ASAP CRN Cloud FAQ

The following document provides answers to frequently asked questions regarding the CRN Cloud.

Table of Contents:

About the CRN Cloud

What is the CRN Cloud?

ASAP CRN Cloud is a platform that allows researchers to access, analyze, and share ‘omics data. Researchers, both within and outside the ASAP Collaborative Research Network (CRN), can access the CRN Cloud at: https://cloud.parkinsonsroadmap.org/

Why was the CRN Cloud established?

The CRN Cloud was established at the recommendation of the CRN Single Cell Multi(omics) Working Group, to enable an environment for researchers around the world to leverage CRN data. The goal of the CRN Cloud is to provide an environment for CRN teams to store ‘omics data, which allows for replication of research findings and enables meta-analyses of the ‘omics work happening within the CRN and across the broader research community.

What ‘omics data types does the CRN Cloud support?

The CRN Cloud supports all ‘omics data. These include transcriptomics, genomics, proteomics, metabolomics, and lipidomics, among others.

Will additional ‘omics datasets be added on the CRN Cloud in the future?

Yes, the CRN Cloud team is continuously adding new ‘omics data.

How is data organized on the CRN Cloud?

CRN Cloud data is organized in a few ways:

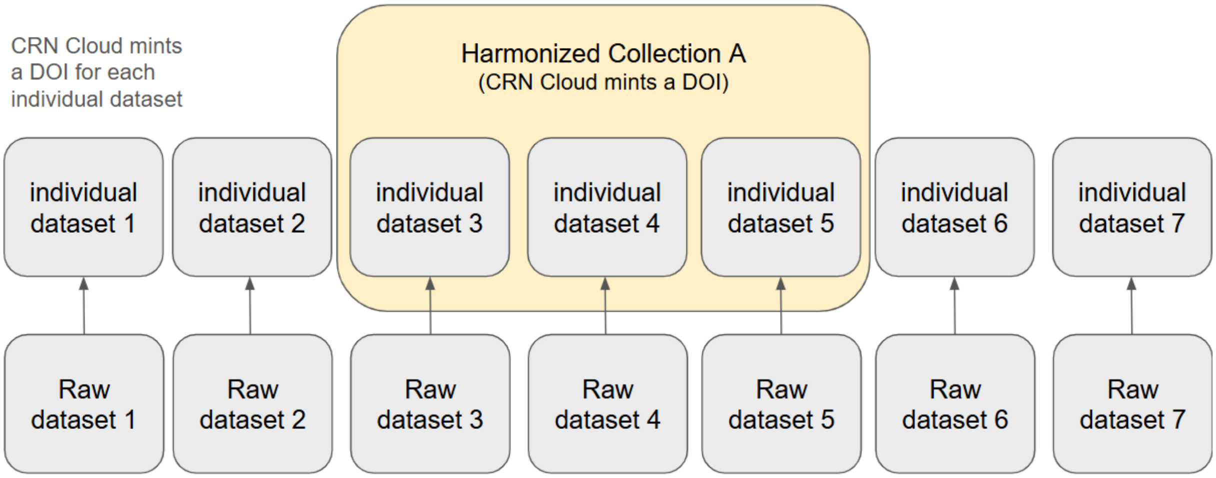
1. **Raw data:** Raw data in its original, unchanged form (i.e., fastqs), as submitted by the data contributors.
2. **Individual datasets:** The CRN Cloud data science team converts each raw data submission into an individual dataset. Each individual dataset is a self-contained collection of metadata, raw data, and processed data. These datasets are unique, with the primary goal of making *submitted* data artifacts accessible and raising awareness of the raw data for others to leverage. Curation for these

datasets focuses on standardizing metadata to align with the CRN Cloud schema and generating ASAP-specific IDs to clearly identify each dataset.

3. **Harmonized collections:** The CRN Cloud data science team combines multiple individual datasets of the same data type into a single harmonized collection. These collections integrate curated data artifacts processed through standardized bioinformatics pipelines, resulting in a unified dataset that aggregates and summarizes data from multiple sources (i.e., Human Postmortem-Derived Brain - snRNA-seq and PolyA RNA-seq Collection). Refer to the collection overview and the README, linked within, for details on the raw data processing and harmonization steps.

   NOTE: *Raw data is released alongside individual datasets.*

The figure below depicts 7 unique datasets organized according to this structure:



What data types are currently being harmonized?

The CRN Cloud team currently supports the harmonization of *transcriptomics* datasets to create larger, unified datasets which we call “Harmonized Collections”, as described in more detail above.

Will harmonization of non-transcriptomics data be established?

Harmonization of non-transcriptomics ‘omics datasets is currently in development and will be rolled out throughout 2026.

Open Science

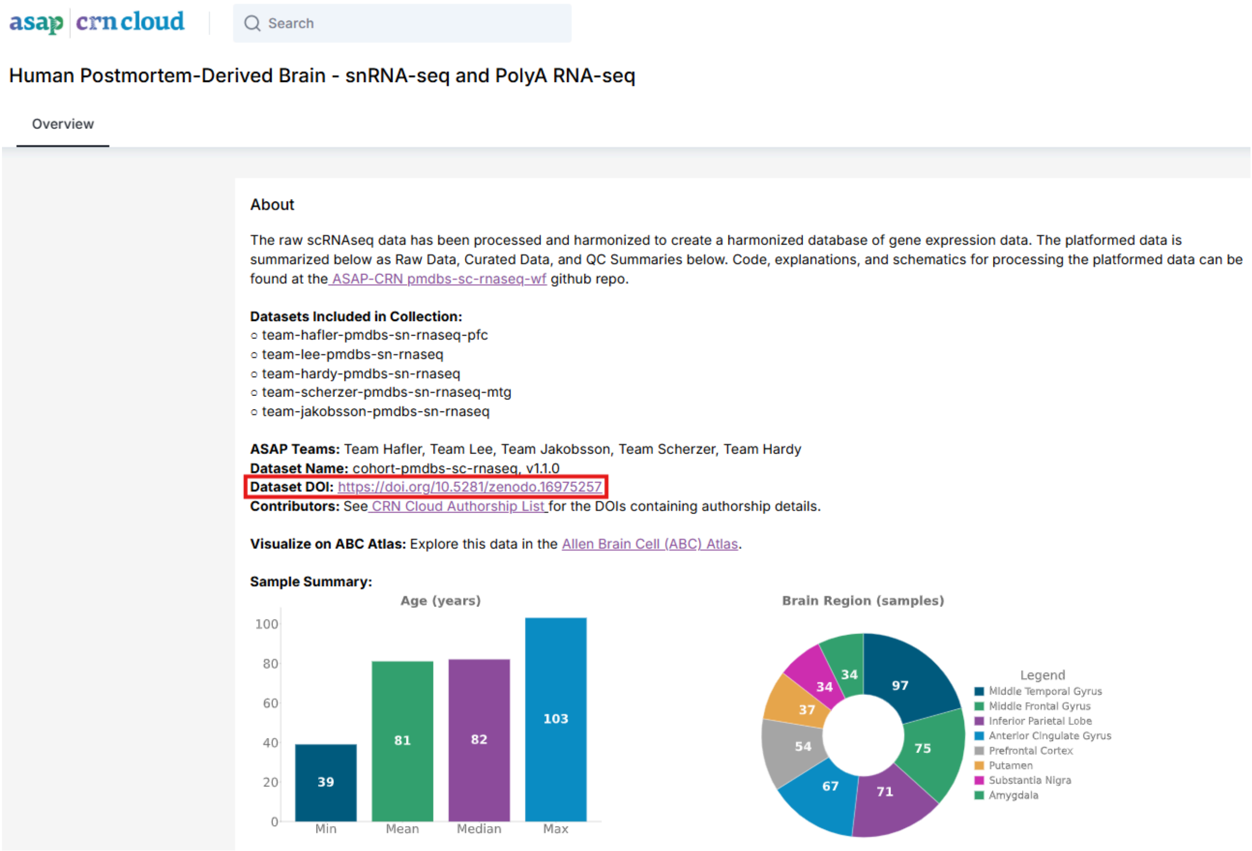
Do datasets submitted to CRN Cloud receive a DOI?

Yes, a citable DOI referencing each dataset is included with its release.

How do I cite a CRN Cloud dataset?

All data released on CRN Cloud can be cited using the DOIs maintained by the CRN Cloud team, and anyone reusing the data must credit the original submitters in their manuscript or acknowledgements section.

The DOI for an individual dataset or a harmonized collection can be found on the 'Overview' tab of the dataset collection card. An example is shown below for reference:



Human Postmortem-Derived Brain - snRNA-seq and PolyA RNA-seq

About

The raw scRNAseq data has been processed and harmonized to create a harmonized database of gene expression data. The platformed data is summarized below as Raw Data, Curated Data, and QC Summaries below. Code, explanations, and schematics for processing the platformed data can be found at the ASAP-CRN pmdbs-sc-rnaseq-wf github repo.

Datasets Included in Collection:
○ team-hafler-pmdbs-sn-rnaseq-pfc
○ team-lee-pmdbs-sn-rnaseq
○ team-hardy-pmdbs-sn-rnaseq
○ team-scherzer-pmdbs-sn-rnaseq-mtg
○ team-jakobsson-pmdbs-sn-rnaseq

ASAP Teams: Team Hafler, Team Lee, Team Jakobsson, Team Scherzer, Team Hardy
Dataset Name: cohort-pmdbs-sc-rnaseq, v1.1.0
Dataset DOI: https://doi.org/10.5281/zenodo.16975257
Contributors: See CRN Cloud Authorship List for the DOIs containing authorship details.

Visualize on ABC Atlas: Explore this data in the Allen Brain Cell (ABC) Atlas.

Sample Summary:

We've also provided an easy reference list of contributors to be used to credit contributing teams when publishing. For guidance on properly citing CRN Cloud and acknowledging data contributors, please refer to terms 12 and 13 of the data use agreement.

Data Access & Usage

What are the data use agreement terms?

The terms of the data use agreement are available here.

How do I access data that is on the CRN Cloud?

The CRN Cloud has a User Manual to guide you through the process of accessing data for analysis. This manual includes detailed instructions on:

- Applying for data access
- Navigating the platform
- Exploring data, and
- Utilizing Verily Workbench for analysis

Can organizations use one account for all users that are members of your research group?

No, the Data Use Agreement contains terms that *each user* must sign.

Are CRN Cloud users required to submit an application for each CRN Cloud dataset they want to access?

No, an approved data access request grants *access to all* CRN Cloud data.

Do CRN Cloud users need to create a Verily Workbench account?

No, each approved user will be added to the CRN Cloud Verily Workbench Reference Workspace. Page 16 of the CRN Cloud User Manual provides guidance on this topic.

How is CRN Cloud data stored?

CRN Cloud data is stored on Google Cloud Platform (GCP), within a system of data buckets provisioned to development, test, and production environments. Only data within production data buckets are exposed/available to CRN Cloud users.

Can CRN Cloud accommodate sensitive data (GDPR compliance)?

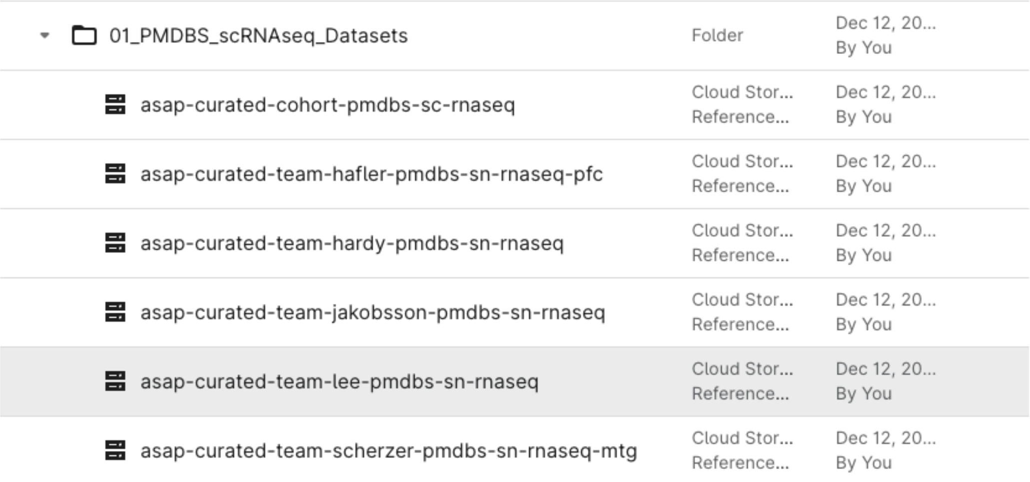
Yes, CRN Cloud is GDPR compliant. The security provided by Google's cloud architecture, our data use policy, and the intentional design of our data governance processes and metadata all support GDPR compliance.

What will be the costs associated with accessing and analyzing this data?

For CRN members, there is no cost associated with:

- Accessing metadata
- Accessing individual dataset and/or harmonized collection outputs from our analysis pipelines including:
  - QC plots, analysis summaries, and gene expression anndata objects,
  - results and reports from Cell Ranger alignment, gene expression tabulation, and CellBender ambient RNA artifact removal.

These outputs are labelled as *'...curated…'*, on Verily Workbench (example screenshot below):



01_PMDBS_scRNAseq_Datasets	Folder	Dec 12, 20... By You
asap-curated-cohort-pmdbs-sc-rnaseq	Cloud Stor... Reference...	Dec 12, 20... By You
asap-curated-team-hafler-pmdbs-sn-rnaseq-pfc	Cloud Stor... Reference...	Dec 12, 20... By You
asap-curated-team-hardy-pmdbs-sn-rnaseq	Cloud Stor... Reference...	Dec 12, 20... By You
asap-curated-team-jakobsson-pmdbs-sn-rnaseq	Cloud Stor... Reference...	Dec 12, 20... By You
asap-curated-team-lee-pmdbs-sn-rnaseq	Cloud Stor... Reference...	Dec 12, 20... By You
asap-curated-team-scherzer-pmdbs-sn-rnaseq-mtg	Cloud Stor... Reference...	Dec 12, 20... By You

For non-CRN members, there is no cost associated with accessing metadata. Non-CRN members are responsible for any costs related to downloading and analyzing individual dataset and/or harmonized collection outputs in the CRN Cloud, as described in more detail above. A Google Cloud Billing Account is required. Please contact workbench-support@verily.com, once your access is granted, to establish a spend pod.

All users, CRN and non-CRN members, are responsible for covering the costs associated with:

- Data transfer when requesting raw data in its original, unchanged form (i.e., fastqs), as submitted by the data contributors of individual datasets.

The raw data is available in Requester Pays buckets. This means that the requester must assume the cost for data transfer when requesting the data. To facilitate handling the cost, you must:

1. Set up a Google Cloud Billing Account owned by you (i.e. the same Google identity you use to log in to Explorer and Verily Workbench)
2. Create a Google Cloud Project and associate it with the Billing Account created in Step (1)
3. Request the raw data using the Project created in Step (2).

Detailed instructions of how to access and analyze raw data are available in Verily Workbench.

Does the CRN Cloud offer visualization tools?

Yes, through partnership with the Allen Institute, the CRN Cloud <> ABC Atlas integration pilot enables researchers to seamlessly visualize and explore the CRN Cloud harmonized single-cell dataset (i.e., Human Postmortem-Derived Brain - snRNA-seq and PolyA RNA-seq Collection) within the Allen Brain Cell (ABC) Atlas environment.

The goal of the pilot is to improve discoverability and interpretation of CRN data by placing it in the context of a widely used reference atlas. This integration also supports collaboration and shared understanding across research teams by providing a common framework for comparing cell types, signatures, and biological patterns. Access the pilot integration here.