Alissa Reporter Using Amazon Web Services

Technical Guide

For Research Use Only. Not for use in diagnostic procedures.

About Alissa Reporter Amazon Web Services

Alissa Reporter allows you to connect to an Amazon Web Services (AWS) S3 storage location containing sequencing data to be analyzed. You can use the connection between Alissa Reporter and AWS to configure Alissa Reporter for AWS automation or for manual upload of files from an S3 location when using the Alissa Reporter upload wizard. Both approaches are described below.

AWS automation

AWS automation allows you to use Alissa Reporter in an automated way. Sequencing data placed in an AWS S3 location that is defined in the settings of the Alissa Reporter customer account are automatically uploaded into Alissa Reporter and analyzed. Results of the analysis can automatically be exported and stored in the same AWS S3 location.

Whenever you have new sequencer files to analyze, deposit them into the account's defined AWS location in a compatible format along with a summary.yaml file containing the needed meta information. Alissa Reporter periodically scans the AWS S3 location for new data. When new data is present, Alissa Reporter imports the files, creates a run, and performs the analysis.

See **"Set up and configure AWS automation"** on page 3 for instructions.

S3 upload wizard

The Alissa Reporter upload wizard, which walks you through the steps for setting up and starting a sequencing run, requires that you upload the sequencing data by selecting the specific files to be analyzed. You can configure Alissa Reporter to be able to access an AWS S3 storage folder containing the sequencing files.

Unlike data files analyzed with AWS automation, this approach does not move any files in the AWS S3 storage folder after analysis and it does not automatically export the analysis results to a selected folder. The AWS S3 location serves only as a folder for file storage that can be accessed by the upload wizard.

See **"Set up and configure AWS S3 storage folder"** on page 15 for instructions.

Set up and configure AWS automation

Step 1. Create an account on AWS

Check with your organization on creating a new AWS account and follow steps 1 through 5 on the AWS website (**AWS Management Console (amazon.com)**).

Follow steps
1 ->5

After account set up
-> Sign in

Step 2. Create bucket using the S3 console

1   Sign in to the console **Amazon Web Services Sign-In**.
2   From the Overview page, click **Services** in the top left corner and select **S3** under **Storage** (**S3 Management Console (amazon.com)**).
3   Click **Create bucket** and provide a bucket name.

The bucket name consists of a mandatory part (`agilent-alissa-c2s-`) followed by the customer's choice of bucket name. For example: `agilent-alissa-c2s-s3demo`.

4 Under **Bucket Versioning**, select **Enable**.

5 Under **Default encryption**, select **Disable** OR **Enable** with the option **SSE encryption**.

The Alissa Reporter application supports unencrypted buckets and SSE encrypted buckets natively via the policy file that can be generated in the Settings and Administration section in Alissa Reporter. KMS encrypted buckets are currently not supported in Alissa Reporter.
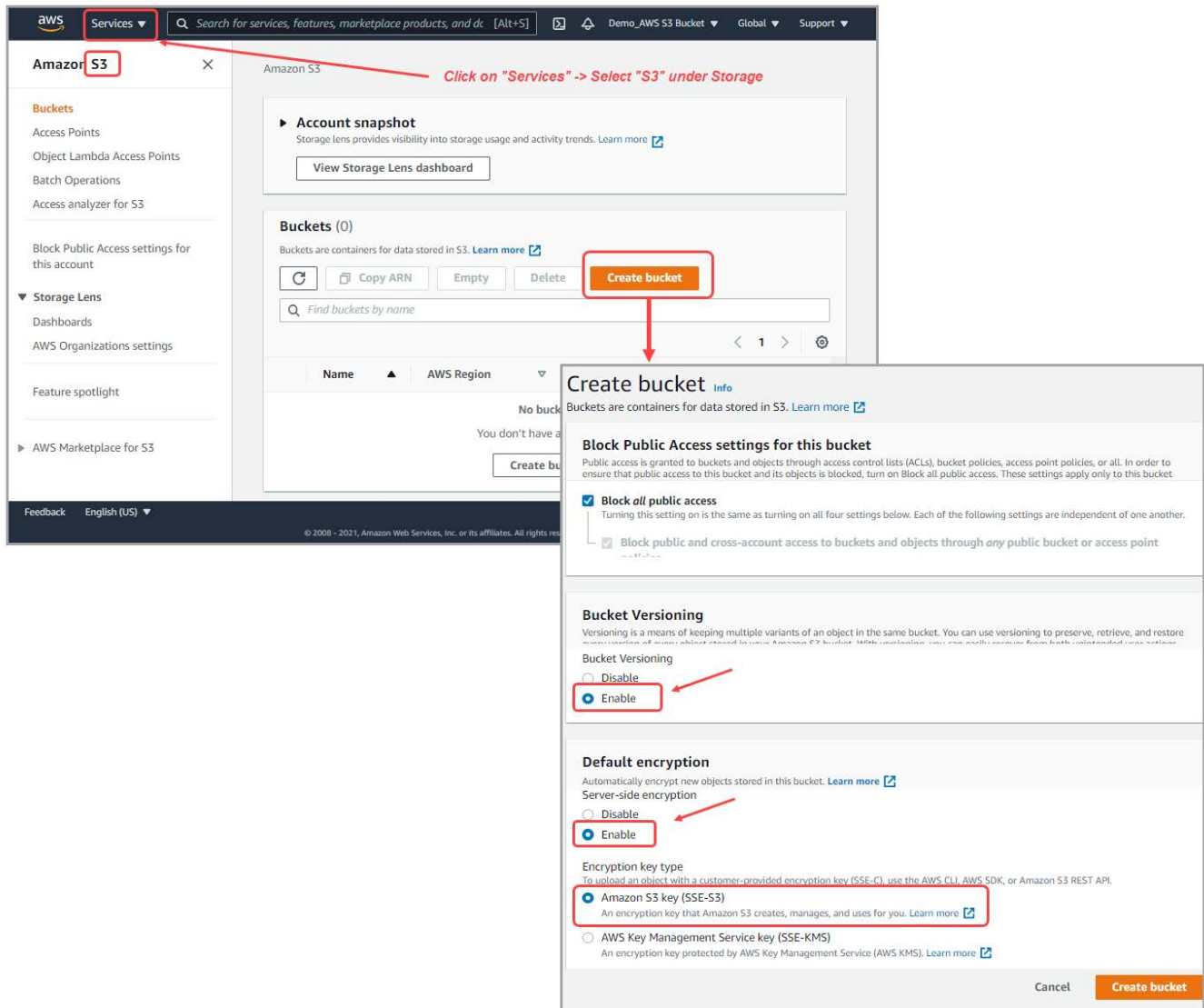
The name of the bucket must be unique (not shared between different accounts). Preferably, the bucket is located within the same region where the Alissa Reporter deployment is located (speed and cost considerations). The following AWS regions are used for Alissa Reporter:

- eu-west-1 region for EU deployment (eu.reporter.alissa.agilent.com): Ireland

- us-east-1 region for US deployment (us.reporter.alissa.agilent.com): N. Virginia

- ap-southwest-2 region for AP deployment (ap.reporter.alissa.agilent.com): Sydney, Australia

Additional information on how to create an S3 bucket is available at the AWS website: **https://docs.aws.amazon.com/AmazonS3/latest/userguide/create-bucket-overview.html**.

Alissa Reporter Using AWS Technical Guide

Step 3. Configure AWS settings in Alissa Reporter and download policy

To activate the link between Alissa Reporter and AWS, a bucket policy needs to be downloaded from Alissa Reporter and pasted into AWS. This policy is based on the account ID and prevents the sharing of AWS credentials between customers and our software framework.

In Alissa Reporter, click **Settings & Administration > My Account > Manage account settings** to open the Manage settings popup. Then, follow the instructions below to create a policy from your account using your AWS bucket settings.
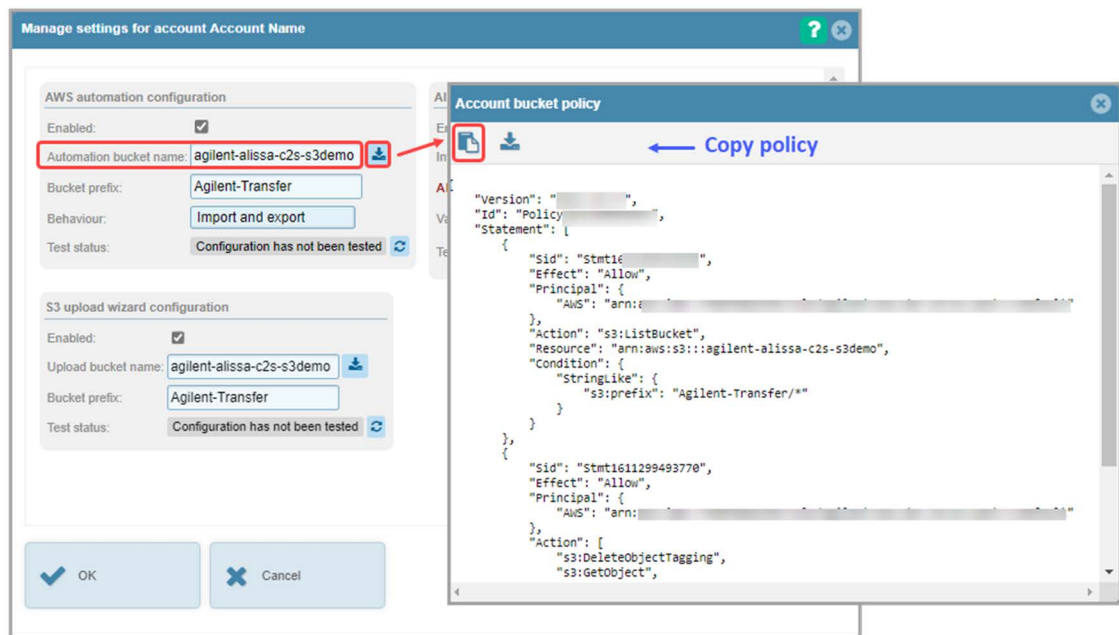
1 Mark the Enabled check box to enable the AWS automation configuration.

2 In **Automation bucket name**, enter the **Bucket name** starting with the mandatory agilent-alissa-c2s- followed by the bucket name as described in **"Step 2. Create bucket using the S3 console"**. For example, **agilent-alissa-c2s-s3demo**.

3 (Optional) Enter the **Bucket prefix**. This optional (but recommended) prefix reflects a subfolder in the AWS bucket. For example, **Agilent-Transfer**.

4 View the policy by clicking the ⬇ arrow and copy the policy from the Account bucket policy popup window.

For additional details on the Manage settings popup, please see the corresponding help pages in Alissa Reporter.

> **NOTE** If you have an account on multiple Alissa Reporter deployments and want to use the same bucket for the different deployments, contact **Agilent Technical Support** for help with updating your bucket policy to allow for that option.

Step 4. Paste the policy in your bucket and disable access control lists (ACLs)

Sign in into the AWS Management Console, go to the Amazon S3 console at **https://console.aws.amazon.com/s3/**, and open your bucket. Then, follow the instructions below.

1 Go to the Permissions tab.

2 Under **Bucket Policy**, click **Edit**.

3 Paste the policy into the provided space.

4 If needed, disable ACLs for the bucket.

For newly created buckets, ACLs are disabled by default. If you just created a new bucket, ACLs are already disabled (as indicated in the Object Ownership settings) and no changes are needed. You can proceed directly to **"Step 5. Test AWS automation configuration in Alissa Reporter"** on page 8.

For buckets created before disabling of ACLs became the default, ACLs may still be enabled. If you are using a bucket that was created previously, follow **step a** through **step d** below to ensure that ACLs are disabled.

a Scroll further down the Permissions tab to the Object Ownership section.

b Under **Object Ownership**, click **Edit**.

The Edit Object Ownership screen opens.

c Make sure that the **ACLs disabled** option is selected.
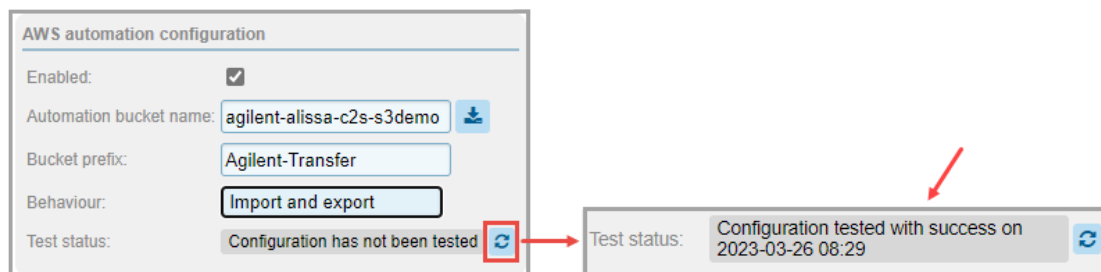
d Click **Save Changes**.

Step 5. Test AWS automation configuration in Alissa Reporter

1   Return to your session of Alissa Reporter with the Manage settings popup open.

2   Verify that all fields are properly filled out (see **"Step 3. Configure AWS settings in Alissa Reporter and download policy"**).

3   Under **AWS automation configuration**, next to **Test Status**, click the test ⟳ button.

The test status indicates testing was successful. The entire setup is correct, permission was granted, and AWS automation is activated.

Alissa Reporter places a "lock" file in the bucket that uniquely links the AWS automation bucket to your Alissa Reporter account/deployment. Removing the lock file from the bucket will stop any existing automation configured for the bucket and allows another account to set up a new bucket using that bucket name.

In case the test status is not successful, verify that all previous steps were performed properly, correct possible errors, and repeat the testing. Contact the Alissa Reporter technical support team at **informatics_support@agilent.com** for assistance if testing remains unsuccessful.

> **NOTE** The creation of the policy file only needs to be done when activating a new bucket. Once permission is granted, it remains active unless configuration changes are made by the user (e.g., modification/removal of the bucket).

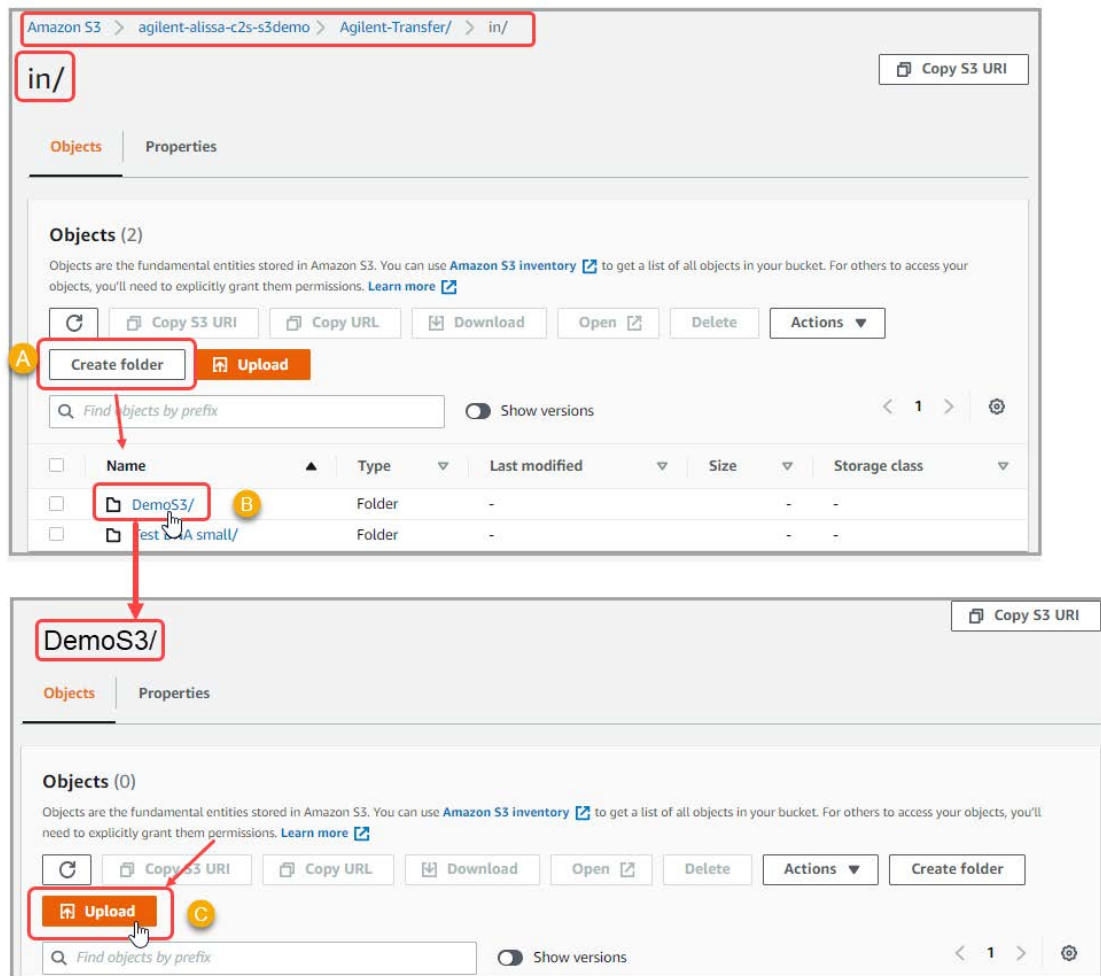## Step 6. Create a run directory and upload sequencer files

The next step is creating a run directory in your bucket. The run directory contains the sequencer run files (FASTQ) and an associated *summary.yaml* file that will be transferred to Alissa Reporter. The run directory name will be used as the **Run ID** in Alissa Reporter.

> **NOTE** Agilent recommends using a unique name for the run directory, in particular when using the export option from Alissa Reporter to AWS (see also **"Step 8. View and consult analysis results in the S3 bucket"**).

The run directory has the following format: *<Bucket Name>/<Bucket prefix>/in/<my_run_name>/*. For example, `agilent-alissa-c2s-s3demo/Agilent-Transfer/in/DemoS3`.

1 Create the run directory and upload your FASTQ sequencer run files under the newly created run directory. Refer to the image above. For details on the format of the sequencer run files, see the corresponding help pages in Alissa Reporter.

After successful upload, the complete path will look as follows:
<Bucket Name>/in/<my_run_name>/<my_sequencer_files>

Or if you entered a prefix (as described in **"Step 3. Configure AWS settings in Alissa Reporter and download policy"**):
<Bucket Name>/<Bucket prefix>/in/<my_run_name>/<my_sequencer_files>

Step 7. Create summary.yaml file and upload file in the run directory

Alissa Reporter requires an associated *summary.yaml* file to initiate the automatic transfer of the run from AWS to Alissa Reporter. The summary.yaml file contains the meta information (e.g., run, samples, in silico filter information, and user account information).

1 Create a summary.yaml file and edit the meta information as shown in the image on the following page.

A template of the summary.yaml file is available for download in Alissa Reporter. From the **Manage account settings** popup of your Alissa Reporter account, open the help page and click the provided link to download the summary.yaml template.

2 Upload the summary.yaml file into the run directory.

Agilent recommends uploading the summary.yaml file after the sequencer run files have been successfully uploaded in the run directory to prevent initiating transfer to Alissa Reporter without the necessary sequencer files being present for transfer.

When the import starts, a preliminary meta information check is done (on the summary.yaml file):

• If successful, the run transfer proceeds. After the full import has finished, the run folder with the files is moved from the folder *in/<my_run_name>* to the folder *finished/<my_run_name>*. Meta tags are added to the summary.yaml file in the bucket.

• If an error occurs, the run folder with the files is moved from the folder *in/<my_run_name>* to the folder *error/<my_run_name>* and meta tags are added to the summary.yaml file in the bucket with the error message.
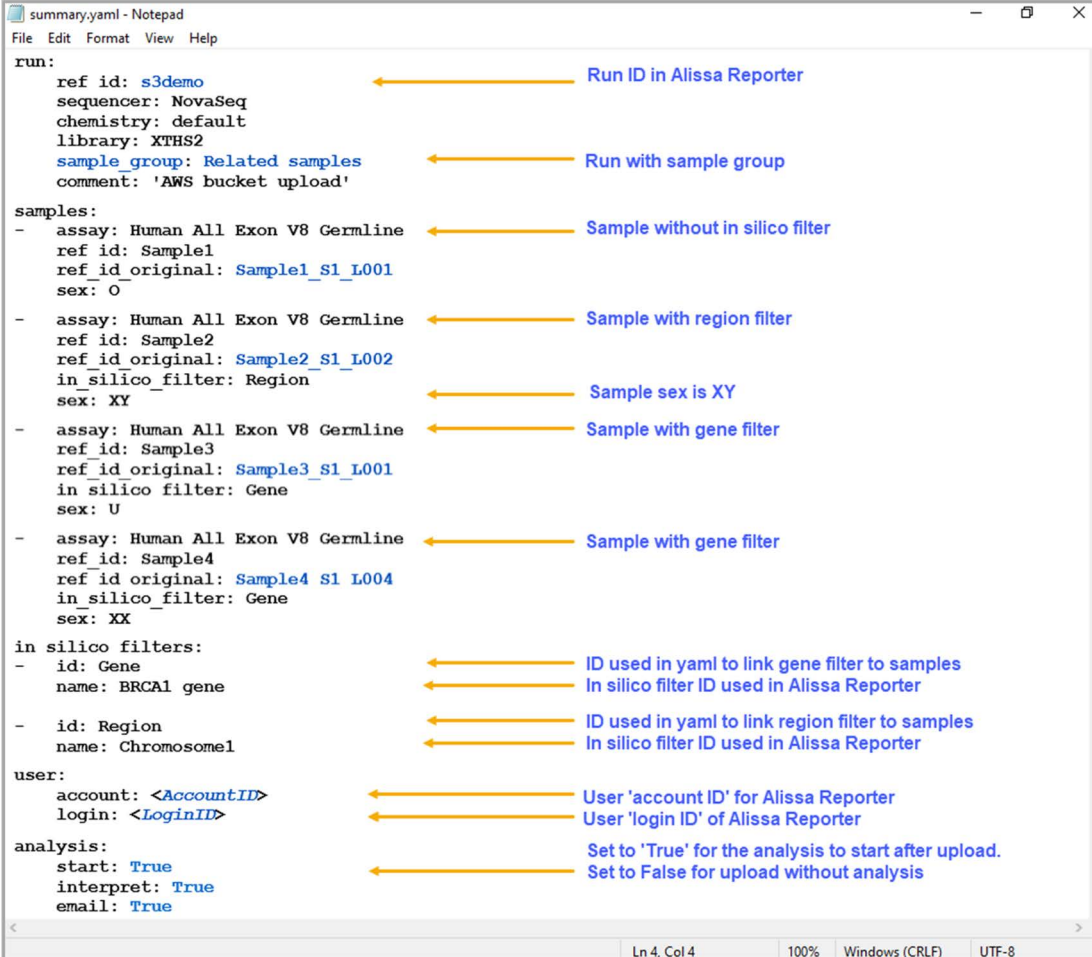
NOTE

- The run is analyzed according to the analysis parameters defined in the Manage options popup of your Alissa Reporter account. These analysis parameters can be set to the default settings or they can have customized settings.
- For somatic applications, a matched/unmatched reference sample can be specified in the summary.yaml file by defining it under the *category* section. Note that if a run only contains reference samples, the analysis will not complete when **Start analysis** is set to **true** in the summary.yaml file.
- In-silico filters specified in the summary.yaml file can either refer to existing in-silico filters in your Alissa Reporter account or they can define entirely new filters. They can be specified on the run level or the sample level in the summary.yaml file.

  To specify new in-silico filters in the summary.yaml, define a *genes_file* (with file extension *tsv*) and/or a *regions_file* (with file extension *tsv*) with the selected genes and/or regions using the same format as the exported files from Alissa Reporter. The specified .tsv file(s) in the summary.yaml file need to be uploaded in the run directory together with the summary.yaml file. If the summary.yaml specifies an in-silico filter with the same name as an already existing in-silico filter, the file information provided in the summary.yaml is ignored.
- A sample group can be specified on the run level or sample level in the summary.yaml file. For each run, a new unique sample group ID is created in the system in the form of **<name> (<ref_id>)** obtained from the summary.yaml.
- Optional analysis settings can be specified in the summary.yaml file including 'start' (to start the analysis after upload completed), 'interpret' (to send the results to your Alissa Interpret instance) and 'email' (to receive an email when analysis completes).

```
summary.yaml - Notepad                                                    —   □   ×
File  Edit  Format  View  Help
run:
      ref id: s3demo                        ←————————— Run ID in Alissa Reporter
      sequencer: NovaSeq
      chemistry: default
      library: XTHS2
      sample_group: Related samples         ←————————— Run with sample group
      comment: 'AWS bucket upload'
samples:
-     assay: Human All Exon V8 Germline      ←————————— Sample without in silico filter
      ref id: Sample1
      ref_id_original: Sample1_S1_L001
      sex: O

-     assay: Human All Exon V8 Germline      ←————————— Sample with region filter
      ref id: Sample2
      ref_id_original: Sample2_S1_L002
      in_silico_filter: Region
      sex: XY                               ←————————— Sample sex is XY
-     assay: Human All Exon V8 Germline      ←————————— Sample with gene filter
      ref_id: Sample3
      ref_id_original: Sample3_S1_L001
      in silico filter: Gene
      sex: U

-     assay: Human All Exon V8 Germline      ←————————— Sample with gene filter
      ref_id: Sample4
      ref_id original: Sample4_S1_L004
      in_silico_filter: Gene
      sex: XX
in silico filters:
-     id: Gene                              ←————————— ID used in yaml to link gene filter to samples
      name: BRCA1 gene                      ←————————— In silico filter ID used in Alissa Reporter

-     id: Region                            ←————————— ID used in yaml to link region filter to samples
      name: Chromosome1                     ←————————— In silico filter ID used in Alissa Reporter
user:
      account: <AccountID>                  ←————————— User 'account ID' for Alissa Reporter
      login: <LoginID>                      ←————————— User 'login ID' of Alissa Reporter

analysis:                                              Set to 'True' for the analysis to start after upload.
      start: True                           ←————————— Set to False for upload without analysis
      interpret: True
      email: True

                                    Ln 4, Col 4        100%   Windows (CRLF)    UTF-8
```

in_silico_filters:
- id: Gene
  name: BRCA1 gene

- id: Region
  name: Chromosome1

Existing in silico filter

in_silico_filters:
- id:
  name:
  genes_file:
  regions_file:

New in silico filter

Step 8. View and consult analysis results in the S3 bucket

If the AWS automation configuration settings in Alissa Reporter are set to include export (as determined by the Behavior setting in the Manage settings popup), then the analysis results are automatically exported to the S3 bucket and can be viewed in the "out" folder under the run directory, e.g.: ***<Bucket name>/out/<my_run_name>***.
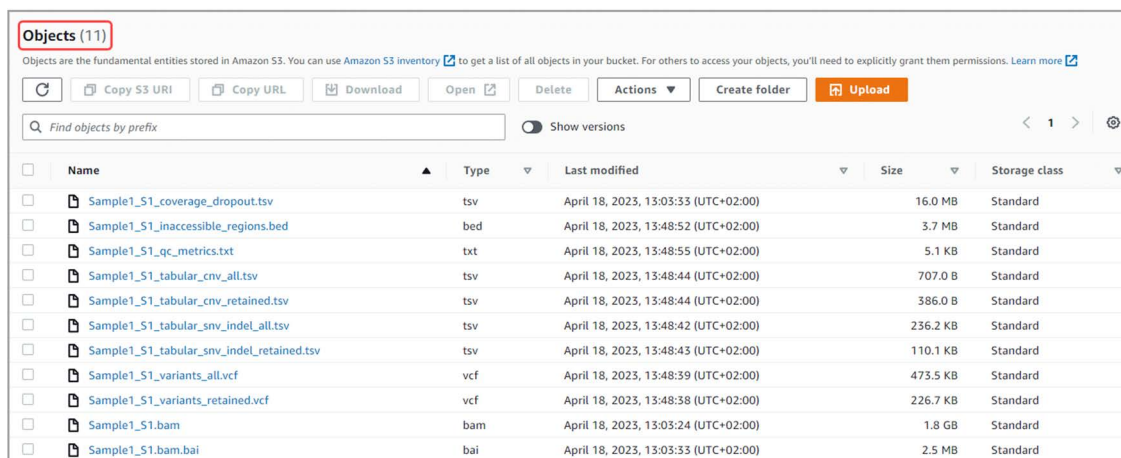
The ***<my_run_name>*** folder name created by AWS is the run ID used in Alissa Reporter. Note that if a folder name already exists in the bucket, AWS automation overwrites existing results and the existing files will be replaced.

Individual subfolders are listed for each successfully analyzed sample of the run and each sample subfolder contains the respective results files. Note that if a sample ID was edited in Alissa Reporter, the edited sample information is used.

An example of a successful Human All Exon analysis is shown below. A total of 11 result files are present.

Each time you have a new set of sequencer files to be analyzed, repeat **"Step 6. Create a run directory and upload sequencer files"** through **"Step 8. View and consult analysis results in the S3 bucket"**.

Alissa Reporter Using AWS Technical Guide

Set up and configure AWS S3 storage folder

Step 1. Create an account on AWS

If you do not already have an AWS account, refer to the instructions in **"Step 1. Create an account on AWS"** on page 3 in the section **"Set up and configure AWS automation."**

Step 2. Create bucket using the S3 console

If you already created a bucket for use in AWS automation, and you intend to use that same bucket for your AWS S3 storage folders, skip this step and proceed to **"Step 3. Configure S3 wizard upload settings in Alissa Reporter and download policy"** on page 15.

If you have not already set up a bucket using the S3 console, or you intend to use a separate bucket for your AWS S3 storage folders, refer to the instructions in **"Step 2. Create bucket using the S3 console"** on page 3 in the section **"Set up and configure AWS automation."**

Step 3. Configure S3 wizard upload settings in Alissa Reporter and download policy

To activate the link between Alissa Reporter and AWS, a bucket policy needs to be downloaded from Alissa Reporter and pasted into the S3 bucket policy. This policy is based on the account ID and prevents the sharing of AWS credentials between customers and our software framework.

In Alissa Reporter, click **Settings & Administration > My Account > Manage account settings** to open the Manage settings popup. Then, follow the instructions below to create a policy from your account using your AWS bucket settings.
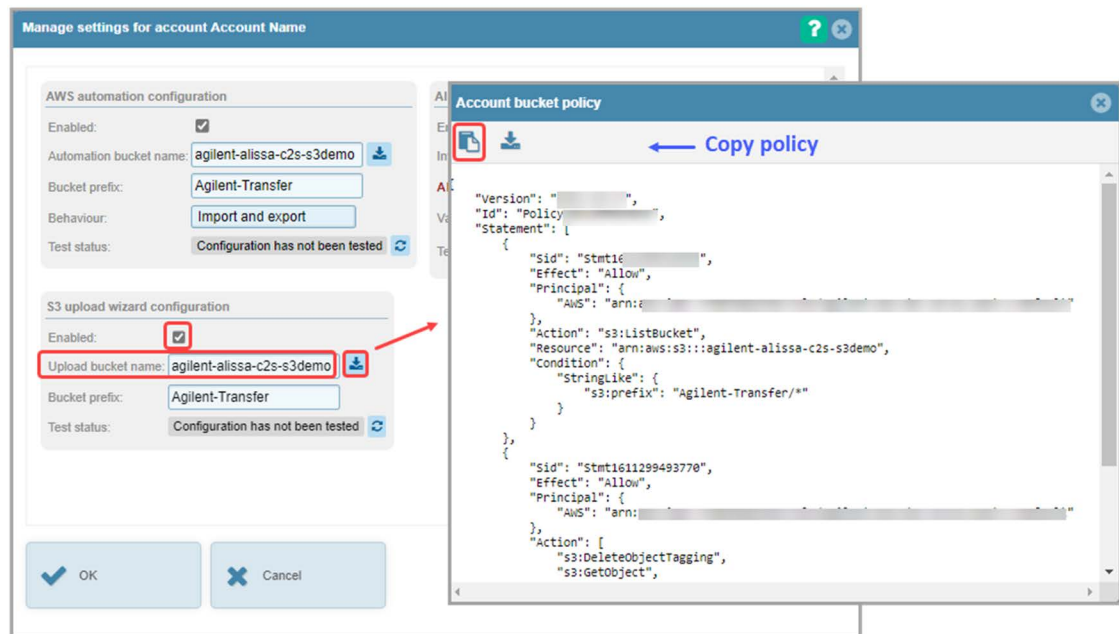
1. Mark the Enabled check box to enable the S3 upload wizard configuration.

2. In **Upload bucket name**, enter the **Bucket name** starting with the mandatory agilent-alissa-c2s- followed by the bucket name as described in **"Step 2. Create bucket using the S3 console"**. For example, **agilent-alissa-c2s-s3demo**.

3. (Optional) Enter the **Bucket prefix**. This optional (but recommended) prefix reflects a subfolder in the AWS bucket. For example, **Agilent-Transfer**.

4. View the policy by clicking the ⬇ arrow and copy the policy from the Account bucket policy popup window.

   Important: If you already created a bucket for use in AWS automation, and you intend to use that same bucket for your AWS S3 storage folders, you do not need to repeat the steps for downloading the policy, pasting it in the bucket, and disabling ACLs. Skip those steps and proceed to **"Step 5. Test S3 upload wizard configuration in Alissa Reporter"** on page 18.

For additional details on the Manage settings popup, please see the corresponding help pages in Alissa Reporter.

> **NOTE**   If you have an account on multiple Alissa Reporter deployments and are using the same bucket on the different deployments, make sure to use a distinct bucket prefix for each deployment, e.g., bucket prefix for deployment A: **Agilent-Transfer/<PrefixA>**, bucket prefix for deployment B: **Agilent-Transfer/<PrefixB>**.

## Step 4. Paste the policy in your bucket and disable access control lists (ACLs)

Sign in into the AWS Management Console, go to the Amazon S3 console at **https://console.aws.amazon.com/s3/**, and open your bucket. Then follow the instructions below.

1 Go to the Permissions tab.

2 Under **Bucket Policy**, click **Edit**.

3 Paste the policy into the provided space.
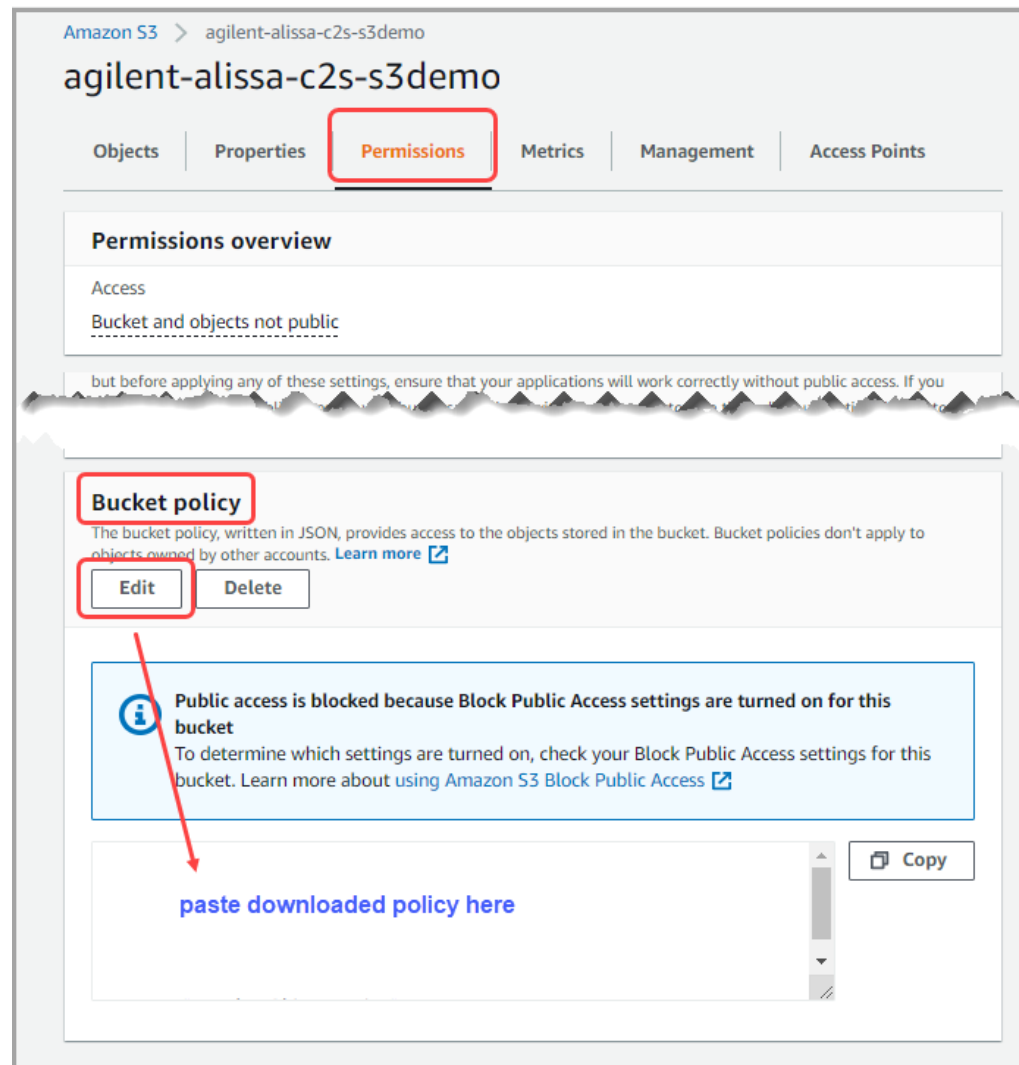
4 Scroll further down the Permissions tab to the Object Ownership section.

5 Under **Object Ownership**, click **Edit**.

The Edit Object Ownership screen opens.

6 Make sure that the **ACLs disabled** option is selected.

7 Click **Save Changes**.
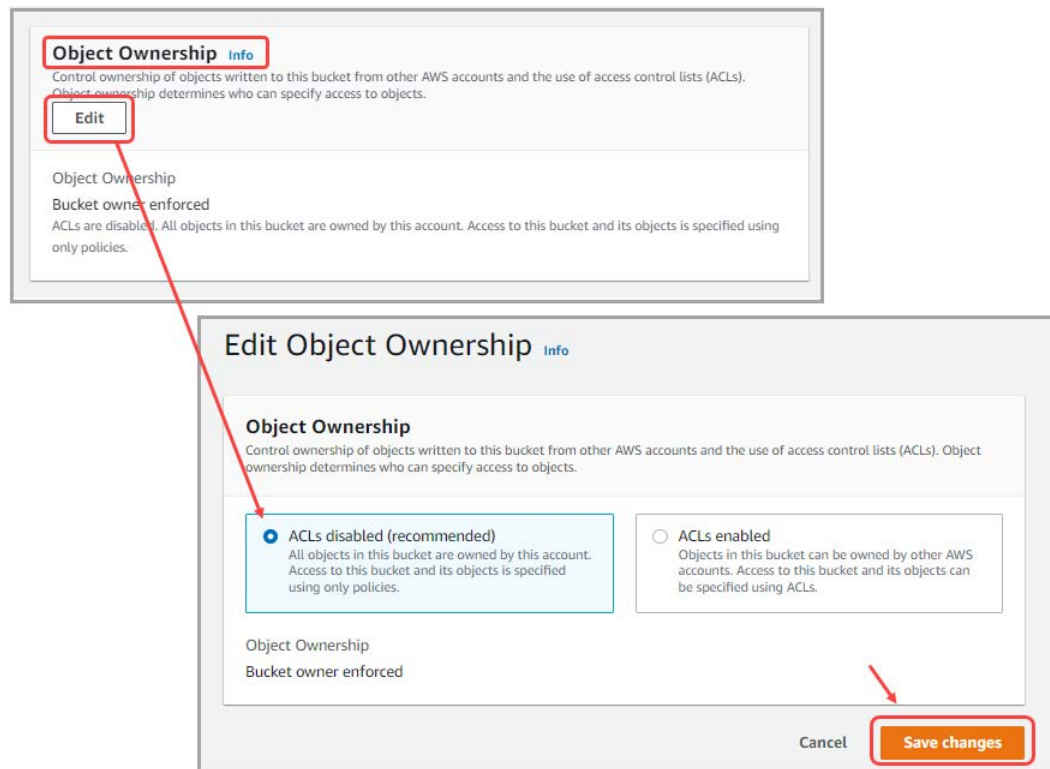
Step 5. Test S3 upload wizard configuration in Alissa Reporter

1 Return to your session of Alissa Reporter with the Manage settings popup open.

2 Verify that all fields are properly filled out (see **"Step 3. Configure S3 wizard upload settings in Alissa Reporter and download policy"**).

3 Under **S3 upload wizard configuration**, next to **Test Status**, click the test 🔁 button.

The test status indicates testing was successful. The entire setup is correct, permission was granted, and access to the S3 folders from the upload wizard is activated.

In case the test status is not successful, verify that all previous steps were performed properly, correct possible errors, and repeat the testing. Contact the Alissa Reporter technical support team at **informatics_support@agilent.com** for assistance if testing remains unsuccessful.
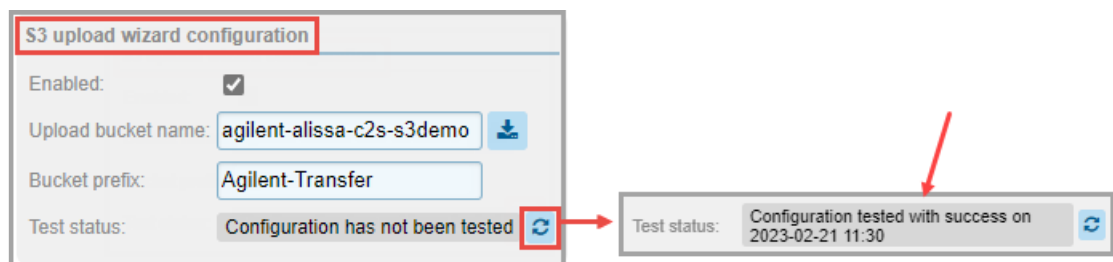
> **NOTE**   The creation of the policy file only needs to be done when activating a new bucket. Once permission is granted, it remains active unless configuration changes are made by the user (e.g., modification/removal of the bucket).
>
> The S3 bucket is not locked and can be used by other Alissa Reporter accounts/deployments.

Step 6. Create an S3 storage folder and upload sequencer files

The next step is creating an S3 storage folder in your bucket. The folder contains the sequencer run files (FASTQ) that can be accessed from the upload wizard.

The storage folder has the following format:
<Bucket Name>/<Bucket prefix>/<my_storagefolder_name>/.

For example, **agilent-alissa-c2s-s3demo/Agilent-Transfer/Samples**, where **Samples** is the S3 storage folder where FASTQ sequencer run files are saved.

1 Create the storage folder and upload your FASTQ sequencer run files under the newly created folder. Refer to the image above. For details on the format of the sequencer run files, see the corresponding help pages in Alissa Reporter.

After successful upload, the complete path will look as follows:
<Bucket Name>/<my_storagefolder_name>/<my_sequencer_files>

Or if you entered a prefix (as described in **"Step 3. Configure S3 wizard upload settings in Alissa Reporter and download policy"**):
<Bucket Name>/<Bucket prefix>/<my_storagefolder_name>/<my_sequencer_files>

The Alissa Reporter upload wizard, shown below at the **Upload run data** step, now includes the **select from S3 bucket** link for connecting to the S3 bucket and selecting data files to upload and analyze.
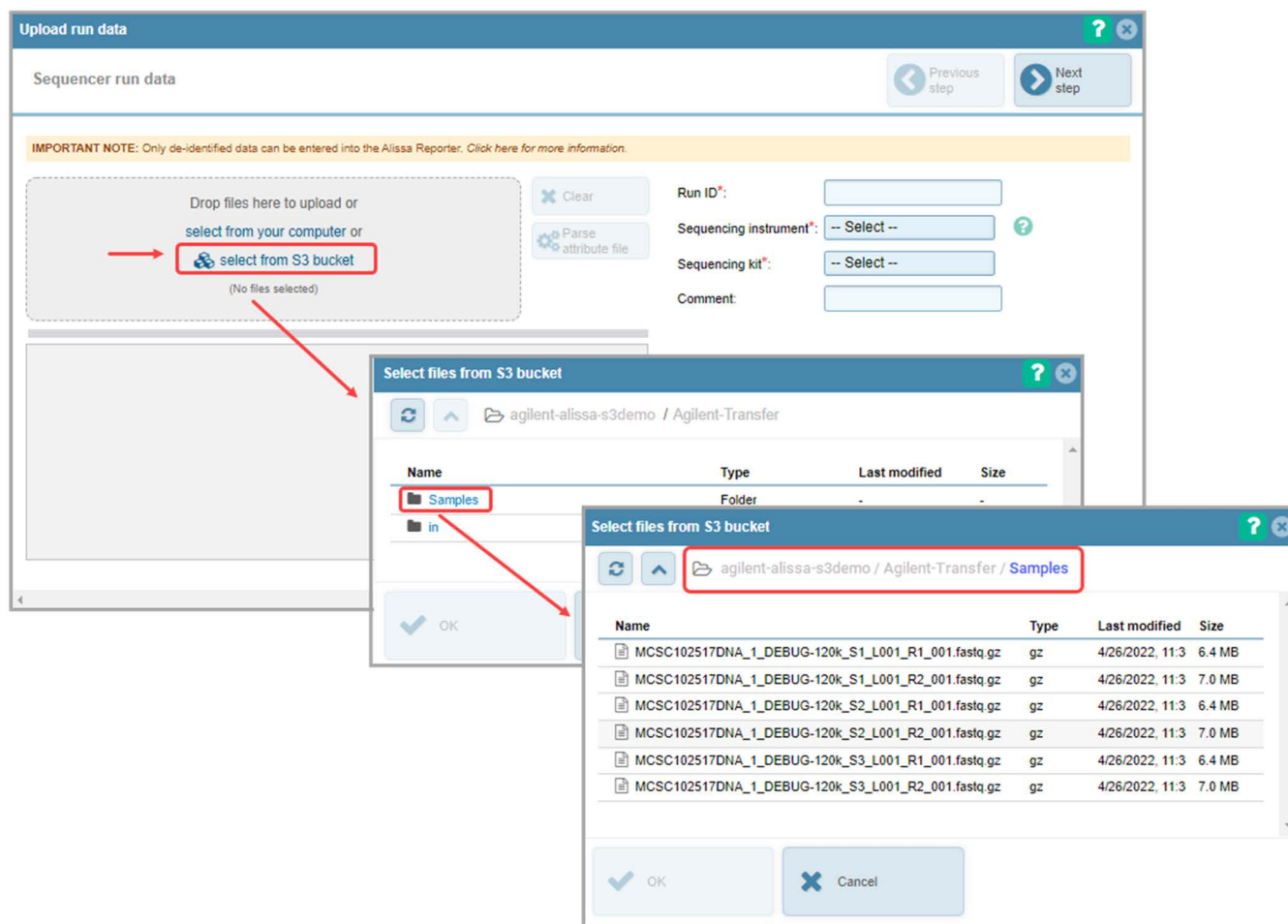
www.agilent.com

Agilent Technical Support

For US and Canada

Call 800-227-9770 (option 3, 5, 4)
Or send an email to **informatics_support@agilent.com**

For all other regions

Agilent's world-wide Sales and Support Center contact details can be obtained at **www.agilent.com/en/contact-us/page**

Revision D0, January 2024

5390-0011

Agilent