

Paste these links (<https://ai-for-life-sciences-1.s3.amazonaws.com/>, <https://ai-for-life-sciences-2.s3.amazonaws.com/>) into any browser to list the contents of the bucket. Eukaryote data can be found in **Bucket 1** while **Bucket 2** contains prokaryote data.

iterating this list each component can be downloaded using your preferred technology, eg, wget https://ai-for-life-sciences-1.s3.amazonaws.com/sra_metadata_5d8517e0-8bb8-4a34-ba02-26c397dea5a6.csv

When you request for instance "<https://ai-for-life-sciences-1.s3.amazonaws.com/>" you will get a XML which includes all paths to all objects - except the csv file. The user would need to create a small python (or whatever language preferred) script looping through the content and downloading each object individually.

The manual way:

Download and install xmlstarlet or xml2.

brew install xmlstarlet

get the full paths without the csv file

curl -s '<https://ai-for-life-sciences-1.s3.amazonaws.com/>' | xml sel -N

w="<http://s3.amazonaws.com/doc/2006-03-01/>" -T -t -m "//w:Key" -o "\${s3url%/}/" -v . -n

output:

/SRR25384227/Lucas0372.1.fq.gz

/SRR25384227/Lucas0372.2.fq.gz

/SRR25384228/Lucas0371.1.fq.gz

...

create a small python script looping though the list items and download each file individually:

starting with the one not in the content list:

wget https://ai-for-life-sciences-1.s3.amazonaws.com/sra_metadata_5d8517e0-8bb8-4a34-ba02-26c397dea5a6.csv

then continue with the list

wget -x -nH <https://ai-for-life-sciences-1.s3.amazonaws.com/SRR25384227/Lucas0372.1.fq.gz>

wget -x -nH <https://ai-for-life-sciences-1.s3.amazonaws.com/SRR25384227/Lucas0372.2.fq.gz>

Please discuss among your teams the most efficient way to proceed with the downloading the data as a team. We advise to download the dataset before the deadline of December 1st 19:00 GMT. We will be turning off the access to the links at that time. Thank you and best of luck with the sequencing data.