Cover Page for Proposal Submitted to the National Aeronautics and Space Administration

NASA Proposal Number

TBD on Submit

NASA PROCEDURE FOR HANDLING PROPOSALS

This proposal shall be used and disclosed for evaluation purposes only, and a copy of this Government notice shall be applied to any reproduction or abstract thereof. Any authorized restrictive notices that the submitter places on this proposal shall also be strictly complied with. Disclosure of this proposal for any reason outside the Government evaluation purposes shall be made only to the extent authorized by the Government.

SECTION I - Proposal Information

Principal Investigator	E-mail Address	Phone Number
Alyssa Goodman	**agoodman@cfa.harvard.edu**	**617-495-9278**

Street Address (1)	Street Address (2)
60 Garden St	**MS 42**

City	State / Province	Postal Code	Country Code
Cambridge	**MA**	**02138-1516**	**US**

Proposal Title : **oldAstronomy: Crowdsourcing the Classification of Historical Astronomical Images**

Proposed Start Date	Proposed End Date	Total Budget	Year 1 Budget	Year 2 Budget	Year 3 Budget	Year 4 Budget
01 / 01 / 2014	**12 / 31 / 2015**	**338,182.00**	**202,139.00**	**136,043.00**	**0.00**	**0.00**

SECTION II - Application Information

NASA Program Announcement Number	NASA Program Announcement Title
NNH13ZDA001N-ADAP	**Astrophysics Data Analysis**

For Consideration By NASA Organization *(the soliciting organization, or the organization to which an unsolicited proposal is submitted)*

NASA , Headquarters , Science Mission Directorate , Astrophysics

Date Submitted	Submission Method	Grants.gov Application Identifier	Applicant Proposal Identifier
	Electronic Submission Only		

Type of Application	Predecessor Award Number	Other Federal Agencies to Which Proposal Has Been Submitted
New		

International Participation	Type of International Participation
No	

SECTION III - Submitting Organization Information

DUNS Number	CAGE Code	Employer Identification Number (EIN or TIN)	Organization Type
082359691	**1NQH4**	**042103580**	**2J**

Organization Name (Standard/Legal Name)	Company Division
Harvard College	

Organization DBA Name	Division Number

Street Address (1)	Street Address (2)
1350 MASS AVE STE 600	

City	State / Province	Postal Code	Country Code
CAMBRIDGE	**MA**	**02138-3846**	**USA**

SECTION IV - Proposal Point of Contact Information

Name	Email Address	Phone Number
Alyssa Goodman	**agoodman@cfa.harvard.edu**	**617-495-9278**

SECTION V - Certification and Authorization

Certification of Compliance with Applicable Executive Orders and U.S. Code

By submitting the proposal identified in the Cover Sheet/Proposal Summary in response to this Research Announcement, the Authorizing Official of the proposing organization (or the individual proposer if there is no proposing organization) as identified below:

- certifies that the statements made in this proposal are true and complete to the best of his/her knowledge;
- agrees to accept the obligations to comply with NASA award terms and conditions if an award is made as a result of this proposal; and
- confirms compliance with all provisions, rules, and stipulations set forth in the two Certifications and one Assurance contained in this NRA (namely, (i) the Assurance of Compliance with the NASA Regulations Pursuant to Nondiscrimination in Federally Assisted Programs, and (ii) Certifications, Disclosures, and Assurances Regarding Lobbying and Debarment and Suspension.

Willful provision of false information in this proposal and/or its supporting documents, or in reports required under an ensuing award, is a criminal offense (U.S. Code, Title 18, Section 1001).

Authorized Organizational Representative (AOR) Name	AOR E-mail Address	Phone Number

PI Name : **Alyssa Goodman**	**NASA Proposal Number**
Organization Name : **Harvard College**	**TBD on Submit**

Proposal Title : **oldAstronomy: Crowdsourcing the Classification of Historical Astronomical Images**

SECTION VI - Team Members

Team Member Role	Team Member Name	Contact Phone	E-mail Address
PI	**Alyssa Goodman**	**617-495-9278**	**agoodman@cfa.harvard.edu**
Organization/Business Relationship: **Harvard College**		Cage Code: **1NQH4**	DUNS#: **082359691**
International Participation: **No**	U.S. Government Agency:		Total Funds Requested: **0.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Co-I	**Alberto Pepe**	**310-600-3929**	**apepe@cfa.harvard.edu**
Organization/Business Relationship: **Harvard College**		Cage Code: **1NQH4**	DUNS#: **082359691**
International Participation: **No**	U.S. Government Agency:		Total Funds Requested: **0.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Co-I/Institutional PI	**Arfon Smith**	**312-322-0525**	**arfon@adlerplanetarium.org**
Organization/Business Relationship: **Adler Planetarium**		Cage Code: **33EH9**	DUNS#: **083081802**
International Participation: **No**	U.S. Government Agency:		Total Funds Requested: **0.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Collaborator	**Alberto Accomazzi**	**617-495-7076**	**aaccomazzi@cfa.harvard.edu**
Organization/Business Relationship: **Smithsonian Institution/Smithsonian Astrophysical Observatory**		Cage Code: **1PPP1**	DUNS#: **003261823**
International Participation: **No**	U.S. Government Agency:		Total Funds Requested: **0.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Collaborator	**Jonathan Fay**	**425-703-6825**	**jfay@microsoft.com**
Organization/Business Relationship: **Microsoft Research**		Cage Code: **N/A**	DUNS#: **N/A**
International Participation: **No**	U.S. Government Agency:		Total Funds Requested: **0.00**

PI Name : **Alyssa Goodman**	**NASA Proposal Number**
Organization Name : **Harvard College**	**TBD on Submit**

Proposal Title : **oldAstronomy: Crowdsourcing the Classification of Historical Astronomical Images**

SECTION VII - Project Summary

This proposal seeks to augment the ongoing ADS All-Sky Survey effort with a citizen science project called "oldAstronomy," and to tightly integrate results into ADS itself.

In 2011, the NASA ADAP program funded the ADS All-Sky Survey (ADSASS), which is turning the NASA Astrophysics "Data" System (ADS), known for its literature holdings, into more of a "data" resource. The ADSASS has two components: 1) an all-sky "heatmap" showing why the sky has been studied, for what reasons, over time; and 2) an "historical image layer," constructed by extracting all the images of the sky from ADS-held journals. The heatmap uses "astrotags" (sourcename-article associations) to calculate the density of articles about each sky position, and the image layer is constructed by "astroreferencing" extracted images in order to place them at the correct sky location.

In order to astroreference (determine the scale and position of) of images on the sky, the current ADSASS pipeline extracts images from articles automatically, and then runs all of them through an automated sky-registration program called "Astrometry.net" that relies on matching star catalogs to stars in the image. In designing the original ADSASS project, we knew that Astrometry.net would succeed only on images that: contain many optically-visible stars; are not heavily annotated; are not split into several subfigures; and are not otherwise subject to graphical manipulations that confound automated systems. In practice, we know today that about 1% of images of the sky in the ADS can be solved in a fully-automated pipeline.

In 2011, we speculated that automatically text-mining figure captions might offer a good path to astroreferencing images that Astrometry.net cannot solve, but our research has now shown that humans are much better than computers at understanding the annotations needed to put a figure back on the sky. In the proposed project, to be carried out using the popular citizen-science Zooniverse platform, we plan to calibrate all the images which cannot be solved in automated fashion via a citizen-science initiative called "oldAstronomy." Zooniverse citizen scientists will be presented with a clear series of engaging tasks that will result in each sky image being correctly astroreferenced inside of, on average, a few minutes. Citizen scientists will be rewarded with special previews of their own astroreferenced images, shown to them within the oldAstronomy web environment using an elegant sky browser made possible by the new HTML5 implementation of WorldWide Telescope from Microsoft Research.

The combination of automatically-solved images resulting from our ongoing ADSASS work and the new proposed oldAstronomy work will result in a database of ~240K astro-referenced images, i.e., images of the sky for which coordinates, orientation, and pixel scale will be available as metadata. Images in that database will be made accessible, via ADS, as coordinate-aware image downloads and as image views offered in a sky context using WorldWide Telescope (WWT). The software development efforts in the proposal will culminate in a sustainable system for exposing the new ADSASS content from within ADS (e.g. from individual article's pages) and as an all-sky (patchwork) layer in WWT. Because WWT already offers all-sky and detailed imagery at all wavelengths, and now runs in any platform's web browser, accomplishing the ADSASS -WWT integration will allow astronomers the world over to study the hundreds of thousands of images previously locked up within ADS's articles in their natural context--on the sky.

CONTENTS

1. INTRODUCTION

Overview

In 2012, the NASA ADAP program funded our group to create the ADS All-Sky Survey (ADSASS), a project designed to turn the NASA Astrophysics Data System (ADS) into a data resource. We are pleased to report that the initial ADSASS will be released as an online resource in early 2014. This initial release will offer "heat maps" of the density of articles on the sky, as well as access to an image layer where astronomical images formerly only available as figures within ADS-held articles are now viewable, in context, on the sky. The work that went into this initial release included mining the "astrotags" (mentions of sources/sky positions) in ADS listings (created in collaboration with CDS[1]) and automatically extracting and then placing ("astro-referencing") tens of thousands of sky images from ADS holdings. With the new work proposed here, we aim to *complete* the image layer in the ADSASS by adding about 236K more images and to *integrate* the survey with other online resources, making it much more accessible and useful in astronomical and historical research.

The initial (2014) release of the image layer, based on *automated* astro-referencing, will contain about 4000 individual images, which is only ~0.1% of ~3.2 million images in the ~800K astronomy articles accessible through ADS (see Figure 4). The remaining images are absent for two reasons: 1) for roughly 93% of the ADS corpus, article copyrights are still held by publishers, and we need to complete negotiations to ensure that it is legal to "re-publish" images in publisher-owned articles as part of the ADSASS; and, 2) more critically, only a small fraction (~1.5%) of ADS figures are optical images of the sky of the kind that can be successfully placed at the right sky location ("astro-referenced") using an automated pipeline. Our current automated astro-referencing pipeline uses Astrometry.net, a service which can algorithmically determine the scale and position of images on the sky. But, as we explained in the original ADSASS proposal, images that do not contain enough optically-visible stars cannot be astro-referenced by Astronomy.net. In addition, many images–even some that could be potentially be solved by Astrometry.net–are heavily annotated, split into several subfigures, or subject to other graphical manipulations that confound automated systems. In sum, roughly one out of every 100 actual astronomical images of the sky in the ADS can be solved in a fully-automated pipeline.

In the funded (2011) ADSASS proposal, we speculated that automatically text-mining figure captions might offer a good path to astro-referencing images Astrometry.net cannot solve. Our research into that approach has now clearly shown *people* to be much smarter than computerized text-mining for this task! As a result, in February 2013, we began a collaboration with the Zooniverse[2] team to turn astro-referencing images from ADS holdings into a Citizen Science project called "oldAstronomy." In this new effort, we plan to harness the collective intelligence of the Zooniverse platform and citizen scientists to: 1) find out which images contain a sky view; and 2) perform astrometric calibration for sky images that cannot be resolved in an automated fashion. Volunteers will be presented with an image extracted from an astronomy paper and asked to per-

---

[1] CDS is the Strasbourg astronomical Data Center, an EU-based organization dedicated to the collection and worldwide distribution of astronomical data and related information.

[2] Zooniverse is a citizen science web portal in which human volunteers actively participate in projects to complete research tasks, see zooniverse.org.

form the following tasks: 1) *subdivide* the image into multiple images in case of a multi-pane image; 2) *classify* the image (e.g., image of the sky, table, drawing, graph); 3) identify if *coordinates* exist on the image and if so mark them; 4) manually *adjust image* contrast and color scale to improve the chances of getting the image calibrated with Astrometry.net; and finally 5) *discuss* the image with experts and *tag* it with useful information such as epoch and bandwidth of the image.

The combination of automatically-solved images resulting from our original ADSASS work and the new proposed oldAstronomy work will result in a potential database of ~240K astro-referenced images, i.e., images of the sky for which coordinates, orientation, and pixel scale will be available in the form of calibration metadata. The resulting image archive, to be hosted by ADS, will be very similar to a telescope data center's archive, in that it will consist of images and metadata describing the properties of those images. Only, in this case, the "metadata" will include not only the coordinates, orientation, and plate scale needed to make the image useful in sky-viewing tools–it will also offer a link to a full published article putting the image into context!

This proposal describes both our plans for the new oldAstronomy aspect of the ADSASS and our plans to create a robust archive and ingest system for future content designed to make the output of ADSASS a key piece of the rapidly-evolving online astronomy research environment.

History of the ADSASS, 2011-2013

The ADS All-Sky Survey, ADSASS (Pepe et al, 2012) is an ongoing NASA-funded effort (NASA NNX12AE11G: ADS All-Sky Survey; PI: A. Goodman) aimed at turning the NASA Astrophysics Data System (ADS), widely known for its unrivaled value as a literature resource for astronomers, into a data resource. With the ADSASS, we are advancing the idea that *the ADS is not a data repository per se, but it implicitly contains valuable holdings of astronomical data, in the form of images, tables and object references contained within articles*. The objective of the ADSASS is to extract these data and make them discoverable and available as data viewable using existing astronomical software. The resulting ADSASS data products (the filterable "all-sky literature heat map" and "historical data layer" described below) will enable new research by tying astronomical literature and data assets into one resource.

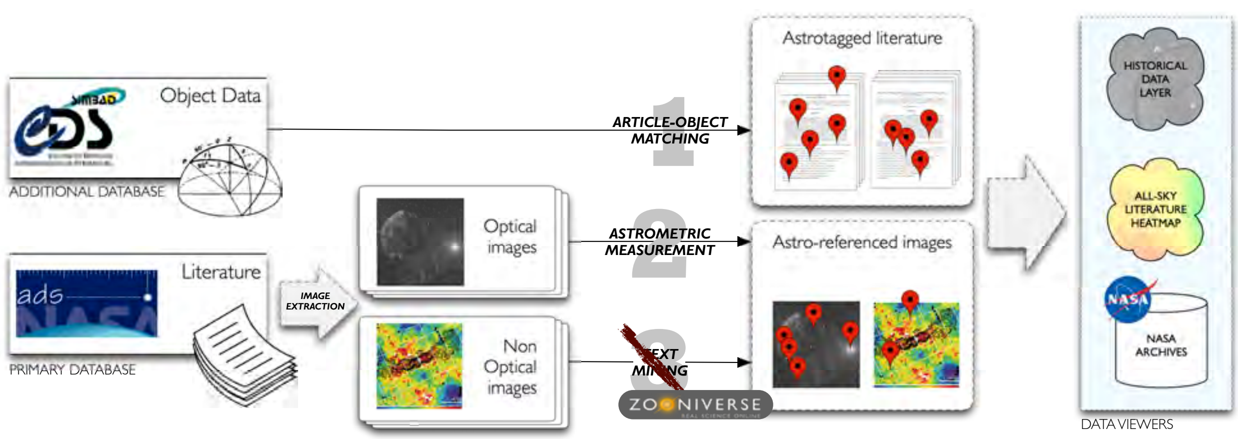


Figure 1 The three steps of the original ADSASS plan are shown labeled with **grey numbers**. Note that the *"TEXT MINING"* step is being replaced with the **oldAstronomy** Zooniverse (citizen-science) effort for which we seek funding in this proposal.

As shown in Figure 1, the ADSASS is composed of three steps. In **Step 1**, we are constructing an **interactive heat map** of the sky, which allows astronomers to visualize what parts of the sky have been cited in the literature, and in what context (see Figure 2). The heat map shows citations to celestial objects. For every article in ADS, CDS staff have curated a list of objects mentioned in that article. Those "mentions" can be thought of as "astrotags" (by analogy to geotags) on the articles, and the heat map is constructed by giving a count to each astrotag associated with an article. The precise content of the heat map can be filtered interactively based on the needs of the astronomer perusing it. For example, an astronomer might want to explore "*who has studied M31, at what wavelength, over time?*" . Video **demonstrations** of the working prototype addressing such questions are available at http://tinyurl.com/adsassmovies, and Figure 2 shows screenshots.

Figure 2 Screenshots of the ADS All-sky literature heat map prototype showing all-sky map of all articles, articles in the vicinity of M31, in different bands and for different time periods.

Step 2 of the ADSASS is the production of an **historical data layer**, constructed from images of the sky automatically extracted from ADS literature, and then astro-referenced using the Astrometry.net program. We have currently extracted roughly 240K images from a sample corpus of over 55K (see Figure 4) astronomy papers, and we have calibrated about 1.5% of them using Astrometry.net (Lang et al, 2010). As explained above, Astrometry.net fails to calibrate many valid images which contain a sky view or coordinates for two reasons. First, Astrometry.net relies fundamentally on comparisons with catalogued stellar positions, so images that do not contain enough stars (e.g. at non-optical/infrared wavelengths and/or with narrow fields of view) cannot be solved. Second, while Astrometry.net is a powerful computational tool, many images contained in articles have been indadvertedly "disfigured" by authors and publishers, thanks to annotation, placement too close to other images, color contrast manipulation, or other graphical manipulations that confound automated systems.

Step 3: In our 2011 ADSASS proposal we discussed text mining image captions as a way to extract metadata for **images that Astrometry.net could not solve.** In many cases, the textual metadata in a figure caption includes a position, source name, a plate scale and/or a waveband. In principle, parsing the captions intelligently could lead to at least an approximate placement on the sky. In the past few months, though, we have experimented with such a text mining approach, with poor results. There are a number of challenges that prevent accurate textual analyses of image captions. These include: 1) the fact that old scanned articles have captions that cannot be converted to digital text, even with advanced OCR methods; and 2) even if names of celestial objects can be easily recognized, interpreting coordinates from free text in an automated fashion is a arduous task due to the ambiguity of language. During our experiments with a corpus of a few thousand images, we came to the realization that automated techniques of text mining do not pose as a solid solution to astro-reference images. On the other hand, manual techniques, that involve human judgement and annotation can prove extremely useful to calibrate images astrometrically: *humans can quickly realize if an image contains a sky field, if it has coordinates, and* *if its caption has references to coordinates, pixel scale, orientation, etc.* For this reason, we decided to approach the task of astro-referencing historical images by harnessing the collective intelligence of citizen scientists: amateur and professional astronomers, as well as non-astronomers, who can peruse images on the web, classify them, tweak them, and annotate them with the intent of providing those images with an astrometric calibration. As indicated in Figure 1, the new citizen-science-based "Step 3" is called **oldAstronomy**.

To design, build, employ, and promote the oldAstronomy citizen-science initiative advanced in this proposal, we are partnering with the Zooniverse[3], the most active and respected citizen science portal on the web owned and operated by the Citizen Science Alliance, which is a collaboration involving top academic institutions (such as the University of Oxford and the University of Nottingham), museums (such as the Adler Planetarium and the National Maritime Museum), and funding bodies (such as NASA, the NSF, and Sloan). The Zooniverse is a portal, a platform, and an aggregator for citizen science projects. At the moment it has a community of over 800K volunteers and it hosts more than a dozen projects in fields such as astronomy, earth science, biology and even the humanities.

Outcomes

The oldAstronomy project proposed here will become a **citizen science initiative** hosted on the Zooniverse platform[4]. Citizen scientists will visit oldAstronomy, and by performing the tasks outlined in the next section, their work will lead to a **database** of roughly 240K astro-referenced images, i.e., images provided with calibration metadata. This database will be integrated as part of our proposed effort into the **ADSASS historical data layer.** Individual images and the full historical data layer will be made available via 1-click access from within ADS, using integrative "virtual observatory" tools, such as CDS's Aladin and Microsoft's WorldWide Telescope.[5]

Figure 3 shows an example of how Astrometry.net-calibrated images are viewable on the web now. On the image-sharing site flickr, an Astrometry.net bot has been implemented that analyzes all images in the flickr group called "astrometry." In the example, an image of IC1396 uploaded to the flickr astrometry group has had coordinate and scale metadata automatically added by the Astrometry.net bot, in the image's flickr page. The last line of the flickr comments posted by the bot offers a "View in WorldWide Telescope" link to see where the image belongs on the sky. After clicking that link, a user can opt to s*ee the IC1396 image on nearly any wavelength background*, because WWT has built-in access to all-sky imagery as well as many regional images, at a variety of wavelengths. The majority of the contextual imagery available in WWT has been derived from NASA-funded missions and resources. In Figure 3, the user has chosen a far-infrared background, to see how the location of this nebula relates to our Galaxy's dust distribution. A user might also notice that WWT, by default, offers preview images of sources of interest within the

[3] The U.S. base of Zooniverse and the Citizen Science Alliance is the Adler Planetarium in Chicago, and all NASA funding requested in this proposal will be spent within the U.S., through agreements with the Adler-based Zooniverse team.

[4] see https://www.zooniverse.org

[5] WWT is developed by Microsoft Research, in collaboration with the PI of this proposal, and the viewer to be used in ADS runs, *independent of operating system*, in any browser, using an HTML5 API.

view field, including, in this case, the "Elephant Trunk Nebula" (tilted inset in Figure 3, left). The user can zoom in on the Elephant Trunk region, shift-click to access WWT's "Finder Scope," and then be offered direct "Research" links to information about this source in the form of ADS references or SIMBAD data links. These WWT services already work, for any position on the sky.

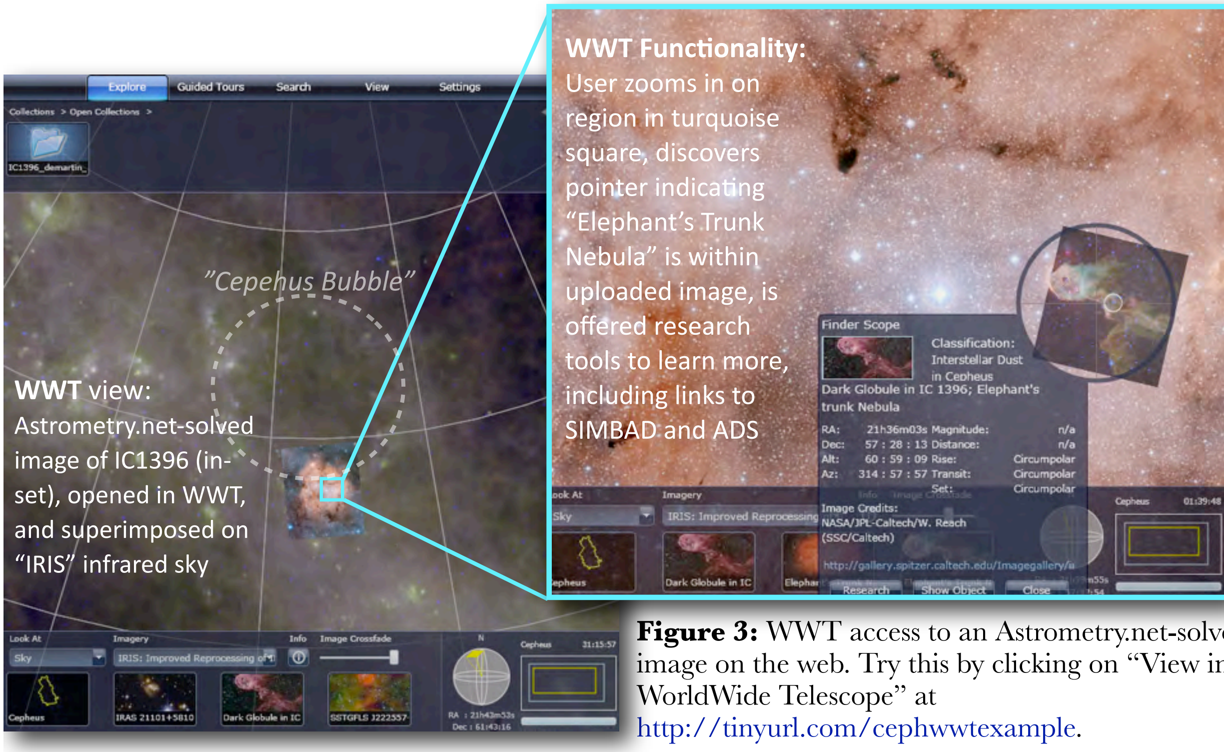


Figure 3: WWT access to an Astrometry.net-solved image on the web. Try this by clicking on "View in WorldWide Telescope" at http://tinyurl.com/cephwwtexample.

Two groups–oldAstronomy citizen scientists *and* all ADSASS users–will all be rewarded with opportunities to **explore the sky** as shown in Figure 3.

In addition to completing the historical data layer, we will establish a pipeline for images within **future papers** added to ADS to be extracted and to become part of the ADSASS. The ultimate outcome of this proposal will be the creation of an integrated **end-to-end** system offering access to all of ADS's articles' images, both present and future. In addition, the heat map of simple *source-mention astrotags* produced under our earlier award will be augmented with separate **heat map**, showing the *coverage of all the solved images* in the historical data layer.

2. TECHNICAL APPROACH AND METHODOLOGY

Data sources

The primary database for this project is the literature contained in the NASA ADS, maintained at the Harvard-Smithsonian Center for Astrophysics (CfA). The ADS maintains three bibliographic databases containing roughly 9 million records, 4.5 million scanned pages, and 2.6 million articles in their full text format (a "full text" is a scanned or digital version of an article). A Venn diagram depicting the organization of ADS' holdings is presented in Figure 4. Of the 2.6 million articles, 800K are astronomy articles. Those 800K articles represent the full corpus from which we can extract images and add metadata to them. At the moment, we are only employing a subset of the entire corpus, comprising the 55K articles which we can legally expose to the public. These include roughly 27K publisher-created PDFs managed by ADS and 28K ADS-generated PDFs with either grayscale or color images. The ADS is the official distributor and repository for these (27K+28K=55K) images; as such, they are not subject to any publisher-imposed copyright restriction. The ADSASS (and oldAstronomy) will ultimately expand to the entire corpus of 800K papers, subject to publisher authorization. Once we offer our first "data release" of the ADSASS (in early 2014), we will seek the needed authorizations. Since allowing inclusion of figures from these "closed" papers in the ADSASS effectively amounts to free advertising for publisher-held articles, we expect the publishers to agree to release the images, and discussions to date show publishers to be amenable.

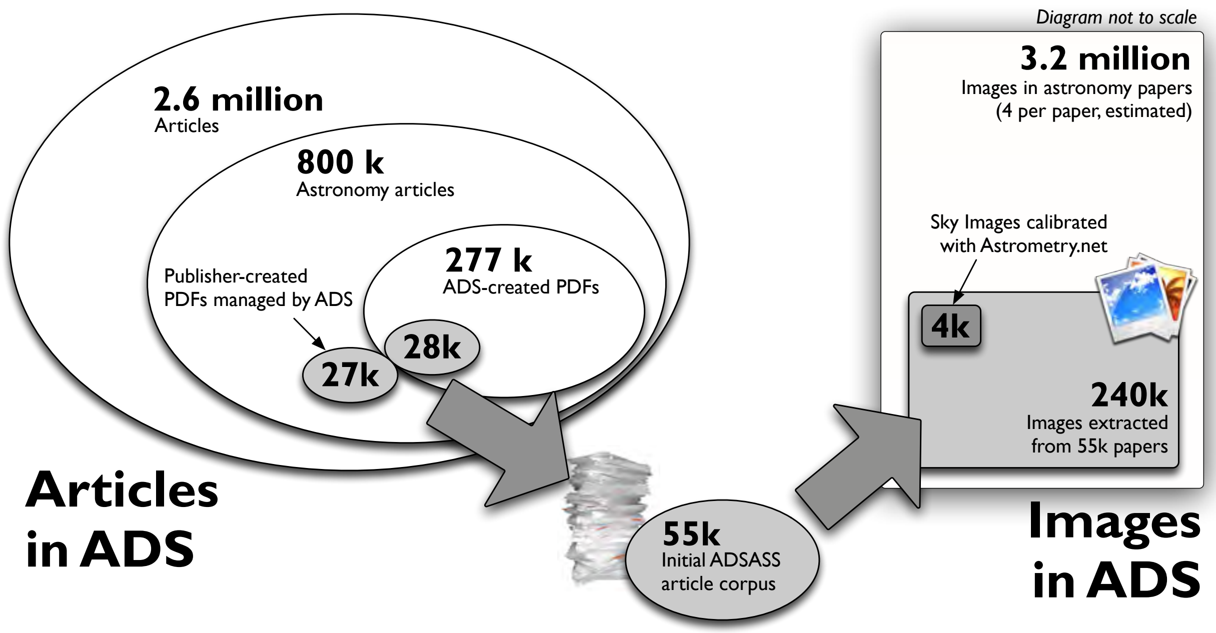


Figure 4 Quantitative breakdown of the holdings of the Astrophysics Data System (articles and images)

For the time being, we have extracted 240K images from the corpus of 55K articles and these images are currently available (archived) on ADS servers. Every extracted image is available both in .ppm format (average size 5MB) and .png (average size 500KB). Each image is also available in its inverted format (e.g., black on white images of the sky are also available as white on black).

As explained above, the **automated process** (Astrometry.net) that we ran on the image corpus of roughly 240K images successfully **placed only ~1.5% of these images** on the sky. The **remaining 98.5%** of the 225K images that were not calibrated by Astrometry.net automatically will be worked on by **citizen scientists**, in the first phase of the oldAstronomy project. These "failed" images vary widely in content. In Figure 5, we present a small collection of the kind of images likely to be found in astronomical papers. Key characteristics of these images are presented in a tabular format in Table 1. One obvious issue is that not all images in the journals give information about the sky (e.g. Figure 5d). Even when images do show the sky, many observations presented in modern articles have been made at wavelengths where stars are not apparent, so Astrometry.net, which relies on star catalogs for its astrometric solutions, cannot succeed.

Further challenges to automatic segmentation and registration[6] are presented when images are heavily annotated, split into panes, not sufficiently high in color contrast, or subject to other similar graphical manipulation (e.g., Figure 5a,b). Of the images shown in Figure 5, we expect that only Figure 5c could ever be solved by Astrometry.net in an automated fashion.

Yet, except for the non-sky image in Figure 5d, humans can facilitate the calibration of all the images in Figure 5. Figure 5a contains clear coordinate information on its axis, and can be solved by most users, after having been split into subfigures. Figure 5b will be solved by expert users (with some knowledge of astronomy), as coordinate information available in the caption has to be interpreted. Figure 5c cannot be manually calibrated, but our experiments show that a human instructed to "adjust contrast and brightness to bring out stars more clearly" does better than a computer program at

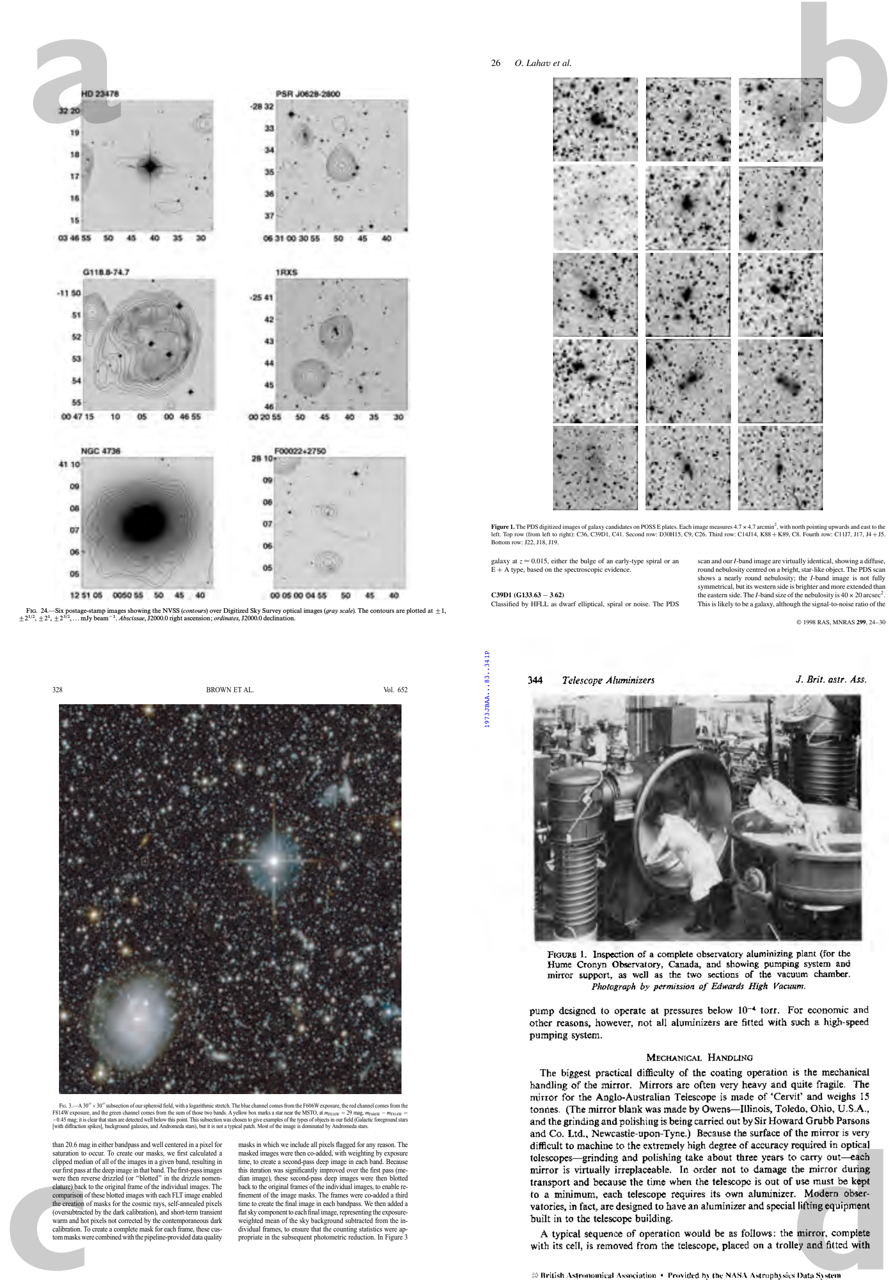


Figure 5 Sample of images contained in the corpus of papers.

TABLE 1: Characteristics of Images in Figure 5	a	b	c	d
sky image?	yes	yes	yes	no
multi-pane?	yes	yes	no	no
coordinates?	yes	yes/no	no	no
solvable by Astrometry?	no	no	maybe	no

[6] The terms "segmentation" and "registration" refer, respectively, to dividing an image into salient features (e.g. sub-panels) and to aligning them to other images (e.g. to a coordinate system on the sky). We sometimes refer to these tasks generically as "calibration" in this proposal.

such a task, to such a degree that Astrometry.net succeeds in solving an "adjusted" image even when it fails on an "unadjusted" version of the same image.

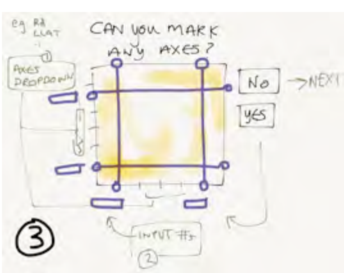
oldAstronomy Workflow

Keeping in mind the diversity of use cases shown in Figure 5 and Table 1, we have developed a generic workflow (detailed below, and in Figure 6) that will allow Zooniverse users to detect and calibrate images. Bear in mind that the kinds of images extracted from the literature may vary widely in content: images of the sky, contour plots, images of people, telescope, *x-y* plots, etc.

Step 0. The very first step is a **tutorial** which will be shown to first time users only. The tutorial involves familiarizing users with the aim, motivation, and tasks of the oldAstronomy project. The tutorial provides an explanation of image classification tasks, manual calibration of images which can be astro-referenced, and an explanation of axis labels and tick labels of astronomical images.

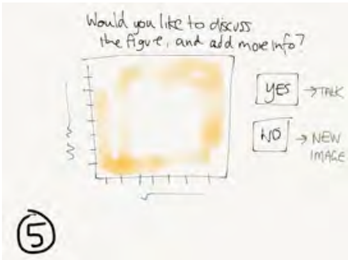
Step 1. User is prompted with an image extracted from the astronomical literature. User is asked: **Is this one image or many?** If user says that there is only one image in view, they will be taken to the next step. If there are multiple images, the user will be asked to draw a rectangle around the image using a selection tool (**cut up image**). Each sub-image will then also re-enter the oldAstronomy workflow cut out as a single image.

Step 2. User is asked: **What sort of image is this?** The aim of this step is to allow citizen scientists to classify the image. Possible classification categories are: a) a picture of the real world, b) a table, c) a drawing, d) a graph, e) a photograph of the sky, f) a contour map of the sky, g) another image of the sky. While only classifications e-g can lead to image calibration (next step), classifications a-d can however inform curatorial efforts of the ADS.

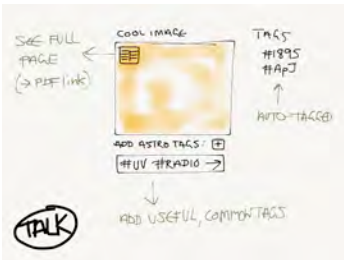


Step 3. User is asked to **mark the axes on the image.** In order to successfully position the image back in the sky w.r.t. coordinates and orientation, we need two separate data points on each *x-y* coordinate plane. We achieve this data collection process by asking users to draw two horizontal straight lines and two vertical lines in the box around the image, and input the coordinates found at those points. Users will be aided in the process by dropdown menus which allow them to select the coordinate system they are dealing with (e.g., RA/DEC).

Step 4. Some genuine images of the sky fail automatic calibration simply because their color contrast is not high enough and Astrometry.net fails to "see the stars". But humans are very good at manually adjusting contrast; thus, in this step, we ask users to adjust all images which they identified as images of the sky to **bring out the stars** using a contrast slider. These image will be sent back to Astrometry.net for calibration in an automated fashion.



Step 5. We finally ask users: **Would you like to discuss this figure and add more information?** It is possible that an image was difficult to calibrate, was intrinsically interesting, or was only partially calibrated (e.g., the user could perform only part of the tasks). In these cases, users can choose to discuss the image in more detail with fellow citizen scientists and experts on a dedicated forum platform called Zooniverse TALK.



Step TALK. The Zooniverse platform has a built-in TALK layer which is used to **discuss anything related to classification tasks**. In the case of oldAstronomy, we will use the TALK layer to allow discussions around interesting images, and to allow users to tag images with extra information, such as journal information, epoch of the celestial coordinates (e.g., B1950 or J2000), and image bandwidth (e.g. uv, visible, infrared, radio).

"Success!" When an image is apparently calibrated successfully using either Step 3 or Step 4, users will be rewarded with their image positioned back in the sky. This will be achieved by overlaying the image (½ transparent overlay of image with DSS in background) in WorldWide Telescope, running in HTML5 in the user's browser. The interrelationship of the steps of this workflow are depicted in in Figure 6.

oldAstronomy in Action: Logistics

Based on discussions with experienced Zooniverse staff, we estimate that having 3 independent users calibrating the same astronomical image with the same metadata is enough to assign safe calibration to an image. Moreover, for testing purposes, images which were previously calibrated automatically using Astrometry.net can (and will) be provided to users to test the accuracy of citizen scientists (i.e., Astrometry.net calibration metadata can be used as ground truth).

Given that we are dealing with an initial corpus of roughly 236K images, ~700K "tasks" will have to be completed in order to achieve our minimal research goal of 3 tasks per image. We estimate, however, that about two thirds of the oldAstronomy tasks will end at Step 2 (see Figure 6) in the workflow (because the image under consideration is not an image of the sky). Thus, the total number of calibration tasks needed is probably in the range of 230K.

The time needed to carry out an astronomical calibration task depends on the content of the image and on the skills of the person performing the task. Experienced astronomers and inexperienced astronomers alike can probably figure out if an image portrays the sky fairly quickly (less than 5 seconds). As explained above, according to our estimates, these images (non-sky) represent the bulk of our corpus (two thirds or more). As for the calibration of sky-images, for an astronomer the entire oldAstronomy task should normally take from 30 seconds to 2 minutes at most. At the other end of the spectrum, for a citizen scientist with very little experience in astronomy, the task can take from 2 to 5 minutes. Estimating an average task time of 5 seconds for 70% of the images (classification of *non*-sky images), and 4 minutes for the other 30% (calibration of sky-images), we expect our sample of 230K images to be completed in 1.3 weeks for the classification of non-sky images assuming one classification task (5 seconds) per citizen scientist **+** 82 weeks for the calibration of sky images assuming 3 calibration tasks for each image and 4-minute tasks for each citizen scientist. The total estimated **human time** required for oldAstronomy is thus **83.3**

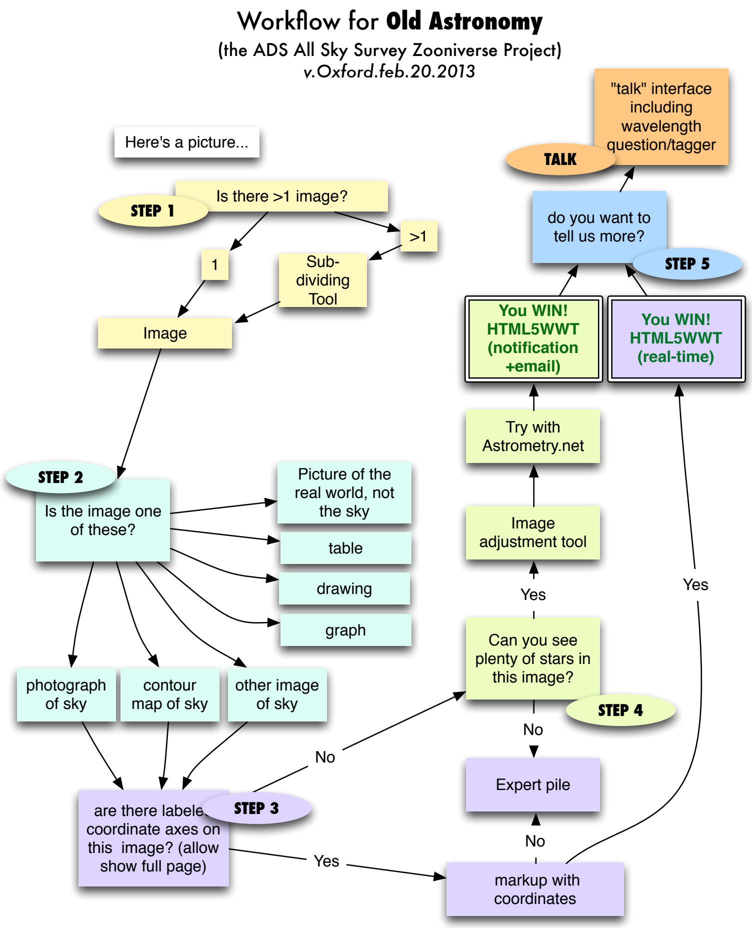


Figure 6 oldAstronomy workflow**.** In **online version** at https://www.cfa.harvard.edu/~agoodman/oldAstronomy/adsass_workflow.html, ovals click through to screen views.

weeks. Assuming a citizen scientist will spend about half hour on the site, we will need **~28K citizen scientists** to complete oldAstronomy. We expect to complete all the classification and calibration tasks in the last quarter of 2014 and first two quarters of 2015 for a total of **~40 weeks of real time**.

Integration with ADS and WWT

As explained in the Plan of Work below, all of the image and metadata output of oldAstronomy, along with ongoing output from the automated parts of the ADSASS will be ingested into ADS, and served to the world both via ADS links and via an historical image layer viewable in WorldWide Telescope.

At present, ADS Labs (adslabs.org) already includes the option[7] to view objects mentioned in search results in context on the sky, using the WWT Silverlight-based web client. For the ADSASS, this kind of integrated WWT link in ADS will use the newer (fully platform-independent) HTML5 version of the WWT web client. ADS pages for articles whose images have been extracted and solved in the ADSASS will offer "view the data on the sky" links, where a user effectively clicks on a figure link to "see" it placed on the sky in WWT, as in Figures 3 and 8. In addition to these new "from the article" links in ADS, the ADSASS images will also be added as a data layer to WWT, so that when a WWT user's field of view includes ADSASS images, those images will show up as selectable for overlay on any other data layer in WWT. The built-in layers in WWT already include all-sky and regional images at a wide variety of wavelengths, and thanks to ADSASS, as of 2014 they will also include the heat map of the full sky based on astrotagging, as shown in Figure 2.

In addition to enabling images-on-the sky functionality, the metadata added by oldAstronomy users will also allow ADS to offer searches for articles containing particular kinds of images. For example, ADS users will be able to find all "images of people" in ADS holdings.

---

[7] **To see this option in action,** click this link, then select "SIMBAD Objects" and then, for example, click the word "Star" and the word "V* PV Cep" and then "WorldWide Telescope."

3. PERCEIVED IMPACT

The ADS All-Sky Survey will become an unprecedented multi-purpose resource for astronomical and historical research, and for the STEM-interested public. The filterable heat maps will allow researchers to understand which parts of the sky have and have not been studied, with what techniques, and for what reasons. The heat map and the historical image layer we aim to complete in this proposal will be used by astronomers for: studies of particular objects; investigations of classes of objects; and, no doubt, in many serendipitous discoveries. Historians of science will be able to use the ADSASS holdings to examine trends in the study of the sky, and to analyze how the quality and quantity of imaging effects the practice and productivity of science over time. And, in rare cases, astronomers who study time-variable objects, will be able to use images recovered from the literature to add key points to time-domain astrophysical investigations.

It is critical to remember that the ADSASS is not *just* about pictures: every astrotag (source mention) in the heat map, and every astronomical image put into context on the sky has an entire journal article attached to it as "metadata." To date, the human capacity for unstructured pattern recognition remains unmatched by any computer algorithm. The visual *patterns* humans see in the ADSASS heat maps, and the visual *context* offered by the historical image layer, will suggest ideas to researchers that they will follow up using the ADS literature resources. The projects that start from "visual" curiosities will lead to some of the most exciting results to be derived from the ADSASS, and it is very had to predict their exact nature without seeing the full ADSASS first!

As one example of the kind of "contextual" work made possible by the ADSASS, consider again the IC1396 region show in Figure 3. Imagine a graduate student studying triggered star formation, where radiation and winds from one generation of stars compresses molecular gas enough to cause gravitational collapse, which engenders a new generation of star formation. In a standard ADS literature search, the student comes across a paper entitled "Origin and Evolution of the Cepheus Bubble," by Patel et al. (2008), Figure 1 of which shows a map of the CO emission (molecular gas) in the region, extending over 10-degrees on the sky (see Figure 7). The graduate student wonders just how big this bubble is, and what other star-forming regions might have been "triggered" by its expansion, but the original paper offers only textual speculation on such questions. Lucky for this graduate student, citizen scientists in the oldAstronomy project were presented with this same image, and they carried out the markup steps described in the previous section and shown here in Figure 7. Thus, the graduate student uses an ADSASS link from the figure's ADS page, to see it on the sky in WorldWide Telescope. Figure 8 (top) shows a view

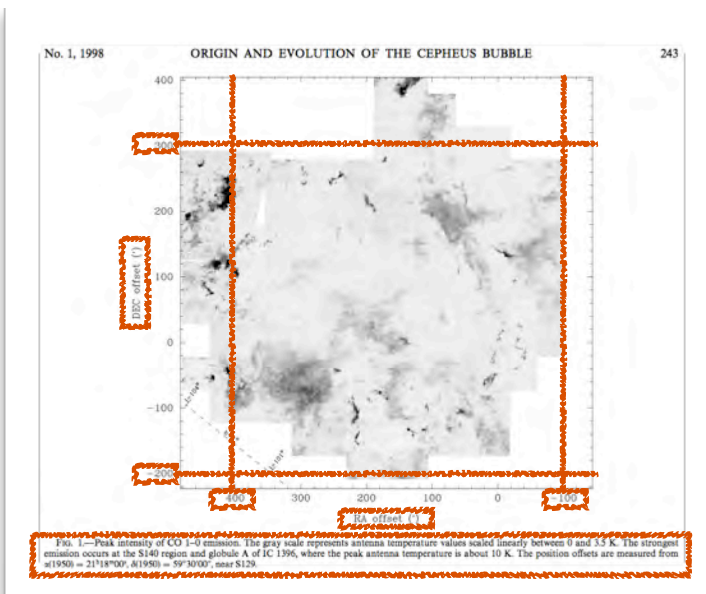


Figure 7 Patel et al. 1998, page 243, Figure 1, with markup (orange) that would have been made by a citizen scientists using oldAstronomy tools.

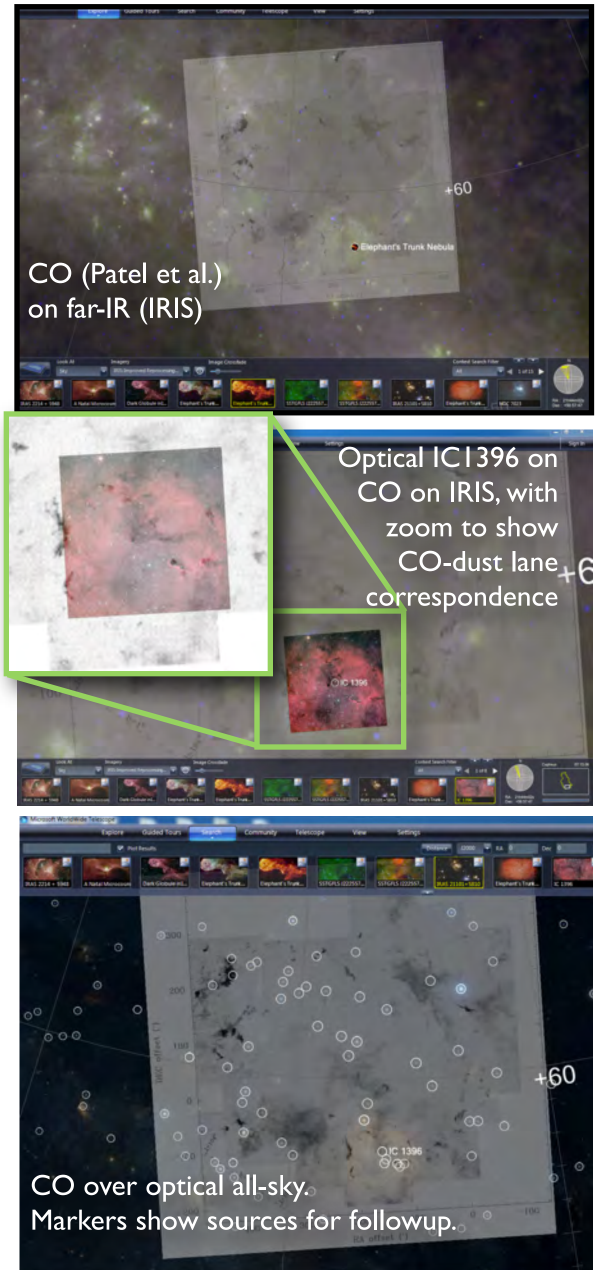


Figure 8 A graduate student's ADSASS exploration of the Cepheus Bubble mapped in CO by Patel et al. in 1998. Each panel shows a successive snapshot, as the student interacts with WWT.

of Patel et al.'s CO map (Figure 7) in WWT, superimposed on the "IRIS" far-IR all-sky survey (Miville-Deschenes, et al. 2005, based on NASA IRAS data). The graduate student creates this contextual view in WWT by just scrolling to a convenient zoom level and choosing IRIS as the background. The graduate student adjusts Patel's figure's transparency with a slider, so as to see how well/poorly the CO-traced molecular gas and the dust emission line up. WWT automatically offers thumbnail-preview images of objects of interest in the present field of view, and our grad student sees a marker for "Elephant's Trunk Nebula" pop up in the lower right part of the CO image, so she zooms in to learn more. She notices that WWT is now offering many preview images (some also from the ADSASS) of the Elephant's Trunk, and its parent nebula, IC1396. She hovers over an IC1396 thumbnail, which adds a third layer to her display (Figure 8, middle), showing an optical image of the IC1396 nebula we highlighted in Figure 7 above. Zooming in on the IC1396 region and adjusting the transparency, the student can then see excellent correspondence between the dark lanes in IC1396 and the CO emission in the Patel et al. image she got from the ADSASS (see green-edged blow-up in Figure 8). Now, she speculates that IC1396 is just one of many star-forming regions along the bubble in Patel et al.'s article. She wonders just how many regions show optical nebulosity (evidence of massive star formation) and good correspondence of dark lanes with CO features (evidence of association of dense material and the bubble). Using the ADSASS heat map , filtered to show "infrared" & "stars," our graduate student creates a custom views showing where studies that would include data on, and discussions of, massive star formation have been done in this region. (Note that the bottom panel of Figure 8 shows a *proxy* for this filtered heat map, where circles simply indicate the most studied sources in this region.) Now, armed with a contextual understanding of the relationship between CO emitting-gas, clusters of massive stars, and emission nebulae in the Cepheus Bubble region, the graduate student is ready to follow up her quick exploration with detailed research. The followup research will involve reading many of the ADSASS astrotagged papers that contributed to the heat map, and the source papers for the relevant images in the ADSASS historical sky layer. Ultimately, as we move toward a future where data and literature are more closely tied, the ADS-hosted papers will also connect back to online databases that allowing the student to do her research more and more "seamlessly."

Some of what the student did in our example above would be possible without the ADSASS. The key question to ask here is: *would* the student–or any researcher–be able to assemble this much context, and glean so much from it, in a such a short time? Starting from an oldAstronomy-solved version of Patel et al.'s CO map, the entire investigation described above would take just a *few minutes* using the ADSASS and WWT. Doing even just the parts of this investigation that are feasible without the ADSASS now would take *hours*, and so our conjecture is that most of the time, researchers proceed without most of the contextual knowledge that would, if readily at-hand, yield many more dramatic scientific results, more quickly.

4. RELEVANCE OF THE PROPOSED WORK

The Astrophysics Data System was initially intended as a data repository, in addition to a literature archive for astronomy (Murray et al. 1992). Our work within the Seamless Astronomy Group (which the PI leads at the Center for Astrophysics) aims to accomplish this early vision, by reuniting astronomical data and publications and making scientific research and communication in Astronomy more seamless. As we discuss in the Overview, we believe that the literature held by ADS represents a **hidden NASA data resource**. With initiatives like the ADS All Sky Survey and oldAstronomy, we aim to extract and make sense of these data. The NASA-sponsored ADS and Microsoft-sponsored WWT already offer fantastic access to NASA-funded astronomy data and literature, and the ADSASS will serve to *integrate* NASA's archival resources in an unprecedented way that will make them even more useful to researchers, students, and the general public.

oldAstronomy is a natural continuation, necessary for **completion of the ADS All Sky Survey**, which began with 2011 ADAP funding. oldAstronomy is built on an established citizen science infrastructure: with millions of visits and registered users, the Zooniverse is by no doubt the largest, most popular, most reliable, and most successful citizen science platform on the web. The Zooniverse infrastructure will make oldAstronomy technically possible, and, at the same time, oldAstronomy will enrich the ecosystem of Zooniverse projects citizen science initiatives overall. While the ADSASS and oldAstronomy are focused solely on astronomical literature for the time being, the tools and workflows which will be built in the context of this proposal are **reusable in, and adaptable to, other disciplines**. The Zooniverse team expects that oldAstronomy will act as a catalyst for similar programs aimed at "liberating" and making sense of image data embedded in articles in fields outside of astronomy, such as the biological and natural sciences and the humanities, which are all currently represented on the Zooniverse platform.

The oldAstronomy initiative is, like many citizen science initiatives, a tremendous resource for **education**, and its output will be a goldmine for the creation of WorldWide Telescope Tours, both about science and about the history of science. WWT Tours are interactive narrated paths through the sky that are created completely within WWT, and they can experienced in a web browser (*Windows OS not required*!). Many sample Tours, including one about "Dust and Us," are available online at https://wwtambassadors.org/wwt/. The PI of this proposal is also PI of the WorldWide Telescope Ambassadors (WWTA, wwtambassadors.org) Program, which uses WWT to teach STEM subjects both in formal (e.g. school classrooms) and informal (e.g. science fairs, online) settings. WWT Ambassadors are trained scientists who volunteer their time to go into schools, attend public events, and create materials online. We expect that many **WWT Ambas-**

sadors will use the results of the ADSASS (including oldAstronomy output) in their presentations, social media updates, and in a set of Tours specifically created to explain ADSASS discoveries to the public. For reference, WWT has been downloaded about 11 million times so far, and the WWT Ambassadors have worked, personally, with about 10,000 learners to date.

5. PLAN OF WORK

Figure 9 shows a Gantt chart detailing our proposed plan of work. Overall, we plan to spend:

1) *the first two quarters of Year 1* developing a working prototype of oldAstronomy on Zooniverse and integrating it, and the WWT HTML5 API, into the existing ADSASS workflow;
2) *the last two quarters of Year 1* launching the beta version of oldAstronomy, collecting preliminary data, and improving the user interface (UI) and experience (UX), by testing the platform on public, and preparing for public launch near the end of Year 1;
3) *the first two quarters of Year 2* promoting oldAstronomy to the public on the Zooniverse platform, and through talks and conferences, collecting both data from citizen scientists and feedback to intermittently improve the platform;
4) *the last two quarters of Year 2,* wrapping up data collection, organizing the data, and integrating outcomes from oldAstronomy into the ADSASS, exposing the historical image layer through links within ADS and within WWT. Throughout the initiative we will present oldAstronomy at conference and workshops, such as the AAS and the dotAstronomy meeting. We also plan to summarize the research results of our effort in an article publication for PLoS One.

The Principal Investigator of this proposal (**PI, Goodman**) will orchestrate the work detailed in Figure 9, in collaboration with a Research Associate (**Co-I, A. Pepe**). Zooniverse staff (**Zoo**), led by **Smith** will work to build the oldAstronomy platform and integrate it with existing ADS technology (**Dev**). Undergraduate Harvard University interns (**UG**) will help Zooniverse staff test and optimize the oldAstronomy platform, and they will help ADS staff (led by **Accomazzi**) test interface options for connecting to ADSASS resources to ADS and WWT. WWT integration work will be done by the developer (**Dev**), in close collaboration with **Fay**, at Microsoft Research. Microsoft and NASA have made significant investments in the Seamless Astronomy and ADS efforts, so it is important to note here that the funding request in this proposal represents only the *leveraging increment needed to establish oldAstronomy and complete the ADS All Sky Survey.*

6. DATA SHARING PLAN

Our Seamless Astronomy[8] group brings together astronomers, computer scientists, librarians, information scientists, data analysts and visualization experts who design, develop and promote the next generation of online astronomical tools. One of the group's primary goals is to increase the effective sharing and use of online data. All of the work on the ADS All Sky Survey and oldAstronomy proposed here *is* effectively a data sharing project, so we will keep this text short. As explained in detail above, oldAstronomy output will be integrated into the NASA ADSASS, the content of which will be made available through NASA's ADS and Microsoft's WWT, both of which are robust, free, online resources accessible by both researchers and the wider public. In addition, the ADSASS layers will be made available through standing agreements (established at the founding of the ADSASS) with CDS, for use in the free Aladin viewer.

---

[8] http://projects.iq.harvard.edu/seamlessastronomy

Figure 9 Gantt chart of proposed oldAstronomy work plan

Task Name	Prmry	Resp	Collab
Core development, working prototype			
Prepare Zooniverse framework	Dev	PI	Zoo
Test/optimize WWT HTML5 API for Zooniverse infrastructure	Dev	PI	Zoo, MSR
Develop image tools: selection, tick marking, contrast	Dev	Co-I	Zoo
Develop image classification system	Dev	Co-I	Zoo
Embed/test oldAstronomy and WWT API in Zooniverse infrastructure	Dev	PI	MSR, Zoo
Develop user interface (UI) and experience (UX)	Dev, UG	Co-I	Zoo, ADS, MSR
Public Beta version			
Test platform with professional and amateur astronomers	Dev, UG	Co-I	Zoo
Improve UI and UX based on feedback	Dev	PI	Zoo
Launch, Release to Users			
Launch oldAstronomy on Zooniverse platform, with WWT API embedded	Dev	PI	Zoo, ADS, MSR
Publicize oldAstronomy via social media, conferences, talks.	PI, Co-I	PI	All
Use feedback from users to constantly improve the platform	Dev	Co-I	Zoo, ADS
Data Collection and Publication			
Collect data on Zooniverse servers	Dev	Co-I	Zoo
Expose astroreferenced images to NASA archives, WWT	Dev	PI	Zoo, MSR, ADS
Integration with ADS All Sky Survey			
Transfer all images not solved automatically by Astrometry.net to Zooniverse	Dev	Co-I	Zoo, ADS
Format metadata for Astrometry.net-calibrated images for compatibility w/ADSASS	Dev	Co-I	Zoo, ADS
Test solution agreement for "standard" images between oldAstronomy & Astrometry.net	Dev	Co-I	Zoo
Coordinate and execute ingestion of images calibrated in oldAstronomy	Dev, UG	Co-I	Zoo
Develop & deploy pipeline to ingest images in "future" ADS deposits (from XML content) into ADSASS	Dev	PI	ADS, Zoo
(Periodically) launch Astrometry.net process on batches of contrast-enhanced images	Dev	Co-I	Zoo
Ingest first bundle of oldAstronomy results into ADSASSS	Dev	Co-I	Zoo, ADS
Ingest second bundle of oldAstronomy results into ADSASSS	Dev	Co-I	Zoo, ADS
Ingest third bundle of oldAstronomy results into ADSASSS	Dev	Co-I	Zoo, ADS
full launch of ADSASS/oldAstronomy results through ADS + WWT	Dev	PI	MSR, ADS
Publications and dissemination			
Meet with Zooniverse team to coordinate launch	PI	PI	All
Publicize oldAstronomy at AAS conference	PI, Co-I	PI	All
Present oldAstronomy at dotAstronomy conference	PI	PI	All
Prepare/publish scholarly article for PLoS One discussing results of oldAstronomy+ADSASS	Co-I, PI, UG	PI	All

7. BIBLIOGRAPHY

Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010. "Astrometry.net: Blind astrometric calibration of arbitrary astronomical images", *The Astronomical Journal,* vol. 137, pp. 1782–1800.

Murray, S.S., Brugel, E.W., Eichhorn, G., Farris, A., Good, J.C., Kurtz, M.J., Nousek, J.A., & Stoner, J.L. 1992, "The NASA Astrophysics Data System: A Heterogeneous Distributed Processing System Application," *Southern Observatory Conference and Workshop Proceedings*, 43, 387.

Miville-Deschenes, M., Lagache, G. 2005, "IRIS: A New Generation of IRAS Maps," *The Astrophysical Journal Supplement Series*, Volume 157, Issue 2, pp. 302-323.

Patel, N.A., Goldsmith, P. F., Heyer, M. H., Snell, R. L., Pratap, P. 1998, "Origin and Evolution of the Cepheus Bubble" *The Astrophysical Journal*, vol. 507, no. 1, pp. 241–253.

Pepe, A., Goodman, A. A., Muench, A. 2012, "The ADS All-Sky Survey". In *Astronomical Data Analysis Software and Systems XXI, ASP Conference Series*.

BIOGRAPHICAL SKETCH FOR ALYSSA A. GOODMAN, MAY 2013

Professional Preparation

Massachusetts Institute of Technology	Physics	Sc.B, 1984
Harvard University	Physics	A.M., 1986
Harvard University	Physics	Ph.D., 1989
University of California, Berkeley	Astronomy	President's Fellow, 1989-92

Appointments

1999-	Professor of Astronomy, Harvard University
2008-2010	Inaugural Scholar-in-Residence, WGBH Boston
2005-2008	Founding Director, Initiative in Innovative Computing, Harvard University
2001-2002	Visiting Fellow, Yale University
1996-1999	Associate Professor of Astronomy, Harvard University
1992-1996	Assistant Professor of Astronomy, Harvard University
1995-1997	Head Tutor, Harvard University Astronomy Department
1995-	Research Associate, Smithsonian Astrophysical Observatory
1989-1992	President's Fellow, University of California, Berkeley

Selected Honors, Awards, and Elected Positions

2009	Fellow, American Association for the Advancement of Science (AAAS)
2008	Microsoft Academic Partner
2007	Apple Science Innovator (one of 10, nationally), Apple Computer
2007	Chair of Astronomy, American Association for the Advancement of Science
1998	Newton Lacy Pierce Prize, American Astronomical Society

Annotated Bibliography

Instead of providing a standard (ADS-style) list of publications, I highlight a small fraction of our group's recent papers where an ADS All-Sky Survey as described in this proposal would have been of especially great value in the study. After each bibliography entry is a "**Q**" followed by questions explaining why the ADSASS, including oldAstronomy results, would have been so relevant.

Goodman, A.A. et al. 2013, *The Bones of the Milky Way*, pre-release version online for comment at https://www.authorea.com/users/23/articles/249/_show_article, to be submitted to ApJ May 2013. **Q** What is the distribution of infrared-dark clouds (IRDCs) on the sky, and is it related to the structure of the Milky Way? What images of IRDCs have been published using NASA Spitzer MIPS/GLIMPSE data, and what do those images look like on all-sky views of the Galaxy's dust distribution?

Beaumont, C.N. Goodman, A.A. Alves, J.A. Lombardi, M., Román-Zúñiga, C.M., Kauffmann, J. and Lada, C.J. 2012, *A simple perspective on the mass-area relationship in molecular clouds*, MNRAS, 423, pp. 2579–2586 **Q** What is the typical angular extent of clouds where the mass-area relationship has been studied? Are they all angularly large, nearby, and thus small, or are some probing much larger size scales?

Arce, H.G. Goodman, A.A., Borkin, M.A., Pineda, J.E., and Beaumont, C.N. 2011. *A Bubbling Nearby Molecular Cloud: COMPLETE Shells In Perseus.* ApJ, 742(2), p.105. **Q** What are the possible progenitors of the expanding shells in Perseus? (Show me a map of the density of papers about young stars near the center of the shells found.) What prior imaging of the Perseus star-forming region has been displayed in the literature–does it show any multi-wavelength evidence for shells?

Goodman, A. A., Pineda, J. E. & Schnee, S. L. 2009, *The "True" Column Density Distribution in Star-Forming Molecular Clouds*, ApJ, 692, 91-103 **Q** Which molecular clouds were mapped for extinction between 2000

and the present, when high-quality NIR data would have been used? Are those clouds at similar height off the Galactic plane?

Goodman, A. A., Rosolowsky, E. W., Borkin, M. A., Foster, J. B., Halle, M., Kauffmann, J. & Pineda, J. E. 2009, *A role for self-gravity at multiple length scales in the process of star formation*, Nature, 457, 63-66 **Q** Have we missed other regions where *all* of the kinds of data (molecular line, sub-mm and mm continuum, near-IR, optical, and Spitzer c2d sources) included in this paper have been studied together? And, if so, are any near Perseus, or in regions like Perseus on the sky, where direct comparisons would be possible?

Rosolowsky, E. W. et al. 2008, *An Ammonia Spectral Atlas of Dense Cores in Perseus*, AJ Supp., 175, 509-521 **Q** Show me the position of every ammonia core ever mapped, on the sky. Are they all in well-studied molecular clouds (defined as clouds having >10 papers written about them within a defined linear extent? Are there clouds that look like they've been "missed" in earlier core surveys and should we suggest targeting them?

Pineda, J. E., Caselli, P. & Goodman, A. A. 2008, *CO Isotopologues in the Perseus Molecular Cloud Complex: the X-factor and Regional Variations*, ApJ, 679, 481-496 **Q** Where has the X-factor (CO to H_2 ratio) been measured on the sky? How significant are large-scale (region-to-region) variations?

Foster, J. B. & Goodman, A. A. 2006, *Cloudshine: New Light on Dark Clouds*, ApJ, 636, L105-L108 **Q** What is the general relationship between regions where dust column density has been mapped out (e.g. using extinction) and known HII regions? How significant is the non-uniform radiation field created by bright local HII regions or B stars in comparison with the general interstellar radiation field in illuminating dark clouds and causing "cloudshine"?

Goodman, A. A. & Arce, H. G. 2004, *PV Cephei: Young Star Caught Speeding?*, ApJ, 608, 831-845 **Q** Show me, on the sky, every paper ever written about stars in the region between NGC 7023 (where PV Ceph is claimed to have come from) and the present location of PV Ceph (10 degrees away)… then filter those according to criteria such "young star" or "B-star."

Management Experience

As (founding) **Director of the Initiative in Innovative Computing (IIC)** at Harvard, AG built a new institution at Harvard to address and answer scientific questions unanswerable without bringing domain scientists and computer scientists into closer collaboration. The IIC created and fostered one dozen new "e-Science" projects at Harvard, all of which are still ongoing. The projects range from mapping out the wiring diagram of the human brain to collaboration systems to astrophysics. AG stepped down from IIC to re-focus on Astronomy and Data Visualization, but she still leads the "Seamless Astronomy" and "Astronomical Medicine" efforts, which grew out of her work at the IIC. The staff of the IIC, which AG supervised when she left, was comprised of more than 30 scientists, technologists, engineers and staff.

Presently, AG leads the "Seamless Astronomy" effort based at the CfA, which is a collaboration amongst 25 people at three institutions, and she was also PI of the COMPLETE Survey of Star-Forming Regions, which involved 20 investigators in four countries. COMPLETE is widely considered one of the most successful coordinated surveys of the physics of star formation to date.

Community Outreach, and Related Activity

AG is principal scientific advisor to the **World Wide Telescope (WWT)** project at Microsoft Research. She is also PI of the "WorldWide Telescope Ambassadors Program" (WWTA; wwtambassadors.org), which uses WWT as a teaching platform online, in schools, and a public science events. WWTA is a collaboration between Harvard and Microsoft Research. The ADSASS, including all the work of oldAstronomy citizen scientists, will be included as a layer in the public release of WWT and therefore distributed for use by the public at large, both informally and through the WWTA program.

BIOGRAPHICAL SKETCH FOR ALBERTO PEPE, MAY 2013

Professional Preparation

University College London, UK	Astrophysics	B.Sc.2002
University College London, UK	Computer Science	M. Sc. 2003
University of California, Los Angeles	Information Studies	Ph. D. 2010
Harvard College Observatory	Astrophysics	Postdoctoral 2013

Appointments

2012-2013	Fellow, Berkman Center for Internet & Society, Harvard University
2010-	Postdoctoral Research Fellow, Center for Astrophysics, Harvard University
2010-	Head Teaching Fellow, School of Engineering and Applied Sciences, Harvard University
2009	Teaching Assistant, Department of Design Media Arts, UC, Los Angeles
2008	Course designer and teaching assistant, Dept. of Information Studies, UC, Los Angeles
2006–2010	Graduate Research Assistant, CENS, Center for Embedded Networked Sensing, LA, CA
2007	Visiting fellow, Digital library Research and Prototyping Group, LANL, Los Alamos, NM
2004–2006	Marie Curie fellow, CERN, Geneva, Switzerland
2004	Research fellow, CINECA, InterUniversity Consortium, University of Bologna, Italy

Selected Honors, Awards, and Elected Positions

2010	American Society for Information Science & Technology, Best Dissertation Award
2006–2010	Microsoft Research, Technical Computing Fellowship.
2004–2006	European Commission, Marie Curie Fellowship.
2004	Consiglio Nazionale delle Ricerche (CNR-Italy), Assegno di Ricerca

Selected Relevant Publications

A measure of total research impact independent of time and discipline. <u>Alberto Pepe</u>, Michael J. Kurtz. PLoS One. Volume 7. Number 11. 2012.

How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads Twitter Mentions and Citations. Xin Shuai, <u>Alberto Pepe</u>, Johan Bollen. PLoS ONE. Volume 7. Number 11. 2012.

The ADS All-Sky Survey. <u>Alberto Pepe</u>, Alyssa Goodman, August Muench. ADASS XXI. 2012.

Collaboration in sensor network research: an in-depth longitudinal analysis of assortative mixing patterns. <u>Alberto Pepe</u>, Marko A. Rodriguez. Scientometrics. Volume 84, Number 3. 2010.

From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. <u>Alberto Pepe</u>, Matthew Mayernik, Christine L. Borgman, Herbert Van De Sompel. Journal of the American Society for Information Science and Technology (JASIST). Volume 61, Issue 3. doi:10.1002/asi.21263. 2010

Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research. Jillian C. Wallis, Christine L. Borgman, Matthew S. Mayernik, Alberto Pepe. International Journal of Digital Curation. Vol. 3. Issue 1. 2008.

Drowning in Data: Digital Library Architecture to Support Scientists' Use of Embedded Sensor Networks. Christine L. Borgman, Jillian C. Wallis, Matthew S. Mayernik, Alberto Pepe. Proceedings of ACM IEEE Joint Conference on Digital Libraries 2007.

Protocols for Scholarly Communication. Alberto Pepe, Joanne Yeomans. Astronomical Society of Pacific Conference Series Vol. 377. San Francisco: ASP, 2007. ISBN 978-1-58381-316-4

Management Experience

Alberto Pepe is the Science-PI of the ADS All-Sky Survey, a NASA ADAP program (NNX12AE11G). In addition of being actively involved in the design and development of the ADSASS All-sky Literature Heatmap and Historical Image layers, AP is also in charge of project management for this program. Duties include hiring staff, managing students, and coordinating workload and tasks with collaborators. AP is also an active member of the "Seamless Astronomy" group at Harvard where he is involved, besides research work, with the organization of workshops and conferences to promote the use of online astronomy tools. In 2011, AP organized, managed and hosted a an invitation-only workshop on the future of scholarly communication entitled Transforming Scholarly Communication. Funded by the Sloan and Moore Foundation and by Microsoft Research, the workshop took place on October 23rd-25th, 2011 at Microsoft's New England Research & Development Center in Cambridge, MA. The workshop hosted 75 experts in scholarly communication from across the academy, the publishing industry, scholarly and learned societies, as well as funding agencies worldwide, to discuss existing and newly developed technologies designed to enhance scholarship and scholarly communication. In 2013, AP is hosting and managing the 2013 dotAstronomy meeting as a local organiser. The event, which brings together an international community of astronomy researchers, developers, educators and communicators to showcase and build upon these web-based astronomy projects, will be held on September 16-18, 2013 at the New England Research and Development Center, in Cambridge, MA.

Dr Arfon McIndoe Smith

Dr Smith is Technical Director of the Citizen Science Alliance, which provides a platform for the development of a wide range of citizen science projects. Since joining the Zooniverse team in December 2008 he has overseen the development of more than 25 difference citizen science experiences. With a background in astrochemistry and software development Dr Smith leads the team of software developers, educators and scientists at the Adler Planetarium in Chicago.

ADDRESS:

Adler Planetarium
1300 South Lake Shore Drive
Chicago, IL 60605

Phone: +1 312 322 0525
E-Mail: arfon@adlerplanetarium.org

PROFESSIONAL PREPARATION:

Undergraduate:	University of Sheffield	Chemistry	BSc	2001
Graduate:	The University of Nottingham	Astrochemsitry	Ph. D.	2006

APPOINTMENTS (Academic)**:**

Adler Planetarium, Chicago	Director of Citizen Science	(2011 –)
University of Oxford	Postdoctoral Researcher	(2008 – 2011)

APPOINTMENTS (Other)**:**

2007 – 2008	Senior software developer, Wellcome Trust Sanger Institute, Cambridge
2009 – Present	Technical Director, Citizen Science Alliance
2009 – Present	Astronomy organizing committee
2010 – Present	Science Hack Day – San Francisco organizing committee

RELATED PUBLICATIONS:

1. **Galaxy Zoo Supernovae :** A. Smith, S. Lynn, M. Sullivan, C. Lintott, P. Nugent, J. Botyanazki, M. Kasliwal, R. Quimby, S. Bamford, L. Fortson, K. Schawinski, I. Hook, S. Blake, P. Podsiadlowski,
2. **Galaxy Zoo : Bar lengths in local disc galaxies,** B. Hoyle, K. Masters, R. Nichol, E. Edmondson, A. Smith, C. Lintott, R. Scranton, S. Bamford, K. Schawinski, D. Thomas MNRAS, 2011, 415, 3627
3. **Galaxy Zoo : Bars in Disc Galaxies,** K. Masters, R. C. Nichol, B. Hoyle, C. Lintott, S. Bamford, E. Edmondson, L. Fortson, W. Keel, K. Schawinski, A. Smith, D.Thomas MNRAS, 2011, 411, 2026
4. **Galaxy Zoo : A sample of blue elliptical galaxies at low redshift** K. Schawinski, C. Lintott, D. Thomas, M. Sarzi, D. Andreescu, S. Bamford, S. Kaviraj, S. Khocfar, K. Land, P. Murray, R. Nichol, J. Raddick, A. Slosar, A. Szalay, J Vandenberg, K. Sukyoung, MNRAS, 396, 818
5. **Galaxy Zoo : The properties of merging galaxies in the nearby Universe – local environments, colours, masses, star-formation rates and AGN activity**, D. Darg, S. Kaviraj, C. Lintott, K. Schawinski, M. Sarzi, S. Bamford, J. Silk, D. Andreescu, P.Murray, R. Nichol, M. Raddick, A. Slosar, A. Szalay, D. Thomas, J. Vandenberg, MNRAS, 401, 1552