A Visualization Portal for ALMA Data

Final Report

Erik Rosolowsky (University of British Columbia, Okanagan Campus)
A. Russ Taylor (University of Calgary)
Alyssa Goodman (Harvard University)

Overview

This report summarizes the progress of the ALMA Development Study "A Visualization Portal for ALMA Data." The visualization tool provides a high-performance browser-based visualization tool that is capable of exploring the large data cubes that are being generated through ALMA. Through the activities funded in this study, we have made significant progress on the visualization tool. We have achieved our study deliverables. Furthermore, through the study process, we have identified several components of the project that will be the focus of our directed effort as we proceed in visualization development. We anticipate submitting a proposal for a full development project through the ALMA development program.

Report Contents:

Grant Activities

This development study had three main activities on which we report here.

Development of the Visualization Tool-- The study supported the efforts of Pavol Federl, a visualization developer working at the University of Calgary. Pavol is the principal software developer working on the visualization tool. Over the course of the development study, Pavol has made several improvements to the visualization tool. These include:

- Refactoring the code from Flex/Flash underlying framework to one based on HTML5/JavaScript. This development goal was required since the use of Flash is on the decline and the new basis will enable the tool to be used on a broader platform of devices, including iOS devices. This version of the tool is still in testing and the Flex/Flash implementation is accessible as described below. Access to the HTML5/JavaScript version is available upon request.

- Coordinate grids for images and one-dimensional profiles -- This includes a robust World Coordinate System (WCS) implementation for both the spatial and velocity/frequency axes.

- Profile synchronizations -- The profiles are one dimensional slices through the data cube. These track the cursor as it moves through the visualization canvas.

- 1D/2D Gaussian fitting -- Our initial meeting with the Expert User Group emphasized that a critical function of the visualization tool is basic region statistics and Gaussian fitting in image planes or along frequency axes. The one dimensional Gaussian fitting implementation, in particular, was well received by the Expert User Group on implementation.

- Separate profiles for region statistics tool -- The tool allows region statistics to be computed along the frequency axes allowing, for example, averaging over a box or calculation of the median plane-wise in a box through a data cube.

- Polarization visualization tools -- As a demonstration of an alternative mode of visualization, we have implemented a visualization toolkit for a set of images with full Stokes imaging data (I,Q,U,V). This provides a linked view across several different data cubes and explores the correlations among data in the four different Stokes parameters.

Metadata Management and VO Integration -- This study supported the efforts of Ken Anderson, a software developer at UBC Okanagan. Ken developed two packages: `mddb` and `metaData`, both written in Python. These packages parse the metadata in ALMA and VLA images as well as a wide variety of other images and radio data cubes. The tools provide a path by which ALMA data can be directly published onto the Virtual Observatory.

Special care was invested to make these parsing packages fault-tolerant so that other user data can be combined and published to the VO. This tool could readily be used to assimilate user-contributed data products into the ALMA archive. The error checking includes several features which edits non-FITS standard compliant metadata into a standardized, compliant format. The package provides robust reporting of metadata

errors and warnings which can be used to by the user to ensure that the error checking used in the tool is reflecting user intent.

These packages are presently integrated into the CyberSKA portal and provide a means by which users can publish their (arbitrary) data directly to the VO using the CyberSKA archive as a hosting tool. As a demonstration, we have published several Science Verification data sets through the ingest pathway.

The development study also supported travel for Ken Anderson to travel to ADASS and meet with members of the JAO software development team. His work on these tools is described in the proceedings of ADASS XXII (Anderson, Rosolowsky & Dowler 2012).

Usage Assessment -- The final activity supported by the study was convening an Expert User Group whose role was to assess the portal framework in the context of what will be required to support the needs of the ALMA community. The group was constituted to bring together members of the community engaged with areas of expertise in one or more of (1) scientific visualization, (2) ALMA data, or (3) scientific software development. Group membership is described in the Table below.

Name	Affiliation	Area of Expertise
Peter Teuben	U. Maryland	Millimeter interferometry, development of tools including MIRIAD
Adam Leroy	NRAO / NAASC	ALMA user and NAASC contact
Chris Beaumont	Harvard	Development of next generation analysis tools with visualization components
Alyssa Goodman	Harvard	Scientific visualization expert and millimeter-wave astronomer with a specialization in star formation.
Alberto Conti	Space Telescope Science Institute	Scientific software developer and manager and responsibility for community engagement for JWST.
Mark Whitehead	NRAO	Radio astronomer and software developer at the Green Bank site.
Jaime Pineda	European Southern Obs.	ALMA user and European ARC member
David Wilner	Smithsonian Astr. Obs.	Millimeter-wave astronomer with a specialization in star formation.

The group had an initial meeting on August 8th and 9th, 2012 in Cambridge, MA where we developed a visualization framework to described visualization tools and then focus on what the user needs for ALMA visualization will be. Specific members of the expert

user group were used to provide feedback to the visualization developer. At the conclusion of the study, the EUG was asked to provide feedback on the visualization tool. The feedback exercise and results are attached to this report.

Assessment of Deliverable Outcomes

The completed deliverables that were achieved under this project are deployed on the CyberSKA web portal. Any functionality described can be explored using the review account (*username*: ALMAreviewer *password*: A+ac@m@). Note that this tool is the Flex/Flash tool at the writing of this report. The HTML5/Javascript implementation has not had the full set of features implemented, but we anticipate the new version will be deployed by Sept. 1, 2014.

Creation of ALMAportal.org

Upon consultation with the NAASC, we have replaced this deliverable with design of a Portal Integration Tool as described below. Creation of a portal or presence independent of the ALMA science portal was deemed unnecessary. Instead, we have developed a plan by which the portal can be integrated into the existing archive and tools (see below).

Distributed data management system

We have deployed a distributed data system with user and group access management, metadata extraction for Measurement Sets and FITS files, and basic search functionality and VO interoperability.   This system is implemented in the CyberSKA portal but can be deployed in the ALMA science portal as desired. The distributed data management tool is built around the Integrated Rule-Oriented Data System (IRODS).

The primary innovation from the study is the metadata extraction and parsing.   This deliverable has been achieved through the `mddb` and `metaData` packages implemented in Python. The basic architecture of this solution is outlined in Figure 1. We have implemented our solution in coordination with the Canadian Astronomical Data
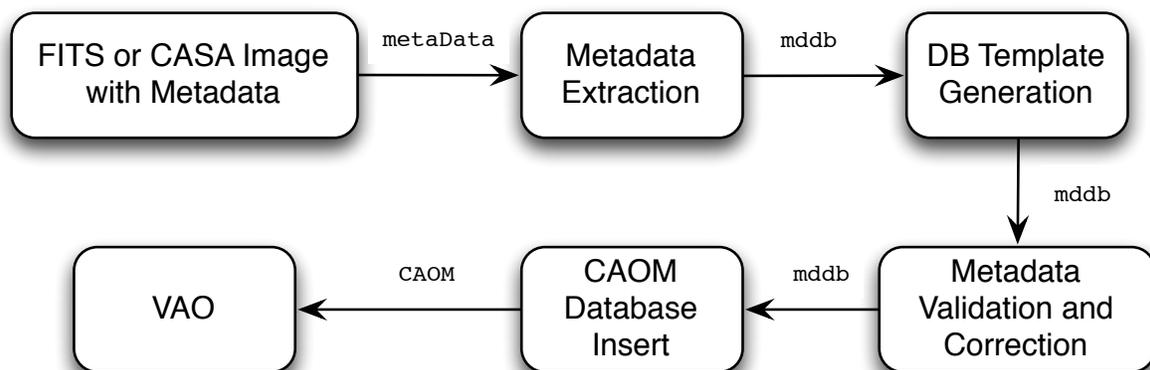

Figure 1: Architecture for the metadata parsing tool. File metadata is extracted from each file, translated into a form that can be inserted into CADC databases and then is available through VO tools.
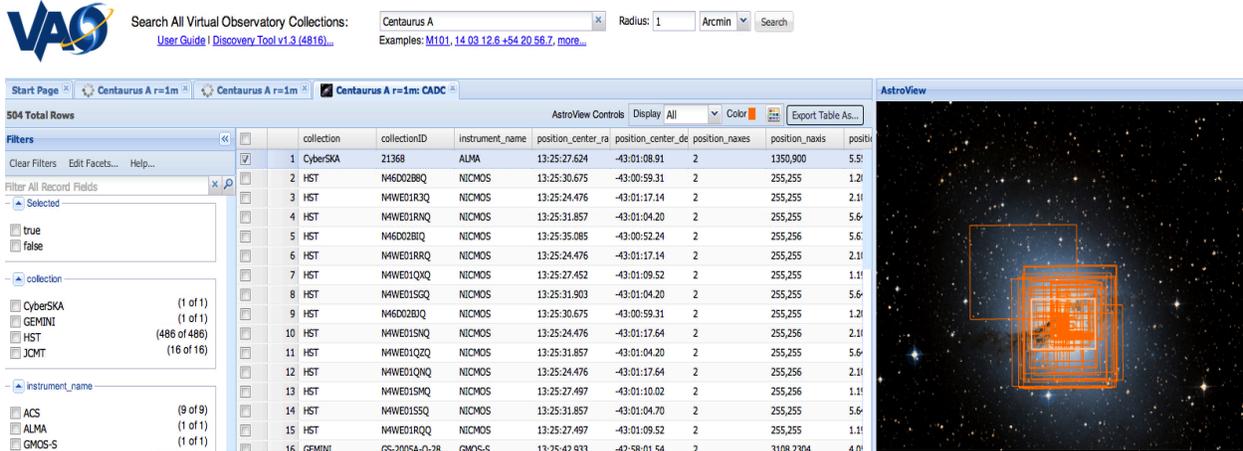
Figure 2: Screenshot of ALMA data discovered through the Virtual Astronomy Observatory (VAO) Data Discovery tool when data on the galaxy Centaurus A is sought. The user can click on the link and retrieve the data. ALMA data can now be published straight to the VO through these tools.

Centre. The package functions by parsing a data file using the metaData package which extracts the metadata into a flexible XML format. This format is then translated through one of several possible filters into a Common Archive Observation Model (CAOM) table entry which is th[...] table is then published to the V[...] hosted on the CyberSKA file a[...] published to the VAO. ALMA data are available in the CyberSKA collection accessed through CADCs Advanced query service[1].

Plane-based visualization tool

This visualization tool is deployed on the CyberSKA site. The visualization tool works on both CASA and FITS formats. Both formats are converted to an optimized storage format for high-performance visualization. Appendix A contains several screenshots of the visualization tool highlighting several of the features.

User Interface and Design of the full visualization tool

Over the course of the study, we have refined the user interface and implemented the basic framework that a final tool would utilize on the CyberSKA portal. The current user interface is designed around two main components: a visualization palette for two and three dimensional images and a user toolkit. Users can launch several windows including spectral and image profiles and Stokes parameters. Several suggestions on refining the user interface were implemented from the recommendation of the Expert User Group. Future priorities (see below) will be developed with interface elements that are evaluated by the Expert User Group.

---

[1] http://www3.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/AdvancedSearch/
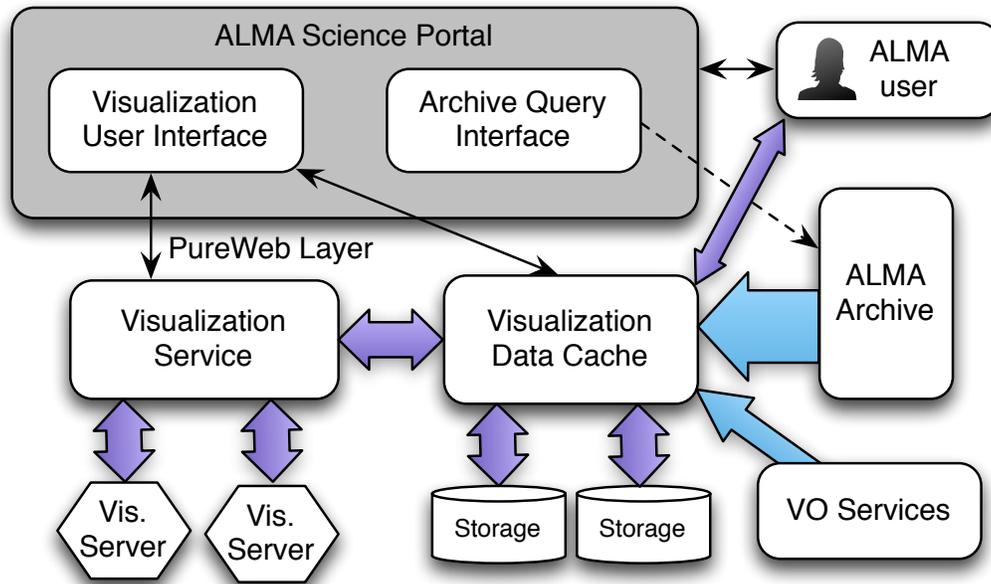
Figure 3: System interactions in the planned ALMA Visualization Tool-- This figure indicates the key components and basic interactions for the components of the visualization tool. Thin, black arrows indicate light-weight data flow (dashed: queries, solid: user interface actions) whereas thick, colored arrows indicate large volume data flow.

Architecture Design and Portal Integration Plan

We have developed a deployment plan for the visualization tool on consultation with members of the the Observatory's software and archive times. In particular, we have consulted with Jeff Kern on software (specifically CASA) integration, Felix Stoehr on Archive structure and Mark Lacy on Portal integration. We have developed the following plan based on their input. The rough outline of the visualization tool deployment is depicted in Figure 3. We are currently deploying the visualization tool into a different science portal (the Canadian Advanced Network for Astronomical Research) to prove our ability to deploy the tool into other frameworks.

A visualization portal will be developed and deployed in tandem with the archives at the the ALMA Regional Centers. The design philosophy of the tool is to place the visualization framework "near" the data and only transport visualization output (screen content) and user interface actions over the internet. Thus, the best results are obtained when the physical infrastructure of the tool is housed in ALMA regional center computation facilities.

The query tool will interact closely with the ALMA Archive Query interface. We propose inserting a "Visualize" option on each of the query results that a user can select instead of "Download." On selecting this option, the data are staged into an intermediate storage array (Visualization Data Cache). In this staging process, the data files are reshaped into a binary format that can be directly memory mapped by the visualization tool for the best performance.

The user is also able to directly upload comparison data sets to their own user space in this data cache (for example an image for overlay). The files in the cache will be deleted soon after the visualization session ends. Given the short time scale for development projects, we do not anticipate that the development activities can be extended to creating a unified local viewer that can run on a desktop machine and interface with the remote visualizer. We anticipate this can be a fruitful long-term development goal. However, since a project must produce a working system at fixed cost, this initially appears beyond the scope of a two-year development project. We are meeting with the CASA team in July 2013 to plan the deployment of this tool.

Once a user triggers the visualization interface for one or more files in the ALMA archive, the data will move through ARC intranet to the Visualization Data Cache which manages storage for the files being actively visualized. Data can also be brought into the data cache through direct user upload or Virtual Observatory services as depicted above. These upload channels may need to be limited to maintain network performance (see Scalability Plan below).

The user interface between the user's browser and the visualization tool is mediated by the PureWeb technology. This is a proprietary tool that enables the high-performance visualization capabilities within the browser. This product designed and built by our industrial partner, Calgary Scientific Inc. (CSI). The PureWeb platform enables web-based access to remote server-side visualization tools from any platforms and was initially developed by CSI primarily to support medical imaging applications. An existing licensing agreement between CSI and the CyberSKA project covered this development study. CSI's policy is that the platform can be used for research purposes and would be able to provide licensing for the entire ALMA scientific community. However, the visualization tool is designed independently from the PureWeb layer. The tool can be decoupled and an open-source middleware layer can be introduced. We anticipate this would require 0.3 person-years.

A system of visualization servers performs the required calculations for the tool. The visualization tool requires 2 processor cores per user with enough storage to handle intermediate data products in the cache. Our model has assumed that power, battery backup and network connectivity would be provided by the existing computing facilities at the ARCs. No backup of the cache systems are needed as the storage intended to be short-term. We have designed a system unit to meet the needs for 12 simultaneous users and these units can be purchased as needed to support the use load. Quotes were obtained without Educational / Research / Bulk discounts or taxes applied.

Dell - $29,000 list	IBM - $29,100
R820 2U rackmount server 5 year next day warranty 4x Intel E5-4617 2.9 Ghz 6 core CPUs 32GB RAM Intel 10GBASE-T 10 Gb network adapter 16TB of disk	System x3750 M4 2U rack mount server 5 year next day warranty 4x Intel E5-4617 2.9 Ghz 6 core CPUs 32GB RAM Intel 10GBASE-T 10 Gb network adapter 16TB of disk

Scalability and Support Plan

We have evaluated the cost to implement the visualization tool at a production scale for the ALMA user base and the cost to maintain the portal should it be implemented by the observatory. We anticipate an ALMA user base of 4000 users based on science portal registration across all ARCs. Of these 4000 registrants, we anticipate 25% will be active in scientific projects using the observatory. Of these, we anticipate that up to 10% will need visualization at any given time when the observatory is in production. Thus, we have designed a system capable of handling ~100 simultaneous users. The visualization tool can be made accessible to the public as well for citizen science usage, but this will increase the anticipated user base and scale the cost of hardware proportionally.

We evaluated three main considerations for the long term deployment of the visualization platform: software maintenance, hardware replacement, network bandwidth. There will additional marginal costs of power and climate control for the data centre, but these are likely to be significantly smaller than costs incurred by other components of the visualization system.

Software: Software maintenance will become integrated into the ALMA Integrated Computing Team. A development project will be conducted supporting members of observatory so the project will be developed in concert with ALMA ICT. Software maintenance will likely require 0.5 FTE with development of additional functionality not being included in this figure. The reason for a high maintenance allocation is because of the constantly evolving nature of web browsers. Ensuring compatibility with current version of the major browsers in the face of a constantly changing feature set will require a significant portion of this position.

Hardware: Given costs of a twelve-user visualization as described above, the costs for hardware will likely be $232,000 every five years anticipating a five year lifetime for the system based on the warranty coverage. We plan to include a hardware budget in our development proposal but the long-term costs will be borne by the observatory.

Network: It is difficult to assess the net effect on network traffic for the ALMA regional centers using this architecture plan. Because the visualization tool will allow users to explore the data without downloading to their own computers, the net outward traffic from the archive will likely be reduced by more precise downloading of optimally useful archive data. Instead, users will likely explore several files, find a subset of files that contain useful data and then extract interesting subregions for download. Acting in the opposite sense, the architecture anticipates user being able to bring supplementary data into the visualizer either from direct upload to the visualization data cache or thorough the Virtual Observatory Services.

User Experience Summary

At the conclusion of the study, we requested feedback from the expert user group on the current feature set and the development priorities. Users responded positively to the basic performance of a web-based tool, especially for a large data cube. Many users noted the aesthetics and layout of the tool as well done and several cited the 1D profile fitting tool as the most impressive aspect of the visualizer not replicated in other

tools. They found that the performance could be improved but the lag was "not unreasonable." The users identified several bugs or minor feature requests that would improve the user experience with minimal development effort.

The major conclusion of the Expert User Group survey has been to establish a design philosophy for the tool. Specifically, the *Visualization tool should be a simple but high-performance image viewer that enables good data discovery followed by flexible export of data for more detailed scientific analysis.* Features such as integrated scripting and analysis tools were deemed by the Expert User Group to be useful but not as essential as tools that facilitate discovery and encourage more detailed follow-up. Also absent from the priorities were the need for three-dimensional visualization methods such as volume rendering. These tools were viewed as interesting, but not essential at enabling discovery. Similarly, the best route to publication quality figures would be to enable agile export of data rather than building another plotting tool.

The user group also articulated their priorities for the tool and suggested several avenues to maximize impact in the ALMA community. Four features which arose consistently in our review were:

Linking views: This feature represents the ability to view different images of the same region simultaneously by, for example, showing an optical image beside the ALMA data or overlaying the data sets on top of each other.   The views on these two images are then linked together so panning/zooming in one image creates the same change in the other.

Flexible Regions of Interest: Currently only rectangular regions of interest are selectable. The user group identified being able to make polygon or circular selections as important.

Export tools: The ability to extract a region of interest and download a relevant subset of data was identified as being the best way to enable more detailed analysis. The tool should posses export tools of both image subsets and the simple analysis data such as statistical profiles.

Slices: The ability to extract a slice along an arbitrary plane in the data was identified as a critical tool to manage three dimensional data.

Future Plans

We regard the study outcomes as largely successful and will be proposing to continue the deployment of this tool as a Development Project. The goal of this development project would be complementary to others. Rather than expanding the types of data the observatory obtains, this project is designed to enable discovery in the data already obtained. A deployed tool would facilitate rapid identification of the salient features in a data set. Rapid exploration will allow users to formulate their analysis strategy before even having the data sets on their computers. By reducing the time spent transferring data and then viewing data on ill-adapted computational resources, the tool will speed publication of ALMA results and enable archive exploration.

## Appendix A: Screen Shots of Visualization Tool Prototype
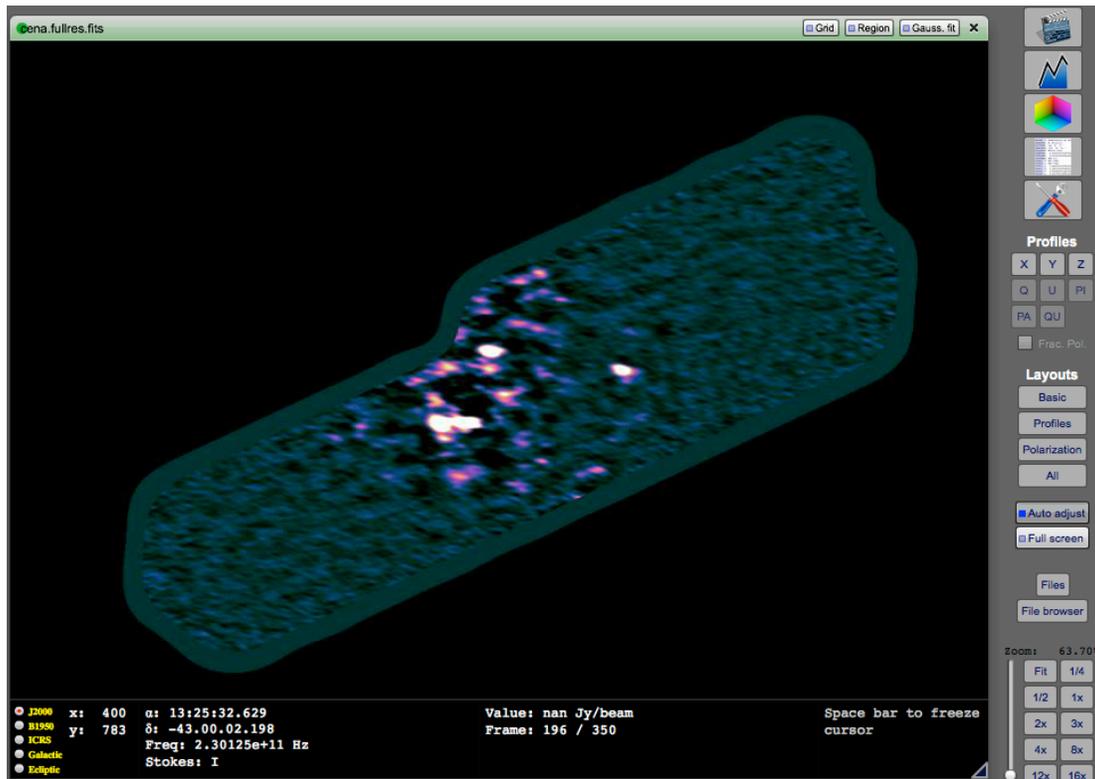
Figure A1: Basic visualization tool examining Centaurus A Science Verification data set. One plane of the image is displayed as well as a palette of options including plane selection tools, histogram control tools, color map tools, astrometric information and user preferences. Several default layouts are also given on the right hand side.
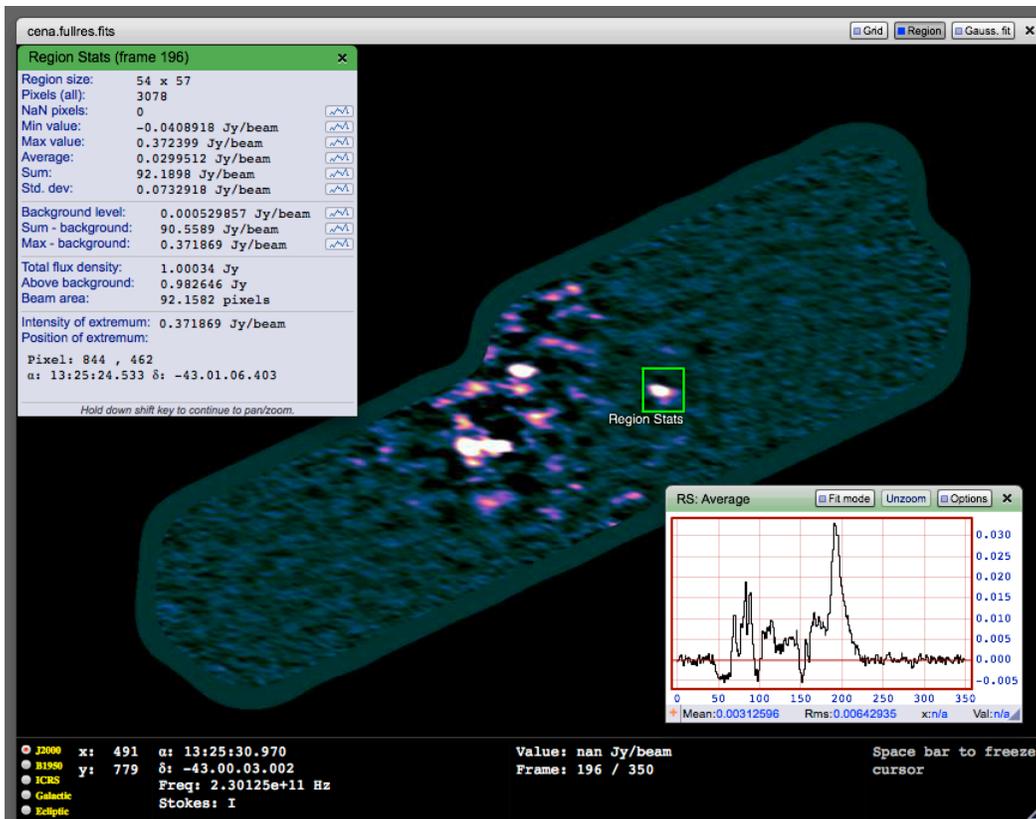
Figure A2: Region statistics mode for a selected feature in the data cube. The statistics are calculated in the highlighted box. Each statistic has a profile over the different channels within the box. The profile for the average is shown in the second window. This is the average spectrum over the box.
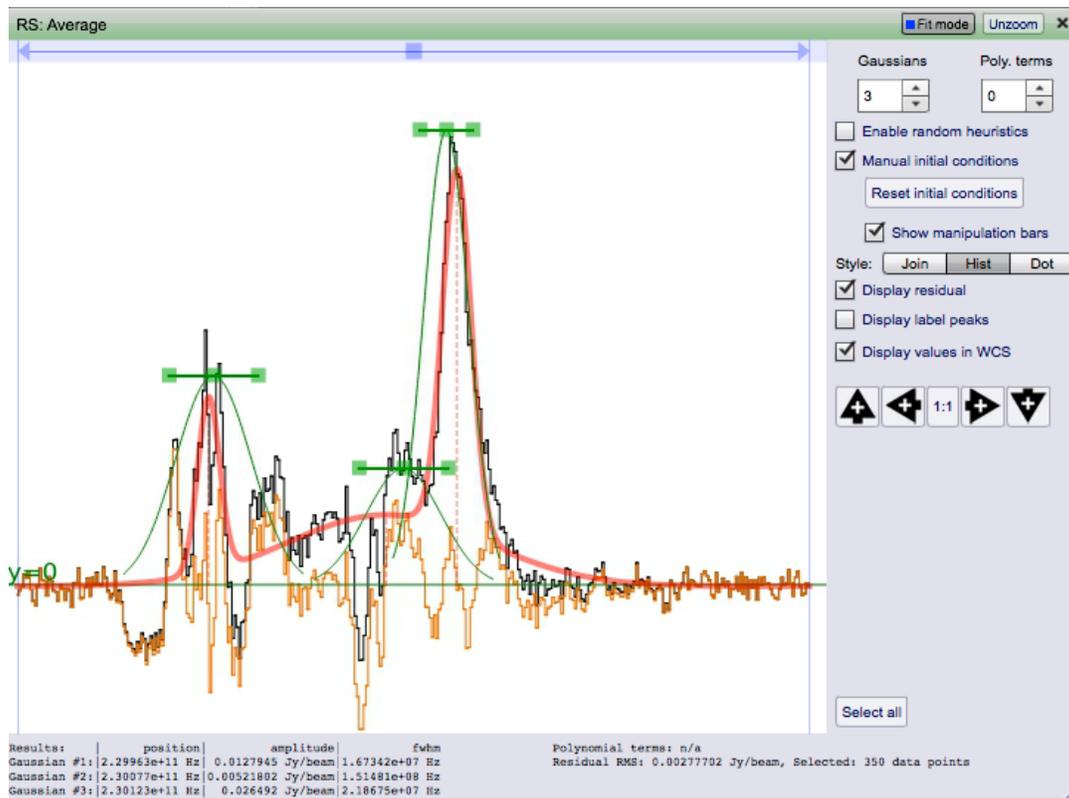
Figure A3: One dimensional Gaussian fitting dialogue. This is the fit is applied to the profile found in the average spectral profile shown in Figure A2. The Manual Initial Conditions dialog allow careful seeding of the fits on spectra fits. The widths and peaks of the Gaussian are reported in the text information box.
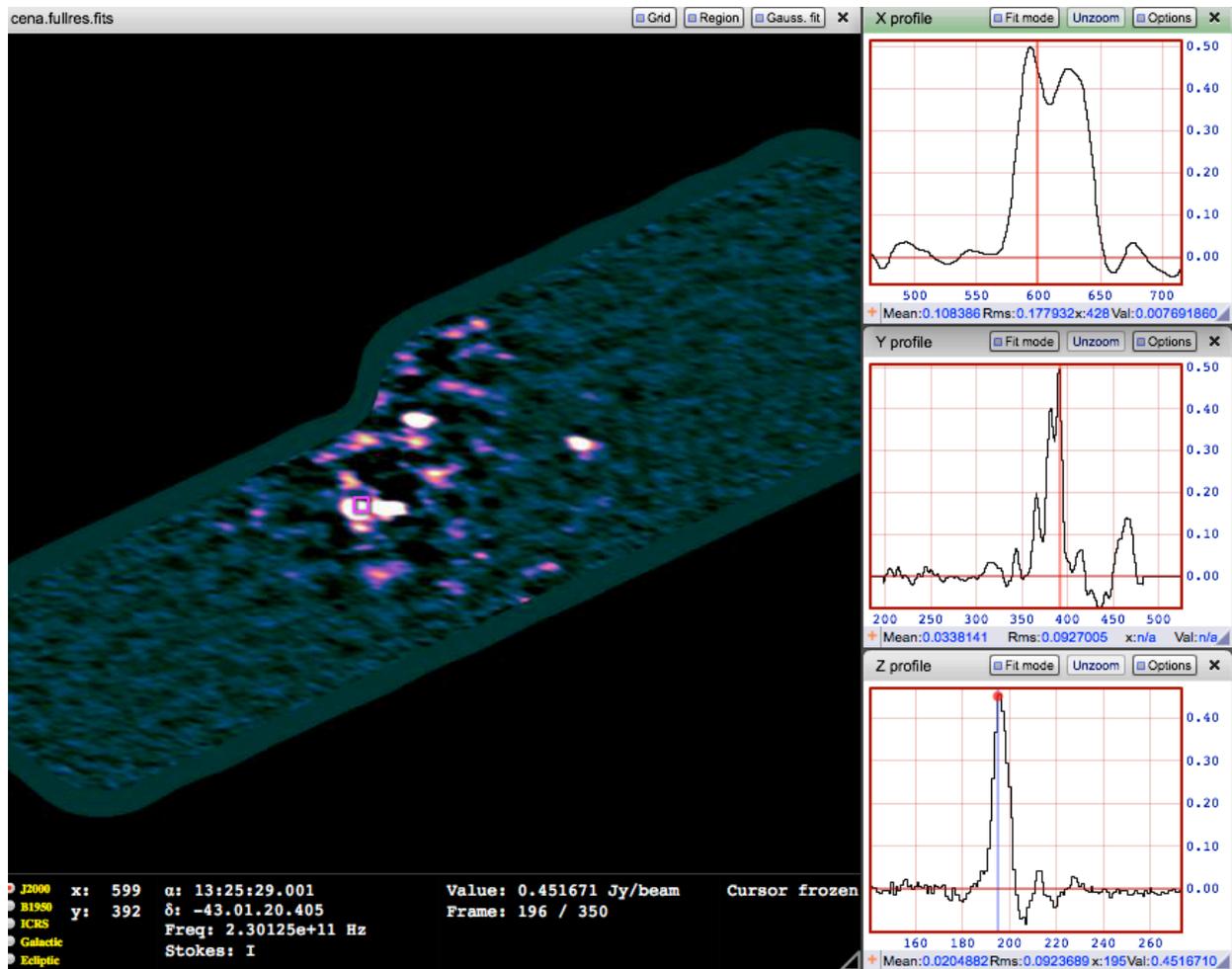
Figure A4: Profile Dialog for the highlighted pixel. This presents linear plots of slices through the data cube along the three principal axes of the data cube.
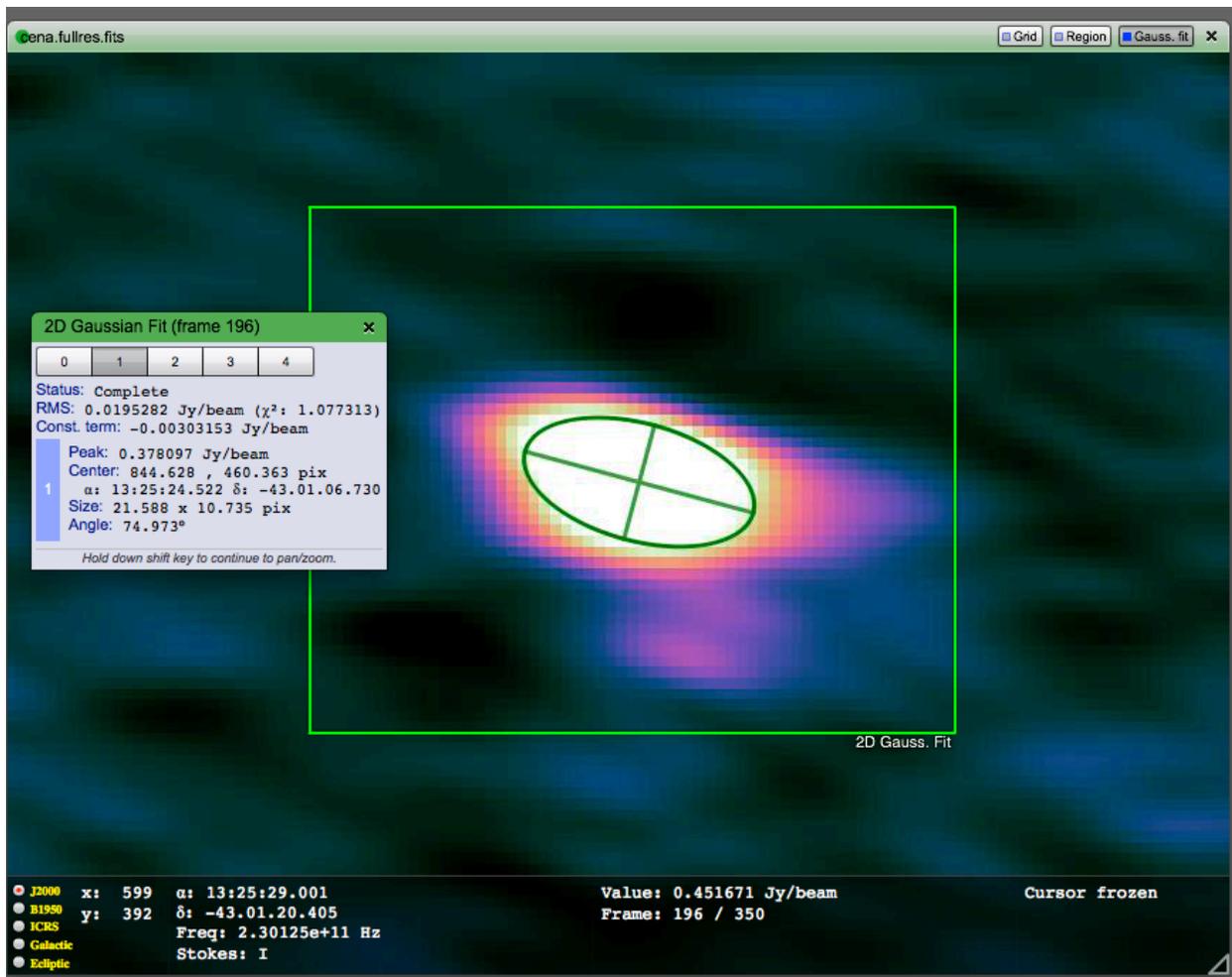
Figure A5: Two dimensional Gaussian fit for the object shown in Figure A2.
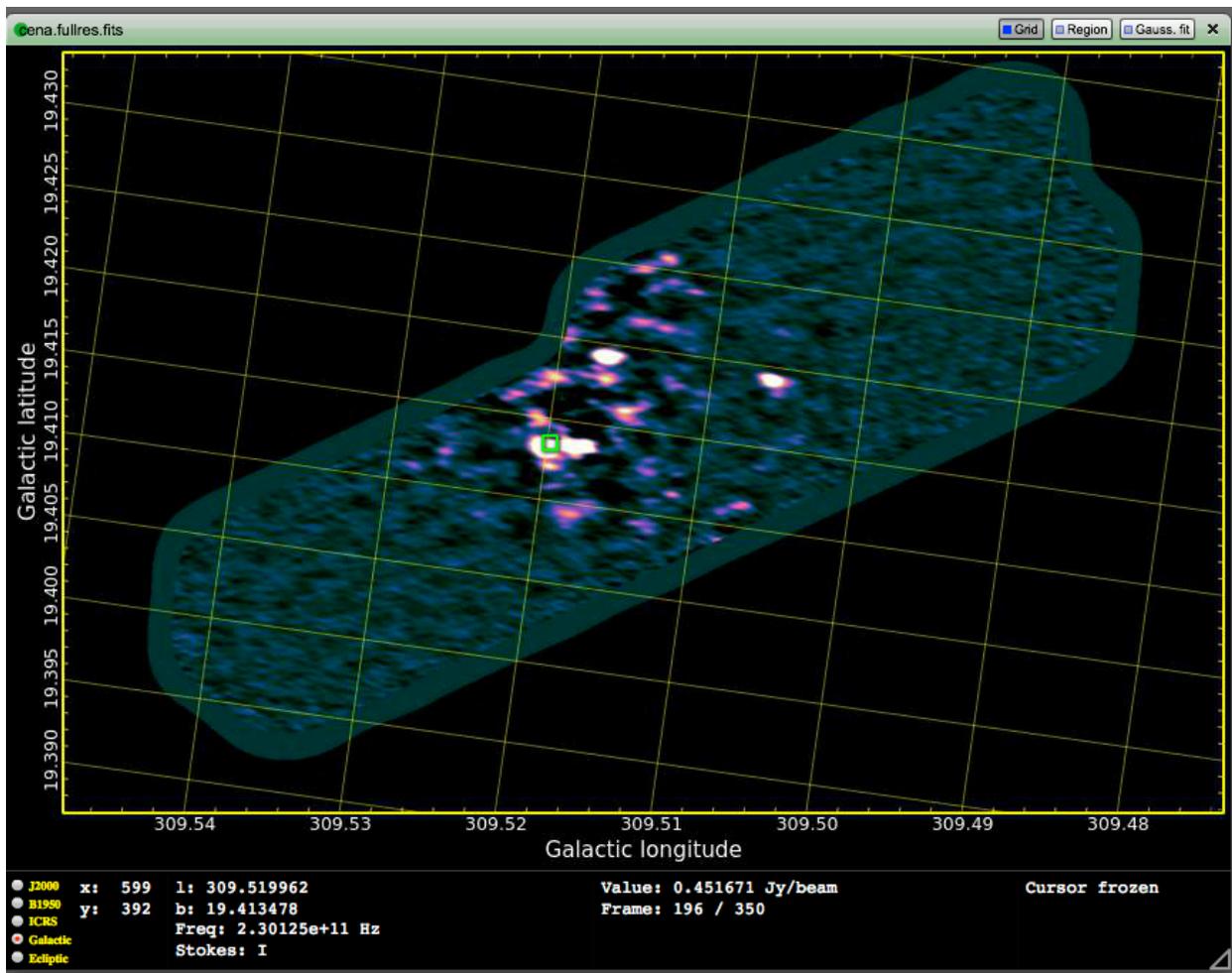
Figure A6: Coordinate grid overlay for the Centaurus A data. Several coordinate systems can be toggled between; this view shows Galactic coordinates even though the native astrometric is Celestial coordinates in J2000.
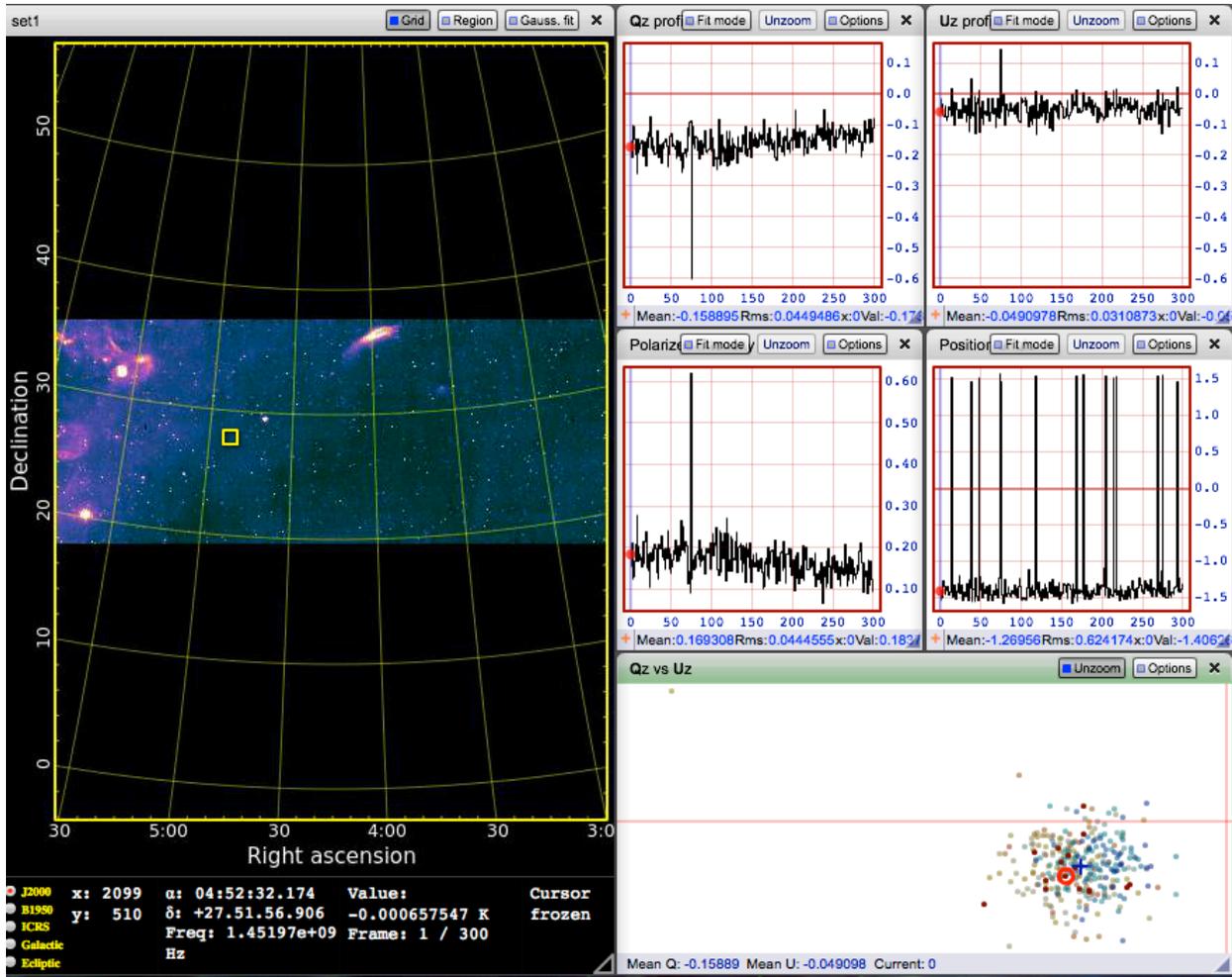
Figure A7: Polarization view for a full stokes IQUV data set. This profile shows simultaneous views through all four data cubes and point-to-point correlations between the Q and U data sets. These data are from Arecibo since full stokes polarization data are not yet available from ALMA.

Appendix B: Feedback Exercise

This document summarizes the exercises that we asked the expert user group to conduct in order to provide feedback on the visualization portal. Their feedback has formed the basis of the content in the report and guided the visualization priorities for a tool.

ALMA Visualization Tool
Critical Review of Tool Prototype

Introduction -- Thank you again for participating in the ALMA Visualization Tool Expert User Group. After our meeting in August, we have developed the visualization tool further, incorporating some of your feedback. At this point, we are soliciting your feedback on the visualization tool prototype as it currently stands. This feedback will be used to articulate the design priorities of a web-based visualization tool that would be incorporated into the ALMA Science portal through an ALMA Development Project.

In what follows, we invite you to explore the visualization tool and provide us with some information about your experience. We anticipate that the feedback exercise will take about one hour and is conducted in two parts. In the first part, we use a "scavenger hunt" to ask you to use the tool in order to answer some questions about two- and three-dimensional image data. This phase is designed to expose you to the current features in the tool and to prompt additional feature requests. Following that, we have a few questions to solicit specific feedback about the tool and future priorities.

Recent Progress -- Since our meeting in August, we have incorporated several additional features into the visualization tool. There is a lengthy development list that we have already created, and the improvements only represent a small portion of our wish list. While not immediately apparent, there have been several back-end improvements in performance, astrometric methods, code refactoring into modules, VO interoperability, and rendering methods. On the user-side there have been significant improvements including:

- Accurate grid overlays for standard astronomical world coordinate systems, including in profile windows.

- Synchronization of profiles following the mouse

- One- and two-dimensional Gaussian fitting

- Region based statistics including plane-wise estimates of the data.

At this point, the development of the current tool has come to an end, based on a decision to move from Flex/Flash to HTML5 and Javascript for the entire tool. The porting project is ongoing and most of the current functionality in the tool will be moved into the new languages within a month. In addition, we are engaged in a large effort to decouple the visualization tool from the CyberSKA portal so that it can be installed in other data portals.   Following those two steps, we anticipate our primary development goals will to polish the user interface and then we will begin implementing new functionality. Several short-term directions have been proposed including:

- Improved 1D Gaussian fitting including parameter constraints (e.g., non-negative amplitudes), exporting fit results and disjoint portions of the x-domain.

- Multiple images with locked views based on world coordinate systems. This includes separate frame views and the ability to overlay images, while adjusting opacity, blinking or mapping into RGB channels.

- Full image cube statistics and the ability to analyze both plane-wise and image-wise probability density functions with, e.g., Gaussian fitting.

- Rotation Measure synthesis (for support of the Square Kilometre Array science that the tool is also being developed for).

- The ability to customize to the user interface via settings and implementing a user-based palette of favourite tools.

ALMA Development Project -- We are planning a proposal to develop this tool into a visualization solution for the ALMA telescope through a Development Project. Given the state of astronomical visualization, we have an initial plan that the best contribution we can make to enabling science is by creating a data-discovery tool for large multidimensional data sets. This tool would be guided by the principles that it should be simple, be able to explore large data sets rapidly, be able to compare with other images, and then to export its results for more careful analysis. Please reflect on this vision as you proceed through your review.

Part 1 -- Using the Visualization Tool. In what follows, there is a list of several steps and questions about files in the Visualization Tool. The steps given here can be vague in hopes that the tool will be sufficiently self-explanatory that detailed instructions are not needed. Where clearer instructions are needed, please indicate so in your record. As you proceed, please make digital notes of (1) bugs and errors, (2) suggestions for changes in the functionality that already exists and (3) features you would like to see in future versions of the visualization tool. We would appreciate the record of your experiences with the tool.

If you encounter difficulties while going through the exercise, please contact Erik Rosolowsky (erosolow@ualberta.ca) or Pavol Federl (federl@gmail.com). In general, direct bug reports can go to Pavol and more general questions about how to use the tool can be addressed to Erik.

1. Log into the visualization tool at www.cyberska.org and select myDashboard from the menubar. There is a single, shared account for this exercise.
    1. Login: ALMAreviewer
    2. Password: A+ac@m@

2. *Optional* View the tool in action through a short screen capture video showcasing some of the features: http://www.cyberska.org/pg/file/read/28236/screen-capture-video-of-visualization-tool

3. Load the M51 B-band image into the viewer. If it is not visible in the File Dashboard tool, click the *My Files* tool to locate it or search for "M51 B-band" in the Search box and narrow your results by selecting "Files" on the sidebar.

1. Use the histogram dialog to adjust the image so that the galaxy is clearly visible. Set the bottom of the displayed range to a value of approximately zero.

2. Change the colormap and select your favourite.

3. Change the default coordinates to Galactic coordinates and turn on a coordinate grid.

4. Try zooming into the Nucleus of M51 using a left-click-and-drag. Unzoom (use the middle mouse button or the slider and presets in the lower-right corner.

5. Change the mouse functionality by selecting the "Region Statistics" button in the menu bar. What is the maximum brightness value of the image in the nucleus of the galaxy?

6. Examine the galaxy NGC 5195 (this the bright galaxy at the top of the M51 image at x=2800,y=9900). Fit a 2D Gaussian to the bright portion of this galaxy to determine the centre and size of the galaxy. Use the "Gauss. Fit" button in the image menubar and select the region that you want to the fit to run over.

4. Load the "Ophiuchus in 12CO(3-2)" data cube into the image viewer. This is a three dimensional data cube.

   1. Use the movie control panel to slide through the data until you find a plane with significant emission. The first plane that you will see when the image is loaded is dominated by noise.

   2. Adjust the colormap to highlight the bright data, setting the bottom of the displayed data range to approximately zero.

   3. Find the maximum value for real emission in the data cube over all the planes. One way to do this is to use the "Max Value" information in the Region Statistics dialog, provided the region encompasses the data and excludes the high noise region at the edge of the data cube What is the maximum value of data in this emission?

   4. Bring up profile dialog using the "Profiles" Layout button and examine the "Z" profile at the position of the brightest pixel in the cube. Use the spacebar to fix the position of the cursor at the brightest pixel.

   5. Launch the fitting profile dialog from the Z profile. Use the tools to decompose this profile into a multiple Gaussian fit (including a polynomial component if needed). Try out the Manual Initial Conditions for a finely tuned fit. What is the line width of the brightest spectral component? *Known issue*: the viewer reports the units of the third dimension of Hz for this data cube even though they are in m/s.

5. *Optional* Upload a FITS file and view it in the visualizer using the Advanced Upload tool. This launches a java client to manage the upload.

Part 2 -- Survey. Now that you have had some experience using the Visualization Tool, please return answers to the following questions along with your log from Part 1.

1. What visualization tools do you use regularly in your astronomical work?

2. Of the tools in question 1, what tool do you use most frequently?

3. In your experience with the Image Viewer, what was the most impressive or useful feature?

4. What aspect of the Image Viewer needed the most improvement?

5. Of the following items, which ones would you identify as the next major prioritys for a Visualization tool for ALMA Data? If you would like to provide a ranked list, that will help our planning.

   • Export of data (subregions, Gaussian fit propertes)
   • Visualization of point-like data (e.g., scatter plots of visibility data)
   • More analysis tools and scripting abilities
   • Overlays and linked views of multiple images / data cubes
   • Flexible region of interest selection
   • The ability to visualize arbitrary slices through data cubes
   • Alternative renderings of three-dimensional data such as volume rendering or isosurface rendering.
   • Improved performance on large data sets

6. Give an example of a use-case that cannot be satisfied with the current tool that you would deem necessary for a viable ALMA data exploration tool.

7. What do you view as the primary obstacle in creating a widely adopted visualization tool for ALMA data?

8. Please comment on what you think this would be the best way to enable scientific discovery with ALMA through visualization tools.

Appendix C: User Commentary

_____

Peter Teuben

1) started at 715pm

http://www.cyberska.org/

2) watched the movie. no comment. looks spiffy how the big cube is viewed, impressed with the speed.

3) m51:   400MB, downloaded it. 8600 x 12200 2D image;   0.05" pixels   10 x 7 arcmin area

histo: changing things there's a noticeable lag (1-2 sec?) histo dialog not so useful for this dynamic range, perhaps too few math functions the % buttons do some of the work, but all the action is in a peak near zero, hard to "zoom in" that part of the histo

4) 740pm: Oph 12co(3-2)   419MB : 314 x 370 x 946 0.001429161045 = 5.1"

looking at splash screen cyberska viewer 1.6.7

histo was (surprising) per plane, so didn't get an option to get to see the peak values very well? ok, back to 100% and then adjust. but would be nice to have a global histo here to guide you on. One problem is that the currently selected % fit is not greyed (i.e. selected). of course ungreyed as soon as user has moved the sliders at the bottom.

rms - how about a robust rms option? region below grey area: can't grab the rescaler, is that a bug?

no colorbar to numbers matching, i found it hard to find the max peak and the correspoinding plane where this happened. would be nice to see in the histo what pixel the min and max were reached.

4b) 8.09pm   m31 mosaic,   4GB image

timeout issues. wouldn't load.

overal comment: where interaction is done, e.g. gaussfitt. would be good to be able to get an output of that region to make it easier to use another tool to do the same kind of guassfit.

5) uploaded my own file.   180 MB I will receive notification. would be nice to get an idea how long this takes.. been sitting here for a few mins, my coffee is getting cold. ah, the dashboard has it.   hmm, maybe email,but i dind't see a notification. could not view image. viewer not compatible. Possibly because the NAXIS=4 and NAXIS4=1....

another overall comment:

  in the My Files, for images and cubes one should see the basic size, e.g. somehting like

  419 MB in 314 x 370 x 946 , maybe even pixel size as well,   to give a quick idea how big we're talking about .

Part 2: 1) ds9, karma (mostly kpvslice because of its unique capabilities,   kvis is powerful, but i tend to use ds9 because of the client   server (xpa) actions in ds9, and karma tools take a lot of button   clicking to get where you want to go. and one wrong click, and   the tool quits.   casaviewer: i recently started to use it. It knows fits and miriad,   very handy for us, as well. tools a bit primitive still.   gipsy: i used to use this a

lot more, but rarely need it. it has   some kpvslice type functionality in one of their tools that is   even more flexible (e.g. slices along a spiral arm or ellipse,   instead of a line)

2) ds9, since it's so super quick for 90% of the viewing you need

3) the interaction of viewer with analysis functions such as a fitter,   both profile fitter, and 2D fitter in an image.

   space bar to freeze is great, the zooming while location of cursor   stays the same (as in the new google maps) is super useful

   the profiles are very nice, but i've seen a tool in gipsy   that creates on the fly XZ and YZ slices that are pasted on   the sides of the XY view,and wiht its 3D cursor give a much   better image view of a cube.   this version is an image in XY, and 3 profiles.

   the region stats would be nice to allow it to morhp into an   arbitrary polygon, so you can fit it around an object you see

   buggy/feauture: you can drag a subwindow off the grid and if that   contains the top bar, you can't get it back   (well, toggle the parent function that got it into existence) oops, histo doesn't do it, profiles does   oops, movie also lost into eternity

4) perhaps more 3D options, slicing along other planes and arbitrary lines?

   i'd also like a colorbar wiht numbers, like ds9 has.   this is the problem if you're used to some other tools, you   tend to look for them here as well, the colorbar i missed, and using the mouse to control stretch and contrast.

5) i guess i'm seeing a lot of things here i alreayd ran into.
- export data definitley high on my listr
- point data not so high
- scripting and analysis integration seems difficult to fit in this environment, and
  reproduce casa... unless they are tied intimitately....
- linked views: YES !!!!!
- overlays , yes, i didn't miss it yet, wasn't part of the homework, but yes, we all use it all
  the time
- - volume rendering I rate not as high, perha[ps mostly because quantiative info is so
  much harder to get from... until we have 3D mice

6) select a 3D region in some way, then train it and repeat over the cube, building a catalog of its output.

7) the prime obstacle has to be speed plus ease of use. both need to   be present. next will be an way to get users to define their own   analysis tools via some kind of plugin.

8) a good way to achieve this is by keeping the viz software simple and modular. or have an API which can make the tool looks   like this. this way simple tools can be glued together to write   complex tools and also effectively interface with compute tools.   it will be important to make that access "alien", i.e. not   just via a CASA interface, but like e.g. the XPA toolset, something   generic for others to take advantage of the programming environment.

  Provice compute services, e.g. to compute a robust median for each   plane is an expensive operation on a bug cube.   Ask the tool to return a 3D wiremesh around some emission. Characterize   it, now find likewise objects in the cube, in the form of a catalogue.   Pie in the Sky :-)

ok, i'm done, this was nearly 2 hours, i've gone over my time.

Jaime Pineda

Connected from Chile on a residential internet connection (aka slow internet)

M51:
-It maybe useful to add some of the ColorBrewer in the colormap options.
-Do you really want to show the coordinate grid outside of the region with data? In my case the M51 image did not fill the display region and obviously it is filled with black pixels. When I turned on the grid it is displayed around the entire display region.
-BUG: When I am in normal mode (not full screen), I can't see all the zoom values. At the moment I do have a scroll bar in my browser but when I use it I only see the gray background. If I go to the full screen mode then I can see all the options for the zoom.

-The maximum value is 78.3459 electrons, but it also reports that the intensity of extremum is 78.0679 electrons (x=3795, y=5260). What is the difference between these two "names" and why are the values different?

-The 2D gaussian fit is quite nice, is it possible to give it an initial guess by hand?

Oph:
-Maximum value=94.439, and intensity of extremum=88.4498K @ x=107, y=226
-BUG: movie control does not use the same convention as the "Max value histogram" to numerate the planes. While movie control shows frame 417, "Max value histogram" reports 416.
-It would be great to click on the "Max value histogram" and bring the image to that plane in the viewer.
-FWHM of brightest component=724Hz (0.724 km/s as per your message)

-The manual initial condition was a good experience. The only issue is that because of my connection I could see a small delay on the initial guess to be updated.

Part 2:
1- I usually use CASA, ds9, IDL, and recently aplpy.
2-CASA
3-The calculation of the image properties, specially for cubes.
4-It is already quite a good data cube explorer. I think that more features are needed to make more useful for the community. The viewer was responsive even when connected from Chile.
5-   -The ability to visualize arbitrary slices through data cubes
       -Flexible region of interest selection
       -Overlays and linked views of multiple images / data cubes
       -Visualization of point-like data (e.g., scatter plots of visibility data)
       -More analysis tools and scripting abilities
       -Export of data (subregions, Gaussian fit propertes)
       -Improved performance on large data sets
       -Alternative renderings of three-dimensional data such as volume rendering or isosurface rendering.

6-   -PV-cuts to study rotation
       -Overplot spectra from different positions to compare the line emission

7- The main problem I see now is the lack of a proper ALMA archive. It is not clear to me that images will be easily provided soon (right now raw-data and images are delivered in a single tar file).


Chris Beaumont

M51 Exercise

The histogram icon looks more like a stock ticker than a histogram (though I figured out the correct button without too much trouble)

I would like to see icons next to each colormap, in addition to their name

The colormap dialog was partially off screen, so I couldn't close it via an "x" button in the upper corner
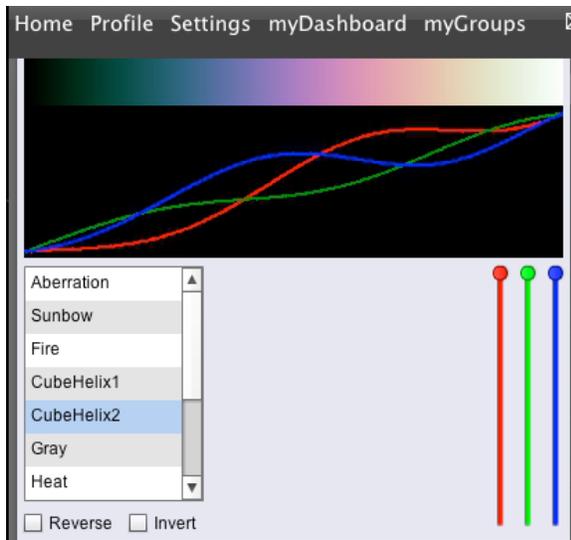
It took me a while to figure out that the 0/1/2/3/4 buttons in the Gaussian fit dialog specified the number of components. (And why is 0 an option?)

The gaussian fit to NGC5195 as a whole didn't converge after ~2 minutes:
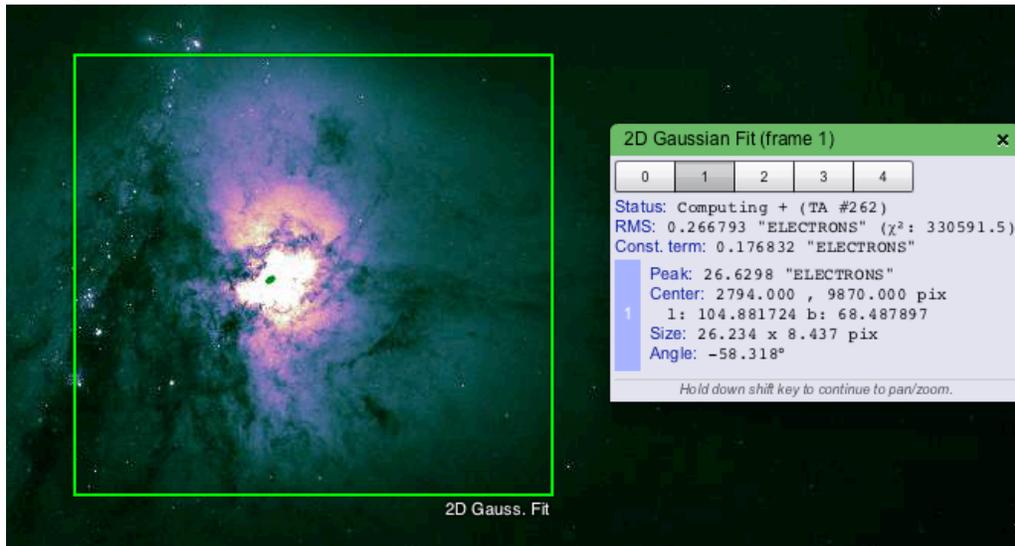
Fitting to the bright core worked well
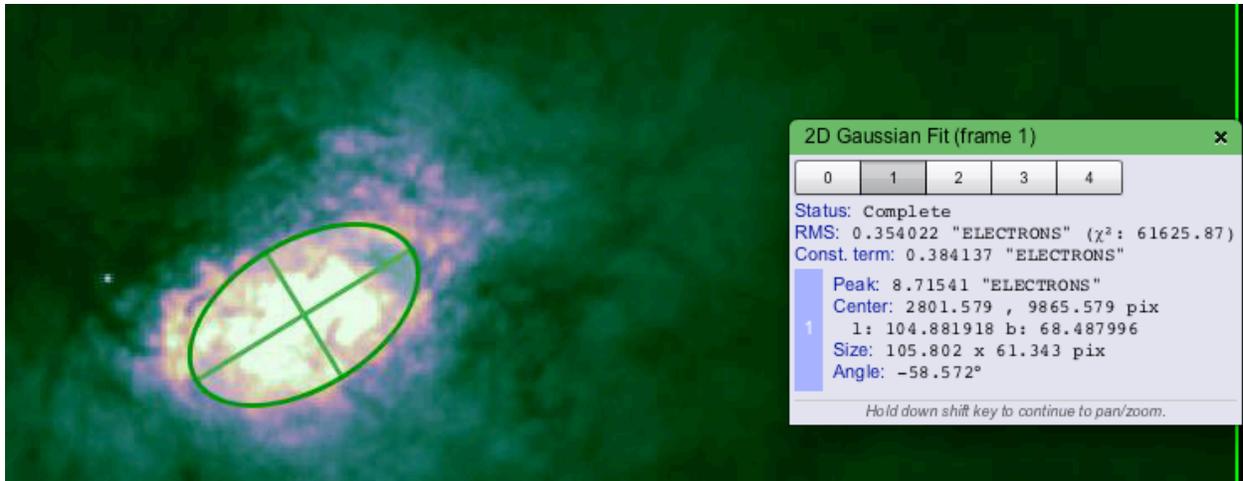
Oph Exercise

The "movie" icon is unintuitive to me

I expected that clicking on a point in the cube would pop open a spectrum (Probably because I'm used to Gaia!)

I don't get how to find the max value in the cube, since the region stats seems to be slice-specific? By sliding through slices, I think the **maximum is around 75K**

The line fitting tool is nice! **FWHM = 1030 m/s**
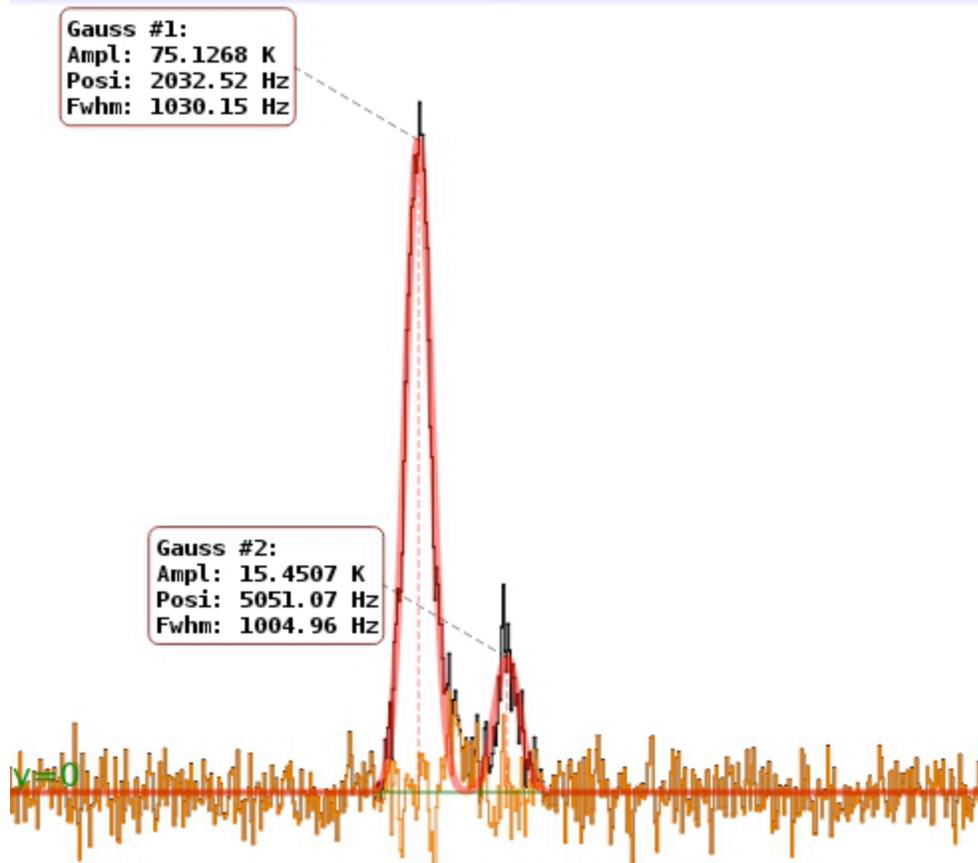
Survey

1) ds9, gaia, topcat, python + matplotlib, idl, glue

2) python, ds9, gaia

3) The line fitting profile worked very well -- much better than the clunky SPLAT setup in Gaia.

4) The icons could be more standardized. There were some latency issues, though I would happily tolerate these if the tool is equally-performant on ALMA-sized data cubes. I'd like to view PV slices

5) My order (top = highest priority)

- Improved performance on large datasets
- Scripting capabilities / more analysis
- Overlays / linked views (especially if dealing with UV data)

- Arbitrary slices
- Flexible ROI
 - point-like viz
 - export (in the context of scripting)
- 3D viz

6) Any "real" analysis I would do would be systematic, and scripted. I have never published a manually-assisted fit, since they aren't reproducible and susceptible to biased initial conditions.

I'd like to derive properties from the Ophiuchus data (kinetic energy, mass, location of outflows, etc). I can't do that now

For ALMA data, I'd probably want some ability to interact with the Fourier plane and the image plane at the same time

7) Fast exploration and analysis on 512-1024^3 data; intuitive exploration of the UV plane

8) One nice feature of having a cloud-based tool is that you could easily enable a "Stack Overflow"-like discussion forum, where users could easily ask other experts about analysis / interpretation issues. You could provide an easy interface within the viz tool to post something like a screencapture along with a question.

David Wilner

1. ds9 is my "go to" viewer for random fits images
   I use the casa viewer for ALMA and VLA data (because it's there).

2. ds9

3. The ability to deal with such a large data cube all at once is very nice!

4. These ame to mind:
   - arbitrary user-defined regions (polygons, circles)
   - examine any 2D plane or 1D slice through a multi-dimensional image
   - some form of 3D rendering and manipulation

5. It's challenging to rank these, as they are all useful/important.

flexible region of interest selection
export of data
ability to visualize arbitrary slices
overlays and linked views
alternative renderings of 3D data
more analysis tools
improved performance on large data sets
visualization of point-like data

6. For ALMA, it would help a lot to be able to create position-velocity diagrams along arbitrary angles.

7. To gain wide use for ALMA, I think the tool will need to seemlessly digest image cubes in the native casa format. In addition, if you can provide useful quantitative analysis algorithms (beyond simple Gaussian fitting) that people find useful, then there will be many adopters. Have you thought about ways that users could (easily) contribute analysis modules to the visualization tool?

8. Astronomers benefit from different ways to "see" their datasets, so this visual exploration approach can have considerable value. I think success in the long run will depend on quantitative analysis, though.

Adam Leroy (NRAO)

Survey then minor picky feedback. It's in *fantastic* shape. I showed it to a few folks around here and we're all excited.

Survey:

(1) In order at this point: CASA Viewer, ds9, kvis. Plus a lot of custom 2-d plots in IDL/python.

(2) CASA Viewer and custom plots.

(3) The speed and simplicity, especially for the profile views and picking a good stretch. It does a really good job of showing me the buttons that I am most likely to use. The Gaussian and polynomial fitting works great and has a very good visualization - this is speaking from having been traumatized by this apparently being really really hard for

our developers. Honestly the polish is at the level that I would already work it into the rotation.

(4) The gaps that I feel most strongly are the lack of collapse functionality (moment maps + threshold + slicing) and the weakness of the astronomy in the z profile. More in the next item.

(5) My feeling is that ALMA's three big challenges are (1) "huge, mostly empty data", (2) "complicated spectra", and (3) "big, complicated emission in a cube." You are making awesome progress on #3 but I would prioritize having the basic answers to #1 and #2 in there. Roughly ranked:

- *Linked views* coupled with basic thresholding/masking and collapse functionality. The linked view of a moment map and a cube is a basic, critical cube tool. Creating these things with some space to define signal is a basic analysis step.

- Stronger spectral handling (rest frequencies, line ids, spectral smoothing). I need some ability to pick over a complicated spectrum. The fits are great but where is the rest frequency/velocity interaction, potentially spectral smoothing, and potentially line identification. I should also be able to get the z-profile / spectrum for a region (right now I get a lot of things like this from the region tool but not the spectrum itself, right?

- Export of data - fits, spectra, subregions, and collapsed images.
- Arbitrary slices
- Flexible ROI
- Scriptability and more tools
- Point like data (primarily if you link it)
- Alternative renders

Throughout, I would stress how much I like the simplicity of the current tool (speaking as someone who has been fighting to simplify the viewer while adding features ...) and suggest that you find ways to add things without losing that.

(6) I can't identify lines or create moment maps. It's not a great tool right now to search for faint stuff. I can't compare the distribution of two lines or say Halpha and a galactic outflow (overlays/linked views). But what is there is awesome.

(7) For this tool I feel like you need to deal with:
- getting the word out
- lower the bar to using it (lots of up/downloads are a pain, it's the biggest obstacle I see)
- exposing hookups to out-of-tool analysis
- you will need to decide how you fit in with publication-quality plots, ultimately you have to fit in path-to-publication (could be just exporting the right slices for matplotlib or whatever)

(8) Just back to my earlier points:
- help people find needles in haystacks (look for signal in noisy, big cubes)
- help make complicated, multi-line cubes simple and tractable
- quick and easy exploration of big, full cubes

Other Feedback:

(1) The tool looks fantastic. Aesthetics are excellent - in particular, I think you are doing an outstanding job of making good decisions balancing general-purpose visualizer and realistic astronomical tool. Specifically, things like the radio button for coord grids, the well-worn choices of histogram percentiles, that kind of thing. Very nice overall. Use of transparency very good. I think this "feels" exactly right.

(2) I had mechanical issues with seeing the bottom of the tool. Chrome, Mac laptop. I only got about halfway down the zoom menu (clipped at 2x 3x) with no way to scroll all the way down. There was a little gray bar at the bottom. Minor mechanical issue, but annoying. Full screen fixes it.

(3) On the wireless at NRAO things are laggy but workable. I had to basically take a deep breath before a color palette change. Not unreasonable. It doesn't hold input while working, which I had to get a bit used to.

(4) It did not visualize the first Gaussian fit of M51b (perhaps because it was very small). It did show the second and third Gaussians. I also apparently have really weird taste and kept getting into a situation where the fit Gaussian (in green) was hard to see.

(5) The 2-D Gaussian fits weren't particularly illuminating...

(6) The mouse / fit / region interaction was not super intuitive without your instructions.

(7) I actually really like the "fit mode" ... I don't know that I would use it for much in some of the cases where I was playing with it (channel stats for Rho Oph) but it's beautiful and responsive and I would for sure use it to play with profiles... very nice. I didn't see how to do this for the pixel values overall (just per plane). Can I do that? I would use that. Anyhow the profile fits are *fantastic.*

(8) Why can't I 'X' out of the color table tool?

(Note--comments on tool functionality have been entered in comments section mostly re:M51 image, and some for Oph and Antennae--see logs.)

Alyssa Goodman (CfA)

1. What visualization tools do you use regularly in your astronomical work?

ds9, WWT, Glue, DataDesk, Photoshop

2. Of the tools in question 1, what tool do you use most frequently?

depends on task!

3. In your experience with the Image Viewer, what was the most impressive or useful feature?

layout is nice aesthetically, and aesthetics are very good overall

4. What aspect of the Image Viewer needed the most improvement?

see comments re:M51--I think serious control over what's being fitted, recorded, constrained are most important

5. Of the following items, which one would you identify as the next major priority for a Visualization tool for ALMA Data?
• Export of data (subregions, Gaussian fit propertes)
• Visualization of point-like data (e.g., scatter plots of visibility data)
• More analysis tools and scripting abilities
• Overlays and linked views of multiple images / data cubes
• Flexible region of interest selection
• The ability to visualize arbitrary slices through data cubes

• Alternative renderings of three-dimensional data such as volume rendering or isosurface rendering.
• Improved performance on large data sets

None of the above. I think that connecting this tool to other tools that can do some of what's on the list above is most sensible. Otherwise, I'd choose 3D volume viz!

6. Give an example of a use-case that cannot be satisfied with the current tool that you would deem necessary for a viable ALMA data exploration tool.

There are plenty. Most important is the abilty to bring several data sets together & see them all at once in a linked-view environment--but we're working on that w/Glue.

7. What do you view as the primary obstacle in creating a widely adopted visualization tool for ALMA data?
Astronomers' luddite ways--so you need to make something so much more capable than what they use already that they'll take notice.

8. Please comment on what you think this would be the best way to enable scientific discovery with ALMA through visualization tools.

more connection to analysis tools, preferably ones people use anyway

Re: image cube:
I am not offered "view movie" the way you were in your screen capture movie, so I can't do what you ask! System seems not to know that this is a 3D cube?

Re: 2D image
- latency in adjusting image was long, so I could not tell when my actions had "taken" -- need progress bar
- could not see where to type comments like these in CyberSKA platform WHILE M51 window was open
- CyberSKA seemed to only open one window at a time, so the work on M51 was lost if I went back to other things (e.g. commenting on the intro movie)
- histogram icon should look like a histogram! also--I don't think of adjusting an image stretch as adjusting a histogram! why not combine the histogram (range/stretch) tool with the color one & just call it "image adjustment"?
- gaussian fitting interface is cool, but unclear--was at first unclear what numbers 0 1 2 3 4 meant--now I see they likely refer to how many Gaussians to add together to make a fit--but since parameters are not user-constrainable, this is weird. How would the program decide which 3 Gaussians to fit in a complicated source? (e.g. two sharp ones and a flat one, or 3 sharp ones?) I'd want to know. Also, is output of results just supposed to be cut & paste? I can't even do that, as text doesn't seem selectable.
- the icon used for "settings" normally means "tools"-- a gear usually means "settings" region statistics is interesting, but with no i/o for constraints (pre-chosen boundary other than a rectangle in, or output of results to savable file or at least cut/paste), not so useful at present, I think
- jump in window appearance when grid is toggled on/off is jarring--the region shown appears to change boundaries--why??
- x,y (,z) profile viewer is cute, but latency makes it hard to use--you need "real time" performance cognitively, for sure; you also need a visual other than just the cursor showing where one is pointing in an image to get this profile--the fixed axis should be indicated, as should be the position along the interactive one
- gaussian fitting does not work (is inactive) if region stats is open at the same time (not good!)
2D gaussian fit of NGC 5195 will not converge for a single Gaussian! (works for other more point-like sources in the image, but not for your chosen example!)