dbGaP data submission

This process assumes that you are considered and extramural investigator and are funded by an NIH grant of some kind. This is also outlined in <https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/> .

1. Figure out who to email. Determine which sub organization of the NIH funded your grant, then find them on the document in this link: <https://osp.od.nih.gov/wp-content/uploads/IC_GPAs.pdf> , or a more up to date version linked in the GPA section on this page <https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/#gpagloss>. The genomic program administrators on that document are responsible for working with you to register the study and upload the data.
2. Get the documents you need to start the process, and email them to the person you identified in step 1.
   1. Institutional certificate. Go here: <https://osp.od.nih.gov/scientific-sharing/institutional-certifications/> and download the certificate relative to your study. Fill this out, and provide it to your institution’s signing official with a copy of the protocol. The signing official and the PI both need to sign this. **It looks like OSU’s signing official should be contacted through Buck-IRB’s portal, but I’m not finding that explicitly stated anywhere. The support sites talk about the institutional certificate on the same page as the IRB protocol.**
   2. Study configuration file. The current version of this form to my knowledge is DSI 2019 v3. The GPA may provide you with a different version of this form to fill out.
   3. Proof that the data are part of a paper accepted for publication, if necessary.
3. Probably spend a few months waiting for and responding to emails from the GPA to work out kinks in registering the study.
4. When the study is approved for registration, the GPA will set up an entry on your submission portal with NCBI. You will be invited to this portal via an email from NCBI, and must log in using the link provided. DO NOT DELETE THIS EMAIL as you will need a secondary link provided there to ever log in again (maybe this is RH’s user error, but I only get to the submission page using the links in the email). From the submission portal more members of the project can be invited to work on the study registration.
5. Provide various files related to the study. There are examples that can be downloaded using links within the submission portal for each of these. There are many optional files you can provide, but these are required. The data dictionary(DD) files provide information about the variables encoded in the data set(DS) files.
   1. Subject sample mapping: DD and DS
   2. Subject consent groups: DD and DS
   3. Phenotype information: DD and DS
   4. Sample attributes: DD and DS
   5. Pedigree information of the subjects, if available: DD and DS (not required, but a default field)
   6. Study config: create this a second time through the submission portal. Do before uploading other files.
6. Submit all the above files and wait for everything to process to provide what you need to upload your molecular data to the SRA. You will be given an SRA contact by your dpGaP study administrator.
7. The SRA will provide you a meta data file to fill out about your sample. This requires information similar to what you’d write in a methods section, with the addition of file names and details.
8. Actually upload samples to the SRA.