OPINION

Mapping context and content: the BrainMap model

Peter T. Fox and Jack L. Lancaster

Functional brain imaging can be used to map the neural systems that underlie human behaviour. To probe the intrinsic complexity of the human mind and brain, a large repertoire of highly sophisticated behavioural challenges are used in imaging research. In patterns as varied and complex as the behaviours by which they are elicited, activations are reported in any and all brain areas. Describing this experimental corpus by context and content is the logical prelude to any attempt to interpret, compare or combine data across studies or centres.

Functional brain imaging is having a profound impact on virtually every area of research on the brain and behaviour. Neural systems that subserve language, music, reasoning, attention, emotions, perception and action, as well as their changes with age, disease, drugs, training and other modulators of behaviour, are being mapped by thousands of investigators in hundreds of laboratories around the world. Because functional brain imaging is intrinsically multidisciplinary, studies are reported in scores of different journals. In this information explosion, simply retrieving the functional imaging literature that is relevant to any given research question is a significant challenge. Understanding the purpose, experimental design, imaging methods and results of such studies is even more labour intensive and uncertain of success. But these are necessary first steps for any use of this literature and its data. Whether the proposed use is simply reading the relevant studies, critiquing them to develop new hypotheses, carrying out a quantitative literature meta-analysis, or — at the furthest extreme — re-analysing from raw data, the starting point is the same. Relevant studies must be identified and assimilated. This points to a pressing need: a database that indexes the context and content of the functional imaging corpus in a comprehensive and readily searchable manner.

Literature indices are inadequate

Existing literature-indexing services (for example, MEDLINE, PubMed and Science Citation Index) are necessarily generic — designed to serve a wide range of scientific and clinical disciplines. Citation data (such as journal, title, authors and dates) are well indexed, readily understood and effectively retrieved. Context — the how and why of a study — is indexed only crudely, relying on a handful of unstructured keywords that are left to the authors' discretion. Content — the findings of the research — is in an even worse state. For functional brain mapping, the results represent activations and deactivations in specific brain areas. At present, the most that an author can accomplish with keywords is to highlight one or two areas that might capture the interest of the reader. As papers commonly report dozens of activation sites per experiment and multiple experiments per paper, keywords cannot index content in a meaningful way.

So, keywords, as they are used at present, are not an adequate system by which to find, filter or sort conceptually related studies (context), and are weaker still as means to analyse content. What we need is an indexing system that maps both context and content as a multidimensional feature space. Given the number and diversity of research domains and the high level of domain-specific expertise that is needed for such an undertaking, it is only reasonable that each discipline take responsibility for developing its own context-mapping scheme. The BrainMap database was designed to provide this service for the human functional brain-imaging literature[1]. As this is an entirely new type of undertaking in scientific information management, BrainMap could also serve as a model for other disciplines.

BrainMap Experiment Coding Scheme

The BrainMap database maps context through a multidimensional, multiscale system of structured keywords that categorize experiments along numerous axes. These axes include intent (for example, normal mapping, disease effects, drug effects and age effects), behavioural domain (such as language, reasoning and memory), experimental paradigm (Stroop task, anti-saccades task and so on), sensory modalities (vision, audition, touch and taste), types of stimulus (words, objects and abstract patterns), motor responses by body part and type (for example, saccadic eye movements), and cognitive task type (detection, discrimination and generation). Similarly, image acquisition modalities (for example, positron emission tomography, functional magnetic resonance imaging (fMRI), magnetoencephalography and event-related potentials), experimental-design category (block or event-related), image normalization and other pre-processing steps, and methods of statistical analysis are all coded in structured fields (BOX 1).

The extensive use of structured keywords — although labour intensive in design, implementation and data input — is enormously powerful for rapid, exhaustive retrieval of all studies in a given domain. Not surprisingly, it is also proving to be an extremely efficient route to understanding a study. Not all authors write with sufficient clarity and completeness to allow the printed literature to be readily understood, even by experts. The BrainMap Experiment Coding Scheme provides an effective shorthand for experiment description, while establishing explanatory standards and conventions. Furthermore, this structured keyword strategy allows rapid retrieval of similar studies and successive filtering of retrieved studies by objective, well-specified criteria, which is a great advantage for formal meta-analysis[2].

Standardized coordinates

Content indexing will be more readily achieved in human brain mapping than in any other research domain because, from the inception of the field, studies have used standardized *x*,*y*,*z* coordinates to analyse and report brain locations[3]. This unusual convention had two initial motivations: to enhance sensitivity of response detection and to create a reporting standard and facilitate communication. Placing images in a common reference space allows images to be averaged across subjects, greatly improving the signal-to-noise ratio and allowing more general conclusions to be drawn than could be from single-subject images[4]. As a reporting standard, coordinates are often more spatially precise than standard anatomical terms. In large part, this precision derives from coordinates being reported to the millimetre (with a standard localization error of 1–2 mm), whereas anatomical terms often apply to much larger regions. This precision also arises by avoiding reliance on surface anatomy to describe the location of functional areas. Evidence is steadily accumulating that surface features (sulci and gyri) are rather crude and unreliable predictors of functional-area locations[5–7]. Furthermore, coordinates

Box 1 | BrainMap context descriptors

- Intent: normative mapping, ageing, development, disease effects and so on
- Subjects: number, gender, handedness, diseases and so on
- Behavioural domain: perception, action, cognition, emotion and so on
- Experimental conditions and contrasts
- Acquisition modality and methods
- Analysis software and methods

improve precision by avoiding national, regional, local or personal idiosyncrasies in the use of anatomical terms. It has been rightly said that scientists would rather share toothbrushes than terminologies. Coordinates entirely circumvent these quaint but irritating impediments to communication.

For BrainMap, content includes *x*,*y*,*z* coordinates of the centre of each activation, signal-intensity percentage change, *z*-score, *p*-value and extent (mm^3) of activated sites in each condition (BOX 2). Location coordinates also provide links to a hierarchical, volume-based set of standardized anatomical descriptors and segmentations of the search space by hemisphere, lobe, gyrus, tissue type and cell structure[8,9]. Having activations mapped by location allows the user to retrieve easily all studies that report activations in a given area or set of areas. Descriptors of activation intensity, extent and *p*-value allow users to filter searches by statistical and physiological significance, bringing the user into the quality-control process.

What level of data to share?

The processing and reduction of raw functional images into a scientifically meaningful and publishable format is a complex, multi-stage, and steadily evolving process. Raw images, especially fMRI images, need to be normalized for global and local sensitivity variations (radio-frequency inhomogeneities and susceptibility artefacts) and corrected for spatial distortions, which are often gross. Functional images contain little anatomical detail, and need to be registered to each subject's high-resolution anatomical images before conversion to standardized coordinates. Subject movements need to be identified and removed. Having completed these steps, statistical parametric images (SPIs) can be computed either per subject or per group. Just as with any type of data, a wide variety of statistical tests can legitimately be applied, producing a wide range of types of SPI. With all these levels of data, the question "what level of data do we share?" is necessarily difficult.

Here's a simple rule for deciding what to share: share the most valuable data type first. Then expand to less valuable data until a point of diminishing returns is achieved. From our perspective, each processing step through which raw images are carried adds value. The authors have laboured to improve the raw data, relying heavily on local expertise and typically using algorithms that are carefully optimized to correct for idiosyncrasies in their imaging systems. The data have also been standardized, making comparisons with data sets from other laboratories more valid and more valuable. So, the simple answer to the question of what to share is to share the most refined data available. For functional brain imaging, the most refined data are tables of activations in standardized space, and with standardized descriptions of intensity and extent. These are the content that BrainMap shares and for which it provides query tools.

Following this rule, group SPIs would be the next data type to distribute, followed by individual-subject SPIs. When properly normalized, both group and per-subject SPIs have two distinct uses: first, in quantitative comparisons across studies; and second, in pooling data into a new mega-SPI to increase statistical power. Both uses, of course, require a detailed understanding of study context, making the BrainMap context-mapping scheme invaluable. The latter use (pooling SPIs) makes sense only if studies that are pooled used identical or closely comparable behavioural manipulations.

Open sharing of raw functional imaging data is far more problematic and of less clear value. The user must assume full responsibility for reprocessing the data from scratch, removing all artefacts without the benefit of local knowledge. Even the proper grouping of data by subjects and trials implies that the user must fully comprehend the experimental design down to the finest detail, including the inevitable experimenter errors and omissions in the laboratory logs. The only truly reliable way to accomplish data reprocessing is with the assistance of the investigators who acquired the data. Given this, it would seem most reasonable to request the data directly from the people who created it, rather than obtaining it from a third party.

Box 2 | BrainMap content descriptors

- Coordinates: centre of activity
- Volume of activation
- Percentage signal change
- Published statistical parameter: *t*-score, *r*-value, *z*-score and so on
- Standardized statistical parameter: *z*-score
- Significance level
- Standard anatomical descriptor: Talairach Daemon labels
- Functional area terms: V1, V2, area MT/V5, supplementary motor area and so on

Despite these reservations about the value of sharing more raw forms of functional imaging data, BrainMap does provide links to sites that provide such data, including submission-specific accession numbers. So, BrainMap users have direct links to more raw data forms of any type and in any location. These sites may be centralized repositories, peer-to-peer sharing networks or ftp sites at individual laboratories. In this way, the context and content search capabilities of BrainMap serve as an infrastructure that supports any and all community data-sharing efforts.

Quality control times three

Quality control is a much debated issue in database development. Three general strategies can be identified: pre-set quality standards, committee review of submissions, and quality descriptors that allow users to set their own quality-control standards. BrainMap uses all three strategies in moderation.

BrainMap has set two firm quality standards. First, all accepted data must be from studies that appeared published (or are in press) in a peer-reviewed journal. Second, all submissions must include activation locations that are reported in standardized coordinates. Coordinate-based entry ensures that data can be rapidly and reliably searched and rigorously compared across studies. As some excellent studies have failed to publish coordinates, coordinates computed after publication will be accepted so as not to be unnecessarily exclusionary. Together, these two simple standards (peer-reviewed publication and standardized coordinates) go a long way to ensure data quality and homogeneity of description.

BrainMap also requires submission review by two members of our Editorial Board (FIG. 1). This review, however, is not a second scientific review. Instead, its purpose is to ensure that the entry is completely and correctly coded and to verify that the two above-stated quality-control standards are met. Of equal importance, the Editorial Board provides an ongoing critique of the BrainMap context-mapping scheme. As the field of functional brain mapping evolves, this scheme will need to be updated.

The detail in which experiments are described in BrainMap empowers users to set their own quality standards. Studies can be selected by experimental design features,

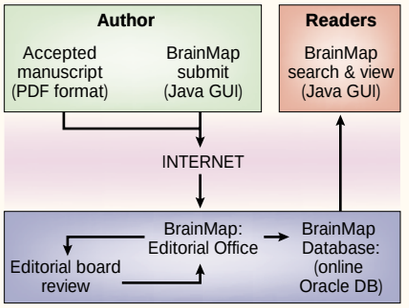

Figure 1 | **Publication in the BrainMap database.** Once a functional brain-mapping paper is accepted in a peer-reviewed journal, it is eligible for submission to BrainMap. This is a three-step process. First, the author encodes the paper using the BrainMap Submit software, which is available at no charge from the BrainMap web site. The Submit software is written in Java so that it runs on virtually all operating systems. The Submit file and the accepted manuscript in PDF format are sent by e-mail to the BrainMap editorial office. Second, review is carried out by two members of the Editorial Board, who have been trained in BrainMap coding and have overseen the coding of the entire published corpus of their laboratories. If necessary, coding is revised at the direction of the Editorial Board and the Editor-in-Chief. Third, the coded research is entered into the BrainMap database. BrainMap is an online Oracle database. The primary site is in San Antonio, Texas. Mirror sites have been tested in St Louis and in Jülich, Germany. Further mirror sites are being planned.

imaging modality, number of subjects (as well as handedness, gender and age) and statistical significance, to name only a few. The query "Report all fMRI studies with more than eight subjects in which words were auditorily presented using an event-related design and activated the posterior third of the superior temporal gyrus to a $p < 0.001$" sets rigorous quality standards, but is readily achieved by existing context and content descriptors.

Another important quality-control feature of BrainMap is that authors code their own data for submission. Extended attempts to carry out context and content coding on papers from other laboratories firmly convinced us that the published literature is grossly underspecified. In most instances, published articles do not provide sufficient detail to fully code a paper for BrainMap submission. So, an emerging role of the BrainMap database is to supplement the printed literature, providing a richness of detail and consistency of description otherwise not available.

Predicted uses and users

The simplest but broadest use for BrainMap is to identify experiments of interest. If you are searching for functional imaging studies, what better route than through an index that is dedicated to this field? BrainMap coding is also an invaluable aid to understanding published studies. For the imaging novice, reading a coded paper is like being guided by the authors themselves through an often dense and occasionally impenetrable thicket of unfamiliar jargon and omitted details. For the imaging expert, it is a welcome short cut through a large and complex literature. Given the enormous interest in brain–behaviour relationships, we expect BrainMap to be a teaching tool that is used by scientists, clinicians, educators and students in many disciplines.

Meta-analysis is another use of BrainMap. Formal meta-analyses, published as experiments in their own right, are making valuable contributions to the functional imaging literature[2,10–14]. On a less grand scale, mini meta-analyses are commonly used to interpret functional imaging results and are being included in the discussion sections of functional imaging experimental reports. We have argued that functional imaging is uniquely well suited to meta-analysis, readily avoiding many of the pitfalls that have plagued meta-analysis in other fields[2].

Longevity — the business model

Databases are fiscally problematic undertakings for funding agencies. Typical research projects last for a few years and then end, freeing up funds for new projects. Successful databases never end. Rather, they grow and grow, seeking ever-larger budgets. So, it might be unreasonable to expect funding agencies to provide long-term support for databases. On the other hand, scientists that contribute to databases should rightly demand some assurance that their efforts in data coding will have permanence. Hence the question, how can a scientific database achieve financial self-sufficiency?

BrainMap is seeking financial self-sufficiency by emulating two information-distribution services: literature-indexing services and scientific journals. Indexing services facilitate paper retrieval. Journals provide new scientific content. BrainMap resembles an indexing service in providing searchable keywords that describe article context and content. BrainMap resembles a journal in requiring author submission and peer review, and in distributing new scientific information (albeit about published studies). Both indexing services and journals achieve fiscal independence by charging subscription fees. For BrainMap, our intention is to make a gradual transition from grant funding to subscription-based funding.

Peter T. Fox and Jack L. Lancaster are at the Research Imaging Center, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, Mail Code 6240, San Antonio, Texas 78229-3900, USA.
Correspondence to P.T.F.
e-mail: fox@uthscsa.edu

DOI: 10.1038/nrn789

1. Fox, P. T. & Lancaster, J. L. Neuroscience on the Net. *Science* **226**, 994–996 (1994).
2. Fox, P. T., Parsons, L. M. & Lancaster, J. L. Beyond the single study: function/location metanalysis in cognitive neuroimaging. *Curr. Opin. Neurobiol.* **8**, 178–187 (1998).
3. Fox, P. Spatial normalization: origins, objectives, applications and alternatives. *Hum. Brain Mapp.* **3**, 161–164 (1995).
4. Fox, P. T., Mintun, M. A., Reiman, E. M. & Raichle, M. E. Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab.* **8**, 642–653 (1988).
5. Watson, J. D. G. *et al.* Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex* **3**, 79–94 (1993).
6. Zilles, K. *et al.* Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. *Hum. Brain Mapp.* **5**, 218–221 (1997).
7. Hasnain, M. K., Fox, P. T. & Woldorff, M. G. Structure–function covariance in the human visual cortex. *Cereb. Cortex* **11**, 702–716 (2001).
8. Lancaster, J. L. *et al.* Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum. Brain Mapp.* **5**, 238–242 (1997).
9. Lancaster, J. L. *et al.* Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* **10**, 120–131 (2000).
10. Fox, P. T. *et al.* Location-probability profiles for the mouth region of human primary motor–sensory cortex: model and validation. *Neuroimage* **13**, 196–209 (2001).
11. Xiong, J. *et al.* Intersubject variability in cortical activations during a complex language task. *Neuroimage* **12**, 326–339 (2000).
12. Paus, T. Location and function of the frontal eye fields: a selective review. *Neuropsychologia* **34**, 475–483 (1996).
13. Picard, N. & Strick, P. L. Motor areas of the medial wall: a review of their location and functional activation. *Cereb. Cortex* **6**, 342–353 (1996).
14. Schulman, G. L. *et al.* Common blood flow changes across visual tasks. I. Increases in subcortical structures and cerebellum but not in visual areas. *J. Cogn. Neurosci.* **9**, 623–646 (1997).

Acknowledgements

The development of BrainMap was supported by the National Library of Medicine and the National Institute of Mental Health, the Human Brain Project and the EJLB Foundation. We thank S. Petersen, J. Xiong and L. Parsons for much helpful advice in BrainMap design and logistics. We thank A. Hecker, P. Kochunov and S. Mikiten for development of software for BrainMap and the Talairach Daemon. We thank S. Fox and M. Crank for extensive testing of the BrainMap coding scheme for designing and developing BrainMap training tools and procedures.

Online links

FURTHER INFORMATION
BrainMap: http://www.brainmapdbj.org/
Encyclopedia of Life Sciences: http://www.els.net/
bioinformatics | biological data centres | brain imaging: localization of brain functions | brain imaging: observing ongoing neural activity | computed tomography | magnetic resonance imaging | mining biological databases
MIT Encyclopedia of Cognitive Sciences: http://cognet.mit.edu/MITECS/
electrophysiology, electric and magnetic evoked fields | magnetic resonance imaging | positron emission tomography
Access to this interactive links box is free online.