NeuroImage 48 (2009) 1–7

Contents lists available at ScienceDirect

NeuroImage

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

Comments and Controversies

Lost in localization: The need for a universal coordinate database

Jan Derrfuss [a,*], Raymond A. Mar [b]

[a] Max Planck Institute for Neurological Research, Gleueler Strasse 50, 50931 Koeln, Germany
[b] York University, Department of Psychology, 4700 Keele St., Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 6 November 2008
Revised 15 January 2009
Accepted 18 January 2009
Available online 5 February 2009

ABSTRACT

One of the great advantages of neuroimaging research is the use of an established and uniform coordinate system. This 3-D coordinate system allows for the comparison of activation locations across studies. In order to capitalize upon this advantage, however, researchers must be able to find relevant studies based upon activation locations. A number of research groups have embarked upon solutions to this problem, but to date there exists no exhaustive, universal coordinate database. In this commentary we outline the nature of the problem, its current solutions, and propose alternate solutions. We close with suggestions on how those in the field can facilitate the process of developing a universal coordinate database.

One of the primary goals of cognitive neuroscience is to establish structure–function relationships in the human brain. That is, cognitive neuroscience aims at understanding what the different areas of the human brain do and how these areas cooperate to produce cognition and action. Functional neuroimaging plays a key role in this endeavour and, accordingly, the rate of empirical publications based on neuroimaging methods has been rapidly increasing for the past two decades (Fig. 1). Most of this increase can be attributed to the development of functional magnetic resonance imaging (fMRI) as a brain mapping technique. By the end of the year 2008 approximately 9400 fMRI studies investigating human cognition and action will have been published in English language journals (Appendix A). An estimated 74% of these studies (Appendix A) report the locations of statistically significant peak activations in a 3-D reference space (Evans et al., 1993; Mazziotta et al., 2001; Talairach and Tournoux, 1988; see Lancaster et al., 2007 for a comparison of the variants of this reference space).

Despite the importance of a standard 3-D coordinate system for functional imaging, to date there is no comprehensive coordinate database. That is, there is no way to find *all* of the published studies that report coordinates in a certain location. The utility of such a database has long been in the minds of researchers (Fox and Lancaster, 1994), and some important and commendable attempts to address the issue have been undertaken (Fox and Lancaster, 2002; Hamilton, 2005; Nielsen, 2003; Van Essen et al., 2008). However, we believe that the current databases are insufficient in the number of articles they include and we are concerned that in their current form they will not be able to keep up with the increasing rate of publication. In this commentary we discuss the need for a universal coordinate database, examine the current solutions available to researchers, and conclude by proposing a number of additional possible solutions to this problem.

How many relevant studies exist for a given location?

In order to establish structure–function relationships, researchers need to identify studies that report activations in similar locations. We can get a sense of the enormity of this seemingly mundane task by estimating the amount of information uncovered by an exhaustive search of this type. To estimate the number of studies that report a given location, we turned to the BrainMap coordinate database. This database contains coordinates that have been uploaded by researchers and students, for numerous studies (discussed below). As of October 2008, the BrainMap database contained 1601 papers and 58,600 coordinates. Dividing the number of activations in the database by the total number of studies gives us the average number of activations reported by a study in BrainMap, which is 37. Extrapolating this to all published articles suitable for inclusion in a coordinate database reveals that around 258,000 activations will have been reported by the end of the year 2008 (Appendix A). Given that the average gray matter volume of the human brain is estimated to be 780 cm^3 (Lüders et al., 2002), this means that on average approximately 330 peak coordinates have been reported by different studies for every cubic centimetre of gray matter in the human brain. (This is assuming that activations are evenly distributed throughout the brain; in reality this number is likely to be more or less depending on location [cf. foci density map in Van Essen et al., 2008].) Or, put differently, if you were to draw a sphere with a radius of 6.2 mm around a single activation from your study, on average, there would be ~330 activations from other studies located within this sphere. It is difficult to estimate how

* Corresponding author.
E-mail address: jan.derrfuss@nf.mpg.de (J. Derrfuss).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.01.053

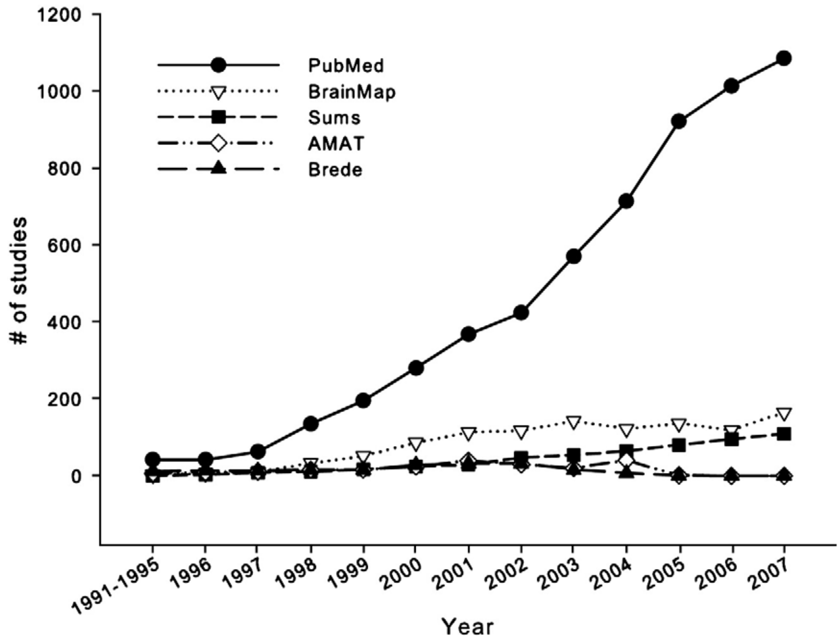

Fig. 1. Total number of published fMRI studies reporting coordinates by year and number of studies included in current coordinate databases. A detailed description of this data is presented in the Appendix A.

many different papers contribute these ~330 foci, as many studies report several different contrasts that often result in similar activations. Even from a conservative standpoint, however, this is likely to represent a large number of relevant papers.

How can we go about finding these studies?

Hundreds of relevant papers for a single activation is an impressive amount of information. Locating these articles in any sort of an exhaustive manner is currently very difficult, and quite possibly untenable, given current means. For the large majority of studies, we have access only to the information indexed by databases like PubMed or ISI Web of Knowledge, which do not include brain location coordinates. If we want to find studies that report activations similar to those in our own studies, there are currently three major ways to search these databases: by structure, by Brodmann area (BA), or by topic. Each of these methods suffers from a number of shortcomings. Structure-based and BA-based searches are unsatisfactory because nearby activations might receive different anatomical labels or different BA designations, and activations quite distant from one another might be given identical anatomical labels or BA designations. For example, activations close to the junction of the inferior frontal sulcus and the inferior precentral sulcus (Brass et al., 2005; Derrfuss et al., 2004, 2005, 2009) might be referred to as lying in the inferior precentral sulcus, the inferior frontal sulcus, the precentral gyrus, the middle frontal gyrus, the inferior frontal gyrus, pars opercularis, Broca's area, the premotor cortex, or the inferior frontal junction. On the other hand, activations in the most ventral and the most dorsal part of the posterior frontal cortex might both receive the label “precentral gyrus” or “BA 6,” although these activations might lie as far as 60 mm apart. The third way to search for relevant studies is to search for articles on the same topic as your own study, or from within the same paradigm class. While this approach might well assist you in locating relevant studies with similar activations, in many ways this constitutes a search for confirmatory information. From this method we may learn of similar studies that found similar activations, but will remain ignorant of dissimilar studies that found similar activations and thus limit our understanding of an area's function.

Searching current article databases like PubMed by structure, BA, or topic, are clearly inadequate methods for finding articles reporting activations close to a particular location. As a result, our functional descriptions of areas are likely to remain biased or incomplete. This problem will not only affect researchers who wish to put their results into context, but also researchers undertaking a meta-analysis or review. The neuroimaging community has become increasingly sensitive to the shortcomings of current search methods, which more often than not result in a very domain-centred approach. The end result is that groups of researchers from different topics all lay claim to a certain brain region–believing it specific to their own process of interest–and ascribe it different functions. By ignoring studies using a different paradigm or investigating a separate topic, one will capture only a very narrow picture of what is associated with a particular location in the brain. This is akin to the parable of the blind men and the elephant, with researchers each describing the function of a region in very idiosyncratic terms based upon the task employed. A number of recent publications illustrate the growing awareness of this issue. The journal *Cortex*, for example, recently published a special issue that attempted to reconcile and integrate various perspectives of Broca's area and the ventral premotor cortex (Schubotz and Fiebach, 2006). Other articles have attempted a similar integration of perspectives with regard to the superior temporal sulcus (Hein and Knight, 2008), temporoparietal junction (Mitchell, 2008), precuneus (Cavanna and Trimble, 2006) and posterior cingulate (Nielsen et al., 2005).

Of course, our best understanding of how the brain operates is bound to come from an understanding of networks of regions, how different brain areas interact and work together, rather than a modularist assignation of single functions to specific regions. That said, an understanding of how individual regions contribute to different networks in order to support very different processes, is likely to aid us in uncovering the underlying processes that contribute to these networks. The ability to easily locate studies that report activations in a particular area will be an essential part of this endeavour. A recent example of how looking across topics can lead to a better understanding of networks can be found in the growing interest in the default network, the collection of brain areas that appear active in the absence of external stimuli (e.g., Raichle et al.,

Table 1
Comparison of four existing coordinate databases.

	AMAT	BrainMap	Brede	SumsDB/Caret
Creator(s)	Dr. Antonia F. Hamilton (University of Nottingham, UK)	Dr. Peter T. Fox and Dr. Jack L. Lancaster (University of Texas, USA)	Dr. Finn Å. Nielsen (Technical University of Denmark/Copenhagen University Hospital, Denmark)	Van Essen Lab (Washington University in St. Louis, USA)
Link	http://www.antoniahamilton.com/amat.html	http://brainmap.org/	http://hendrix.imm.dtu.dk/services/jerne/brede/brede.html	http://sumsdb.wustl.edu/sums/index.jsp
Scope (as of Oct., 2008)	212 papers, 675 contrasts, 5379 foci	1601 papers, 7338 contrasts, 58,600 foci	186 papers, 586 contrasts, 3912 foci	1039 papers, 31,052 foci[a]
Most recent study from	2005	2008	2005	2008
Coordinate submission	Send data (.csv format) to creator	Scribe (Java GUI)	Matlab interface (via Brede Neuroinformatics Toolbox), send data (.xml format) to creator	Import data (.csv format) into Caret and SumsDB[b]
Search interface	Matlab	Sleuth (Java GUI)	Internet	Caret or Internet[c]
Additional information	Requires SPM2	Meta-analyses and coordinate transformation with GingerALE	Database also included in Brede Neuroinformatics Toolbox	

[a] This includes studies and foci from imaging modalities other than fMRI or PET (e.g., morphometric studies).
[b] Submission to SumsDB is still under development, but a beta version of the submission process has recently been released (a tutorial and instructions are available on the SumsDB website).
[c] Internet search interface offers restricted search options only.

2001; Mason et al., 2007). It has been empirically demonstrated that a number of different processes appear to draw upon this same network of brain areas (Spreng et al., in press), and theorists have proposed that it may support a single set of processes, such as self-projection (Buckner and Carroll, 2007) or scene construction (Hassabis and Maguire, 2007). Advances such as these would be greatly facilitated by a simple method of searching the entirety of relevant studies for a given area, based upon coordinate location.

Given that peak activations are reported in a common coordinate system, producing a database that associates these coordinates with the studies that report them would seem to be useful, necessary, and achievable. In practice, however, developing such a coordinate database has proven difficult and elusive, for reasons that we explore below.

Current coordinate databases

To date, a few research groups have set out to create a coordinate database akin to what we describe above; we briefly discuss four of the most popular databases here. All of these databases allow for coordinate-based searches and are freely available. Apart from these commonalities, the databases differ substantially in the number of articles included, the information available about these articles, the submission procedure, and a number of other relevant features (see Table 1 for an overview). It is apparent that BrainMap (Fox et al., 1994; Laird et al., 2005b) is the most comprehensive database (Fig. 1) followed by SumsDB; AMAT and Brede contain far fewer articles than these two databases. BrainMap and SumsDB also offer the greatest diversity of search options (Table 2) and, together with Brede (Nielsen, 2003), give the most detailed search results (Table 3). With AMAT (Hamilton, 2005), contributing new studies to the database is simple and fast, but as a result of low demands on the contributor the information provided by the database to users is rather limited. BrainMap and SumsDB, on the other hand, offer very detailed information about studies but this level of output means that a very time-consuming submission process must be completed by contributors.

All of the databases discussed are an important step in the right direction. However, even BrainMap contains only about 19% of the total number of fMRI articles published by the end of 2007 that are suitable for inclusion in a coordinate database (Appendix A). In a 2005 publication, Laird et al. (2005b) estimated that BrainMap would be able to keep up with the rate of publication for neuroimaging papers by shifting from a focus on voluntary submission to student coding of published papers. At the beginning of the year 2005, BrainMap contained approximately 500 articles (Fox et al., 2005). Since then the database has been growing at a mean rate of ~300 articles per year. Although this is a large number of articles and an impressive improvement over previous submission rates, this does not appear to be sufficient to keep up with current publication rates. Currently, around 1000 eligible articles are published per year, and this rate is rising annually. Moreover, it is unclear how studies are being selected for inclusion in BrainMap, an important concern since only a subset of the total number of articles is being included. Unless a new solution is proposed and implemented, we fear that BrainMap and other coordinate databases will be unable to provide a representative or exhaustive database of relevant studies.

The issue of how new studies can be included or submitted to a database seems to us to be the key obstacle for creating an exhaustive database, one that includes all of the relevant neuroimaging research. This, in turn, will affect the likelihood that such a database will be useful to researchers. We now move to a discussion of possible solutions to this problem.

Possible solutions to the problem of a coordinate database

As shown above, even the most comprehensive database we have to date contains only about a fifth of the relevant studies. The question thus arises as to what can be done to create an exhaustive database. In our view, the two major questions regarding this issue are whether a

Table 2
Comparison of search options for existing coordinate databases.

Search options	AMAT	BrainMap[a]	Brede	SumsDB/Caret[a]
Author	✓	✓	(✓)[b]	✓
Title	✗	✗	(✓)[b]	✓
Abstract	✗	✗	✗	✓
Keywords	✗	✓	✗	✓
Year	✓	✓	(✓)[b]	✓
Journal	✗	✓	(✓)[b]	✓
Subject characteristics	✗	✓	✗	✗
Stimulus/response types/modalities	✗	✓	✗	(✓)
Paradigm class	✗	✓	✗	✓
Coordinate search	✓	✓	✓	✓
Adjustable range	(✓)[c]	✓	✗[d]	✓
Multiple coordinate search	✗	(✓)[e]	(✓)[f]	(✓)[f]
Brodmann areas	✗	✓	✗	✓
Anatomical structures	✗	✓	✗	✓
Logical operators	✗	✓	✗	✓
PubMed ID	✗	✓	✗	✓

Notes. (✓) indicates search options with restricted functionalities.
[a] Sleuth and SumsDB/Caret offer additional search options; for brevity, only a subset of relevant options is included in the table.
[b] Via Google.
[c] The number of neighboring coordinates to be retrieved can be entered.
[d] The 30 closest coordinates will be retrieved.
[e] Only an OR search is possible.
[f] The 20 most similar studies (Nielsen and Hansen, 2004) are automatically retrieved.

Table 3
Comparison of search results for existing coordinate databases.

Search results	AMAT	BrainMap	Brede	SumsDB/Caret
Author	✓	✓	✓	✓
Title	✓	✓	✓	✓
Citation	(✓)[a]	✓	✓	✓
Abstract	(✓)[a]	✗	(✓)[a]	✓
Contrast	✓	✓	✓	(✓)[b]
Experimental Conditions	✗	✓	✗	✗
Distance between entered and retrieved coordinate(s)	✓	✗	✓	✗
Visualization of coordinate locations	✓	✓	✓	✓
Significance value	✗	✓	✗	✓
Volume of activation	✗	✓	✗	✓
Brodmann area	✓	✓[c]	✗	✓
Anatomical structure	✓	✓[c]	✓	✓
Brain template	✓	✓	✗	✓
Related volumes	✗	✗	✓[d]	✗
Number of subjects	✗	✓	✓	✗
Imaging modality	✗	✓	✓	✓
Scanner type & field strength	✗	✗	✓	✗
Export search results	✗	✓	✓	✓
Link to PubMed	✓	✓	✓	✓
Link to DOI	✗	✗	✓	✓

Notes. (✓) indicates restrictions in search results.
[a] Via Pubmed.
[b] Not yet provided for all studies.
[c] Via Talairach Daemon.
[d] Nielsen and Hansen (2004).

new database should be created and who should submit the relevant information to the database. Below we outline two approaches to answering these questions. These proposals should be considered merely examples of a possible solution and we acknowledge that a number of other workable solutions are likely possible.

The bottom-up approach

One approach to solving the current issue is to adopt a bottom-up strategy, in which an existing database is expanded and authors submit their own information to the database. As BrainMap is the largest coordinate database, it might be the best choice for such an expansion.

What advantages would this solution have? Obviously, a major advantage of this option is that no new database would have to be created, which means little effort and expense would need to be invested into this solution. The other main advantage is that authors know their studies best and thus appear to be in an ideal position to describe the tasks and analyses employed. Finally, we hope that a database created in this way will remain freely available, so universities and other research institutions would not have to pay for access to this valuable resource.

What problems would this solution face? In our view, the main problem with this option is ensuring that relevant articles are submitted to the database. We think that the limited success of current databases shows that an exhaustive database is not likely to be achieved with voluntary submission (Laird et al., 2005b). For this reason, we would argue that mandatory submission to the database is necessary if this approach is adopted. More specifically, neuroimaging journals could adopt a policy in which authors are required to submit their results (i.e., locations of peak activations for contrasts) to BrainMap as just another step in the publication process, along with signing the transfer of copyright and submitting final versions of figures. This would ensure that new articles are represented in the database.

If such a policy was adopted, however, a second problem arises related to the amount of information required to contribute to BrainMap. Currently, BrainMap requires a time-consuming submission process via Scribe. We suspect that the time and effort associated with this process is one of the reasons why this sophisticated database has not seen more submissions. The question arises whether this procedure could be changed in a way that would maintain the character of the database, but make submission easier and faster. In our view, a possible way to achieve this would be to explicitly define core information that must be entered into Scribe. In our view, this core information would encompass the complete citation (including the abstract), coordinates with their significance value and test statistic (e.g., *F*-value), the imaging modality, the name of the brain template employed, a short description of the task, the type of contrast computed (e.g., subtraction, parametric, functional connectivity), the number of subjects, the gender of the subjects, and whether the data come from an empirical study or a meta-analysis. Other entries could be made optional, with the possibility of updating a submission at a later point in time. This optional information might, for instance, include the size of the activation cluster, stimulus and response modalities, stimulus and response types, and the analysis software used. A further change that would significantly decrease time for submission to BrainMap would be the possibility to upload text files containing the relevant information (e.g., tables in comma-separated-values [.CSV] formatting).

Along these lines, to facilitate the creation of such a database, it would make sense for a standardized data format to be created that would allow for easy uploading and sharing of neuroimaging results. Analysis software could then provide a toolbox for exporting results into this data format, and these files would then be available for easy upload to a database. Once this format has been established, it is easy to imagine that users will create tools for creating custom databases, as well as develop new methods for search and data manipulation.[1]

We recognize that the mandatory submission policy described in this scenario might be a concern for some researchers because of the additional effort required for data submission. In our view, making database submission as fast and as easy as possible will be paramount for the database to gain acceptance in the neuroimaging community. Furthermore, researchers should consider submission to the database in their best interests, as doing so increases the likelihood that an article will be discovered by another researcher and thus cited.

To this point we have focused on articles that are to be published in the future. A separate problem is how to ensure that studies which have already been published are entered into the database. With the bottom-up approach, one solution is to appeal to researchers to submit all their previous studies. Societies such as Human Brain Mapping or Society for Neuroscience might encourage their members to participate in this undertaking for the good of the discipline, or perhaps by providing concrete incentives. Such a solution might lead to the exclusion of papers by researchers who are no longer active, however. Another possible solution is for societies to hire individuals, or perhaps recruit volunteers, who would work to enter old papers into the database. In line with this idea, one possibility is to establish permanent funding for the BrainMap database. This funding could then be used to ensure that the current backlog of published but excluded studies is eliminated, with BrainMap overseeing the volunteers and staff needed to enter these articles. Money could be contributed through donations, by societies who collect membership fees, or perhaps even by journals (who will pay to have their back catalogue entered into the database).

While we acknowledge that this bottom-up approach would require a great deal of organization and collaboration on the part of the neuroimaging community, this level of initiative and commitment is not without precedent. The establishment of the Neuroimaging Peer Review Consortium, for example, demonstrates the same level of collaboration on the part of neuroscience journals as would be required to organize mandatory author-based submission to a database.

[1] An anonymous reviewer is kindly thanked for suggesting this idea.

The top-down approach

Another likely option is the creation of a new database by a private company or a government institution that would then be responsible for entering coordinate information into the database. For example, such a database could be created by an indexing company (e.g., Thomson Reuters, Ovid Technologies), by a consortium of neuroimaging journals, or by the National Library of Medicine (akin to the Genome database accessible via *Entrez*).

What advantages would this solution have? Arguments in favour of involvement by government institutions or private companies include the fact that these organizations are already in the business of making scientific publications searchable and have far more experience and resources at their disposal than any single researcher, journal, or even scientific society. This means that a more ambitious approach can be taken toward the database, without being hindered by a paucity of resources. Also, for companies that already index scientific articles, a lot of the relevant information regarding previously published studies already exists in their own databases; adding a coordinate-based search capability to these databases seems to be a manageable undertaking. These companies are already in the business of indexing research articles, and have in place the infrastructure necessary to support this database through subscriptions. Capitalizing upon these existing resources appears very attractive. Another advantage is that researchers will not have to spend their own time uploading their results to a database. This information will be collected from the published article by the indexing body, along with the information already indexed by these companies (e.g., abstract, keywords, etc.).

What main problems would this solution face? Provided that one of the above-mentioned organizations recognizes the need for a coordinate database and is willing to establish it, how long it would take until this new database could be made accessible is an open question. Apart from the necessary technical requirements, this solution would require a number of decisions to be made regarding the nature of the database, hopefully in consultation with neuroimaging researchers. Also, if a private company was to create the database, this would necessarily result in a database that requires some sort of subscription fee. If this fee is small, and not excessively prohibitive, we feel that this will be only a small hurdle for the database since academic institutions should be willing to shoulder the burden for researchers. However, if only a small number of neuroimaging researchers exist at an institution it might reduce the chances that a library will subscribe, especially in light of the current economic hardships faced by academic institutions. Clearly, cost will be a key factor in the success of a top-down approach.

How to search?

No matter what type of solution is adopted, some decisions will need to be made regarding how the database will be searched. This will inevitably be determined in part by the type of information ultimately included in the database. In principle, we think that all the information entered in the database should be searchable. At bare minimum, searches should be possible using a variety of different methods, including single coordinates, spheres of possible coordinates, keywords, title, and authors. It should also be possible to conduct searches for sets of coordinates, so that it is possible to search for networks (i.e., find papers that report activations within a set of locations). In addition, we consider it important that quantitative meta-analyses (Chein et al., 2002; Laird et al., 2005a; Turkeltaub et al., 2002; Wager and Smith, 2003) be included in the database and it be possible to restrict searches to these types of papers. This would allow for the identification of relevant papers on a meta-level.

Information overload?

A concern of some researchers might be how to deal with all the studies that are bound to result from searching an exhaustive coordinate database. Isn't it possible that researchers will avoid using the database once they realize that this entails sorting through hundreds of papers for each coordinate? We fully acknowledge that searches will inevitably result in a lot of information that needs to be digested, but we believe that this is not a good reason to ignore the problems we have outlined. This information already exists and remains relevant to our own studies. Currently, we ignore this information because it is easy to do so. If we want to understand the function of a brain area, however, we need to look at all of the available information. As scientists we have a responsibility to seek out all the information that informs our research and a universal database will help us to fulfill this obligation.

We also believe that with the large number of articles that will be included in the database, the developers of this index will inevitably find new ways to summarize this information and make it accessible to the user. As an example, the Brede database already incorporates an algorithm that identifies related functional volumes (Nielsen and Hansen, 2004), and can also present frequently mentioned words associated with the chosen coordinates. Meta-analysis techniques employing, for example, replicator dynamics to identify functional networks could further assist in organizing this information (Neumann et al., 2005).

How should we proceed from here?

In our opinion, a universal coordinate database such as the one proposed here can only be successful if its utility is recognized by the neuroimaging community. For this reason, we believe that a survey that asks neuroimaging researchers their opinion on the utility of such a database and what sort of features they would find most useful would be advantageous. To this end we have created a survey to collect this information, so that future discussions of this issue may be informed by the community of neuroimagers. We would greatly appreciate it if readers completed this brief online questionnaire, and encouraged others to do the same. It can be found at http://www.tinyurl.com/db-survey (this URL redirects to Qualtrics.com where the survey is hosted). The results of this survey could help shape the development of a policy regarding database submission for neuroscience journals, or motivate article indexing companies to include this information in their own substantial databases.

Other initiatives that might prove helpful to initiate the process of developing a universal coordinate database include a symposium or open forum at major neuroimaging conferences. This could perhaps be followed by the formation of a task force to examine this issue more closely and generate a more complete set of possible solutions that can then be potentially voted upon.

Conclusion

We would like to conclude by emphasizing that creating a workable universal database would benefit everyone involved in neuroscience research. Authors and journals will benefit from an increased likelihood of an article being cited, researchers will gain access to a wealth of relevant information, and the field as a whole is likely to progress at a faster rate by encouraging scientists to look outside of their own paradigm.

Of all the possible solutions that we have explored above, it is unclear what the best way to proceed may be. In our minds, however, we favour a solution that involves current indexing corporation in consultation with working neuroscientists. These organizations already have so much of the information required in their own databases, as well as the resources to undertake a project such as this and the infrastructure to make it sustainable. Overall, a solution that involves these corporations appears to embody a number of positives (e.g., increased likelihood of older studies being indexed) with few negatives. As neuroscientists, however, we must not forget that it is our responsibility to facilitate a solution to this important problem, be it through advocacy or innovation.

Appendix A

The following search syntax was used to identify potentially relevant papers in PubMed:

("fMRI" OR "functional MRI" OR "functional magnetic resonance imaging") AND humans[MH] AND "magnetic resonance imaging"[MH] NOT review[PT] AND ("psychological phenomena and processes"[MH] OR "behavior and behavior mechanisms"[MH]) AND english[LA] AND<year>[DP].

Explanation of the search syntax

1. ("fMRI" OR "functional MRI" OR "functional magnetic resonance imaging"): searches for these terms in all fields (e.g., title, abstract, keywords).
2. AND humans[MH]: restricts search to studies with the medical subject heading (MeSH, http://www.nlm.nih.gov/mesh) "humans".
3. AND "magnetic resonance imaging"[MH]: this term was included in addition to the first search term to make sure that only fMRI studies are found; by this search term we excluded studies using other techniques (e.g., near-infrared spectroscopy, NIRS) that might say something like "Previous studies using fMRI have shown that… Here, we use NIRS to…" in the abstract.
4. NOT review[PT]: excludes articles of the publication type (PT) "review".
5. AND ("psychological phenomena and processes"[MH] OR "behavior and behavior mechanisms"[MH]): uses MeSHs to restrict hits to articles that study psychological phenomena or behavior; PubMed automatically expands the search to subordinate MeSHs (e.g., Psychological Phenomena and Processes → Mental Processes → Cognition).
6. AND english[LA]: only articles in English.
7. AND<year>[DP]: restricts search to a particular year.

The number of hits that resulted from the search using the above search syntax is shown in Table A1. To estimate the number of articles published in 2008, we fitted a polynomial function of degree 4 to the existing data. This resulted in an estimated 1502 articles for 2008. Adding this number to the 7908 hits from 1991 to 2007 resulted in 9410 articles estimated to be published until the end of 2008.

However, not all fMRI studies report coordinates in 3-D coordinate space. Some perform only ROI analyses or present pictures without listing coordinates. For this reason, it was necessary to estimate the percentage of papers that report Talairach coordinates. We chose the hits from the year 2007 and drew a random sample of 100 studies. (Online access was not available for three of the 100 papers originally chosen, so these were replaced by other randomly chosen papers.) Using this sample, we then examined whether each study reported coordinates; this was the case for 74 papers. This percentage was then used to estimate the fraction of papers reporting Talairach coordinates and is shown in Fig. 1 and Table A1.

To estimate the number of coordinates reported by the end of the year 2008, we took 74% of the 9410 papers (6963 papers) and multiplied this number with 37 (the mean number of coordinates per study in the BrainMap database). This resulted in 257,631 coordinates.

For Fig. 1, the databases were searched on a yearly basis, up until the end of 2007. For BrainMap and SumsDB/Caret, the search was limited to fMRI studies. The BrainMap search was performed on October 31st, 2008. The search of SumsDB/Caret was based on the October 2008 version of the stereotaxic foci database (archive: Human.PC-CC_STEREOTAXIC_FOCI_COMPOSITE_31K_ASSIGNED_Oct08.73730.spec, downloaded from http://sumsdb.wustl.edu/sums/directory.do?id=6529195). The search was performed with the search option "Data type = fMRI," limiting search results to fMRI studies. Numbers for AMAT and Brede include both PET and fMRI studies as these databases do not offer an option that restricts search to a particular imaging modality.

The estimate of BrainMap containing about 19% of the total number of fMRI articles published by the end of 2007 is based on the numbers shown in the two rightmost columns of Table A1.

Table A1
Number of articles found in PubMed, estimated percentage of these articles reporting Talairach coordinates, and number and percentage of fMRI studies in BrainMap.

Year	# of hits	74% of hits	# of studies in BrainMap	% of studies in BrainMap
1991–1995	57	42	3	7
1996	57	42	6	14
1997	85	63	10	16
1998	183	135	33	24
1999	263	195	51	26
2000	379	280	86	31
2001	497	368	113	31
2002	573	424	117	28
2003	770	570	142	25
2004	964	713	122	17
2005	1245	921	136	15
2006	1369	1013	119	12
2007	1466	1085	164	15
Total	7908	5852	1102	19

References

Brass, M., Derrfuss, J., Forstmann, B., von Cramon, D.Y., 2005. The role of the inferior frontal junction area in cognitive control. Trends Cogn. Sci. 9, 314–316.
Buckner, R.L., Carroll, D.C., 2007. Self-projection and the brain. Trends Cogn. Sci. 11, 49–57.
Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129, 564–583.
Chein, J.M., Fissell, K., Jacobs, S., Fiez, J.A., 2002. Functional heterogeneity within Broca's area during verbal working memory. Physiol. Behav. 77, 635–639.
Derrfuss, J., Brass, M., von Cramon, D.Y., 2004. Cognitive control in the posterior frontolateral cortex: evidence from common activations in task coordination, interference control, and working memory. NeuroImage 23, 604–612.
Derrfuss, J., Brass, M., Neumann, J., von Cramon, D.Y., 2005. Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. Hum. Brain Mapp. 25, 22–34.
Derrfuss, J., Brass, M., von Cramon, D.Y., Lohmann, G., Amunts, K., 2009. Neural activations at the junction of the inferior frontal sulcus and the inferior precentral sulcus: interindividual variability, reliability, and association with sulcal morphology. Hum. Brain Mapp. 30, 299–311.
Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L., Peters, T.M., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Proc. IEEE-Nuclear Sci. Symp. Med. Imaging Conf. 1813–1817.
Fox, P.T., Lancaster, J.L., 1994. Neuroscience on the net. Science 266, 994–996.
Fox, P.T., Lancaster, J.L., 2002. Opinion: mapping context and content: the BrainMap model. Nat. Rev., Neurosci. 3, 319–321.
Fox, P.T., Mikiten, S., Davis, G., Lancaster, J.L., 1994. BrainMap: a database of human functional brain mapping. In: Thatcher, R.W., Hallett, M., Zeffiro, T., John, E.R., Huerta, M. (Eds.), Functional Neuroimaging. Academic Press, San Diego, pp. 95–105.
Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., Koenig, S.F., Lancaster, J.L., 2005. BrainMap taxonomy of experimental design: description and evaluation. Hum. Brain Mapp. 25, 185–198.
Hamilton, A.F., 2005. Retrieved October 31, 2008, from http://www.antoniahamilton.com/amat.html.
Hassabis, D., Maguire, E.A., 2007. Deconstructing episodic memory with construction. Trends Cogn. Sci. 11, 299–306.
Hein, G., Knight, R.T., 2008. Superior temporal sulcus—it's my area: or is it? J. Cogn. Neurosci. 20, 1–12.
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.
Lancaster, J.L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum. Brain Mapp. 28, 1194–1205.

Lüders, E., Steinmetz, H., Jäncke, L., 2002. Brain size and grey matter volume in the healthy human brain. NeuroReport 13, 2371–2374.
Mason, M.F., Norton, M.I., Van Horn, J.D., Wegner, D.M., Grafton, S.T., Macrae, C.N., 2007. Wandering minds: the default network and stimulus-independent thought. Science 315, 393–395.
Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R. Soc. Lond., B Biol. Sci. 356, 1293–1322.
Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex 18, 262–271.
Neumann, J., Lohmann, G., Derrfuss, J., von Cramon, D.Y., 2005. Meta-analysis of functional imaging data using replicator dynamics. Hum. Brain Mapp. 25, 165–173.
Nielsen, F.Å., 2003. The Brede database: a small database for functional neuroimaging. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY, NeuroImage, 19 (2), Elsevier. Available on CD-Rom.
Nielsen, F.Å., Hansen, L.K., 2004. Finding related functional neuroimaging volumes. Artif. Intell. Med. 30, 141–151.
Nielsen, F.Å., Baslev, D., Hansen, L.K., 2005. Mining the posterior cingulate: segregation between memory and pain components. NeuroImage 27, 520–532.
Raichle, M.E., Macleod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl Acad. Sci. U.S.A. 98, 676–682.
Schubotz, R.I., Fiebach, C.J., 2006. Integrative models of Broca's area and the ventral premotor cortex. Cortex 42, 461–463.
Spreng, R.N., Mar, R.A., Kim, A.S., in press. The common neural basis of autobiographical memory, prospection, navigation, theory of mind and the default mode: a quantitative meta-analysis. J. Cogn. Neurosci.
Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Georg Thieme Verlag, Stuttgart.
Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. NeuroImage 16, 765–780.
Van Essen, D., Reid, E., Harwell, J., Gu, P., 2008. Search foci using Caret: mining neuroimaging data reported in stereotaxic coordinates. Retrieved December 31, 2008, from http://sumsdb.wustl.edu/sums/directory.do?id=6617378 and dir_name=STEREOTAXIC_FOCI.
Wager, T.D., Smith, E.E., 2003. Neuroimaging studies of working memory: a meta-analysis. Cogn. Affect. Behav. Neurosci. 3, 255–274.

NeuroImage 48 (2009) 8–10

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 12 May 2009

ABSTRACT

Commentaries by Derrfuss and Mar (Derrfuss, J., Mar, R., 2009. Lost in localization: the need for a universal coordinate database. Neuroimage.) and Nielsen (Nielsen, F.A., 2009. 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.

Recent commentaries by Derrfuss and Mar (2009) and Nielsen (2009) 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) 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) 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.

References

Derrfuss, J., Mar, R.A., 2009. Lost in localization: The need for a universal coordinate database. Neuroimage 48, 1–7.
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.Å., 2009. Lost in localization: A solution with neuroinformatics 2.0? Neuroimage 48, 11–13.
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.

NeuroImage 48 (2009) 11–13

Contents lists available at ScienceDirect

NeuroImage

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

Comments and Controversies

Lost in localization: A solution with neuroinformatics 2.0?

Finn Årup Nielsen *

Center for Integrated Molecular Brain Imaging, Copenhagen, Denmark
DTU Informatics, Technical University of Denmark, Lyngby, Denmark
Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 3 March 2009
Revised 19 May 2009
Accepted 26 May 2009
Available online 2 June 2009

ABSTRACT

The commentary by Derrfuss and Mar (Derrfuss, J., Mar, R.A., 2009. Lost in localization: The need for a universal coordinate database. NeuroImage, doi:10.1016/j.neuroimage.2009.01.053.) discusses some of the limitations of the present databases and calls for a universal coordinate database. Here I discuss further issues and propose another angle to the solution of a universal coordinate database with the use of wiki technology.

The commentary by Derrfuss and Mar (2009) in NeuroImage shows how far the present coordinate databases lag behind the published literature and the two authors call for a universal coordinate database. For the studies of methodologies for meta-analysis we have undertaken a universal coordinate database has not been necessary. On the other hand, for neuroscience research a universal coordinate database will greatly help to identify relevant studies, in the execution of meta-analyses and to obtain more unbiased interpretations of the prior literature. Several databases exist for coordinates and each has different advantages from 'minimal' approaches to efforts with a high degree of annotation (Hamilton, in press; Van Essen, 2009; Laird and Fox, 2009). Since 1998 we have gained some experience in the area by working both with the BrainMap and the Brede coordinate databases, and apart from the issues mentioned by Derrfuss and Mar I would like to raise three additional concerns: ownership, extensibility and community involvement.

One key obstacle identified by Derrfuss and Mar is how new studies can be submitted and included in the database. For some time limitations of the Brede Database have been apparent to us, and recently we have explored wiki-solutions to counter some of them setting up the MediaWiki-based *Brede Wiki*: http://neuro.imm.dtu.dk/wiki/. (MediaWiki is the software that also runs Wikipedia). A wiki presents an open interface for anyone to edit and read where the entered data becomes immediately available. It is even possible to built scripts that automatically add information to a wiki. One interesting project in this domain is *Gene Wiki* where researchers populate Wikipedia with genetics information (Huss et al., 2008). For the Brede Wiki a simple Matlab script allows for formatting coordinate lists from the SPM program in the style suitable for inclusion in the Brede Wiki. Another issue pointed to by the two authors arises when the large amount of information needed to be entered forms a contribution barrier. The user needs to learn all aspects of the data entry and go through a time-consuming submission process before the user can contribute. As an important element in recruiting new users a wiki system typically has a low contribution barrier (Bryant et al., 2005)—e.g., it only takes two mouse clicks and a keypress to correct a comma error. In a wiki the data entry task can be broken down as information can be entered incrementally by different editors, e.g., one may start with core information like bibliographic information and coordinates and then later on add, e.g., subject information and imaging modality.

The MediaWiki software has built-in text search facilities and means for categorizing pages. However, it lacks more complex means of query. Recent research effort has gone into semantic or fielded wikis that represent information with types, and the so-called templates of MediaWiki enable this. These systems may build taxonomies and ontologies such that computer programs can recognize, e.g., that 'happiness' is an 'emotion' or that a specific number is an *x*-coordinate rather than a Brodmann area or a page number. Examples of semantic wikis are SNPedia (http://snpedia.com) databasing genetic variations and NeuroLex (http://neurolex.org/) that organizes a neuroscience lexicon. Once data is entered within templates tools can extract it. One large-scale effort is DBpedia (http://dbpedia.org/) that extracts data from Wikipedia and presents on-line services for queries on the structured data (Auer et al., 2008). We have also been able to extract the structured data from Wikipedia and perform statistical analysis (Nielsen, 2007), and for the Brede Wiki we extract the templates and build an SQL database, that is used for searching after nearby coordinates to a given query. More complex queries can be formed such as 'Find all fMRI papers published between 2003 and 2008 with more than 12 subjects and with coordinates appearing less than 5 mm from (−40, 0, 30)'. However, the completeness of the result depends on whether the papers are completely described: If the subject information is not (yet) entered then it cannot be searched, and if the experimental conditions are not defined for the paper it is not possible to search this particular data. MediaWiki templates and categories are defined by the editors rather than the wiki administrator, so this

* DTU Informatics, Richard Petersens Plads, Building 321, DK-2800 Lyngby, Denmark. Fax: +45 45 87 25 99.
E-mail address: fn@imm.dtu.dk.

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2009.05.073

system has inherent extensibility. It is relatively easy to define new templates for, e.g., neuroimaging studies that report their results with respect to brain regions rather than peak coordinates. The MediaWiki software may dump its complete content to an XML file and an application programming interface enables on-line queries. Raw dumps of the Brede Wiki as well as an SQL file with information from its templates are available on the Web. Wikis do typically not support visualization directly. However, extensions to MediaWiki may enable the generation of visualizations, see e.g., the generation of graphs by Dengler et al. (2009). At present, the Brede Wiki calls external Web services, such as ICBM View, for the visualization of individual coordinates.

A recent dispute about data presented by Shmuel and Leopold (2008) has brought forth the issue of ownership to primary neuroimaging data within-laboratory (Fox et al., 2009) but disputes with neuroimaging data sharing between laboratories have a longer history (Aldhous, 2000). The issue of ownership of analysis results and meta-data, such as bibliographic information, would also be present for a coordinate database. Coordinates themselves are probably not covered by copyright since they do not reach the threshold of originality: Two researchers working independently would get the same set of coordinates given same subject data and the same analysis method. Bibliographic data may be copyrightable. Abstracts can probably be copied as research-based fair use, but they can probably not be copied without permission in commercial contexts, see also the discussion by Dunckley (2009). Other bibliographic data, such as year of publication and page number, can hardly be copyrightable *per se*, however when presented in a database it might gain database rights—at least in certain jurisdictions. There is no international agreement on the issue of database rights (Sanders, 2006). A 1996 European Union directive protects the maker of a database where the creation constitutes a 'substantial investment' so that a 'substantial part' of the database content cannot be copied without permission. This also covers data that is not copyrightable on its own. In Germany a court ruled that an alphabetic list of just 251 weblinks was protected, and in Denmark a court barred a Web service from deep linking and systematic copying of headlines from news web-sites (Mercado-Kierkegaard, 2006). On the other hand the United States Supreme Court accepted—in the so-called Feist case—that telephone subscriber information could be copied (Sanders, 2006). U.S. National Library of Medicine claims ownership and imposes restrictions on the use of PubMed and may terminate its license (U.S. National Library of Medicine, 2008). Since the NLM license may be terminated NLM data can probably not be merged with data from sources published under non-revokable Creative Commons licenses (http://creativecommons.org). The Brede Database has not yet a formal license but the entire database is available from its homepage for others to include and indeed Hamilton's AMAT database (Hamilton, in press) has added content from the Brede Database. There has been a tradition in the functional neuroimaging community for relatively open sharing, e.g., of software. The most popular analysis software, SPM, uses a so-called copyleft license. This kind of license encourages mutual sharing and numerous SPM extensions have been written by third-parties. In a database context a copyleft license will ensure that users share their version of and additions to the database if they copy it. This is the notion of share-alike. The only major difference between copyleft and the Creative Commons notion of share-alike is that share-alike may prohibit commercial reuse (the CC-by-nc-sa license) while copyleft licenses always allow it (corresponding to by-sa Creative Commons license). Properly copylefted databases may aid database federation since data can move freely between databases provided that they are all copylefted. A private company will not likely use a copyleft license, and users will be left with the search interface that the company can provide, since the data cannot be freely delivered through a third-party search engine. It would be an unfortunate development if neuroimaging result data is hidden behind subscription fees and restrictive licenses, and it will seriously impede the development of novel retrieval and analysis methods. We have setup part of the Brede Database with information from PubMed and included abstracts copyrighted by publishers or authors, so it is questionable if we can put the entire database under a copyleft license. As we have entered the coordinates ourselves the database rights for this part belongs to us and we can issue copyleft licenses for that part. When Hamilton uses coordinate data from Brede Database in her AMAT database and publishes AMAT under a copyleft license her reuse of our database is exactly as we intended. Copyleft and other open licenses have targeted the development of software code and text. Recently, a license has been drafted specifically for databases (http://www.opendatacommons.org/licenses/odbl/) (Miller et al., 2008), and this will be a suitable license for the Brede Wiki.

Wikipedia has shown a tremendous growth. It will be optimistic to think that a specialized wiki such as the Brede Wiki will experience similar growth. However, editing in the Brede Wiki is a big leap forward in comparison with the tedious submission procedure for the Brede Database, which requires downloading a Matlab program, understanding all the data fields, entering the data, storing an XML file and submitting this file.

As neuroscience data are complex it is not entirely clear how neuroinformatics databases should be structured. The original BrainMap database structure as described by Fox et al. (1994) inspired the design of our Brede Database. In many cases the BrainMap framework is sufficient and provides solid performance for standard neuroimaging meta-analysis (Fox et al., 2005). Yet extensions to the framework with the definition of an ontology for brain functions within the Brede Database allowed for automatic coordinate-based mass meta-analysis for all brain functions listed in the ontology (Nielsen, 2005). A brain region ontology allowed us to data mine across all brain regions (Nielsen et al., 2006) as well as link to the CoCoMac database (Kötter, 2004). This latter ontology was not anticipated during our initial design of the Brede Database but was a later addition. The *extensibility* of the database makes such data mining efforts possible.

Mandatory submission could be a goal, but in my opinion mandatory submission should not be undertaken before a system has been setup that is sufficiently easy to use and where the data can move freely between different databases. The Journal of Cogntive Neuroscience at one point required the submission of imaging data to the fMRI Data Center for studies published in the journal. This is no longer required, and it provides an illustration of the issues with mandatory submission of neuroscience data.

Wikipedia has shown how powerful commons-based peer production can be when Web-based technology enables it. Furthermore, collaborative information aggregation and collective prediction can be quite effective and in certain cases better than that of experts: We have experienced it in such diverse cases as humans in a web game and an ensemble of mathematical models (Pennock et al., 2001; Hansen et al., 2001). The neuroimaging community should embrace the notions of open knowledge and collective intelligence for community involvement in managing neuroinformatics data.

Acknowledgments

Thanks to the Lundbeck Foundation for funding and Lars Kai Hansen and Daniela Balslev for comments.

References

Aldhous, P., 2000. Prospect of data sharing gives brain mappers a headache. Nature 406, 445–446.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z., 2008. DBpedia: a nucleus for a web of open data. The Semantic Web, Volume 4825 of Lecture Notes in Computer Science. Springer, Heidelberg/Berlin, pp. 722–735.

Bryant, S.L., Forte, A., Bruckman, A., 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work. ACM, New York, NY, USA, pp. 1–10.

Dengler, F., Lamparter, S., Hefke, M., Abecker, A., 2009. Collaborative process development using Semantic MediaWiki. In: Hinkelmann, K., Wache, H. (Eds.), WM2009: 5th Conference of Professional Knowledge Management, Volume P-145 of Lecture Notes in Informatics. Bonner Kollen Verlag, Solothurn, Switzerland, pp. 97–107.

Derrfuss, J., Mar, R.A., 2009. Lost in localization: the need for a universal coordinate database. NeuroImage 48, 1–7.

Dunckley, J., 2009. Why you can't copy abstracts into Wikipedia. http://www.cotch.net/blog/20090126_1005. Accessed 6 May 2009.

Fox, P.T., Mikiten, S., Davis, G., Lancaster, J.L., 1994. BrainMap: a database of human function brain mapping. In: Thatcher, R.W., Hallett, M., Zeffiro, T., John, E.R., Huerta, M. (Eds.), Functional Neuroimaging. Technical Foundations. Academic Press, San Diego, California, pp. 95–105.

Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., Koenig, S.F., Lancaster, J.L., 2005. BrainMap taxonomy of experimental design: description and evaluation. Hum. Brain Mapp. 25, 185–198.

Fox, P.T., Bullmore, E., Bandettini, P.A., Lancaster, J.L., 2009. Protecting peer review: correspondence chronology and ethical analysis regarding Logothesis vs. Schmuel and Leopold. Hum. Brain Mapp. 30, 347–354.

Hamilton, A., in press. Lost in localization: a minimal middle way. NeuroImage. doi:10.1016/j.neuroimage.2009.05.007.

Hansen, L.K., Nielsen, F.Å., Strother, S.C., Lange, N., 2001. Consensus inference in neuroimaging. NeuroImage 13, 1212–1218.

Huss III, J.W., Orozco, C., Goodale, J., Chunlei, Batalov, S., Vickers, T.J., Valafar, F., Su, A.I., 2008. A gene wiki for community annotation of gene function. PLoS Biol. 6, e175.

Kötter, R., 2004. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2, 127–144.

Laird, A.R., Lancaster, J.L., Fox, P.T., 2009. Lost in localization? The focus is meta-analysis. NeuroImage 48, 18–20.

Mercado-Kierkegaard, S., 2006. Clearing the legal barriers — Danish court upholds 'deep linking' in Home v. Ofir. Comput. Law Secur. Rep. 22, 326–332.

Miller, P., Styles, R., Heath, T., 2008. Open data commons, a license for open data. In: Bizer, C., Heath, T., Idehen, K., Berners-Lee, T. (Eds.), Proceedings of the WWW2008 Workshop on Linked Data on the Web, Volume 369 of CEUR Workshop Proceedings. RWTH Aachen University, Aachen, Germany.

Nielsen, F.Å., 2005. Mass meta-analysis in Talairach space. In: Saul, L.K., Weiss, Y., Bottou, L. (Eds.), Advances in Neural Information Processing Systems 17. MIT Press, Cambridge, MA, pp. 985–992.

Nielsen, F.Å., 2007. Scientific citations in Wikipedia. First Monday 12.

Nielsen, F.Å., Balslev, D., Hansen, L.K., 2006. Data mining a functional neuroimaging database for functional segregation in brain regions. In: Olsen, S.I. (Ed.), Den 15. Danske Konference i Mønstergenkendelse og Billedanalyse. The Department of Computer Science. University of Copenhagen, Copenhagen, Denmark.

Pennock, D.M., Lawrence, S., Nielsen, F.Å., Giles, C.L., 2001. Extracting collective probabilistic forecasts from web games. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. SIGKDD, Association for Computing Machinery. ACM, New York, NY, USA, pp. 174–183.

Sanders, A.K., 2006. Limits to database protection: fair use and scientific research exemptions. Res. Policy 35, 854–874.

Shmuel, A., Leopold, D.A., 2008. Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: implications for functional connectivity at rest. Hum. Brain Mapp. 29, 751–761.

U.S. National Library of Medicine, 2008. License agreement for NLM data. http://www.nlm.nih.gov/databases/license/license.pdf. Accessed 14 May 2009.

Van Essen, D.C., 2009. Lost in localization — but found with foci?! NeuroImage 48, 14–17.

NeuroImage 48 (2009) 14–17

Contents lists available at ScienceDirect

NeuroImage

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

Comments and Controversies

Lost in localization — But found with foci?!

David C. Van Essen *

Department of Anatomy & Neurobiology, Washington University in St. Louis, 660 S. Euclid Avenue, St. Louis, MO 63110, USA

ARTICLE INFO

Article history:
Received 20 April 2009
Revised 11 May 2009
Accepted 14 May 2009
Available online 27 May 2009

ABSTRACT

Commentaries by Derrfuss and Mar [Derrfuss, J., Mar, R.A., 2009. Lost in localization: the need for a universal coordinate database. Neuroimage (doi:10.1016/j.neuroimage.2009.01.053).], Nielsen [Nielsen, F.A., 2009. Lost in localization: a solution with neuroinformatics 2.0? Neuroimage.], Hamilton [Hamilton, A., 2009. Lost in localization: a minimal middle way. Neuroimage.], and Laird and Fox [Laird, A.R., Fox, P.T., 2009 Lost in localization? The focus is meta-analysis. Neuroimage.] agree on the need for a comprehensive database of published stereotaxic coordinates but offer diverse views on how best to achieve this objective. Here, I summarize recent enhancements to the SumsDB database that increase its utility and decrease the impediments to data submission, thereby making it attractive as a resource that can approach comprehensive content in a realistic time frame.

In their commentary "Lost in localization: the need for a universal coordinate database", Derrfuss and Mar (2009), argue cogently for a comprehensive database that would provide efficient access to the hundreds of thousands of stereotaxic coordinates that summarize key experimental findings in an estimated 10,000 neuroimaging studies. They noted that existing coordinate databases contain only a modest fraction of the relevant data and also that none (at that time) was matching the pace at which new coordinate data are being published. The core problems are that submitting coordinate data requires substantial time and effort and that the benefits from submitting such data have not inspired widespread voluntary participation by the neuroimaging community. The keys to alleviating this bottleneck are to reduce the effort entailed and to increase the benefits of data submission. This is precisely the objective of recent improvements to the SumsDB database (http://sumsdb.wustl.edu/sums/).

Key features of SumsDB

It is useful to summarize key features of SumsDB and the associated visualization software (Caret and WebCaret), especially since many enhancements were implemented after Derrfuss and Mar submitted their commentary. Features that make SumsDB useful for data mining of coordinates ('foci' in our terminology) fall into five main categories.

* Fax: +1 314 747 3436.
E-mail address: vanessen@brainvis.wustl.edu.

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.05.050

Flexible search options

The 'Quick-Search' repository in SumsDB (April, 2009) currently contains ~40,000 foci from ~1300 studies (Fig. 1A) and supports searches based on many types of metadata. These include:

- Spatial location (*x*, *y*, *z* coordinates; cortical sulcus, or cortical area)
- Functions or tasks specified in individual tables or other 'experiment-specific metadata'
- Information about the study as a whole (e.g., abstract, title, keywords).

This enables searches that address questions of the following types:

- "What is known about the function of region 'X', such as area MT+ (Fig. 1B)?"
- "What brain regions are involved in processing related to function 'Y', such as music (Fig. 1C) or task 'Z' (e.g., a pitch discrimination task)?"
- What brain regions show abnormalities in structure or function in a particular disease or disorder such as autism (Fig. 1D) or schizophrenia?

Each focus is associated with extensive metadata, immediately viewable by clicking on that focus (arrow in Fig. 1E). There are also direct links from each focus to PubMed and to the online article. Thus, while the initial search results typically include many foci that are irrelevant to the primary question posed, information that is close at hand allows screening of extraneous foci and selection of just the relevant foci.

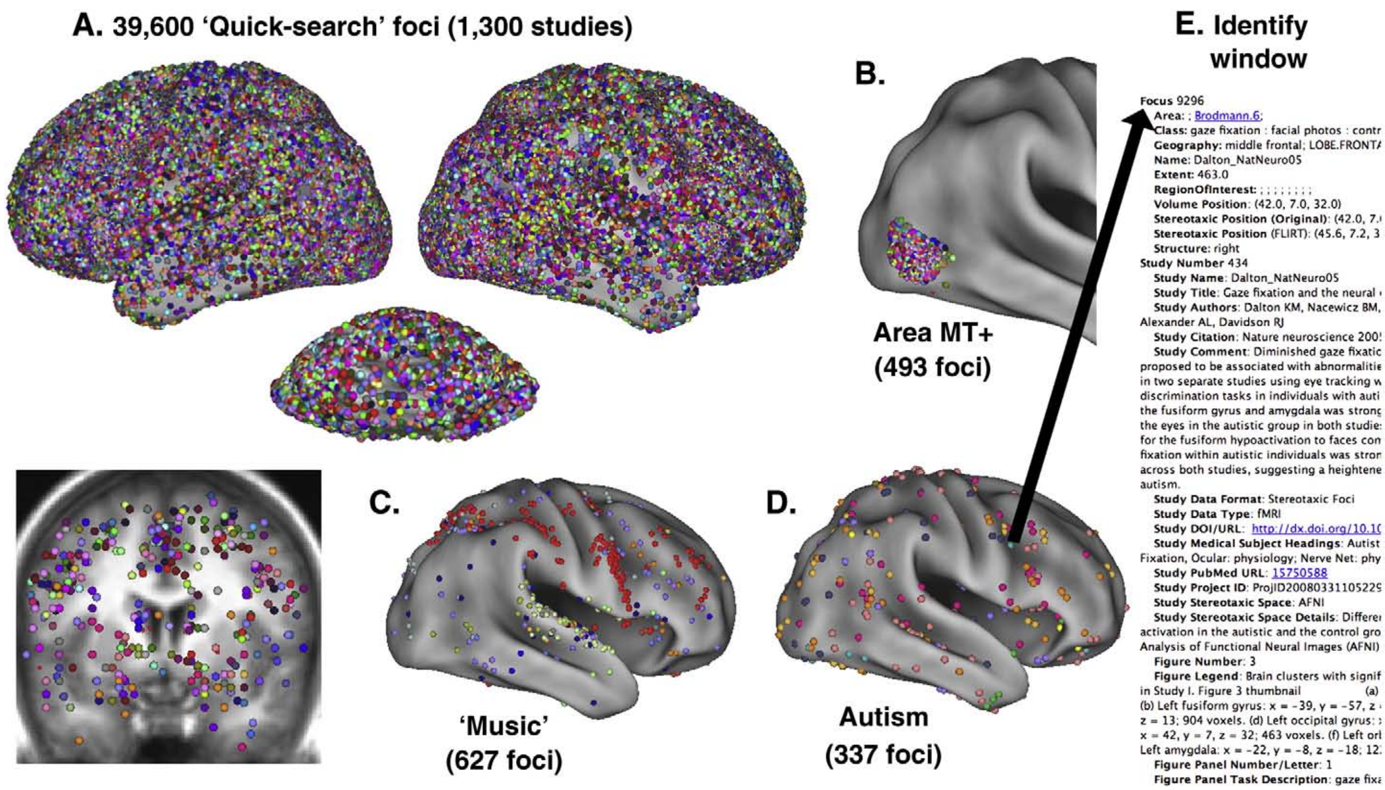

Fig. 1. (A) 39,600 foci in the SumsDB Quick-Search repository, projected onto atlas surfaces (inflated PALS-B12 cerebral cortex and Colin cerebellum) and a coronal slice through the atlas volume. Each focus is colored according to the study from which it originated. (B) 493 foci localized to area MT+, displayed on the PALS-B12 right hemisphere. (C) 627 foci associated with studies containing the term 'music' in the abstract, keywords, or other metadata. (D) 337 foci associated with studies referring to autism. (E) Extensive metadata from a particular focus related to autism reported in the WebCaret 'Identify window'.

Visualization and analysis

SumsDB is linked to online (WebCaret) and offline (Caret) software that enable visualization of search results on a human brain atlas. Important attributes of the atlas and the visualization software include:

- *Surfaces and volumes*. Search results can be viewed on surfaces and/or volumes, thereby capitalizing on the inherent complementarity of these formats (Fig. 1A). Surfaces include the population-average PALS-B12 atlas of cerebral cortex plus the 'Colin' individual atlas of cerebellar cortex (Van Essen and Dierker, 2007b). Volumes include the PALS-B12 average structural MRI.
- *Many stereotaxic spaces*. The PALS-B12 atlas surfaces and volumes are available in numerous stereotaxic spaces that are widely used in neuroimaging studies (AFNI, FLIRT, all SPM versions, and several others). Coordinates originally reported in a diversity of spaces can be automatically projected to any desired target space while preserving access to the original reported coordinates (Van Essen and Dierker, 2007a). The ability to switch easily from one space to another facilitates accurate comparisons with fMRI activations and other volumetric data reported in different stereotaxic spaces.
- *Reference maps*. Extensive reference data aid in localizing coordinates relative to cortical areas (defined by any of several partitioning schemes), cortical sulci, cerebellar lobules, and various other regions of interest. Reference maps can be updated as additional information becomes available from newly published studies.
- *Online and offline data mining*. WebCaret software allows SumsDB search results to be viewed immediately, without downloading software or data. Specific search results (or the entire Quick-Search Repository) can be downloaded for offline analysis using Caret. This facilitates comparisons with other studies and with experimental data from the investigator's own lab, whether these are viewed as additional foci or as data mapped onto surfaces or volumes.

Links to online publications

Rapid access to the original publication is frequently important, in order to assess the relevance of various foci to whatever question the investigator has in mind or to critically evaluate exactly what experiments were done and how they were analyzed and interpreted. To facilitate this process, SumsDB, WebCaret, and Caret all provide links from specific search results to the corresponding online publications (via PubMed, and also directly when possible). For most foci, the table, figure, or page number is specified, thereby allowing the relevant section of the study to be found quickly.

Study collections and meta-analyses

SumsDB contains a 'Study Collection Library', in which each study collection points to a thematically unified group of studies and associated foci. Some study collections represent a formal meta-analysis linked to a published review or research article on that topic, such as 'deception' (Christ et al., 2008). Others represent informal, non-comprehensive collections that provide useful entrees to various topics of interest (e.g., 'face perception'). Study collections can be kept private or made public, and they can be easily modified (e.g., to add newly identified studies related to the study collection topic).

Federation with other resources

The Neuroscience Information Framework portal (NIF, http://neuinfo.org/) allows efficient searching of a wide range of neuroscience-related resources available on the web. For SumsDB and many other resources, NIF supports ‘deep’ data mining, wherein relevant database contents (not just the home page) can be directly accessed by queries initiated in NIF. For SumsDB, NIF-initiated queries report specific search results and in addition allow users to immediately link out and view the results using WebCaret.

Efficient data submission

Data submission to SumsDB entails entering three types of data into two curated ‘libraries’.

- ‘Core’ metadata for each study are entered into the Master Study Library and are mainly extracted automatically from PubMed.
  This includes the authors, title, citation, abstract, stereotaxic space, species, and data type (e.g., fMRI or PET).
- ‘Experiment-specific’ metadata, generally including succinct task characterizations extracted from published tables (and their subheaders), figures and page references, are also entered into the Master Study Library.
- The *x*, *y*, *z* coordinates of each focus are entered into the Foci Library, along with ‘focus-specific’ metadata (e.g., cortical area or sulcus) and the table (and subheader), figure, or page number.
  Besides the assignments extracted from the original publication, a standard set of assignments derived from reference maps on the PALS atlas is added by a post-submission process carried out by the curators.

For any given data entry, the Master Study Library and Foci Library support multiple versions that differ in their metadata content, providing useful flexibility and updating capabilities. The Quick-Search repository used to expedite routine searches is a distillation that includes a single entry for each focus and each study.

Tutorial and instruction documents (accessible via ‘Foci Data Mining’ on the SumsDB home page) show how to enter coordinate-related data into SumsDB. Training takes 5–10 h, depending on initial familiarity with Caret software. After training, data submission typically takes 30–60 min per study. This is only modestly slower than the ~20 min needed for the AMAT database with its ‘minimal’ metadata requirements (Hamilton, 2009) and is much faster than the extensive task characterizations required by the BrainMap database (Fox et al., 2005; Laird and Fox, 2009). Thus, SumsDB provides an important middle ground, with large value added for a modest data entry effort.

Data submission to SumsDB offers multiple benefits to the submitter:

- submitting foci from publications of your own lab will increase their visibility, through data mining initiated in SumsDB or NIF;
- submitting relevant studies from your research subfield will facilitate cross-study comparisons and promote broader awareness of research in that area;
- individual contributors are recognized by ‘provenance’ assignments for each study (or version) entered into SumsDB.
- SumsDB libraries can also be used to store foci and study collections for ongoing projects that are not yet published. (Data in these libraries are not made public until requested by the submitter and then vetted by a curator in the Van Essen lab to insure conformance to basic metadata description standards.)

Nielsen (2009) has proposed a wiki-based approach to submitting coordinates and metadata. As noted by Hamilton (2009) and Laird and Fox (2009), this approach may encounter difficulties if it lacks an enforced, coherent metadata structure. Indeed, our experience with SumsDB is that a robust and carefully designed infrastructure is necessary for dealing with various technical complexities. These include avoiding unwanted duplication of foci and studies already in the database while allowing multiple versions of foci and studies when they differ meaningfully in metadata content (e.g., in the description of behavioral tasks). Providing for these and other important needs adds a modest overhead to the data submission process, but yields major benefits in making this a user-friendly resource.

Scaling up and catching up — volunteers needed!

SumsDB libraries have nearly tripled in content over the last 16 months, from 14,000 (~500 studies) in January 2008 to ~40,000 foci (~1300 studies) in April 2009. This approaches the rate at which new studies reporting coordinates are published and makes SumsDB the fastest-growing of the existing coordinate databases. We anticipate being able to sustain this pace through ongoing curation efforts in the Van Essen lab. However, to accelerate the process and substantially reduce the large backlog, it is vital to enlist volunteers from the neuroimaging community. An attractive and feasible model is for one or two individuals (students, postdocs, or knowledgeable technicians) from each of many laboratories to enter data published by their own laboratory plus selected topics related to that lab's research interests. For example, if 50 volunteers each added ~20 studies per year (15–30 h per volunteer, including training), the current rate of submission would approximately double, and about half of the relevant literature would be covered in ~5 years.

Psychology and neuroscience courses offer another way to promote data submission. For example, a classroom project to explore a specific aspect of brain function might include analysis of existing studies in SumsDB plus addition of relevant studies from the literature that are not yet in SumsDB. Bearing in mind that our central curation process prevents flawed or invalid entries from becoming public, such efforts can be undertaken by graduate students or even undergraduates in a supervised instructional setting.

Datasets underlying published meta-analyses constitute a particularly attractive source for database submissions. A PubMed search for ‘neuroimaging AND meta-analysis’ reports 180 meta-analyses, half of which were published since 2006. Most of these meta-analyses (based on inspection of the 20 most recent ones) involved extraction of coordinate data directly from the literature, not from any database. Instructions for handling existing meta-analyses are included in the documentation for submitting coordinate-related data into SumsDB, so as to capitalize on the work already done in extracting coordinate data and to generate a study collection that can be linked to the published meta-analysis.

Looking around the bend

Several developments could further accelerate the pace at which coordinate data are made accessible and useful to the community.

Accelerating foci submission

Greater sharing of data across existing coordinate databases would reduce duplication of effort. Already, SumsDB contains over 5000 foci (228 studies) extracted from the AMAT database (with permission from and acknowledgment to A. Hamilton); similarly, AMAT includes many entries initially entered in the BREDE database (Nielsen, 2003). Data from the SumsDB Foci Library and Master Study Library are freely available for data mining or incorporation into other databases, with the expectation that usage of SumsDB will be appropriately acknowledged. Open sharing of data would allow each database to capitalize on its unique capabilities. The two largest databases (Brain Map and SumsDB) each have extensive visualization and analysis capabilities that differ greatly, making these resources more complementary than competing. Volunteers would presumably be more willing to contribute coordinate data to their preferred database if they anticipate that the data will soon populate other databases and thus be of broader utility. SumsDB allows credit to be allocated to the database of origin as well as to individual contributors, thus sharing the credit appropriately. Because databases differ in data format and metadata content, sharing can be expedited by providing a schema that characterizes the database structure, as we learned through the process of federating SumsDB with NIF. More generally, the recommendation that the neuroimaging community should embrace open knowledge for data mining (Nielsen, 2009) is timely and meritorious.

Greater standardization of how results are tabulated and reported in journal articles could increase the efficiency of extracting coordinate data and metadata. This can begin with modest, incremental steps, though the ultimate objective should be to enable automatic or semi-automatic data extraction. However, without buy-in from the neuroimaging community, journal editors are unlikely to impose steps that might be even modestly burdensome to authors. On the other hand, systematization of how coordinate data are reported and described would benefit authors and journal readers alike (Poldrack et al., 2008 — see 'Figures and tables should stand on their own' section). In short, the potential benefits of data standardization are large, but the timing and the process for building community support need careful attention. Organizations such as the Organization for Human Brain Mapping, and Society for Neuroscience, and the International Neuroinformatics Coordinating Facility could help catalyze this process in response to inputs from their membership.

The iceberg beneath the tip

The conciseness of x, y, z coordinate data is both a strength and a limitation of this data format. Obviously, much greater information about complex spatial and temporal patterns of brain activation is available in the volume and surface data from which foci are extracted. In an explicit comparison of approaches, Salimi-Khorshidi et al. (2009) demonstrated that image-based meta-analyses (IBMA) are more sensitive than coordinate-based meta-analyses (CBMA) in revealing patterns of activation that are consistent across studies. For IBMA to become widely used for meta-analyses, a searchable database repository for image (volume) data from a large number of studies would be extremely useful. SumsDB can already handle volume as well as surface data and thus provides an existing option for this purpose. Indeed, a growing number of studies directly link to datasets in SumsDB, providing access to the underlying data and enabling WebCaret visualization of scenes that replicate the published figures (see 'Publications with links to SumsDB' on the home page). To facilitate efficient data mining, we intend to establish a Surface Library and a Volume Library in SumsDB, analogous to the existing Foci Library, that will house curated sets of published surface and volume data and associated metadata.

In conclusion, it is useful to place this and other emerging neuroimaging-related efforts in neuroinformatics into perspective relative to the burgeoning field of bioinformatics. Powerful data mining tools for analyzing genome and protein sequence data have dramatically advanced our understanding of genetics and molecular biology over the past two decades. Progress has been slower in making neuroinformatics approaches useful to the neuroscience community for a variety of technical and sociological reasons (Gardner et al., 2008). Coordinate data from human neuroimaging studies represent 'low-hanging fruit', now ripe for exploitation using increasingly powerful neuroinformatics tools that can accelerate progress in elucidating brain function in health and disease.

Acknowledgments

I thank Erin Reid for yeoman's work in data entry and curation, John Harwell and Ping Gu for excellent software design and development, Jessica Cohen, Jeff Phillips, Shawn Christ, and Donna Dierker for comments, and Susan Danker for help in manuscript preparation. Supported by NIH Grant R01-MH-60974, funded by the National Institute of Mental Health, the National Institute for Biomedical Imaging and Bioengineering, and the National Science Foundation, and by the Neuroscience Information Framework NIH Subcontract.

References

Christ, S.E., Van Essen, D.C., Watson, J.M., Brubaker, L.E., McDermott, K.B., 2008. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analysis. Cerebral Cortex doi: 10.1093. electronic publication.

Derrfuss, J., Mar, R.A., 2009. Lost in localization: the need for a universal coordinate database. Neuroimage 48, 1–7.

Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., Koenig, S.F., Lancaster, J.L., 2005. BrainMap taxonomy of experimental design: description and evaluation. Hum. Brain Mapp. 185–198.

Gardner, D., Akil, H., Ascoli, G.A., Bowden, D.M., Bug, W., Donohue, D.E., Goldberg, D.H., Grafstein, B., Grethe, J.S., Gupta, A., Halavi, M., Kennedy, D.N., Marenco, L., Martone, M.E., Miller, P., Muller, H.-M., Robert, A., Shepherd, G.M., Sternberg, P.W., Van Essen, D.C., Williams, R.W., 2008. The neuroscience information framework: a data and knowledge environment for neuroscience. Neuroinformatics 6, 149–160.

Hamilton, A.F.d.c., 2009. Lost in localization: a minimal middle way. Neuroimage 48, 8–10.

Laird, A.R., Lancaster, J.L., Fox, P.T., 2009. Lost in localization? The focus is meta-analysis. Neuroimage 48, 18–20.

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

Nielsen, F.Å., 2009. Lost in localization: a solution with neuroinformatics 2.0? Neuroimage 48, 11–13.

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.

Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., Nichols, T.E., 2009. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. Neuroimage 45, 810–823.

Van Essen, D.C., Dierker, D., 2007a. On navigating the human cerebral cortex: response to 'in praise of tedious anatomy'. Neuroimage 37, 1050–1054 discussion 1066-1058. PMCID: PMC2045137.

Van Essen, D.C., Dierker, D.L., 2007b. Surface-based and probabilistic atlases of primate cerebral cortex. Neuron 56, 209–225.

NeuroImage 48 (2009) 18–20

Contents lists available at ScienceDirect

NeuroImage

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

Comments and Controversies

Lost in localization? The focus is meta-analysis

Angela R. Laird *, Jack L. Lancaster, Peter T. Fox

Research Imaging Center, University of Texas Health Science Center San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA

ARTICLE INFO

Article history:
Received 29 April 2009
Revised 4 June 2009
Accepted 16 June 2009
Available online 25 June 2009

ABSTRACT

The recent commentary by Derrfuss J, Mar RA. (2009). Lost in localization: the need for a universal coordinate database. Neuroimage, In Press proposed a universal coordinate database to archive functional neuroimaging results. In this response, we discuss our strategy in developing the BrainMap database, which was created as a mechanism to promote coordinate-based meta-analysis methods.

Recent commentaries by Derrfuss and Mar (2009), Hamilton (2009), Nielsen (2009), and Van Essen (2009) all advocated the creation of a universal coordinate database as a means to aggregate functional neuroimaging results published in standardized coordinate form, which are growing at a rapidly accelerating pace. The BrainMap database (Fox and Lancaster, 2002) currently contains 1729 papers, 8007 experiments, and 64,940 foci (as of April 2009); Derrfuss and Mar estimate this to be approximately 20% of the relevant literature, making it the largest coordinate-based database in functional neuroimaging to date. BrainMap is a freely accessible community database in which reported activations can be searched for within user-defined ROI boundaries, thus offering the opportunity to relate behavioral functions to specific brain locations.

Reconciling new results to those previously published can be overwhelming, particularly when the relevant studies pertain to different research domains. Derrfuss and Mar proposed that a coordinate database be used to comprehensively identify published studies reporting activation in a given brain region, so that researchers can compare papers reporting foci proximate to their own results. Given the extremely large amount of neuroimaging results that have been reported thus far, the BrainMap project has elected to focus on coordinate-based meta-analysis methods to synthesize this data and provide a means to ascribe a set of functions to a given set of brain regions (Fox et al., 2005a). Derrfuss and Mar calculated that approximately 330 coordinates have been reported in the literature for every single cubic centimeter of gray matter, which is an impressive statistic that conveys the enormity of the task of results summation. Without the aid of meta-analysis, users of a universal coordinate database who query regions of interest will be left with long lists of published studies, the contents of which must be manually filtered and interpreted. BrainMap's approach to investigating function–location correspondences has been to reduce this burden of labor by developing and promoting quantitative meta-analyses of peak coordinates and their associated meta-data. The BrainMap database offers the ability to not only retrieve studies returned by regional searches without domain-specific biases, but also provides the means to synthesize the search results into coherent brain networks using the *GingerALE* meta-analysis application (Laird et al., 2005a).

Archiving coordinates and meta-data

Coordinate databases offer an opportunity to localize brain activation from a number of different studies that employed a wide array of tasks. The easiest and most rapid path to achieving a comprehensive coordinate database is to archive only coordinates and citation information; however, the range of potential inferences to be made from this type of database is limited. Systematically establishing function–location associations requires that function must somehow be defined in relation to the archived coordinates. To accomplish this, meta-data for each focus must be extracted from the published studies. From 1992 to 1998, developers of the BrainMap database held a series of annual workshops in which leaders of the field debated the structure for a taxonomy of functional neuroimaging experiments. Much of the debate focused on determining the appropriate level of detail for what eventually evolved into the BrainMap coding scheme. These meta-data allow each coordinate to be linked with how the observed activation was experimentally derived, a formulation that lends itself to rich data mining possibilities. BrainMap's power to capture knowledge associated with function–location relationships is due to both the quantity and quality of meta-data that is archived. But the ability to perform complex analyses of coordinate data in BrainMap comes at the cost of manually extracting meta-data from each publication. Peer-reviewed

* Corresponding author. Fax: +1 210 567 8152.
E-mail address: lairda@uthscsa.edu (A.R. Laird).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2009.06.047

publications can be submitted to BrainMap by the original authors (uncommon) or by investigators performing a meta-analysis (very common); two BrainMap research assistants also enter data on a full-time basis. All entries are reviewed for quality control by BrainMap staff and faculty before being entered into the database to ensure the accuracy and consistency of coding.

In addition to citation information, the current BrainMap coding scheme contains meta-data descriptions on subjects, experimental conditions (stimulus, response, instructions), paradigms, and behavioral domains. Derrfuss and Mar suggest that a greater volume of the literature could be more effectively archived by a reduction of BrainMap's required submission fields. However, their recommended list of necessary core fields is nearly comprehensive to BrainMap's current structure. Reduction of the design and results of an entire neuroimaging experiment into a small set of meta-data fields is a complex neuroinformatical dilemma, with agreement rarely observed across investigators as to which are the truly critical components. We agree that BrainMap's data entry procedure can be time-consuming (Laird et al., 2005b). It takes a research assistant approximately 30–60 min to enter the details of a single publication into our data entry application, *Scribe*. However, we argue that the depth of the current coding strategy is what provides diverse data mining opportunities and hence increases the value of the database. Examination of published studies reveals that the BrainMap taxonomy performs well in matching to search filters applied by meta-analysis authors, thereby reducing the time needed for manual searches of the literature (Fox et al., 2005b). The current depth of the BrainMap coding scheme represents our instantiation of a compromise between a rapid data entry procedure and a sufficient level of meta-data to yield useful data mining results.

In the BrainMap taxonomy, a structured keyword system has been favored over free text entry to reduce redundancy due to alternative or competitive terminology. Only an ontology for functional neuroimaging experiments will fully prevent the loss of information associated with alternative vocabularies; however, a complete ontology does not currently exist in this domain despite increasing acknowledgement that one is necessary (Toga, 2002; Price and Friston, 2005; Poldrack, 2006; Binder et al., in press). Such an ontology would also enable classification of studies in BrainMap at a deeper level of detail. Poldrack (2006) argues that BrainMap's divisional structure for behavioral domains is too coarse and does not allow for experiments to be coded at a sufficient level of detail to enable meaningful structure–function mappings. Yet until an ontology of cognitive processes is developed and adopted by the functional neuroimaging community, the use of alternative terminologies will result in the dilution of concepts over many domains, thereby interfering with both the data submission and retrieval procedures in coordinate-based databases. At this time, it is unwarranted to assume that functional neuroimaging results will reveal an organizational structure of the human mind that conforms to theoretical cognitive models, such as those associated with cognitive architectures (Langley et al., 2009) or conceptual primitives (Mandler, 2004). Until a more finely detailed behavioral domain hierarchy is tested and validated against neuroimaging data, BrainMap will continue to utilize a broader approach to behavioral classifications. This domain structure was designed to group like studies, rather than segregate them based on an unproven classification system.

The need for meta-analytic tools

A comprehensive coordinate database would undoubtedly be a welcome addition to the neuroimaging community, as this would allow researchers to expedite their literature searches and streamline the collation of relevant coordinates. Although not comprehensive, BrainMap contains a significant percentage of the literature and "provides a broad enough sample of different studies to provide a useful proof of concept" (Poldrack, 2006). While we aim for a database that contains 100% of eligible studies, funding restrictions dictate that the BrainMap project pursue a joint strategy of both data archival and tool development. We are unwilling to redistribute our efforts to focus solely on data entry, as this would undermine our ability to develop and improve meta-analysis tools. In this way, we favor an ideology in which scientific contributions are valued over the convenience provided by a comprehensive database. For example, we recently implemented several modifications of the activation likelihood estimation (ALE) method (Eickhoff et al., in press), which is the coordinate-based meta-analysis method supported by BrainMap. ALE now includes: estimates of the between-subject and between-laboratory variability, to more explicitly model the spatial uncertainty associated; and weighs each study by the number of included subjects. The method of testing for statistical significance in ALE was also modified, resulting in a transition from fixed-effects to random-effects meta-analyses. In addition, we have developed and validated coordinate conversion algorithms that reduce the disparity between MNI and Talairach coordinates (Lancaster et al., 2007). These corrections for varying spatial normalization techniques affect both data retrieval results for regional queries of coordinates, as well as meta-analytic results since more accurate coordinate corrections result in tighter, more coherent nodes of concordance. BrainMap automatically applies these corrections to incoming database entries as part of the data submission process, a feature that directly resulted from our commitment to the development of meta-analysis methods and neuroinformatics tools.

It is our aim that this commitment will allow further extension of how these methods are applied, such that coordinate-based meta-analyses are not limited to the simple pooling of studies utilizing the same experimental task. We strive for a wider scope of applications in line with the BrainMap's intended goal of facilitating the creation of a functional brain atlas. BrainMap's search capabilities can support various types of queries, such as "for a given function, what regions are typically engaged?", "for a given region, what tasks elicit activation?", or "for a given region, what other regions are coactivated?". Using meta-data archived in BrainMap these correspondences (function-to-regions, region-to-tasks, or region-to-network) can be constructed in a data-driven manner. For example, BrainMap is capable of generating function-to-regions associations by creating whole-brain meta-analytic maps for each behavioral domain category, which can then be decomposed into sub-networks based on different levels of the domain hierarchy. Region-to-network correspondences can be constructed by analyzing which foci coactivate with coordinates located in an anatomically defined region of interest, as a meta-analytic analogue of functional connectivity studies (Koski and Paus, 2000; Postuma and Dagher, 2006; Toro et al., 2008). Applying high-level filters from the entire BrainMap coding scheme to either the domain or coactivation meta-maps may be an effective strategy for refining their spatial specificity. Thus, while paradigm class and behavioral domain have been established as important meta-data fields in the BrainMap coding scheme, other fields, such as stimulus modality and response type, also have the potential to assist in unraveling the brain's systems and their interactions.

In conclusion, we agree with Derrfuss and Mar in their discussion of the value offered by a universal coordinate database, but caution database developers in designing neuroinformatics tools with limited applicability. A large-scale archive of foci is only as useful as the corresponding meta-data that it contains. In our view, databases in functional neuroimaging have not yet fully realized their potential for knowledge discovery in mapping human brain function. The meta-analytic applications made possible through BrainMap will evolve and grow more powerful as development of the database continues, perhaps leading to a multi-layered probabilistic functional brain atlas of meaningful mappings between function and structure.

Acknowledgments

The authors would like to acknowledge support from the NIMH to PTF (R01-MH074457) and ARL (R01-MH084812).

References

Binder J.R., Desai R.H., Graves W.W., Conant L.L., in press. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex.

Derrfuss J., Mar R.A., 2009. Lost in localization: The need for a universal coordinate database. Neuroimage 48, 1–7.

Eickhoff S.B., Laird A.R., Grefkes C., Wang L.E., Zilles K., Fox P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp.

Fox, P.T., Lancaster, J.L., 2002. Mapping context and content: the BrainMap model. Nat. Rev. Neurosci. 3, 319–321.

Fox, P.T., Laird, A.R., Lancaster, J.L., 2005a. Coordinate-based voxel-wise meta-analysis: dividends of spatial normalization. A report of a virtual workshop. Hum. Brain Mapp. 25, 1–5.

Fox, P.T., Laird, A.R., Fox, S.P., Fox, M., Uecker, A.M., Crank, M., Koenig, S.F., Lancaster, J.L., 2005b. BrainMap taxonomy of experimental design: description and evaluation. Hum. Brain Mapp. 25, 185–198.

Hamilton A.F.d.c., 2009. Lost in localization: A minimal middle way. Neuroimage 48, 8–10.

Koski, L., Paus, T., 2000. Functional connectivity of the anterior cingulate cortex with the human frontal lobe: a brain-mapping meta-analysis. Exp. Brain Res. 133, 55–65.

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.

Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M, Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum. Brain Mapp. 28, 1194–1205.

Langley, P., Laird, J.E., Rogers, S., 2009. Cognitive architectures: research issues and challenges. Cogn. Syst. Res. 10, 141–160.

Mandler, J., 2004. The Foundations of Mind. Origins of Conceptual Thought. Oxford University Press, New York.

Nielsen F.Å., 2009. Lost in localization: A solution with neuroinformatics 2.0? Neuroimage 48, 11–13.

Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. 10, 59–63.

Postuma, R.B., Dagher, A., 2006. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. Cereb. Cortex 16, 1508–1521.

Price, C.J., Friston, K.J., 2005. Functional ontologies for cognition: the systematic definition of structure and function. Cogn. Neuropsychol. 22, 262–275.

Toga, A.W., 2002. Neuroimage databases: the good, the bad, and the ugly. Nat. Rev. Neurosci. 3, 302–309.

Toro, R., Fox, P.T., Paus, T., 2008. Functional coactivation map of the human brain. Cereb. Cortex 18, 2553–2559.

Van Essen D.C., 2009. Lost in localization — But found with foci?! Neuroimage 48, 14–17.