and similarly sized datasets. The long-term goal of AXS is to enable petascale catalog and stream analyses by individual researchers and groups.

P1.19: Machine Learning for Quasar Candidate Selection

Theme: Machine Learning in Astronomy

Author(s):
Zhang Yanxia (National Astronomical Observatories,CAS)
Jin Xin (National Astronomical Observatories,CAS)
Zhang Jingyi (National Astronomical Observatories,CAS)
Zhao Yongheng (National Astronomical Observatories,CAS)

Contact: *zyx@bao.ac.cn*

With the big data era of astronomy coming, machine learning becomes more popular in various astronomical aspects, for example, classification of celestial objects, physical parameter measurement, rare object detection and so on. Quasar candidate selection belongs to the classification problem. Based on a number of present photometric and spectroscopic sky survey projects, pre-selecting quasar candidates from galaxies and stars is a hot issue. It is easy to separate quasars and stars from galaxies by morphology while it is difficult to distinguish quasars from stars due to their similar morphology. Nevertheless, researchers have figured out various approaches on quasar candidate selection. We will introduce them in detail.