1st INCF Workshop

on

Global Portal Services for Neuroscience

September 3-4, 2007 - Stockholm, Sweden

www.incf.org

1st INCF Workshop on Global Portal Services for Neuroscience

September 3-4, 2007
International Neuroinformatics Coordinating Facility Secretariat
Stockholm, Sweden

Authors

Jaap van Pelt and Gordon Shepherd

Scientific Organizer

Gordon Shepherd, Yale University, New Haven, USA

Workshop Participants

Jan G. Bjaalie, INCF Secretariat, Stockholm, Sweden
Chinh Dang, Allen Institute for Brain Research, Seattle, USA
Mark H Ellisman, University of California San Diego, La Jolla, USA
Daniel Gardner, Weill Medical College of Cornell University, New York, USA
Michael F Huerta, National Institute of Mental Health, NIH, Bethesda, USA
Colin D Ingram, University of Newcastle, Newcastle-upon-Tyne, UK
Kathie L Olsen, National Science Foundation, Arlington, USA
Raphael Ritz, INCF Secretariat, Stockholm, Sweden
Andrea Schenone, University of Genova, Genova, Italy
Gordon Shepherd, Yale University, New Haven, USA
Arthur W Toga, University of California Los Angeles, Los Angeles, USA
Shiro Usui, RIKEN BSI, Wako, Japan
David C Van Essen, Washington University, St. Louis, USA
Jaap van Pelt, VU University, Amsterdam, the Netherlands (Rapporteur)
Robert W. Williams, University of Tennessee, Memphis, USA

Workshop observers

Jim Austin, UK
Esther P Gardner, USA
Andreas Herz , Germany
Luciano Milanesi, Italy
Elli Chatzopoulou, INCF Secretariat
Pontus Holm, INCF Secretariat
Ylva Lillberg, INCF Secretariat

Supported by the EU Special Support Action INCF, the INCF Central Fund and the Swedish Foundation for Strategic Research

Contents

1. Executive Summary

Coordinating Development of Portals/ Databases

INCF was established by definition to coordinate activities in neuroinformatics, i.e., coordination and promotion of databasing/sharing, analytical tools and computational models. Portals were considered to be basic tools not only in neuroinformatics activities in the field but also for the INCF mission itself. As the amount of neuroscience data explodes, so do the numbers of databases and associated portals, and the relations, often overlapping, between them. It is urgent that these activities be coordinated to make them accessible and useful to the neuroscience community. The committee determined that it is essential that the INCF take a lead role in this coordination.

1. INCF should be proactive in establishing a list of maintained and developing neuroscience portals/databases. This list should include links to development sites and descriptions of the objectives of the project. The level of disclosure of any new development project should be regulated by the developer.

2. INCF should: i) act as a broker to bring together individuals with expertise in particular areas of development, ii) provide guidance on the availability of development resources, and iii) advise on current "best practice", including infrastructure recommendations, design processes, and scalable source control. This "best practice" should be conducted to enable interoperability via improved and pervasive ontology, have accessible API/scriptable interface, and have exposed database schemas.

3. INCF should generate a policy statement on open source and accessibility of projects in development.

4. INCF could provide a repository or library for the code base, data, etc., of projects no longer in development which might be reused in developing new resources.

5. INCF representatives should be observers at NSF, NIH, DOD, Creative Commons, UK e-Science All Hands, and other meetings on core technologies related to cyberinfrastructure, and pass appropriate information on to developers.

6. INCF should play a role in identifying and facilitating the development of unique value-added projects, such as an international database of neuronal connectivity.

7. INCF should identify and coordinate requirements (standards) that assist in the acquisition of data, metadata, and development of databases.

In addition, there needs to be a standard practice/policy on the release of statements and information. Methods and routes for dissemination vary between internal documents within nodes and those for external consumption. How do you measure/evaluate success and how are people to be held accountable? Success measures must be considered in the context of on-going funding/support from the National Nodes. Annual reports can fulfill some of these requirements both in the preparation and its dissemination. This could include summaries of node activity, which will help with good practice.

Last but not least, opportunities for running undergraduate courses in neuroinformatics should be considered as well as including neuroinformatics in medical training. Educational material could also be made available via the portal.

2. Introduction

The goal of the workshop was phrased as *to map out existing portal services for neuroscience, identify their features and future plans, and outline opportunities for synergistic developments*. *The workshop aimed to discuss alternative formats of future global and integrated portal services*.

The workshop covered two full days and followed a program of talks and discussions. The first day included brief presentations by each of the participants, identification of several key areas for discussion and development, and an initial meeting of workgroups in these areas. Individual presentations focused on topics of direct or indirect relevance for the overall workshop goal and included areas such as: perspectives on how to organize neuroscience data on a global scale, examples of large scale portals existing and under development, examples of large, individual databases, examples of networks of databases, and institutional and funding perspectives. The program on the second day included a continuation of discussions within the workgroups and preparations of conclusions and presentations, followed by plenary reports of the workgroups, a general discussion, and working on a final draft for the recommendations to INCF on coordinating neuroscience portals and databases for further dissemination.

The workshop was considered to be timely in orienting the effort toward a new era of global organization of portals and databases for accessing neuroscience information.

An overview of portal facilities presented at the workshop is given in Appendix A1. The discussions focused mainly on the role of INCF in global portal services. The outcomes of these discussions are summarized in the next section.

3. Workshop Discussions

3.1 Role of INCF in global portal services for neuroscience

Discussions led to the idea of a 'portal of portals' as an instrument for INCF to overview available portal services, to make bridges between them, and to promote and advise in the development of new portal services. INCF could thus provide a portal of portals, with emphasis on coordination (such as mirroring of sites in the case of heavy traffic, or providing a unified look and feel of computational environments). In providing directories of portals and databases, INCF should not just duplicate what exists but give added value, for instance, by adding short descriptions and characterizations of databases that will be helpful to the users.

Discussions resulted also in the view that INCF should adopt both a supportive role when different initiatives are taken in the field, and a proactive role in identifying issues and facilitating new directions.

The committee discussed a wide range of issues that relate to the role of INCF in coordination and support of global portal services. A general model for this coordination is summarized in the next section.

3.2 Defining a portal

Role/Content of portals

The question "What is a portal?" was raised. Is it just a website with links, or a system that integratively connects databases? Portals can function (i) as a resource registry/repository (GEO, boring but necessary), (ii) as a sophisticated search engine (Google, NCBI, UCSC, NDG, NIF), (iii) as a knowledge generator/integrator (BIRN, ABA, GeneNetwork), (iv) as a playground and resource bazaar for experts only (Bioconductor for array data code), (v) as a dynamic encyclopedia (Wikipedia), (vi) as a local community resource tool (example of ALZforum, CTC society site). Expectations of a portal were phrased as *getting good answers to standard questions* and *getting good answers to questions we didn't even know we could ask*. Portals may also function as grid-based e-science environments with Biolab and CARMEN as typical examples. Participants agreed that an INCF portal of portals should be in the service of the Neuroscience Community in the broadest sense, both in research and in education. As different user communities are generally not mutually exclusive, neither are portals servicing these communities. Thus, while showing much complementarity, portals may also have overlapping functionality. It was noted that a vacuum exists in using databases and models to generate testable hypotheses.

Lessons learned

A significant contribution came from Arthur Toga, who summarized his experiences with a variety of portal projects for large, multisite imaging consortia (such as The International Consortium for Brain Mapping (ICBM), The Alzheimer's Disease Neuroimaging Initiative (ADNI), The Biomedical Informatics Research Network (BIRN-Mouse, Morph, F), The Huntington's Disease Neuroimaging Initiative (HDNI), The fMRI Data Center (fMRIDC), The NIH MRI Study of Normal Brain Development (NIHPD), ACE).

Mission – The mission of the project should be clearly defined. An important issue is whether the project's aims are creating databases or doing science.

Data – There is a general consensus about the data to be made available through the portal.

Composition of the consortium – It is important to know the people participating in the project, and to allow them to decide their own participation on the basis of expectations, skills, and complementarity. The required expertise within the consortium should be present and proven.

Design – The portal needs to be designed for longitudinal use.

Technology/tools – The technology for data sharing must be in place and available for the implementation of the portal functions. An upgrade policy is needed in view of future technological developments in order to guarantee stability in portal operations. Otherwise, the project follows a moving target, controlled by technological developments.

Scientific focus – The scientific focus of the project needs to be well defined (which does not always happen to be the case). Most projects are too broad, with the risk of reducing the impact of the project. However, there are also examples of good projects with less focus.

Size of the consortium – Different rules appear to be needed to maintain coherence within large or small consortia. Large consortia appear to raise many issues of competence.

Duration of the project – Long-lasting projects will inevitably face many changes in their context and be confronted with continuous requests for additions, with the risk that the project loses focus.

License issues – License issues need to be defined at the beginning of the project. One has to be aware that 'open source' has many meanings.

Leadership – It must be known who is running the show. A committee may have diverging views.

Funding – Funding is essential for continuity of the portal, otherwise the game is over. This issue inevitably becomes pertinent at the end of a project and raises many problems.

Sharing – It is important to agree on further differentiation of data and tool sharing with respect to who (the people to share with), how (the mechanism of sharing), when (the conditions for sharing), what (the content to be shared), etc. For instance, the people to share with can be differentiated according to their function (such as home, guest, member, leader, manager) and the mechanisms of sharing can also be differentiated (such as Browse, Download, Upload, Access content, Project info). The Table below gives an example of sharing differentiation.

	Home	Guest	Member	Leader	Manager
Browse	*	*	*	*	*
Download		*	*	*	*
Upload			*	*	*
Access content				*	*
Project info					*

Table 1: Example of sharing differentiation

3.3 Developing a portal

Global collaboration

An important goal, outlined by Kathie Olsen of NSF, is expanding support of collaboration on a global scale with international partners. For instance, INCF may assist groups of database developers in Japan in finding partners, and give support for the international embedding of Japanese resources.

Enabling technologies

Portal services are strongly driven by enabling technologies such as high-performance computing, GRID infrastructures and visualization tools. Examples of such portals are BIRN, CARMEN and Biolab. It is strongly recommended that INCF builds on the opportunities arising from these technological developments.

Sustainability

Questions were raised concerning the lifetime of data, archiving the data, and how to maintain and support database facilities over the long term, requiring high investments in data centers. These questions essentially concern sustainability issues. The importance of sustainability was recognized by INCF, which is organizing a dedicated INCF Workshop on Sustainability of Neuroscience Databases, December 13-14, 2007, at the INCF Secretariat in Stockholm, Sweden.

Interoperability

INCF has a primary role in promoting interoperability. Adherence to interoperability requirements may lead to convergence on common architectures. INCF will build on "best practices" from other experienced web communities such as ATOM (http://www.atomenabled.org/), OAI-PMH (http://www.openarchives.org), RDF (http://www.w3.org/RDF), OWL (http://www.w3.org/TR/owl-ref), and OpenID (http://openid.net). Interoperability makes it possible to put findings of different projects together, which may reveal surprising new correlations (an example mentioned by Arthur Toga is that studies of cortical maturation show that it begins in childhood at the central sulcus, while independent studies of ageing show that degeneration occurs in reverse order; database interoperability brings these two complementary findings together).

Evaluation

The question concerning the organization around portals was discussed. Japan has adopted a scheme of annual evaluations of platforms and databases, and this was determined to be a good model for INCF. Criteria must be formulated for these evaluations.

Steering committee for portal services

It was agreed that a steering committee is needed for supporting this INCF coordination role. The present workgroup on Portal Services could become such an INCF steering committee on portal services, acting in the first instance through the organization of annual workshops.

3.4 INCF role in coordination & support

Many issues of INCF coordination concern data-related concepts, such as data models, data formats, definitions of terms, ontologies, etc., which are the formalisms that allow computation upon, and organization and querying of, the data. By coordinating data-related concepts, informatics tools and informatics resources will become increasingly accessible and interoperable. It must be recognized that such coordination will be an ongoing and dynamic process.

When to coordinate?

INCF should carefully select the areas for further stimulation. The community to address needs to be selected as being best poised, and be willing to be coordinated and to provide the required input of data and information. The community needs to be identified (for instance, by the type of data being used or the domain of research or the existence of digitized data or the use of neuroinformatics tools). An educated estimation of the extent of the task is needed (i.e., how big is the bite?). It is recommended that INCF formulates a list of criteria that helps selecting areas for coordination. Also the existence of adjacent communities, where neuroinformatics tools are already routinely used, is important. It was deemed crucial for INCF to find and coordinate selected areas for pilot projects, which make something happen and stimulate new developments.

Initiation of new projects/developments

INCF implements its coordination role by organizing workshops, providing portal services, stimulating ontology development, and supporting topical program areas such as large-scale modeling and digital atlasing. The workshops are the instruments to recommend projects and themes for further development. Recommendations may also concern new services to be supported by a portal, or even setting up new portals. Recommendations for new initiatives, such as building databases, should be based on actual overviews of current developments in the areas concerned, such as databases in development. INCF has a primary role in producing these overviews. INCF may initiate unique value-added projects, such as facilitating data mirroring through collaborations. A concrete example is the collaboration with the Allen Brain Institute to set up a mirror site for the Allen Brain Atlas.

Assistance

INCF can provide assistance in the development and maturation of portals and databases, in the formulation of open-source datasharing/databasing policies, and in the formulation of criteria for acceptance. Such assistance can be targeted to developers, to funding agencies and journals, and to industry, etc. INCF can play a role in periodic evaluations of databases and providing constructive feedback. INCF can function as an honest broker to help individual investigators in searching for the availability of existing databases and databases in development for sharing, or in finding alternative data sources. INCF can provide logistical support. There are potentials for INCF toward atlas consensus, spatial reconciliation, and directions for finding data to test datasets for segmentation.

User communities

User communities are the *raison d'être* of portals and databases. They define the spectrum of integrated services to be provided by the portal, such as information about PhD programs, conferences, meetings, etc, and with links to databases and other portals, to tools and technical developments. For INCF to play a meaningful role it is crucial to have information about the communities serviced by the portals, i.e. who are the people or committees involved in the portals. Services supporting social interactions are also important, as was demonstrated by Shiro Usui for Japan, where a variety of tools is already available via the internet for organizing feedback and user interaction, such as blogging, tagging, planets, clouds —examples from industry, such as Amazon, can also be cited.

National Nodes

It was discussed whether each National Node should develop its own portal, and what role INCF should play in their coordination. Because National Nodes have a coordinating role at the local level, they may wish to develop their own portals to accommodate services for their "local" communities. INCF could/should watch over their global embedding.

Technical development of portals

INCF can provide advice and resources in the technical development of databases. Assistance can be given to implement interoperability requirements as well as standards on data and metadata, ontology "standards" and the use of markup languages (such as NeuroCommons Mashup Browser in ABA, for example), the integration in semantic webs, open-source policies, etc. INCF can also advise on infrastructure, e.g., with reference to the NSF cyberinfrastructure, and possibly with support from NSF observers. It was noted that many policies and standards are currently in various phases of development themselves, yet that is no reason to wait until their stabilization. The graphical toolbox of Shiro Usui could be helpful in visualizing the relations and homologies in different ontology and nomenclature schemes. These tools may also help visualize relationships between portals.

3.5 Opportunities

Industry

INCF may coordinate processes toward new standards in instrumentation as a service to communities and industry.

Education and training

Portals may additionally be instrumental in the coordinating role of INCF with respect to education and training in neuroinformatics. INCF may promote the development of targeted E-science programs and virtual lab facilities, for instance via EU funded Initial Training Networks.

Link portals with journals

Transformations in the scientific publication process may dramatically impact how neuroscientists share and mine experimental data. A recent PubMed Plus conference in St. Louis, MO, USA, June 2007, on "New Directions in Publishing and Data Mining" discussed the formation of a new Neuroscience Peer Review Consortium for sharing reviews between neuroscience journals. This Consortium has been operational since January 1, 2008. INCF could also advise journals by notifying them when data is suitable for database storage, and which databases are available for different kinds of data.

4. Recommendations

4.1 Recommendations for a General Coordination Model and Roles

As the amount of neuroscience data explodes, so do the numbers of databases and associated portals. It is urgent that these activities be coordinated to make them accessible and useful to the neuroscience community. It is essential that INCF takes a leading role in this coordination.

What is being coordinated by INCF? Fundamentally, it is data-related concepts such as data models, data formats, definitions of terms, ontologies, etc., which are the formalisms that allow computation upon, and organization and querying of, the data. By coordinating data-related concepts, informatics tools and informatics resources will become coordinated and increasingly interoperable. It is recommended that a general model for this coordination would include the following elements:

- Recognition that coordination will be an ongoing and dynamic process

- Roles in coordination

  - INCF provides organizational and logistical support

  - Leadership and implementation comes from the research community
    - Thought leaders
    - Community-at-large

  - Iterative process between products by thought leaders and modification based on input from community-at-large

  - Implementation by those developing informatics tools and resources (e.g., databases)

- Parsing neuroscience into subcommunities

  - Based on data types (e.g., genetic, circuits, neuroimaging, etc.)

  - Identify opportunities to integrate across these (see bullet below)

- Keeping in mind potential hooks to informationally-adjacent subcommunities (e.g., genetic data-defined community and protein structure data-defined community) as coordination within a subcommunity develops, so conventions and standards that emerge in one community can be leveraged by, and serve as a potential venue of integration with, another community

- Identification of which subcommunities are well poised or ill poised for such coordination—work first with subcommunity(s) that are best poised

  - Develop criteria to identify which is well or not well poised. Potential criteria might include:
    - Are data widely gathered in digital form?
    - Do members use informatics and other computational tools?
    - Is there evidence of the need for coordination of informatics efforts (e.g., via editorials, recommendations from meetings, etc.)

- INCF coordination might be at the level of disseminating or otherwise facilitating the use of particular, already generally accepted (if only by major players) data-related concepts (e.g., data models, data formats, vocabularies, ontologies) OR at the level of sharing information about developing and emerging data-related concepts.

4.2 Specific recommendations to the INCF

- INCF should identify relevant portals worldwide and evaluate their complementarity or overlap. National Nodes should be encouraged to participate in this analysis. There is an urgent need to coordinate this vast array of portals and data.

- INCF could serve as an honest broker across neuroinformatics efforts. INCF could not only point to ongoing neuroinformatics tools and resources, but also characterize these, using objective criteria, providing this information to potential users. Such criteria could relate to specific aspects (such as ease of data entry and suitability of particular databases for sharing data in particular communities, etc.).

- The criteria developed to characterize neuroinformatics tools and resources could be posted and the research community could be encouraged to indicate how well a particular tool or resource satisfies each criterion.

- INCF could educate investigators about the availability and utility of neuroinformatics tools and resources useful for their research. INCF might also mediate between investigators and database managers.

- INCF should develop and sustain expertise about databases suitable to allow data sharing connected to journal publications.

- INCF could work with journals to facilitate data sharing by increasing awareness of databases appropriate to their respective journals. Reviewers and editors could encourage authors to submit their data to relevant databases when appropriate.

- INCF should help investigators to deposit their data into appropriate databases.

- INCF could serve as a coordinator of policies related to the sharing of neuroscience data, and neuroinformatics tools and resources across different funding agencies (public and private) across different countries—perhaps by stating best practices and promulgating those across these agencies and countries. A useful starting point would be for INCF to ask each node to provide information about such policies of major neuroscience and neuroinformatics-funding organizations in their country.

- INCF should facilitate training opportunities across information science and neuroscience at undergraduate student, graduate student, and postdoctoral levels.

- INCF should establish a steering committee to provide advice on the implementation of these recommendations. It is critical to have follow-up that would include physical meetings of this committee at appropriate venues.

Appendix A: Summary of portals discussed during the workshop

	URL addresses	Remarks
Japan node and platforms	http://www.neuroinf.jp http://nijc.brain.riken.jp	Organization and quality - Steering committee, Platform coordinating committee, annual evaluation, each platform services own community
NIF Neuroscience Information Framework	http://neurogateway.org/	NIH - catalog of electronic and non-electronic neuroscience research resources,
SumsDB Surface Management System Database	http://sumsdb.wustl.edu/sums/index.jsp	Repository of brain-mapping data
NDG Neuroscience Database Gateway	http://ndg.sfn.org/	Selection criteria for inclusion (level of maturation, functionality, how rapid is information obtained)
SenseLab	http://senselab.med.yale.edu/	Models of neurons and neural systems
Neurondb The neurophysiology database	http://neurodb.dertech.com/	Database for neurophysiological data storage, retrieval, and analysis
BrainPharm	http://senselab.med.yale.edu/BrainPharm/	Database to support research on drugs for the treatment of different neurological disorders.
Bio-Lab Laboratory for bioimages & bioengineering	http://www.bio.dist.unige.it/	DIST – Genova – portal function, grid approach, clinical applications. Integration of data and services, complexity hidden for users
Neurodatabase.org	http://neurodatabase.org	Neurophysiology database + tools anatomical visualization
NeuroMorpho.Org	http://neuromorpho.org/	Inventory of digitally reconstructed neurons
ABA Allen Brain Atlas	http://www.brain-map.org	Image database of gene expression in the mouse brain.
CARMEN Code Analysis, Repository and Modelling for e-Neuroscience	http://www.carmen.org.uk	EPSRC e-science Pilot Project Including database for electrophysiology, virtual laboratory, grid enabled, repositories, toolkit, integration of electrophysiological and morphological data
BIRN Biomedical Informatics Research Network	http://www.nbirn.net/	A National Information Infrastructure to Enable and Advance Biomedical Research
CCDB Cell Centered Database	http://ccdb.ucsd.edu/	High resolution 2D, 3D and 4D data from light and electron microscopy,
The GeneNetwork	http://www.genenetwork.org	Resources for systems genetics.
NITRC The Neuroimaging Informatics Tools and Resources Clearinghouse	http://www.nitrc.org/	NIH - neuroimaging resources for fMRI and related structural analyses

Appendix 2: Workshop Program

September 3, 2007

08.30 – 09.00	Introductions (Bjaalie and Shepherd)
09.00 – 18.00	Scientific presentations and discussions
Gordon Shepherd	Introduction and orientation
David C. van Essen	Opportunities and challenges in navigating
Shiro Usui	Japan-node portal and the platforms
Andrea Schenone	Some experiences in sharing data and services through e-science portals
Gordon Shepherd	Experiences building the Neuroscience Database Gateway
Daniel Gardner	The Neuroscience Information Network: An open source solution and template for web neuroscience
Chinh Dang	Creating a community-driven neuroscience webware
Jaap van Pelt	Database opportunities for the Netherlands
Colin D. Ingram	CARMEN: Developing an e-science portal to support collaborative research in time-series neuroscience data
Mark H. Ellisman	The Biomedical Informatics research Network (BIRN) and the Cell Centered Database (CCDB) – Building collaborative environments for integrating and sharing neuroscience knowledge
Arthur W. Toga	Multi-site imaging consortia: Lessons learned
Robert W. Williams	Interoperability among mouse genomic, neurogenetic, and neuroimaging web resources
Raphael Ritz	The INCF Neuroinformation Community Portal
Michael F Huerta	Neuroscience and informatics: Intersecting priorities at NIH
Kathie Olsen	NSF: Perspectives on cyberinfrastructure and neuroscience
20.00 –	Dinner

September 4, 2007

08.30 – 12.00	Discussion
12.00 – 13.30	Lunch
13.30 – 18.00	Discussion

Each presentation was approximately 20 minutes, including questions.

www.incf.org

INCF Secretariat
Karolinska Institutet
Nobels väg 15 A
SE-171 77 Stockholm
Sweden

Tel: +46 8 524 87 093
Fax: +46 8 524 87 150
E-mail: info@incf.org

incf International Neuroinformatics Coordinating Facility