ARTICLE IN PRESS

YNIMG-06245; No. of pages: 3; 4C:

NeuroImage xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Comments and Controversies

Lost in localization: A minimal middle way

Antonia F. de C. Hamilton

School of Psychology, University of Nottingham, Nottingham, UK

ARTICLE INFO

Article history:
Received 19 March 2009
Revised 5 May 2009
Accepted 7 May 2009
Available online xxxx

ABSTRACT

Commentaries by Derrfuss and Mar (Derrfuss, J., Mar, R., 2009-this issue. Lost in localization: the need for a universal coordinate database. Neuroimage.) and Nielsen (Nielsen, F.A., 2009-this issue. Lost in localization: a solution with neuroinformatics 2.0? Neuroimage.) outline the need for a universal coordinate database and some possible approaches to creating one. I highlight the issue of minimal or maximal database scope and advocate a bottom-up approach to this problem.

© 2009 Elsevier Inc. All rights reserved.

Recent commentaries by Derrfuss and Mar (2009-this issue) and Nielsen (2009-this issue) discuss the need for a comprehensive database of neuroimaging publications organised by coordinates. I agree with this proposal and would like to highlight two key issues: the level of detail needed in a neuroimaging database and the different possible ways of creating and maintaining the database. Level of detail could be either minimal, with just the critical information extracted from each publication, or maximal with a taxomony of studies and all reported variables. Implementation could be achieved by a bottom-up collaboration lead by a small group of neuroscience researchers, by a top-down commercial organisation or by an entirely end-user contributed 'neuroinformatics' approach. Derrfuss and Mar favour the top-down solution while Nielsen argues for the user-driven approach. Based on my experience of developing a small coordinate database for neuroimaging results, I would like to argue for a minimal, bottom-up approach.

In 2004/5, I developed AMAT (a meta-analysis toolbox, http://www.antoniahamilton.com/amat.html) a Matlab-based, open source interface for searching fMRI coordinates together with a simple database of coordinates. The AMAT database is deliberately designed to be minimal. Effectively, the database reproduces the tables of *XYZ* coordinates which are common in fMRI papers. Each coordinate is associated with the anatomical label given by the authors of the original paper, a flag for Talaraich or MNI coordinates, a very brief description of the description of the functional task or contrast which activated this coordinate, and the PubMed ID of the published paper. The latter links directly to the abstract in PubMed and allows the user to retrieve the original publication. Anatomical information labelling a coordinate as a particular Brodmann area or functional region is optional, and is normally only included if the authors of the original paper included these labels. No other information is stored. The only criterion which must be met for a coordinate to be eligible for entry into AMAT is that it must be published in a peer-reviewed paper.

This minimal format has three major advantages. First, it means that data entry is fast and does not require any subjective interpretation of the methods of a paper. AMAT deliberately does not include data on the number of participants in the study or the significance or size of the cluster or other 'analysis' variables, because these are hard to compare across studies. Similarly, it does not include information on stimuli or responses or tasks. To input these, the data entry clerk must read and understand the text of a paper and make a decision about how to best describe an experimental paradigm in a standardised format, which makes data entry slow and intellectually demanding. In contrast, minimal data entry can be carried out by people without training in neuroimaging and takes 20–30 minutes per publication. Automisation could reduce data entry time further. Tables in published papers could be read using software which extracts data from pdf files, while data from newly submitted papers could be converted automatically to database format when the publication is accepted by a journal. Standard journal formats for tables would facilitate this process (Poldrack et al., 2008). Critically, reducing data entry time substantially reduces the cost of generating and maintaining the database.

Second, the minimal format facilitates searching by coordinate. The typical user might be a researcher who has found an unexpected result in an fMRI study and wants to know — who else found activations near to 32, 24, 78? The AMAT interface provides a list of coordinates, ordered by Euclidian distance from the search location coordinate and allows the user to find each research paper. Other functions such as activation likelihood meta-analyses (Laird et al., 2005a) would be a secondary use but are possible when a minimal database is used in conjunction with other search tools.

Third, AMAT's minimal description of the contrasts in each paper deliberately forces the user to refer to the complete original manuscript for full details of a study, rather than relying on the AMAT database alone. This means that users cannot be mislead by a database

E-mail address: antonia.hamilton@nottingham.ac.uk.

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.05.007

summary of the contrasts or experimental conditions which might not reflect a critical feature of the original study. More importantly, users cannot be lazy and use a database as a substitute for reading original papers. The purpose of the database is to guide the user towards published work, not to replace the need for studying and understanding the scientific literature.

The alternative to this minimal approach is a ‘maximal’ approach, of which several examples exist. An extreme maximal approach was taken by the Dartmouth fMRI Data Centre which archives complete raw fMRI datasets (Van Horn et al., 2004). Only 122 datasets are available and the database has not been accepting new data for the last two years, so it seems that data archiving on this scale has yet to live up to its initial promise. In contrast, the BrainMap database (Laird et al., 2005b) contains 1721 papers and stores coordinates as well as subject characteristics, stimulus and response details and a classification of a paper coded in a taxonomy of research domains. Similarly, the Brede database (Nielsen, 2003) stores information from 186 papers on task, scanner type, analysis software and significance levels as well as the critical coordinates and publication data. Some of this information is valuable for meta-analysis but not all is necessary or even helpful. For example, any taxonomy or ontology of research topics forces a user to fit their ideas into the categories given, and may find it hard to deal with newly emerging areas such as social cognition (which is not in the BrainMap taxonomy). Finally, both of these databases are severely limited by the time (maybe a couple of hours) and skill (graduate level studies in neuroimaging) required to convert a journal publication into the database format, because both require a detailed interpretation of the original paper to appropriately fill the database fields. Requiring all this information means these database would need more funding to generate and maintain their data than a minimal database like AMAT.

The choice between a maximal and minimal format depends to some extent on what question you want a database to answer. Questions of the form “what activates this brain coordinate?” can be easily addressed from a minimal database which refers the user to the relevant papers in PubMed. Questions of the form “where in the brain does function X occur” might seem to be more suited to a maximal format which includes information on research domain, tasks, stimuli and responses. However, use of a minimal database in conjunction with a PubMed search to identify papers in the appropriate field would yield useful results even for this type of question.

Based on my experience with AMAT, I suggest that a minimal approach to neuroimaging databasing can be valuable and economical. The AMAT database encompasses the minimum amount of neuroimaging data per publication which is necessary to be useful. This would not preclude additional information being stored in some cases, but to require excess detail and classification of research at the outset would only add to database cost. Devising a flexible and extensible database format with a minimal set of core requirements and scope for expansion is likely to be central to the success of future databasing efforts.

A second critical question is whether and how such a database could be established at all. Derrfuss and Mar (2009-this issue) describe both a bottom-up approach organised by a small group of researchers and a top-down approach in which a commercial organisation would shoulder the financial burden of organising data storage and would charge a subscription for researchers to access data. They suggest that the latter is more feasible. Nielsen (2003) proposes a ‘neuroinformatics’ approach in which end-users voluntarily add information to an open platform and have the capacity to download and remix the data as they wish. In considering the merits of these options, the experience of AMAT may provide some insight.

AMAT is most similar to the end-user driven approach described by Nielsen, though it lacks a web interface. A major aim in AMAT was to minimize data entry time, in the hope that this would encourage users to add to the database and allow the database to grow in an organic fashion like Wikipedia. Though AMAT has been frequently downloaded, it has not received user contributions in the way I had hoped. In particular, while AMAT provides a reasonable sampling of fMRI datasets prior to 2005, it does not yet contain any more recent data. The experience of AMAT suggests that purely user-driven contributions are not substantial. An improved interface and better support might increase the rate of user contributions, but my experience suggests that Nielsen (2009-this issue) is optimistic when he promotes a wiki or user-driven solution to the data entry problem. Most wikis also lack the enforced, coherent structure that is needed to make a database fully searchable. The success of a database depends to a large extent on its completeness, and it is unlikely entirely voluntary data entry could achieve the required level of submissions.

Does that mean the only appropriate solution is a top-down, commercially backed, subscription only service as suggested by Derrfuss and Mar? As Nielsen highlights, such a system would likely restrict searches and analyses to those specified by the company's own interface. Copyrighting of the data itself and the need for institutional subscriptions would further limit access. This would severely reduce the potential for researchers to develop new meta-analysis tools or new ways to browse and visualize data as they might with an open access database. Restricting innovation in this way and tying neuroimaging data to the whims (and profit margins) of a commercial organisation cannot be in the interests of the neuroimaging community.

I suggest that the neuroimaging community can learn from the successful database efforts carried out by researchers in other fields. The National Library of Medicine maintains databases of genes, proteins and macromolecular structures as well as the PubMed database of abstracts. These databases developed out of the efforts of researchers within those fields to make sense of their own data, and a similar collaboration must be within the capacity of neuroimaging researchers. Many journals (e.g. Science and Nature) and grant agencies (e.g. ESRC & MRC in the UK, NIH in the USA) already have data sharing policies which require researchers to submit data to an appropriate database or archive when a paper is published. The problem for neuroimagers is that no database exists. If an appropriate body were to establish a basic database, and if major journals in the field (e.g. NeuroImage and Human Brain Mapping) were to require submission of published coordinates to the database by the authors of each study, then it is likely that other journals would follow suit and the database would quickly grow to accommodate the future of neuroimaging. Adding the back-catalogue of past studies could be a more gradual process, with contributions both from authors keen to see their past work cited more frequently and from paid data entry clerks. Such an effort would, of course, require funding but building a minimal database and using open source software would mean that the initial outlay need not be large. Given the value of such a database, a strong case could be made to suitable funding bodies.

Derrfuss and Mar describe a similar bottom-up solution in their commentary but suggest that lack of resources would make this hard to achieve. I maintain that we must achieve this. An open access database of author submitted results endorsed and supported by the major players in the field will have the benefits highlighted by Nielsen of vast potential for innovations in meta-analyses as well as providing an authoritative source of information for all researchers. A cooperative effort to develop this database would represent the maturation of the field of neuroimaging and reflect the importance of brain mapping data in answering fundamental questions about how humans are able to interact with each other and the world around us.

Acknowledgments

AH is supported by ESRC grant RES-061-25-0138 and by Autism Speaks.

ARTICLE IN PRESS

References

Derrfuss, J., Mar, R., 2009-this issue. Lost in localization: the need for a universal coordinate database. Neuroimage.

Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Turkeltaub, P.E., Kochunov, P., Fox, P.T., 2005a. ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. Hum. Brain Mapp. 25, 155–164.

Laird, A.R., Lancaster, J.L., Fox, P.T., 2005b. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics 3, 65–78.

Nielsen, F.A., 2003. The Brede database: a small database for functional neuroimaging. Neuroimage 19, Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003.

Nielsen, F.A., 2009-this issue. Lost in localization: a solution with neuroinformatics 2.0? Neuroimage.

Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E., 2008. Guidelines for reporting an fMRI study. Neuroimage 40, 409–414.

Van Horn, J.D., Grafton, S.T., Rockmore, D., Gazzaniga, M.S., 2004. Sharing neuroimaging studies of human cognition. Nat. Neurosci. 7, 473–481.

Please cite this article as: Hamilton, A.F.C., Lost in localization: A minimal middle way, NeuroImage (2009), doi:10.1016/j.neuroimage.2009.05.007