

2020-05-08 ICDC-NCBI Data Handling Meeting notes

Date

08 May 2020

Attendees

- [Beyers, Matthew \(NIH/NCI\) \[C\]](#)
- Amit, Adam Stine, Durga, Erika, Phil, Rick Lapoint

Goals

1. Review plan for overall access. [NCBI-ICDC-DCF_agreedplan.pptx](#)
2. Decide on transfer plan
3. Decide who (or team) will write SOP

Discussion items

Time	Item	Who	Notes
	review plan	Matt	reviewed plan. Question: do we want to include BAM indexes and BAMs, or just BAMs? Erika says we should include indexes as well. Adam says we will include BAM index as well (separate from SRA proprietary format). Amit wants to rename files to eliminate prefix and replace _ with -. Adam agrees. BAM indexes allow users to read through BAM files faster. .BAI files. BAM and CRAM files should include generated indexes. BAM indexes would have to be acquired (from submitter) or generated. Do not currently exist in SRA. NCBI will want index files listed in metadata if they exist. NCBI is ok with us just taking AWS format uncompressed data.
	transfer plan	All	NCBI is happy to transfer files - need to grant write access to the ICDC bucket. Amit is imagining an S3 sync process, from within an EC2 instance within NCBI. NCBI agrees with S3 sync. ICDC creates bucket for Glioma study. Create IAM user and give writes to NCBI and then NCBI transfers data. Maybe with file ownership issues, NCI should do transfer. AWS S3 copy should move file.
	SOP	All	Amit, Adam

Action items

- ☐ Adam and Amit to setup meeting to sort out the last details.
- ☐ After data transfer is done, setup meeting to discuss additional studies and how the flow will work.
- ☒ Phil will check with Samir to see if he can find the original index files. File names for indexes will need to match BAMs.
- ☒ Checkin as a group in 2-3 weeks.