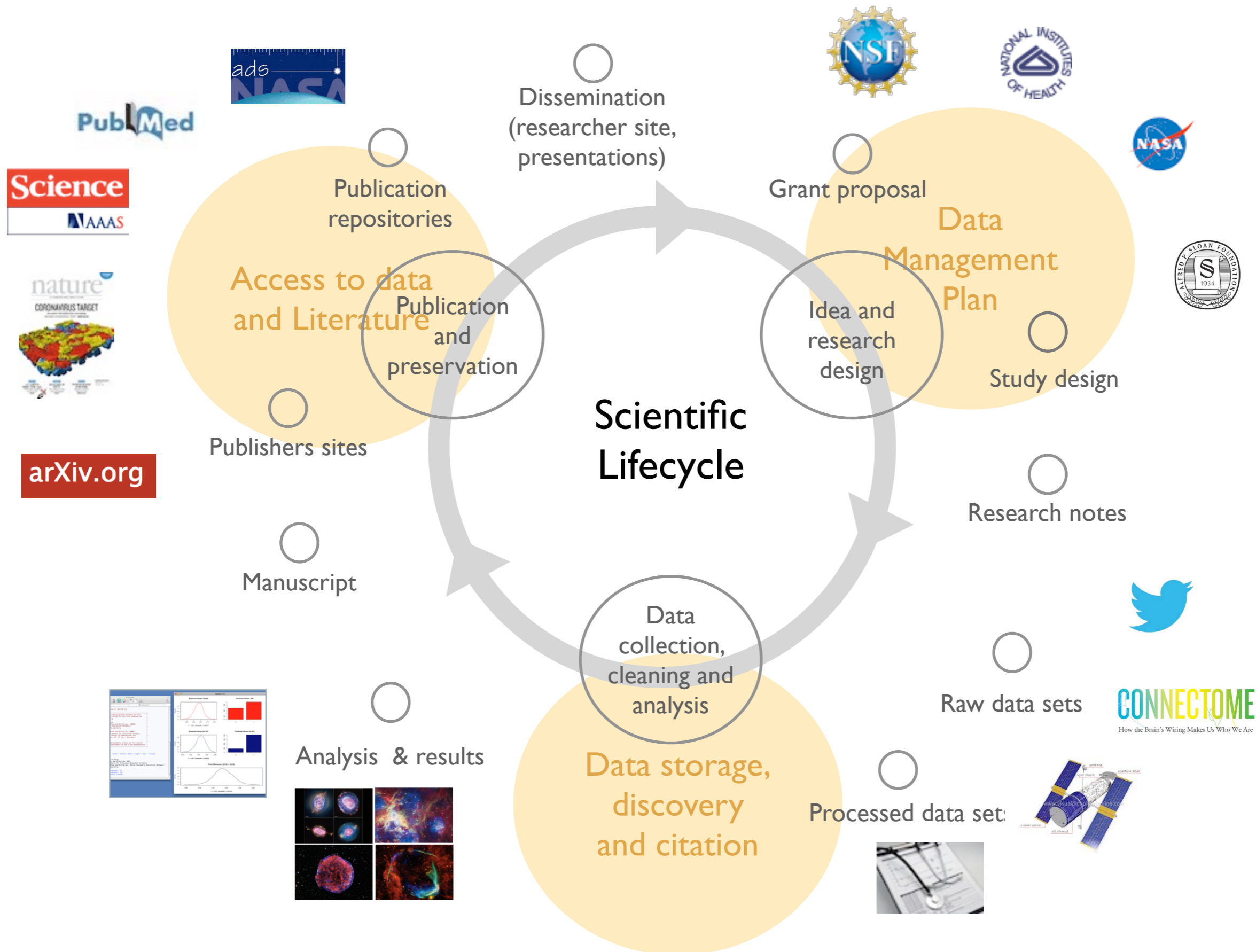


# DATA CITATION: PRINCIPLES & PRACTICES

Mercè Crosas, Ph.D.  
IQSS, Harvard University  
 @mercecrosas

Micah Altman, Ph.D.  
MIT  
 @drmaltman

Beyond the PDF 2, Amsterdam, March 19-20



## THE COMPLETE SURVEY OF OUTFLOWS IN PERSEUS

HÉCTOR G. ARCE<sup>1</sup>, MICHELLE A. BORKIN<sup>2</sup>, ALYSSA A. GOODMAN<sup>3</sup>, JAIME E. PINEDA<sup>3</sup>, AND MICHAEL W. HALLE<sup>4,5</sup><sup>1</sup> Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520, USA; [hector.arce@yale.edu](mailto:hector.arce@yale.edu)<sup>2</sup> School of Engineering and Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138, USA<sup>3</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA<sup>4</sup> Surgical Planning Laboratory, Department of Radiology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA<sup>5</sup> Initiative in Innovative Computing, Harvard University, 60 Oxford Street, Cambridge, MA 02138, USA*Received 2009 October 7; accepted 2010 April 9; published 2010 May 7*

## ABSTRACT

We present a study on the impact of molecular outflows in the Perseus molecular cloud complex using the COMPLETE Survey large-scale  $^{12}\text{CO}(1-0)$  and  $^{13}\text{CO}(1-0)$  maps. We used three-dimensional isosurface models generated in right ascension–declination–velocity space to visualize the maps. This rendering of the molecular line data allowed for a rapid and efficient way to search for molecular outflows over a large ( $\sim 16 \text{ deg}^2$ ) area. Our outflow-searching technique detected previously known molecular outflows as well as new candidate outflows. Most of these new outflow-related high-velocity features lie in regions that have been poorly studied before. These new outflow candidates more than double the amount of outflow mass, momentum, and kinetic energy in the Perseus cloud complex. Our results indicate that outflows have significant impact on the environment immediately surrounding localized regions of active star formation, but lack the energy needed to feed the observed turbulence in the *entire* Perseus complex. This implies that other energy sources, in addition to protostellar outflows, are responsible for turbulence on a global cloud scale in Perseus. We studied the impact of outflows in six regions with active star formation within Perseus of sizes in the range of 1–4 pc. We find that outflows have enough power to maintain the turbulence in these regions and enough momentum to disperse and unbind some mass from them. We found no correlation between outflow strength and star formation efficiency (SFE) for the six different regions we studied, contrary to results of recent numerical simulations. The low fraction of gas that potentially could be ejected due to outflows suggests that additional mechanisms other than cloud dispersal by outflows are needed to explain low SFEs in clusters.

*Key words:* ISM: clouds – ISM: individual objects (Perseus) – ISM: jets and outflows – ISM: kinematics and dynamics – stars: formation – turbulence

*Online-only material:* color figures

in IC 348 (HD 281159) is confirmed to reside in the Perseus cloud, but there might be a few other high-mass stars that interact with the cloud (through their winds and/or UV radiation) even though they were not necessarily formed in the cloud complex (see, e.g., Walawender et al. 2004; Ridge et al. 2006a; Kirk et al. 2006; Rebull et al. 2007). There is also a large number of nebulous objects associated with outflow shocks (i.e., HH objects and H<sub>2</sub> knots) that have been identified in the cloud complex (Bally et al. 1996b, 1997; Yan et al. 1998; Walawender et al. 2005b; Davis et al. 2008).

The whole Perseus region was first surveyed in <sup>12</sup>CO by Sargent (1979), and since then has been mapped in CO at different angular resolutions (all with beams > 1') by a number of other authors (e.g., Bachiller & Cernicharo 1986; Ungerechts & Thaddeus 1987; Padoan et al. 1999; Sun et al. 2006). These maps show a clear velocity gradient in the Perseus molecular cloud complex where the central cloud (LSR) velocity increases from about 4.5 km s<sup>-1</sup> at the western edge of the cloud to about 10 km s<sup>-1</sup> at the eastern end. The large velocity gradient in the gas across the entire complex and the fact that different parts of the Perseus cloud appear to have different distances (see above) could possibly indicate that the complex is made up of a superposition of different entities. Recently, the Perseus molecular cloud complex was also observed (and studied) in its entirety in the mid- and far-infrared as part of the “From Molecular Cores to Planet-forming Disks” (aka c2d) *Spitzer* Legacy Project (Jørgensen et al. 2006; Rebull et al. 2007; Evans et al. 2009).

## 2. DATA

In this paper, we use the <sup>12</sup>CO(1–0) and <sup>13</sup>CO(1–0) data collected for Perseus as part of the COordinated Molecular Probe Line Extinction Thermal Emission (COMPLETE) Survey of Star Forming Regions,<sup>6</sup> described in detail by Ridge et al. (2006b). The <sup>12</sup>CO and <sup>13</sup>CO molecular line maps were observed between 2002 and 2005 using the 14 m Five College Radio Astronomy Observatory (FCRAO) telescope with the SE-QUOIA 32-element focal plane array. The receiver was used with a digital correlator providing a total bandwidth of 25 MHz over 1024 channels. The <sup>12</sup>CO  $J = 1-0$  (115.271 GHz) and the <sup>13</sup>CO  $J = 1-0$  (110.201 GHz) transitions were observed simultaneously using an on-the-fly (OTF) mapping technique. The beam telescope at these frequencies is about 46". Both maps of <sup>12</sup>CO and <sup>13</sup>CO are essential for a thorough study of the outflow and cloud properties. The <sup>12</sup>CO(1–0) is a good tracer of the cool and massive molecular outflows and provides the information needed to study the impact of these energetic phenomena on the cloud. The <sup>13</sup>CO(1–0) provides an estimate of the optical depth of the <sup>12</sup>CO(1–0) line and can be used to probe the cloud structure and kinematics.

Observations were made in 10' × 10' maps with an effective velocity resolution of 0.07 km s<sup>-1</sup>. These small maps were then patched together to form the final large map of Perseus, which is about 6:25 × 3°. Calibration was done via the chopper-wheel technique (Kutner & Ulich 1981), yielding spectra with units of  $T_A^*$ . We removed noisy pixels that were more than 3 times the average rms noise of the data cube, the entire map was then resampled to a 46" grid, and the spectral axis was Hanning smoothed<sup>7</sup> (necessary to keep the cubes to a size manageable by

the three-dimensional visualization code, see below). During the observations of the Perseus cloud, different OFF positions were used depending on the location that was being mapped. Some of these OFF positions had faint, though significant, emission which resulted in an artificial absorption feature in the final spectra. Gaussians were fitted to the negative feature in regions with no gas emission, and the fits were then used to correct for the contaminating spectral component. The resulting mean 3 $\sigma$  rms per channel in the <sup>12</sup>CO and <sup>13</sup>CO maps are 0.25 and 0.20 K, respectively, in the  $T_A^*$  scale. Spectra were corrected for the main beam efficiencies of the telescope (0.49 and 0.45 at 110 and 115 GHz, respectively), obtained from measurements of Jupiter.

## 3. COMPUTATIONAL MOTIVATION AND THREE-DIMENSIONAL VISUALIZATION

This study allows for a test of the effectiveness of three-dimensional visualization of molecular line data of molecular clouds in R.A.–decl.–velocity ( $p$ – $p$ – $v$ ) space as a way to identify velocity features, such as outflows, in large maps.<sup>8</sup> The primary program used for three-dimensional visualization is 3D Slicer,<sup>9</sup> which was developed originally at the MIT Artificial Intelligence Laboratory and the Surgical Planning Lab at Brigham and Women’s Hospital. It was designed to help surgeons in image-guided surgery, to assist in pre-surgical preparation, to be used as a diagnostic tool, and to help in the field of brain research and visualization (Gering 1999). The 3D Slicer was first used with astronomical data by Borkin et al. (2005) to study the hierarchical structure of star-forming cores and velocity structure of IC 348 with <sup>13</sup>CO(1–0) and C<sup>18</sup>O(1–0) data.

We divided the Perseus cloud into six areas (with similar cloud central LSR velocities) for easier visualization and outflow search in 3D Slicer (see below). The borders of these areas are similar to those named by Pineda et al. (2008), who also based their division mainly on the cloud’s central LSR velocity. The regions, whose outlines are shown in Figure 1, overlap between 1 and 3 arcmin to guarantee complete analysis. This overlap was checked to be sufficient based on the fact that new and known outflows which crossed regions were successfully double-identified.

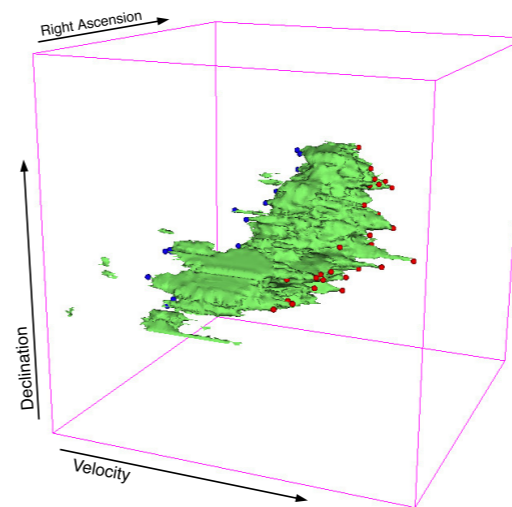
For each area, an isosurface (constant intensity level) model was generated in 3D Slicer, using the <sup>12</sup>CO(1–0) map. The threshold emission intensity level chosen for each isosurface model was the lowest level of emission above the rms noise level for that particular region. This creates a three-dimensional model representing all of the detected emission. The high-velocity gas in this three-dimensional space can be identified in the form of spikes, as shown for the B5 region in Figure 2, which visually stick out from the general distribution of the gas. These sharp protrusions occur since one is looking at the radial velocity component of the gas along the line of sight, thus causing spikes wherever there is gas at distinct velocities far away from the main cloud velocity. Instead of having to go through each region and carefully examine each channel map, or randomly scroll through the spectra by hand, this visualization allows one to instantly see where the high-velocity points are located (see also Borkin et al. 2007, 2008).

<sup>8</sup> This work is done as part of the Astronomical Medicine project (<http://am.iic.harvard.edu>) at the Initiative in Innovative Computing at Harvard (<http://iic.harvard.edu>). The goal of the project is to address common research challenges to both the fields of medical imaging and astronomy including visualization, image analysis, and accessibility of large varying kinds of data.

<sup>9</sup> <http://www.slicer.org/>

<sup>6</sup> See <http://www.cfa.harvard.edu/COMPLETE>.

<sup>7</sup> See <http://www.cfa.harvard.edu/COMPLETE/projects/outflows.html> for a link to the molecular line maps.

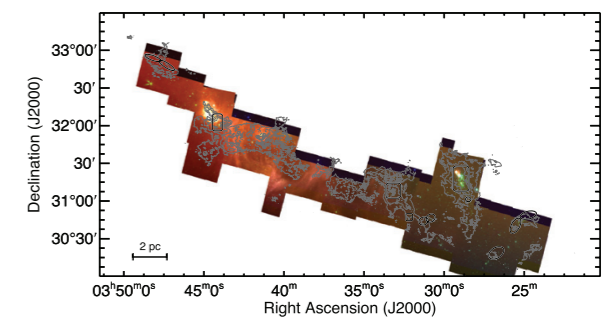


**Figure 2.** Three-dimensional rendering of the molecular gas in B5 (i.e., Area VI in Figure 1), using 3D Slicer. The gray (green) isosurface model shows the <sup>12</sup>CO emission in position–position–velocity space. The small circles show the locations of identified high-velocity points (with the color in the online version representing whether the point is blue- or red-shifted).

(A color version of this figure is available in the online journal.)

## 4. OUTFLOW IDENTIFICATION

A total of 218 high-velocity points were visually identified in 3D Slicer for all of Perseus in <sup>12</sup>CO. We checked the position of each high-velocity point against the locations of known outflows (based on an extensive literature search) to determine if the point is associated with any known molecular outflow. From the 218 high-velocity points found, a total of 36 points were identified as associated with known molecular outflows. Figure 3 shows the approximate regions where previously known <sup>12</sup>CO(1–0) outflows lie. The number of high-velocity points associated with a single outflow varies depending on its size and intensity. For example, the parsec-scale B5 IRS1 outflow is a conglomerate of six high-velocity points whereas the HH 211 outflow, which is only ~0.1 pc long, is identified by only one point. We inspected each of the remaining 182 high-velocity points to verify whether they are outflow related or caused by other velocity features in the cloud. To determine if a high-velocity point is outflow related, we checked the spectrum by eye to look for outflow traits (e.g., high-velocity low-intensity wings) and verified its proximity to known outflows and outflow sources (Wu et al. 2004), HH objects (Walawender et al. 2005b), H<sub>2</sub> knots (Davis et al. 2008), candidate young stellar objects (YSOs) from the c2d *Spitzer* survey (Evans et al. 2009) and other known outflow sources and YSOs. We also checked the velocity distribution and morphology of the gas associated with each high-velocity point to verify whether the velocity and structure of the gas were significantly different from that of the cloud in that region. From the remaining 182 high-velocity points found, a total of 60 points were classified as being outflow candidates based on the criteria mentioned above. For 97% of these outflow candidates, the maximum velocity away from the cloud velocity is equal to or greater than the escape velocity in that region of the cloud. We note that we purposely chose not to be too restrictive in the definition of outflow candidate (e.g., we identified outflow candidates even without a solid outflow source identification, see



**Figure 3.** *Spitzer* IRAC (color) image of the c2d coverage of the Perseus cloud made from 3.6, 4.5, and 8.0  $\mu$ m images of the region (Evans et al. 2009). The color code is blue (3.6  $\mu$ m), green (4.5  $\mu$ m), and red (8.0  $\mu$ m). Ellipses and squares with rounded corners show the approximate regions where previously known outflows in Perseus lie. The gray contours show the 4 K km s<sup>-1</sup> level of the <sup>13</sup>CO(1–0) integrated intensity map (not corrected for the FCRAO beam efficiency).

(A color version of this figure is available in the online journal.)

below). Using our broad, yet realistic, definition we can calculate the maximum possible impact from all plausible molecular outflows to the cloud. Out of the remaining 122 points, 17 points were discarded due to too much noise or being pixels cut off by the map’s edge and the other 105 points are thought to be caused by a number of other kinematic phenomena, including clouds at other velocities in the same line of sight unrelated to the Perseus cloud and spherical winds from young stars that produce expanding shell-like structures in the molecular gas (as opposed to the discrete blob morphology observed in the 60 outflow candidates). The distribution and impact of these expanding shells on the cloud will be discussed further in a subsequent paper (H. G. Arce et al. 2011, in preparation).

We visually inspected the velocity maps in the area surrounding each of the 60 high-velocity points identified as outflow in origin (but unrelated to known outflows) and chose an area (in R.A.–decl. space) and velocity range that included all or most of the emission associated with the kinetic feature. The integration area and velocity ranges were conservatively chosen to include only the emission visibly associated with the outflowing material, thus avoiding cloud emission. The high-velocity gas associated with these 60 points shows discrete morphologies in area and velocity. Hereafter each of these high-velocity features is referred to as a “COMPLETE Perseus Outflow Candidate” (CPOC) and we list their positions and other properties in Table 1.<sup>10</sup> In Figure 4, we show the velocity ranges of all CPOCs, in comparison with their local cloud (LSR) velocity.

Our outflow-detection technique proved to be reliable, as we detect high-velocity gas associated with all published CO(1–0) outflows (see Figure 3). However, it is very probable that the catalog of new molecular outflows generated for this paper is an underestimate of the true number of previously undetected molecular outflows due to the resolution of the CO maps and other limitations of our outflow-detection technique. Unknown outflows that are smaller than the beam size of our map (i.e., 0.06 pc at the assumed distance of Perseus) or that have weak high-velocity wings (i.e., with intensities less than twice the rms of the spectra at that particular position) cannot be detected by our technique. Outflows with maximum velocities too close to

<sup>10</sup> See <http://www.cfa.harvard.edu/COMPLETE/projects/outflows.html> for a link to the fits cubes and the integrated intensity fits files of the CPOCs, as well as a list of the YSO candidates, HH objects, and H<sub>2</sub> knots in the cloud.

in IC 348 (HD 281159) is confirmed to reside in the Perseus cloud, but there might be a few other high-mass stars that interact with the cloud (through their winds and/or UV radiation) even though they were not necessarily formed in the cloud complex (see, e.g., Walawender et al. 2004; Ridge et al. 2006a; Kirk et al. 2006; Rebull et al. 2007). There is also a large number of nebulous objects associated with outflow shocks (i.e., HH objects and H<sub>2</sub> knots) that have been identified in the cloud complex (Bally et al. 1996b, 1997; Yan et al. 1998; Walawender et al. 2005b; Davis et al. 2008).

The whole Perseus region was first surveyed in <sup>12</sup>CO by Sargent (1979), and since then has been mapped in CO at different angular resolutions (all with beams > 1') by a number of other authors (e.g., Bachiller & Cernicharo 1986; Ungerechts & Thaddeus 1987; Paduan et al. 1999; Sun et al. 2006). These maps show a clear velocity gradient in the Perseus molecular cloud complex where the central cloud (LSR) velocity increases from about 4.5 km s<sup>-1</sup> at the western edge of the cloud to about 10 km s<sup>-1</sup> at the eastern end. The large velocity gradient in the gas across the entire complex and the fact that different parts of the Perseus cloud appear to have different distances (see above) could possibly indicate that the complex is made up of a superposition of different entities. Recently, the Perseus molecular cloud complex was also observed (and studied) in its entirety in the mid- and far-infrared as part of the "From Molecular Cores to Planet-forming Disks" (aka c2d) *Spitzer* Legacy Project (Jørgensen et al. 2006; Rebull et al. 2007; Evans et al. 2009).

## 2. DATA

In this paper, we use the <sup>12</sup>CO(1-0) and <sup>13</sup>CO(1-0) data collected for Perseus as part of the COordinated Molecular

Probe 1  
vey of  
et al. (2  
observe  
Radio A  
QUOIA  
with a c  
over 10  
<sup>12</sup>CO J  
taneous

beam telescope at these frequencies is about 46". Both maps of <sup>12</sup>CO and <sup>13</sup>CO are essential for a thorough study of the outflow and cloud properties. The <sup>12</sup>CO(1-0) is a good tracer of the cool and massive molecular outflows and provides the information needed to study the impact of these energetic phenomena on the cloud. The <sup>13</sup>CO(1-0) provides an estimate of the optical depth of the <sup>12</sup>CO(1-0) line and can be used to probe the cloud structure and kinematics.

Observations were made in 10' × 10' maps with an effective velocity resolution of 0.07 km s<sup>-1</sup>. These small maps were then patched together to form the final large map of Perseus, which is about 6.25 × 3'. Calibration was done via the chopper-wheel technique (Kutner & Ulich 1981), yielding spectra with units of T<sub>b</sub>. We removed noisy pixels that were more than 3 times the average rms noise of the data cube, the entire map was then resampled to a 46" grid, and the spectral axis was Hanning smoothed<sup>6</sup> (necessary to keep the cubes to a size manageable by

the three-dimensional visualization code, see below). During the observations of the Perseus cloud, different OFF positions were used depending on the location that was being mapped. Some of these OFF positions had faint, though significant, emission which resulted in an artificial absorption feature in the final spectra. Gaussians were fitted to the negative feature in regions with no gas emission, and the fits were then used to correct for the contaminating spectral component. The resulting mean 3σ rms per channel in the <sup>12</sup>CO and <sup>13</sup>CO maps are 0.25 and 0.20 K, respectively, in the T<sub>b</sub> scale. Spectra were corrected for the main beam efficiencies of the telescope (0.49 and 0.45 at 110 and 115 GHz, respectively), obtained from measurements of Jupiter.

## 3. COMPUTATIONAL MOTIVATION AND THREE-DIMENSIONAL VISUALIZATION

This study allows for a test of the effectiveness of three-dimensional visualization of molecular line data of molecular clouds in R.A.–decl.–velocity (*p–p–v*) space as a way to identify velocity features, such as outflows, in large maps.<sup>6</sup> The primary program used for three-dimensional visualization is 3D Slicer<sup>7</sup> which was developed originally at the MIT Artificial Intelligence Laboratory and the Surgical Planning Lab at Brigham and Women's Hospital. It was designed to help surgeons in image-guided surgery, to assist in pre-surgical preparation, to be used as a diagnostic tool, and to help in the field of brain research and visualization (Gering 1999). The 3D Slicer was first used with astronomical data by Borkin et al. (2005) to study the hierarchical structure of star-forming cores and velocity structure of IC 348 with <sup>12</sup>CO(1-0) and C<sup>18</sup>O(1-0) data.

We divided the Perseus cloud into six areas (with similar cloud central LSR velocities) for easier visualization and outflow

threshold emission intensity level chosen for each isosurface model was the lowest level of emission above the rms noise level for that particular region. This creates a three-dimensional model representing all of the high-velocity gas in this three-dimensional space in the form of spikes, as shown in Figure 2, which visually stick out from the main cloud. These sharp protrusions occur in regions where the radial velocity component of the outflow is thus causing spikes whenever the outflow is far away from the main cloud velocity. We randomly scroll through each region and carefully inspect the data to see where the high-velocity points are located (see also Borkin et al. 2007, 2008).

<sup>6</sup> This work is done as part of the Astronomical Medicine project (<http://amc.harvard.edu>) at the Initiative in Innovative Computing at Harvard (<http://iic.harvard.edu>). The goal of the project is to address common research challenges to both the fields of medical imaging and astronomy including visualization, image analysis, and accessibility of large varying kinds of data.  
<sup>7</sup> <http://www.slicer.org/>

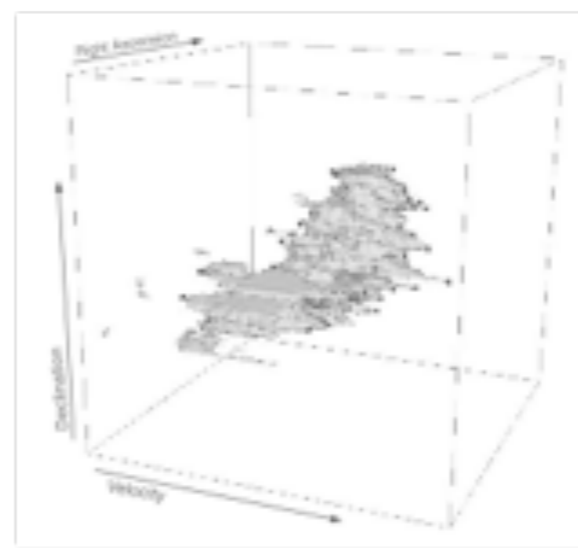


Figure 2. Three-dimensional rendering of the molecular gas in R.A.–decl.–velocity (*p–p–v*) space (i.e., Area VI in Figure 1), using 3D Slicer. The gray (green in color) isosurface model shows the <sup>12</sup>CO emission in position–position–velocity space. The small circles show the locations of identified high-velocity points (with the color in the online version representing whether the point is blue- or red-shifted).

(A color version of this figure is available in the online journal.)

## 4. OUTFLOW IDENTIFICATION

A total of 218 high-velocity points were visually identified in 3D Slicer for all of Perseus in <sup>12</sup>CO. We checked the position of each high-velocity point against the locations of known outflows (based on an extensive literature search) to determine if the point

was a molecular outflow. From the 218 total of 36 points were identified as molecular outflows. Figure 3 shows the locations of these outflows, which were previously known <sup>12</sup>CO(1-0) high-velocity points associated with outflows, including on its size and intensity. For the HH 211 outflow is a conglomerate of outflows, which is not identified by only one point. We inspected each high-velocity point to verify whether

they are outflow related or caused by other velocity features in the cloud. To determine if a high-velocity point is outflow related, we checked the position, but we do not look for outflow

60 points were classified as being outflow candidates based on the criteria mentioned above. For 97% of these outflow candidates, the maximum velocity away from the cloud velocity is equal to or greater than the escape velocity in that region of the cloud. We note that we purposely chose not to be too restrictive in the definition of outflow candidate (e.g., we identified outflow candidates even without a solid outflow source identification, see

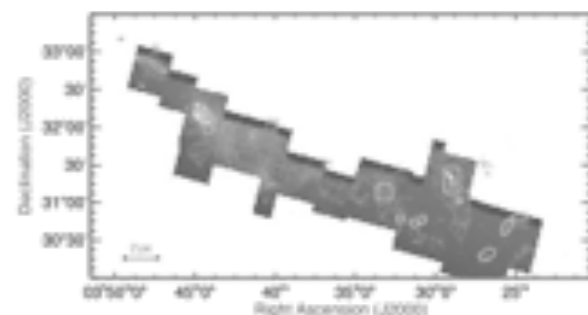


Figure 3. *Spitzer* IRAC color image of the c2d coverage of the Perseus cloud made from 3.6, 4.5, and 8.0 μm images of the region (Evans et al. 2009). The color code is blue (3.6 μm), green (4.5 μm), and red (8.0 μm). Ellipses and squares with rounded corners show the approximate regions where previously known outflows in Perseus lie. The gray contours show the 4 K km s<sup>-1</sup> level of the <sup>12</sup>CO(1-0) integrated intensity map (not corrected for the FCRAO beam efficiency).

(A color version of this figure is available in the online journal.)

below). Using our broad, yet realistic, definition we can calculate the maximum possible impact from all plausible molecular outflows to the cloud. Out of the remaining 122 points, 17 points were discarded due to too much noise or being pixels cut off by the map's edge and the other 105 points are thought to be caused by a number of other kinematic phenomena, including clouds at other velocities in the same line of sight unrelated to the Perseus cloud and spherical winds from young stars that produce expanding shell-like structures in the molecular gas (as opposed to the discrete blob morphology observed in the 60 outflow candidates). The distribution and impact of these expanding shells on the cloud will be discussed further in a subsequent paper (H. G. Arce et al. 2011, in preparation).

We visually inspected the velocity maps in the area surrounding each of the 60 high-velocity points identified as outflow in origin (but unrelated to known outflows) and chose an area (in R.A.–decl. space) and velocity range that included all or most of the emission associated with the kinetic feature. The integration area and velocity ranges were conservatively chosen to include only the emission visibly associated with the outflowing material, thus avoiding cloud emission. The high-velocity gas associated with these 60 points shows discrete morphologies in area and velocity. Hereafter each of these high-velocity features is referred to as a "COMPLETE Perseus Outflow Candidate" (CPOC) and we list their positions and other properties in

Table 1. We list the positions of all CPOCs, as well as the positions of all CPOCs.

city. We note that this is not a complete list of all CPOCs, as we only inspected the area around the outflows. It is possible that there are other outflows in the cloud that were not detected in this paper. We note that we purposely chose not to be too restrictive in the definition of outflow candidate (e.g., we identified outflow candidates even without a solid outflow source identification, see below). Outflows with maximum velocities too close to the cloud velocity are not detected by our technique. Outflows with maximum velocities too close to

<sup>8</sup> See <http://www.cfa.harvard.edu/COMPLETE/projects/outflows.html> for a link to the fits cubes and the integrated intensity fits files of the CPOCs, as well as a list of the YSO candidates, HH objects, and H<sub>2</sub> knots in the cloud.

<sup>6</sup> See <http://www.cfa.harvard.edu/COMPLETE>.

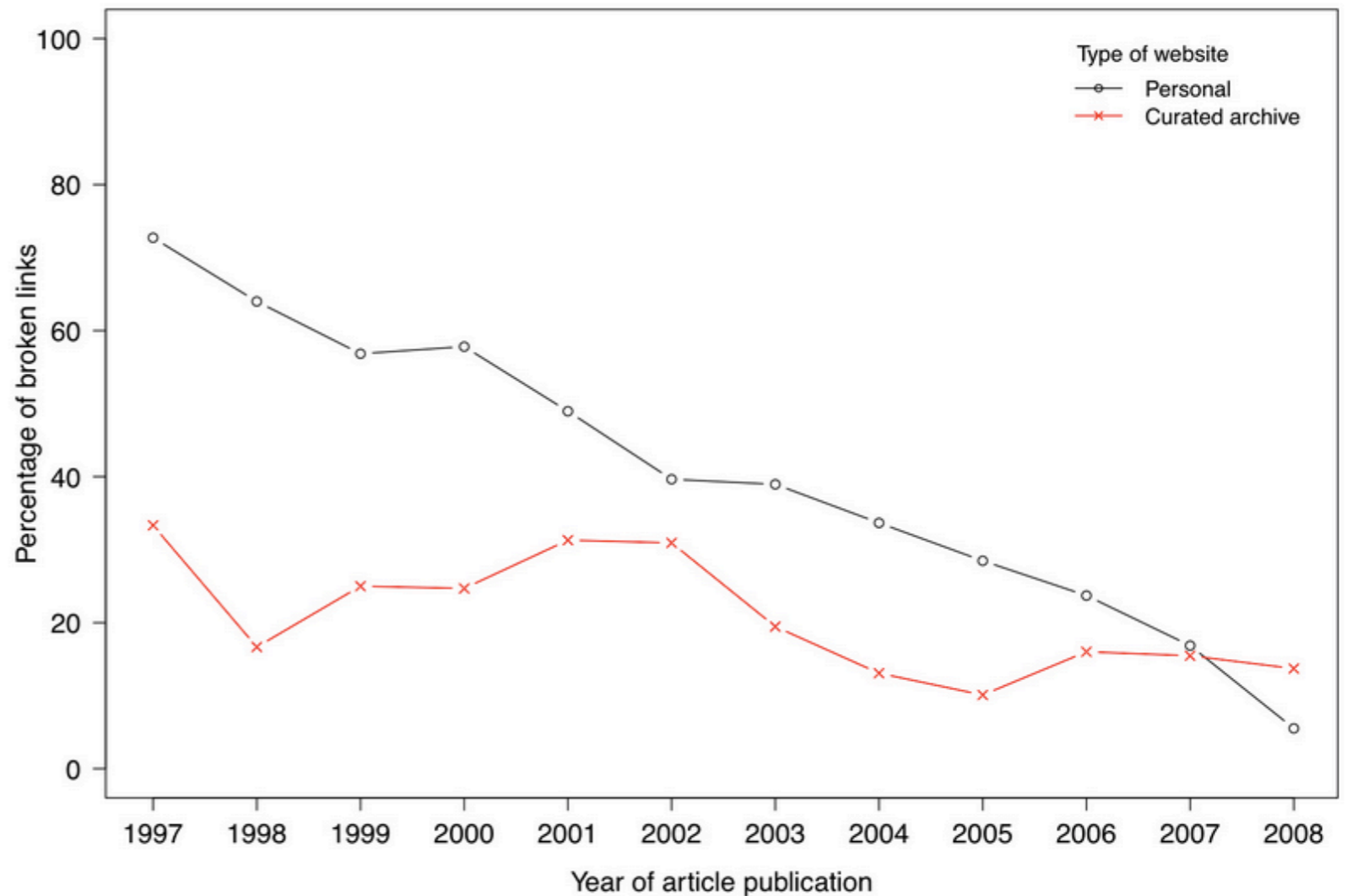
<sup>7</sup> See <http://www.cfa.harvard.edu/COMPLETE/projects/outflows.html> for a link to the molecular line maps.

<sup>10</sup> See <http://www.cfa.harvard.edu/COMPLETE/projects/outflows.html> for a link to the fits cubes and the integrated intensity fits files of the CPOCs, as well as a list of the YSO candidates, HH objects, and H<sub>2</sub> knots in the cloud.



# PROBLEMS

- No standard guidelines for data citation
- No proper attribution, lacking persistent author identifiers
- Links to personal sites broken after some time



Pepe, Goodman, Muench, Crosas & Erdmann, "Handling, Archiving and Citing Data in Astronomy," Forthcoming

From a corpus of 7641 publications in Astronomy:  
> 50% of links published prior to 2001 are broken

# Data Citation Principles Workshop

May 16 – May 17, 2011, IQSS at Harvard University



The goal of the Data Citation Principles workshop is to facilitate collaboration among leaders in publishing, data archiving and data citation research and to articulate common answers to core questions such as: Should data that is used to support published research results be cited in peer-reviewed articles presenting those results? Should each citation include sufficient information to allow a unique dataset to be identified over time? Should citations to data appear in a consistent place within publication? Should citations to data be indexable so that they can be used for linkage and impact analysis?



“The first principle for citing data, as identified by workshop participants was *data citations should be first class objects for publication.*”





# ADDITIONAL PRINCIPLES

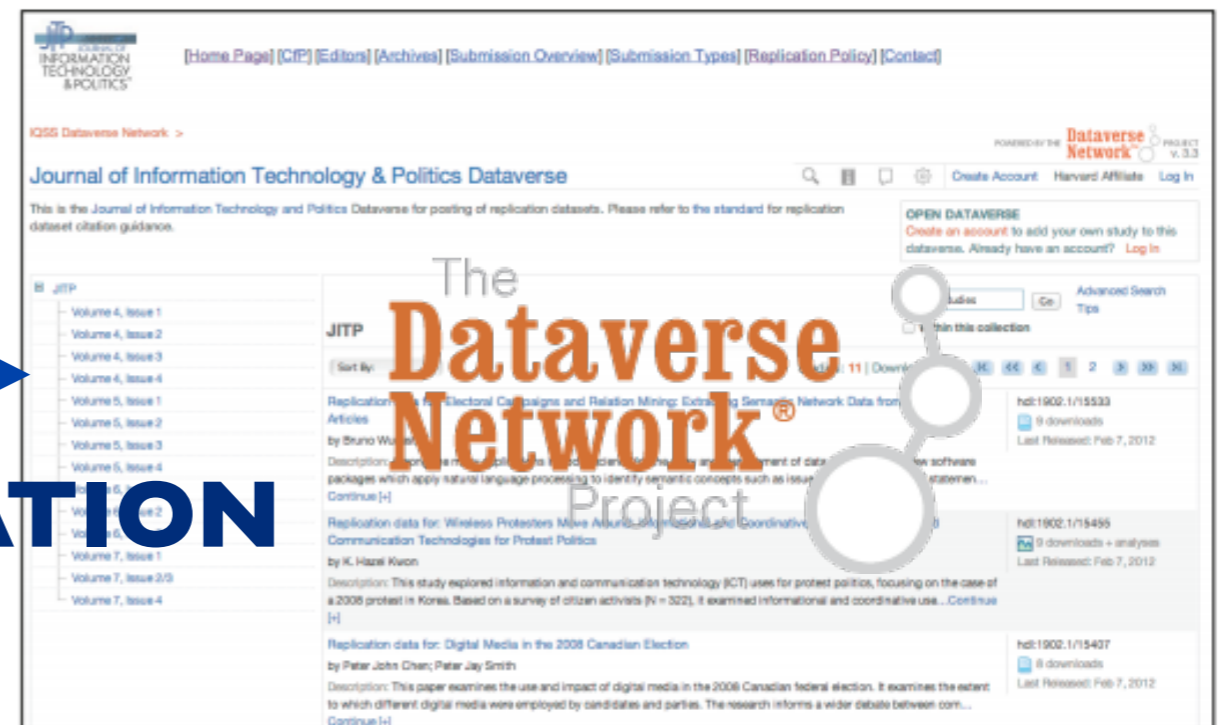
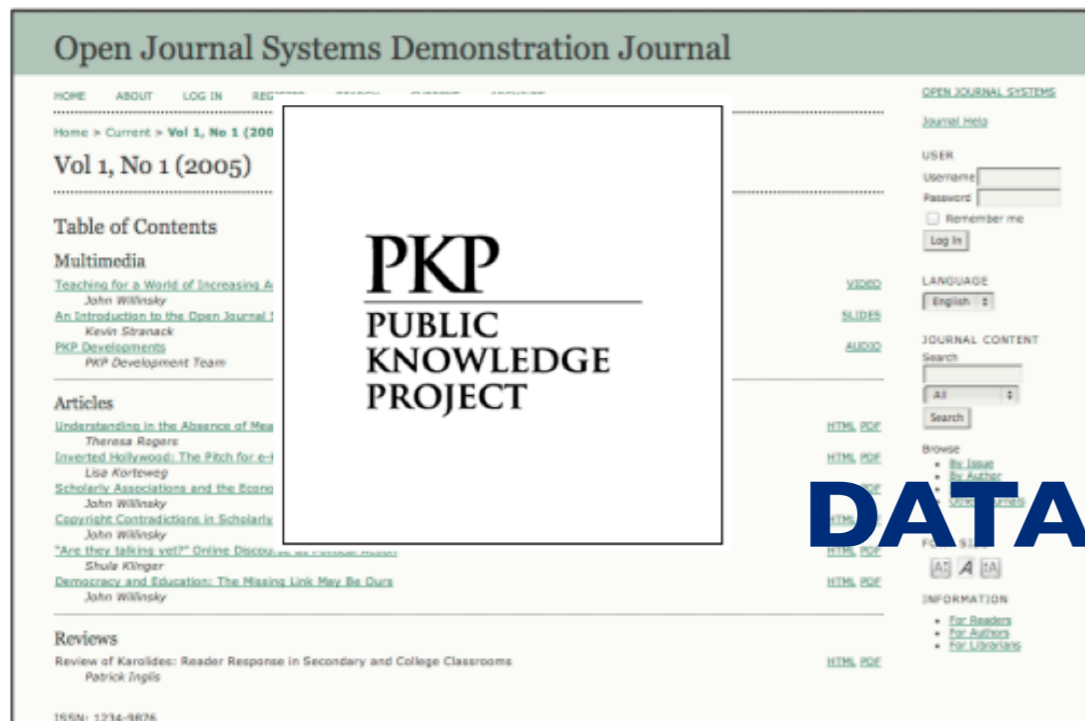
At minimum, all data necessary to understand, assess, extend conclusions in scholarly work should be cited.

- Citations should persist and enable access to fixed version of data at least as long as the citing work exists.
- Data citation should support unambiguous attribution of credit to all contributors



# PRACTICES

A two year project to connect papers to data with PKP, IQSS @ Harvard, Micah Altman @ MIT



DATA CITATION

- Authors submit paper + research data through PKP's Open Journal System (OJS)
- 40 journals participating in the pilot project to test plugin
- Research data + metadata automatically deposited to journal's dataverse
- SWORD API from OJS to Dataverse

# DATA CITATION

A data citation standard on the basis of provenance, replicability and attribution, generated automatically by the Dataverse:

*Authors, Year, Title, **Persistent Identifier (handle/DOI)**,  
**Universal Numerical Fingerprint (UNF)**, Distributor, Version,  
[+ Optional Fields]*

Altman & King, 2007 “A Proposed Standard for the Scholarly Citation of Quantitative Data”

# WHICH WORKFLOW?

1



Submit  
paper



Approve  
paper +  
ask for  
data



Publish paper  
in OJS +  
release data in  
Dataverse

2



Submit  
paper +  
data



Approve  
paper +  
review  
data



Publish paper  
in OJS +  
release data in  
Dataverse

3



Submit  
paper +  
data



Reject  
paper +  
review  
data



Release data in  
Dataverse

+ 4, 5, 6, ....

More at:

<http://projects.iq.harvard.edu/ojs-dvn>

<http://thedata.org>

Mercè Crosas

<http://iq.harvard.edu/merce-crosas>

 @mercecrosas

Micah Altman

<http://micahaltman.com/>

 @drmaltman

THANKS