

The Need for Open-access, Structured Data in Clinical Brain Research

Adomas Bunevicius

Neuroscience Institute and Department of Neurosurgery, Lithuanian University of Health Sciences, Kaunas, Lithuania
<http://lsmuni.lt/en/structure/medical-academy/-neuroscience-institute/>

ARTICLE INFO:

RECEIVED: 22 Sep 2014
REVISED: 02 Nov 2014
ACCEPTED: 14 Nov 2014
ONLINE: 04 Jan 2015

KEYWORDS:

Open Access
Dataset
Brain Diseases
Disease Models
Data Sharing

ABSTRACT

Neurological and psychiatric disorders are considered major public health problems associated with a significant number of suffering patients and an enormous socioeconomic burden to societies. Therefore, significant research funds have been allocated towards improving outcomes of patients suffering from brain disorders. In recent years, researchers have been facing a constantly growing worldwide pressure to make their datasets freely and readily available to other researchers and the public. The major goal of this movement is to allow re-use of already collected data in order to facilitate medical progress. There are numerous ways that clinical neuroscience researchers and research organizations commonly employ to make their databases available. Institutional databases and publicly available research and clinical data repositories and registries are the most common ways to share data among researchers and organizations in order to facilitate collaboration. In addition, numerous traditional neuroscience peer reviewed journals have implemented various data sharing policies. Furthermore, a few traditional clinical neuroscience peer review journals have recently started accepting dataset papers. Finally, there is a constantly growing number of database journals which accept clinical neuroscience research datasets for publication. Potential for improving of citation metrics is among the important advantages for authors considering publishing their databases in dataset journals.



Creative Commons BY-NC-SA 4.0

© 2015 Hosting by Procon Ltd. All rights reserved.

The human brain is a susceptible organ to numerous disorders that may be attributed to both external and internal causes and develop across a lifespan. Brain disorders comprise a vast and heterogeneous group of diseases affecting both central and peripheral nervous systems at both structural and functional levels. Neurological disorders and mental illness are considered two major groups of brain disorders. Neurological disorders include but are not limited to neurodevelopmental disorders, neurodegenerative disorders, cere-

brovascular diseases, nervous system malignancies, infections, trauma and functional disorders. Neurological disorders affect more than 1 billion people worldwide and account for 6.8 million deaths, which is 12 percent of total deaths.¹ Population based studies carried out in Western societies have estimated that more than one quarter of individuals met the diagnostic criteria for at least one mental disorder within a 12-month period.^{2,3}

From an individual patient perspective, neurological and psychiatric disorders are considered among the most serious and debilitating conditions and are often associated with severe functional disability, significant reduction of health related quality of life and poor prognosis. Due to the lack of effective treatment modalities, brain disorders are often incurable disorders resulting in gradual and progressive functional decline. From a public health perspective, neurological and psychiatric disorders are considered the major public health problem associated with significant socio-economic burden for societies that are mainly attributed to significant financial efforts required for treatment of patients with brain disorders and progressive decline in patient productivity. For example, the economic cost of brain disorders was estimated at about 798 billion Euros in 2010 in Europe.⁴ According to the World Health Organization, neurological and neuropsychiatric disorders should be considered among the most serious public health problems causing greater disability than infectious disorders, malignancies, ischemic heart disease, respiratory diseases and digestive diseases.⁵ It has been estimated that neurological disorders contributed to 92 million of disability-adjusted life years (DALYs) lost in 2005.⁵ The same study has projected a 12% increase of DALYs accounted by neurological disorders in 2030. The global burden of neurological and psychiatric disorders is expected to increase significantly in the future mainly due to increased longevity and improved patient survival.⁵ As a result, studies aiming to elucidate the underlying biological mechanisms of brain disorders and to improve outcomes of patients suffering from neurological and psychiatric conditions remain an important research funding priority.^{6,7}

These research efforts have significantly advanced our understanding about underlying biological mechanisms of brain disorders and improved outcomes of patients suffering from neurological and psychiatric disorders. Nevertheless, much more remains to be done in order to improve recognition and expand treatment option of patients suffering from brain disorders. To this end, there is a rising worldwide movement encouraging sharing of scholarly research data with investigators and other interested parties.⁸ Data sharing is expected to be the next giant step allowing re-using already collected scholarly data in order to re-validate current knowledge and generate new hypotheses leading towards improved patient outcomes. Researchers as well as academic research institutions are facing a consistently growing worldwide push from societies,⁹ re-

search funding agencies¹⁰ and political and licensing authorities¹¹ to make their raw structured scholarly research data openly and readily accessible for other researchers and public. For many researchers and research organizations it is no longer a question whether scientific data should be shared and that the amount of shared data will increase exponentially in the nearest future. Aside from well expected benefits to society via accelerated improvement of outcomes of patients suffering from brain disorders, other potentially important advantages of data sharing for individual researchers and/or research organizations should be acknowledged. Importantly, data sharing is associated with greater citation count. For instance, it has been demonstrated that papers with publicly available data have received significantly greater number of citations relative to papers without publicly available datasets.¹² Furthermore, data sharing is expected to increase visibility of researcher and institutional research output, improve publication and data quality and transparency, increase collaboration opportunities among researchers and research organizations, reduce research costs allowing to re-use already collected data and fulfil data sharing requirements imposed by certain funding agencies and peer-reviewed journals. Also, data sharing can have an important moral benefit making each individual member of the scientific community feel more as a team player of a scientific community with an ultimate goal of more expedited advancement of research progress that would eventually translate into reduced disease burden and patient suffering.

However, data sharing is still in its infancy stage facing numerous technical, institutional, ethical and cultural hurdles that preclude from universal implementation of data sharing culture.¹³ From an individual researcher perspective, a paucity of peer-reviewed journals accepting raw scientific biomedical research datasets for publication, lack of knowledge on how the raw data should be handled and presented, and uncertainty about potential uses and/or misuses of the data can be considered as important obstacles for data sharing. In addition, lack of knowledge of universally accepted ethical guidelines and policies regarding raw research data sharing, especially when considering studies involving human subjects, is another important barrier for data sharing. Furthermore, storage of large datasets can impose significant challenges to traditional publishers since data storage and sharing is associated with a need to improve technological infrastructure that can consequentially increase publishing costs to the journal.

To date, there are numerous ways that researchers and clinicians working in the field of neuroscience employ to share their research and clinical data (see Table 1). Deployment of clinical and research databases in institutional data repositories is the most commonly employed way by researchers and research organizations to store research data. Importantly, some organizations and researchers' groups are interested in data sharing with other interested parties. The major goals of such data sharing are to improve visibility of organization, support reliability and transparency of their research findings and attract new collaborators for possible future research endeavours. For example, the Clinical Research in Neurology Registry of the Emory University is a prospectively collected longitudinal database of individuals suffering from various neurological disorders that includes over 2500 clinical and biological variables. According to the program's website, the program is interested in collaboration and interested parties are encouraged to contact the organization. Another example of similar approach is the Clinical and Research Database for Persistent Schizophrenia of the Department Of Psychiatry of University Of Cambridge. Researchers interested in schizophrenia research can contact a responsible person. However, such data storage does not provide with immediate benefits for investigators via traditional citations metrics.

Data from clinical trials can be obtained upon request from study investigators once the dataset has been released. Sharing of clinical trial data is important from clinical perspective since findings from such studies are often used to approve new drugs and devices in standard patient care. Sharing of clinical trial data is strongly encouraged by funding and licensing authorities. An example of such clinical trial data sharing is the *National Institute for Neurological Disorders (NINDS) Recombinant Tissue Plasminogen Activator (rt-PA) Stroke Trial* that was a pivotal trial published in 1995 demonstrating efficacy of intravenous rt-PA for management of acute stroke patients.¹⁴ Results of this trial have led to intravenous rt-PA approval for management of acute ischemic stroke patients. The original research data has become available more than 10 years after the initial publication of major findings and anyone willing to obtain original dataset should contact the NINDS and purchase the CD-ROM containing the data for US\$79. However, some authors have recently noted that access to the trial data was complex and some data appeared to be missing.¹⁵ Another example of such data sharing is the International Stroke Trial database that is freely available for

download under the Open Data Commons Attribution License. Finally, a number of psychiatry clinical trial datasets are available from the NIMH after completion of required documentation.

There are numerous national and international clinical registries that prospectively collect clinical and outcome data of patients suffering from brain disorders. Important goals of such registries are improving patient care and optimizing allocation of available healthcare resources. For example, The National Neurosurgery Quality and Outcomes Database (N²QOD) was established by the NeuroPoint Alliance, a non-profit organization of the American Association of Neurological Surgeons, and is currently managed by the Vanderbilt Institute of Medicine and Public Health. The data from such clinical data registries is usually available for participating institutions. Also, there is an increasingly growing interest towards establishing regional, national and/or international prospective clinical registries or databases of individuals suffering from rare neurological disorders that could potentially allow to pool sufficient number of patients for clinical research studies and facilitate collaboration of scientists and physicians.¹⁶ According to the ORPHANET, the portal of rare disorders funded by the French government and the European Commission, there is a total of 641 registries of which 74 have been classified as global. These registries usually contain clinical data and an access to such registries requires permission. Contributors of clinical data registries can usually receive traditional authorship credits by participating in writing scientific papers stemming from such registries.

Numerous promising public initiatives have been developed towards promoting clinical and basic neuroscience research data sharing.¹⁷ For example, the Neuroscience Information Framework (NIF) is a project funded by the NIH Blueprint for Neuroscience Research project and maintains the largest searchable collection of neuroscience data.¹⁸ The NIF is a user-friendly platform that contains over 174 available datasets and databases. However, it should be noted that the majority of datasets include basic neuroscience, namely neuroimaging and genetics data. Aside from numerous important advantages of the NIF and similar platforms, authors do not receive deserved credits via traditional bibliographic metrics.

Several traditional peer-reviewed clinical neuroscience and psychiatry journals require authors to make their datasets publicly available for reviewers and readers upon acceptance of the manuscript. *The Journal of Cognitive Neuroscience* has pioneered the field by requesting authors of

Table 1. Common ways of clinical neuroscience research data sharing.

<i>Sources</i>	<i>Comment</i>	<i>Example(s)</i>	<i>Major advantages for authors</i>
Institutional databases	Continuously collected registries that include patient clinical and biological data. Data can be made available for researchers upon request.	Clinical Research in Neurology Registry of the Emory university, Clinical and Research Database for Persistent Schizophrenia	Improved collaboration and networking opportunities. Enhanced author and institutional visibility and transparency.
Raw clinical trial data	Usually includes clinical trial data that can be obtained upon request. Analysis of such data can be complex.	The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, International Stroke Trial database	Enhances reliability and transparency of research data. Required by funding and licensing agencies.
Publicly available research data repositories	Contains large number of datasets and databases. Mostly neuroimaging and neuro-genetic data.	The Neuroscience Information Framework OpenfMRI, XNAT Central	Enhanced author visibility and networking opportunities, reliability of research data
Clinical data repositories and registries	Registries that include clinical and outcome data with a goal towards improving patient care. The data is usually available for research purposes. Allows to aggregate larger sample sizes of patients suffering from rare disorders of the nervous system	England's Compendium of Neurology Data, The National Neurosurgery Quality and Outcomes Database (N ² QOD), Registries of rare disorders of the brain	Collaboration and networking opportunities
Traditional peer reviewed journals with compulsory data sharing policies	Authors have to share datasets that can be either submitted as supplementary material or deposited in public repositories	PLOS, European Journal of Neuroscience, The BMJ	Increasing transparency of research findings
Peer reviewed journals with optional data sharing policies	Authors can optionally choose to share dataset upon manuscript acceptance	European Psychiatry, Biological Psychiatry	
Traditional peer reviewed journals accepting dataset papers	Peer-review journals publishing traditional research articles. Database articles describing datasets are also considered for publication. Datasets are stored in external repositories and should be made freely available for readers.	BMC Neurology, BMC Research Notes	Increasing number of publications and improving citation metrics
Specialized database journals	Consider only database papers. Papers describing dataset are published in the journal. Datasets are made freely available for download from the journal web-page or from public data repository.	The Biomedical Data Journal, Scientific Data, Journal of Open Public Health Data, Journal of Open Psychology Data, GigaScience, Dataset Papers in Science	Increasing number of publications and citation metrics; user-friendly interface; facilitated mechanisms of data sharing.

accepted manuscripts to make their datasets freely available for reviewers and readers. The *PLOS* journals and the *European Journal of Neuroscience* are examples of journals considering brain research studies for publication that require authors of accepted papers to make their databases used for manuscript writing freely available by depositing their databases in public data repositories or submitting them as a supporting document, depending on file size. Numerous other top journals in the fields of neurological and psychiatric research, such as the *European Psychiatry and Biological Psychiatry*, have adopted policies to deposit dataset used for preparation of accepted manuscripts as an optional requirement. *The BMJ* requires authors of drug and device clinical trials to deposit their data in public data repositories. The major goal of such policies is to enhance reliability and transparency of research findings presented in traditional peer-reviewed journal papers. However, authors do not receive a citation credit for the deposited databases.

Only a few traditional peer-reviewed neurology and psychiatry journals included in the major bibliographic databases consider database articles for publication. Recently, the *BioMedCentral* group have started considering dataset articles for publications in their journals.¹⁹ For example, the *BMC Neurology* and *BMC Psychiatry* accept database articles describing “a novel biomedical database likely to be of broad utility”. In the latter journals, papers describing dataset are published in the journal while the datasets are stored in external repositories and should be made readily accessible without restrictions for non-commercial researchers upon acceptance of the manuscript. Additional more direct advantages for authors interested in publishing their dataset article in traditional peer-reviewed journals are stemming from the potential to improve traditional bibliographic and citation indexes, such as increasing the total number of publications, and from the potential for citation count increase. However, the previously mentioned potential technical issues are important limitations for traditional publishers willing to consider dataset papers for publication.

Specialized database journals are expected to fill in the majority of gaps imposed by the above mentioned currently available research and clinical data outlets. Most importantly, authors publishing their datasets in specialized database journals will receive immediate well-deserved credits, such as increasing publication count and potential for greater citation number. Also, specialized database journals are expected to have adequate

technological infrastructure that would provide authors with user-friendly interface and with more sophisticated user-friendly tools to share (i.e., upload and download) databases and metadata. Albeit the number of specialized biomedical database journals currently remains limited, the field is rapidly expanding. Researchers working in the fields of clinical neuroscience and psychiatry can consider the following journals: *Scientific Data Journal* (published by the Nature Group), *Journal of Open Public Health Data* and *Journal of Open Psychology Data* (both journals are published by the Ubiquity Press), *GigaScience* (published by the BioMed Central and supported by BGI, a Chinese non-profit organization) and *Dataset Papers in Science* (published by Hindawi). Despite the inter-journal variations in journal scope, author guidelines and management of datasets, database journals have adopted open-access policy to databases and papers describing the database acquisition and design.

In this context, we are delighted to launch *The Biomedical Data Journal* (BMDJ) that is an open-access specialized biomedical dataset journal published by the Procon Ltd. in collaboration with the European Commission funded OpenScienceLink project. The BMDJ together with the OpenScienceLink platform is expected to provide individual researchers and research institutions with a reliable outlet for sharing structured research data. In addition, the OpenScienceLink platform will provide authors with numerous additional services aiming to facilitate research collaboration, enhance the visibility of their research findings and follow their research utilization by employing novel citation metrics. The BMDJ will consider data papers describing original datasets from various clinical neuroscience disciplines, including psychiatry, psychology, neurology, neurosurgery and neuroradiology. The journal will also consider original scientific investigations pertaining to brain health at all levels, ranging from epidemiological research studies to clinical investigations concerned with diagnosis, management and prevention of brain disorders. Interdisciplinary studies on interrelated disciplines of genetics, immunology and endocrinology integrating knowledge about brain health will also be considered for publication. Papers from clinical investigations studying mental health and neurologic comorbidities in patients with somatic disorders are also strongly encouraged.

Computational modelling techniques have provided important tools that have significantly expanded our understanding how electrical and chemical impulses (and their interaction) affect

brain functioning under normal conditions and in disease states.^{20,21} Computational modelling techniques have an enormous potential for various clinical applications in patients suffering from neurological and mental disorders.^{22,23} However, computational modelling studies require large amounts of experimental and clinical data. Towards this end, an important goal of the BMDJ is to gather a sufficient amount of biological and experimental data that could subsequently be used as an important source for computational modelling studies.

Acknowledgement

The preparation of this manuscript was supported by the OpenScienceLink project (Grant agreement #318652).

References

- ¹ The World Health Organization. The world health report 2007: a safer future: global public health security in the 21st century. Geneva, Switzerland: WHO Press; 2007. Available from: www.who.int/whr/2007/en.
- ² Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005;62(6): 617–27. DOI: 10.1001/archpsyc.62.6.617.
- ³ Wittchen HU, Jacobi F. Size and burden of mental disorders in Europe – a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 2005; 4: 57–76. PMID: 15961293.
- ⁴ Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jönsson B; CDBE 2010 study group; European Brain Council. The economic cost of brain disorders in Europe. *Eur J Neurol*. 2012 Jan;19(1):155-62. DOI: 10.1111/j.1468-1331.2011.03590.x.
- ⁵ Dua T, GarridoCumbreira M, Mathers C, Saxena S. Global Burden of Neurological Disorders: Estimates and Projections. In: *Neurological Disorders: Public Health Challenges*. Geneva, Switzerland: WHO Press;2006. p. 26-39.
- ⁶ National Institutes of Health. Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC). 2014 March 7. Available from: http://report.nih.gov/categorical_spending.aspx.
- ⁷ Sobocki P, Lekander I, Berwick S, Olesen J, Jönsson B. Resource allocation to brain research in Europe (RABRE). *Eur J Neurosci*. 2006 Nov;24(10):2691-3. DOI: 10.1111/j.1460-9568.2006.05116.x.
- ⁸ Uhler PF, Schröder P. Open Data for Global Science. *Data Sci J*. 2007;6:OD36–OD53.
- ⁹ Hiltz PJ. A Law Opening Research Data Sets Off Debate. *The New York Times*. 1999 Jul. Available from: www.nytimes.com/1999/07/31/us/a-law-opening-research-data-sets-off-debate.html.
- ¹⁰ National Institute of Health. NIH Data Sharing Policy and Implementation Guidance. Available from: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- ¹¹ U.S. Food and Drug Administration. OpenFDA. Available from: <https://open.fda.gov/about>.
- ¹² Piwowar HA, Day RS and Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2007 Mar 21;2(3):e308. DOI: 10.1371/journal.pone.0000308.
- ¹³ Wicherts JM, Borsbom D, Kats K and Molenaar D. The poor availability of psychological research data for reanalysis. *Am Psychol*. 2006 Oct;61(7):726-8. DOI: 10.1037/0003-066X.61.7.726.
- ¹⁴ The National Institute of Neurological Disorders and Stroke, rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Eng J Med*. 1995 Dec 14;333(24):1581–7. DOI: 10.1056/NEJM199512143332401.
- ¹⁵ Dachs RJ, Burton JH, Joslin J. A user's guide to the NINDS rt-PA stroke trial database. *PLoS Med*. 2008 May 20;5(5):e113. DOI: 10.1371/journal.pmed.0050113.
- ¹⁶ Rare Disease Registries in Europe. Orphanet Report Series. Rare Diseases collection, 2014 Jan. Available from: <http://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf>.
- ¹⁷ Poline JB, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, et al. Data sharing in neuroimaging research. *Front Neuroinform*. 2012 Apr 5; 6:9. DOI: 10.3389/fninf.2012.00009.
- ¹⁸ The Neuroscience Information Framework. About NIF. Available from: <https://neuinfo.org/about/ind ex.shtml>.
- ¹⁹ Hrynaszkiewicz I. A call for BMC Research Notes contributions promoting best practice in data standardization, sharing and publication. *BMC Res Notes*. 2010 Sep 2;3:235. DOI: 10.1186/1756-0500-3-235.
- ²⁰ Nakagawa TT, Jirsa VK, Spiegler A, McIntosh AR, Deco G. Bottom up modeling of the connectome: linking structure and function in the resting brain and their changes in aging. *Neuroimage*. 2013 Oct 15;80:318-29. DOI: 10.1016/j.neuroimage.2013.04.055.
- ²¹ Dauvermann MR, Whalley HC, Schmidt A, Lee GL, Romaniuk L, Roberts N, Johnstone EC, Lawrie SM, Moorhead TW. Computational neuropsychiatry - schizophrenia as a cognitive brain network disorder. *Front Psychiatry*. 2014 Mar 25;5:30. DOI: 10.3389/fpsy.2014.00030.
- ²² Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B. Alzheimer's Disease Neuroimaging Initiative. How early can we predict Alzheimer's disease using computational anatomy? *Neurobiol Aging*. 2013 Dec;34(12):2815-26. DOI: 10.1016/j.neurobiolaging.2013.06.015.
- ²³ Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry* 2014 Jul;1(2):148-158. DOI: 10.1016/S2215-0366(14)70275-5.