

# FANC (Female Adult Nerve Cord) Reconstruction Community

## Automated reconstruction proofreading and data ownership guidelines

by Jasper Phelps and Brandon Mark  
with approval from John Tuthill and Wei-Chung Allen Lee  
Written April 2, 2021 / Last updated July 8, 2021

See also:

- [FANC Reconstruction Community – Information Home Page](#)
- [Reconstruction Community Guidelines](#)

**Introduction:** Over the past year, we (Zetta AI and the Lee and Tuthill labs) have worked to produce an automated segmentation of all the neurons in FANC. Zetta AI has built a neuroglancer-based user interface for proofreading the segmentation (very similar to FlyWire), and this interface is now ready for release to you, members of the FANC reconstruction community. Proofreading automated reconstructions is much faster than manually tracing neurons, and with this speed-up, you will be able to reconstruct more neurons and larger circuits. This increases the likelihood that different labs' interests will overlap, necessitating updated policies and more frequent communication between labs. Through the #neuron-ownership-discussion channel, we have asked for your input on developing these policies. This document includes:

1. Some general information about the status of our work on the automated reconstruction project.
2. Rules that mandate open communication about goals and interests, to ensure we identify potential overlaps as early as possible.
3. Guidelines for resolving situations where more than one lab is interested in including some piece of data in a publication.

**Publication policy:** We will prepare a manuscript reporting the automated neuron segmentation, automated synaptic link predictions, and examples of some discoveries enabled by these data. Our goal is to have a preprint of our manuscript posted to bioRxiv by the end of 2021. You may not submit manuscripts that include FANC automated segmentation data to preprint servers or peer-reviewed journals until our preprint has been posted or until Jan 1, 2022, whichever comes first (but see UPDATE JULY 2021 below). You may submit a manuscript to a peer-reviewed journal after this embargo ends, but please attempt to delay the date that a journal would publish your paper until after a journal has published ours, or until after April 1, 2022, whichever comes first. We understand publication timelines can sometimes be tricky and hard to control but we plan to communicate openly about our publication timeline and expect members of the community to do so as well.

If your project benefits significantly from having access to the automated segmentation data before our preprint is posted to bioRxiv, you are encouraged to add Jasper Phelps, Brandon Mark, John Tuthill, and Wei-Chung Allen Lee as authors to your conference abstract or paper in

recognition of the effort it took to make the data available to you and to organize and coordinate the reconstruction community.

UPDATE JULY 2021: We have decided to allow labs in the community to submit manuscripts containing FANC automated reconstruction data to a journal before we post our preprint to bioRxiv, under the condition that the FANC data is not a central focus of the manuscript (to be judged on a case by case basis), and under the condition that the following members of the reconstruction team are added as middle authors to your manuscript: Manuel Castro, Jay Gager, Akhilesh Halageri, Dodam Ih, Nico Kemnitz, Kisuk Lee, Ran Lu, Thomas Macrina, Marwan Tammam, Derrick Brittain, Forrest Collman, Sven Dorkenwald, Casey Schneider-Mizell (in an order to be determined later). (This list is in addition to adding Jasper, Brandon, John, and Wei as described in the section above). Conference abstracts including FANC automated reconstruction data may also be submitted before we post our preprint to bioRxiv, under the condition that you add “Zetta AI LLC” as an author to your abstract in addition to adding Jasper, Brandon, John and Wei.

**Overall proofreading approach:** We expect that most labs will want to first dive into reconstructing circuits of interest in order to answer specific questions relevant to their lab’s work (e.g. the goals listed in the proposal that labs submitted when they joined this community). We encourage and support these kinds of targeted efforts. Along the way, we also encourage everyone to correct obvious segmentation errors in neurons’ backbones that they see as they work. Such corrections often take just a few seconds each, and they will benefit the whole community and the eventual effort to proofread the backbone of every neuron.

After completing their initial proofreading projects, some labs may be interested in contributing to a more comprehensive/unbiased proofreading effort where we collaboratively attempt to correct every error that affects any neuron’s backbone morphology. Such a project would allow us to produce a draft of the whole female VNC connectome and to publish this as a resource, similar to the *Janelia* hemibrain. Labs that think they would be interested in dedicating proofreading time toward this type of centralized and comprehensive effort, which would merit authorship on a publication reporting the draft connectome, should contact Wei. Our decision of when to organize and launch a centralized effort depends strongly on how big of a team is interested in being involved.

**Identifying neurons of interest through exploratory proofreading:** Groups who have already been tracing in CATMAID will likely already know where their neurons of interest are in FANC. For other groups, you will want to look around in the automated segmentation in hopes that you can find your neuron(s) of interest by eye without needing to do any proofreading. However, in order to find your neuron(s) of interest it’s likely that you will need to do some number of merge and split operations to proofread some neuronal backbones. We refer to this as the exploratory proofreading phase. Some of your exploratory proofreading will end up enabling your formal proofreading projects (see next section), but some will not, and instead they may eventually enable projects by other groups. You should not expect that the time invested in exploratory proofreading efforts will lead to authorship on another lab’s paper, even if a lab ends up doing a

project that benefits from the exploratory proofreading you did. Doing exploratory proofreading is simply part of working in a collaborative reconstruction community. Keep in mind that other groups' exploratory proofreading may benefit you in the same way that your exploratory proofreading may benefit them.

**Starting and finishing formal proofreading projects to ensure data ownership:** Once you identify your neuron(s) of interest, you must post on the #proofreading-projects channel of [fanc-reconstruction.slack.com](https://fanc-reconstruction.slack.com) to describe generally why you are interested in this neuron and what kind of data you think you might want to include in a paper. Here are two examples of what we imagine these posts might look like:

1. I am interested in knowing the {presynaptic/postsynaptic} partners of {some type of neuron, for which you share a light microscopy image if you have one}. I have identified this type of neuron in the FANC segmentation: {neuroglancer link}. I am planning to proofread this neuron and its {presynaptic/postsynaptic} partners in order to find its strongest {upstream/downstream} partners. If anyone else has interests that overlap with this, please let me know.
2. I am interested in characterizing the morphologies of all the {lineage X / some other description of a cell type} neurons. I have identified some neurons of this type in the FANC segmentation: {neuroglancer link}. I am planning to proofread the backbones of all neurons of this type to see how many there are and to characterize the different morphological subtypes. If anyone else has interests that overlap with this, please let me know.

If no one else responds to indicate that they have an interest that overlaps with yours, you have *preliminary ownership* of the data that you intend to collect through your proofreading. You may then embark upon this formal proofreading project that your post described with the expectation that you will have priority to publish the data that your project produces.

Once you finish your proofreading project, that is, you have results in hand that you expect to eventually include in a publication, you must announce this on Slack to finalize your *full ownership* of these results. To do this, please reply to your original #proofreading-projects post with a simple message that looks something like this: "I completed the proofreading project described in the post above. We now know {all postsynaptic partners of DNz99 / the morphologies of all the T1 LinX neurons}." Please check the box next to "Also send to #proofreading-projects" before sending your message so that everyone can see your post. In this post you do not have to announce **what** you found, but please reiterate **what kind of data you collected**, i.e. which neuron(s) are involved and whether you've reconstructed presynaptic partners, postsynaptic partners, backbone morphologies, etc.

We discourage individual people from simultaneously undertaking multiple formal proofreading projects. We expect that after you start a proofreading project, you either finish it or announce that you are no longer going to work on it (in which case the *preliminary ownership* expires). If you have a good reason for wanting to be involved in multiple proofreading projects at once, please post about this on #proofreading-projects and we can make decisions on a case-by-case basis. On the other side of the coin, you are certainly free to have multiple people involved in the same proofreading project, but try to make it clear in your Slack messages who the "lead" is, if there is one person overseeing the proofreading project.

Posts on #proofreading-projects will serve as important records of what proofreading projects are ongoing and which ones have been completed. Because they are records, do not use Slack's "Edit message" feature on any posts in #proofreading-projects. (We will keep separate copies of these posts as a backup, but still, this rule will make life easier.) If you need to correct or add to a post, you can do so by posting a reply message to your initial post (instead of using the "Edit message" button).

**"No locked neurons" policy:** At all times, you are free to proofread any neurons in the dataset, regardless of if that neuron is part of another lab's ongoing or completed proofreading project. There are a few reasons for this policy.

1. If we allow labs to "lock" their neurons of interest, where no other labs are allowed to make edits to them, the number of "locked" neurons across the dataset may grow to a point that makes it a huge headache to actually do any work. We strongly wish to avoid this situation.
2. We do not yet have infrastructure in place that would make it easy to know which neurons are part of another lab's proofreading project and which aren't. Therefore it is not logistically feasible to know when you are making an edit that would affect a neuron of interest to another lab.
3. We want the community to get used to the fact that segmentations are always changing. For instance, at some point we may generate a new and improved automated segmentation of the whole dataset, and we want labs to be prepared for that possibility. Specifically, the "no locked neurons" policy encourages that labs design their analyses in ways that are robust to neurons continuing to change in slight ways over time.
4. There is an easy-to-use "time travel" tool that straightforwardly allows you to see the state of the dataset at a past time. This means that labs always have the option to note down a specific time point that they want to use for analysis, and that analysis will then not be affected in any way by other users continuing to improve the reconstruction of that neuron. While this is an option, we discourage this and instead hope that labs will always make use of the most up-to-date data for their analyses.

While our policy is that there are no "forbidden edits" or "locked neurons", if you happen to notice that you have made an edit to a neuron of interest to another lab, sending the relevant person a Slack message about your edit may be appreciated, and so we encourage you to do this. Don't bother doing this for edits that add or remove a tiny fragment, but, for instance, if you merge a reasonably large branch that was missed, sending a message about that edit would likely be appreciated.

### **Rules about use of reconstructions produced by a different lab's proofreading project:**

(Some of the text below is directly inspired by [FlyWire's "Fairness" principle](#).)

Before using unpublished reconstructions in a presentation or scientific publication, users must obtain agreement from every principal investigator whose lab undertook a formal proofreading project that contributed to those reconstructions. Labs may agree to share their work freely, may request acknowledgement or co-authorship, or may request that the work not be shown or used for the time being. Negotiations between principal investigators may be necessary to reach an agreement, but the bottom line is that an agreement must be negotiated. If at all possible, be generous with sharing data and be generous in including others as authors when they share data

with you. Realize that working in a collaborative community means being more liberal with sharing data and including others as authors than you might be used to from other contexts.

To be explicit, *full ownership* of data (which results from completing a proofreading project) does not mean you get absolute unilateral control over the data – however, we expect that no one will make aggressive requests to use data that another lab has full ownership of. If a request comes up while you have only *preliminary ownership* of some data (that is, your proofreading project is in progress), we encourage you to view this as an opportunity to collaborate toward accomplishing a proofreading project more quickly. However, in all cases, if someone makes a request that you object to, you have every right to decline the request and attempt to negotiate an alternative.

If an agreement cannot be reached, a committee composed of Wei-Chung Lee, John Tuthill, and a third member of the community (to be determined) will arbitrate the dispute. (In disputes involving the Lee or Tuthill labs, an additional community member will take their place on the arbitration committee.) The following are three of the main principles that will be taken into consideration during arbitration:

1. The guidelines in the sections above clearly state our expectation that you communicate early and communicate often about the proofreading projects you're working on within the EM dataset. If a conflict arises between two groups and it is clear that more open communication from one group could have helped avoid the conflict, that will weigh heavily on the committee's decision.
2. Work done to determine a neuron's identity (e.g. by comparing with light microscopy data), often done during the exploratory proofreading phase and posted on Slack when a proofreading project is started, is seen as an important contribution and will weigh into the committee's decision.
3. The more time that a lab invested in a proofreading project, the more the committee will be willing to protect the rights of the lab that performed the proofreading. For example, a lab that undertakes a formal proofreading project that requires only a few days of work is unlikely to be tolerated if they insist that other labs cannot use the resulting data in any way until it is published. However, a lab that undertakes a proofreading effort that takes six months of work may be seen as justified in asking another lab to delay their use of the data for a significant period of time, potentially until the first lab gets a paper accepted.

We view arbitration as a last resort. We hope to never need it, but we understand it may be the most logical way to resolve some situations. In the unlikely case that a single lab is repeatedly involved in disputes that require arbitration, we reserve the right to revoke access to the automated segmentation for specific labs.

**Synapses:** Automatically predicted synaptic links for the whole VNC will be made available soon. We do not currently have plans to support proofreading of the automated synapse predictions within neuroglancer. See [here](#) for an argument about why manually proofreading synapses one-by-one is not believed to be a good use of time (even if we had the infrastructure for performing synapse proofreading, which we currently don't). If necessary, synapses can be manually proofread after importing into CATMAID (see below).

**What role will CATMAID play?:**

1. For projects already well underway in CATMAID, there are two main options:
  - a. You are welcome to complete your projects by continuing to work in CATMAID. It is possible to export automatically segmented neurons as skeletons and import those into CATMAID to manually join them with your existing tracing, but we have not done this yet ourselves. (Greg Jefferis's group has done this some in FAFB and may be able to provide code and advice if you're interested in doing this.) Note that if you use the automated segmentation in this way, you will then be subject to the publication embargo, whereas if you just complete your project in CATMAID (while looking at the automated segmentation to guide your tracing efforts if you find that useful), you will not be subject to it.
  - b. An alternative plan would be to start a segmentation proofreading project targeting the neurons you have been tracing in CATMAID. Groups wishing to do this will certainly get priority on undertaking such projects. (That is, if you've been spending lots of effort tracing some neurons in CATMAID, we will not tolerate other groups attempting to work on the automated segmentation version of your neurons before you're able to complete your project.)
2. We highly recommend that groups not start new manual tracing efforts using CATMAID.
3. While it is possible to export neurons from the automated segmentation as skeletons and then import them into CATMAID where various skeletons can then be connected together and extended via manual tracing, we strongly discourage this approach, except for the situation where projects are already well underway in CATMAID (detailed above in 1a). We believe everyone will be better off if we agree to work collaboratively on proofreading the automated reconstructions, instead of importing fragments into CATMAID and doing merges and cuts there.
4. Once a neuron has been proofread essentially to completion in neuroglancer and is nearing ready for publication, we can then help import it into CATMAID where final edits can be made (e.g. correcting a few synapses or twigs). We require that you make these skeletonized versions of proofread neurons available on Virtual Fly Brain as part of your publication, as has been standard for papers that include skeleton reconstructions from FAFB. We can help you with this process.

### **Amendments to this document**

We expect to add new sections or make modifications to this document as interests, infrastructure, and technology related to FANC reconstruction evolve. We invite the community's input about when additions and/or modifications to these policies are needed. Please feel free to use the #neuron-ownership-discussion channel for ideas that you wish to get community feedback on, or email Wei directly. Ultimately, we reserve the right to make final decisions about proofreading and data ownership guidelines, but our main goal is to enable scientific progress by keeping the community harmonious, productive, and free of conflict.