Turning neuroscience data into insight in the 21st century

The modern research and industrial approach to neuroscience involves an increasingly vast assortment of data collection methods and data handling approaches. While data collection and accumulation are expanding, approaches to data management and insight gathering are still lacking.

Readers of this white paper will learn how three major phases of activity: 1) Visualize, 2) Collaborate, and 3) Build, when done correctly, can massively expand an organization's ability to extract value from neuroscience data. Through examples presented through interface images and diagrams, the reader will arrive at a clear sense of the modern state of the technological landscape available for this challenge.

Authored by Stephen D. Larson, Ph.D., CEO MetaCell

KEY TAKEAWAYS

- Visualization of neuroscience data enables multi-modal data to be recombined and taken apart in novel ways
- Collaboration built around neuroscience data is critical for unifying interdisciplinary teams around data
- An effective, consciously-designed process for building digital technologies around data maximizes insight value and speed

INTRODUCTION

"How can a three-pound mass of jelly that you can hold in your palm imagine angels, contemplate the meaning of infinity, and even question its own place in the cosmos?" -- V.S. Ramachandran

Neuroscience is a field of study that confronts some of the deepest mysteries that human beings have ever grappled with. What is the nature of our consciousness? What goes wrong when mental disease strikes? How do we heal injuries of the brain? While scientists have begun to answer these questions, many answers are still far from our grasp.

From new molecular probes, to magnetic resonance imaging, to high-powered light and electron microscopes, to electrical recording, today, neuroscience is a discipline awash in data. While the field continues to devise increasingly powerful imaging and recording techniques, generating insight from these data now also requires pushing the limits of information science, mathematics, visualization, interaction, and software engineering. Why is this so complicated?

Because the brain itself is so intricate in structure and dynamics, no single method of data collection captures all of its functional aspects. In addition, to gather data from the brain, non-overlapping scientific disciplines such as physics, chemistry, electrical engineering, and genetics are required. As a result, experimental data is often collected from a single dimension of measurement. This means that as soon as the data are collected, they are already in need of being interpreted in the context of some other data. For example, scalp EEG data is collected from multiple electrodes, often with the need to have the exact positions of those electrodes on the head. Analogously, an image of brain tissue taken from a fiber-optic microscope often requires knowing the 3D location in the brain where the image was taken. In these cases and in many others, the context provided by a second or third modality of data is very often critical to deriving the full benefit from the first.

As a result of the fragmented nature of neuroscience data, an end user will often want to visualize and interact with those data in special ways. In addition, users will want to collaborate with others around their data. Tools that are used to process such data can be arcane and require special knowledge to work with. Moreover, when data sets are small, they can be transported more easily, but as their size grows, this can be more difficult. Alternatively, a research group that produces a large volume of data, even if small, may quickly find themselves losing track of what data they have generated without basic systems for organizing their results that is archivable and easy to work with.

The better able researchers are to work with neuroscience data, recombine them with other data, share them, and drill into them, the faster our progress will be toward understanding the brain. There is so much still to learn from the brain; unlocking its full breadth of mysteries awaits those who can bring the greatest skill set to bear on working with its data.

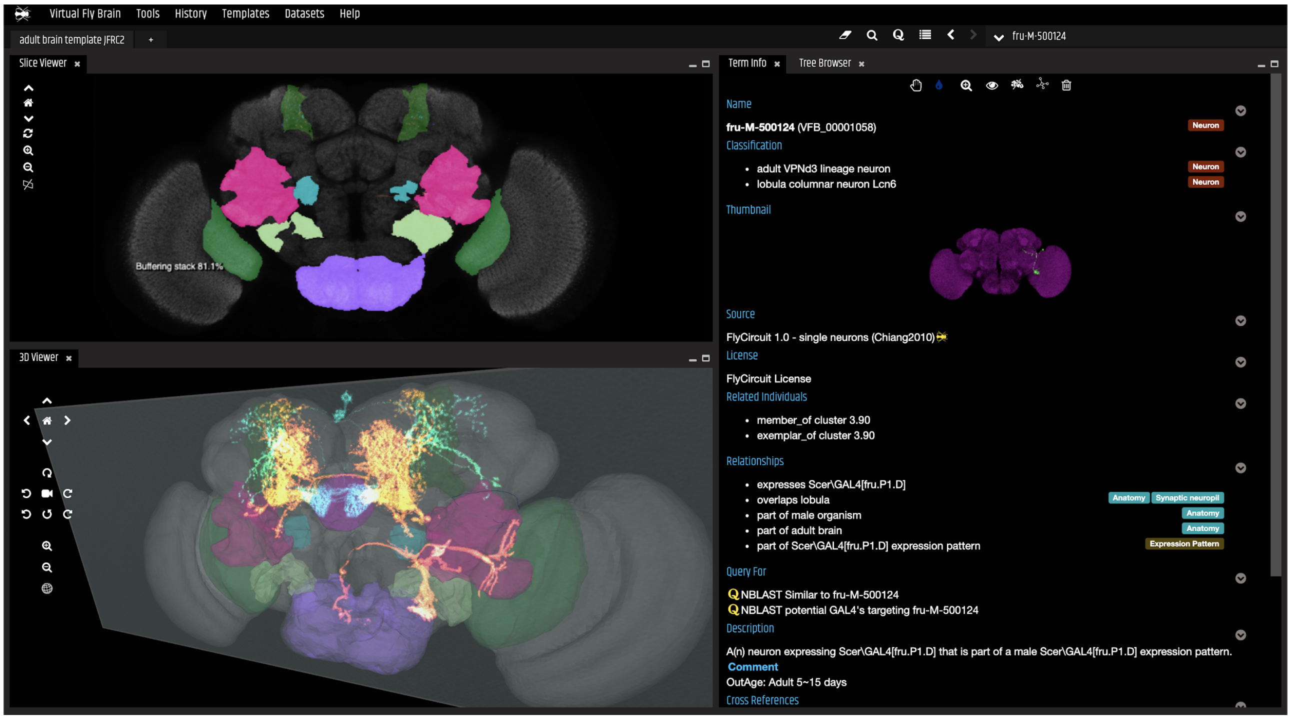

FIGURE 1 - A complex neuroscience scene with many entities, including expression patterns and segmented neurons. The 2D stack viewer is synchronized with the scene's 3D canvas that has the same colors. The 2D side panel that floats on top of the 3D view shows the 2D versions of the data displayed beneath it. A grey plane that crosscuts the whole brain surface models indicates the reference plane from which the 2D datasets are drawn. Two examples of images are shown. A purple and green gene expression image corresponding to the reference plane is shown, where the green highlighted area in 2D corresponds to the pink highlighted neuron structure in the 3D view. Below this, a blue, green, orange and gray image shows the slice from the whole brain data set that corresponds to the location of the reference plane. The highlighted blue, green, and orange areas also appear with the same colors in the 3D view. Both of these images are much higher resolution than shown in the side panel, and can be expanded further in order to examine the high-resolution version in greater detail. Additional information appears in the side panel in the form of text. Specifically, identifiers for individual data are shown in order to be able to track down the unique provenance of that dataset. English descriptions of the data being shown further help to specify what the user is looking at. Type information allows the user to search for other data containing structural features of the same type. Finally, there are links to the original sources of the data, allowing for attribution to be properly assigned.

The keys to using neuroscience data to their fullest potential involve three approaches. First of all, carefully designed data visualizations, which capture the key dimensions of the data, should allow users to intuitively explore and "reach into" the data in ways that are appropriate to that data, rather than a “one size fits all” approach. Secondly, collaborations built around key data that leverage cloud infrastructures and reuse open source code enable more eyes to look at the data, reducing the burden of dealing with vast and complex datasets through sharing. Lastly, an efficient process for building experiences around data that leverages strong cross-disciplinary teams is critical for ensuring that data experiences are robust. In the rest of this article, we will explore these approaches -- visualize, collaborate, and build -- in more detail to better understand how they are the foundations of progress in neuroscience data in the 21st century.

VISUALIZE

In the nervous system, function is deeply tied to structure. Imaging tools allow scientists to probe nervous system structure in a comprehensive manner. Each modality of imaging reveals some dimensions of nervous system structure while excluding others. By reassembling the modalities synthetically inside digital visualization environments, it is possible to provide integrated views of nervous system structure that are otherwise impossible to see through direct imaging. These integrated views allow a user to better understand the relationships between structural features visible in different modalities. This allows for insight that is not possible with one or the other modality alone.

This approach to creating synthetic digital experiences around neuroscience data can be used to great effect. In Figure 1, above, five different types of neuroscience data are brought together on a single screen, drawing from a rich library of hundreds of individual data items. This environment combines gene expression patterns expressed as 3D image stacks targeted to neurons, segmented neurons expressed as 3D wireframes, a whole brain slice volume that has been segmented by its major anatomical features, whole brain surface models, and textual descriptions associated with the various data elements.

These data types are displayed in a 3D synthetic view as well as separately in a 2D side panel floating on top of the synthetic view (Fig. 1.). In the 3D view, an integrated representation of the data can be rotated and zoomed to view the different data in a unified spatial context. Making up the main 3D structure of the visualization are 3D models of the surfaces of the anatomical brain regions identified in the fruit fly. These whole brain surface models are represented as wireframes, inside which other data can be scaled and fit together. Appropriately-scaled gene expression data are visualized within the spatial context of the whole brain wireframes. Portions of the gene expression data that do not have signal are made transparent, revealing the structure of single neurons. In addition, some neurons have been segmented, enabling a wireframe structure to be created that describes the neuron in greater detail. Such segmented neuron models can also appear in the spatial context of the whole brain.

With the structural data assembled in this way, we can go one step further and overlay functional data. Figure 2 shows the same data set from Figure 1, but with the camera pulled in closer to examine a single neuron wireframe. In this example, experimental traces recorded from the neuron via the patch clamp method are called up and superimposed as a line trace. This neuron has been reconstructed as a conductance-based model. In the second plot to the right of the experimental data is the result of a computer simulation of activity through the conductance-based model.

What kinds of questions can be answered with data assembled in this way? Some example investigations that can be made possible from the software in Figures 1 and 2 include:

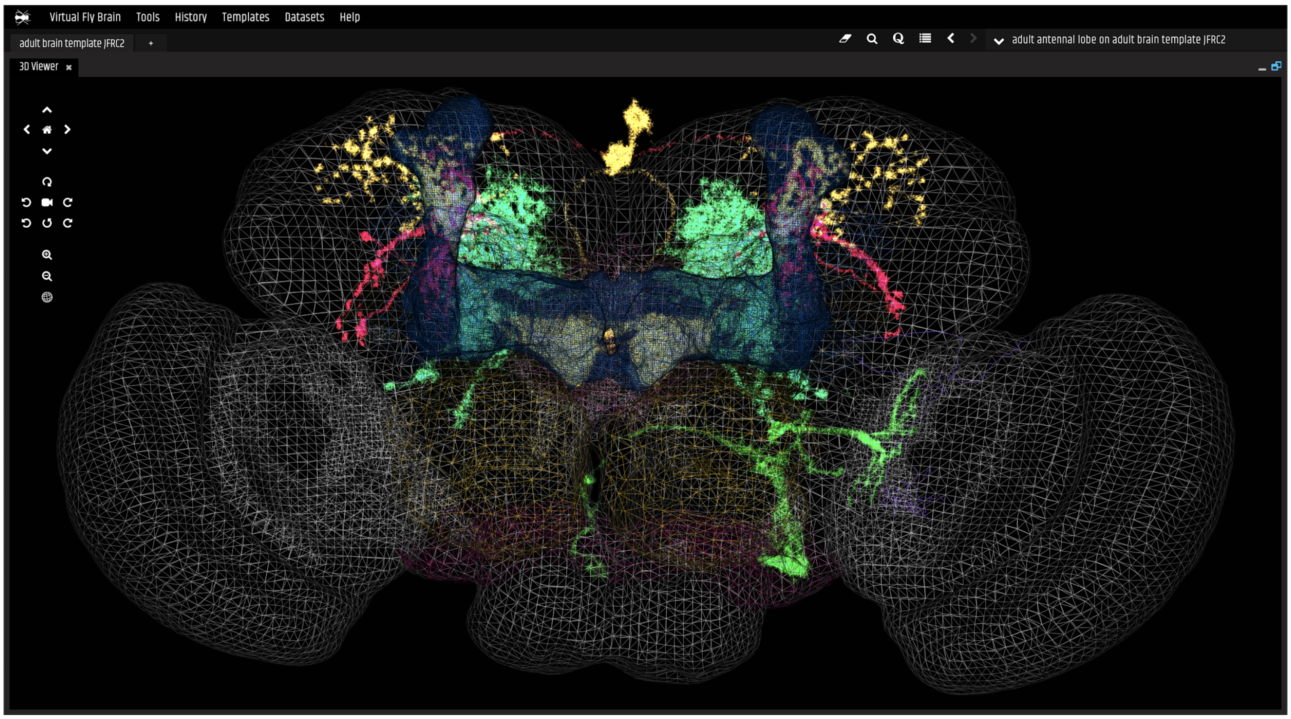

FIGURE 2. Visualization of non-skeletonized data as a volumetric point cloud in Geppetto. The data is superimposed on the surface data. All neuroscience data, including these expression patterns, can be easily added by the user in real time on their browser.

- Search for a single neuron of interest and see all data related to that neuron. Data can be related via different dimensions such as:
    - Same type in different locations
    - Spatial proximity
    - Gene expression
    - Found in same anatomical region
- Build hypotheses about the interactions between two target brain regions by displaying all neurons that are found to have an arbor contained within both of them

- Build detailed single-cell neuron models from gene expression data while keeping both in the anatomical context of the whole brain's coordinate system

These examples show a very powerful use case for investigating neuronal structure and function in a unified data visualization. Once visualizations allow deep navigation of data sets, additional interactivity can be added. Through the use of state-of-the-art user interface and user experience techniques, effortless exploration of extremely complex and embedded data can be achieved.

Furthermore, this visualization is not locked to a single high-powered workstation but is made available through web-based technologies, well suited for collaboration. In the next section, we will further examine the implications of collaboration on detailed data sets like this.

COLLABORATE

Neuroscience has always been a highly interdisciplinary area of scientific and R&D investigation. The skill sets that are required to conduct research in this field are extremely difficult to find in a single individual. Consequently, effective collaboration between individuals of complementary skill sets is essential. As both academic and industry organizations acquire and groom the best human resources, the need for those teams to efficiently share complex scientific data becomes critical.

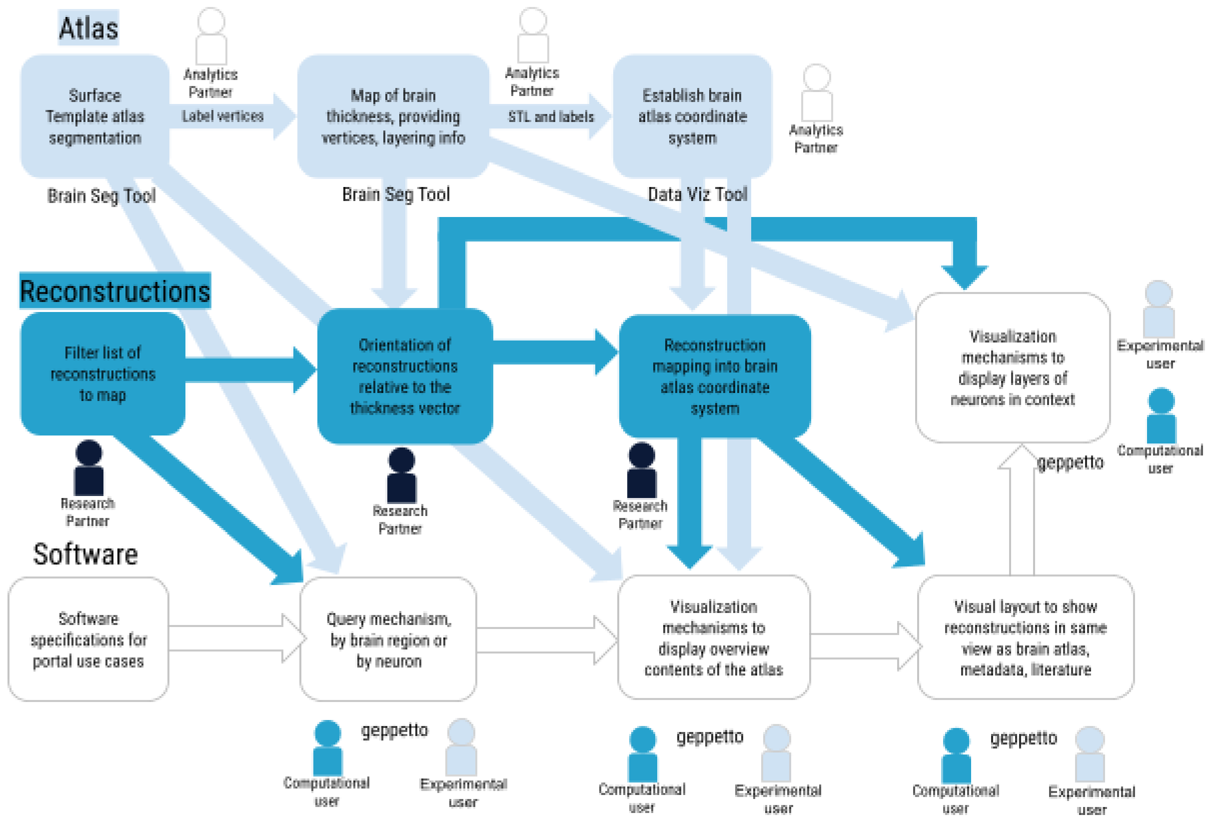

FIGURE 3 - Collaboration occurring between individuals with different roles around software with various data transformation sets and user interface elements. An example of a data and process flow diagram highlighting aspects from a brain atlas, a series of reconstructions, and the software features that enable the data to be visualized and interacted with to a high degree. Personnel appearing near process elements shows members of a collaboration who have access, respectively. Individuals in this case must be carefully coordinated around specific software processes that allow the data to be assembled correctly. The software interfaces (white boxes) that enable collaboration require an atlas (light blue boxes) and various reconstructions (dark blue boxes) to have been appropriately combined, requiring the collaboration of research partners and analytics partners.

As mentioned earlier, in the past, scientific and R&D investigation could proceed easily with one or two different kinds of data collection. In these "single-modality" areas, data could be handled with more conventional methods. In the 21st-century, however, modern computing has brought us firmly into a multi-modal world of research. No discipline exemplifies this better than neuroscience.

Thus, interdisciplinary methods of data collection have produced data that require broad skill sets to interpret. For example, Magnetic Resonance Imaging (MRI) is a device that requires both an understanding of electromagnetism (EM) and cellular neuroscience to fully interpret the data it produces. The combination of Functional Magnetic Resonance Imaging (fMRI) with conventional anatomical MRI takes advantage of different fundamental principles of EM to generate very different images of the same brain. A third data set and modality can be added if we include post-mortem cellular-level high-resolution images.

Even when an MRI, fMRI, and post-mortem slice images are taken from the same subject, making maximum use of these data is complicated. Additional data processing steps such as warping, registration, and down-sampling are required to create apples-to-apples comparisons of such images. As shown in an example in Figure 3, this may require that analytics partners and research partners play a role in transforming data to serve them up to users via appropriate user interfaces. The right digital collaboration environment can enable these data processing steps and make the difference between successful and unsuccessful harnessing of these data sets.

The more people that can work with a high-value data set, the more potential a team has to derive value from it. For example, if an experimentalist and a computational user are both able to work with the same data, a more rapid R&D cycle can occur. Lowering the barrier to entry for all team members can be enabled by well-chosen user interfaces. When done effectively, good user interfaces can bridge the context that is needed for team members from different backgrounds to quickly interpret data. Techniques such as rapid context sharing, linking out to background materials, and intuitive interface with online help can rapidly enable interdisciplinary teams to be “on the same page” while working with complex interdisciplinary data. Otherwise, because ideas in these disciplines are often quite deep and require significant background knowledge to interpret, the data may mean very different things to different team members, and effective idea generation and communication will be limited.

We have seen that diverse teams are needed in neuroscience and that multi-modal data are often required to ask hard questions. Eventually, all team members will want to use data that they are not experts in. In average teams, many team members will not be able to work with these data. However, in the best teams, organizations will build collaboration environments around data that lower the barrier to entry for all team members to understand all data.

BUILD

Now that we know the value of visualization in complex data fields and we understand some of the collaboration environments that can maximize the value of the data, how do we ensure that these collaboration environments get constructed? Without the right convergence of skills in the team that is assembling the software, it will be impossible to have the software reach a level of quality where it can achieve these ambitious goals.

A team with a diverse skill set is needed to build software to fully take advantage of opportunities in modern neuroscience data (Fig. 4).

Firstly, tight process control and use of standards is critical. Even very talented programmers need a process put in place that ensures that they are building what the customer needs. A formal requirements-gathering phase and use case definition period is critical to create shared expectations between the software team and the users. An “agile” development process that involves continual client check-ins throughout the process of software construction is similarly critical.

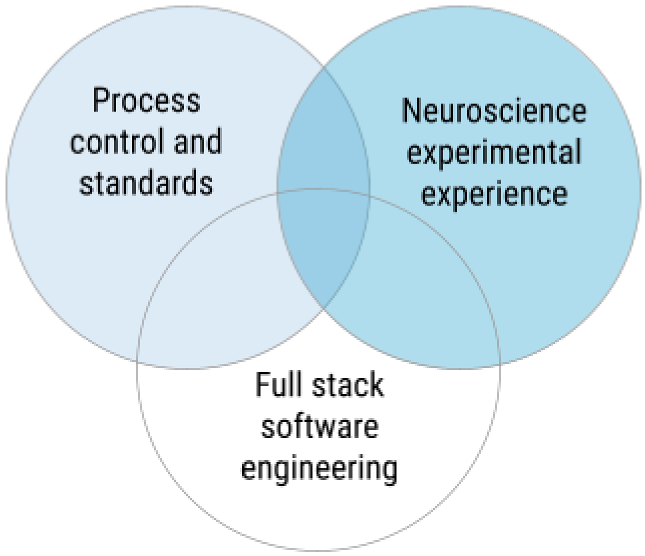

FIGURE 4 - In order to build the most effective software for interdisciplinary teams to collaborate, three key elements must converge: tight process control, neuroscience experimental experience, and full stack software engineering.

As part of that process, open source and academic community aspects are valuable. Because active open source communities continually evolve the code bases they work on, reusing open source libraries combats the natural progression of ever-increasing software releases breaking software systems. Moreover, keeping control of the underlying source code enables continual iteration and customization. Inevitably, requirements will change and expand as the needs of research and development evolve.

Tie-ins with academic standards initiatives ensures that new developments in data tools are incorporated into the software. For example, ontologies have continually found a role in neuroscience, and therefore academic standards such as OWL, less known in industrial circles, are valuable to keep track of and have expertise in.

Secondly, a team capable of full stack software engineering is critical. Full stack software engineering refers to the ability to create applications that have components both on the server side and the client side. This means taking advantage of modern web services, i.e., the creation of lightweight mobile and web apps. One of the key software best practices needed for applications such as the ones we are describing is modularity, the ability to structure software programs into components with manageable complexity. This ensures that as the software evolves, expanding its functionality is a manageable, maintainable process.

Lastly, team members cross-trained with neuroscience experimental experience are critical---and difficult to find. The complexity of experiments and details of experimental procedure mean that the interpretation of data is difficult without personal expertise. For example, time varying data requires using skills from computer science that relate to data streams, whereas image-based data requires using matrix-based computer science. Teams that appreciate the underlying data science of neuroscience data will be much more adept at building robust software applications that make maximal use of data.

CONCLUSIONS

In summary, we have explored what it takes to create value from data in the 21st century in the area of neuroscience. We have seen that visualization is a powerful way to deal with multi-modal data, which is becoming the keystone of modern neuroscience R&D. Furthermore, the creation of cross-disciplinary teams to evaluate and work with that data requires building collaborations that are enabled by powerful and context-heavy user interfaces. Lastly, we've seen that, in order to create such applications, key skill sets must converge: neuroscience experience, full-stack software engineering, and process control.

Organizations that learn these lessons of neuroscience data will end up with powerful, enabling software technologies that will lead to the creation of high-value research.

REFERENCES

Larson, Semantic and Spatial Multi-Scale Information Models of the Nervous System, Ph.D Dissertation, UCSD, 2012.
Ghosh et al., Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. Nature, 2011.
Akil, Martone, Van Essen, Challenges and Opportunities in Mining Neuroscience Data. Science, 2011.
Virtual Fly Brain http://virtualflybrain.org
FlyLight: Jenett et al., 2012, Cell Rep. 2(4): 991--1001 http://dx.doi.org/10.1016/j.celrep.2012.09.011
FlyCircuit: http://dx.doi.org/10.1016/j.cub.2010.11.056
Images from FlyCircuit were obtained from the NCHC (National Center for High-performance Computing) and NTHU (National Tsing Hua University), Hsinchu, Taiwan.
BrainName paper (Ito et al., 2014) http://dx.doi.org/10.1016/j.neuron.2013.12.017

ABOUT METACELL

MetaCell is a software company that puts neuroscience data online, creating compelling visual and collaborative experiences. Our software applications unlock the true value of neuroscience data and models including microscopy and MRI images, EEG and electrophysiology data, computer simulations, and much more.

We work at the intersection of neuroscience and software engineering, presenting a "best of both worlds" translational approach between the two. We see great opportunity to take best-in-breed information technology and apply it to the world's most challenging neuroscience questions, from mechanisms of complex diseases that still elude us to the mysteries of the brain.