Minutes of the Meeting “Joint AIDA Internal Science Team - Developers meeting” on Nov. 18, ESAC

Participants:

Mark Allen, Christophe Arviset, Deborah Baines, Isa Barbarisi, Andrey Belikov, Francois Bonnarel, Marcus Demleitner, Juan Gonzalez, Raul Gutierrez, Evanthia Hatziminaoglou, Andrea Laruelo, Keith Noddle, Pedro Osuna, Paolo Padovani, Carlos Rodrigo, Jesus Salgado, Gert Sikkema, Enrique Solano, Mark Taylor, Gijs Verdoes Kleijn

Sebastian Derriere on teleconference during the photometry discussion

General:

* Undo (or undo last)
* Record actions(1)
* Save actions

(1) Though it might be difficult to implement, it can be extremely useful unless we want the users to learn all the scripting "languages" for the individual tools. Aladin has most of the functionality already implemented: each click is translated to a "command" on the console, the user can easily copy-paste or customise later on. MT commented that from a TOPCAT point of view, this is quite complex.

There was general agreement that the ‘Undo’ option is the most required.

Queries:

* No tool currently offer a rectangle search and without area limitations (missing infrastructure?)
It was commented that this could be solved with footprints.

* Multi-object, multi-resource queries:
  - Only Aladin in scripting mode (+python scripting, e.g. AstroGrid) allows for this but the output is very difficult to combine in a single master table
  - TOPCAT latest release is a great step forward; even though it allows for a single resource at a time, by performing repeated queries and appending the results to the previous catalogue one ends up building the master table in a reasonably fast way

A discussion then followed: Handling different sources in the same table can be easy but processing it from the client part can be hugely complex. EH proposed to create a single master table for all the fields associated to one source when a multi-target option is implemented.
JS mentions that concatenating different fields from different resources for the same object in the same table doesn't make sense (for SSA and SIA searches). A better way to do it is to have one table per resource. It’s not recommended to have one single table with the same null value for the majority of the columns, and some fields have only meaning within its resources context.
This would also be very difficult to implement from a visualization point of view.

AB proposed having an extension to the VOTable Schema to allow multiple VOTables inside a VOTable.

- Queries such as "which of my objects has been observed in the X-rays and in the B-band" are not possible right now, unless if you select all X-ray catalogues and all B-band images/catalogues

A discussion followed on the lack of mandatory parameters necessary to get sources from a determinate region. Which VO services are providing data in x ray? Actually this question cannot be answered, because the necessary parameters are not compulsory.
If the registry had these parameters as mandatory, it would be possible to make queries such as "which of my objects has been observed in X-rays and in the B-band".
JS mentions a way to modify user queries with an automatic band to spectral range conversion (band resolver), to make the search of a certain band easier for the user.
This is an historical problem: the quality of data.
Historical Question: should we be policing the data or not?
CA mentions the DALValidater, which checks the compliance of services and, as a next step, the validation level flag (flag between 0 and 5) so that VO tools can choose to display only services with a certain grade of validation. It currently needs further agreement with people involved.
It was proposed that these issues should be raised in the relevant IVOA working groups.

Cross-correlations:

* We still do not have a tool to perform cross-correlation of very large catalogues (some people even ask for all sky; EH thinks we should not have such a tool and that such things should be seen as services. Both EH and JS suggest the actual match should be done once and for all and then people could go and retrieve whatever they need from the cross-matched catalogue.
* STILTS can not cope with tables larger than ~106 rows, not only for cross-correlation but also for e.g. format conversion. EH tried to convert a .VOT of half a million lines to TST and even with the -disk flag could not do it. MT suggests he could increase the size of the cross matching (>2 million or more) but he's not convinced that scientists really want to do it.
* NOTE: positional cross-matching is fine; anything else would be extremely dangerous, as it all depends on the SED of the objects, the characteristics of the instruments, the code, the physics, the science one wants to do.
ES mentions that currently if you want to do cross matching with Aladin, you have to do it in many steps. For him, the larger the region to cross match the better and suggests this is desirable from a scientific point of view.
The current limit is probably a couple of million, because of memory problems.
PO mentions that the cross match issue has been a problem since the beginning, if it is done locally it needs huge resources, one possible solution to this could be the GRID.
PO mentions that in the current VO scheme, this type of all-sky cross-matches are not easy to implement.
General issue raised: VO is not meant for this massive comparison, it's preferable to have the complete cross-matched catalogue and then perform the query.

Cutouts:

* Is there any VO tool that can do this?
- TOPCAT can create them on the fly (Activation action), ***EXTREMELY*** useful functionality but

no longer maintained (still, never crashed!) but the cutouts can not be saved (?) and are only provided for a very restricted set of images (which is fine, because TOPCAT is not an Image analysis tool!)

A discussion followed on scripting functionalities and legacy tools:
The importance of scripting: many scientists have many programs that they have written themselves, Are they going to import VO into their environment or are they going to use new VO tools?
ES mentions that VO tools don't have to provide support to legacy data, because legacy tools can easily adapt themselves to VO, just implementing the discovery capability and the ability to handle VOTables. ES mentions that legacy tools should be more exploited and reused; we shouln't reinvent the wheel.
PO mentions that VOSpec implements certain functionalities because they have been required by the community. PO agrees that Legacy applications should be adapted to VO, as nobody wants to spend resources in reinventing the wheel. However, it looks like legacy applications are _not_ adapting to VO standards (e.g., SAMP). This is something to be discussed in a wider context, and actions should be taken to approach applications developers to enroll the VO standards.

Registries:

* Different tools deal with the registry in distinct ways. e.g. Aladin seems to be largely hardwired and somewhat arbitrary. Who decides what goes into "Aladin Image Service" and how often is this being updated (if ever)? Same with "Mission" logs. When was the last time the list and the contents of the logs were updated? "All VO" only reaches a restricted number of resources; VOSED also accessed a very restricted number of catalogues with no description
* The same keyword query on any two registries will return different lists of resources; technically there is no way of returning them in the same order (which is fine) but returning ***different*** resources is confusing and gives an image of lack of consistency
* The EURO-VO registry seems to handle keyword search in a different way and overall, it seems to be the least flexible registry (e.g. the only one that is not possible to query from TOPCAT)
* Why is VODesktop incapable of retrieving the metadata for the vast majority of resources?

A discussion then followed on the issues raised above: Aladin suggested they would look into the issues raised above
Should different tools use different registries?
CA mentions that there should be only one registry (EuroVO), at least for European partners.
CA suggests EuroVO applications should use only this one. In this way all the applications will be coherent, retrieving the same resources. Also, once the validation is implemented, all applications will be have the resources passed though a curation level.
Some issues are reported with the EuroVO Registry. CA informs ESAVO will look into them and solve them all so European partners can make full use of the EuroVO registry.
It is impossible to assure the same level of quality with different registries in different applications.
CA suggests checking the problems with the registry and that the harvesting is working correctly
There is also an issue with duplicated services, and how one can delete a service from a registry.
Services should be marked as deleted, not deleted completely from a registry.
It was proposed that these issues should be raised in the relevant IVOA working group and meetings.

SSA:

* Whenever possible (i.e. if the information is included in the data), aperture size and quality should be queriable parameters. Aperture and Data.FLuxAxis.Quality are now optional parameters in the SSA protocol but some services (e.g. INES) do provide this information. Although the implementation of data quality flags is not a VO job, the VO should ensure that this information (whenever available) is propagated correctly.
--> Suggestion for VO-tools: implement these types of queries
--> Suggestion for data providers: provide information on aperture/quality if available.

PO mentions that this is a data provider issue. Can we help?
ES says sometimes the SVO contact the provider and give suggestions.
PO says that the ESAVO team have also investigated this and suggest they can send the results of the SSA servers investigation to the meeting participants.

* Some data from the ESO Science Archive Spectrum Service cannot be plotted using VOSpec (try, for instance RA:52.96238, DEC:-27.68781 SIZE:0.001) as they do not incorporate the Dataset.SpectralAxis and Dataset.FluxAxis information. "These parameters should be provided for native data to make it possible for a client to interpret such data in a basic fashion without having to understand the details of each project-specific native data format." (SSAP document).
ESAVO explained that the ESO service is supplying the same spectrum in 3 formats: CSV, fits and VOTable. VOSpec does plot spectra from the ESO service with the VOTable format and in fits format. It doe snot open them in CSV format. Probably queries within VOSpec would thus have to be done only with the relevant "format=fits" and "format=votable" or similar, avoiding results in CSV to appear in the tree/table results.

* Typically a SED is made up of different spectral pieces. Joining these pieces implies having flux and wavelength calibrated in an absolute way. However Fluxcalib and Wavecalib are "OPT" parameters in the SSA protocol so there is no way of knowing in advance which services should be queried to build a properly calibrated SED.
--> Suggestion for VO-tools: make Fluxcalib and Wavecalib queriable parameters in a clearer way (VOSpec and SPLAT actually allow these queries).
--> Suggestion for SSAP: Make these parameters mandatory (e.g. Fluxcalib, Wavecalib, Exposure times, Resolution).

A discussion on SSA followed:
ES: There is also an issue with Time Series spectra. It is not possible to distinguish time series spectra at the registry level, but should be possible by ResourceType.
ES also wants the tools to be able to recognize Timeseries.
JS mentions that TimeSeries shouldn't have been included in SSAP.
JS suggests ES to propose, as a recommendation, to reserve something specific for TimeSeries. For example the same solution with theoretical services (type:simulation).

Issue on quality information: Data.FluxAxis.Quality metadata should be used but its meaning has not been well defined yet.
In the VO tools it would be useful to use the resolution metadata information.

KN suggests subclassing the SSAP.
PO mentions that it is already being done for region searches (position is mandatory but it can be ignored).

Normalized spectra are not well managed in the protocol and users cannot trust it.
The ESAVO propose to look into adapting VOSpec so that users can be better aware of whether they are handling calibrated or non-calibrated data.

TOOLS:

Aladin:

* Some graphics issues:
  - Load: 'by l...', 'by t...', 'B...', 'All c...', 'Whole c...', coordinates
  - (on mac) in order to refresh the windows one has to click inside
* Catalogues; difficult to explain, try the following: a "random" catalogue, e.g. Hatziminaoglou+, J/A+A/384/81. This catalogue is in fact the concatenation of 3 different catalogues with different columns. In a previous version, Aladin would display the contents of the first sub-catalogue and then would stop and the first sub-catalogue would be the one to send to TOPCAT, too; in a following version Aladin would show all 3 catalogues concatenated and would still send the first catalogue to TOPCAT; now it only shows a limited number of columns (even though I select all columns and whole catalogue), common to all 3 sub-catalogues and does not send anything to TOPCAT!
* (Minor:) load another catalogue that covers a completely different area of the sky; it can only be seen if brought at the bottom of the stack. To see all catalogues, the one with the largest coverage has to be the one at the very bottom of the stack - quite confusing
* Astrometry problem with the images coming from the Aladin image server solved, nice!
* FoV: why is the list so limited?

STILTS:

* Still issues with large catalogues (see cross-correlation section)

TOPCAT

* The EURO-VO registry is the only one that can not be queried from TOPCAT (see also Registries section)
this is more of a wish-list rather than faulty functionalities
* Delete a newly created sample
* Delete a newly created column
* Delete a loaded table and reclaim the memory

VODesktop:

* It is a very "heavy" application (interface-wise) just to filter resources, and the interface is quite

complex for users who don't know about registries.
* What is the max allowed number of resources for the creation of a Smart list?
* Multi-object query, dodgy output: create a smart list with Service capability is cone, Description contains RS CVn (Nov 12, 20 resources); select all and fire a query around NGC1068 in a 0.1deg radius; Astroscope informs you that it only finds matches with the RASS BSC USNO A2 (10 matches); send the output .vot file to TOPCAT; it is a file containing all columns from the first resource in the list (that contains 0 matches) and no rows and the 10 matches from ROSAT are not there.
* The only resources that actually have available table metadata (when looked for from VODesktop) are VizieR tables; all other registries do not seem to be harvested properly
* Is ADQL query supposed to work? I tried it on the Hipparcos catalogue after G. Rixon's talk at the IVOA and the Execute button read "Unavailable". If the functionality is unavailable for this particular service [which seems to be the case here], it should not appear as an option; TAP Query is available but not functional, shouldn't it be hidden till it actually becomes functional?
* The task runner is extremely complex to use; there are known applications in there (e.g. SExtractor, HyperZ, Merlin Imager) and then some entries with e.g. ConeSearch capability (e.g. 2dF object catalogue); tasks run on files that have been saved in VOSpace or locally but it is not possible to run a task on a file that was created from with VODesktop (i.e. a retrieved image or catalogue); are there default configuration files or does the user have to provide them? Never really managed to run any of these tasks (and maybe it is too ambitious even having them there)

Comment: TAP Queries have not yet been released, it's in a beta version

VOSED:

* How is the list of SSAs selected? It doesn't match that of VOSpec, and it doesn't match the green VOSpec list, either; is it hardwired? If so, how often is it being updated? (again, see Registries section)
* Photometric data: again, how was this list selected? what are the "default" magnitudes used? 2MASS PSC is not adequate for nearby extragalactic objects (i.e. extended)
* (Single object search) "Fitting to theoretical data" is misleading. The theoretical data mentioned here are stellar models (which ones?) and therefore the tool is more limited what is suggested by e.g. the field "redshift" at the very top
* (Single object search) Display SED open a new VOSpec every time, in a non VO-compliant way; Multi-object search results in a tar file; it would be very useful to use one of the technologies proposed during the IVOA to link the browser and the application and by clicking on an object from the list, display the SED in VOSed
* The interface is minimalistic and the lack of documentation makes it difficult to use (also underlined text is misleading, as there are no links behind it)
* **VOSA** very complicated and it gets stuck, not much documentation.

The SVO commented that VOSED is not using SAMP, there is a way to do it but it's not very straight forward.

VOSpec:

* The table output of the query should 1) be the default output instead of the tree and should be able to be sent to e.g. TOPCAT as a VOTable
* The symbols are re-set to their default values (points - not lines, very user-unfriendly especially since one is usually dealing with spectra and not SEDs) every time a new selection is made; in general

"retrieve" and "view" can become confusing, as sometimes neither of the two seems to be doing what the user wants, which is to display a set of selected spectra
* I visualise a list of spectra, change the plotting symbols and then all but the recently changed one vanish; I then have to 1) change all the symbols one by one and then tick them all one by one and View
* Graphics with Mac: show disappearing text in query form; I have to kill my browser in order to recover the window!
* I select a line to fit, fit the gaussian and when it's done the view goes back to the full range
* Fitting stellar models does not show how many models have actually been selected by setting the min and max parameters; or how many are left
* Does Normalisation work on two spectra or only on one spectrum and a model? Can't normalise two spectra (try NGC1068, HST spectra)
* Never managed to send a spectrum from VOSpec to SPLAT
* I retrieve a list of spectra but none of them is ticked in the Graphic mode window; which one is supposed to be sent to e.g. SPLAT? ("SAMP->Send Spectrum To")
* Multi-object, multi-resource query

ESAVO explains how VOSpec works for some of the above issues, which are already solved. For the ones not yet solved, will look into them and solve them.

Overall actions

It is agreed to take note of the following actions:

Action on PP to bring these inputs to the CSP
Action on PO to bring these issues to the TCG
Action on partner leaders to estimate possibilities to implement discussed functionalities