

Guiding Principles for the ALMA Cycle 4 Large Program

This program involves a large volume of complicated data, with huge community interest in the results. Moreover, we have expressly committed to providing a set of “deliverables” to the public on a relatively short timescale. Since the scope of this work and these commitments are considerably grander than a typical observing program, we must be incredibly organized to ultimately be successful. This living document is designed to lay out some principles that will establish ground rules and keep us organized as we execute this work.

This document contains five sections: (1) Team Members; (2) Data Access, Calibration, and External Use; (3) Work Assignments; (4) An Open Work Environment; and (5) Publication Policy.

1 - Team Members

For the sake of centralization, the overall program will be managed by the **PI**, Sean Andrews, with significant input from the **co-PIs**, Laura Pérez, Andrea Isella, and Kees Dullemond.

The team working on this program will inevitably grow (and indeed already has). The PIs will establish by consensus the participation of new members and the prioritization of work among the team. The original co-Is of the team will have some special status in several aspects of the project, as will be detailed below. The original **co-Is** are: John Carpenter, David Wilner, Zhaohuan Zhu, Meredith Hughes, Karin Oberg, Xuening Bai, Til Birnstiel, and Luca Ricci.

High priority will be placed on promoting the work of **junior scientists** that are working on this program. These members must have the prior approval of the PIs for joining the program. At the top of that list will be (A) graduate students that have their thesis work primarily devoted to this program, followed by (B) postdoctoral fellows that have their primary focus on this program, and then on (C) graduate students and (D) postdoctoral fellows that will execute smaller projects or roles based on this program. The members associated with these categories are:

(A) Jane Huang (Andrews);

(B) ?

(C) Jason Ling (Isella) *[might be moved to (A)]*; Alba Lumbreras (Pérez)

(D) Viviana Guzman (Carpenter); Yann Boehler (Isella)

The PIs will consider including more senior scientists as program **affiliates** in special circumstances. This will be done sparingly and with considerable discretion, since we feel the functionality of the program benefits significantly from a smaller team of active participants. That said, we have already included one such affiliate: Myriam Benisty will be in Chile for 3 years and will lend her expertise to improve our potential synergy with scattered light imaging

observations (which was notably the only critique on the proposal we received from the ALMA review panel).

If you would like to consider an additional team member in any career stage with this program, you must **contact the PI** with the request and a justification.

2 - Data Access, Calibration, and External Use

2a - Data Access

Our intent is to provide open access to raw and calibrated datasets, as well as supplementary end-products (images, cubes), for everyone on the team.

For *new data* obtained through this program, notification will be sent from the NAASC to the PI when the data first become available and the proprietary period (12 months) is started. The team will be granted immediate access to the delivered data (raw ASDM files, logs and associated content, any quicklook data products, etc.). We have set up a file transfer system at the CfA to permit sharing of more advanced data products (e.g., calibrated MS files). A webpage (https://cfa.harvard.edu/ALMA_disks/), which will end up being password-protected (details to follow), can be used to download these files. To upload your own files, we will use the CfA's anonymous ftp server: instructions will be sent to the team. An automated script will grab those files and link them onto that webpage.

For *archival data* obtained in other programs, all relevant data for this program should already be publicly accessible (at least in raw format) through the ALMA archive (a list of these datasets is available on the program **WIKI** pages; see Section 4).

2b - Calibration and Data Products

A designated *Calibration and Imaging* (CI) group will lead and organize efforts to vet the quality of the delivered data, modify and re-calibrate as necessary, explore and optimize self-calibration strategies, appropriately combine datasets, and generate finalized data products (MSs, images, cubes, etc.). This work will be accessible to the larger team *in real time*, following the open access requirements outlined in Section 4 below. The annotated scripts, supplementary notes, and finalized data products that are ultimately signed off on by the CI group will be posted in an online repository accessible to all members (and eventually the community at large).

The *CI group* includes PIs Andrews, Pérez, and Isella (de facto), and will be lead by Huang. Others that are *committed* to doing the work should sign up by contacting the PI. We will organize a face-to-face, interactive meeting of the CI group to better coordinate the work.

We want to explicitly encourage multiple people to calibrate and image the same datasets, to help work out optimal strategies and potential pitfalls. However, those efforts need to be coordinated (by the CI group) and ideally should happen roughly simultaneously, to minimize the need to cycle back and repeat efforts. The leadership of these efforts for individual datasets will be assigned by the PIs based on the the guidelines in Section 3 of this document.

2c - External Use

During the proprietary period and before any mandated releases, distribution of any data products or associated content can only occur under extenuating circumstances and with the unanimous consent of the co-PIs (after the full team is notified and its comments are carefully considered). Furthermore, **only the PI** may distribute such material (e.g. data products or associated content, including but not limited to scripts, notes, etc.) to someone outside the team.

Images, figures, and results derived from the program's data are encouraged to be used in science talks (colloquia, conferences, etc.), observing proposals, funding proposals, committee presentations, or other internal contexts (in your department, etc.), but only after permission is granted by both the PI (in consensus with the co-PIs) and the designated Project Leader (see below) for the associated dataset. Such permission must be sought at least 1 week in advance of the proposed usage (unless the images, figures, or results have already been published). Please maintain this schedule even after any public release (for *unpublished* content).

An important reminder: we have committed to providing calibration scripts and final data products (images, cubes, etc.) to the general community at the first publication for each dataset or the end of its proprietary period, *whichever comes first*.

3 - Work Assignments

The PIs will distribute leadership responsibility for different aspects or targets of the program, pending a thorough strategic discussion in light of the quick-look data products. The resulting studies will have priority in presenting data for the first time in the literature. Subsequent usage (for example in a theoretical/modeling context or a follow-up collation of multiple targets, etc.), even if worked on simultaneously, will be approved by the PIs.

The person given responsibility for each project (hereafter **Project Leader**) agrees to

- (a) manage the data calibration (with the assistance and/or guidance of the CI group);
- (b) plan/execute the analysis *in a timely fashion* (and, ideally, do a lot of the work);
- (c) follow and enforce the rules of the Open Work Environment described in Section 4;
- (d) prepare deliverables to the general community; (cf., the CI group); and
- (e) write an article describing the results *in a timely fashion*.

Of course, it is difficult to predict in advance what a reasonable work flow and schedule are for individual projects at this point. But our over-arching goal should be to *publish first results that include all the data during the proprietary period*. To meet that goal, a nominal benchmark schedule should look like this (all times relative to delivery date):

- <4 months - final data products available (i.e., well before the Cycle 6 proposal deadline)
- <8 months - primary analysis (sufficient to guide writing) finished
- <10 months - complete draft article distributed to co-authors (see Section 5)

If you cannot commit to a schedule like this, **do not** volunteer or agree to lead a project! Project Leaders should agree to place this work at the top of their (independent research) priority list.

The PIs will confer to allocate Project Leadership for the initial set of projects, primarily based on the work commitment of themselves and junior members of the team. Descriptions of those projects will then be distributed to the full team to encourage participation. Aside from these projects, any member of the team can “pitch” a project idea following this procedure:

- Email a description of your idea and the data required to (all of) the PIs;
- The PIs will then confer to mitigate conflicts with other projects, decide on feasibility and priority, and make suggestions on scope and content;
- After any necessary iteration with the PIs, the (approved) project will be circulated to the full team to solicit comments and participation.

4 - An “Open” Work Environment

All work in this program will be conducted completely *in the open* (not to the public, at least during the proprietary / pre-delivery period, but to all team members). Much of that open-ness will be facilitated through a `github` repository (hereafter “repo”). If you have not already, create an account with `github` and then contact the PI with your username to receive access. Inside the `github` repo is a collection of **WIKI** pages, which will be used for additional practice and policy information and communication. There is a link to a `git` tutorial for those of you unfamiliar with its use; we will post a cheatsheet that should be suitable 99% of the time.

Each project, as well as the CI group, will set up its own directory in the repo. The Project Leaders are responsible for ensuring that participants are regularly (at least daily, when work is being done) pushing their work into the repo. A tutorial on the **WIKI** will describe best practices.

We expect the following list of items (not exclusive) to be included in the repo:

- Working notes, particularly regarding calibration and imaging
- Calibration and imaging scripts
- Post-processing scripts to generate any usable data products

- All code for analysis work (see exceptions below)
- All code for figure generation (see exceptions below)
- All related proposal or article text and figures
- All slides from related scientific presentations

Of course, not everything can/should be done via `github`. Some notable exceptions include:

- No large (>20 MB) datafiles (alternative solutions will be worked out)
- No large / legacy code (e.g., no MHD or radiative transfer codes necessary; considerations for proprietary code will be made in exceptional cases; ask the PI)

5 - Publication Policy

We are a small team and are working on a project of limited scope/duration. That means we should have very few issues with publication rights. To maintain a standard practice, we have identified a brief set of rules in three related subjects:

5a - Authorship:

- The Project Leader (i.e., the person that does most of the work) should be the Lead Author on any article or proposal related to that specific project. The Project Leader will determine the *order* of the authorship list. If there are any disagreements that cannot be resolved by the Project Leader, the PIs will make the decision.
- All original team members (PIs + cols) should be asked to be co-authors on any article or proposal directly related to this program.
- Junior scientists and affiliates should be asked if they wish to be included on the author list for any article or proposal directly related to this program in which they provided a substantial contribution (as deemed by the PIs and/or Project Leader).
- The bare minimum amount of work expected for authorship is a serious review of a completed draft manuscript on a reasonable timescale (see 5c below). The PI's preference would be that the acknowledgments section of the article include a brief description of every author's actual contribution (e.g., SA helped edit the manuscript), akin to what is done in, for example, *Nature* (despite his vitriol for that magazine).

5b - Procedure:

- All writing work will be done in open source (preferably via the `github` repo, although exceptions will be made for online sharing if there are sufficient resources available for the full team). If in the repo, each push should include a PDF of the current draft.

- All team members need to be made aware of the expected contents of any publication before the start of its drafting. This should come in the form of a section by section rough outline, so that an approach can be vetted *before* it is encased in writing.
- We will adopt a set of standards in the context of presentation (in figures, tables, etc.).

5c - Timelines:

- When a draft is complete, co-authors have 14 calendar days to comment on the preliminary version (if someone does not respond at all within that period, they will be removed from the author list);
- After any internal revision(s) after that initial internal vetting, authors shall comment on the revisions within 5 business days (or less, if all authors approve). The article can be submitted only when all authors approve;
- All authors will be notified when a formal peer review cycle is completed; suggestions for responses, etc. will be solicited at this time. When a revision is available, the authors will have 5 business days to offer additional critiques or comments before submission.