

Instructions for contributing data to CRCNS.org

Version 0.6 (March 14, 2018)

Background

The mission of CRCNS.org is to make neuroscience data publicly available so it can be widely used for research, teaching and learning purposes. CRCNS stands for “Collaborative Research in Computational Neuroscience.” The name describes one of the goals of the initiative, which is to facilitate collaborative research between experimental neuroscientists, who gather neuroscience data, and computational neuroscientists, who develop theories of brain function. Data at CRCNS.org is also used by researchers in other disciplines including: mathematics, statistics, machine learning, computer science; and also by instructors and students in courses. Since neuroscience experiments are costly to perform, it is important to make sure that gathered data is made available to all who can use it.

Requirements for contributing data

To be considered for acceptance for submission to CRCNS.org, the data or resources must satisfy the following requirements:

- Consists of experimental neuroscience data or resources (such as software or simulations), which could help analyze experimental neuroscience data.
- Contributed data must be high quality and potentially useful to individuals other than those who were involved in the experiments or initial analysis.
- All contributions must be well documented.

To offer to contribute data, submit the form at:

<https://crcns.org/request-to-contribute-data>

Within a few business days after the form is submitted, an administrator at CRCNS.org will get send you an email responding the to the form submission. If the data contribution is approved, follow the instructions given below to prepare the contribution.

General guidelines

Before sharing the data, try to organize and document it so that it will be useful to the wide spectrum of people who could potentially be interested in your data. Think of someone highly interested and reasonably intelligent but unfamiliar with your experiments and data. Please try to make it as easy as possible for that person, by including not only what is necessary, but also material which would be helpful.

What to include

If possible, the following items should be included:

Primary data. If possible, include the data in the original, unprocessed format. For example, if analysis was done to sort spikes, include the original recordings that were used as input to the spike detection and sorting software. If original data are not available, try to include data that is as close to the original data as possible, for example, waveforms of detected spikes along with the local field potential (LFP). A possible exception to this is if the original raw data are extremely large, and the initial processing that is done to reduce the size of the data is a well-solved problem, then it may not be worthwhile to include the primary data. If you are in doubt about whether or not primary data should be included, please ask staff at CRCNS.org.

Derived data. Include data that was derived from the primary data and used to produce any publications from the data. Some examples of this include results of spike sorting, computed properties of neurons, regions of interests in calcium imaging.

Stimuli. If possible, include the files or other information necessary for generating any stimuli that were used.

Example scripts. If the data is not in Matlab or NWB format, include example scripts that show how to read the data using either Python or Matlab (ideally both). Regardless of the format, include at least one “demo” script in either Python or Matlab that demonstrates loading data, performing some type of analysis, and making a plot.

Other Software. If other software tools were used to process the data, links to the software should be provided (if the software is available online) or if the software is not available online, if possible, the software itself should be included.

Summary files. If there are files, such as spreadsheets or Matlab files, or relational database tables that have summary information about the data, include them. Examples are: a table containing metadata about all of the sessions, or a table including

information about units found and their properties. Even if such information is already in individual session files, it can be a big help to have such information in summary form. If data is to be shared using a relational database, SQLite (<https://www.sqlite.org>) is recommended.

Documentation files. If there are any files that document the experiments, data format, software, or anything else related to the experiments, include them. It is not necessary to include copies of materials available online if the link to the materials is persistent (i.e. will always be valid, such is a DOI to a published paper). In that case, including the link to the documentation is sufficient. If there is any doubt about whether or not materials will be persistently available online, include a copy of the files.

Format and organization of data

If you can use a standard format, and in particular the NWB - neurophysiology format (described at <http://nwb.org>), that would be best, but is not required. Proprietary formats (which require proprietary software to read the data) should be avoided unless software to load the data into either Matlab or Python, and which works on all common platforms (Windows, Linux, Mac) can be included with the data set. Whatever data format used must be well documented. (Instructions for documentation are provided in the “Data description Document” section below).

If your data consist of many files and you already have a transparent system to organize them, just use that for the uploaded data. However, if your data files are stored in a scheme that is hard to understand, please try to reorganize the files to make it easier. For example, if there are lots of files, place related files into separate directories and if helpful, use subdirectories to arrange the files hierarchically.

Choose names for files and directories that are somewhat descriptive of the contents, but if possible, reasonably short. Feel free to use intuitive abbreviations. Examples: "exp23" could be the name of a directory containing results for experiment number 23 and "software" might be a directory containing programs used to analyze the data.

Documentation files that apply to the data set as a whole should be placed in a directory named "docs". Documentation files that apply to the content of only one subdirectory, can be placed in that subdirectory, but must be mentioned in the overall documentation file (the “data description document,” described below).

Packaging the files

Before uploading the data set, if there are many files, organize them into directories (as described above) then package them into archive files using either zip or compressed tar archives (tar.gz). tar.gz is the preferred method, and should always be used if the size of the compress file will be above 4 GB, because most versions of zip software will not work with files above 4GB in size. To create a tar.gz file for the contents of a directory named “dirname” using a Linux or Mac OS command line, enter:

```
tar -cfz dirname.tar.gz dirname
```

Ideally, the maximum file size should be from 3 to 20 GB to allow for easier download. If the data files are all much larger than this, then a small sample file should be provided that users can download to get an idea of what the larger files contain, without having to download them.

Data description document

In addition to any other documentation files, there must be included a single document that serves as the starting point for someone to learn about the data set. This document (called the “data description”) must be created in Microsoft Word format because administrators of CRCNS.org will update parts of it before it is made public. (When it is made public, a PDF version will also be made public, so Microsoft Word will not be required to view it).

To make the data description document, download the “crcns_data_description_template” (MS Word Version) which is available at: <https://crcns.org/contribute> and save a copy using the name “crcns_data_description” (remove the “_template” suffix). Then edit the copy by replacing every occurrence of text in double angle brackets, e.g. << Some text ... >> with the information that is describe by the enclosed text, (or remove the section if instructions in the angle brackets say to do that, or if it is not needed). When complete, put the file in the “docs” directory in the data set.

The creation of the data description document will be somewhat time consuming, but worthwhile because it is the main guideline for using the data set. In addition, parts of the data description document will be used to help setup a webpage and create a DOI for the data set.

Uploading files

Note: *Do NOT upload data before approval has been given for the contribution (as a result of submitting the “request-to-contribute-data” form) and until the other instructions given above have been completed.*

Data at CRCNS.org is hosted at the National Energy Research Scientific Computing Center (NERSC, <http://www.nersc.gov>). Any of the following methods can be used to upload data to NERSC:

1) Using FTP. If you have access to a fast broadband Internet connection and the data size is under 300 GB or so, FTP is probably most convenient method. (If you are not sure if uploading via FTP will work for you, try it). To use FTP, fill out the form at:

<https://www.nersc.gov/users/storage-and-file-systems/transferring-data/nersc-ftp-upload-service/>

to create an FTP account. Specify "jteeters" as the account to receive the files. Then upload them via FTP using the created account. When finished uploading the data, send an email to jteeters@berkeley.edu indicating that the files have been uploaded. (Or if you were unable to upload the files via FTP, send an email indicating that to the same address).

2) Via sharing service. If the data is already available online through a cloud sharing service, for example Dropbox or Google Drive, it may be possible to use those services to provide the data to administrators at CRCNS.org, who will then copy the data to NERSC. To inquire about doing this, send email to the administrator at CRCNS.org who responded to your submission of the request-to-contribute-data form.

3) Via shipping a hard drive. If neither of the other two methods will work, you can provide the data by shipping it via hard drive(s). To request doing this, send an email to the administrator at CRCNS.org who responded to your submission of the request-to-contribute-data form.

Complete a webpage describing the data set

After the data and documentation is uploaded, an administrator at CRCNS.org will do the following:

- a. Peruse the uploaded files and the data description document to make sure everything seems complete. Perhaps some changes will be suggested.
- b. Choose a short ID for the data set (e.g. “pvc-4”).

- c. Create a DOI for the data set using the title and list of contributors given in the data description document.
- d. Fill in the data description document with the short ID and DOI. Generate the completed document in both Word and PDF format.
- e. Setup the uploaded files and data description document on NERSC so they can be downloaded.
- f. Create and make public a prototype webpage for the data set at CRCNS.org. When this is ready an administrator at CRCNS.org will ask you to review the prototype webpage and suggest or make any changes.

Follow-up activities

At this point, the data contribution process is complete. However, there may be some follow-up activities required to help support the users of the data:

- If there are question about your data set on the CRCNS.org forum, you will be notified so you can respond to them.
- If there are any updates required for your data set or documentation (for example as a result of feedback from users or the publication of a new paper that should be referenced in the documentation) notify an administrator at CRCNS.org to coordinate making the updates.

Feedback to CRCNS.org

If you have any feedback or suggestions about these instructions and the contribution process or about any aspect of CRCNS.org, please provide them to a CRCNS.org administrator, either by emailing the admin that worked with you on the data contribution, or by sending an email to admin@crcns.org.

Thanks for sharing your data through CRCNS.org!