

DS6: Data Exploration

Chair: Bob Mann (*DS6 Lead*)
University of Edinburgh

DS6 Review Session Agenda

Introduction and Overview – *Bob Mann* (15 min)

Demos and Highlights (60 min)

PaperScope – *Mark Holliman*

CAMEA – *Francois Bonnarel*

Aladin with PLASTIC – *Thomas Boch*

AstroWeka and PLASTIC – *Brian Walshe*

VisIVO – *Ugo Becciani*

VO-Neural – *Peppe Longo*

Scalability experiments – *Mark Taylor*

Summary, Future Plans & Questions – *Bob Mann* (5 min)

DS6 in the Project Plan

Assessment of “a range of data mining and visualization algorithms & packages, with a view to determining

how they can be run as distributed services,

how they can be made VObs-compliant and

how they can be extended to extremely large datasets.”

DS6 in the Project Plan

Assessment of "a range of data mining and visualization algorithms & packages, with a view to determining

CEA-wrapped FastKDA, VOTECHBroker

how they can be run as distributed services,


Aladin, TOPCAT, VisIVO, Weka, Astroneural - PLASTIC


how they can be made VObs-compliant and

Eirik, column-ordering, file mapping, k-d trees

how they can be extended to extremely large datasets."

DS6 Deliverables

- Data Exploration Study Report
 - Originally due Dec 2006
 - Revised plan
 - Preliminary Report - Jan 2007 
 - Final Report - during 2008

- DS6 Software Releases
 - 1st: Mar 2007 – included in baseline release 
 - 2nd: Dec 2007 – rescheduled to Mar 2008
 - 3rd: Sep 2008

Data Exploration *Preliminary* Study Report

wiki.eurovotech.org/twiki/bin/view/VOTech/DS6StudyReport

- Very substantial piece of work - 85 pages
- Progress to date and plans for rest of DS6
- Some sections need revision for final report



The VOTech Project

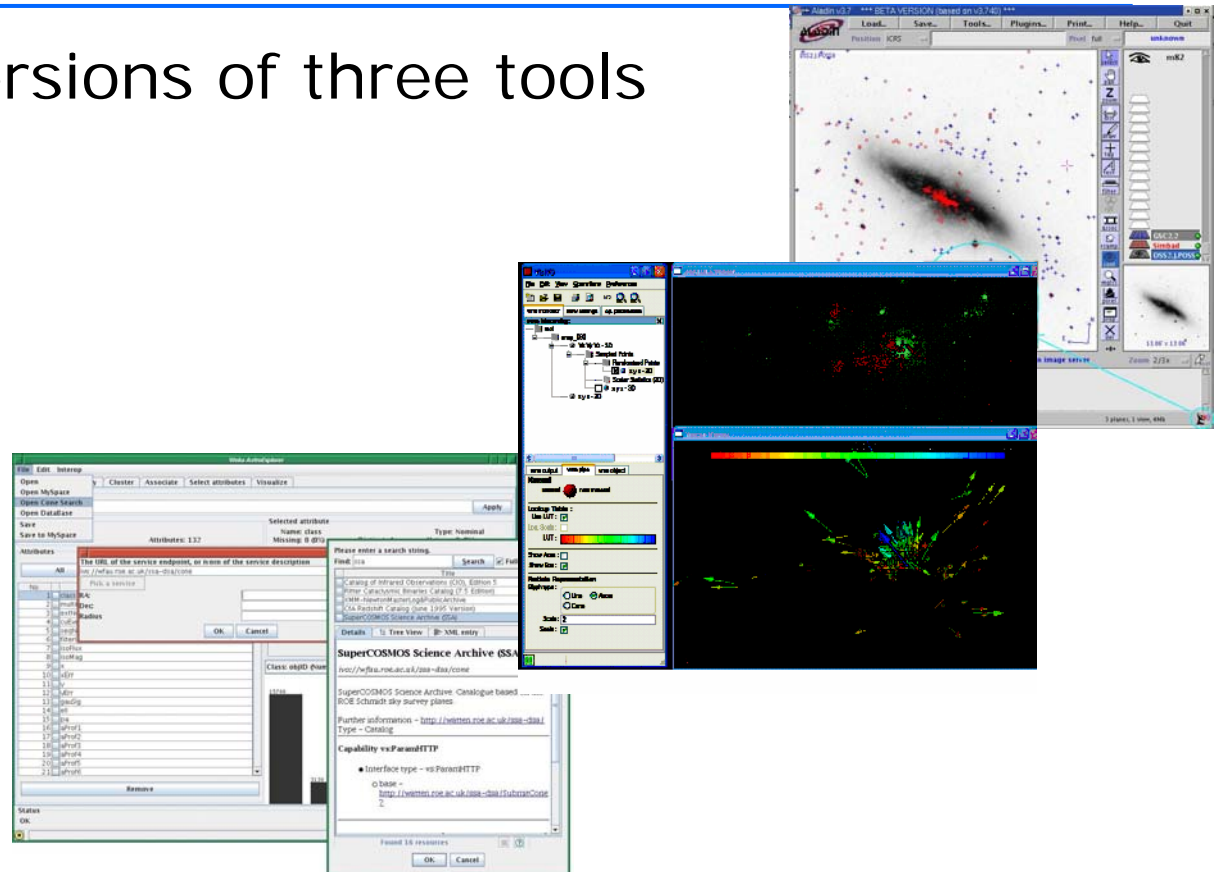
Design Study 6: Data Exploration Preliminary Study Report

January 2007



Software Release 1

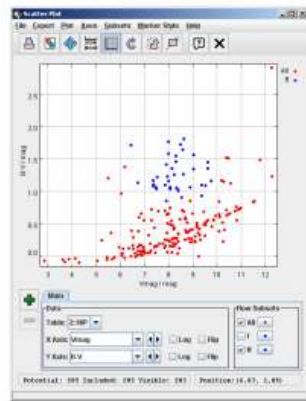
- VO-enabled versions of three tools
- Aladin v4
- VisIVO v1.1
- AstroWeka
- All interoperable through **PLASTIC**



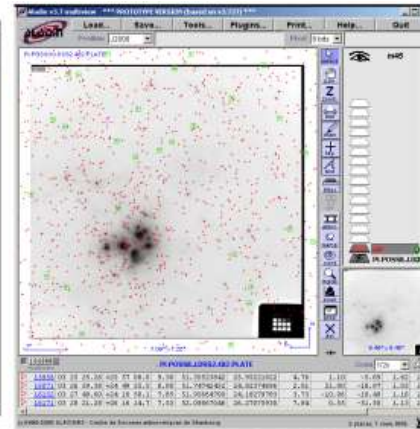
PLASTIC

- No tool can do offer all functionality needed
 - “Interoperability, not integration”
- PLASTIC: inter-app messaging protocol
 - Tools collaborate in exploring dataset and astronomers get the sum of their functionality

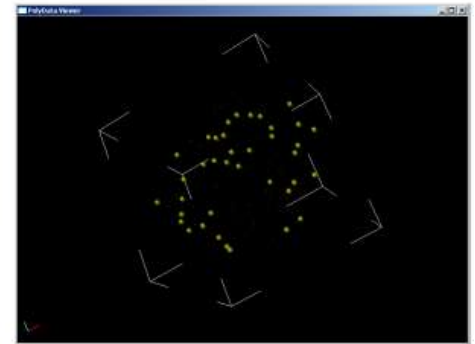
Linked views
via PLASTIC



TOPCAT



Aladin



VisIVO

Demos and Highlights

- Information Visualisation
 - PaperScope, CAMEA
- VO-enabling existing tools interoperably
 - Data mining: AstroWeka, VO-Neural
 - Visualisation: Aladin, VisIVO
- Scalability experiments
 - Column-ordering, file mapping

Summary

- Informal
 - Large quantity of high-quality work completed
 - Collaboration worked well – e.g. PLASTIC
- Formal
 - Study Report:
 - reflects, and reports on, significant progress towards understanding Euro-VO data exploration requirements
 - First Software Release: March 2007
 - VisIVO v1.1, Aladin v4, AstroWeka
 - PLASTIC hubs in Aladin & AstroGrid Workbench

Future Plans

- Turn PLASTIC into IVOA standard
- Second software release: March 2008
 - PaperScope, VO-neural_MLP, CAMEA, VisIVO v1.3
 - PLASTIC hub in AstroGrid VOExplorer
- Further development of specific tools
 - VisIVO, Aladin, VO-neural

Put it all together & see what doesn't work

- Take realistic science case through to publication

Photometric quasar selection testbed

- Richards et al (2004)
 - Look at SDSS DR1 sources in 4D colour space
 - Kernel Density Estimation, based on k-d trees
 - Bayesian classifier for stars/quasars
 - Using photometric data, but trained with spectroscopy
 - >95% reliable – as good as spectroscopy
- Re-do Richards et al, SDSS DR5 + UKIDSS near-IR, and all done within the VO
- *Can we do publishable-quality research in the VO yet?...what is needed to do that?*