CosMx™ SMI Data Export Guide

MAN-10179-01

This guide describes the steps to export CosMx Spatial Molecular Imager (SMI) data from the AtoMx™ Spatial Informatics Platform (SIP).

Choose Destination for Data Export

To transfer files to your own AWS S3 bucket, follow instructions in Option 1. Setup for Export to an AWS S3 Bucket on page 2. To transfer files to a NanoString storage bucket accessible through secure file transfer protocol (SFTP), follow instructions in Option 2. Setup for Export to NanoString Storage using SFTP on page 11. Export to your own AWS S3 bucket is preferred since it allows greater control over and independence in accessing the data, without time constraints (data in the NanoString storage bucket is kept for 2 weeks).

Option 1. Setup for Export to an AWS S3 Bucket

The following steps are provided to supplement AWS user documentation. Refer to Getting Started with Amazon S3 for more comprehensive instructions and technical support and/or or speak with your institution's Informatics or IT teams. AWS Free S3 5 GB plan will suffice for the transfer and storage of most studies (excluding raw data files). (Transfer of larger files is permitted with this plan, but will incur a modest cost). NanoString is gathering data to inform on the best plan options for different use cases.

1. Create an AWS account.

2. Sign in as a Root user (Figure 1).

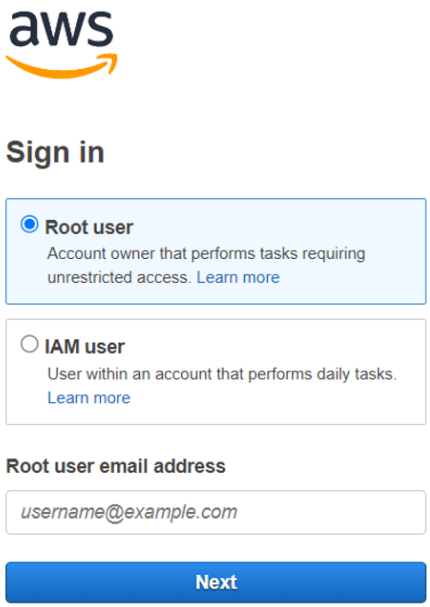

Figure 1: Sign in as root user

3. Click on **Services** in the top left of the screen, then click **Storage**, then **S3** (Figure 2).

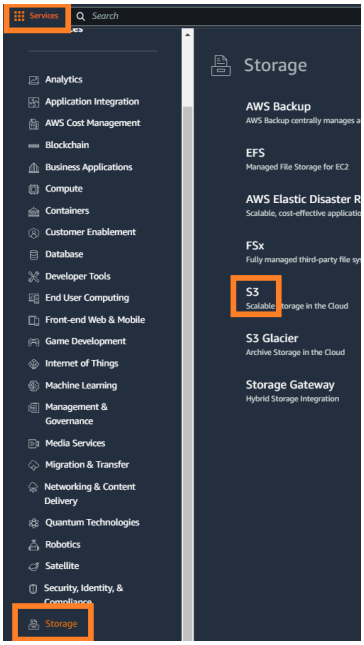

Figure 2: Storage window: S3

4. Click **Create bucket** (Figure 3; your view may be different).

Figure 3: Create bucket button

5. Fill in **Bucket name** – do not use spaces, uppercase or special characters (Figure 4). Choose the **AWS region** that matches the AtoMx SIP AWS (if known) (e.g. US East, EU, etc). Leave all other options on this page as the default values. Click **Create bucket.**

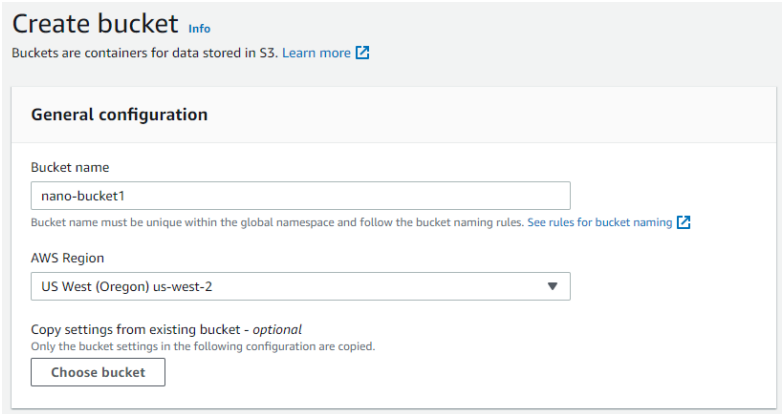

Figure 4: Fill in bucket name and choose AWS region

6. Return to **Buckets** and note the AWS region associated with your bucket (Figure 5). Click on the name of your newly created bucket to access it.

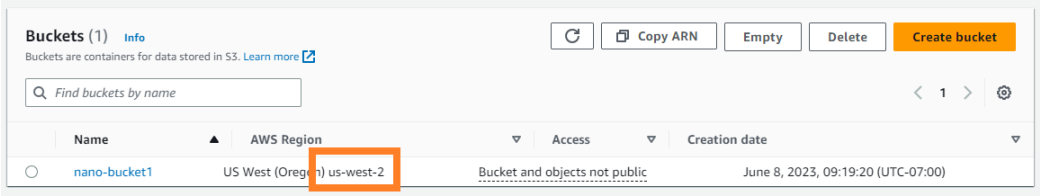

Figure 5: Buckets window

7. Click **Create folder** to make a new destination folder for AtoMX SMI data export (Figure 6).

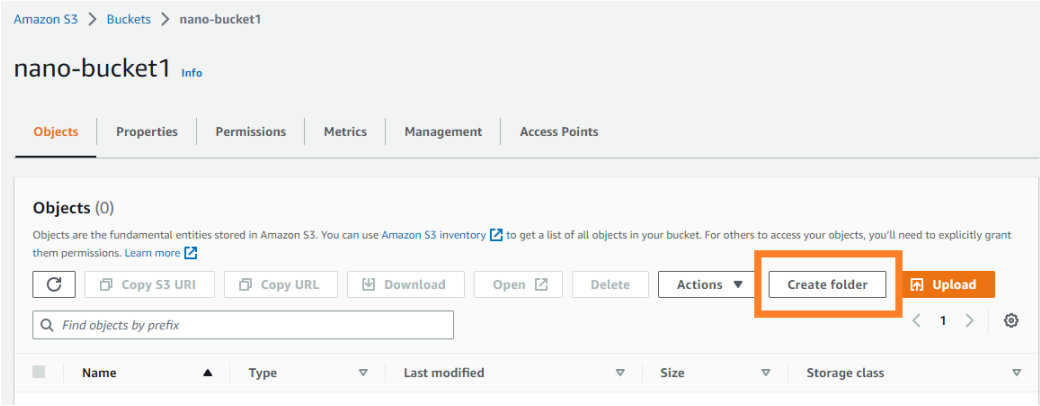

Figure 6: Create folder button

8. Fill in the **folder name (do not use spaces or special characters).** Select **Encryption key type: Amazon S3 managed keys**. Click **Create folder** (Figure 7).

9. Check the box to the left of your newly created folder and click **Copy S3 URI** (the URI should be in the format: s3://atomxtest/S3demo/) (Figure 8). This is the **destination S3 file path** which you will input to the CosMx SMI Export module in AtoMx SIP.

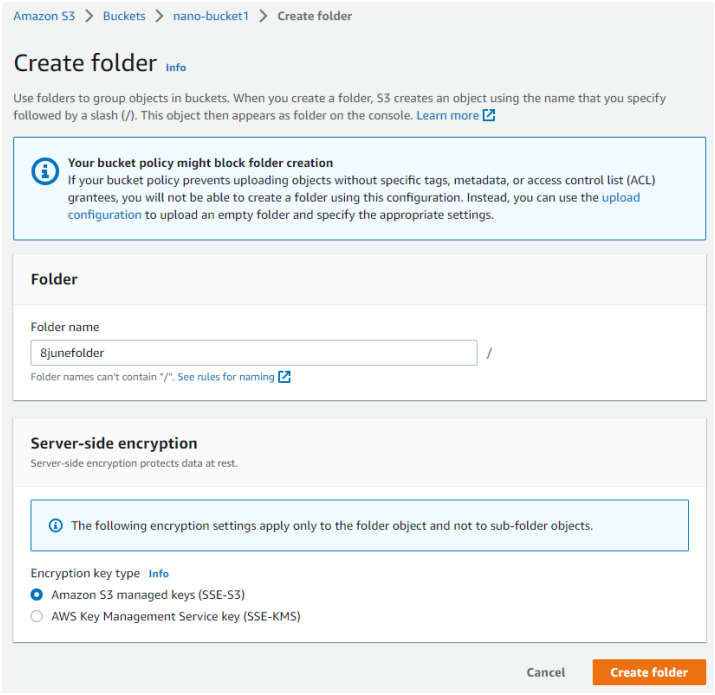

Figure 7: Enter a folder name

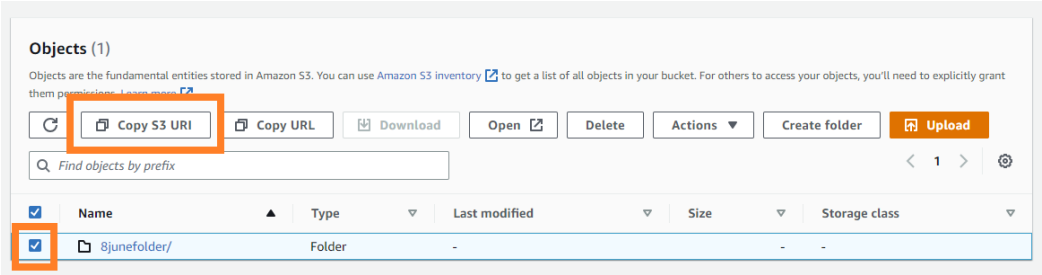

Figure 8: Copy S3 URI button

10. Next, you will generate access keys to access this S3 bucket. Click on **Services**, then **Security, Identity, & Compliance,** then select **IAM** (Figure 9).

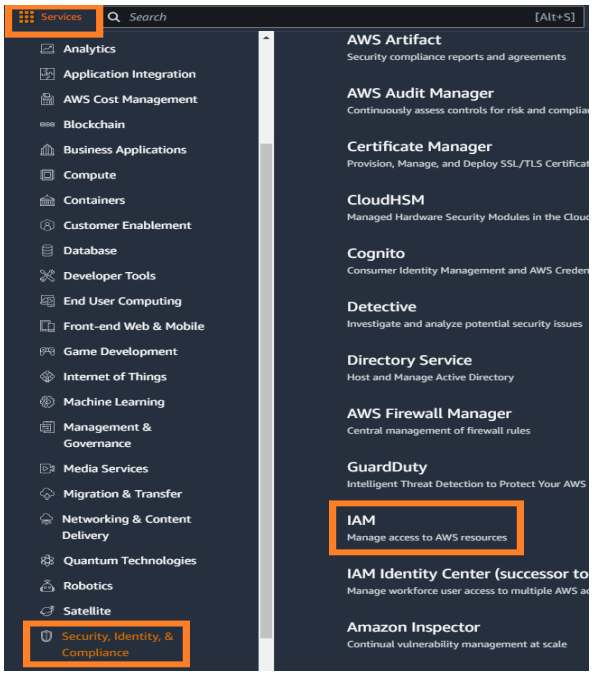

Figure 9: IAM settings

11. Select **Users** from the left menu (Figure 10).

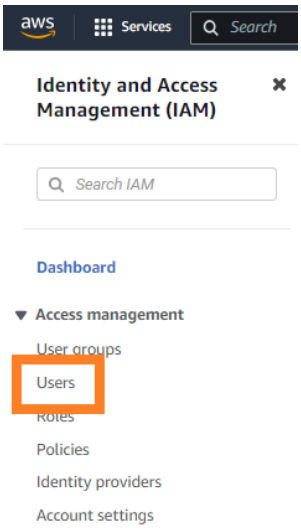

Figure 10: Select Users

12. Click **Add users** (Figure 11).

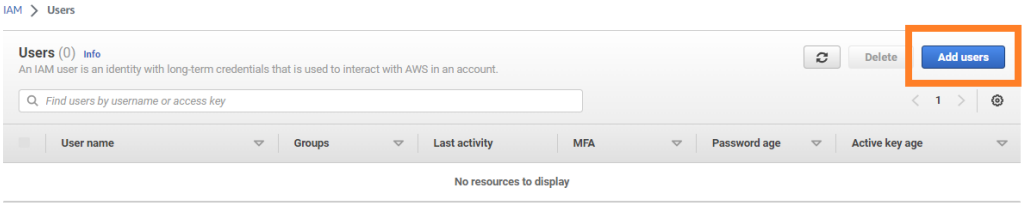

Figure 11: Add users button

13. In the User details window, specify a user name (Figure 12). Click **Next.**

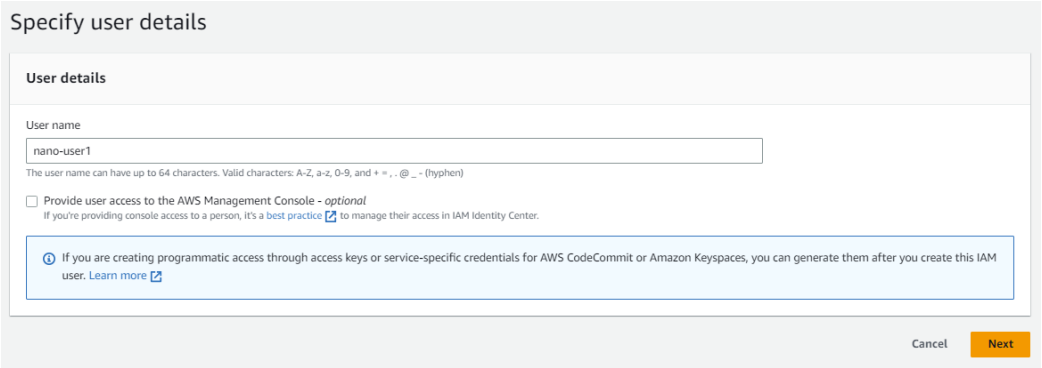

Figure 12: Specify a new user name

14. In the **Set permissions** window, select **Permission options: Add user to group**, then click **Create group** (Figure 13).

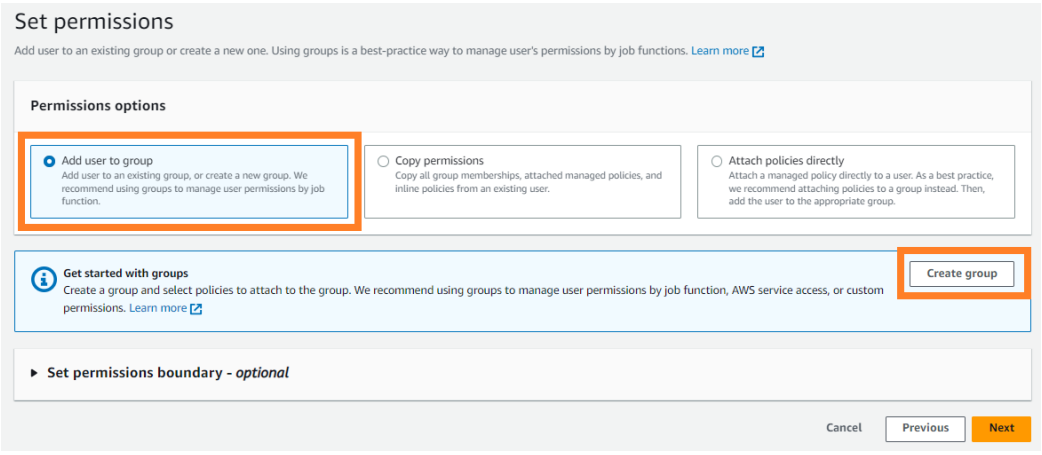

Figure 13: Set permissions window

15. In the window **Create user group**, create a user group name without spaces (Figure 14). Under **Permissions policies**, search for **S3** then check the box for **AmazonS3FullAccess**. Click **Create user group**.

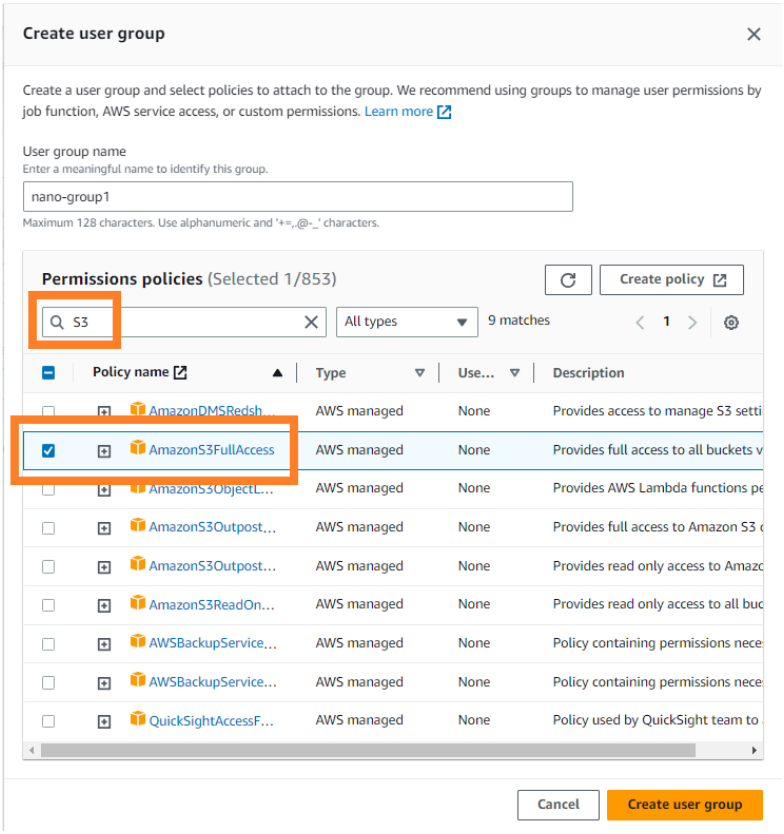

Figure 14: Create user group window

16. You should see the User group you've just created in the list **User groups** (Figure 15). Check the box to the left of the name, then click Next.

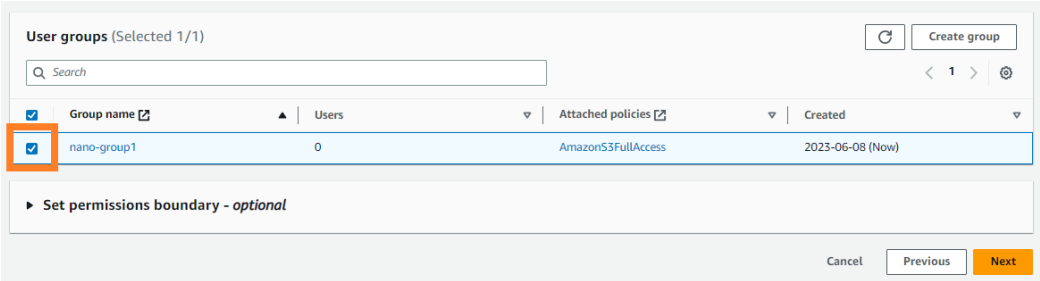

Figure 15: Check the box next to the user group name

17. In the **Review and create** window, confirm that the group name is listed in the Permissions Summary (Figure 16). Click **Create User.**

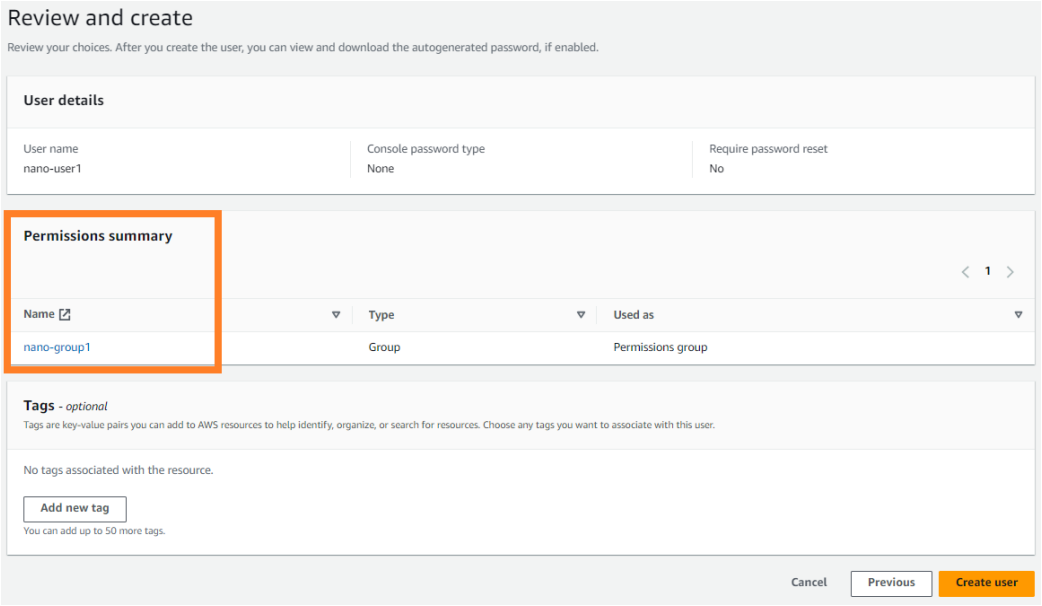

Figure 16: Create user - permissions summary

18. A pop-up message indicates that the user is created successfully (Figure 17). Click **View user.**

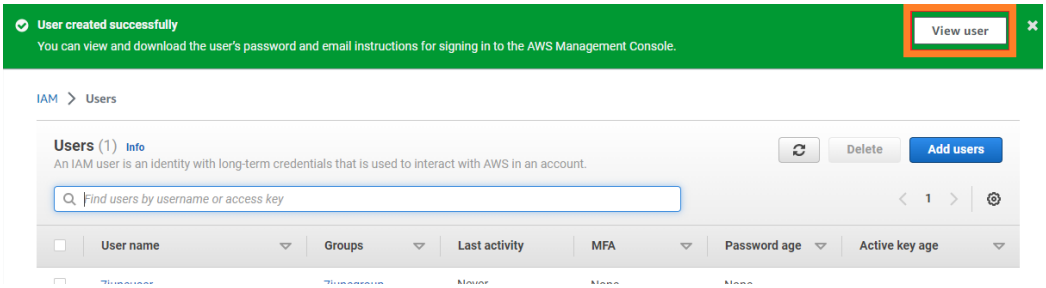

Figure 17: View user button

19. Click on the tab **Security credentials** (Figure 18).

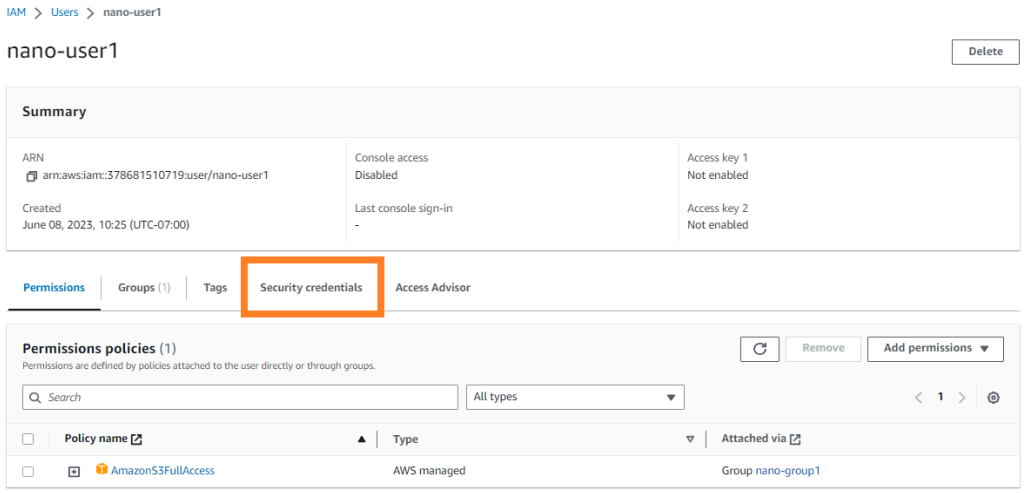

Figure 18: Security credentials from username menu

20. Scroll down to **Access keys** and click **Create access key** (Figure 19).

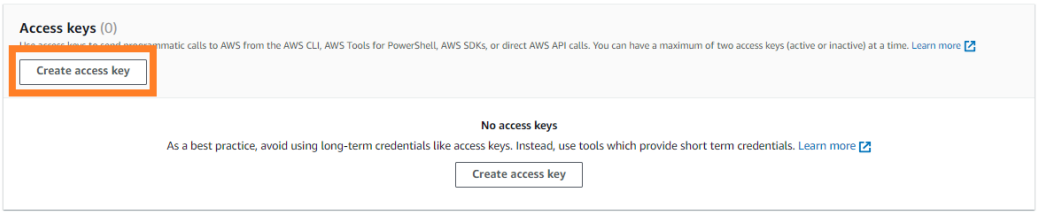

Figure 19: Create access key button 1

21. You may be presented with a list of alternatives to access keys (Figure 20). Select **Other** and click **Next**.

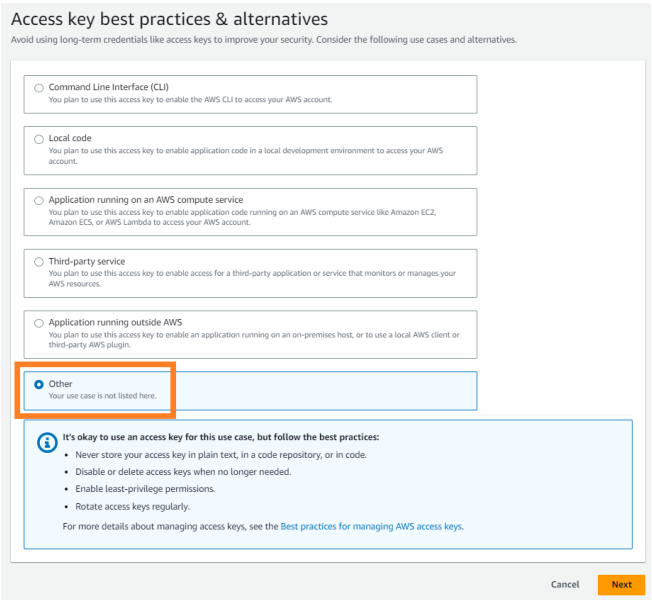

Figure 20: Alternatives to access keys

22. Optionally enter a description tag for the key, then click **Create access key** (Figure 21).

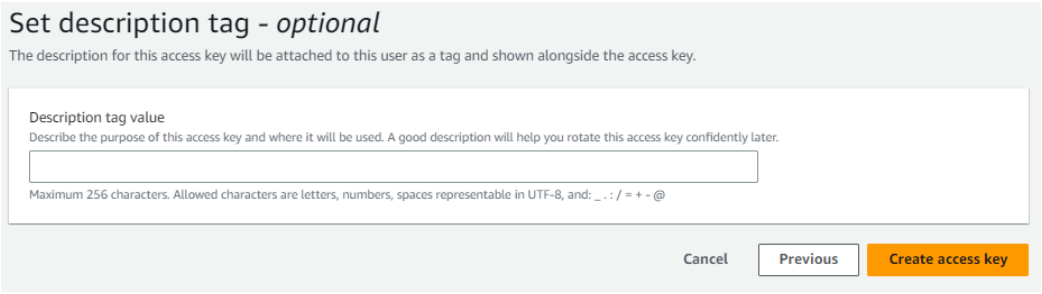

Figure 21: Optional description for access key

23. Copy the access key and secret access key by clicking each copy icon (Figure 22) and pasting them in a safe place. Click **Download .csv file** for additional security backup to document your S3 access key and secret access key.

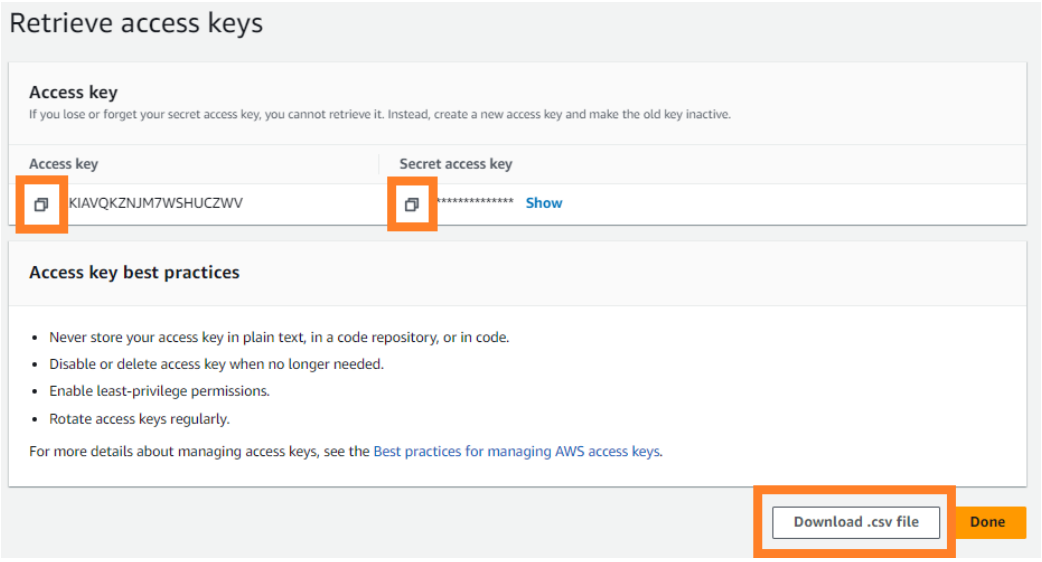

Figure 22: Retrieve access key credentials

24. The destination S3 file path (or S3 URI), access key, secret key, and region are then entered into the fields of the CosMx SMI Data Analysis Export module (refer to instructions on page 15).

Option 2. Setup for Export to NanoString Storage using SFTP

If it is not feasible to export data to a user's own AWS S3 bucket, CosMx SMI data may be securely exported to a storage bucket managed by NanoString, accessible through SFTP. This bucket is configured securely to only accept data from the AtoMx environment, and credentials are secured as such to access data provided by one tenant.

To gain access credentials to the secure storage location, please contact NanoString Support at support@nanostring.com and request **FRM-11152 Credentials to Access AtoMx SFTP Site for Data Export**. Please include your CosMx SMI instrument serial number in your request. The form will be populated with access credentials specific to your tenant.

Input the information from the form when setting up the Export Module (see Export Data on page 12).

Please be aware that data in the NanoString storage location will be retained for 2 weeks.

Export Data

Use the CosMx SMI data export module and instructions (below) to copy data from AtoMx SIP to a user-specified location: the user's own AWS S3 bucket, or NanoString's storage location for transfer by SFTP. Export to a user's own AWS S3 bucket is preferred since it allows greater control over and independence in accessing the data, without time constraints (data in the NanoString storage bucket is kept for 2 weeks).

1. From AtoMx SIP, open the study of interest.
2. If you have already used the Export module to export data, then the script is already loaded in the software; skip to step 3. To load the Export module for the first time,
   a. Click the **custom modules** icon in the Pipeline Run panel to open Custom Modules.
   b. Click **Add Module** to open the Add new module window (Figure 23).
   c. Name the module **Export.** Upload the **Export script** obtained from NanoString.
   d. In the Module Description field, enter text to be displayed in the Export module parameters pane at a later step. For example, click on the **Code** tab of the custom module window and copy the text in lines 4-7, describing the parameters. Paste this text into the Module Description field. This helpful information will then be displayed when you are making selections from the module parameters.
   e. Select the maximum **CPU Cores** from the dropdown menu. Select **RAM (GB)** and **Max run (h):** for most exports, 128 GB RAM and 24 hours is sufficient. For a very large dataset (e.g. the fullSeuratObject with many pipeline runs or analysis steps), increase the RAM. Note that increasing RAM causes a reduction in maximum number of computing nodes allowed. If exporting data from many FOVs or with a large number of transcripts, increase the time limit to avoid a time-out error.
   f. In the section **Arguments** (Figure 23), click the **+** button to add the following arguments (see Table 1). Ensure **Name** is spelled exactly as shown, as this value is the variable in the script.
   g. Click **Save.** The Export module should now appear in the dropdown list of Custom Modules. Close the Custom Modules window.

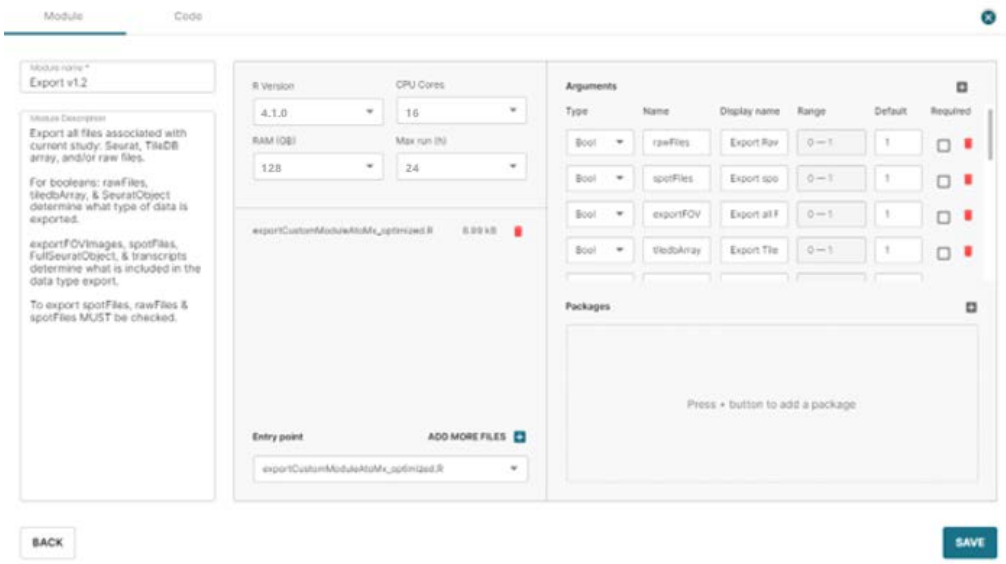

Figure 23: Loading the Export module into Custom Modules

Table 1: Arguments in the Export module

Type	Name	Display name	Range	Default	Reqd.
Bool	rawFiles	Export raw files	0-1	1	☐
Bool	spotFiles	Export spot files (warning: large data)	0-1	1	☐
Bool	exportFOVImages	Export all FOV images (warning: large data)	0-1	1	☐
Bool	tiledbArray	Export TileDB array	0-1	1	☐
Bool	SeuratObject	Export a Seruat Object	0-1	1	☐
Bool	FullSeuratObject	Seurat object will contain all previous module output	0-1	1	☐
Bool	transcripts	Seurat object will contain transcript coordinates (warning: large data)	0-1	1	☐
String	studyName	Output Folder Name	min-max		☑
String	outPath	Destination S3 file path	min-max		☑
String	access_key	Destination AWS access key	min-max		☑
String	secret_key	Destination AWS secret key	min-max		☑
String	s3Region	Destination AWS region	min-max		☑
String	session_token	Destination AWS session token, if configured	min-max		☐

3. Click the Pipeline icon in the Pipeline Run panel to open the Run Pipeline window.
4. If you choose to incorporate the Export module to an existing pipeline, select the pipeline of choice from the Pipeline dropdown menu, then click the pencil icon to edit it. Alternatively, create a new pipeline with the Export module, click **Create Pipeline** to open the Create New Pipeline window.
5. Drag the Export module from the list at left to the gridded workspace and connect it to an existing module (Figure 24). The Export module will export all data in the study, regardless of placement in a pipeline.

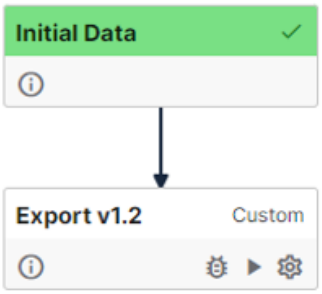

Figure 24: Export module in the pipeline workspace

6. Click the gear icon on the Export module (Figure 24) to input the parameters for exporting:
   a. Confirm Module version: Version 1 (Figure 25) .

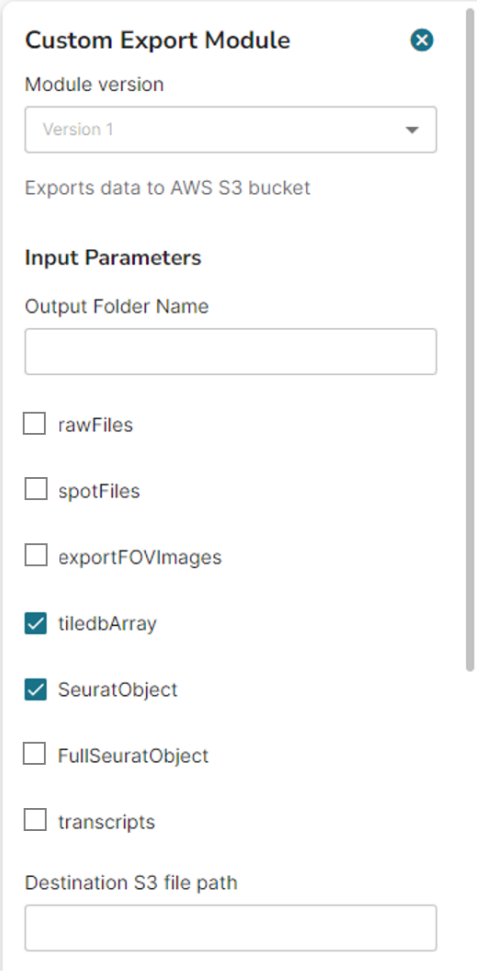

Figure 25: Export module parameters

b. Check the boxes of the files you wish to include in the Export (Figure 25). Keep in mind these considerations:

- rawFiles: large data - will significantly increase the export file size. If exporting raw files, you may also choose to export:

  • spotFiles: very large data - include **only** if you will repeat Target Decoding outside of the software.

  • exportFOVImages: include **only** if you need the raw images on the FOV level. Otherwise, export composite RGB images using Image Viewer (see the CosMx SMI Data Analysis User Manual (MAN-10162)).

- tiledbArray: exports a default tertiary analysis structure. NanoString is developing a vignette or tutorial on the tiledbArray; please inquire with your Product Application Scientist.

- SeuratObject: exports a Seurat object comprised of counts, metadata, and dimensional reduction outputs. If exporting SeuratObject, you may also choose to export:

  • FullSeuratObject: if selected, the exported Seurat object will additionally include all analysis results from modules run, like pathway scores, cell type log probabilities, and DE outputs.

  • transcripts: if selected, the exported Seurat object will additionally include transcript coordinates.

c. In **Output Folder Name** (Figure 26) create a name for the output folder, which will appear within the destination folder of the S3 bucket. **Do not use spaces or special characters** in the folder name (underscores are acceptable).

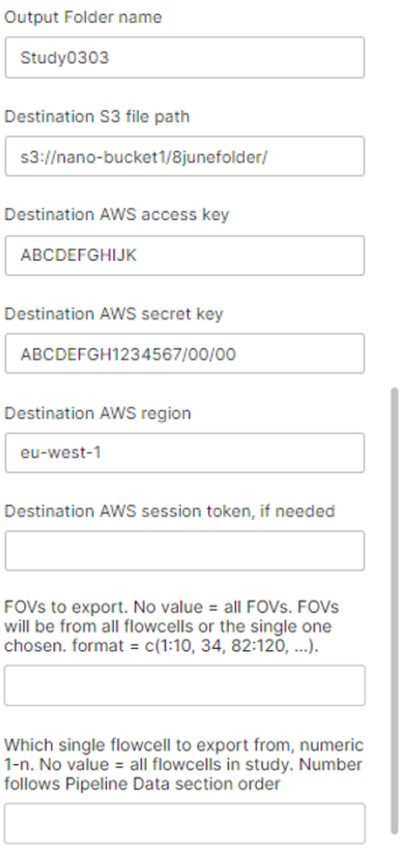

Figure 26: Export module parameters, continued (example inputs shown)

d. In **Destination S3 file path,** enter the S3 URI of the destination folder in the S3 bucket.

- If exporting to your own S3 bucket, copy the S3 URI from the AWS S3 console (select the destination folder using the check mark on the left, then click **Copy S3 URI**). The format should be S3://... **Do not use spaces or special characters in the destination file path.** Refer to Option 1. Setup for Export to an AWS S3 Bucket on page 2 for details.
- If exporting to NanoString's storage location for SFTP, copy the Destination S3 file path from the tenant-specific information provided on form **FRM-11152 Credentials to Access AtoMx SFTP Site for Data Export**. The format should be S3://... **Do not use spaces or special characters in the destination file path.** Refer to Option 2. Setup for Export to NanoString Storage using SFTP on page 11 for details.

e. In **Destination AWS access key** and **Destination AWS secret key,** enter the AWS keys corresponding to your S3 bucket or the NanoString storage location. Refer to Option 1. Setup for Export to an AWS S3 Bucket on page 2 or Option 2. Setup for Export to NanoString Storage using SFTP on page 11 for details.

f. In **Destination AWS region,** enter the region name of your S3 bucket as it appears in your Amazon S3 console under the tab **Properties,** or the region name of the NanoString storage location as it appears on form FRM-11152, in the format us-west-1.

g. In **Destination AWS session token,** enter the S3 session token (only if your AWS S3 bucket has been configured for it; read more at https://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp_use-resources.html).

h. In **FOVs to export,** leave blank to export all FOVs, or input a subset of FOVs as shown in the following example: to export FOVs 1-10, 34, and 82-120, enter c(1:10, 34, 82:120). (c is the R function to combine multiple values into vectors).

> NOTE: If selecting a subset of FOV or flow cells, be aware that the data dimensions for the exported cell metadata and RNA assay count matrices may appear mismatched, since the cell metadata captures all the cells with a transcript of at least 1 across the entire study (including various controls) as opposed to the RNA assay count matrices, which rely only on RNA assay counts.

i. In the final field, leave blank to export from all flow cells or enter a numerical value to identify one of the flow cells in this study.

j. Click **Save** or **Save & Create Run.**

7. Back in the Pipeline Structure panel, click **Run All.** For a study of 4 FOV (1000 plex panel) without raw files, export should take 5-10 minutes.

8. When the Export run is completed,
   - If exporting to your own S3 bucket, check the contents of your S3 bucket from the S3 console to confirm the transfer of files.
   - If exporting to NanoString's storage location for SFTP, access the SFTP site with the personalized credentials provided on form FRM-11152 to confirm the transfer of files.
9. Note that the AWS **access and secret keys may still be visible** in the Export module parameters (Figure 23). It is recommended to delete the text from these fields, close the module parameters pane, and refresh the browser to maintain the privacy of the keys.

From there, you can manage the data in the AWS environment or, if desired, download the data to the local environment. Refer to Download CosMx SMI Files After Export on page 18 for addtional support.

AWS provides comprehensive user documentation. See, for example:

- https://docs.aws.amazon.com/AmazonS3/latest/userguide/creating-buckets-s3.html
- https://aws.amazon.com/s3/pricing/

Export Output

The output folder structure is shown here (Figure 27).

- Seurat object
- TileDB array
- Raw files (if included with export)

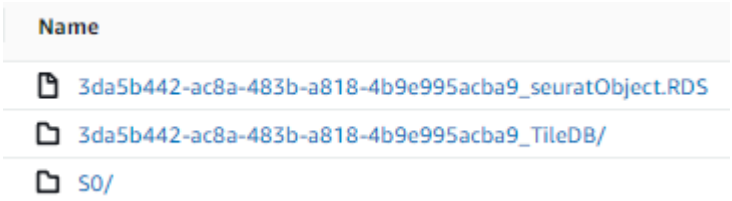

Figure 27: Example export output folder structure

FOV images, if included in the data export, are in the Morphology2D subfolder within the raw files folder. These stacked TIFF files show each channel's image, in the order: blue, green, yellow, red, UV. Refer to the Image Viewer - Render Settings tab to see which marker is in each channel for your experiment.

The following computational packages are required to interact with the exported data.

- Seurat: R Toolkit for Single Cell Genomics. Install in RStudio using `install.packages("Seurat")`
- tiledbsc: an R implementation of the Stack of Matrices, Annotated (SOMA). Install in RStudio using `remotes::install_github("tiledb-inc/tiledbsc", force = TRUE)`
- tiledbr: an R interface to the storage engine of TileDB. Install in RStudio using `remotes::install_github("TileDB-Inc/TileDB-R", force = TRUE)`

Download CosMx SMI Files After Export

From Your AWS S3 Bucket

NOTE: It may be preferable to manage your data in AWS rather than download it to a local environment. Downloading from S3 (file egress) incurs costs according to AWS pricing structure. Please refer to AWS documentation for additional support managing your data in AWS.

Files can be downloaded using the S3 console; folders must be downloaded using command line interface (CLI). The following steps may be performed using AWS CLI (preferred) or Anaconda Prompt.

IMPORTANT: Before beginning, ensure that the local environment has enough free space to accommodate the files.

1. Install the AWS CLI script package on your computer (or, if using Anaconda Prompt, install the Anaconda Client).
   a. Navigate to the AWS Command Line Interface website and download the appropriate installer from the links on the right (Figure 28).
   b. Follow the prompts to install CLI to your computer.

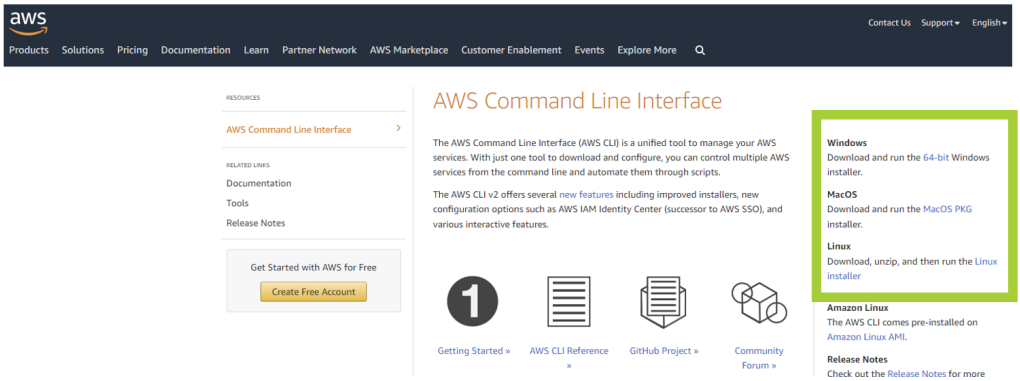

Figure 28: Download AWS CLI Installer

2. Once install is finished, open Command Prompt on Windows by pressing the Windows key on your keyboard and typing ‘command’ (Figure 29). Click the Command Prompt to open.

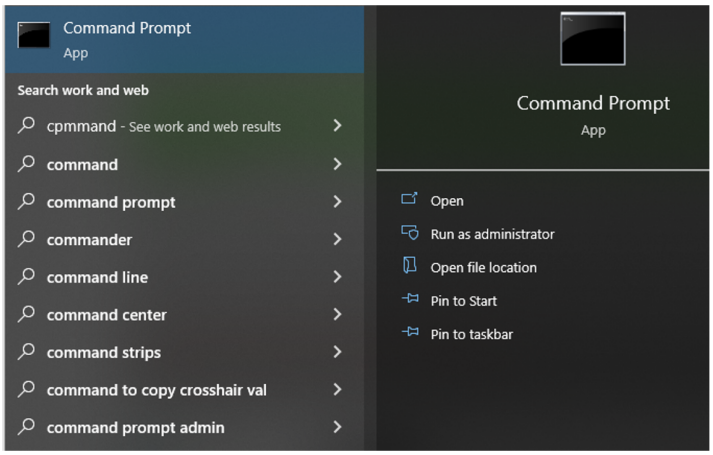

Figure 29: Open Command Prompt

3. Type in `aws configure` (Figure 30).
   a. When asked to provide the Access Key, enter the specific key for the IAM user.
   b. When asked to provide the Secret Key, enter the specific key for the IAM user.
   c. Set as region: enter the region of the S3 bucket, such as `us-west-1` or `eu-central-1`. This information can be found in the main folder of your S3 bucket.
   d. Set as output format: `text`

```
C:\>aws configure
AWS Access Key ID [None]:
AWS Secret Access Key [None]:
Default region name [None]: eu-west-1
Default output format [None]: text
```

Figure 30: Input for AWS Configure command

4. Type in `aws s3 ls` to check whether the bucket is configured correctly. The output should show the name of your S3 bucket.
5. Within the command prompt, navigate to the folder to which to download the data. For example, if the target folder is 'testdownload' on your C: drive, type `cd testdownload` (Figure 31). The command `cd..` navigates back one step in the C: drive folder hierarchy.

```
C:\>cd testdownload

C:\testdownload>
```

Figure 31: Navigate to target folder

6. Within your AWS S3 account, navigate to the folder to download and select it with a checkmark (Figure 32). Click on **Copy S3 URI** to copy the folder location.

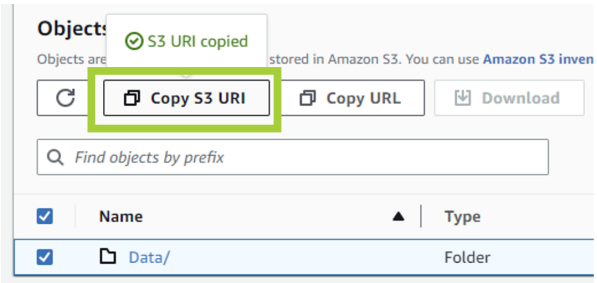

Figure 32: Copy S3 folder URI

7. In the command prompt, type `aws s3 sync` then paste the URI copied from the S3 folder in Step 7. Following the pasted URI, type a space and then period `.` (Figure 33).

```
C:\testdownload>aws s3 sync s3://mholperts3/Test/Data/ .
```

Figure 33: Sync CLI to S3 folder

8. The download will begin once you press Enter. The duration of download time depends on the size of the files.

From NanoString's Storage Location using SFTP

Once the Export pipeline run has completed, the data will appear in the designated SFTP site. Access the site using the credentials provided in form **FRM-11152 Credentials to Access AtoMx SFTP Site for Data Export**. Ensure that the local environment has enough free space to accommodate the files, and that your IT organization permits this type of traffic. Use any standard SFTP program such as WinSCP, Cyberduck, or FileZilla to transfer the files to a local drive.

Study data will be retained in the SFTP site for 2 weeks.

For technical support, please contact your NanoString representative or email NanoString Technical Support at support@nanostring.com.