Development Upgrades of the Atacama Large Millimeter/submillimeter Array (ALMA)

Project Proposal Form

PROJECT PROPOSAL TITLE: ALMA DATA MINING TOOLKIT

PRINCIPAL INVESTIGATOR: **LEE G. MUNDY**

INSTITUTION: UNIVERSITY OF MARYLAND

ADDRESS: ASTRONOMY DEPARTMENT, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD

PI CONTACT INFORMATION:

Telephone Number: 301 405-1529

Email address: lgm@astro.umd.edu

1.0 Co-Investigator(s) and Collaborating Institution(s)

Name	Institution	Email / Telephone
Jeff Kern	**NRAO**	**jkern@nrao.edu / 575 835-7391**
Adam Leroy	**NRAO**	**aleroy@nrao.edu / 434 244-6807**
Leslie Looney	**University of Illinois**	**lwl@illinois.edu / 217 244-3615**
Anthony Remijan	**NRAO**	**aremijan@nrao.edu / 434 244-6848**
Amitabh Varshney	**University of Maryland**	**varshney@umiacs.umd.edu/301405-6722**

2.0 Subcontractors

2.1 Company/Institution: University of Illinois

Address: Office of Sponsored Programs and Research Administration
1901 South First Street, Suite A
Champaign, Il 61820

Subcontracted parts and/or services:

Participation in software tasks assigned in the WBS

2.2 Company/Institution: NRAO

Address: **Anthony Turner**
NRAO, Charlottesville, VA,
aturner@nrao.edu
434 296-0343

Subcontracted parts and/or services:

Software interface and expertise with the molecular line database Splatalogue.
Software and management expertise and interface to the NRAO CASA group

Institutional Letters of Commitment are in Appendix A

3.0 Project Plan

3.1 Science Case

The production of rich datasets is in the DNA of ALMA. Mining the science from them is fundamental to unleashing the full power of ALMA as an instrument. We propose to create a value-added software package which integrates with the ALMA archive and CASA to provide scientists with quick access to traditional science data products such as moment maps, and with new innovative tools for exploring data cubes. The top-line goals of the software package are to: (1) make the scientific value of ALMA data more immediate to all users, (2) create an analysis infrastructure that allows users to build new tools, (3) provide new types of tools for mining the science in ALMA data, and (4) increase the scientific value of the rich data archive that ALMA is creating. The ALMA Data Mining Toolkit (ADMIT) provides capabilities, missing in the current ALMA baseline system, that are key to achieving ALMA's full science potential for both proposers and archival data users.

The proposed software toolkit will increase the accessibility of ALMA data to the broad astronomical community by making traditional data products directly available from the archive. These products can also be made available through the archive database and/or Virtual Observatory services, if ALMA choses. The toolkit is also a platform for innovate tools which build on the ADMIT infrastructure. We provide examples of two such tools: overlap integrals and descriptor vectors.

3.2 Project Overview

The ADMIT creates an infrastructure and tools which enable scientists to better mine the science potential of their data, and to better mine the gold in the growing archive of ALMA data. The ADMIT, through the automated production of basic data products will significantly enhance the accessibility of data in the ALMA archive which will have major positive impact on the total scientific productivity of ALMA. This project assists ALMA in two of its major goals: maximizing the scientific value of ALMA, and maximizing the value of ALMA to the broadest community of scientists.

3.3 Project Scope

The project will deliver to ALMA ADMIT software which runs robustly within the framework of CASA and the archive. The ADMIT software will: (1) create basic data products for the ALMA archive, (2) create a software infrastructure system for ADMIT tools to run in CASA on the user's local machine, (3) provide a graphical user interface for inspecting data products, (4) provide selected new tools for mining data cubes. We have an understanding that the ALMA archive will support the ingestion and serving of ADMIT Basic Data Products and that CASA will support distribution of ADMIT software as an add-on package. ADMIT will be designed to be compatible for data product ingestion into the ALMA archive database and compatible with use of CASA Viewer for displaying appropriate data products; but we will not be developing tools for either use. The project will deliver the ADMIT software at the end of 2 years from project commencement for just under $1 Million.

3.4 Project Measures of Success and Risk

The ADMIT measures of success will be:

(1) benchmark Basic Data Products (BDP) on test maps with different observational configurations bands, and correlator setups with 100% repeatability and accuracy,
(2) successful creation of complete BDP for >95% of ALMA projects with graceful failure modes that provide available data products.
(3) Create BDP from pipeline maps using <45 min of cpu time for > 50% of projects,
(4) delivery of ADMIT metadata XML files for archive ingestion for > 95% of projects
(5) user survey of pilot ADMIT GUI release to have a 75% approval for usability.

We will track the performance metrics for items 2, 3, and 4 by monitoring the statistics as the modules are tested and enter utilization. The final measures of success will be the statistics during the last month of the proposal performance period. Item 5 will be measured with a user survey in Quarter 1 of 2015 and a follow-up survey in Quarter 3 of 2015.

The designed implementation of ADMIT provides robust mitigation of most risks. As ADMIT relies on ALMA pipeline image cubes, the largest risk is delay in the ALMA pipeline full-band imaging products (see Table 7.0). The ALMA project goals include production of full maps of all of the bands so this is only a question of timing. Significant pipeline delay is a low risk during the duration of our project; the primary mitigation is to process whatever pipeline products are available which will decrease the value of the ADMIT BDP in the archive. One of the key strengths of ADMIT is the ability of the user to run the entire package on their local machines. A user can make their own images with their own parameters and run the ADMIT analysis on their image cubes to recoup the full analysis. This is a second mitigation path.

A second risk is that the archive user interface will not be upgraded to allow for easy identification of the ALMA image cubes for each band in the archive. The baseline plan is to do this identification within the ADMIT XML files; if this is not possible, the risk is mitigated by the user manually matching the appropriate archived files with the ADMIT products. We will stay in contact with the archive team over the period of this proposal to create the best possible solution.

Finally, the last risk is that some desirable analysis will be too CPU extensive for the computer resources available at the ARCs. Again, all tools will be included in the ADMIT package and the users can run any of the analysis tools on their own machines. This risk should also be reduced as the observatory goes through refresh cycles with the pipeline computer hardware.

Note that in all cases, the risks are completely retired without impact to cost.

3.5 Project Deliverables

3.5.1 Hardware: no deliverables

3.5.2 Software

The following are software products to be delivered:

- ADMIT Python main module
- Test suite and inline documentation for main module
- ADMIT Core Tools
- Test suite and inline documentation for each core tool
- ADMIT Graphical User Interface

These constitute the components needed for the ADMIT infrastructure, BDP pipeline, user display of BDP, and the tools discussed in the science and technical sections.

3.5.3 Services: no significant deliverables

3.5.4 Documentation

- Monthly "4-square" Progress Reports
- Software design documentation

    - ADMIT design documents
- Interface Control Documents (ICDs)
- Technical manuals and procedures
    - ADMIT User Guide
    - ADMIT Tool Writer's Guide
- Quality Assurance procedures
    - Software test and acceptance procedures
- Final Report
- Outcomes Report

3.6 Site Location Impact Statement (if applicable)

We anticipate no significant impact on existing facilities. The proposed ADMIT procedures to create the BDP would run at ALMA facilities on the ALMA pipeline image cubes. We will target ADMIT procedures to take no more than 5% of the cpu time required for the pipeline to process corresponding data cube. The ADMIT Basic Data Products would be stored in and served from the ALMA data archive; the BDP for a project is expected to be 5-40MB, which less than 1% of the typical project data size. Similarly due to the small size of BDP, there is minimal additional load on the archive server.

The ADMIT package is proposed to be an add-on package to CASA distributions.

4.0 LEVEL OF EFFORT AND COST BREAKDOWN

Title	Key Personnel *(leave blank if inapplicable)*	FTE	Duration (Months)
Principal Investigator	Lee Mundy	0.14	24
Scientific Lead	Leslie Looney	0.14	24
Software Lead	Peter Teuben	0.70	24
Science		0.4	24
Software Development		2.16	24
Management	Marc Pound	0.40	24
Test and Document		0.75	18
	TOTAL FTEs	**4.69**	

4.1 Period of Performance

The requested performance period is 24 months. Due to the stated limits on spending in Year 1, we have allocated our FTE's to be roughly 1/3 in Year 1 and 2/3 in Year 2.

4.2 Level of Effort:

*Table 1.0; **Offerer's Labor Estimate**.*

Item	Labor Category	FTE
1	Science	0.44
2	Software Engineering	3.00
3	Management	0.50
4	Documentation and testing	0.75
	Total Labor	**4.69**

4.3 Cost Breakdown:

*Table 2.0; **Offerer's Cost Estimate**.*

Item	Cost Category	USD ($)
1	Labor	917,247
2	Materials & Services	15,163
3	Travel	58,750
4	Miscellaneous	0
	Total Cost ($)	**991,160**

*Table 3.0; **Total Cost of Project**.*

Project Participants	USD ($)
Expected contribution from the NA ALMA Development Program	0
Offerer's Total Cost	991,160
Total Cost of Project ($)	**991,160**

4.4 Cost Distribution

*Table 4.0; **Project Cost Distribution**.*

FY2014 Cost ($)	FY2015 Cost ($)	Total Cost ($)
325,284	**665,876**	**991,160**

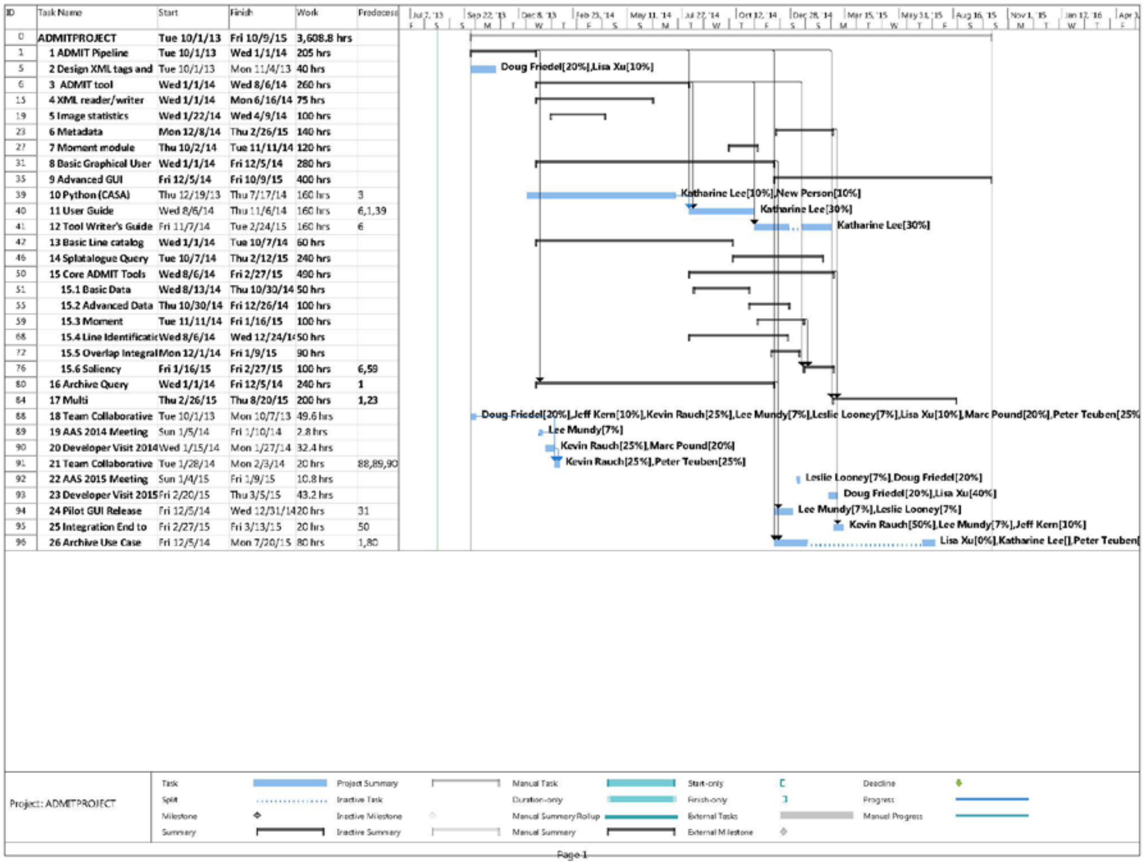
4.5 Budget Summary

*Table 5.0; **Project Budget Summary**.*

WBS No.	Task Description	Labor ($)	Materials & Services ($)	Travel ($)
1.0	Admit Pipeline	51250	14910	
2.0	Design XML tags and structure	1000		
3.0	Admit tool architecture	65000		
4.0	XML reader/writer	18750		
5.0	Image statistics	25000		
6.0	Metadata summarizer	35000		
7.0	Moment Module	30000		
8.0	Basic graphical user interface	70000		
9.0	Advanced GUI	100000		
10.0	Python CASA documentation	40000		
11.0	User Guide	40000		
12.0	Tool Writer's Guide	40000		
13.0	Basic Line Catalog query	15000		
14.0	Splatalogue Query	60000		
15.0	Core ADMIT tools	122500		
16.0	Archive Query	60000		
17.0	Multi-source operations	50000		
18.0	Team collaboration visit 2014	12500		8000
19.0	AAS 2014 Meeting	750		5000
20.0	Developer visit 2014	8000		10000
21.0	Team collaboration visit 2015	5000		10000
22.0	AAS 2015 Meeting	2750		9000
23.0	Developer Visit 2015	10750		14000
24.0	Pilot GUI Release	5000		
25.0	Integration end-to-end testing	5000		
26.0	Archive use case review	20000		
26.0	Prepare closeout report	18000		
	SubTotals ($)	**920250**	**14,910**	**56,000**
			TOTAL COST ($)	**991,160**

Material & services are mainly computer connection and services fees which spread over all tasks.

4.6 Schedule Summary



ID	Task Name	Start	Finish	Work	Predecess
0	ADMITPROJECT	Tue 10/1/13	Fri 10/9/15	3,608.8 hrs	
1	1 ADMIT Pipeline	Tue 10/1/13	Wed 1/1/14	205 hrs	
5	2 Design XML tags and	Tue 10/1/13	Mon 11/4/13	40 hrs	
6	3 ADMIT tool	Wed 1/1/14	Wed 8/6/14	260 hrs	
15	4 XML reader/writer	Wed 1/1/14	Mon 6/16/14	75 hrs	
19	5 Image statistics	Wed 1/22/14	Wed 4/9/14	100 hrs	
23	6 Metadata	Mon 12/8/14	Thu 2/26/15	140 hrs	
27	7 Moment module	Thu 10/2/14	Tue 11/11/14	120 hrs	
31	8 Basic Graphical User	Wed 1/1/14	Fri 12/5/14	280 hrs	
35	9 Advanced GUI	Fri 12/5/14	Fri 10/9/15	400 hrs	
39	10 Python (CASA)	Thu 12/19/13	Thu 7/17/14	160 hrs	3
40	11 User Guide	Wed 8/6/14	Thu 11/6/14	160 hrs	6,1,39
41	12 Tool Writer's Guide	Fri 11/7/14	Tue 2/24/15	160 hrs	6
42	13 Basic Line catalog	Wed 1/1/14	Tue 10/7/14	60 hrs	
46	14 Splatalogue Query	Tue 10/7/14	Thu 2/12/15	240 hrs	
50	15 Core ADMIT Tools	Wed 8/6/14	Fri 2/27/15	490 hrs	
51	15.1 Basic Data	Wed 8/13/14	Thu 10/30/14	50 hrs	
55	15.2 Advanced Data	Thu 10/30/14	Fri 12/26/14	100 hrs	
59	15.3 Moment	Tue 11/11/14	Fri 1/16/15	100 hrs	
68	15.4 Line Identificatio	Wed 8/6/14	Wed 12/24/14	50 hrs	
72	15.5 Overlap Integral	Mon 12/1/14	Fri 1/9/15	90 hrs	
76	15.6 Saliency	Fri 1/16/15	Fri 2/27/15	100 hrs	6,59
80	16 Archive Query	Wed 1/1/14	Fri 12/5/14	240 hrs	1
84	17 Multi	Thu 2/26/15	Thu 8/20/15	200 hrs	1,23
88	18 Team Collaborative	Tue 10/1/13	Mon 10/7/13	49.6 hrs	
89	19 AAS 2014 Meeting	Sun 1/5/14	Fri 1/10/14	2.8 hrs	
90	20 Developer Visit 2014	Wed 1/15/14	Mon 1/27/14	32.4 hrs	
91	21 Team Collaborative	Tue 1/28/14	Mon 2/3/14	20 hrs	88,89,90
92	22 AAS 2015 Meeting	Sun 1/4/15	Fri 1/9/15	10.8 hrs	
93	23 Developer Visit 2015	Fri 2/20/15	Thu 3/5/15	43.2 hrs	
94	24 Pilot GUI Release	Fri 12/5/14	Wed 12/31/14	20 hrs	31
95	25 Integration End to	Fri 2/27/15	Fri 3/13/15	20 hrs	50
96	26 Archive Use Case	Fri 12/5/14	Mon 7/20/15	80 hrs	1,80

*Figure 1.0; **Level 1 Project Schedule.***

5.0 Delivery Schedule and Terms of Payment

*Table 6.0; **Project Milestone and Payment Schedule**.*

Item	***Project Milestone***	***Time of Completion***	***Payment (USD)***
1	Milestone 0: Project commencement	T_0	240,000
2	Milestone 1: XML and pipeline structure defined and prototyped	T_0 + 6 months	85,284
3	Milestone 2: Basic line catalog query and ADMIT tool infrastructure completed, XML structure completed	T_0 + 12 months	250,000
4	Milestone 3: Successful creation, ingestion and delivery of BDP utilizing ALMA Archive, testing of basic GUI	T_0 + 15 months	150,000
5	Milestone4: Full BDP created for >90% of projects. Overlap integral and descriptor vector tools completed	T_0 + 18 months	150,000

	and under use testing, Field tested User Guide delivered		
6	Milestone 5: All capabilities delivered and undergoing use testing for robustness against unusual dataset and user errors, Field tested Tool Writer's Guide delivered.	T_0 + 21 months	115,876
		TOTAL FIRM FIXED PRICE ($)	**991,160**

6.0 COMMITMENT

Having read all documents listed in and annexed to the Call for Development Project Proposals, and having assessed the situation and the nature and difficulties of the proposed services, the undersigned hereby offers the "***ALMA Data Mining Toolkit***" in accordance with the provisions of the present Call for Development Project Proposals and, if awarded the Agreement, undertakes to carry out the work required according to best trade practices, within the prescribed time limits, and at the price set out in this Proposal Form.

Name: Lee Mundy

Institution: University of Maryland

Signature: Lee G Mundy

Date: August 16, 2013

University of Maryland institutional letter of commitment is in Appendix A

Development Upgrades of the Atacama Large Millimeter/submillimeter Array (ALMA)

PROJECT PLAN

ALMA Data Mining Toolkit

PRINCIPAL INVESTIGATOR: LEE MUNDY

INSTITUTION: UNIVERSITY OF MARYLAND

ADDRESS: ASTRONOMY DEPARTMENT, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742

PI CONTACT INFORMATION:

Telephone Number *301 405-1529*

Email address *lgm@astro.umd.edu*

ABSTRACT

We propose a software package which integrates with the ALMA Archive and CASA software system to provide scientists with immediate access to traditional science data products such as moment maps, and innovative tools for exploring data cubes. The principal goals of the software package are to:

(1) make the scientific value of ALMA data more immediate to all users,

(2) create an analysis infrastructure that allows users to build new tools,

(3) provide new types of tools for mining the science in ALMA data, and

(4) increase the scientific value of the rich data archive that ALMA is creating.

These capabilities, missing components in the current ALMA baseline system, are vital additions for achieving the full science potential for both proposers and archival data users.

The ALMA Data Mining Toolkit (ADMIT) consists of two main components. The core is the infrastructure layer that defines data structures and programmatic pipelines; provides methods for data selection, and computation of scientifically relevant quantities; extracts and organizes scientific metadata from the cubes; and defines the application programming interface for higher-level tools. Built on the infrastructure layer are the ADMIT Tools that provide advanced functionality for scientific analysis. ADMIT defines a standard architecture that allows use of existing analysis tools and creation of custom analysis tools. We will deliver a core set of tools that employ both simple (data summary, moment) and advanced algorithms (line identification and cutout, descriptor vectors). These tools will enable science mining for a broad range of projects and are flexible enough to be tailored by the user to specific science objectives and bulk processing needs.

The ADMIT will increase the scientific productivity of ALMA for the proposing scientists and for future archival work. Increasing access to the scientifically interesting data is fundamental to achieving ALMA's goal of being a transformative scientific instrument, and of being inclusive of the broad community of astronomers. The innovative tools that are proposed here, and that users can grow on the infrastructure of ADMIT, are essential to uncovering the full range of science hidden in the information rich datasets, and to freeing ALMA to achieve its full potential.

TABLE OF CONTENTS

1.0 SCIENCE CASE

The production of rich datasets is in the DNA of ALMA. ALMA's sensitivity, resolution, frequency coverage, and flexibility enable transformative research in a wide range of scientific fields because its vast, complex datasets will contain unparalleled science opportunities. Mining the full range of science from the data is fundamental to unleashing the full power of ALMA as an instrument. The objective of this proposal is to provide cutting-edge tools for mining the science from ALMA releasing the ancillary treasures in the ALMA archive and adding value to all science projects.

Why is data mining important to ALMA? The majority of pixels in an astronomical data cube are in fact noise, with an abundance of scientific information spread among that noise. Consider the typical molecular spectral line observation of a galactic molecular cloud or nearby galaxy. The ALMA correlator provides 4 basebands of 2 GHz width per polarization with up to 4096 channels per baseband per polarization. With this capacity to gather large quantities of data, the typical proposal will have several science goals that require observations of three or more lines and the underlying continuum dust emission. Often, the science goals would require observations of several different target locations or target sources to, for example, mosaic image a region or gather a modest sample of 5-20 objects. Such projects generate tens of data cubes with one or more molecular lines of interest in each, and possibly several serendipitous lines in each cube as well. The scientific results need to be identified and mined efficiently out of this mountain of data and also have the flexibility for unexpected scientific discoveries.

What are the benefits of improved tools for mining the science? First and foremost, it will increase the scientific productivity of the instrument for the proposing scientists and for future archival work. Second, increasing access to the scientifically interesting data is fundamental to achieving ALMA's goal of being a leading instrument for the broad astronomical community. And, third, innovative tools are needed to discover the full range of science hidden in the information rich datasets.

We propose to create a value-added software package that integrates with the ALMA archive and CASA to provide scientists with immediate access to traditional science data products such as moment maps, and provide innovative tools for exploring data cubes. The proposed package is called ADMIT, for ALMA Data Mining Toolkit, a software suite that creates value-added data products from ALMA data image cubes directly from the ALMA pipeline for ingestion into the archive and provides a standalone infrastructure and applications for mining user-downloaded cubes. The top-line goals of the ADMIT are to:

(1) make the scientific value of ALMA data more immediate to all users,

(2) create an analysis infrastructure that allows users to build new tools,

(3) provide new types of tools for mining the science in ALMA data, and

(4) increase the scientific value of the rich data archive that ALMA is creating.

ADMIT provides capabilities, missing in the current ALMA baseline system, that are key to achieving ALMA's full science potential for both proposers and archival data users.

Goal 1: to provide an immediate data access capability. This will be accomplished by creating a set of basic data products (BDP) which are included into the archive and available to the users for quick overviews of their data. The BDP are a combination of XML files, JPEG images, and FITS files that will be produced by ADMIT software (a combination of Python scripts and CASA programs) from the ALMA's approved pipeline image cubes (all correlator bands). The BDP consist of harvested header information about the source and observational setup, statistics of the data cube, spectra through emission peaks, identification of strong molecular lines present in the data, and integrated intensity, velocity centroid, and velocity dispersion maps of each identified line. The BDP images will be associated with XML metadata that contain the information about their creation. The BDP, in the range of 5-40 MB in size, will be ingested into the ALMA archive concurrent with ingestion of the ALMA pipeline image cubes. The ADMIT BDP will be available to the archive user as an independent item that can be selected and downloaded separately from the large image cubes. On the user's local machine, ADMIT software (distributed as a package in CASA) is used to view and manipulate the BDP. Since the BDP are primarily XML and JPEG, the information can be accessible to the ALMA archive database and CASA Viewer if desired by ALMA. The small size and summary scope of the BDP are ideal for (1) quickly exploring what is in the data set, (2) deciding what full cubes should be downloaded, and (3) cross comparing detections or emission properties across multiple sources. An ALMA example is examination of results for a spectral line survey of 25 sources. Without downloading any of the data cubes, the scientist could quickly view the moment maps to classify sources based on line detections, line widths, and kinematics. Then, the full data cubes for sources of the most interest could be downloaded to further explore populations. The BDP assist in this process because they include the information about where the lines are located within each cube.

Goal 2: to provide a simple infrastructure that can be easily adopted by users to create new data mining tasks. We will support this capability by building the infrastructure from small re-usable unit tasks, documenting the pieces and creating example use cases. Since ADMIT is a combination of XML, Python, and CASA, the component languages are all familiar to moderately experienced CASA users. Users can add features directly to create their own custom ADMIT pipeline and add new tools to the toolkit that can be shared among collaborators and the community. A custom ADMIT pipeline is especially useful for large surveys where astronomers need to uniformly process the sample, perhaps repeating the analysis many times to iteratively adjust parameters after examining the results.

Goal 3: Use the BDP and ADMIT infrastructure as groundwork for building innovative tools for mining information from the data cubes. We propose to build several specific tools as examples of the potential of ADMIT. First, with the spectral lines identified, it is possible to calculate overlap integrals between velocity channels to see how the emission changes within an individual line, and/or to calculate the overlap integral between different lines in a dataset to see how the spatial distribution of the emission is correlated. We will write a CASA task to do this with the outputs of CASA image and XML/JPEG data products. An example use case is an observation of a hot core source where the data cubes have scores of detected lines. The overlap integrals will show how the spatial distributions of the lines compare.

Past these familiar analysis tools, we will develop analysis based on a descriptor vector approach to characterizing the observed emission. The idea here is taken from contemporary computer science image analysis techniques where the image is divided into a set of sub-regions (for example Nyquist-sampled beams), each characterized by an N-dimensional vector according to a set of prescriptions. The distance between vectors can then be calculated to identify which parts of the image are most similar or most different, or to search for which parts of the image are most like a selected region. For instance, in the image recognition world it is possible to define a vector that well-characterizes the image of a person within a multi-gigapixel data cube, and then find all of the people in the data cube by identifying all vectors with similar characteristics. An ALMA example would be to analyze spatial molecular line distributions in images of a local galaxy. Starting with a dataset of N line images, a descriptor formulation can build vectors for each Nyquist-sampled region (or a chosen resolution) across the image with significant emission in any line. These vectors would characterize the emission in each molecular line in the project (all bands and multiple correlator setups if observed) by, for example, intensity and line width. ADMIT would then be able to identify regions throughout the image that have vector values close to a user selected reference position; for example place where both CO and SiO were present but not CH_3OH. With current software, this comparison would require tedious visual inspection of the cubes. For this proposal, we will write infrastructure for the descriptor vector task and create descriptor formulations for several standard ALMA applications. We will document the descriptor formulation procedure so that users have the option to build on this approach, share vector formulations, and build libraries of formulations.

This innovative tool aspect of ADMIT can also be a compute layer for creating new image products that can be fed into an independent visualization tool. With XML, JPEG, and FITS/CASA images as the ADMIT products, it will be easy to design-in compatibility with CASA Viewer to enhance its use. We would also look forward to working with any new ALMA visualization tool that might be funded to define an XML/image interface.

Goal 4: to optimize the long term science value of the ALMA archive. The key here is to make the archival data easily accessible to science-based data-mining investigations. The BDP for each project is a first step in this direction. It allows the scientist to download the overview information to see if the project contains data of interest. The second step is that the ADMIT scripts to create BDP can be executed on the local machine to create a new and, if desired, custom version of the BDP. Thus, the scientist could download 10 or 50 data cubes from a single or many projects onto a directory on their local disk and then run the ADMIT script to generate new BDP that contain information for all of the image cubes in the directory. A future possibility would be for the BDP XML metadata to be ingested into an archive database tool to enable searches based on user requests; this would be an optional outcome to be implemented in collaboration with the ALMA Archive team if they choose. An ALMA use case example is a scientist downloading the BDP of two or three large surveys of Class 0 protostellar sources. From examining the BDP, the scientist can decide what full images cubes to download from the archive. With all of the cubes on the local directory, the custom ADMIT scripts can be run to create a uniform set of BDP for the selected sources which sets the stage for comparisons and further analysis.

It is important to emphasize that the ALMA archive will be a tremendous resource for science after only a few years of operation. Tools that maximize the science from the ALMA archive amplify the impact of the instrument at little cost. As an example, Figure1shows that the Hubble archive is extremely productive, producing more archival-data based papers per year than papers from current observations. In addition, science papers with archival datasets have been found to have about as much impact in the community to original papers from the same datasets [1]. It is crucial improve the end-to-end user access to the ALMA data even at this early stage because ALMA data, with so many spectral channels, is much richer scientifically than the early HST data.

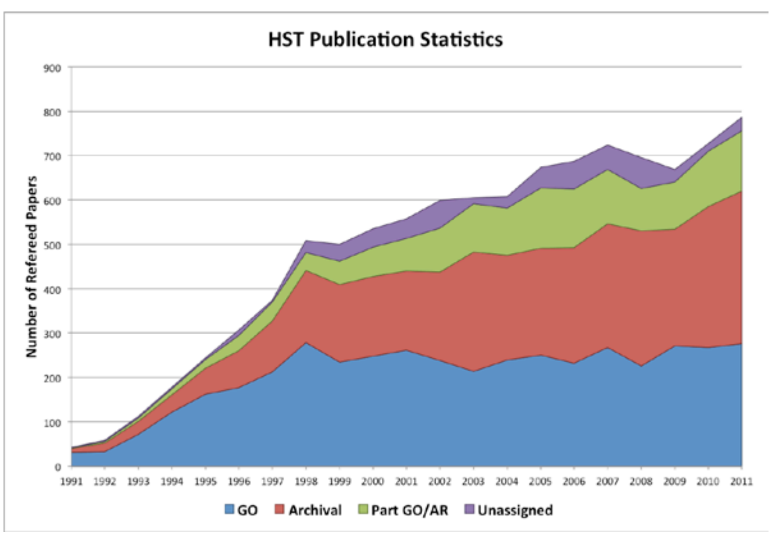


Figure 1: Number of HST GO and archival papers published as a function of year (reproduced from [1]).

1.1 Technical Approach

1.1.1 Overview

ADMIT is an add-on software package that interfaces to the CASA software system and ALMA archive—implemented both at the server side of the archive and as a desktop application with a full GUI. The ADMIT plan is the outcome of a funded ALMA Study proposal from last cycle. The Study focused on data access and data mining options, and how to improve the scientists' access to the science in ALMA data. In the course of the study, we prototyped tools for examining data, interacted with computer science colleagues to identify new methodologies to bring into astronomy, and interacted with the members of the NAASC, CASA, and ALMA Archive groups to hear about the needs and the near-term plans of the groups. The plan in this proposal is based on the knowledge and insights gained from the Study; the examples in this section were made with prototype tools and scripts. The end-result of this proposal will be to create a robust, well-documented ADMIT system for deployment. We understand the difference between prototypes for expert-use and reliable, robust, daily-use tools; significant intellectual effort and testing in the proposed work will go toward achieving that goal, and that is reflected in our measures of success.

The ADMIT core is the infrastructure layer that defines XML structures and programmatic pipelines, extracts and organizes scientific metadata from image cubes, defines the application programming interface (API) for higher-level tools, provides for I/O interaction with the XML, and computes scientifically relevant quantities. Built upon the infrastructure layer are specific pipelines to produce Basic Data Products (BDP) and the ADMIT Tools that provide advanced functionality for scientific analysis. ADMIT is targeted at both a novice user (via the convenient GUI), as well as an experienced user (via a Python toolkit within the CASA framework), and is designed to be extensible.

The interface to CASA is accomplished primarily through XML files and Python scripts. Where possible, ADMIT will call CASA tasks via the XML parameter files to accomplish computations. The new tools in the proposal will be Python or CASA tasks, depending on the required level of access to CASA data and whether the task requires speed optimization through more direct access to the CASA core routines. ADMIT will be designed and tested to work in the CASA environment.

The interface to the ALMA Archive will be simply through the creation of an ADMIT data package: a tarball or other standard package format as requested by the Archive. The ADMIT data package will be ingested, archived, and served to the user as a single item. The user downloads the ADMIT data package by selecting it through the standard ALMA Archive interface. This approach has been discussed with members of the ALMA archive team (Felix Stoehr and Mark Lacy) and is straightforward. Since ADMIT outputs are standard XML and JPEG images, it is a future option for the Archive Database to ingest the ADMIT information to enhance the archive search capability; it is not a baseline requirement for ADMIT but it is designed as a feature.

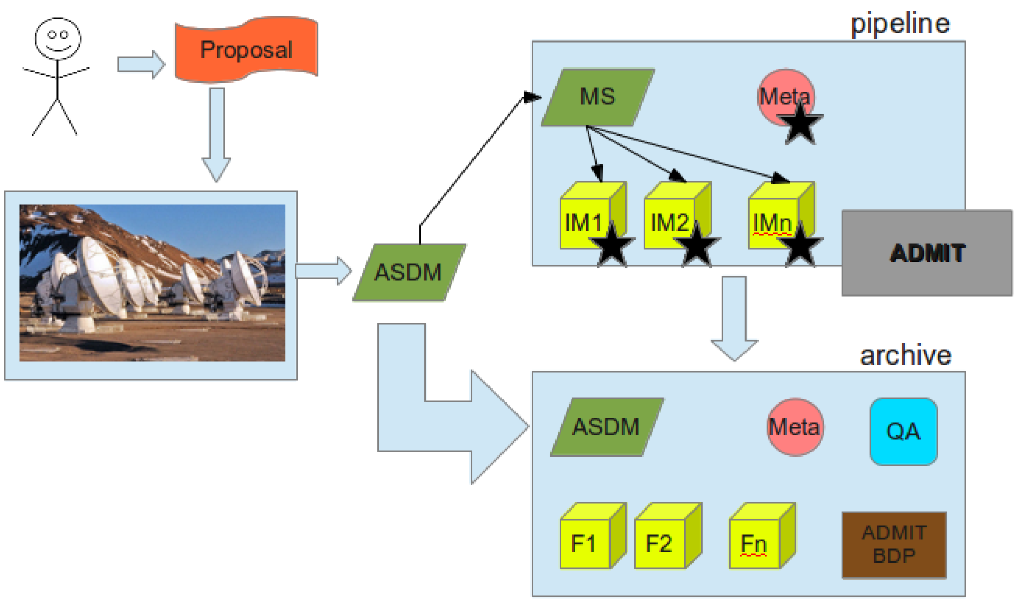


Figure 2. ADMIT and the ALMA Archive Interface. ADMIT is an add-on program that works on the data cubes output from the ALMA pipeline and their associated metadata (all marked with star symbols).

Figure 2 shows how ADMIT slots into the Reduction Pipeline and ALMA Archive. The ADMIT core processes (gray box) are executed after the observatory certifies that the image cubes meet proposal science requirements. Information for the BDP is gathered from the image cubes, relevant ALMA pipeline metadata outputs, and calculated quantities. These are packaged into a file that is ingested (brown box). Figure 3 shows the ADMIT BDP in the archive. Through the standard ALMA Archive browser interface, they can be downloaded to a target machine and viewed either through the ADMIT GUI or the Python toolkit interface.

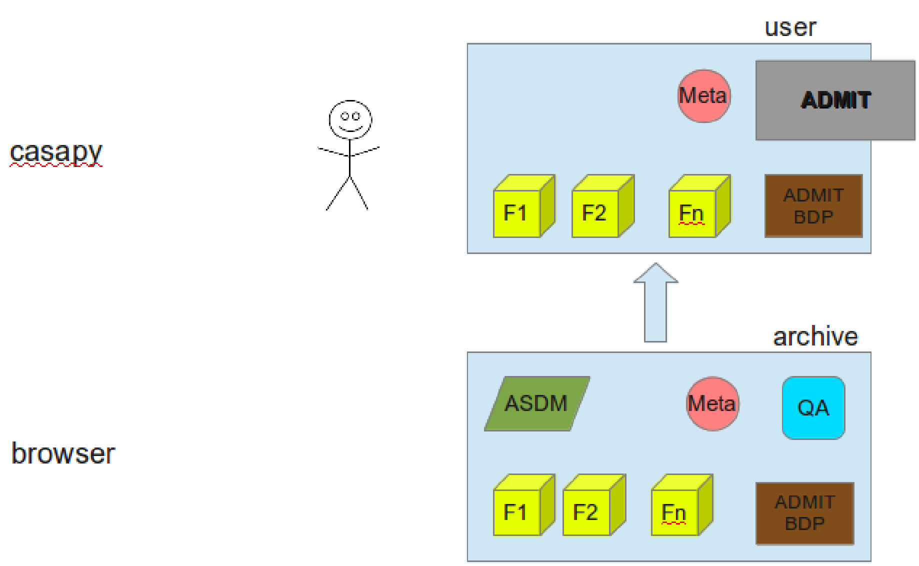


Figure 3. ADMIT Basic Data Products in the archive browser and desktop views.

1.1.2 Infrastructure Layer and Pipeline

The infrastructure of ADMIT consists of a CASA-compatible Python module of access routines and an XML-format metadata file. The XML metadata file, created in the BDP or by any re-running of the ADMIT script locally, contains information about the observational configuration, source, data statistics, and discovery information about the identifications of the lines in the observed bands, channels with detected lines, and data products. Simple routines access this metadata to retrieve information as needed for the current operation. For interfacing to CASA tasks, the metadata is used to construct XML parameter files which are used to run the task. Each ADMIT operation builds additional detail into the XML file creating a more comprehensive, scientifically useful description.

The ADMIT pipeline is a series of operations that create metadata then utilize the metadata to create higher level products. This approach is simple to expand to add further operations or to customize for specific applications. The default ADMIT pipeline consists of the summary information, noise levels, spectra, line identifications, moment maps, overlap intervals, and image saliency information. ADMIT can work with images in both FITS and CASA format. The outputs of the pipeline are wrapped into a single, self-describing tar file, which any ADMIT Tool can parse and manipulate. A typical tar file might contain a descriptive XML file and associated small FITS and JPEG files.

ADMIT runs at the archive, but it can also be called by a user from within the CASA Python environment as individual commands or a script. This flexibility allows a user to tweak the ADMIT settings, add user functionality for a specific survey or project, or create more complete metadata. The infrastructure can encompass multiple projects and sources, and perform the same operations on each source. This can be done either in parallel or recursively. It can then extract information from each project and in essence mine a large suite of data, allowing linked data techniques to visualize the extracted information and provide new insight on what is common or different in the sources.

1.1.3 The Toolkit

Once the core ADMIT components have created the XML metadata, ADMIT uses the toolkit component to explore the metadata and make it direct assessable the user (both novice and expert). This is possible because ADMIT defines a standard architecture that allows users to use existing analysis tools or to create and “plug in” their own analysis tools to ADMIT. ADMIT provides an API for these tools to call existing CASA commands, manipulate images or metadata, compute new quantities, return data structures to casapy, and store results in the tar file with new metadata written to the XML. Below we described the initial set of tools accessed in the GUI that will typically be run as a desktop application (see Figure 3). These tools will deliver science mining applicable to a broad range of sources and encapsulate the best methodologies arrived at through the community’s many years of experience.

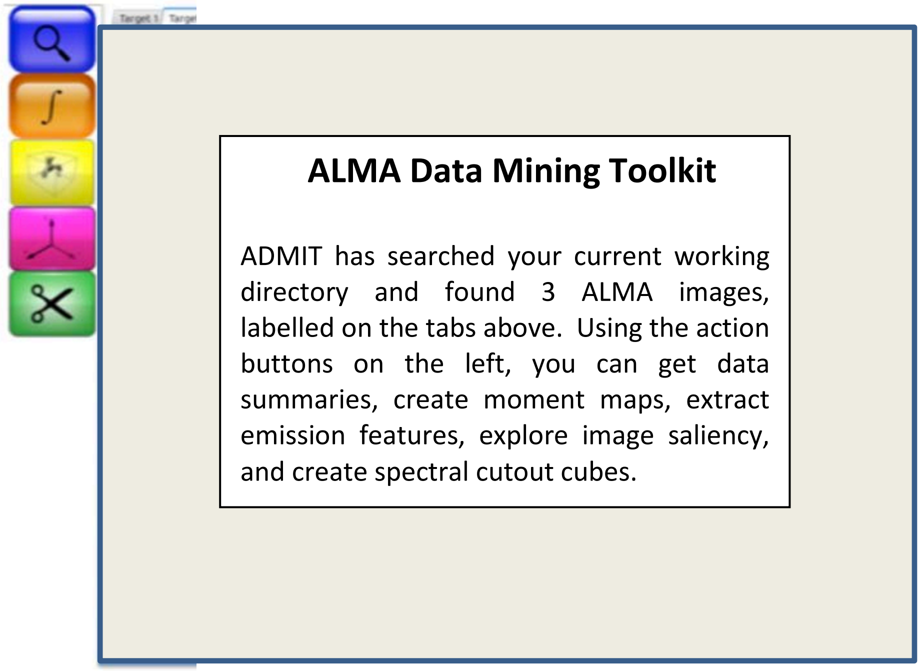


Figure 4. Mockup of the start screen of the ADMIT GUI (desktop application). On the left from top to bottom are buttons to invoke ADMIT Tools (Data Summary, Moment, Feature Extraction, Saliency, and Line Cube Cutout) for the target source selected in the middle panel. On the top row, tabs are visible for a number of collected sources.

1.1.4 Data Summary

The Data Summary gives the scientist a quick, visual answer to the question, "What's in my data?" ADMIT answers that question, but as a powerful addition, the Data Summary also provides the answer to that question at any point or step in the analysis process. ADMIT collects metadata and writes them to an XML file based on a set of ADMIT schema. The metadata are gathered from the image cube headers, any metadata produced by the ALMA pipeline, and computed quantities from the ADMIT pipeline. When run locally, the default is to gather the XML data from all ALMA image files in the current directory. Some components of these metadata contain project information (e.g., sources, sky positions, ALMA bands used), others may be computed values (e.g., image statistics such as min, max, mean, robust median, RMS per channel, etc.), still others will be determined by any previous workflow from ADMIT analyses. Each ADMIT Tool operation will update the metadata so that an up-to-date summary of everything that has been gathered and computed is quickly presented (Figure 5). Selecting a particular band, line, or source would open up more detailed views of the BDP (Figure 6).

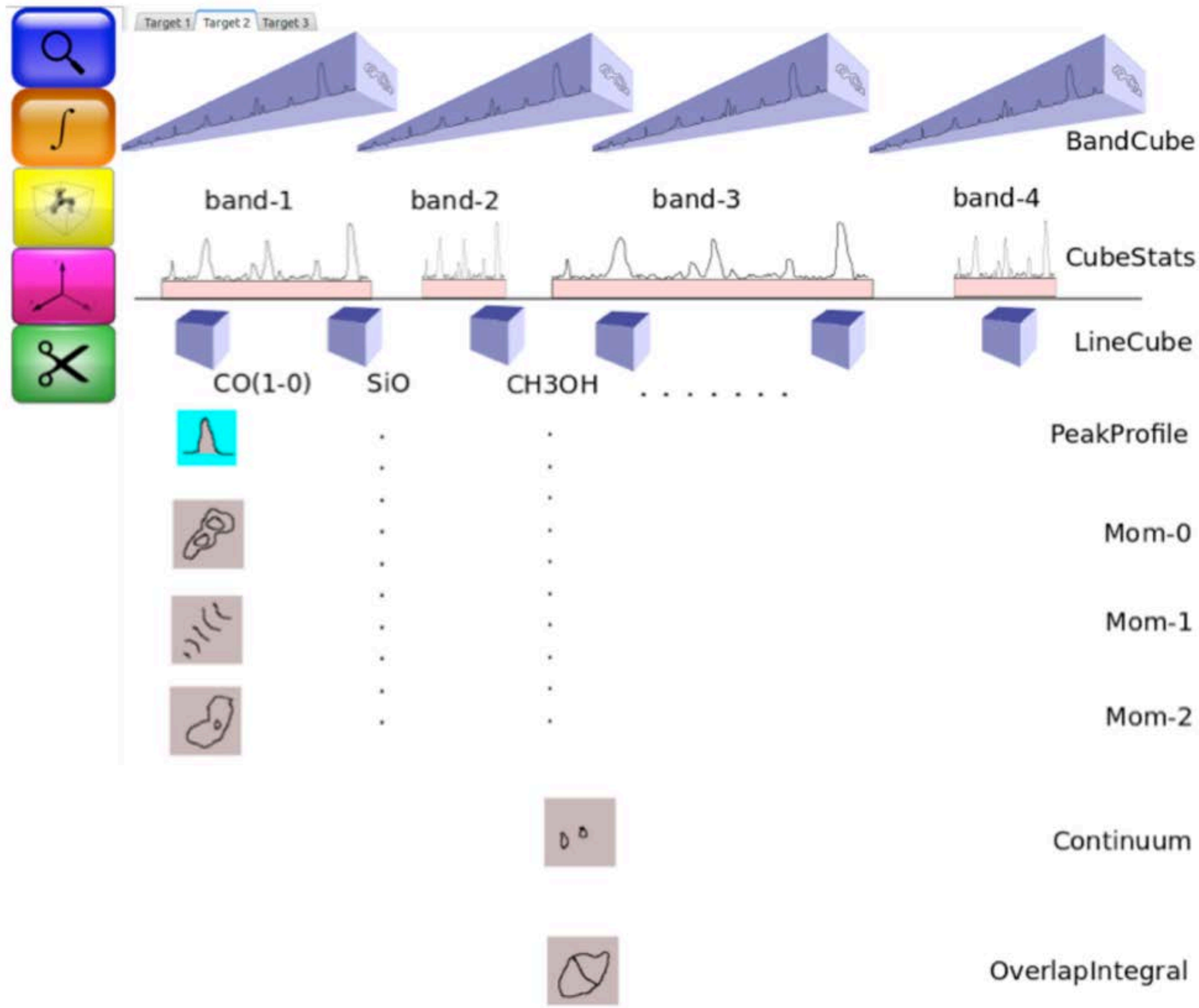


Figure 5. A mockup of the Data Summary that would be presented to the user in the ADMIT GUI. Each row shows the output of an ADMIT Tool operation listed on the right, giving an easy-to-understand visual summary of the science data and tool outputs for the target source selected in the middle panel. On the top row, tabs are visible for a number of collected sources.

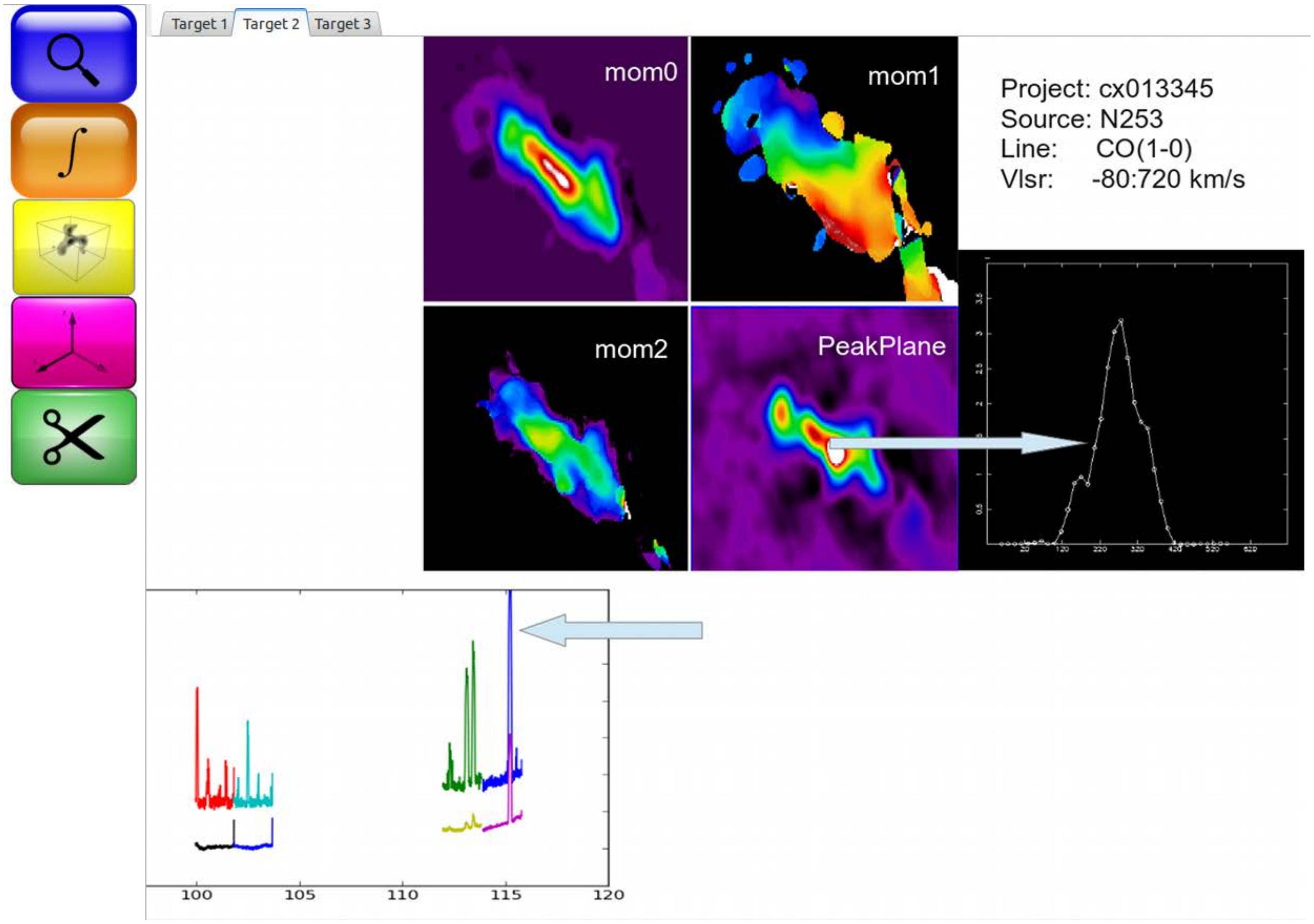


Figure 6. Mockup of the detailed view of some of the data products of a single source and spectral line, as produced by the Data Summary tool for the target source selected in the middle panel. In this scenario, the user has clicked on the CO data cube icon in Figure 5. On the top row, tabs are visible for a number of collected sources. The lower arrow indicates which line is the CO line in the full-band spectra. The upper arrow indicates the reference position for the spectrum shown.

1.1.5 Line Identification and Line Cutout

The Line ID Tool will compare the rest frequency coverage of the image cubes with a database of common line retrieved from Splatalogue to identify the appropriate channel/cube locations of potential lines. A line-strength vs. frequency table is the essential input into a line identification procedure (here we will leverage Tony Remijan's datasplat project). The input data for the procedure will be improved by the previously computed per-channel cube statistics and aided by cross-correlation techniques in a position-velocity diagram (Figure 7) or in the cube itself. The output will be a series of identified and unidentified lines across the different bands in a simple table of line, frequency, channel range, and detection probability.

Based on the line identification line tables, line cutout cubes can be extracted, with now the third axis in Doppler velocity space, to ensure that we can compare the different lines and molecules on the same basis. Depending on the number of lines in the bands, keeping only the line cubes can significantly cut down user disk usage.

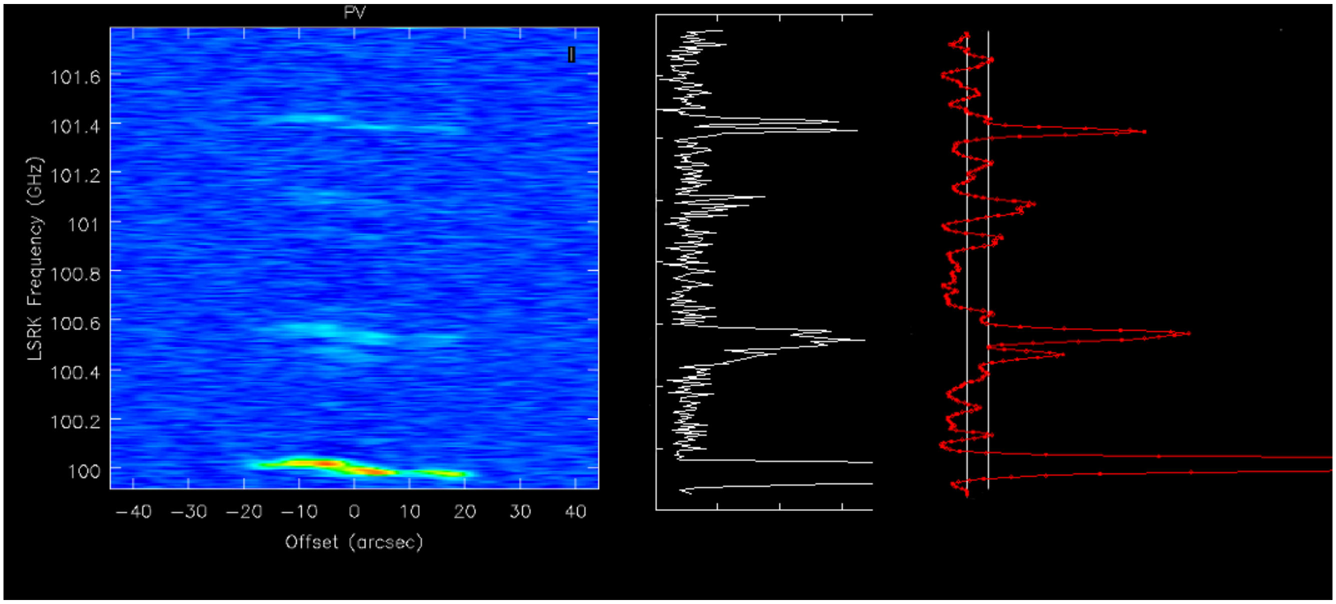


Figure 7. Line identification can be tremendously improved in a position-velocity diagram by cross-correlating the signal in the velocity direction. Compare the more traditional noisy peak/rms plot in the middle (in white) with the smooth cross-correlation technique on the right (in red). The lines are better resolved using the cross-correlation technique. The two vertical lines denote zero and a 3 sigma clipping, above which we deem the line significant.

1.1.6 Moment Maps

Moment maps are a simple, yet powerful tool for examining global properties of an object's emission. The most commonly used maps are integrated flux (moment 0), centroid velocity (moment 1), and velocity dispersion (moment 2). The Moment Map tool takes the line cubes produced by the Line ID Tool and creates the three moment maps for each spectral line (see Figure 6). The moment can be clipped based on simple RMS cutoff (already computed by the Data Summary tool!), but also could employ more sophisticated methods such as local smoothing to increase the signal to noise to create a more robust clipping mask. Moment maps can also be produced using parametric shapes, e.g. a Gaussian fit, which would then store the total emission, mean velocity and width of the spectrum at each location.

1.1.7 Overlap Integral

The overlap integral provides an overview of the emission properties across all spectral lines. For each detected line, one assigns a bit in an integer, and then logically ORs them across all lines, creating a bit-mask map of the emission present at this location. One can do this in a map, or even in a cube, to determine in which spatial or velocity regions certain lines are present or absent in relation to other lines. An example using NGC 253 spectral line data from ALMA Cycle 0 is shown in Figure 8. This technique can also be applied to individual channels of a single spectral line to examine the velocity pattern of emission, similar to a renzogram.

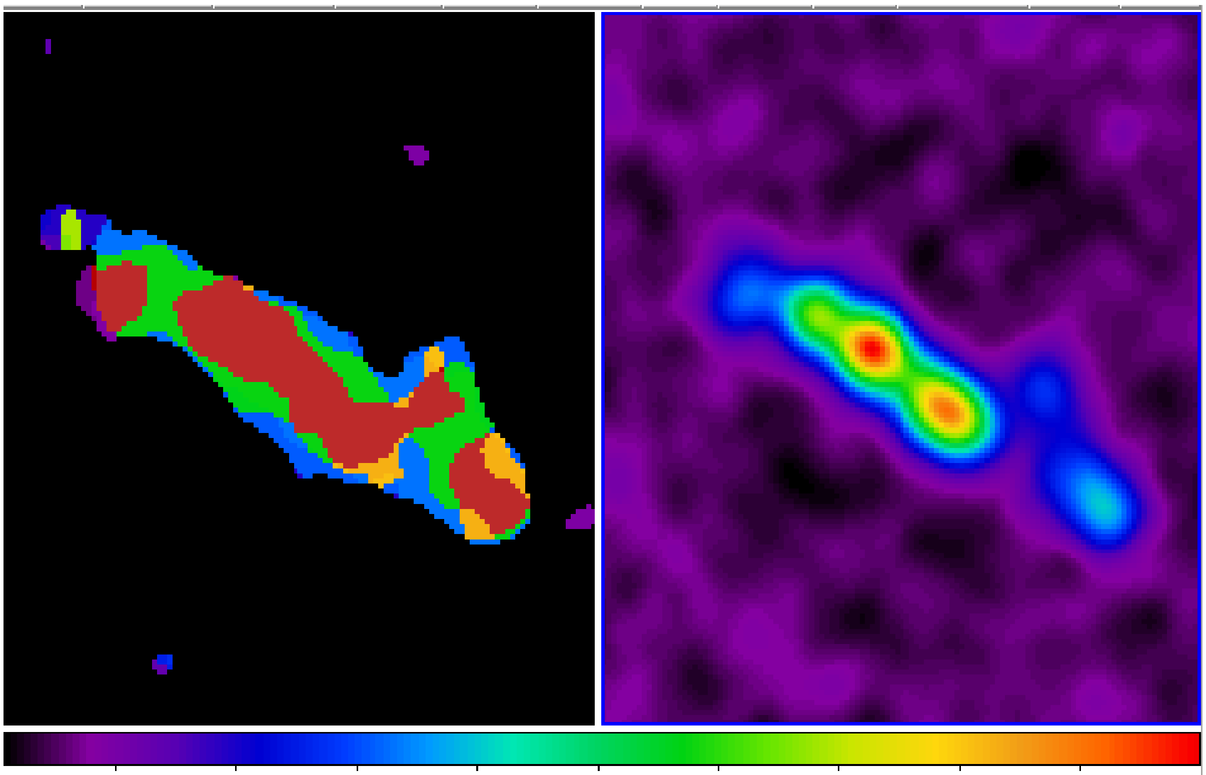


Figure 8. An example of an overlap integral using ALMA Cycle 0 data of NGC 253 [6]. Right: Integrated flux (moment 0) of CO(1-0). Left: The combined overlap integral - valued 0 to 255 - of 8 spectral lines, each using a single bit in an 8 bit integer. As can be seen, most of the peaks have all 8 lines present. The spectral lines used to create this overlap integral are CO, C17O, CN(F=1), CN(F=2), HC3N, CH3OH, H2CS(F=1), and CH3C2H

1.1.8 Descriptor Vectors

To enable the creation of an interactive visual analysis, reasoning, and discovery environment, new methodologies are needed for effective visual presentation of salient features and their characteristics in large datasets. Saliency in an image is based on segmenting the pixels in different ways and assigning an n-dimensional description vector to the features found. This provides an application-independent, purely mathematical measurement of information [2,3]. For astronomical images, description vectors can be chosen from any number of properties of the emission. An example in vision research utilizes a color histogram, which enables a very fast cataloging of feature types in an image, as well as finding the features similar to a selected one, a type of machine learning [4,5]. However, it is important to note that although this may sound similar to Principal Component Analysis, the descriptive vectors do not need to be orthogonal or even have the same units. For this same reason, data mining the saliency of sources can be an extremely powerful technique, allowing a faster comparison of features that might be missed otherwise. The comparison of sources is measured by the distance between vectors of those objects, the closer the more similar and the farther the more dissimilar. During our study, we evaluated this technique's potential for achieving science goals. The key to success is discovering the best description vector for the specific science case [2]. For example, in one case the descriptive vector that best describes a specific science goal may be the properties of the certain molecular lines and the overlap interval, while another may be the location of the molecular peaks with respect to the continuum. ADMIT will provide the infrastructure to write various descriptor vectors, provide some of the vectors (based on BDPs) that will most likely to cover standard science goals, and provide a task that determines best vector choices based on user inputs. We will also document the descriptor formulation procedure so that ALMA has the option to build on this approach in the future.

1.1.10 Example Use Case

In this use case scenario, the user has already downloaded a number of projects from the ALMA archive, and all have the basic ADMIT data included. The hypothesis to be checked is if the line width in CO correlates with the presence of a selected number of molecular lines, in the sense that for wider profiles some of the molecules would be destroyed, and thus the overlap integral would not achieve the maximum value (where all molecules are present). In this case the user selected CO as the reference line, and $C^{17}O$, CS, and $H_{13}CN$ as probes. In this example the user is more interested in a global value per object, than a point by point comparison in the map.

The code loops over all projects, and first ensures that all lines are represented. It estimates the line width from the PeakProfile (precomputed by ADMIT) in the CO line cube as a representative value for the line width for the object. This particular user is conservative and wants to recompute the integrated flux (moment 0) maps with a 3-sigma clip. These new moment maps are then turned into an overlap map (one of the ADMIT tools) and via some simple NumPy math the fraction is computed from number of pixels with value 15 (where all lines are present) divided by the number of pixels with any signal in any of the lines (overlap map value more than 0). The line width is then plotted against this fraction, with the expectation that it starts near 1 and then suddenly drops at some line width. In the example code is shown below, one can see the power of ADMIT: in just a few lines of simple Python, a scientific hypothesis can be created and examined based on ALMA images.

```
import admin, math                              # grab some needed python modules
import numpy as np                              #

adm = admit.ADMIT()                             # initialize ADMIT

projects = adm.query_dir('.')                   # look for ADMIT projects here

lines = ['co', 'c17o', 'cs', 'h13cn']           # first line is reference line
omax = math.pow(2,len(lines))-1                 # max in overlap map, 15 in this case

s_list = []                                     # accumulate line widths
f_list = []                                     # accumulate fractions

for p in projects:
  adm.setdir(p.dirname)                         # move into the proper project directory
  line_cubes = {}
  for c in p.linecubes:                         # loop over line cubes and grab the line name
    if lines.count(c.line):                     # if we got one of the lines we wanted
       line_cubes[c.line] = c                   # store reference to that line cube

  if len(lines) != len(line_cubes):
    print "Skipping ",p.dirname
    continue

  c_ref   = line_cubes[lines[0]]                # reference to the CO cube
  x       = c_ref.peakprofile('vlsr')           # spectrum X in KM/S
  y       = c_ref.peakprofile('value')          # spectrum Y in JY/BEAM
  x_mean  = (x*y).sum()/y.sum()
  x_width = (x*x*y).sum()/y.sum() - x_mean*x_mean

  s_list.append(x_width)                        # accumulate for later

  m = []                                        # accumulate maps for new overlap
  for l in lines:
    m0 = adm.moment(c,[0],'clip',rms=3*c.rms)  # compute new moment maps
```

```
    m.append(m0)

  o = adm.overlap(p,m)                          # get overlap image
  oval = o.array()                              # get a numpy reference for work
  f_all = len(np.where(oval == omax)[0])
  f_sig = len(np.where(over > 0)[0])

  f_list.append(f_all/f_sig)
  #
#
adm.plot2d(s_list,f_list)                       # scatterplot of accumulated data
```

1.2 Technical Personnel

We have a very experienced technical team at the University of Illinois and University of Maryland. The key technical people are long term employees of the universities. The paragraphs below give more detail on their experience. In addition, our team brings in valuable expertise from NRAO through the co-I's: Jeff Kern is a leader in the CASA group; Adam Leroy is a scientist at the North American ALMA Science Center with strong software interests; and, Anthony Remijan is a scientist at the North American ALMA Science Center who lead the molecular line database effort.

At the University of Illinois, the Laboratory for Astronomical Imaging (LAI) and the National Center for Supercomputing Applications (NCSA) have been partners in leading the techniques for astronomical data archiving. LAI/NCSA has been developing and operating the data archives for the Berkeley Illinois Maryland Association (BIMA) and the Combined Array for Research in Millimeter-wave Astronomy (CARMA) millimeter arrays for the last 20 years. In addition to the archive expertise, Illinois has developed a data pipeline for CARMA that provides near science quality calibrated datasets and images.

Prof. Leslie Looney has over 18 years of interferometric experience with special emphasis on statistical techniques (e.g., Bayesian Monte-Carlo Markov Chain analysis) and multi-wavelength comparisons. He is the Director of the Laboratory for Astronomical Imaging, which has been a leader of astronomical software development. Looney was on sabbatical from July 2012-July 2013 at NRAO in Charlottesville, allowing for a good personal connection between the proposal team and NAASC scientist and Co-Is. Currently, he is also a member of the ALMA North American Science Advisory Committee.

Dr. Friedel has more than 12 years of astrochemistry experience, especially detecting low signal-to-noise molecular line emission using single dish and interferometric techniques. Friedel is the LAI lead of astronomical software with direct experience with the CARMA archive, CARMA data reduction pipeline, interferometric data formats, and data visualization. He is a skilled developer and designer with detailed astronomical applications using Python, C/C++, Java, Fortran, and Perl. He has been involved with this project during the study stage, and has already developed the prototype of the XML write/read scheme.

Lisa Xu has an M.S. degree in Computer Science from the University of Illinois and 8 years of experience in radio astronomy data management as the lead developer for the CARMA archive at the NCSA at Illinois, as well database administration, and operations management for both the CARMA and BIMA archives. She is a skilled developer and designer, with technical proficiency in a range of programming languages. Xu is also proficient is testing and integration methodologies and required documentation standards. She is the prime community contact for the CARMA and BIMA archives.

The final person on the Illinois team has not yet been named, and the exact person will depend on personnel demand at NCSA. (But as the VO program is being ramped down at the NCSA, we expect access to high quality personnel.) We will recruit a professional programmer with experience in astronomical techniques and especially Python who will be able to work closely with the team to fully develop the software, as well as contribute to integration, testing and documentation.

At the University of Maryland, the Laboratory for Millimeter-wave Astronomy (LMA) has been involved in millimeter-wave interferometry and software for 25 years associated with first BIMA and now CARMA. The LMA has a history of involvement in a wide variety of software and hardware projects ranging from the CARMA data analysis package MIRIAD to a wideband correlator for the Green Bank telescope and the FPGA programming for the CARMA's current correlator and the next generation correlator under construction.

Prof. Lee Mundy has nearly 30 years of experience in millimeter wavelength interferometry and a broad background in interferometric techniques. He is the Director of the LMA.

Peter Teuben has over 25 years of experience in software development in general and radio interferometry specifically. He has been maintaining MIRIAD for the past 15 years. He also contributed to the CARMA online system, AIPS++ and CASA, in particular the carmafiller, which converts MIRIAD data into a CASA Measurement Set. In addition, he has experience building and maintaining a number of packages in C, C++, Fortran, and Python. He is one of the three original authors of the NEMO package, which contains a large number of tools to interoperate between N-body Snapshots, Orbits and Images to compare simulations and observations.

Dr. Marc Pound has over 25 years of experience in single dish and interferometric radio astronomy, with scientific emphasis on Milky Way molecular clouds. For the past 8 years, he has been the CARMA software coordinator, prioritizing and overseeing the work of developers at CARMA's partner institutions. He is experienced in software project management and modern software engineering practices from design through implementation and testing. In addition to the management responsibilities, Pound designed and developed some of CARMA's key software subsystems.

Dr. Kevin Rauch's scientific expertise is in the dynamics of nearly-Keplerian systems and in non-Newtonian gravitations problems. His work develops innovative software implementations to solve problems numerically. He joined the LMA in 2001 as a specialist in software development for instrumentation, working on the "Zpectrometer" correlator now deployed at Green Bank. In the past decade, he has played the central role in design and development of FPGA software for the CARMA digital correlators, and the interface of the correlator software to higher level software. Rauch has the deep analytical and mathematical expertise required for innovative algorithm development.

Prof. Amitabh Varshney is the Director of the Institute for Advanced Computer Studies (UMIACS) and Professor of Computer Science at the University of Maryland at College Park. Varshney's research focus is on exploring the applications of graphics and visualization in engineering, science, and medicine through advances in geometry processing, illumination models, perceptual rendering, and high-performance visual computing. He has also contributed towards development of new algorithms for a number of research areas in graphics and visualization including visual saliency, summarization of large visual datasets, automatically generating multi-resolution hierarchies, procedural textures, and rendering with points, images, meshes, and volumes. He is currently exploring several applications in general-purpose high-performance parallel computing using clusters of CPUs and Graphics Processing Units (GPUs).

2.0 PROJECT DELIVERABLES

2.1 Hardware: No deliverable hardware.

2.2 Software:

The following are software products to be delivered:

- ADMIT Python main module
- Test suite and inline documentation for main module
- ADMIT Core Tools
- Test suite and inline documentation for each core tool
- ADMIT Graphical User Interface

These constitute the components needed for the ADMIT infrastructure, BDP pipeline, user display of BDP, and the tools discussed in the science and technical sections.

2.3 Services: No specific deliverables

2.4 Documents:

- Monthly "4-square" Progress Reports
- Software design documentation
  - ADMIT design documents
- Interface Control Documents (ICDs)
- Technical manuals and procedures
  - ADMIT User Guide
  - ADMIT Tool Writer's Guide
- Quality Assurance procedures
  - Software test and acceptance procedures
- Final Report
- Outcomes Report

3.0 INTERFACES TO ALMA

The ADMIT package will be an add-on to CASA working under its Python environment. We will interface to existing CASA tasks through XML parameters files. We will design our XML and image files to be compatible with CASA Viewer so that there is an option for its use with ADMIT data products. In a couple of cases, we may create new CASA tasks. These tasks would follow CASA programming and testing requirements; the tasks would not require any changes to the existing CASA package.

ADMIT will run on the ALMA pipeline imaging system. It would run on a project basis as imaging cubes are approved in the ALMA system and readied for the ALMA archive. The ADMIT scripts operate on the image cubes and do not require interaction with the ALMA pipeline.

The ADMIT Basic Data Products would be ingested into the ALMA archive along with the data cubes in the project directory. There is one ADMIT tarball for each project directory at the time of ingestion. We have discussed this approach with the ESO ALMA Archive Scientist, Felix Stoehr and the NRAO ALMA Archive User Support person, Mark Lacy; they see no problem with ingesting and serving our tarball from the ALMA Archive.

The ADMIT package would be distributed as an add-on package to CASA. Jeff Kern, co-I on this proposal, finds this acceptable.

4.0 SITE LOCATION IMPACT STATEMENT *(if applicable)*

We anticipate no significant impact on existing facilities. The proposed ADMIT procedures to create the BDP would run at ALMA facilities on the ALMA pipeline image cubes. We will target ADMIT procedures to take no more than 5% of the cpu time required for the pipeline to process corresponding data cube. The ADMIT Basic Data Products would be stored in and served from the ALMA data archive; the BDP for a project is expected to be 5-40MB, which less than 1% of the typical project data size. Similarly due to the small size of BDP, there is minimal additional load on the archive server.

The ADMIT package is proposed to be an add-on package to CASA distributions.

5.0 PERIOD OF PERFORMANCE

We request a performance period of 24 months, from October 4, 2013 to September 30, 2015.

6.0 STAFFING

The effort for this project will include a number of key personnel with experience with radio interferometry data and images from years of involvement with the Combined Array for Research in Millimeter-wave Astronomy (CARMA). The PI, Lee Mundy, just stepped down as the Director of CARMA and remains Director of the Laboratory for Millimeter-wave Astronomy at University of Maryland. The lead scientist, Leslie Looney, has over 10 years of experience with CARMA and is the Director of the Laboratory for Astronomical Imaging at University of Illinois. The software lead, Peter Teuben, has over 25 years of radio interferometry and software development experience; he is the lead scientist for the CARMA data analysis software and has extensive experience with scripting and programming, and supports the CARMA interface to CASA. Marc Pound, the management lead, has been the CARMA software coordinator for the past 8 years and has extensive experience with radio interferometry and real time array control systems. Doug Friedel and Lisa Xu who will be contributing to the XML and pipeline programming have extensive experience from creating and operating the CARMA archive and imaging pipeline. Kevin Rauch who will be contributing to the algorithms, is an experienced in implementing mathematical algorithms for astronomical applications.

Amitabh Varshney is the Director of the University of Maryland Institute for Advanced Computer Studies and a computer scientist within interests in exploring the applications of graphics and visualization in engineering, science, and medicine. Jeff Kern is the project manager for CASA, and Anthony Remijan leads the molecular line database effort (Splatalogue).

The PI and science lead will provide leadership of the science aspects of the project but the majority of our personnel have scientific expertise relevant to the project and are expected to contribute to the scientific direction as part of their software duties. The software development will be led by Peter Teuben; Doug Friedel, Lisa Xu, Kevin Rauch, and the new hire near the end of year 1 will be contribute to the software development. Katherine Lee (a postdoc starting in the fall at UMD) and the new hire at Illinois will have primary responsibility for user and technical documentation , respectively.

6.1 Offerer's Staffing:

*Table 1.0; **Offerer's Labor Estimate**.*

Title	Key Personnel *(leave blank if inapplicable)*	FTE	Duration (Months)
Principal Investigator	Lee Mundy	0.14	24
Scientific Lead	Leslie Looney	0.14	24
Software Lead	Peter Teuben	0.70	24
Science		0.4	24
Software Development		2.16	24
Management	Marc Pound	0.40	24
Test and Document		0.75	18
	TOTAL FTEs	**4.69**	

6.2 External Staffing:

Complete Table 2.0, below.

*Table 2.0; **External Staffing and Contact Information**.*

Title	Name	Institution	Email	Telephone
Co-Investigator	Jeff Kern	NRAO	jkern@nrao.edu	575 535-7391
Co-Investigator	Adam Leroy	NRAO	aleroy@nrao.edu	434 244-6807
Co-investigator	Tony Remijan	NRAO	aremijan@nrao.edu	434 244-6848

7.0 COST BREAKDOWN

BUDGET for University of Maryland and Subcontracts

	Year 1		Year 2		TOTAL
PERSONNEL	Effort	Amount	Effort	Amount	
PI, Lee Mundy	8.33%	[redacted]	8.33%	[redacted]	[redacted]
Peter Teuben	25.00%	[redacted]	50.00%	[redacted]	[redacted]
Kevin Rauch	0.00%	$0	50.00%	[redacted]	[redacted]
Marc Pound	20.00%	[redacted]7	30.00%	[redacted]	[redacted]
Postdoc	15.00%	[redacted]	30.00%	[redacted]	[redacted]
FRINGES		$18,850		$46,203	$65,053
Salary+Fringes		**$88,354**		**$207,478**	**$295,832**
TRAVEL					
Domestic		$10,236		$12,725	$22,961
Foreign		$0		$0	$0
OTHER DIRECT COSTS					
Sub to U. Illinois		$117,075		$293,386	$410,461
Subcontract to NRAO		$24,959		$24,959	$49,917
Materials & Supplies		$0		$0	$0
Computer Services		$1,173		$2,602	$3,776
Publication Costs		$0		$0	$0
TOTAL DIRECT COSTS		**$247,407**		**$550,017**	**$797,424**
F& A					
Direct Costs Base		$149,763		$222,805	$372,568
On campus rate		$77,877		$115,859	$193,736
Total		**$325,284**		**$665,876**	**$991,160**

Budget for University of Illinois Subcontract:

Proposed Budget
University of Illinois at Urbana-Champaign

	Man-Months Per Period	Year 1 10/1/13 9/30/14	Year 2 10/1/14 9/30/15
Salaries and Wages			
Institutional PI/Looney	1.0 / 1.0	[redacted]	
Research Programmer/Friedel	2.4 / 2.4		
Research Programmer/Xu	2.4 / 4.8		
Research Programmer/TBD	1.2 / 12.0		
Fringe Benefits			
42.94% of salaries		18,721	52,694
Total Salaries and Benefits		**$62,318**	**$175,410**
Equipment			
None		0	0
		$0	**$0**
Travel			
Collaborative visits to UMD (4 people)		8,000	0
Domestic Conference (4 people)		0	7,000
		$8,000	**$7,000**
Other			
Consumable Research Supplies		1,000	1,030
Information Technology Supplies		1,000	0
IT Support Services		1,500	1,545
		$3,500	**$2,575**
Total Direct Costs		**$73,818**	**$184,985**
Indirect Costs --			
58.6% of Modified Total Direct Costs		**$43,257**	**$108,401**
TOTAL AMOUNT REQUESTED PER YEAR		**$117,075**	**$293,386**
TOTAL AMOUNT REQUESTED			**$410,461**

.

Budget for NRAO subcontract:

NRAO Budget for ADMIT

	FY2014	FY 2015 2%	2-year cost
New Hire 7500*2	7,500.00	$7,500	$15,000
Jeff Kern Labor ($43*2080)	[redacted]		
Leave pool 20%			
Benefits 35%			
CCR 30%			
AUI IDC 4.84%			
AUI Management Fee 2%			
8% of Jeff Kern			
CCR+IDC+Mgnt on new hire			
Total Cost	$30,569	$33,826	$64,395

7.1 Offerer's Cost:

*Table 3.0; **Offerer's Cost Breakdown**.*

WBS No.	Task Description	Labor ($)	Materials & Services ($)	Travel ($)
1.0	Admit Pipeline	51250	14910	
2.0	Design XML tags and structure	1000		
3.0	Admit tool architecture	65000		
4.0	XML reader/writer	18750		
5.0	Image statistics	25000		
6.0	Metadata summarizer	35000		
7.0	Moment Module	30000		
8.0	Basic graphical user interface	70000		
9.0	Advanced GUI	100000		
10.0	Python CASA documentation	40000		
11.0	User Guide	40000		
12.0	Tool Writer's Guide	40000		
13.0	Basic Line Catalog query	15000		
14.0	Splatalogue Query	60000		
15.0	Core ADMIT tools	122500		
16.0	Archive Query	60000		
17.0	Multi-source operations	50000		

18.0	Team collaboration visit 2014	12500		8000
19.0	AAS 2014 Meeting	750		5000
20.0	Developer visit 2014	8000		10000
21.0	Team collaboration visit 2015	5000		10000
22.0	AAS 2015 Meeting	2750		9000
23.0	Developer Visit 2015	10750		14000
24.0	Pilot GUI Release	5000		
25.0	Integration end-to-end testing	5000		
26.0	Archive use case review	20000		
27.0	Closeout Report	18000		
	SubTotals ($)	**920250**	**14,910**	**56,000**
			TOTAL COST ($)	**991,160**

7.2 Collaborating Institution / Subcontractor Cost:

Table 4.0; ***Collaborating Institution / Subcontractor Cost****.*

Collaborating Institution / Subcontractor	***USD ($)***	***In-Kind Contribution Value in USD ($)***
University of Illinois	$410,461	
NRAO	$64,395	
TOTAL COST ($)	**$474,856**	
TOTAL VALUE of IN-KIND CONTRIBUTIONS ($)		***0.00***

7.3 Total Cost:

Table 5.0; ***Total Project Cost.***

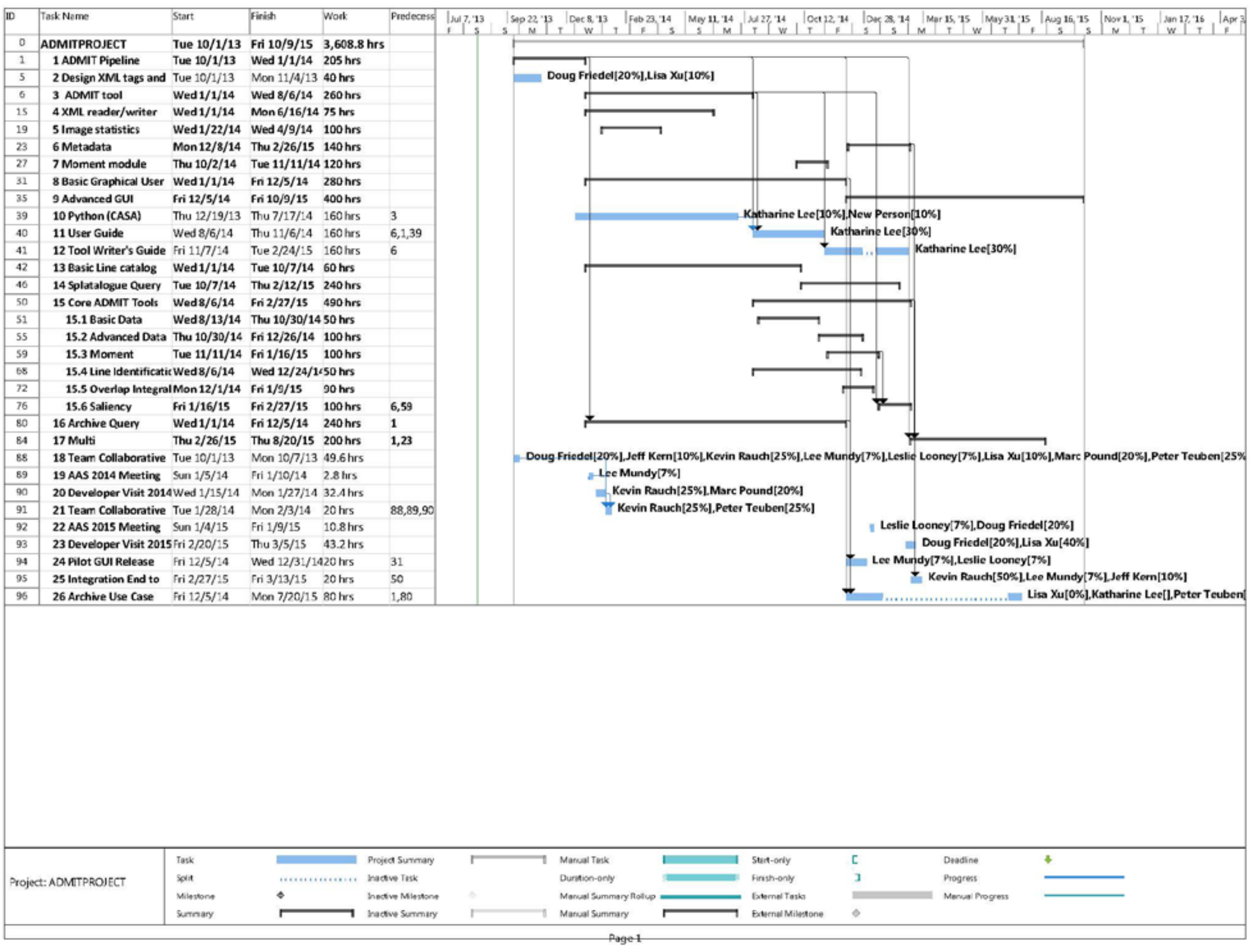
Project Participants	***USD ($)***
Offerer's Total Cost *(Table 3.0)*	$991,160
Collaborating Institution / Contractor Total Cost *(Table 4.0)*	$0
Expected contribution from the NA ALMA Development Program	
Total Cost of Project ($)	**$991,160**

7.4 Total Value:

Table 6.0; ***Total Project Value.***

Category	*USD ($)*
Total Cost of Project *(Table 5.0)*	$991,160
Total Value of In-Kind Contributions *(Table 4.0)*	
Total Value of Project ($)	**$991,160**

8.0 SCHEDULE



ID	Task Name	Start	Finish	Work	Predecess
0	ADMITPROJECT	Tue 10/1/13	Fri 10/9/15	3,608.8 hrs	
1	1 ADMIT Pipeline	Tue 10/1/13	Wed 1/1/14	205 hrs	
5	2 Design XML tags and	Tue 10/1/13	Mon 11/4/13	40 hrs	
6	3 ADMIT tool	Wed 1/1/14	Wed 8/6/14	260 hrs	
15	4 XML reader/writer	Wed 1/1/14	Mon 6/16/14	75 hrs	
19	5 Image statistics	Wed 1/22/14	Wed 4/9/14	100 hrs	
23	6 Metadata	Mon 12/8/14	Thu 2/26/15	140 hrs	
27	7 Moment module	Thu 10/2/14	Tue 11/11/14	120 hrs	
31	8 Basic Graphical User	Wed 1/1/14	Fri 12/5/14	280 hrs	
35	9 Advanced GUI	Fri 12/5/14	Fri 10/9/15	400 hrs	
39	10 Python (CASA)	Thu 12/19/13	Thu 7/17/14	160 hrs	3
40	11 User Guide	Wed 8/6/14	Thu 11/6/14	160 hrs	6,1,39
41	12 Tool Writer's Guide	Fri 11/7/14	Tue 2/24/15	160 hrs	6
42	13 Basic Line catalog	Wed 1/1/14	Tue 10/7/14	60 hrs	
46	14 Splatalogue Query	Tue 10/7/14	Thu 2/12/15	240 hrs	
50	15 Core ADMIT Tools	Wed 8/6/14	Fri 2/27/15	490 hrs	
51	15.1 Basic Data	Wed 8/13/14	Thu 10/30/14	50 hrs	
55	15.2 Advanced Data	Thu 10/30/14	Fri 12/26/14	100 hrs	
59	15.3 Moment	Tue 11/11/14	Fri 1/16/15	100 hrs	
68	15.4 Line Identificatic	Wed 8/6/14	Wed 12/24/1	50 hrs	
72	15.5 Overlap Integral	Mon 12/1/14	Fri 1/9/15	90 hrs	
76	15.6 Saliency	Fri 1/16/15	Fri 2/27/15	100 hrs	6,59
80	16 Archive Query	Wed 1/1/14	Fri 12/5/14	240 hrs	1
84	17 Multi	Thu 2/26/15	Thu 8/20/15	200 hrs	1,23
88	18 Team Collaborative	Tue 10/1/13	Mon 10/7/13	49.6 hrs	
89	19 AAS 2014 Meeting	Sun 1/5/14	Fri 1/10/14	2.8 hrs	
90	20 Developer Visit 2014	Wed 1/15/14	Mon 1/27/14	32.4 hrs	
91	21 Team Collaborative	Tue 1/28/14	Mon 2/3/14	20 hrs	88,89,90
92	22 AAS 2015 Meeting	Sun 1/4/15	Fri 1/9/15	10.8 hrs	
93	23 Developer Visit 2015	Fri 2/20/15	Thu 3/5/15	43.2 hrs	
94	24 Pilot GUI Release	Fri 12/5/14	Wed 12/31/14	20 hrs	31
95	25 Integration End to	Fri 2/27/15	Fri 3/13/15	20 hrs	50
96	26 Archive Use Case	Fri 12/5/14	Mon 7/20/15	80 hrs	1,80

Project Schedule.

9.0 PROJECT MANAGEMENT

9.1 Systems/Configuration Control

9.1.1 Systems Requirement and Specification Control

Development engineering and design activities will be conducted in accord with established ALMA Systems Engineering policies, practices and procedures.

9.1.2 Documentation Control

All shared documents will be dated and bear a revision level number.

9.1.3 Product & Quality Assurance Control

Development engineering and design activities will be conducted in accord with established ALMA PA/QA policies, practices and procedures. A unique Product Assurance Plan is unnecessary.

9.2 Performance to Schedule

The Principal Investigator has primary responsibility for schedule development and performance to schedule. The NA ALMA Development Program office will provide support to the PI in establishment of a revision-controlled Project schedule and monthly preparation of performance to schedule status. In the event of a schedule variance, the PI and the NA ALMA Development Program Manager will assess the impact and develop the appropriate recovery action(s).

9.3 Performance to Budget

The Principal Investigator has primary responsibility for intra-project budget allocation and cost performance. The NA ALMA Development Program office will provide support to the PI in establishment of cost accounts, budget load, and the preparation of a revision-controlled, monthly Budget Status Report. In the event of a cost variance, the PI and the NA ALMA Development Program Manager will assess the impact and develop the appropriate recovery action(s).

9.4 Measures of Success

The ADMIT measures of success will be:

(1) benchmark Basic Data Products (BDP) on test maps with different observational configurations bands, and correlator setups with 100% repeatability and accuracy,
(2) successful creation of complete BDP for >95% of ALMA projects with graceful failure modes that provide available data products.
(3) Create BDP from pipeline maps using <45 min of cpu time for > 50% of projects,
(4) delivery of ADMIT metadata XML files for archive ingestion for > 95% of projects
(5) user survey of pilot ADMIT GUI release to have a 75% approval for usability.

We will track the performance metrics for items 2, 3, and 4 by monitoring the statistics as the modules are tested and enter utilization. The final measures of success will be the statistics during the last month of the proposal performance period. Item 5 will be measured with a user survey in Quarter 1 of 2015 and a follow-up survey in Quarter 3 of 2015.

These five measures will be used to assess risk to the project (see the Risk Assessment in the Proposal Form for the foreseeable risks) and to prioritize new risks as they develop to mitigate any issues.

9.5 Risk Management

The designed implementation of ADMIT provides robust mitigation of most risks. As ADMIT relies on ALMA pipeline image cubes, the largest risk is delay in the ALMA pipeline full-band imaging products (see Table 7.0). The ALMA project goals include production of full maps of all of the bands so this is only a question of timing. Significant pipeline delay is a low risk during the duration of our project; the primary mitigation is to process whatever pipeline products are available which will decrease the value of the ADMIT BDP in the archive. One of the key strengths of ADMIT is the ability of the user to run the entire package on their local machines. A user can make their own images with their own parameters and run the ADMIT analysis on their image cubes to recoup the full analysis. This is a second mitigation path.

A second risk is that the archive user interface will not be upgraded to allow for easy identification of the ALMA image cubes for each band in the archive. The baseline plan is to do this identification within the ADMIT XML files; if this is not possible, the risk is mitigated by the user manually matching the appropriate archived files with the ADMIT products. We will stay in contact with the archive team over the period of this proposal to create the best possible solution.

Finally, the last risk is that some desirable analysis will be too CPU extensive for the computer resources available at the ARCs. Again, all tools will be included in the ADMIT package and the users can run any of the analysis tools on their own machines. This risk should also be reduced as the observatory goes through refresh cycles with the pipeline computer hardware.

Note that in all cases, the risks are completely retired without impact to cost.

Table 7.0; ***Project Risk Assessment.***

No.	Primary Risk(s)	Prob. (%)	Impact ($)	Mitigation
1	ALMA pipeline delayed in producing full imaged cubes	Low	None	ADMIT will be designed with full (all bands and complete channels) image cube expectations but the system will work on whatever image cubes are available. In the worst case, the user can create their own cubes and execute ADMIT locally.
2	Planned upgrades to archive user interface are delayed making identification of image cubes difficult	Low	None	User can manually identify appropriate image cubes in archive
3	Some tools require more computer resources than available at ALMA pipeline infrastructure	Low	None	User can run all tools on local machines
	Total Cost Impact ($)		**0.00**	

9.6 Communication Plan and Progress Reporting

A monthly, "Four-square" progress report will be prepared by the Principal Investigator in accord with Observatory Program Management practices and procedures. Informal reviews will be conducted by the NA ALMA Development Program Manager upon the completion of Level I milestones.

10.0 PROJECT CLOSEOUT

Upon conclusion of this Project, the NA ALMA Development Program Office will coordinate the orderly closeout of activities; or, the transition of activities to implementation. At a minimum, this will include the following:

- verification of compliance with established procurement policies and procedures;
- verification of Purchase Order final payments;
- cost and schedule variance analysis;
- preparation of a Final Report;
- preparation of a Project Outcomes Report; and
- archiving of Project records

REFERENCES

[1] White et al. 2010, The High Impact of Astronomical Data Archives, ASTRO2010 position paper. http://archive.stsci.edu/hst/bibliography/pubstat.html

[2] M. Chen and H. Janicke. An information-theoretic framework for visualization. IEEE Transactions on Visualization and Computer Graphics, 16(6):1206 –1215, 2010.

[3] H. Janicke and M. Chen. A salience-based quality metric for visualization. Computer Graphics Forum, 29(3):1183–1192, 2010.

[4] C. Y. Ip and A. Varshney. Saliency-assisted navigation of very large landscape images. IEEE Transactions on Visualization and Computer Graphics, 17(12):1737– 1746, 2011.

[5] C.Y. Ip, Ph.D. Thesis, University of Maryland, 2014.

[6] A. Bolatto et al., Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind, Nature, 499, 450-453, 2013.

APPENDIX A - REFERENCE DOCUMENTS

APPENDIX B – CURRICULUM VITAE OF KEY PERSONNEL

Biographical Sketch – August 2013

Jeffrey S. Kern (jkern@nrao.edu)
(575) 835-7391
1003 Lopezville Road
Socorro NM 87801-0387

Professional Preparation

June, 1995 B.A., Astronomy and Mathematics, Colgate University, Hamilton, NY.
June, 1998 M.S., Physics, University of Vermont, Burlington, VT.
June, 2004 Ph.D., Astrophysics, New Mexico Institute of Mining and Technology, Socorro NM.

Appointments

2010-	Common Astronomy Software Applications (CASA) Group Lead, National Radio Astronomy Observatory (NRAO)
2009-2010	ALMA Computing Control and Correlator Subsystem Lead, NRAO
2009 –	ALMA Computing North American Deputy Lead, NRAO
2008-2010	ALMA Computing Control Subsystem Lead, NRAO
2006-2008	ALMA Test Facility Software Team Lead, NRAO
2005-2006	Software Developer ALMA Control Software Team, NRAO

Relevant Publications

Extending the 3D capabilities of the CASA viewer, M. Kümmel, H. Kuntschner, D. Schiebel, M. Westmoquette, D. Petry, D. Mehringer, P. Ballester, J. Kern, 2011, ADASS XXI Proceedings, in publication.

Integration and Testing in the ALMA Control Subsystem, Hiriart, R.; Kern, J.; 2009, Astronomical Data Analysis Software and Systems XVIII, ASP Conference Series, **411**, 414.

Additional Significant Publications

An ALMA Beamformer for VLBI and Phased Array Science, Doeleman, S.; Webber, J.; Lacasse, R.; Inoue, M.; Alef, W.; Honma, M.; Nagar, N.; Baudry, A.; Kern, J.; Crew, G.; Cappallo, R.; 2011 ,Proceedings of the XXXth URSI General Assembly.

The Brightest Pulses in the Universe: Multifrequency Observations of the Crab Pulsar's Giant Pulses, Cordes, J.M.; Bhat, N.D.R.; Hankins, T.H.; McLaughlin, M.A.; Kern, J.; 2004, Ap.J., **612**, 375-388.

Nanosecond radio bursts from strong plasma turbulence in the Crab pulsar, Hankins, T.H.; Kern, J.S.; Weatherall, J.C.; Eilek, J.A.; 2003, Nature, **422**:6928, 141-143.

Synergistic Activities

Event Horizon Telescope: Participating in international collaboration to enable millimeter VLBI using the ALMA telescope.

Professional Interests

Software development for the efficient acquisition, reduction, and analysis of scientific data. Incorporation of modern software engineering practices with scientific reduction systems. Digital signal processing of astronomical signals, particularly as applied to short lived phenomena.

CV for Adam K. Leroy

Education

Harvard University – B.A. in Astronomy & Astrophysics, Magna Cum Laude 1999
University of California at Berkeley – M.A. in Astrophysics, May 2002
University of California at Berkeley – Ph. D. in Astrophysics, December 2006

Employment

2011 – Present	Assistant Astronomer at NRAO North American ALMA Science Center
2009 – 2011	Hubble Fellow at NRAO, Charlottesville
2006 – 2009	Postdoctoral Researcher, Max Planck Institute for Astronomy

Sample Publications

"Molecular Gas and Star Formation in Nearby Disk Galaxies" Leroy, Adam K.; Walter, Fabian; Sandstrom, Karin; Schruba, Andreas; Munoz-Mateos, Juan-Carlos; Bigiel, Frank; Bolatto, Alberto; Brinks, Elias; de Blok, W. J. G.; Meidt, Sharon; Rix, Hans-Walter; Rosolowsky, Erik; Schinnerer, Eva; Schuster, Karl-Friedrich; Usero, Antonio, AJ, 147, 2

"The CO-to-H_2 Conversion Factor From Infrared Dust Emission Across the Local Group" Leroy, A. K., Bolatto, A., Gordon, K., Sandstrom, K., Gratier, P., Rosolowsky, E., Engelbracht, C., Mizuno, N., Corbelli, E., Fukui, Y., Kawamura, A., 2011, ApJ, 737, 12

"The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively" Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., Thornley, M. D., 2008, AJ, 136, 2782

"The Resolved Properties of Extragalactic Giant Molecular Clouds" Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., Blitz, L., 2008, ApJ, 686, 948

"Bias-free Measurement of Giant Molecular Cloud Properties" Rosolowsky, E. & Leroy, A., 2006, Publications of the Astronomical Society of the Pacific, 118, 590-610.

Synergistic Activities

I am an NRAO Assistant Astronomer with functional duties supporting North American ALMA science as part of the North American ALMA Science Center (NAASC). In that context I have expertise in both the workings of the ALMA observatory and ALMA data. I led our efforts to teach ALMA/CASA data reduction to Cycle 0 PIs via data reduction workshops and the Synthesis Imaging school. I serve as contact scientist for a diverse set of projects and have been heavily involved in data reduction and CASA testing. I have also worked heavily with the CASA team on goals and testing of viewer and image analysis capabilities. This range gives me a good interface with the NAASC activities directly relevant to this proposal.

My scientific focus is studying the processes that lead to star formation and shape the interstellar medium in nearby galaxies. Within this area, my approach is to generate cutting-edge radio and millimeter data sets and then to exploit these using a broad suite of multiwavelength data. To that end I have been an active user of a wide set of radio/millimeter facilities: the IRAM 30-m, the IRAM PdBI, CARMA (and earlier BIMA), the VLA and EVLA, and now ALMA. Exploiting these data has involved the development of novel analysis techniques, including: the identification of giant molecular clouds and extraction of their properties from CO data cubes (Rosolowsky & Leroy 2006, Blitz et al. 2007, Bolatto, Leroy et al. 2008); searches for faint molecular emission from dwarf galaxies (Leroy et al. 2005, 2006, 2007); the implementation of novel molecular-line stacking techniques using HI priors (Schruba, Leroy et al. 2011, 2012). I work with molecular and atomic line, radio continuum (Leroy et al. 2011), and dust continuum (Leroy et al. 2007, 2009, 2011) data in a wide variety of systems and scales, experience that is key to understanding the analysis challenges facing ALMA.

Graduate and Postdoctoral Advisors

Leo Blitz (UC Berkeley) and Fabian Walter (MPIA)

Curriculum Vitae
Leslie W. Looney (lwl@illinois.edu)

Professional Preparation

May, 1992	B. S., Electrical Engineering, with honors, University of South Florida
May, 1992	B. S., Physics, with honors, University of South Florida
Dec., 1998	Ph.D., Physics, University of Maryland

Appointments

2010-	Director of the Laboratory of Astronomical Imaging, University of Illinois
2009-	Associate Professor, University of Illinois
2012-2013	Sabbatical Visitor at NAASC, NRAO
2002-2009	Assistant Professor, University of Illinois
2001-2002	Project Scientist FIFI LS, Staff, Max-Planck-Institute (MPE)
1998-2001	Project Scientist FIFI LS, Fellow, MPE
1993-1998	Research Assistant, University of Maryland
1992-1993	Research Assistant, Goddard Space Flight Center, NASA
1989-1991	System Engineer, Kennedy Space Center, NASA

Expertise

Interferometry: 15 years of experience in both hardware and software

ALMA: Member of the ALMA North American Science Advisory Committee and sabbatical visitor to NASC for 1 year (July 2012-July 2013)

Far-Infrared: Development of a far-infrared spectrometer for SOFIA

Star Formation Science: Multi-wavelength observations of young stars (both low and high mass) and their environments; model fitting of interferometric data to models using multiple statistical techniques

Director of the Laboratory of Astronomical Imaging at Illinois, which leads archiving and data pipeline development for CARMA

Select Recent Publications

"Earliest Stages of Protocluster Formation: Substructure and Kinematics of Starless Cores in Orion", Lee, K., Looney, L.W., Schnee, S., & Li, Z.-Y. 2013, ApJ, 772, 100

"The Magnetic Field Morphology of the Class 0 Protostar L1157-mm", Stephens, I., Looney, L.W., Kwon, W., Hull, C.L.H., et al. 2013, ApJL, 769, 15

"A 0.2-solar-mass protostar with a Keplerian disk in the very young L1527 IRS system", Tobin, J.J., Chiang, H.-F., Wilner, D.J., Looney, L.W., Loinard, L., Calvet, N., & D'Alessio, P. 2012, Nature, 492, 83

"The Envelope and Embedded Disk around the Class 0 Protostar L1157-mm: Dual-wavelength Interferometric Observations and Modeling", Chiang, H.-F., Looney, L.W., & Tobin, J.J. 2012, 756, 168

"Resolving the Circumstellar Disk of HL Tauri at Millimeter Wavelengths", Kwon, W., Looney, L.W., & Mundy, L.G. 2011, ApJ, 741, 3

Biographical Sketch – August 2013

Lee G. Mundy (lgm@astro.umd.edu)

Professional Preparation

June, 1977	B. S., Astronomy, California Institution of Technology
Dec, 1984	Ph.D., Astronomy, University of Texas, Austin

Appointments

2007-	Director of Laboratory for Millimeter-wave Astronomy
2006-	Director of Cooperative Agreement for Research in Exploration Space Science and Technology
2002-2007	Chair of Astronomy Department, University of Maryland
1999-	Professor, University of Maryland
1993-1999	Associate Professor, University of Maryland
1988-1993	Assistant Professor, University of Maryland
1987-1988	Senior Research Fellow, California Institute of Technology
1984-1987	Research Fellow, California Institute of Technology

Areas of Expertise

Star formation, interferometry techniques, molecular cloud structure, radiative transfer, molecular excitation, infrared observations.

Current Ph.D. Students

Shaye Storm (anticipated Ph.D. 2015), Maxime Rizzo (anticipated Ph.D. 2016), Arnab Dhabal (anticipated Ph.D. 2017)

Selected Recent Publications

"Misalignment of Magnetic Fields and Outflows in Protostellar Cores", Hull, C.L.H., and 24 authors including L. Mundy, 2013, Ap.J., 768, 159.

"The Coordinated Radio and Infrared Survey for High-mass Star Formation. II. Source Catalog", Purcell, C.R., and 25 authors including L. Mundy, 2013, Ap.J.Sup., 205, 1P.

"Design and status of the Balloon Experimental Twin Telescope for infrared interferometry (BETTII): an interferometer at the edge of space", 2012, Rinehart, S.A., and 15 authors including L. Mundy, SPIE, 8445, 08R.

"racking near-infrared fringes on BETTII: a balloon-borne, 8m-baseline interferometer", Rizzo, M.J., Rinehart, S.A.; Barry, R.K.; Benford, D.J.; Fixsen, D.J.; Kale, T.; Leisawitz, D.T.; Lyon, R.G.; Mentzell, E.; Mundy, L. G.; Silverberg, R. F., 2012, SPOE 8445, 1TR.

"DIRT: Dust Infrared Toolbox", Pound, M.W., Wolfire, M.G., Mundy, L.G., Lord, S., 2011, http://adsabs.harvard.edu/abs/2011ascl.soft02021P.

"Resolving the Circumstellar Disk of HL Tau at Millimeter Wavelengths", Kown, Woojin; Looney, Leslie W.; Mundy, Lee G., 2011, Ap.J., 741, 3K.

"Benford D.J., Mundy, L.G., "WISPIR: a wide-field imaging spectrograph for the infrared for the SPICA Observatory", 2010, SPIE, 7731, 22B.

"Grain Growth and Density Distribution of the Youngest Protostellar Systems", Kwon, Woojin; Looney, Leslie W.; Mundy, Lee G.; Chiang, Hsin-Fang; Kemball, Athol J., 2009, Ap.J., 696, 841-852.

"Deep JHKs and Spitzer Imaging of Four Isolated Molecular Cloud Cores", Chapman, Nicholas L.; Mundy, Lee G., 2009, Ap.J., 1866-1882.

"The Mid-Infrared Extinction Law in the Ophiuchus, Perseus, and Serpens Molecular Clouds", Chapman, Nicholas L.; Mundy, Lee G.; Lai, Shih-Ping; Evans, Neal J., 2009, Ap.J., 690, 496, 511.

"The Spitzer c2d Legacy Results: Star-Formation Rates and Efficiencies; Evolution and Lifetimes", Evans, N.J. II, and 17 other authors including L. Mundy, 2009, Ap.J.Sup., 181, 321.

"Constraining the Earliest Circumstellar Disks and their Envelopes", Chiang, Hsin-Fang; Looney, Leslie W.; Tassis, Konstantinos; Mundy, Lee G.; Mouschovias, Telemachos Ch., 2008, Ap.J., 680, 474-482.

ANTHONY J. REMIJAN

Address and Telephone Number:
National Radio Astronomy Observatory
520 Edgemont Rd., Charlottesville, VA 22903
Tel: (434) 244-6848; Fax: (434) 296-0278
E-mail address: aremijan@nrao.edu

Employment:

2011-Present	NRAO Associate Scientist – Division Head, Scientific User Support and Research
2011-2013	NRAO Assistant Scientist – Manager, Observatory Science Operations
2008-2011	NRAO Assistant Scientist – ALMA Commissioning and Science Verification

Postdoctoral Appointments:

2006-2008	ALMA Postdoctoral Research Associate, National Radio Astronomy Observatory
2004-2006	Resident Research Associate, NASA Goddard Space Flight Center

PhD Education:

1998-2003 Ph.D. in Astronomy, University of Illinois at Urbana-Champaign
Ph.D. Thesis: "Observations of Large Biologically Important Interstellar and Cometary Molecules"
Advisor: Prof. Lewis E. Snyder (Univ. of Illinois)

Students Advised:

Joanna Corby, University of Virginia, 2010 - current, Thesis Advisor
Brett McGuire, Caltech, 2012 - current, Thesis Committee Member
Ryan Loomis, University of Virginia, 2011 - current, Undergraduate Thesis Advisor

Selected Publications in Refereed Journals (CY 2013):

Observational Results of a Multi-Telescope Campaign to Detect Interstellar Urea [$(NH_2)_2CO$], Anthony J. Remijan, Lewis E. Snyder, Brett A. McGuire, Hsin-Lun Kuo, Leslie W. Looney, Douglas N. Friedel, G. Yu Golubiatnikov, Frank J. Lovas, V. V. Ilyushin, E. A. Alekseev, S. F. Dyubko, Benjamin J. McCall, & Jan M. Hollis, 2013, ApJ, submitted

A Search for l-C_3H^+ and l-C_3H in Sgr B2(N), Sgr B2(OH), and the Dark Cloud TMC-1, B. A. McGuire, P. B. Carroll, R. A. Loomis, G. A. Blake, J. M. Hollis, F. J. Lovas, P. R. Jewell, & A. J. Remijan, 2013, ApJ, accepted

The Detection of Interstellar Ethanimine (CH_3CHNH) from Observations Taken during the GBT PRIMOS Survey, Ryan A. Loomis, Daniel P. Zaleski, Amanda L. Steber, Justin L. Neill, Matthew T. Muckle, Brent J. Harris, Jan M. Hollis, Philip R. Jewell, Valerio Lattanzi, Frank J. Lovas, Oscar Martinez, Jr., Michael C. McCarthy, Anthony J. Remijan, Brooks H. Pate, & Joanna F. Corby, 2013, ApJ, 765, L9

Detection of E-Cyanomethanimine toward Sagittarius B2(N) in the Green Bank Telescope PRIMOS Survey, Daniel P. Zaleski, Nathan A. Seifert, Amanda L. Steber, Matt T. Muckle, Ryan A. Loomis, Joanna F. Corby, Oscar Martinez, Jr., Kyle N. Crabtree, Philip R. Jewell, Jan M. Hollis, Frank J. Lovas, David Vasquez, Jolie Nyiramahirwe, Nicole Sciortino, Kennedy Johnson, Michael C. McCarthy, Anthony J. Remijan, & Brooks H. Pate, 2013, ApJ, 765, L10

Acetone in Orion BN/KL. High-resolution maps of a special oxygen-bearing molecule, Peng, T.-C., Despois, D., Brouillet, N., Baudry, A., Favre, C., Remijan, A., Wootten, A., Wilson, T. L., Combes, F., & Wlodarczak, G. 2013, A&A, 554, 27

Biographical Sketch: Amitabh Varshney

Contact Information

4407 AV Williams Building, University of Maryland, College Park, MD 20742
Phone: (301) 405-6722, FAX: (301) 314-9658
Email: varshney@umiacs.umd.edu, Web: http://www.cs.umd.edu/~varshney

Professional Preparation

- *B. Tech. in Computer Science and Engg*, Indian Institute of Technology, Delhi, 1989.
- *M.S. in Computer Science*, University of North Carolina, Chapel Hill, NC, 1991.
- *Ph.D. in Computer Science,* University of North Carolina, Chapel Hill, NC, 1994.

Appointments

- Director, University of Maryland Institute for Advanced Computer Studies (UMIACS), University of Maryland, College Park, MD, *2010 – present*.
- Director, NVIDIA CUDA Center of Excellence, University of Maryland, College Park, MD *2010 – present.*
- Professor, Department of Computer Science and UMIACS, University of Maryland, College Park, MD, *2006 – present.*
- Associate Professor, Department of Computer Science and UMIACS, University of Maryland, College Park, MD, *2000 – 2006.*
- Assistant Professor, Department of Computer Science, State University of New York, Stony Brook, New York, *1994 – 2000*.

Five Relevant Publications

1. *Saliency-Assisted Navigation of Very Large Landscape Images*, C. Y. Ip and A. Varshney, IEEE Trans. on Visualization and Computer Graphics, Vol 17 (12), 2011, pp 1737-1746.
2. *Mesh Saliency*, C. H. Lee, A. Varshney, D. Jacobs, ACM Trans. Graphics (SIGGRAPH) 24(3), 2005, pp 659 – 666.
3. *Saliency-guided Enhancement for Volume Visualization*, Y. Kim and A. Varshney, IEEE Trans. Vis. Comp. Graphics, 2006, 12(5), pp 925 – 932.
4. *Persuading Visual Attention through Geometry*, Y. Kim and A. Varshney, IEEE Transactions on Visualization and Computer Graphics, Vol. 14(4), pp 772 – 782, July 2008.
5. *Mesh Saliency and Human Eye Fixations*, Y. Kim, A. Varshney, D. Jacobs, and F.Guimbretière, ACM Transactions on Applied Perception, Vol. 7, No.2, 2010, pp 1 – 13.

Five Significant Publications

1. *Simplification Envelopes*, J. Cohen, A. Varshney, D. Manocha, G. Turk, H. Weber, P. Agarwal, F. P. Brooks,Jr., W. V. Wright, Proceedings SIGGRAPH 96, pp 119–128, 1996
2. *Optimizing Triangle Strips for Fast Rendering*, F. Evans, S. Skiena, and A. Varshney, Proceedings of the IEEE Visualization '96, pp 319 –326., 1996
3. *Adaptive Real-Time Level-of-detail-based Rendering for Polygonal Models*, J. Xia, J. El-Sana, and A. Varshney IEEE Trans on Vis and Comp Graphics, 3(2), 1997, pp 171–183
4. *Generalized View-Dependent Simplification*, J. El-Sana and A. Varshney, Computer Graphics Forum (Eurographics'99), 18(3), 1999, pp 83–94
5. *High-throughput Sequence Alignment using Graphics Processing Units*, by M. C. Schatz, C. Trapnell, A. Delcher, and A. Varshney, BMC Bioinformatics, 8:474, December 2007

Selected Recent Professional Service

- *Chair,* IEEE Visualization and Graphics Technical Committee, 2008 – 2012.
- *General Chair,* Innovative Parallel Computing (InPar) 2012.
- *Editorial Board*, Graphical Models, since 2011.
- *Papers Co-Chair*, ACM Symposium on Interactive 3D Graphics and Games, 2011.
- *General Co-Chair*, ACM Symposium on Interactive 3D Graphics and Games, 2010.
- *Keynote Speaker,* Symposia on Point-based Graphics & Volume Graphics, Prague, CZ, 2007
- *Best Papers Award Committee*, IEEE Visualization 2009, 2007, 2002
- *Steering Committee,* IEEE Visualization Conference, since 2006
- *Conference Chair*, IEEE Visualization 2006 and 2007
- *Editorial Board*, IEEE Transactions on Visualization and Computer Graphics, 1999 – 2003
- *Papers Chair,* IEEE Visualization 2000 and 2001
- *Technical Program Committees* for several conferences including Shape Modeling Intl, Computer Graphics Intl, Eurographics, Symposium on Geometry Processing.

Synergistic Activities

My research interests are in high-performance scientific visualization of very large datasets. We have worked on interactive methods for visualization using meshes and volumes for several application domains, including astronomy, proteins, computational fluid dynamics, medical imaging, plasma physics, and mechanical CAD. We have worked on computational models of saliency, mesh processing, and geometric modeling of large visualization datasets. We have explored new techniques for visual depictions of interesting regions in landscape and hyperspectral imagery. We have been working on mapping and characterizing a variety of applications on streaming architectures of GPUs (graphics processing units). We have also been exploring how best to interleave computation and visualization on CPUs and GPUs for very large-scale datasets in cluster environments.

Collaborators with Affiliations

1. **Collaborators and Co-Editors:**
   Ricardo Barni (UFRGS, Brazil), Joao Comba (UFRGS, Brazil), Ricardo Farias (Federal University of Rio de Janeiro, Brazil), Francois Guimbretiere (Cornell University), Aswin Sankaranaryanan (Rice University), George Stantchev (SAIC), Qin Wang (Rhythm and Hues Studios), Chris Wyman (University of Iowa).

2. **PhD Dissertation Advisor:** Frederick P. Brooks, Jr., Univ. of North Carolina at Chapel Hill.

3. **Graduated Ph.D. Students:**
   Chu-Fei Chang (ITRI, Taiwan), Jihad El-Sana (Ben-Gurion University, Israel), Francine Evans (Schlumberger Research, TX), Xuejun Hao (Columbia University, NY), Aravind Kalaiah (NVIDIA, CA), Youngmin Kim (Knight Equity Partners, NYC, NY), Chang Ha Lee (Chung-Ang University, Korea), Andre Maximo (Institute of Pure and Applied Math (IMPA), Rio de Janeiro, Brazil).