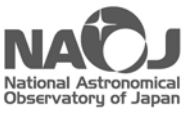
Doc 7.23, ver. 1.0 | April, 2020

Standard for Externally Contributed ALMA Large Program Data Products



www.almascience.org

User Support:

For further information or to comment on this document, please contact your regional Helpdesk through the ALMA User Portal at www.almascience.org
. Helpdesk tickets will be directed to the appropriate ALMA Regional Center at ESO, NAOJ or NRAO.

Revision History:

Version	Date	Editors
2.0	22 April 2020	**Felix Stoehr, Mark Lacy, Brenda Matthews, Kuo-Song Wang, Rie Miura, Adele Plunkett, Andrew Lipnicky, Patricio Sanhueza, iSOpT, integrated Science Operations Team**

In publications, please refer to this document as:

F. Stoehr et al. 2020, Standard for Externally Contributed ALMA Large Program Data Products, ALMA Doc. 7.23 v2.0.

Table of Contents

1. Introduction

The award of an ALMA Large Program (hereafter LP) carries with it the responsibility to deliver enhanced data products and documentation describing the enhanced data products to ALMA within one year of the delivery of the final QA2 calibrated data products. These enhanced data products are "value added" and supplement the standard ALMA products that are generated by the ALMA observatory and already stored in the ALMA Science Archive (ASA). This document details the standards for these enhanced products to ensure their proper ingestion into the ASA.

The standards provided herein assure that the following key goals for the enhanced products are met. That:

- Users can download the LP enhanced data products individually or packaged into tar files.
- When the ASA has the capability to provide services such as previews, cut-outs or remote visualization for contributed FITS files, that the LP enhanced data products can take advantage of these capabilities as they will follow the necessary ALMA standards.
- PIs of LPs can have flexibility in the files they provide back to the ALMA Observatory. In particular they can provide products that are natural for their survey structure and goals and may be very different from standard ALMA data products.

PIs are required to abide by several criteria when delivering data back to ALMA. Specifically,

- The PIs of the LPs are fully responsible for the quality and scientific validity of the enhanced data products.
- A README file is provided to accompany the data delivery. This file is distinct from the individual "Data Description Documents" that should accompany each upload.
- The enhanced data products should ideally be associated with at least one published or accepted paper and must be accompanied by a "Data Description Document" as described in Section 3.2.
- All product deliveries to ALMA must be packaged "per LP Group ObsUnitSet (OUS)" or "per LP Member OUS" and must not contain subdirectories. In case there are multiple Member OUSs in a LP's Group OUS, the default is the packaging at the Group OUS level. For more information on the ALMA data hierarchy, PIs should refer to Section 8.1 of the Technical Handbook. PIs should contact the ALMA Helpdesk for questions related to the packaging of the data.
- The uploaded files must be preceded by a prefix and follow the ALMA product naming convention (see Section 13 below).
- Where applicable, the uploaded FITS files must include standard FITS headers (Section 5).
- An image showing a result from the LP (in PNG format, 400x300 pixels) with a paragraph explaining the content of the image must be provided. This image will be displayed on the landing page for the Project on the ALMA Science Portal (see next section).

2. Landing Pages for Large Programs

Large Programs have a central page on the ALMA Science Portal (SP): https://almascience.org/alma-data/lp. From this page, individual pages for each program will be linked. A program can collect data from several ALMA LPs or even data from other facilities over the years. Each program will have its own landing page,

generated by the ALMA Project. For instance, one of the earliest programs can be found at: https://almascience.org/alma-data/lp/DSHARP.

The team's page should give a high level description of the program a general summary of the (types of) targets observed. The specific sections of the Landing Page are listed below. In addition, each landing page will be prefaced with a statement summarizing the content of the pages and a standard disclaimer about the formatting of the enhanced data products.

The sections of the Landing Pages for Large Programs are:

- Program Abstract (max 300 words) and associated image
- Table of ALMA LP project codes, their titles and their PI names so far obtained
- Team members
- List of publications
- Main deliveries
- Links to other resources (e.g. data at other facilities)
- (optional) main science drivers
- (optional) observing strategy

Example wording for the Main deliveries

"<LP> delivers images and source lists (moment maps, spectra, etc. as specific for each program).

Please note that when a Large Program team delivers enhanced data products to ALMA, the data files are checked to assure that the necessary metadata are present and that naming conventions formatting has been followed. Other than that, the files are provided as-is. Specifically, no other Quality Assurance assessment is made by ALMA staff. The PI of the LP is responsible for the content and quality of the contributed products. "

3. Basic Requirements

Some basic requirements must be met in the enhanced data product delivery. Enhanced data delivered to ALMA must be carefully checked by LP PIs before upload to ensure compliance with these requirements.

3.1 README File

One README file must be provided for each set of enhanced data uploaded (corresponding to a Group or Member OUS) in ascii (.txt) format. If a published paper on the LP data products and/or processing techniques exists, a link to that publication should be provided. A Data Description Document (see Section 3.2) must also be provided in PDF format, regardless of whether data processing descriptions exist in publications.

The README file must conform to the following template:

Project Code: (ALMA project code)
ObsUnitSet: (LP Group (or Member ObsUnitSet) UID to which the data products are associated)

Data Description Document: Name of the file provided to describe in detail the processing of the data and the products.

Relevant publication: (link to publication, if none, the name of the data description document in the package)

Description of products: (Brief description of the supplied products, with a list of files in the package)

e.g. member.uid___A001_Xbc4641_X20.lp_sandrews.README.txt

Project Code: 2016.1.00484.L
GOUS: uid://A001/Xbd4641/X1f
MOUS: uid://A001/Xbc4641/X20

Data Description Document: member.uid___A001_Xbc4641_X20.lp_sandrews.description.pdf

Relevant Publication: https://ui.adsabs.harvard.edu/abs/2018ApJ...869L..41A/abstract

Description of products:
member.uid___A001_Xbc4641_X20.lp_sandrews.gw_lup.cube.I.pbcor.fits
enhanced primary beam corrected continuum image
member.uid___A001_Xbc4641_X20.lp_sandrews.gw_lup.cube.I.pb.fits
primary beam
member.uid___A001_Xbc4641_X20.lp_sandrews.gw_lup.cube.I.mask.fits
mask file
member.uid___A001_Xbc4641_X20.lp_sandrews.gw_lup.mfs.I.pbcor.fits
continuum image
member.uid___A001_Xbc4641_X20.lp_sandrews.scriptForImaging.py
imaging script file

The README file must not go into detail about how the products were generated. It just contains a list of the files contained in the enhanced data delivery. Details about the generation of the products (e.g. specific CASA commands etc.) belong in the Data Description Document (see section 3.2).

If the enhanced data contained in a program file comes from a LP and other ALMA sources (PI projects) or even other facilities, the README file should note these products and provide the corresponding dataset identifiers. In case of ALMA data, the ALMA project code(s) is sufficient. The enhanced data products provided can include FITS files containing data from additional project IDs, but the content of these FITS files should be clearly noted. As many files as needed can be uploaded, each clearly associated with one LP Group OUS (or Member OUS), and packaged and named accordingly.

In the case where a PI of a LP is about to contribute enhanced data products for an OUS of that LP back to ALMA and the products have been created also using data from a previous ALMA LP, the PI has two options. Either only a reference to the previous LP's enhanced data products is made (as indicated above), or alternatively, the enhanced data products from the previous LP that have been used can be included in the new LP contribution. Of course, these products will have to be prefixed with the new Group (or Member) OUS ID of the intended contribution.

3.2 Data Description Document

A detailed Data Description Document must accompany the Group (or Member) OUS enhanced data package, which provides the user with a complete list of all the uploaded data products and a description of their content. The document must include the project title and project code and provide detailed descriptions of data files (and their provenance) contained in the delivery. This file should provide sufficient detail for a prospective user to reproduce the data reduction and understand the (potentially multiple) data sets within the products. The level of description in most cases will exceed that of the "observations" or "data processing" section of a publication. The document should be less than 50MB in size.

As products can be generated that contain data from ALMA alone or from ALMA in combination with other facilities, the origin of all the data contained in the products should be fully described (see below).

The Data Description Document can contain textual entries, tables and figures. PIs may wish to include details of the content in the filenames, but all files should include a short descriptor regardless of the amount of detail in the filename itself.

Structure of a data delivery document

1 Project title, PI, Project ID
2 Introduction: This section describes the data included in the delivery, including whether data from more than one facility is included. A general description of the observations and overall (scientific) goal of the project should be included.
3 Content of the data delivery: This section contains more details about the individual files that are delivered. If any file naming conventions are used, they should be described in this section. The delivered files should be listed and the nature of each file described (image, table, spectrum etc.). A table of the sources included in the file(s) should be included. The equinox of the positions should be explicitly given.
4 The methods of data processing should be described. Any individual notes on specific sources should be included here. Different sections can be used to describe different components of the data (e.g., continuum, spectral line, polarization). Data reductions scripts can be included. Techniques should be described in enough detail that researchers can recreate the enhanced products if desired (specifying smoothing kernels, thresholds etc. if these are not obvious in the script provided).
5 Any graphics or tables can be added to illustrate the content of the data.
6 The quality assessment or science validation that has been done on the script and/or products by the LP team should be described.

3.3 Recommended files

As previously stated, the uploaded enhanced data products must include, at minimum, the README and the data description files.

However, it is strongly recommended, that if the team delivers the .pbcor. FITS files according to ALMA standards (see Section 5 below), then the corresponding .cont., .mask., and .pb. FITS files are also delivered.

It is also recommended that the deconvolution commands or scripts that were used to generate the products are provided. If PIs of LPs judge it useful, they also can provide the calibrated, split-out Measurement Set files (ms) file(s) in .tgz format.

4. Data Associations

As previously mentioned, every enhanced data product delivered must be uniquely associated with an ALMA Group (or Member) OUS of the LP. This is true of both the ALMA-only enhanced data products and data products that result from the combination of ALMA data with data from other facilities.

If PIs have trouble associating the created enhanced data products with unique OUSs, help can be provided through the ALMA Helpdesk.

5. Images

Images must be supplied in FITS format. Images must comply with the definition of the FITS standard and must be readable with the CFITSIO library (available from the FITS Support Office at NASA Goddard Space Flight Center) and with WCStools (from the Smithsonian Astrophysical Observatory). These conditions can be verified with FITSVERIFY (see Section 10). Files that use custom variants of the standard or that must be read with special software cannot be accepted.

FITS images or data cubes made only out of ALMA data must have the standard four dimensions (two spatial dimensions, frequency and polarization in the standard order). The main data must be associated with the primary header data unit (HDU) and not be placed into a FITS extension. The FITS files must follow the ALMA FITS format of recent ALMA Pipeline products.

If PIs have trouble generating the appropriate FITS images, help can be provided through the ALMA Helpdesk.

Required header information

The FITS keywords for images derived solely from ALMA data should provide the same FITS keywords as current ALMA Pipeline products. Note that data processed through the ALMA Pipeline will have the correct keywords by default.

Mandatory keyword and values:

- OBJECT, DATAMIN, DATAMAX and WCS keywords.
- TELESCOP must be 'ALMA'.
- BMAJ, BMIN and BPA should be specified in the header. BUNIT must be given.
- RADECSYS must be 'ICRS', and the velocity frame (VELREF) must be 'LSRK' or 'SOURCE' (for solar system objects), DATE-OBS, VELREF.

Recommended keywords and values:

We strongly recommend following the ALMA FITS keywords and values as present in recent ALMA Pipeline products.

FITS files containing the full set of mandatory keywords and the ALMA FITS data-structure will be accessible to server-side services for visualization.

FITS files from non-ALMA data or FITS files that do not represent images or data cubes should satisfy the basic FITS standards by passing FITSVERIFY.

6. Tables

Tables may be supplied in CSV or FITS format. Positional information should include the system (ICRS very strongly recommended, Galactic [equinox]). Time system information (UTC, MJD, etc…) should be included for time series.

We strongly recommend that supplied tables follow the standards described at http://cds.u-strasbg.fr/doc/catstd.htx.

7. Spectra

Spectra may be supplied as either tables (see above) or 1D images.

8. Models and Simulations

PIs may submit model and simulation data along with observations provided they conform to the formats listed at the end of this document and are fully described in a published paper or Data Description Document.

9. Reduction Scripts

Data reduction and analysis scripts can be provided in .tgz format. Reduction scripts are not required, but ALMA endorses the delivery of a reduction script with the data. Reduction scripts should clearly specify the required software, including version numbers/dates. Scripts running casapy are strongly preferred.

10. FITS Verification

FITS files (images or tables) should be validated using the FITSVERIFY validator available from HEASARC and any problems with the files must be fixed BEFORE delivery of the enhanced data products. If there are only a small numbers of files, it may be more convenient for PIs to verify the FITS files online: http://fits.gsfc.nasa.gov/fits_verify.html.

11. Data Origin

Sufficient information should be present in either/both the Data Description Documents or the headers of the provided enhanced data files to understand the provenance of the data contained in the enhanced data products. This includes the facility and instrument descriptors as applicable.

12. Directory Structure

Products for each OUS must be uploaded in a flat directory structure (i.e. no subdirectories below the OUS level are allowed). In cases where the PIs of Large Programs absolutely require a data tree that needs to be

browsed through the addition of a sub-directory structure, PIs should contact the ALMA Helpdesk for assistance and recommendations.

13. Filenames

The ALMA project will provide a mandatory file-prefix identifying the OUS with which the files are associated. This prefix must be prepended to each delivered enhanced data file to ensure a unique filename in the archive. Prefixes have the form - member.uid___A001_x005F_X1234_X45676.lp_almaid._<ALMA user id>. The "lp" designation identifies the data file as part of a contributed dataset from a large program, the <ALMA user id> is the ALMA userid of the PI of the LP. Where possible, the ALMA file-naming convention should be followed, which is described in the ALMA QA2 Data Products document, and which can also be obtained by submitting a request to the ALMA Helpdesk.

Only the characters [A-Z],[a-z],[0-9],.,_,- are allowed in the filenames. Lower-case characters should be used everywhere in the filenames except for the README, source names and stokes parameter.

14. Replacing Products

If needed and justified, enhanced data products that have been delivered to ALMA from a LP Team can be replaced. In these cases, the entire data package of an OUS will be updated. This is true even if not all the files need to be replaced, or if files are simply being added. A note describing the changes must be added to the README file if a dataset has been updated. The new products will replace the old products stored for that OUS.

The OUS packages of a given LP can be uploaded at different times.

15. Allowed Data Formats

Accepted file formats are: fits, tgz, pdf, png, csv, txt, jpg, py and mp4.

16. How to Upload Products

Once the PI has assembled an OUS enhanced data products package, including a README and data descriptor file(s), the PI should contact the ALMA Helpdesk for assistance in transferring data files.

17. Additional Assistance

If PIs or LP teams find issues with any of the recommendations for enhanced data product submission, PIs should contact ALMA through the ALMA Helpdesk.

The Atacama Large Millimeter/submillimeter Array (ALMA), an international astronomy facility, is a partnership of the European Organisation for Astronomical Research in the Southern Hemisphere (ESO), the U.S. National Science Foundation (NSF) and the National Institutes of Natural Sciences (NINS) of Japan in cooperation with the Republic of Chile. ALMA is funded by ESO on behalf of its Member States, by NSF in cooperation with the National Research Council of Canada (NRC) and the Ministry of Science and Technology (MOST) in Taiwan and by NINS in cooperation with the Academia Sinica (AS) in Taiwan and the Korea Astronomy and Space Science Institute (KASI).

ALMA construction and operations are led by ESO on behalf of its Member States; by the National Radio Astronomy Observatory (NRAO), managed by Associated Universities, Inc. (AUI), on behalf of North America; and by the National Astronomical Observatory of Japan (NAOJ) on behalf of East Asia. The Joint ALMA Observatory (JAO) provides the unified leadership and management of the construction, commissioning and operation of ALMA.