



## [Quick Guide to Dataset Creation in AHED](#)

Before starting to gather and prepare your own data, it is worth getting familiar with the elements and organization of AHED datasets.

In AHED, 'data' consists of one or a combination of the following: a file, a collection of files, a link, or several links to other repositories or online resources (uniform resource locator URLs and digital object identifiers) containing relevant data. To become an AHED dataset, 'data' must be described and labeled using the Astrobiology Resource Metadata Standard (ARMS). A description of ARMS can be found [here](#). We think labeling data with ARMS should be relatively simple (but we are interested to get feedback about this). The AHED portal includes a dataset creation tool that will guide you through the process of adding data and metadata when you click 'Contribute'. Links to 'related works' describing the data or external data resources are an optional part of AHED datasets. Digital Object Identifiers (DOI's) are automatically assigned when creating a new dataset.

After adding or linking data and completing ARMS, new AHED datasets can be published. When published, datasets will become visible on the AHED Portal via search interface. A landing page is automatically generated for each dataset: an example is shown below.

[Back To Search](#)

## The NASA Ames PAH IR Spectroscopic Database

Last Revision Date: Feb 16, 2022 | Published Date: Feb 16, 2022

DOI: <https://doi.org/10.48667/6p1n-w007>



Image credit: The PAHdb Team

### Description

The astronomical emission features, formerly dubbed the unidentified infrared bands, are now commonly ascribed to polycyclic aromatic hydrocarbons (PAHs). The laboratory experiments and computational modeling done at the NASA Ames Research Center to create an unrivaled collection of PAH IR spectra relevant to test and refine the PAH hypothesis have been assembled into a spectroscopic database. This database contains the laboratory-measured and density-functional-theory-computed spectra of over 4,000 PAHs and PAH clusters. Online and offline Python/IDL tools are available to work with the data.

### Data

File or URL	Description/Data Usage	File Size	Date
<a href="https://www.astrochemistry.org/pahdb/">https://www.astrochemistry.org/pahdb/</a>	Database of over 4,000 laboratory-measured and density-functional-theory-computed spectra of PAHs and PAH clusters.		Feb 16, 2022

### Related Works

DOI	Description
<a href="https://doi.org/10.3847/1538-4365/abc2c8">https://doi.org/10.3847/1538-4365/abc2c8</a>	The NASA Ames PAH IR Spectroscopic Database: The Laboratory Spectra
<a href="https://doi.org/10.3847/1538-4365/aaa019">https://doi.org/10.3847/1538-4365/aaa019</a>	The NASA Ames PAH IR Spectroscopic Database: Computational Version 3.00 with Updated Content and the Introduction of Multiple Scaling Factors
<a href="https://doi.org/10.1088/0067-0049/211/1/8">https://doi.org/10.1088/0067-0049/211/1/8</a>	THE NASA AMES PAH IR SPECTROSCOPIC DATABASE VERSION 2.00: UPDATED CONTENT, WEB SITE, AND ON(OFF)LINE TOOLS
<a href="https://doi.org/10.1088/0067-0049/189/2/341">https://doi.org/10.1088/0067-0049/189/2/341</a>	THE NASA AMES POLYCYCLIC AROMATIC HYDROCARBON INFRARED SPECTROSCOPIC DATABASE: THE COMPUTED SPECTRA

### Summary

**Lead Investigator(s)** [Boersma, Christiaan](#)

**Funding Sources** CAN | NPP | other

**Mission/Projects** James Webb Space Telescope (JWST) | Other | Spitzer Space Telescope | Stratospheric Observatory for Infra-Red Astronomy (SOFIA) | Other | Gemini Observatory

**Themes** Characterizing Environments for Habitability and Biosignatures | Coevolution of Life and the Physical Environment | Abiotic Building Blocks of Life | Early Life and Increasing Complexity | Constructing Habitable Worlds

**Keywords** RNA world | polycyclic aromatic hydrocarbons (PAH) | haze | polycyclic aromatic hydrocarbons (PAH) | infrared spectroscopy | haze | National Aeronautics and Space Administration (NASA) | National Aeronautics and Space Administration (NASA)

**Disciplines** planetary science | astrochemistry | astronomy | biochemistry | physical chemistry | spectroscopy | astrophysics

**Team Members** [Allamandola, Louts J.](#) | [Bauschlicher, Charles W.](#) | [Hudgins, Douglas M.](#) | [Mattioda, Andrew L.](#) | [Ricca, Alessandra](#)

### Fieldwork Location(s)

No Locations Given

### Contact

[Click here to contact the team](#)

### How to Cite

When using this dataset, please cite the data package provided here as well as original publications, when available. [License terms and conditions apply](#).

**Dataset:** Boersma, C; Allamandola, L; Bauschlicher, C; Hudgins, D; Mattioda, A; Ricca, A. The NASA Ames PAH IR Spectroscopic Database, Astrobiology Habitable Environment Database. Retrieved Date: Mar 11, 2022. <https://doi.org/10.48667/6p1n-w007>

For data available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.



### [Dataset examples](#)

AHED was not designed to be a ‘one-size-fits-all’ repository and we recognize that some astrobiology related data may be better suited to storage on other venues. We do not want this to be a barrier to discovering relevant astrobiology data and resources in AHED, therefore some AHED datasets simply provide links, information, and citations to other resources that can be discovered using AHED search tools. One example, is the [The NASA Ames PAH IR Spectroscopic Database](#) (see figure above), which provides a link to a dedicated database page, links to related literature, and summary of information about the database.

Another example dataset is the [CheMin database](#). This dataset contains links to resources that host CheMin instrument X-ray diffraction data, collected during the Mars Science Laboratory Missions at Gale Crater. The dataset provides links to NASA’s Planetary Data System, as well as a custom database published using the Open Data Repository Platform (the software which is also used to create AHED datasets).

Other datasets can include both data files and URLs to other resources, such as in the [ThermoBase dataset](#), that currently houses comprehensive descriptions for 603 thermophilic or hyperthermophilic organisms

Publishers are increasingly asking authors to accompany journal articles with the raw data and supporting materials needed to replicate research in recognized repositories. Creating an AHED dataset containing this kind of supporting material, perhaps based on a recent publication, would make an ideal way to test the system for the Pilot exercise. An example can be found [here](#).

If you have more questions that come up when thinking about the type and scope of data to put into an AHED dataset please feel free to contact us before the Pilot. We can also discuss this during the webinar.



Tips for preparing your data for AHED

<p style="text-align: center;"><u>Collect your data</u></p> <p>Consider including all forms of data that would help others to understand and reuse your dataset.</p> <ul style="list-style-type: none"> <li>• Files       <ul style="list-style-type: none"> <li>○ Unprocessed raw data.</li> <li>○ Processed data.</li> <li>○ Organized tables and spreadsheets.</li> <li>○ Code for data processing and visualization.</li> <li>○ Manuscripts.</li> <li>○ Graphs, images, etc...</li> </ul> </li> <li>• URL       <ul style="list-style-type: none"> <li>○ Publicly available repositories where your data is already stored.</li> <li>○ Project website.</li> <li>○ DOI of publications related to your dataset.</li> <li>○ GitHub repository.</li> </ul> </li> </ul>	<p style="text-align: center;"><u>Recommended formats</u></p> <ul style="list-style-type: none"> <li>• Choose open file formats, when possible, to facilitate the use and preservation of the data. If you include proprietary files, such as instrument vendor's data files, consider also providing the same data in an open format.</li> <li>• Plain text formats are recommended.</li> </ul>
<p style="text-align: center;"><u>Verify that your data can be shared publicly</u></p> <ul style="list-style-type: none"> <li>• Make sure that you have the rights to distribute any data uploaded to this site and that the data is appropriate to distribute. You should not upload any information considered:       <ul style="list-style-type: none"> <li>○ Classified, Sensitive But Unclassified (SBU)/Official Use Only.</li> <li>○ Copyrighted (without retaining proper copyright notices and without permission where not the owner)</li> <li>○ ITAR/EAR restricted.</li> <li>○ Proprietary without permission.</li> </ul> </li> <li>• For URL to other repositories, make sure that the data is publicly available and not password protected.</li> </ul>	<p style="text-align: center;"><u>Organize data logically</u></p> <ul style="list-style-type: none"> <li>• Create a clear and logical file structure.</li> <li>• Name files descriptively and consistently.</li> <li>• For large collections of files (e.g., spectra, diffractograms, chemical information), bundle organized files into compressed file archives (such as .zip). Verify they can be opened and not corrupted.</li> <li>• Omit unneeded files.</li> <li>• Try to keep individual files or compressed files smaller than 10GB to facilitate data download.</li> </ul>
<p style="text-align: center;"><u>README file</u></p> <p>Describing your dataset with a README.txt file is highly recommended. It will facilitate the re-use and replication.</p> <ul style="list-style-type: none"> <li>• Write it clearly and concisely for a broad audience.</li> <li>• Describe data collection methods, processing, and analysis steps.</li> <li>• Define variables and allowable values.</li> <li>• Describe software used to process, visualize, analyze, and compress your data.</li> </ul>	



[AHED site registration](#)

Although you can explore much of the AHED Portal without logging in, dataset creation and participation in the Pilot requires you to sign in.

- If you already have a NASA identity you may use Launchpad to do this with your PIV card and Pin or your NASA Launchpad credentials (username & password).
- If you **\*do not\*** have a NASA ID you will need to create a *NASA guest account*. We recommend using an institutional email when you set up your NASA Guest account.

Our recommended way to create a NASA Guest account is to click on the “Login” button in the top right corner of the AHED homepage, then click the “Sign Up” button followed by the “CREATE GUEST ACCOUNT” button. You will then be guided through setting up a NASA Guest account.

Once the NASA Guest account has been created, it needs to be registered with AHED to allow access. To register a newly created NASA Guest account, click on Sign in with Launchpad button on the login page. Select Agency User ID option and use the email and password on your NASA Guest account.

Fill out and submit the AHED registration form. Our team will review the information and contact you shortly via email to confirm registration.

Once your account has been approved, use Sign in with NASA Launchpad to log in into AHED.

More instruction on how to login can be found here: [https://ahed.nasa.gov/assets/downloads/LOGIN\\_Instructions\\_Rev.3.pdf](https://ahed.nasa.gov/assets/downloads/LOGIN_Instructions_Rev.3.pdf)