Open, Industry-Standard File Format for Neurophysiological Data

Initial Organizational Meeting on November 3, 2000

Time and Location

November 3, 2000, 2:00pm to 6:00pm, in Dauphine Room I,
Hampton Inn & Suites (next to New Orleans Convention Center)
1201 Convention Center Boulevard, New Orleans, LA 70130
(504) 566-9990 or (800) 292-0653

Purpose and Background

The purpose of this meeting is to start the process of developing an industry-standard file format for neurophysiological experiment data. The meeting is open to all companies and academic groups that develop data acquisition hardware and analysis software for neuroscience research. Each participating group is asked to limit their attendance to two delegates. This meeting is being organized by Bionic Technologies, Inc. and meeting facilities are being sponsored by the NIH NINDS under the guidance of Bill Heetderks.

Meeting Agenda

The meeting will begin with introductions and sign-in for all participants. The remainder of the meeting will be conducted in a discussion format (led by Dr. Brian Hatt, President of Bionic Technologies, Inc.). It will focus on answering the following questions:

1. **Is the concept of the jointly developed file format for neurophysiological data feasible?**
   - 1.1. What is the level of interest and energy among the attending companies?
   - 1.2. What are the possible problems and conflicts that are likely to arise?
   - 1.3. Will people use the format once it's completed?
2. **How should the development work be organized and coordinated?**
   - 2.1. Should we conduct the remaining development through a small, representative working group and web-based review forum open to the public?
   - 2.2. Can a consensus be reached on who should be in the working group?
   - 2.3. Are there any groups that should be involved with this effort that are not present?
3. **What should be the basic design and scope for the proposed file format?**
   - 3.1. How will the format be used in the data collection and analysis process?
   - 3.2. What types of data should be included in the format?
   - 3.3. Which data modalities should be in separate files and which should be combined?
   - 3.4. How can the format be structured to accommodate different processor platforms?
   - 3.5. What types of data should be included in headers and how can it be standardized?
   - 3.6. How can the format be made flexible for inclusion of future data types?

Contact Information

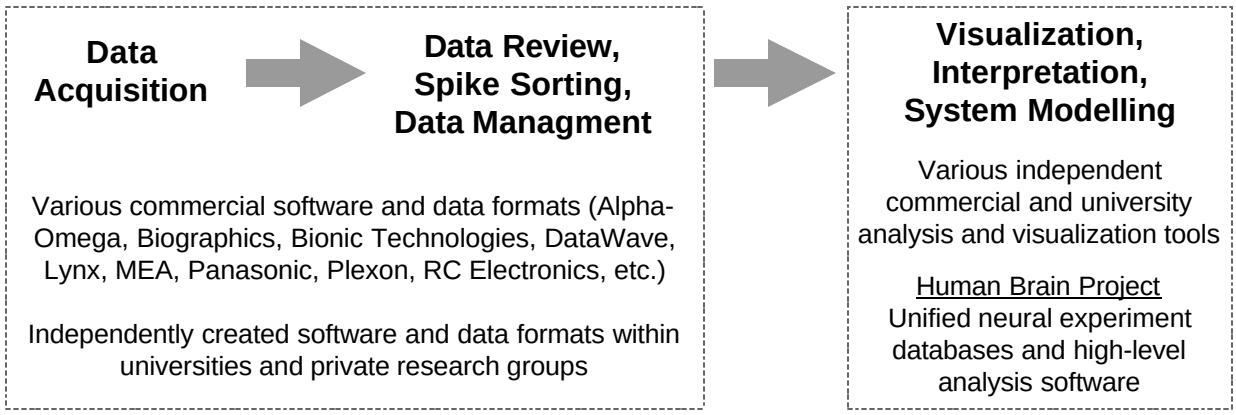
If you have questions or suggestions for the meeting, please contact Brian Hatt (brian@bionictech.com) or Shane Guillory (shane@bionictech.com) at (801) 582-5533.

Motivation

There are many high-quality data acquisition systems available for neuroscientists interested in multi-channel neural recording. However, all of the commercial systems presently available use proprietary, independently-developed file formats for the acquired data. This is natural for a young industry, but it creates several practical problems for neuroscientists. For example:

- Very little interoperability is available at the acquired data level for users wanting to use instruments, data processing, and spike classification tools from different vendors.
- There are considerable difficulties for labs wishing to directly share data when they are using different hardware, software, and/or experimental arrangements.
- Shared software tools for low-level data processing and spike classification written by university labs are usually written to support a limited set of acquired data file formats. Moreover, these groups often create their own intermediate file format if they have a data type that is not supported by existing formats.
- Neuroscientists wishing to develop their own custom data acquisition hardware must choose between data file formats supported by desired tool sets or create their own data format if they have unsupported data types or experimental variables.

Most all data acquisition equipment vendors provide access to higher level analysis environments such as Stranger, NeuroExplorer (NEX), or MatLab. Moreover, there are many development efforts for higher-level visualization and database-ready representations of neural data being catalyzed by programs such as the NIMH-sponsored Human Brain Project. But there is a clear need for data portability at the acquisition and review level:



We believe that the solution to these limitations is to create an open, industry-standard file format for neurophysiological data and a web-based forum to support this standard with documentation, software utilities, and source code. In addition, the development program should include NIH guidance to insure that the development process for the standard is open, fair and inclusive. The neuroscience equipment industry is now mature enough to support such a standard, but still small enough for standardization to be useful and widely implemented.

A good example for the impact of low-level file formats is the JPEG format for images. This standard has encouraged the development of many public domain utilities as well as high-quality commercial software packages for image processing. We have identified a few past efforts to create standardized neural data formats such as the SPike Interchange File Format (SPIFF) group at the University of California San Francisco. These efforts have produced useful internal formats for small groups of labs, but they have not incorporated wide enough industry and user feedback to make them universal enough for general neuroscience work.

The file format specifications would be made freely available and implementations of this standard would be completely voluntary and determined by individual companies. Participation in the format development process should imply no commitment to use the format. Initially, companies that support the standard are likely to do so with small translation programs for their proprietary formats. However, the format should be designed with enough efficiency to allow its direct incorporation into future instrument and software products.

Proposed Scope

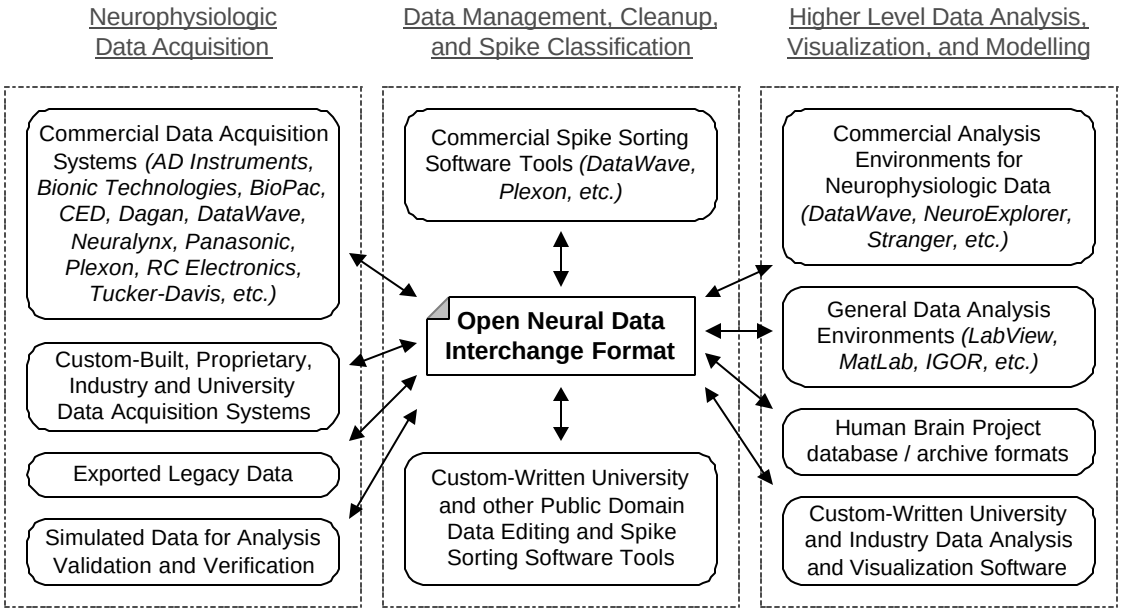
Although the data file formats from neuroscience equipment companies are different, there are only a few general types of information stored during neurophysiological experiments. These general types include:

- ***Headers, labels, and annotations*** for experiment conditions, electrode and equipment configuration, user notes and comments, and data file editing markers.
- ***Continuous high sampling-rate signals*** including extracellular and intracellular recordings and experimental information such as auditory signals or fast stimuli.
- ***Continuous low sampling-rate signals*** including Local Field Potentials, EEG, ECG, EOG, EMG, and experiment information such as force, pressure, flow, positional, orientation, eye-tracking, temperature, pH, oxygen level, and chemical concentration.
- ***Sampled waveform segments*** for extracellular spike recordings from individual or groups of electrodes and trial-delineated recording traces.
- ***Discrete experiment event markers and timing markers for synchronization*** of the recorded data with other time series data files such as audio, video, and MRI images.

It is important to limit the scope of the file format to the level of data acquisition and manipulation so that the format will be compact and practical. There is a common temptation to make new file formats into universal file formats that are all-encompassing and overly complex. As a result, the formats become difficult to use in real-time systems. The shared industry knowledge and experience of companies in this area will provide a pressure to produce a complete, yet pragmatic, data format. Moreover, this industry-wide knowledge and experience base will encourage a more generic format focused on neurophysiological data instead of equipment biases from individual companies.

Presently, there are higher-level, complex data formats being developed under the NIMH-sponsored Human Brain Project for distributed analysis and database storage of neural data. These include the Common Data Format being developed by the Cortical Neuron Net Database group at Cornell University, and the NeuroCore Time Series DataBase development at the University of Southern California. In addition, the NSF is sponsoring the development of an

abstracted Time Series Data Format by the NTSA Workbench group at the University of Illinois at Urbana-Champaign. There are also several generic abstract data exchange models and methodologies for multi-platform network sharing that have been developed by companies such as Sun Microsystems and Microsoft. These high-level formats are extremely powerful for portability and processing flexibility, but are largely impractical for implementation with real-time data acquisition hardware and data editing tools. The proposed format should provide standardization at the lower data-management level and provide direct data sharing capabilities that would complement ongoing higher-level standardization and sharing methods.



The format should also provide a common portal to higher level analysis tools and environments. Commercial neurophysiological analysis programs such as Stranger (from BioGraphics) and NeuroExplorer (NEX), can import data from an impressive, but limited, number of neural data file formats. The new file format should be complete enough for representation of legacy formats with little or no data loss. As such it would be capable of serving as an intermediate translation layer between other formats with low support and higher-level analysis tools.

In order to distribute the file format specification documents, we propose the creation of a vendor-neutral, web-based forum. This forum would also include areas for revision suggestions, software applications, source-code, and links to users and suppliers supporting the standard. We would like for the NIH will have ultimate control of this web site and would like maintenance of the format to be administered by the NIH and a small industry/user consortium. The completed file format should also be promulgated by possible journal publications and workshop sessions at conferences such as the Society for Neuroscience annual meeting.

Proposed Development Plan

The work for specifying the file format will be conducted by two groups. The first will be a general user/industry advisory group that will meet once at the beginning of the program to determine overall characteristics, experimental data types, and usage scope for the file format. The second will be a smaller working group selected at the advisory group meeting. This

smaller working group will be responsible for the initial detail-level design decisions and the actual writing of the draft specification. Once completed, the draft will undergo revisions via an open web page forum for peer review and comment. Revisions will be managed by the working group, and the group may meet after the revision period to finalize the specification. Program coordination, word processing, and facilities support will be provided by Bionic Technologies as needed throughout this program.

Tasks
1 – Solicitation of users and companies willing to participate
2 – Registration and Deployment of the Web Site
3 – General user/industry advisory group meeting
4 – Initial draft specification by working group
5 – Iterative draft revision process
6 – Compilation of final draft by working group
7 – NIH debriefing and final document review before release
8 – Creation and submission of C/C++ import/export library
9 – Web page finalization and final document release

Proposed Deliverables for the Development Effort

- A file specification document (on the order of 100 pages) detailing a binary data format and summaries of the design philosophy and decisions made for its components.
- A vendor-neutral web site to support and freely distribute the format specification with areas for documents, suggestions, source code, software applications, and links to NIH and neuroscience research groups, and links to industry supporters of the format.
- A C/C++ code library for importing and exporting data in the open format.
- Promotion of the format in the form of web links on neuroscience related pages, journal publications and letters, and a small workshop at the Society for Neuroscience meeting.

Note: The software library and all shared software created in this development effort will be made freely available under the open source model and published on the format web site.