Gardner D, Shepherd GM. : A gateway to the future of neuroinformatics. *Neuroinformatics*. 2004;2(3):271-274.

Development of a Database of Databases: the Neuroscience Database Gateway (NDG)

Thomas Morse[1], Luis Marenco[1,2], Daniel Gardner[3], Gordon Shepherd[1]

[1]Department of Neurobiology, [2]Center for Medical Informatics,
Yale University School of Medicine, New Haven, CT
[3]Department of Physiology and Biophysics, Weill Medical College of Cornell University, NY

Tom.Morse@Yale.edu

This talk is based on an article that recently appeared in Neuroinformatics (Gardner and Shepherd 2004). The Society for Neuroscience recently announced a Neuroscience Database Gateway (NDG), accessible at http://www.sfn.org/ndg which provides a centralized gateway for accessing neuroscience-related databases (fig 1).

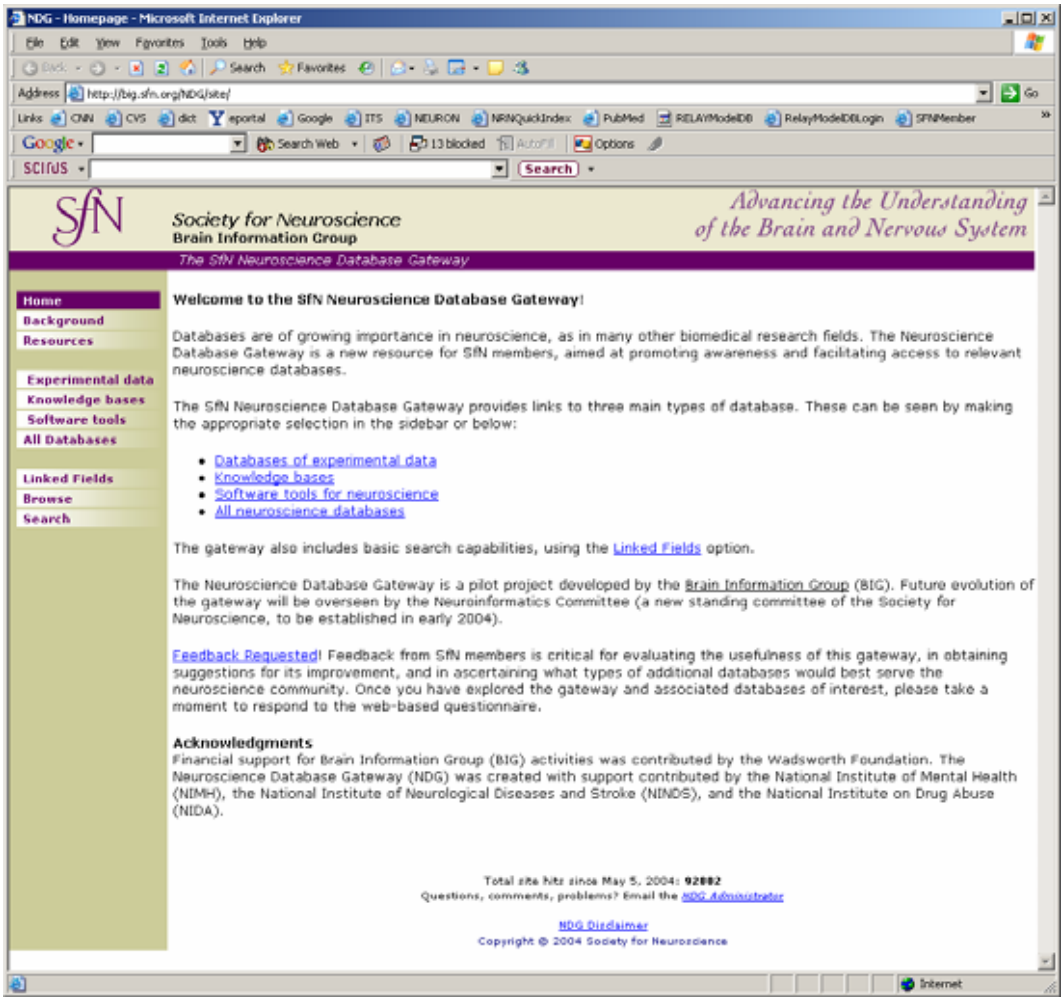

Figure 1. The Neuroscience Database Gateway (NDG) homepage.

In order to assess the significance of this event, it is useful to briefly indicate the background leading up to the NDG, and the challenges it presents for the future.

1. The Origins

Although the history of neuroinformatics now stretches back over a decade (an eon in the present climate of current biomedical science), anyone who has been in the field during that time knows that establishing this as a new field and having its value recognized by neuroscientists has been an uphill battle. This was predictable; even the most successful bioinformatics resources, such as GenBank (http://www.ncbi.nih.gov/Genbank), for sequences, and the Protein Data Bank (http://www.rcsb.org/pdb), for protein structures, met initial resistance before becoming critical research tools for biomedicine.

Although individual databases, mostly for brain atlases, began to be created in the late 1980s, neuroinformatics as a new discipline was launched in a formal way by the Human Brain Project. This was an initiative, led by the National Institute for Mental Health and incorporating other agencies and NIH Institutes, which began accepting applications to carry out research combining neuroscience and informatics in 1993, the same year that the first full featured browser, Mosaic, appeared and took the internet by storm. The early creation of the Human Brain Project while the web was so nascent shows the foresight of the early NIH working group. Over the succeeding decade that effort has grown to include nearly 40 projects, many with web-accessible databases (*see* http://www.nimh.nih.gov/neuroinformatics/researchgrants.cfm), covering subfields from molecular biology through cell morphology and physiology, to neural circuits, brain atlases, brain scans, clinical neurology, and behavior. With the realization that the new era of neuroinformatics was rapidly unfolding, in July 2003 the Society for Neuroscience (SFN), under then-President Huda Akil, convened a Brain Interest Group (BIG) chaired by Floyd Bloom to discuss how the Society could begin to play a role in helping to organize and support the development of neuroscience databases for the benefit of its members and the field. Members of the BIG were largely drawn from the Human Brain Project and the Editorial Board and contributors to *Neuroinformatics*.*

Over several meetings, what evolved was a joint discussion as to how neuroinformatics and neuroscience together can advance our understanding of nervous systems, and the role of the Society in informing the broader neuroscience community of such advances and resources.

2. The Gateway

The initial success of this collaboration led to a proposal to set up a gateway which would give coherent access to the burgeoning number of databases. A working group, led by David Van Essen, and building on resources of the Human Brain Project, developed the NDG. The pilot HBP database, HBPDB, had been created previously by Gordon Shepherd and his collaborators that stored information on the projects supported by recent grants from the Neuroinformatics Human Brain Project of the NIH (fig 2).

The HBPDB only contains 9 databases, but influenced the development of the NDG whose initial version contains 70 databases (fig 3) and allied resources.

A structured access method enables the user to navigate the many databases disciplines, levels of organization, and laboratories represented (fig 4).

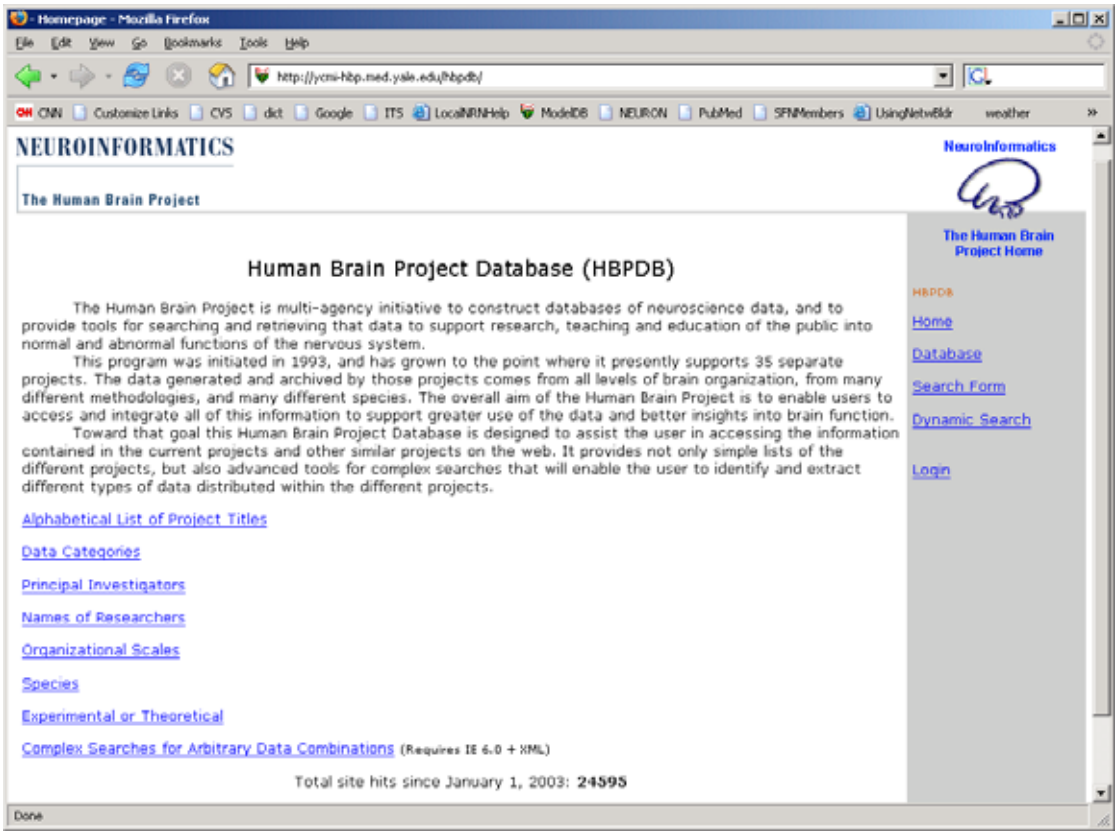

Figure 2. The Human Brain Project Database (HBPDB) homepage.

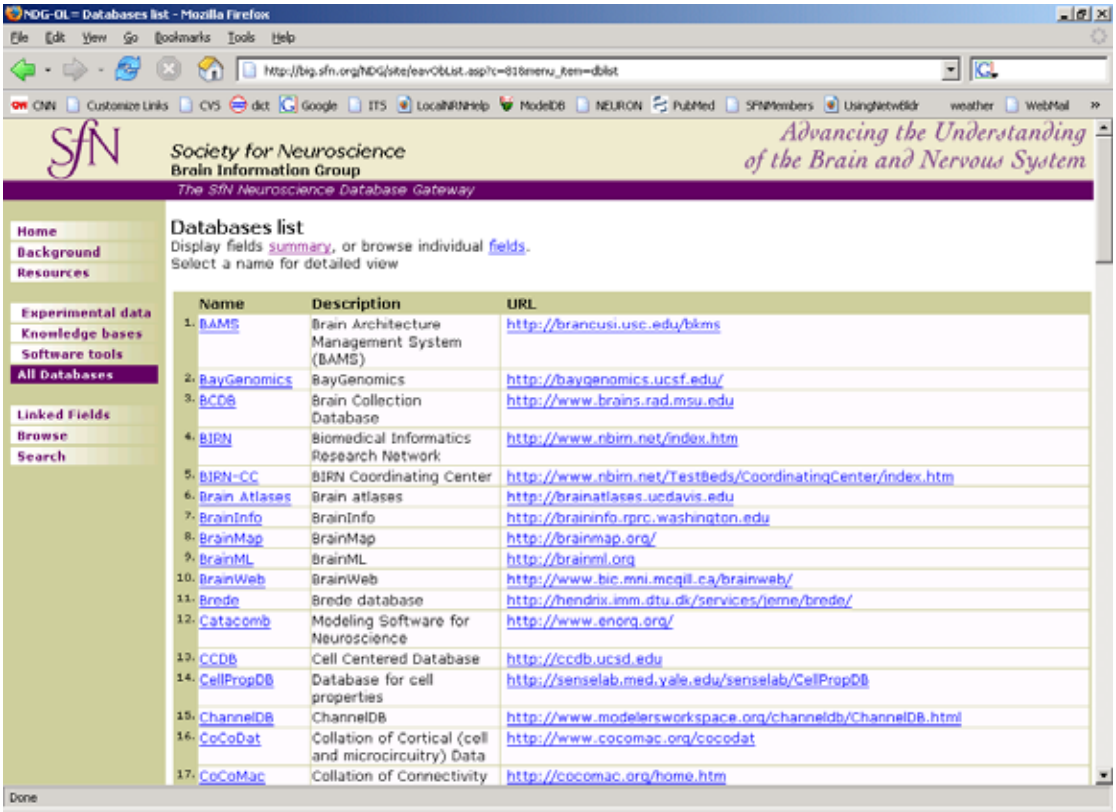

Figure 3. Over 80 databases have entries in the NDG database.

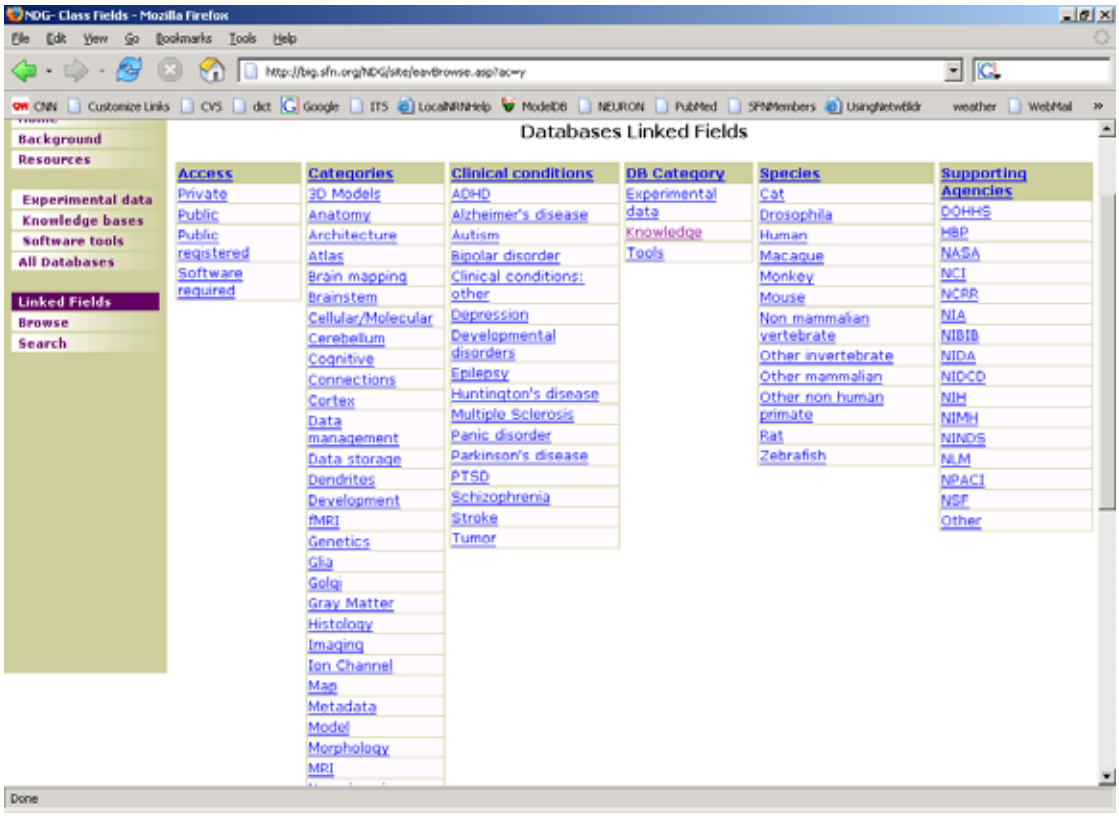

Figure 4. Multiple fields assist navigation of the NDG.

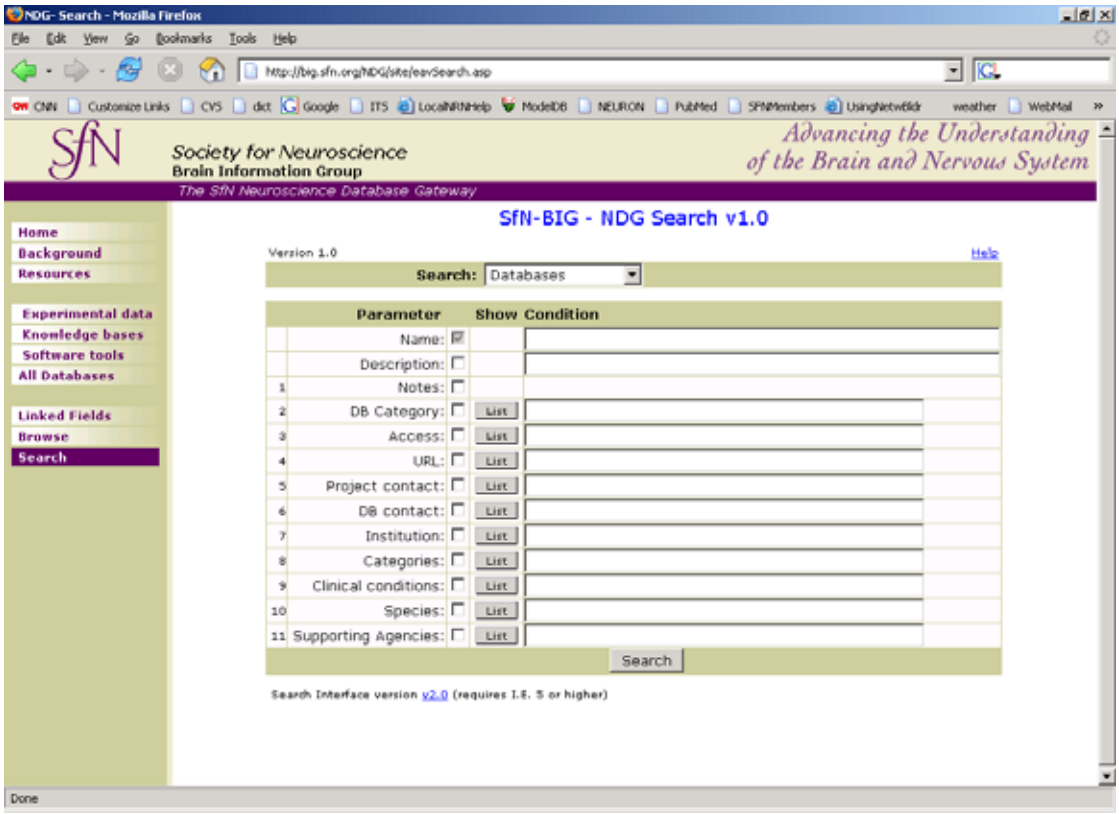

Figure 5. NDG complex search tool.

Procedures are being set up to admit further databases and to provide for increasingly sophisticated tools to enable users to access information through complex Boolean searches.

Although this effort is a significant achievement, it is really only the beginning. Most importantly, the challenge now facing the neuroscience community is to begin to bring itself into the digital era. This will involve a series of challenges, many of which are recognized in the other success of the BIG, a comprehensive White Paper that balances both opportunities and cautions for neuroinformatics (http://web.sfn.org/content/Programs/NeuroscienceDatabaseGateway/whitepaper.html).

3. The Future

A joint goal for the Society and the neuroinformatics community is to bring the 30,000 members of the SFN into the new age of neuroinformatics, so that our data do not remain restricted to use by the originating labs. For this, the Gateway is only the first step in enabling neuroscientists to leverage developments in informatics as successfully as have many other biomedical communities. Despite progress, acceptance by neuroscientists has been slow. This is largely owing to the difficult nature of neuroscience data, which go far beyond simple sequence data to include morphological images and physiological recordings, and which require multilevel descriptive metadata. Such technical difficulties have slowed development of databases serving specific subfields of neuroscience.

Moreover, neuroscientists have been reluctant to follow the lead of molecular biologists and deposit their data to available databases at the time of journal publication. For neuroinformatics and neuroscience to integrate fully and to develop tools equivalent to Genbank and the PDB, more databases must be developed and more users encouraged. Neuroscience database development must recognize that many researchers consider their data inappropriate for distribution or difficult to prepare for sharing. Indeed, many types of data are heavily nuanced and qualified by both experimental protocols, external conditions, behavioral contexts, and technical specifications.

This is a technical challenge and a call for resources that should be recognized, adopted, and implemented. Databases alone are insufficient; tools and interfaces that support and ease data entry and acquisition are also needed. Isolated databases will be less useful than those that provide interoperability through standard, community-supported descriptors.

We note also that neuroscience databases can both complement and enhance the utility of journal publication. As access to sharable and reanalyzable data is facilitated through development of databases, interest will be spurred in published descriptions of initial research carried out with such data, as well as the context in which they were recorded or derived.

Development and availability of databases is a prerequisite for, but not a guarantee of, community acceptance and widespread use. Policies and incentives are needed to support populating these databases. Sharing data by submission to a database at the time of publication is needed for neuroscience data, similar to the standard procedures for sequence data. A major step in this direction was taken by the *Journal of Cognitive Neuroscience*, which in 2000 began requiring deposit of published functional brain images and the datasets on which they were based to a brain-image database (http://www.fmridc.org) at the time of publication. Although this was initially

met with considerable opposition in the field, the initial setup received a strong stamp of approval by the scientific advisory board representing the brain imaging field, and this has now become routine procedure. The *Journal of Neuroscience*, the official publication of the SFN, in 2002 recommended to its authors that they similarly submit their images and datasets to the database at the time of publication (Shepherd, 2002). The *Journal of Computational Neuroscience* recently published an editorial that strongly encouraged its contributors to deposit their models to ModelDB (Richmond, 2004).

The NIH has recently instituted a policy of requiring that proposals totaling more than $500,000 in direct costs provide for sharing of the data generated (http://grants.nih.gov/grants/policy/data_sharing/). A growing number of neuroscientists support efforts in this direction (Gardner et al., 2003).

The strengths of the Society are its size and its breadth, effectively incorporating essentially all of US neuroscience, and representation beyond as well, and the Society should leverage these strengths:

- The major goals are dual: to make neuroinformatics a fundamental component of neuroscience and to extend the benefits of neuroinformatics to all aspects of neuroscience. Together, these build bridges between neuroinformatics and neuroscience.
- Integration towards a unified understanding of brain function will require coordinated vocabulary development across fields. We note that the Society was an early leader in this area with the adoption of a comprehensive keyword list developed by Bernice Grafstein for abstract submission.
- The Annual SFN Meeting can, in principle, serve as a complement to the all-neuroinformatics HBP spring meeting. SFN meeting abstract themes/topics should be expanded beyond the current bioinformatics to explicitly include neuroinformatics.
- The Society should consider establishing a section of the journal for neuroinformatics.
- The parallel field of computational neuroscience has many resources to offer.
- Readers of *Neuroinformatics* are encouraged to inspire by example and to promote neuroinformatic exploration and technique adoption and development by the broader neuroscience community.

To carry this synthesis further, the SFN is forming a standing committee on neuroinformatics.

The great challenge for all will be to apply neuroinformatic expertise to specific and varying needs of different communities and techniques within neuroscience in order to produce a synthesis as valuable for neuroscience as a whole as existing resources are for molecular and structural biology. At this early date most neuroinformatics databases are populated with data created from or collected by the same group that developed the database. Careful attribution of scientific work provided from databases, the possibility of becoming a parent of subsequent derived work, and the joy of transparency, openness, and pride will increasingly motivate outside contributors to participate. The coming era's web-shared diverse data, protocols and methods, mathematical tools, and models will be the enhanced foundation upon which new cathedrals of science will be built.

*In addition to ourselves, Douglas M. Bowden, Gwen A. Jacobs, Edward G. Jones, Maryann E. Martone, David Van Essen, and Robert W. Williams.

References

1. Gardner, D., Shepherd, G.M. (2004) A Gateway to the Future of Neuroinformatics. Neuroinformatics 2(3), 271-274.
2. Gardner, D., Toga, A.W., Ascoli, G.A., et al. (2003) Towards effective and rewarding data sharing. Neuroinformatics 1, 289–295.
3. Richmond, B. (2004) Editorial Commentary. Journal of Computational Neuroscience 17 (1): 5-5, July - August, 2004
4. Shepherd, G.M. (2002) Supporting databases for neuroscience research. J. Neurosci. 22, 1497.