

**The American Heart Association Precision Medicine Platform (PMP)**

**User Guide**

[**https://precision.heart.org/**](https://precision.heart.org/)

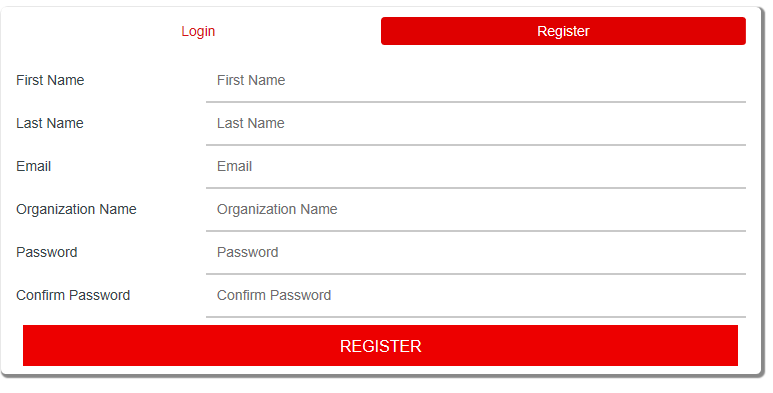
**January 2020**

The Institute’s Precision Medicine Platform (PMP) is a cloud-based technology solution that enables the medical research community to accelerate breakthroughs in cardiovascular and brain diseases by providing streamlined access to diverse datasets, a community vetted framework for data harmonization and management, secure workspace environments equipped with state of the art tools – such as artificial intelligence – for data analysis, and collaboration capabilities for both internal and global teams. Amazon Web Services (AWS) provides the cloud computing capabilities and REAN Cloud, a software engineering firm located just outside of Washington, DC, provides the ongoing software development and managed services under the leadership of The Institute.

The PMP was developed specifically for data scientists in the medical research field who are engaged in the discovery process. The platform’s robust capabilities range from the ability to identify relevant datasets, gain access to datasets efficiently, leverage a broad set of analytics tools in workspaces with built-in security and compliance measures, collaborate with internal and external colleagues, and publish the research results to the scientific community if appropriate.

**Registering**:

A new user must Register prior to being given access to the PMP. Click on Register Now or on the Sign In page, then click on Register to complete the form.



The password must have at least 12 characters, to include one upper-case, one lower-case, one number, and one special character. Both the email address and password are case sensitive. Once the form is completed click on Register. The system will send an email with a code to the email address used in the registration process for verification. Follow the instructions within the email to finalize the registration process.

**Signing In:**

After successfully registering, click on the Sign In page and use your new login credentials to sign into the portal. Remember, the email address and password are case sensitive. Upon signing in the first time, you will be taken to the Search page to begin the process of requesting a workspace. *Please note, should you leave the session idle for more than 20 minutes, it will time out.*

**Change Password:**

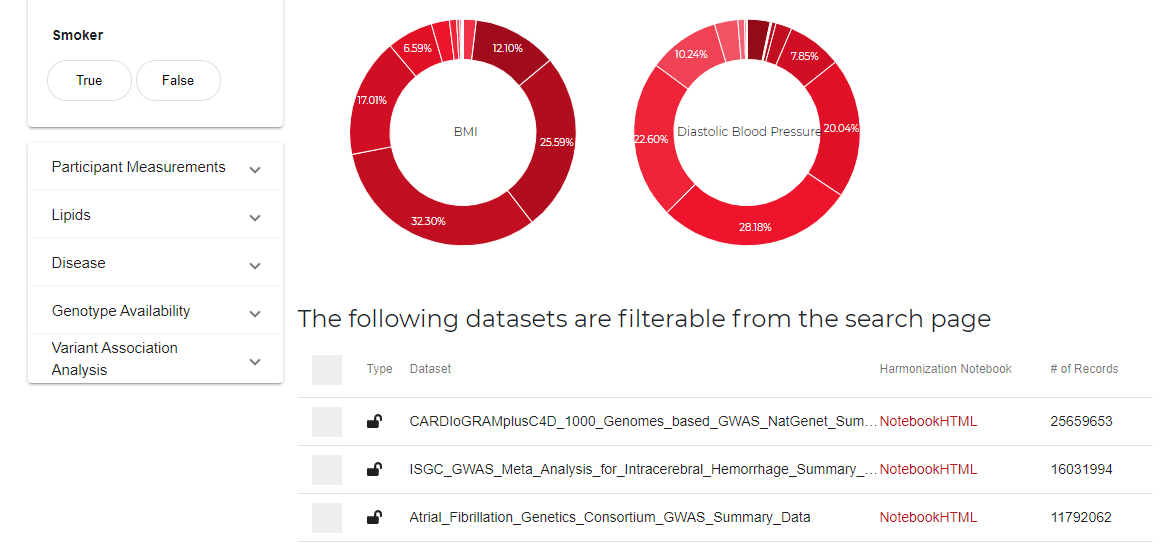
To change your password, when logged in, go to  and click on Profile then click on Change Password.

**Search:**

The PMP enables registered users to efficiently search across the datasets stored in the platform to quickly determine the data needed to address a specific research question. Researchers can take advantage of the platform’s filtering capabilities to find the data efficiently. You can also use the notebooks that accompany each dataset to learn more about the data, including how it was harmonized. The basic search capability is free of charge to all registered researchers.

The search page acts as a data marketplace and is designed to aid researchers in finding the datasets they need to conduct research, perform analyses or build analytic pipelines and tools. The layout consists of ring charts in the center, participant characteristic filters on the left, and the corresponding available datasets at the bottom of the page. The ring charts represent the total number of participants, by characteristic, that are searchable in the datasets listed at the bottom of the screen.





To begin the process of determining which dataset has the type of data needed for your study, use the filters in the Participant Characteristics panel on the left. You can choose a variety of participant characteristics within the panel to better define the dataset using the criteria of your study. Move the end points of the red lines to define those characteristics that have ranges. When you hover over the red lines the chosen range is displayed. For other characteristics, click on the criteria that apply to your research.

When using the filtering capability, the ring charts reflect the number of participants in the datasets that apply to that filter. When you hover over a ring chart segmentation, the number of individuals within that segmentation is displayed. The datasets are also updated to accurately reflect the total number of participants that match the filtered criteria.

The Harmonization Notebook link next to each dataset can be used to understand the science behind the data, the publications associated with the data, information about the data contributor, how to access extra documentation, how each variable is harmonized, and the distribution and summary statistics of all the variables available in the dataset. Data harmonization is a way of organizing and categorizing the many nuanced intricacies in data so that they may be useable for other researchers or shaped into a form that can be useable with other datasets.

The lock icons in the Type column indicates if the dataset is public or private; the data contributor of private datasets must approve access via a formal request process. The AHA serves as a trusted third party to ensure data security and privacy by facilitating the exchange between the data user and the data contributor. The process is streamlined through the Data Use Oversight System (DUOS) that allows users to request access to multiple private datasets at one time. DUOS is a semi-automated management service for compliant secondary use of human genomics data. The system ensures that researchers using genomics data honor these restrictions. DUOS interfaces with the various data access committees (DAC) to evaluate data access requests requiring manual review.

**Workspace**:

The PMP’s workspaces are secure environments equipped with a variety of analytic tools and computational languages, such as Jupyter notebooks, Python, and R. When researchers request access to data in the platform, they are required to “subscribe” to a workspace, which becomes their own “walled garden” in which to perform analyses. Researchers may also upload their own data to the platform for analysis. PMP workspace users may invite team members – whether local or global -- to share their workspaces for project collaboration. Generally, users may also customize their workspaces with their own tools, although there are sometimes questions about software licenses that must first be resolved.

The AHA Precision Medicine Platform provides a friendly web UI that allows you to write code in various languages (for example, Python, R, Scala), execute the code, and view the results as they are processed.

A workspace is a cloud-based compute cluster built on Hadoop/AWS EMR that you can interact with directly in your browser. The workspace is based on the concept of notebook files, where each notebook contains one or more code blocks (called cells). The content you write in your notebook is saved, even if you pause your workspace and come back to it later.

Cloud computing provides computational power beyond the ability of a personal computer. PMP workspaces make the process of accessing and utilizing cloud computing highly efficient. The Institute helps clients estimate how much cloud computing is likely to be necessary for their research projects however computing power in the workspaces will adjust on the fly as necessary. Data security is the highest priority of The Institute. The PMP is HIPAA (Health Insurance Portability and Accountability Act) compliant and has also been certified through the Federal Risk and Authorization Management Program (FedRAMP) for cloud computing security. For more information go to the [About Workspaces](https://precision.heart.org/workspace/about).

Jupyter Notebooks, formerly known as the IPython notebooks, are a flexible tool that helps you create readable analyses, as you keep code, images, comments, formulae and plots together. Jupyter is quite extensible and supports many programming languages. When working with Python in Jupyter, the IPython kernel is used, which gives you access to IPython features from within the Jupyter notebooks. For more information about Jupyter Notebooks go to the [About Jupyter Notebooks](https://precision.heart.org/jupyter_notebooks/about).

We have included Spark libraries for various languages (e.g. Pyspark, if you're writing Python code) to leverage the full parallel computing power of the Hadoop/AWS EMR platform. You can find examples and code snippets in the sample notebooks that are included on each workspace. Go to the [About Workspaces](https://precision.heart.org/workspace/about) for a current list of languages and tools pre-installed in the workspaces.

**Request a Workspace:**

A registered user can request a workspace by clicking on the Request a Workspace button at the bottom of the Search page. You will be asked to agree to the [Terms of Service](https://precision.heart.org/tos) and to indicate the amount of data you expect to bring on to the platform. Click Next and, if you are a single-payer, you’ll need to complete the billing information. Click Submit.AHA Grant Applicants are to complete the grant application related fields as well as the billing information, however, applicants and grantees will not be charged usage fees as these are included in your AHA grant.

To request access to a dataset along with a workspace, check the box to the left of the dataset(s) and then click on Request a Workspace at the bottom of the Search page. This will initiate the data access consent (DUOS) process as well as the workspace provisioning process. The first step is to confirm the datasets you wish to have access to and agree to the [Terms of Service](https://precision.heart.org/tos). You will also need to indicate the amount of data you expect to bring to the platform, if any, outside of the datasets you are requesting access to. Click Next.

You will be taken to the Research Details portion of the consent process when a chosen dataset is locked (private) and therefore approval is required prior to the release of the dataset.

**Research Details**:

* Research User Statement (RUS): a RUS is a brief description of the applicant’s proposed use of the dataset(s). The RUS will be reviewed by all parties responsible for data covered by the Data Access Request. The RUS should be one or two paragraphs in length and include research objectives, the study design, and an analysis plan. Please include the phenotypic characteristics that will be tested for association with genetic variants. If you are requesting multiple datasets, please describe how you will use each of them.
* Non-Technical Summary: Describe the RUS using layman’s terms, suitable for understanding by the general public.
* Type of Research: Provide additional comments as needed and indicate the type of research by choosing from:
  + Disease-related studies
  + Methods development and validation studies
  + Controls
  + Population structure or normal variations studies
* Disease Area(s) of Focus: choose from the drop-down menu the disease area(s) your study focuses on.

**Research Purpose**: A list of Yes/No questions are presented. Respond by clicking in the appropriate box.

**Attestation Statement**: Review the statements in this section and check the box to the left of the “I attest to the following.”

From there you are taken to the final page of the form, and if you are single-payer, you will be asked to provide your billing address as the last step of the process before you click on Submit.

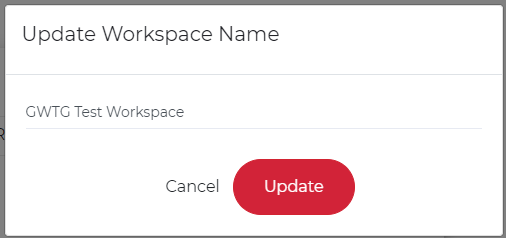
The user will be notified via email when the workspace has been provisioned and then again when you are granted access to any requested datasets. It takes approximately 24 to 36 hours for a new workspace to be provisioned and, if the dataset is public, the dataset should be made available at the same time.

To view the status of workspace requests, go to your Profile and then to the Workspace Requests section.

Once the workspace has been provisioned, on the Workspace Dashboard page, a Launch Workspace button will be presented allowing you to then launch from the platform onto the AWS Jupyter Notebooks to begin your computations and research analysis.

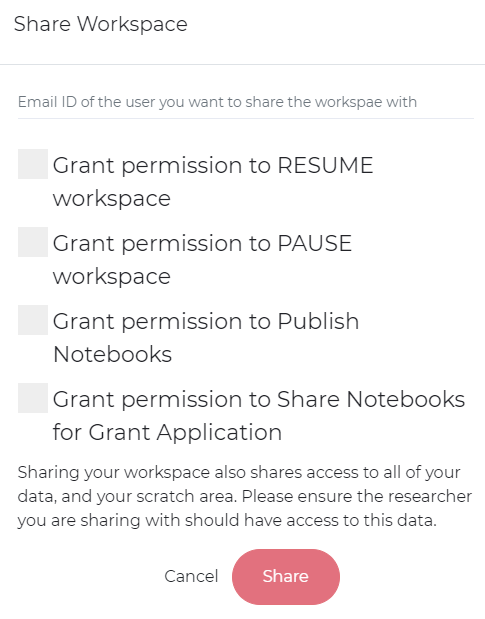
While on the Workspace Dashboard page, you can manage your workspace by editing the workspace name, configuring the view on the portal, and sharing the workspace with other users.

Update Workspace Name**:** When the workspace is provisioned, the name of the workspace is the workspace owner’s User ID (which is normally their email address). To update the workspace name, click on the name and complete the form in the pop-up.

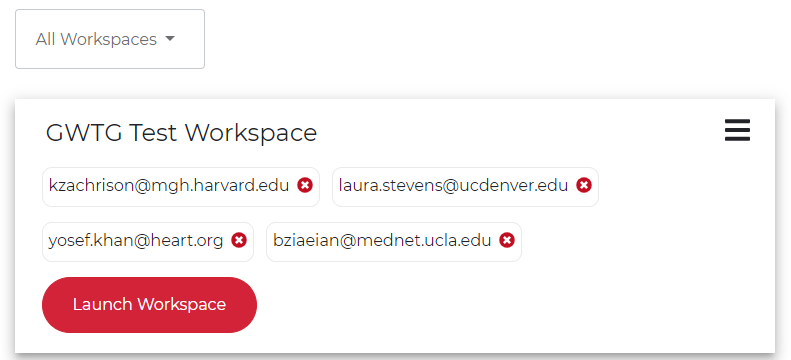


**Config**: This allows the user to determine if the billing history is to be displayed on the workspace landing page. *Only Paying customers will have this option.*

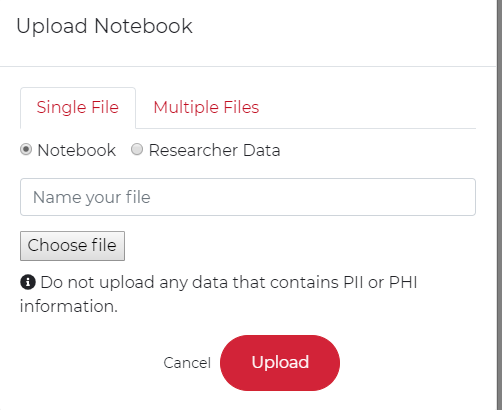
**Share Workspace:** The PMP’s collaboration capabilities drive a highly efficient research process. Researchers can share their workspaces with team members who may rotate in and out of academic departments, companies, or other organizations – including on a global basis. Even as team composition changes, the data, the coding, and the version records for the projects remain in the workspace. This eliminates the challenge of tracking down former colleagues to determine why something was done or not done in a particular way in a research project. The collaboration capabilities also ensure that researchers who want to reuse their data for a different project will have all the necessary information stored in their workspaces for use at any time.



As the owner of a workspace, you can share your workspace with others regardless of their location or organization. The only requirement is that the workspace is shared with a registered user of the PMP platform. To share your workspace, go to the Workspace page on the portal, click on the three lines  and then on Share and input the email address of the registered user in the pop-up box. (Remember, email addresses are case sensitive.) At this point, you can grant the user the ability to resume and pause the workspace, should you decide this is necessary. The owner of the workspace also has the ability to grant permissions to publish or share notebooks from within the workspace. Click on Share and an email will be sent to the user notifying them that they now have access to your workspace. The email addresses of the users with shared rights to your workspace will appear on your Workspace Dashboard. At any time, you can stop sharing the workspace by clicking the X next to the person’s email address.



**Upload data**: To upload files onto the platform, click on the  Upload Notebook icon. On the pop-up screen, you have the option to upload a single file or multiple files. They can be in any format, to include compressed, however it is recommended that the files be in Parquet format. Indicate if the upload is Researcher Data or a Notebook, name the file, and then choose the file to be uploaded and finally, click Upload.

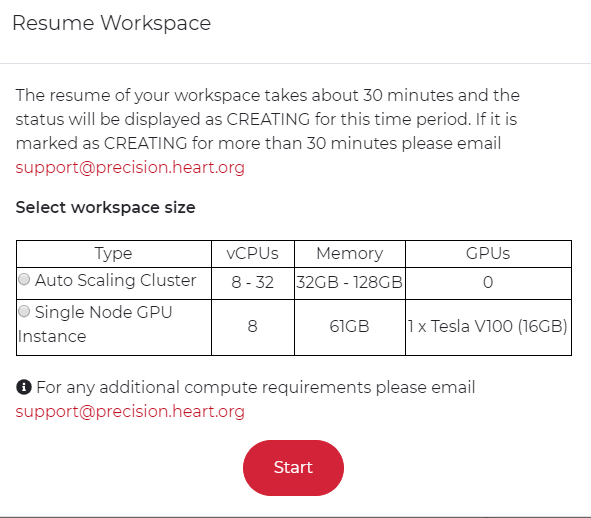


To upload multiple files or directory, you need to download an upload script file first. Click on the link to download the upload script and then go to the directory where you have downloaded the fileupload.sh file and use the appropriate (Mac OS or Windows) command to upload multiple files. The command will upload all files in the directory to S3 under a folder given by name "folder name". Refresh the page to see the updated notebook list. Keep in mind that the credentials are valid for 1 hour once you have initiated this process.

Once the data is uploaded, it will appear in the My Notebooks & Data section on the Workspace Dashboard page. It will then become available to anyone you have shared the workspace with.

***You are restricted from uploading Personally Identifiable Information (PII) or Protected Health Information (PHI) data on to the platform.***

**Pause**: This function provides you with control on how your workspace uses the Amazon Web Services cloud computing capabilities. If you are running a demanding computational process that uses a lot of high-performance computing, this function allows you to pause and resume this as need be. It also provides you with the ability to resize the cluster that has been loaded – to increase or decrease the size as required – by pausing the workspace and then resume it with different size parameters.



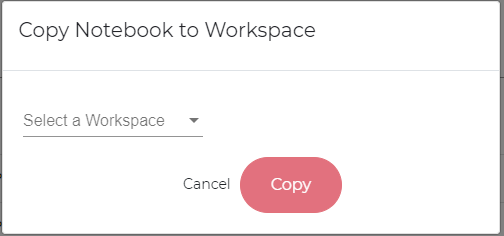
***It is best practice to pause the workspace when the compute function is finished or will not be used in the next few days.***

**Publishing/Sharing**:

The PMP’s publishing capabilities deliver a robust solution for enabling reproducibility and reusability of research results. Researchers can publish their results on the platform’s Learn page as well to external sources, like Github, or to provide a URL for use in a document or journal publication.

Share a Notebook to the Learn -> Shared Notebook page: To share a notebook in your workspace to the PMP Learn community, to the right of the notebook listed on your Workspace landing page, click on this icon 

Copy Notebook to Workspace: To copy a notebook that’s been shared on the Learn -> Share Notebooks page, click on this icon and then in the pop-up box, indicate which workspace to copy the notebook to and click Copy.



Publish File to generate a static URL: To create a static URL link to a notebook in your workspace – for publishing or sharing outside of the PMP community – click on this icon and complete the form with the descriptive information of the notebook that will go along with the link. Click on Publish and then OK. You will receive an email from [pmp\_admin@precision.heart.org](mailto:pmp_admin@precision.heart.org) with the static link.

Sharing a Notebook for a Grant Application: To share a notebook as part of a grant application, click on this icon  and then click Yes

**Resources**:

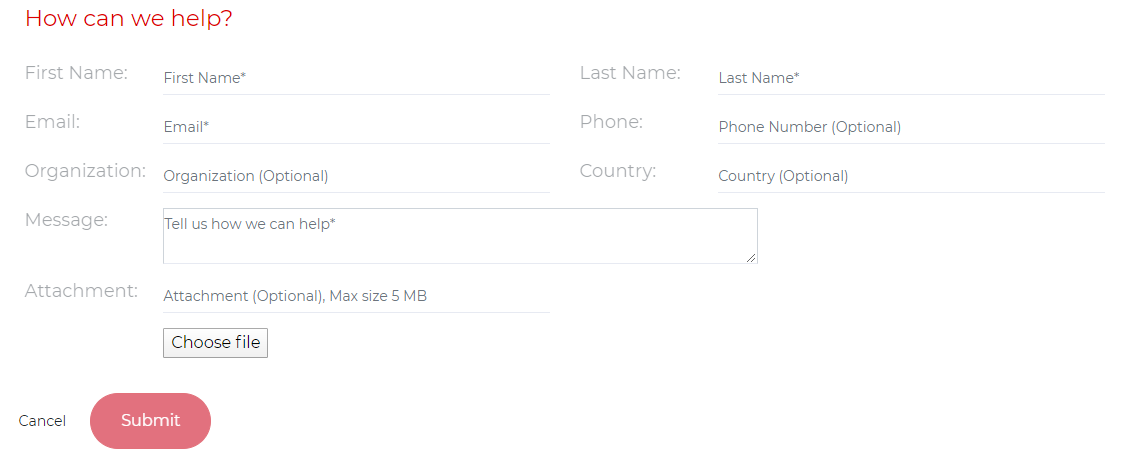
The Resource page provides the user with links to the [Terms of Service](https://precision.heart.org/tos), [Security Information](https://precision.heart.org/security-information) and [HIPAA Information](https://precision.heart.org/security-information).

Learn -> About PMP: On the Learn -> About PMP page you will find how-to tutorial videos for your reference. The videos range from 2 - 9 minutes in length and they provide you with high level instructions for how to use the tool. In addition to the videos, there is a repository of instructional information to include how-to documents and published material from other researchers

Learn -> Shared Notebooks: This is where PMP users shared notebooks can be found (see above).

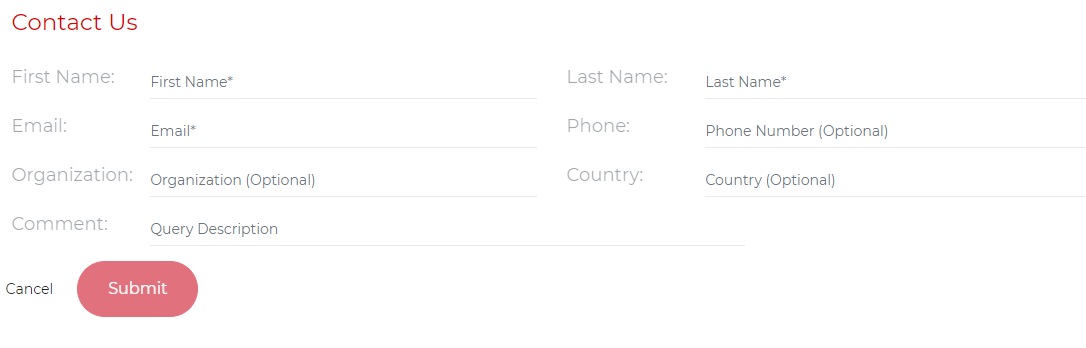
**Technical Support:**

If you are encountering technical issues while on the PMP, go to About -> Technical Support to submit a ticket to the technical support team. Complete the form, providing as much detail as possible, to include screen shots of error messages. Support is provided 24X7.



**Contact Us:**

Should you have general questions regarding the PMP portal or Workspace, please go to About -> Contact Us to submit your questions. You will be contacted within 24 hours, during the business work week.



**Frequently Asked Questions:**

**Functionality:**

1. Can I access my workspace outside of the US?
   1. Yes, if you have internet access, you are able to access the PMP.
2. Why am I not able to use the AWS Marketplace as I do with my university AWS infrastructure?
   1. The PMP is built using Hadoop and Jupyter framework as an easier transition to move to the cloud. Because of the design and security decisions, not all the offerings on AWS Marketplace are available within the workspace.
3. Can I use SAS or other licensed products in conjunction with the PMP?
   1. Yes, however, you must have an active license for the tool prior to loading it on the platform. Go to Technical Support for assistance loading tools; when completing the form, include details of the type of software you want to upload onto your workspace, including your active license.
4. Is there anything I need to do to prepare myself and my work to go on to the cloud?
   1. We recommend familiarizing yourself with R and Python and consider converting your files into the Parquet format, although Parquet is not a required format.
5. I know nothing about Jupyter Notebooks. Do you provide training on how to use the notebooks?
   1. No, however, there are many resources online that can help you get up to speed on how to use Jupyter Notebooks. Go [here](https://www.bing.com/videos/search?q=jupyter%20notebook%20tutorials%20for%20beginners&qs=AS&form=QBVR&sp=3&pq=jupyter%20notebook%20tutorial&sk=HS1RI1&sc=8-25&cvid=2CE84C953129498CB37725101C38BA94) to look at the various tutorial videos online and this [link](https://jupyter-notebook.readthedocs.io/en/stable/notebook.html) takes you to the official Jupyter Notebooks site with detailed user documentation.
6. I normally Terminal mode, does the PMP have a Terminal mode?
   1. Yes! You can access the Terminal through Jupyterlab or by clicking the *New* button on the top right corner of Jupyter and selecting Terminal on the drop down

**Data**:

1. What types of data are allowed on the PMP?
   1. There are no restrictions to the type of data you load onto the PMP, however, only de-identified data is allowed on the PMP. No PHI or PII data is to be loaded on to the PMP.
2. Are there limits to the amounts of data I can upload onto the workspace?
   1. No, the only limitations will be with your internet connection and the security timeout of one hour. Should you have a really large dataset, please reach out to us via Technical Support (About -> Technical Support) and an SFTP can be set up for a one time use.
3. How do I find my data after I clicked Upload?
   1. Once the data has been uploaded, the new data file will be found in the My Notebooks and Data section on the Workspace.
4. If I upload data onto my workspace, does it then show up on the Search page?
   1. No. Any data you upload onto your workspace is private and only visible to you and those you share your workspace with. If you decide that you want to share your data to the PMP community for others to use, follow the steps in the Publishing/Sharing section above and follow the Share instructions.
5. How does the system know if the people I share my workspace with meet the appropriate access requirements for each dataset?
   1. Currently the system does not provide this function. It is up to the workspace owner to ensure that the people you are sharing the data with meet the proper access requirements. However, if your workspace has access to a Private/Protected dataset, when you click on the Submit button to share access to your workspace, a formal request to the DAC will be initiated via the DUOS. Access to the workspace will only be provided upon approval by the DAC. (2020 Functionality)
6. I still don’t understand what data harmonization is and why it is necessary, can you provide me with an example of this?
   1. Yes! Data harmonization looks at how data is represented in a dataset to find commonly used variables and then reassigns those variables to a ‘more common’ label. The goal of harmonization is to make more datasets interoperable and available for researchers. An example of this is how a person’s sex could be listed as Male or Female or M and F, however the ‘more common’ label for a person’s sex is Male and Female. If the dataset uses M and F as the labels, the data harmonization process would change these to Male and Female and then make note of the change. By harmonizing the data within datasets, researchers can pull data from multiple datasets and be confident that the terms used to label variables and the variables themselves are consistent across datasets.

**General Questions about the AHA PMP platform**

1. What exactly is the cloud?
   1. The cloud is a network of remote servers hosted on the internet and used to store, manage, and process data in place of local servers and personal computers. This technology allows us to work on large, complex datasets in a much faster time span than was previously possible with conventional computing.
2. How is the AHA Precision Medicine Platform different from what my university provides me for my computational requirements?
3. The PMP does not rely on local servers managed by a university. The PMP uses Amazon Web Services (AWS) to provide its users an on-demand secure cloud services platform – a virtual cluster of computers – offering compute power, database storage, content delivery and other functionality to help institutions scale and grow their research exponentially faster than they might on their own. This provides users access to large-scale computing capacity quickly and inexpensively without the need to schedule or coordinate access as one might at an institution.
4. Why isn’t the PMP on the list of approved repositories for the AHA Open Science Policy?
   1. Not only do you need to register as a user of the PMP to gain access to data, some of the PMP datasets require approval for use/access, hence, the PMP does not meet the guidelines of an ‘Open Data’ repository where everyone has access to the data.

**Credits:** (for Grantees only)

1. Can I keep my unspent credits after my award ends?
   1. In some situations, a grantee will be allowed to continue to use the PMP after their award ends, if this should happen, use of available credits is allowed.
2. Do I have to pay for the workspace after my award is over?
   1. Should you want to continue to use the PMP after your grant has ended you would need to switch to a paying user. Depending on the situation, the AHA may be able to offer discounted pricing to aid researchers in the continuation of their work on the PMP.