The Brede database and new tools for meta-analysis

Finn Årup Nielsen

Lundbeck Foundation Center for Integrated Molecular Brain Imaging
at
Department of Informatics and Mathematical Modelling
Technical University of Denmark
and
Neurobiology Research Unit,
Copenhagen University Hospital Rigshospitalet

August 1, 2011

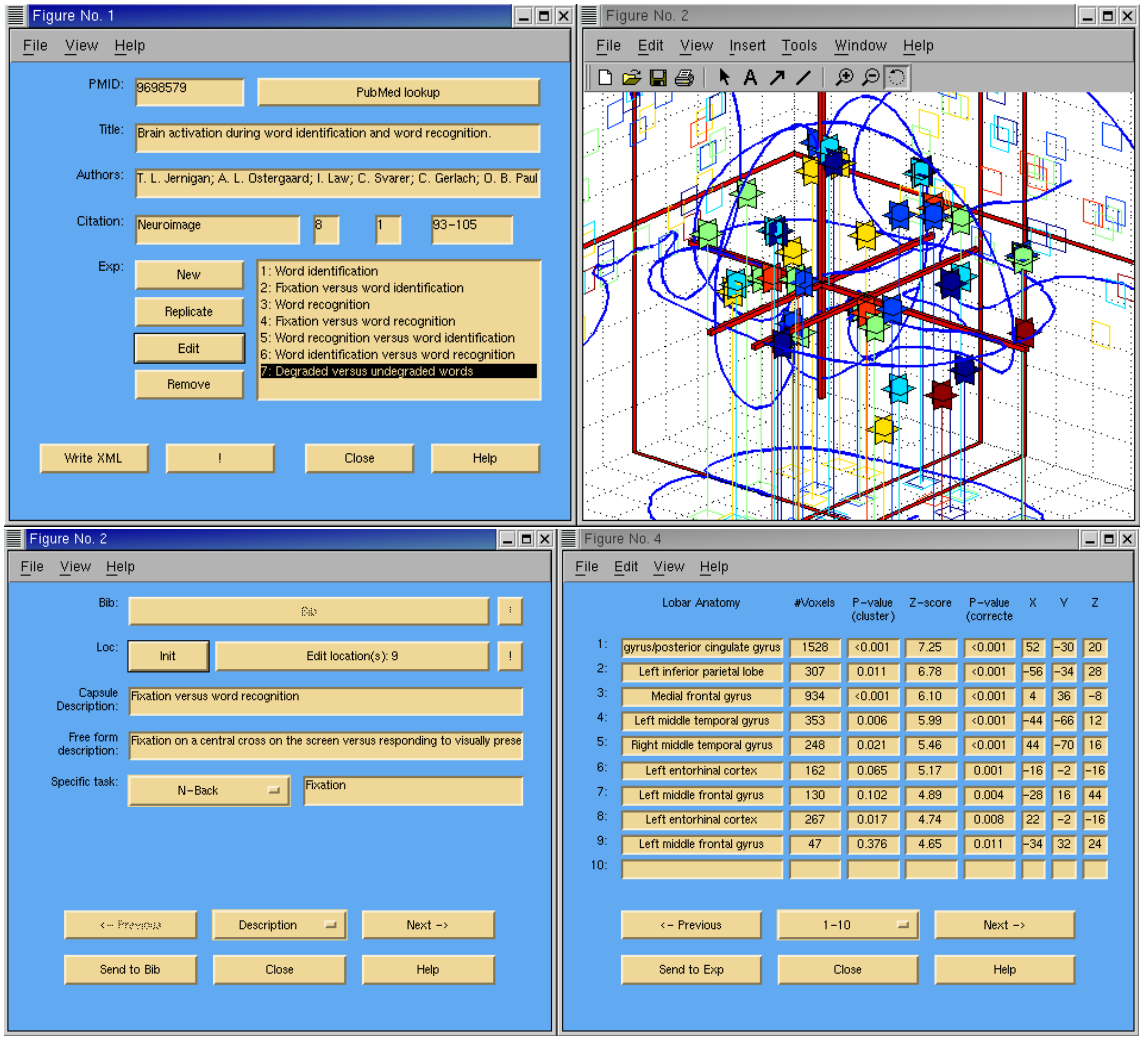
The 3 Brede tools

Brede Toolbox: A program package primarily written in Matlab. Handles visualization, linear modeling, multivariate analysis, locations (Talairach coordinates), volumes, papers, texts.

Brede Database: Basically a collection of XML files with data from neuroimaging papers as well as ontologies. Distributed with the Brede Toolbox. "Output" and query services to the Brede Database (generated with the Brede Toolbox) is available on the Internet: http://neuro.imm.dtu.dk

Brede Wiki: A wiki with data from neuroimaging papers as well as ontologies. Both freeform text and "semantically" organized within MediaWiki templates.

Brede Toolbox with the Brede Database

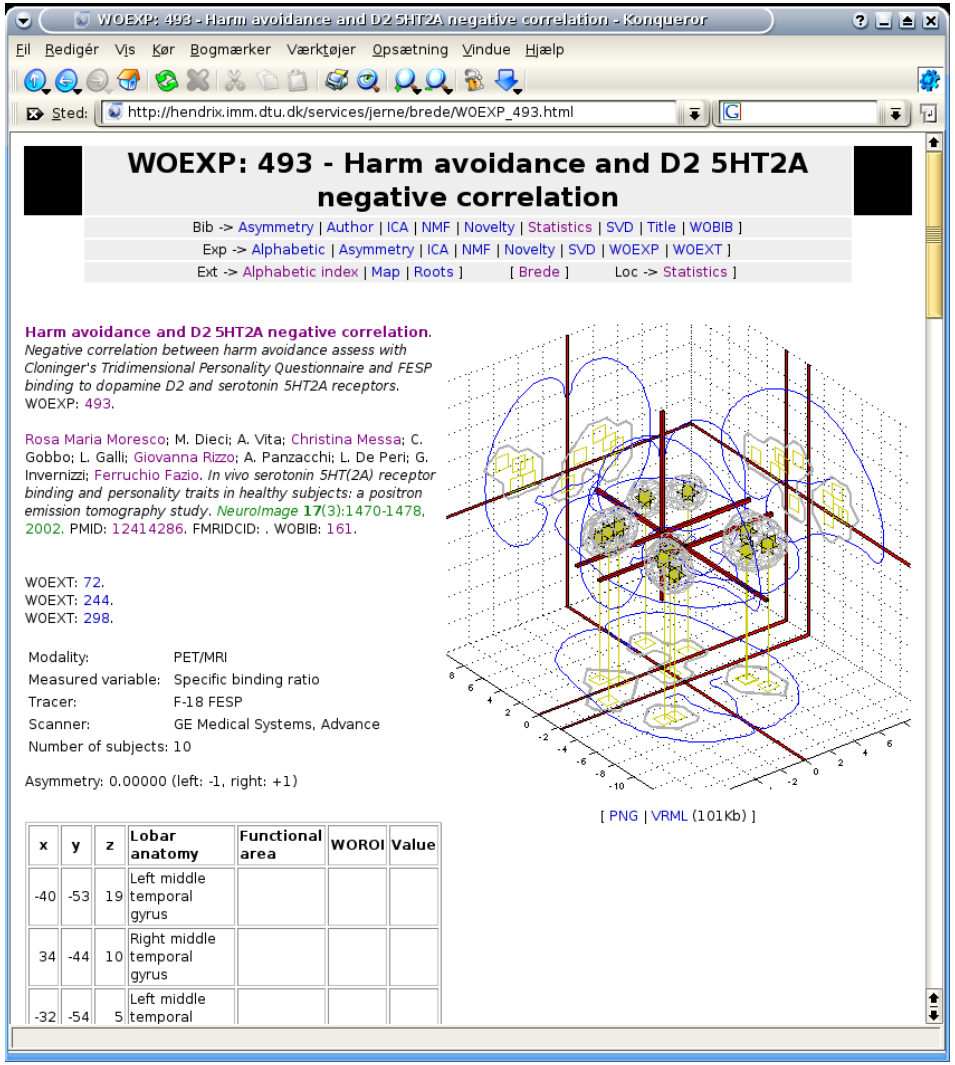


Graphical user interface of Brede Toolbox used to enter data into the Brede Database.

Brede Database: A database with results from published neuroimaging studies as well as ontologies for, e.g., brain regions and brain functions (Nielsen, 2003).

Data stored in XML available on the Web

The Brede Database on the Web

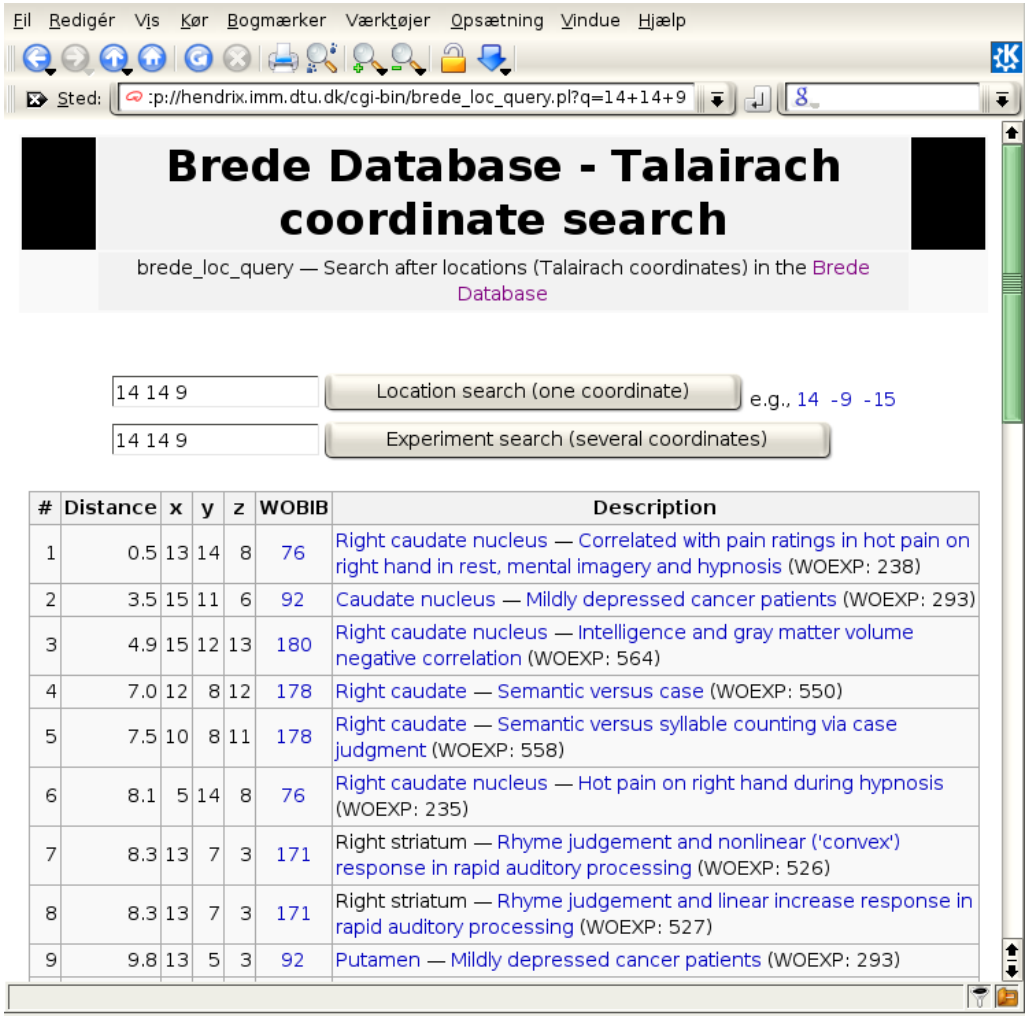


Presentation on the Web

Off-line meta-analysis and generation of indices and visualization in static HTML.

Interactive search on coordinates from Web page or within a image analysis program (Wilkowski et al., 2009).

Searching on Talairach coordinate



#	Distance	x	y	z	WOBIB	Description
1	0.5	13	14	8	76	Right caudate nucleus — Correlated with pain ratings in hot pain on right hand in rest, mental imagery and hypnosis (WOEXP: 238)
2	3.5	15	11	6	92	Caudate nucleus — Mildly depressed cancer patients (WOEXP: 293)
3	4.9	15	12	13	180	Right caudate nucleus — Intelligence and gray matter volume negative correlation (WOEXP: 564)
4	7.0	12	8	12	178	Right caudate — Semantic versus case (WOEXP: 550)
5	7.5	10	8	11	178	Right caudate — Semantic versus syllable counting via case judgment (WOEXP: 558)
6	8.1	5	14	8	76	Right caudate nucleus — Hot pain on right hand during hypnosis (WOEXP: 235)
7	8.3	13	7	3	171	Right striatum — Rhyme judgement and nonlinear ('convex') response in rapid auditory processing (WOEXP: 526)
8	8.3	13	7	3	171	Right striatum — Rhyme judgement and linear increase response in rapid auditory processing (WOEXP: 527)
9	9.8	13	5	3	92	Putamen — Mildly depressed cancer patients (WOEXP: 293)

Result after search for nearest coordinates to (14, 14, 9) with the Brede Database.

Translation of the data from XML to SQL (Szewczyk, 2008)

Perl + SQLite web-script

Similar searches possible in Antonia Hamilton's AMAT programs, BrainMap, SumsDB and Brede Wiki.

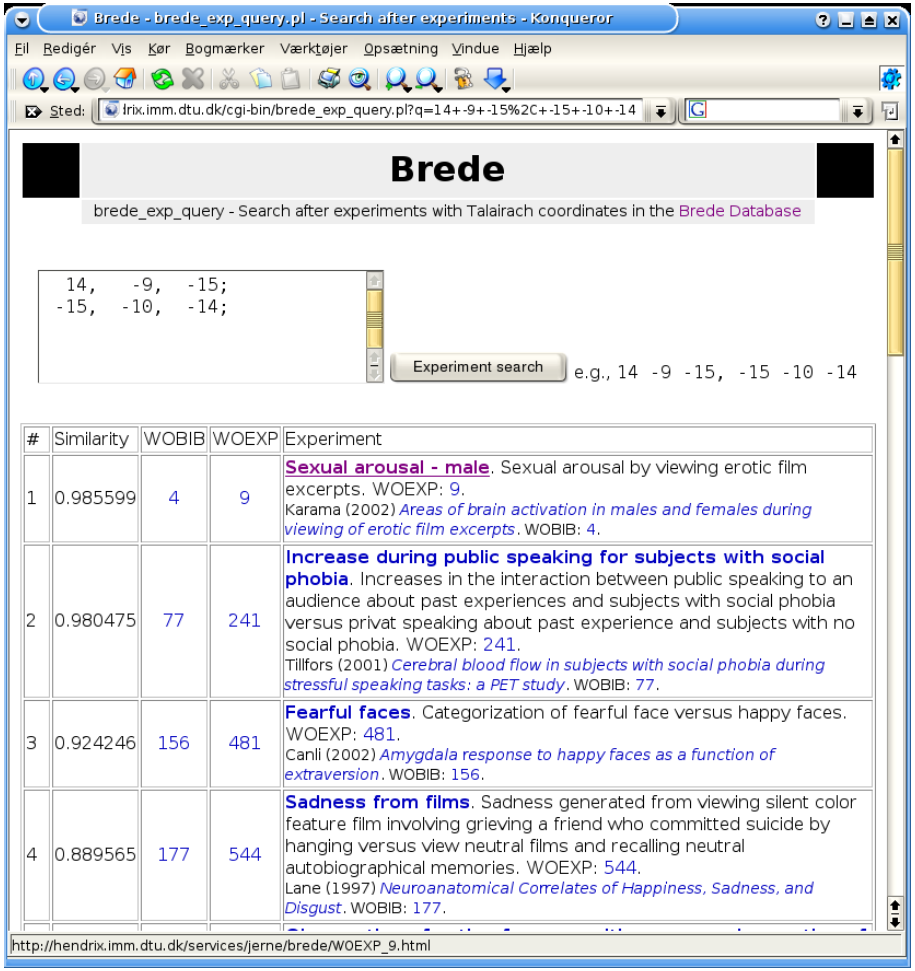
Online experiment search (multiple coordinates)

Online search on two coordinates in left and right amygdala in the experiments recorded in the Brede Database.

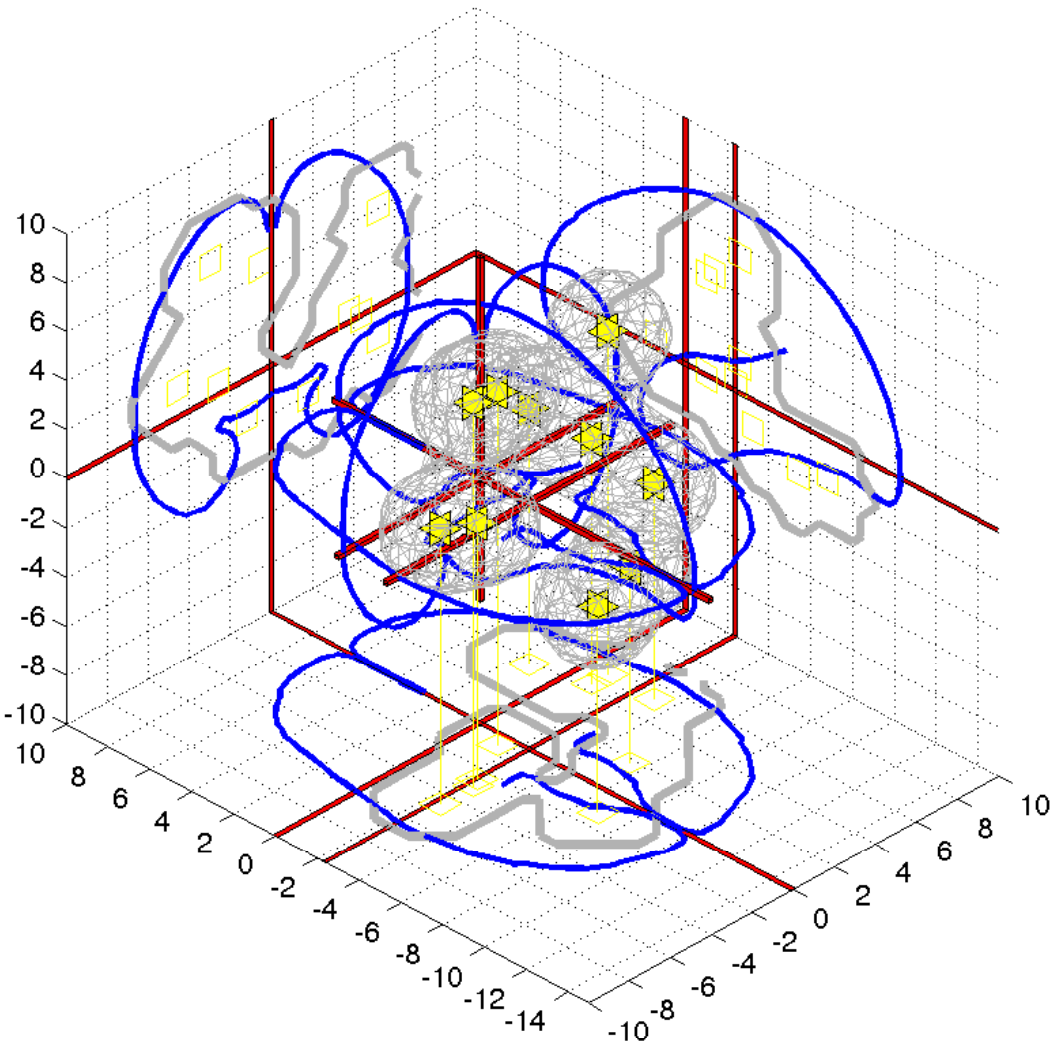
"Related volume" also available from the "original" BrainMap database (Nielsen and Hansen, 2004):

http://neuro.imm.dtu.dk/services/jerne/ninf/

Search available to the Brede Database from SPM plugin (Wilkowski et al., 2009).



Coordinates-to-volume transformation

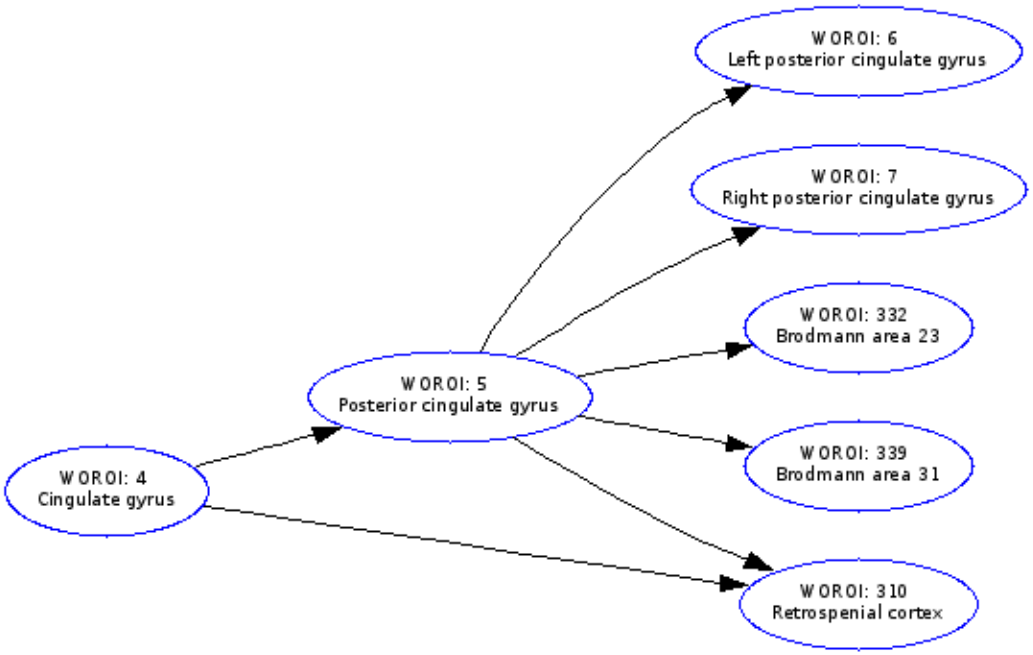


Coordinates in an article converted to volume-data by filtering each point (kernel density estimation) (Nielsen and Hansen, 2002b; Turkeltaub et al., 2002)

One volume for each article or one volume for a set of coordinates in multiple articles.

Yellow coordinates from a study by (Blinkenberg et al., 1996), with grey wireframe indicating the isosurface in the generated volume

Brede brain region taxonomy/ontology



Taxonomy of neuroanatomical areas with items linked in a hierarchy with "Brain" in the top root and smaller areas in the leafs. WOROI is the ID.

Records parent region, child region, naming variations,

Links to other brain region ontologies

Links to digital brain atlases (AAL, Claus Svarer, Alexander Hammers)

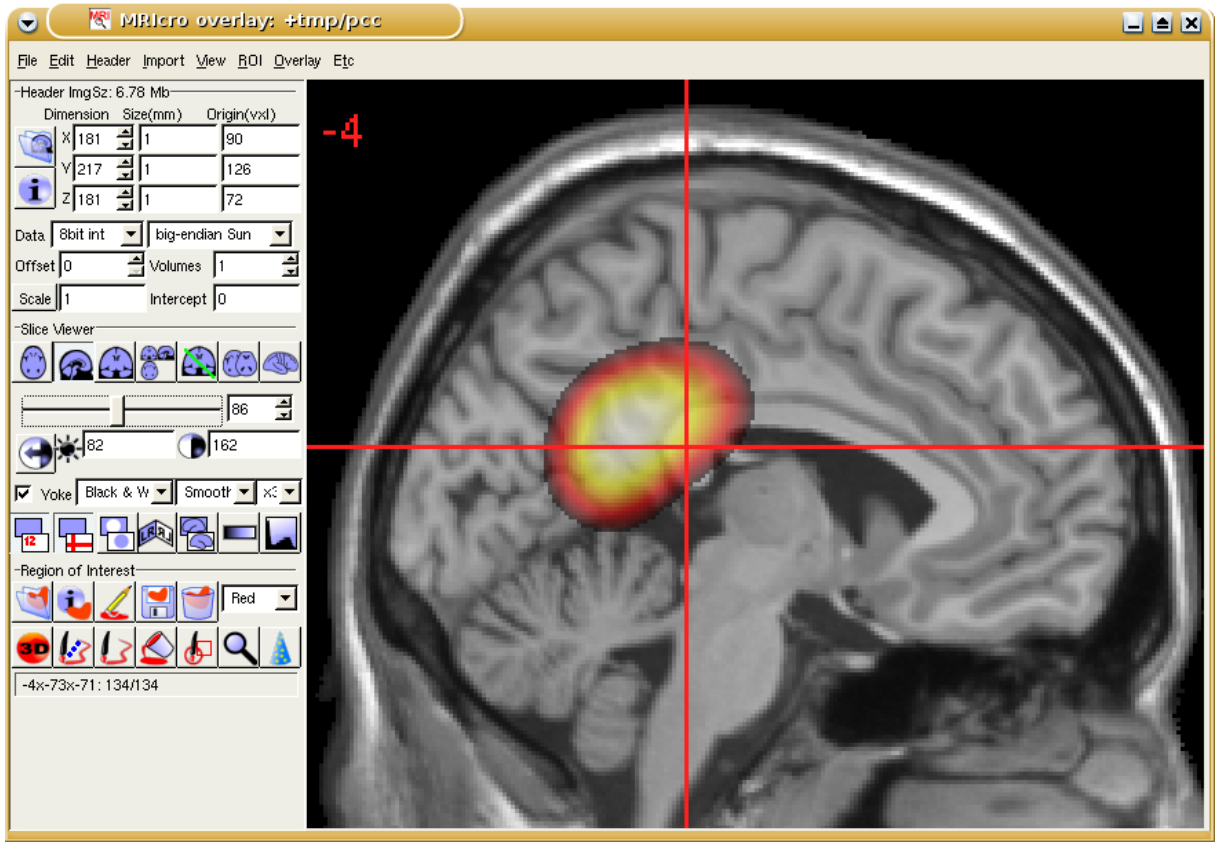
Example with Brain region ontology

The ontology enables one to get all names for PCC and its subregions. Output is (24 names in total):

```
'Posterior cingulate gyrus'
'Posterior cingulate'
'Posterior cingulate area'
'Posterior gyrus cinguli'
'Posterior cingulate cortex'
'Left posterior cingulate gyrus'
'Left posterior cingulate'
'Posterior cingulate gyrus, left'
...  e.g., BA23, retrosplenial, ...
```

Suitable for text mining where you identify as many occurrences in a corpus that is not using a controlled vocabulary, such as ordinary scientific articles.

Example: Get PCC locations



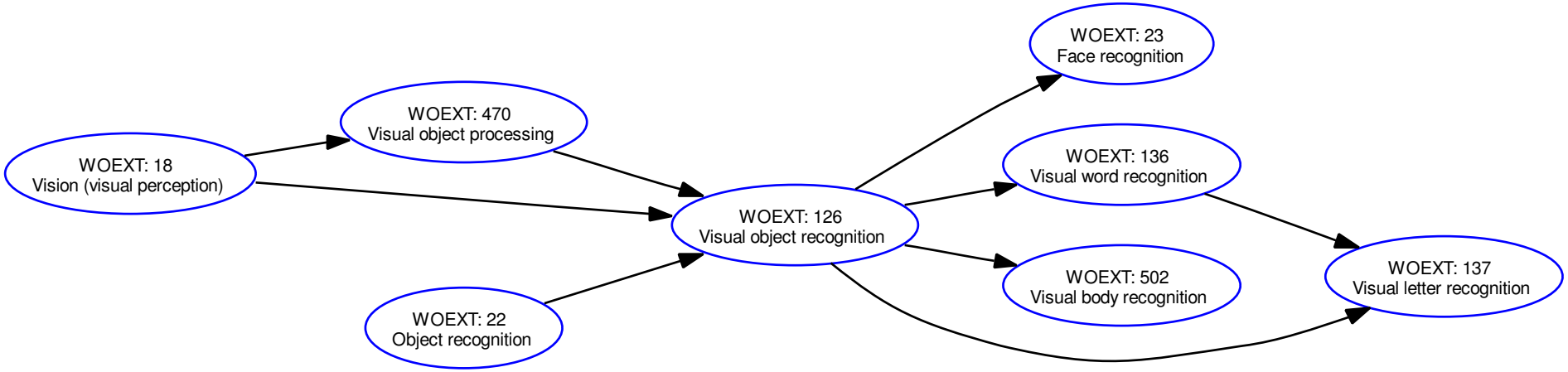
Get all posterior cingulate locations that match on of the naming variation for the regions and its subregion.

Model the locations with kernel density estimation, and convert the density to a probability.

Volume written to an Analyze file

Viewed in the external MRIcro program

Topics ontology



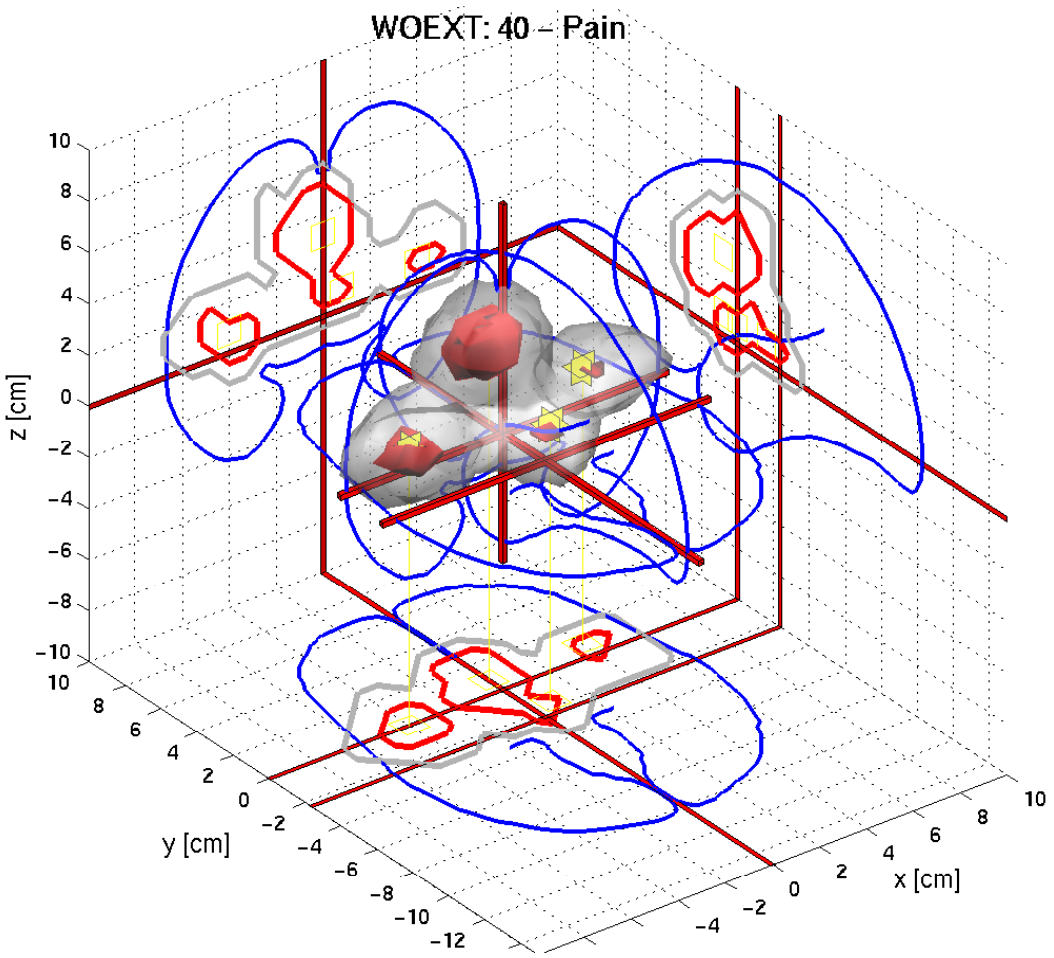
Topics, such as brain functions and mental disorders, organized in a hierarchy. Example: episodic memory retrieval, OCD, 5-HT2A receptor.

Used to label each neuroimaging experiment

Other efforts: MeSH (too coarse), BrainMap, Cognitive Atlas (Poldrack), Cognitive Paradigm Ontology (Turner and Laird, 2011)

Cognitive components are "open to interpretation"

Supervised data mining



Volume for a specific taxonomic component: "Pain" Volume threshold at statistical values determined by resampling statistics (Nielsen, 2005). Red areas are the most significant areas: Anterior cingulate, anterior insula, thalamus. In agreement with "human" reviewer (Ingvar, 1999).

Implementations of supervized datamining in the Brede Toolbox and in GingerALE.

Text representation: a "bag-of-words"

	'memory'	'visual'	'motor'	'time'	'retrieval'	...
Fujii	6	0	1	0	4	...
Maddock	5	0	0	0	0	...
Tsukiura	0	0	4	0	0	...
Belin	0	0	0	0	0	...
Ellerman	0	0	0	5	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Representation of the abstract of the articles in "bag-of-word". Table counts how often a word occurs

Exclusion of "stop words": common words (the, a, of, ...), words for brain anatomy, and a large number of common words that appear in abstracts. Mostly words for brain function are left. More advanced extraction: Match to ontologies

Grouping of words from articles

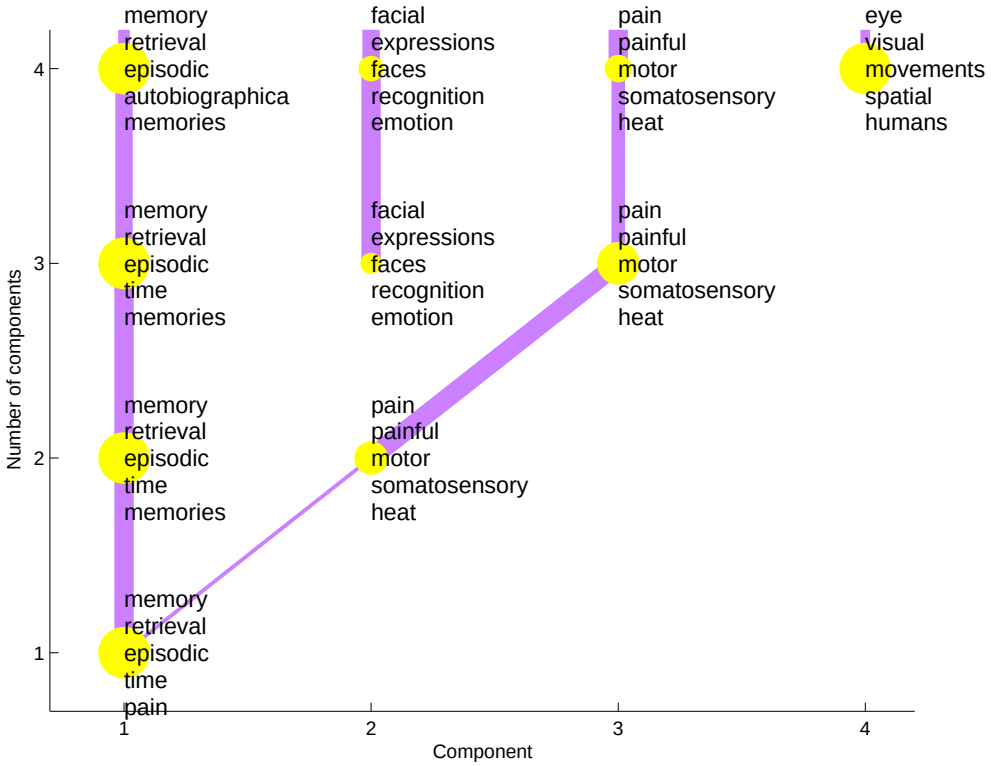


Figure 1: Grouped words.

Multivariate analysis (NMF) of the text in *posterior cingulate* articles to find "themes", which can be represented with weights over words and articles (Nielsen et al., 2005).

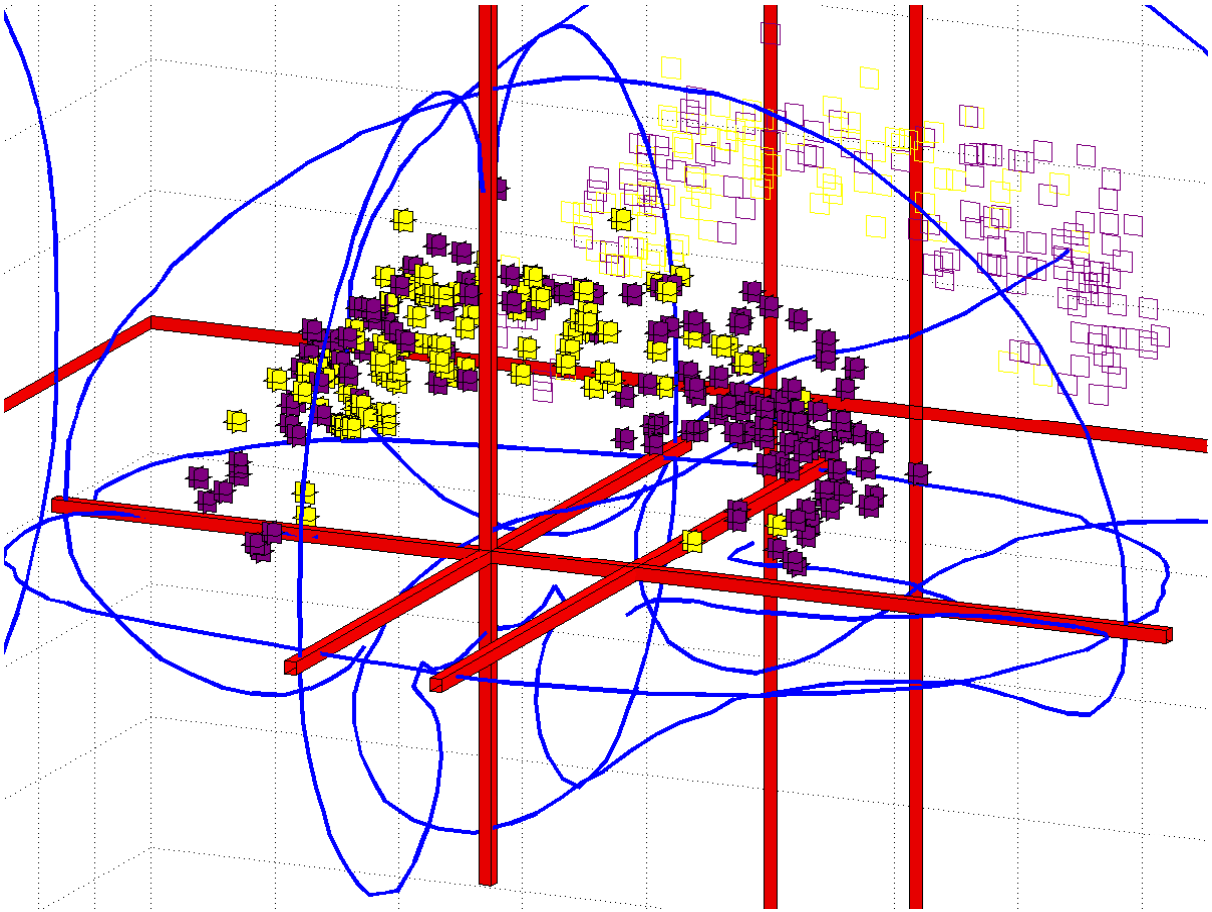
Most dominating words: memory, retrieval, episodic

pain, painful, motor, somatosensory

facial, expressions, faces,

eye, visual, movements

Combining text analysis and coordinates



Is there a difference in how brain functions distribute in the cingulate gyrus?

Possible to find the corresponding articles for the coordinates — and text mine these articles for clustering and label the coordinate according to cluster.

Sagittal plot of memory (magenta) and pain (yellow).

Text and volume: Functional atlas

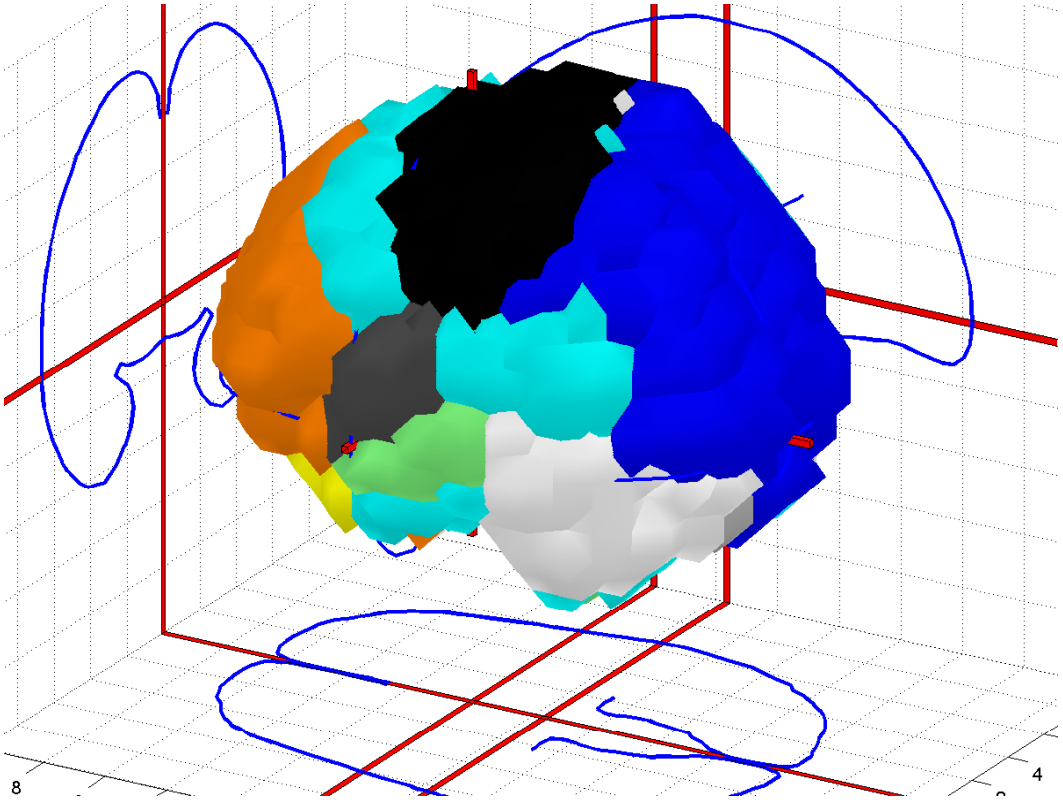


Figure 2: Functional atlas in 3D visualization.

Automatic construction of functional atlas, where words for function become associated with brain areas

Two matrices: Bag-of-words matrix, matrix from voxelization of coordinates. NMF on the product matrix.

Example components: Blue area: visual, eye, time. Black: motor, movements, hand. White: faces, perceptual, face.

Problems

Difficult to add new information to the Brede Database

Difficult to do incremental additions.

Problems

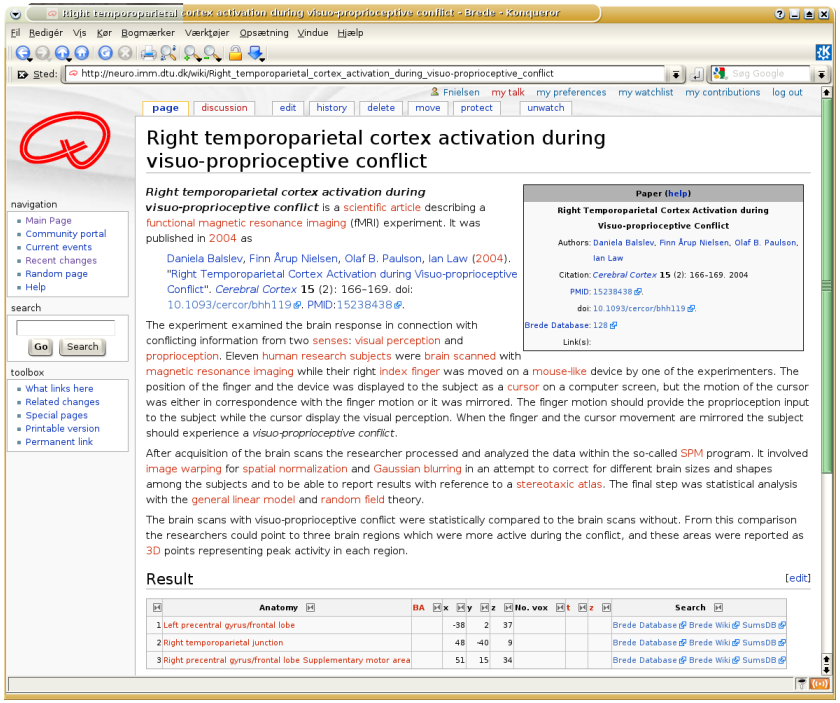
Difficult to add new information to the Brede Database

Difficult to do incremental additions.

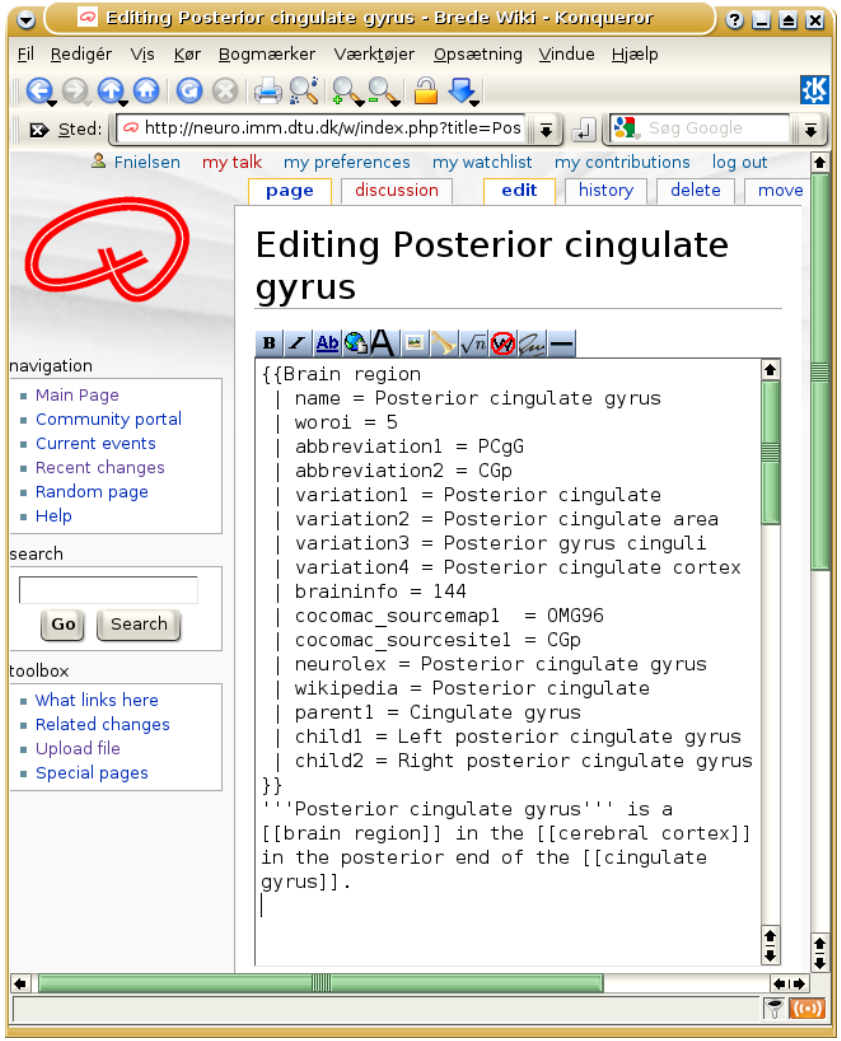
Solution?

Wiki with structured data

Brede Wiki = MediaWiki templates + Extraction + SQL + Neuroscience



Principles of the Brede Wiki



Structured information is stored in the so-called "templates" of Mediawiki.

Template use simple so it is easy to convert data all template instantiations to an SQL representation: No wiki formating in field values, non-nested templates, lower case field names (a one-to-one mapping of MediaWiki templates and ontology classes). (Nielsen, 2009)

Link as much as possible in the template values.

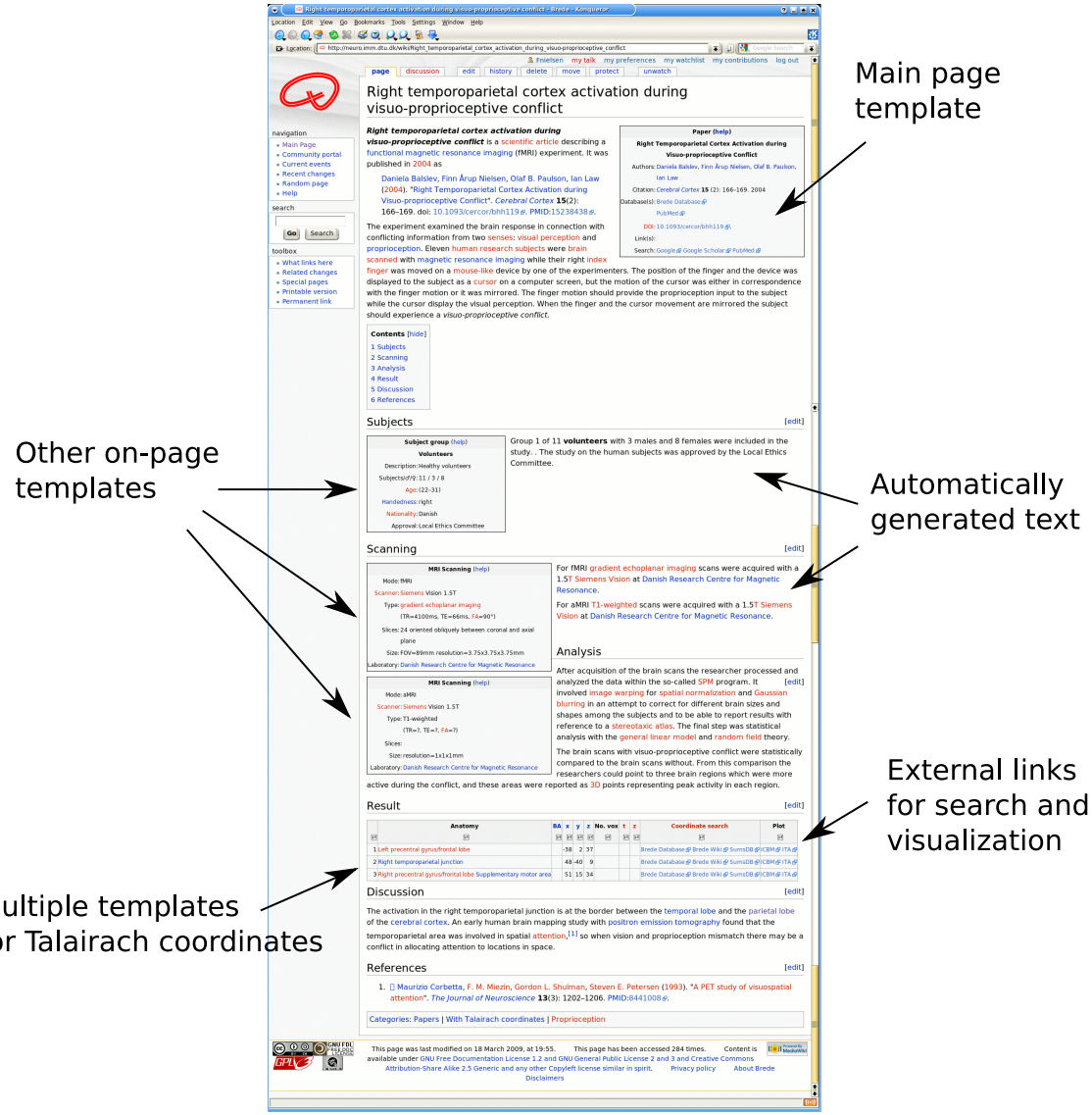
Link to external sites whenever possible.

Brede Wiki templates

Templates may describe a paper with bibliographic information or a researcher or journal.

Hierarchical templates: Brain regions, Topics, Organizations, Software.

Multiple templates on each page, e.g., to describe subject group, brain scan, experimental condition, Talairach coordinate, brain volume, gene personality association.



Storing of volumes

Scanning

MRI Scanning (help)

Mode: fMRI

Scanner: Philips Achieva 3T

Type: Gradient-echo echo-planar

(TR=3000ms, TE=35ms, FA=?)

Slices: 49 (thickness=3mm) oriented Horizontal

Size: FOV=240 x 147 x 240mm resolution=2.5 x 2.5 x

Laboratory: missing *laboratory*

For fMRI Gradient-echo echo-planar scans were acquired with a 3T Philips Achieva.

For aMRI T1-weighted scans were acquired with a .

MRI Scanning (help)

Mode: aMRI

Scanner:

Type: T1-weighted

(TR=?, TE=?, FA=?)

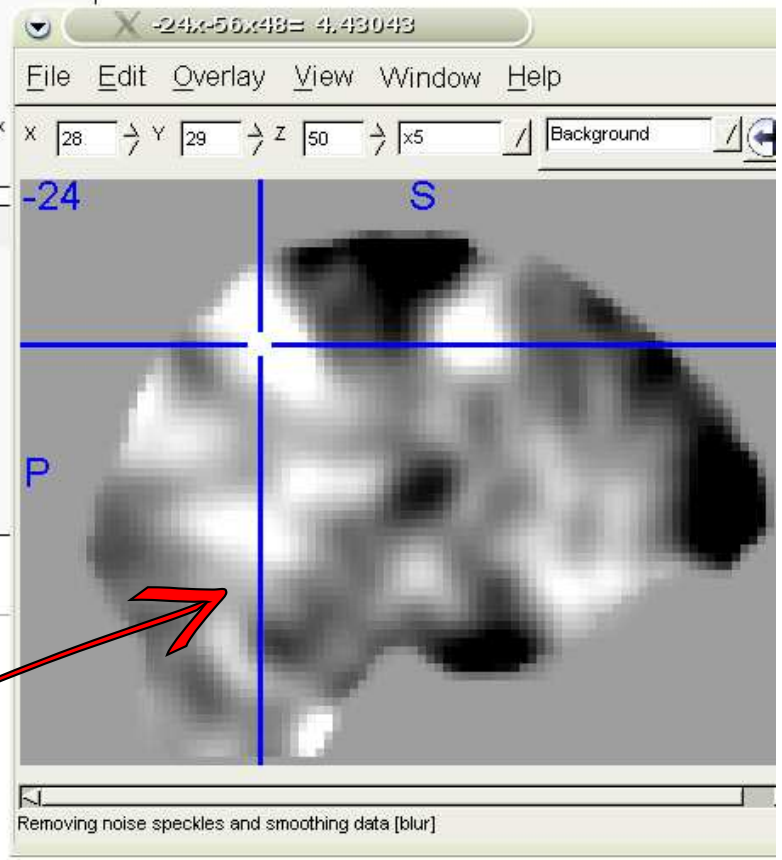
Slices:

Size: resolution=1 x 1 x 2mm

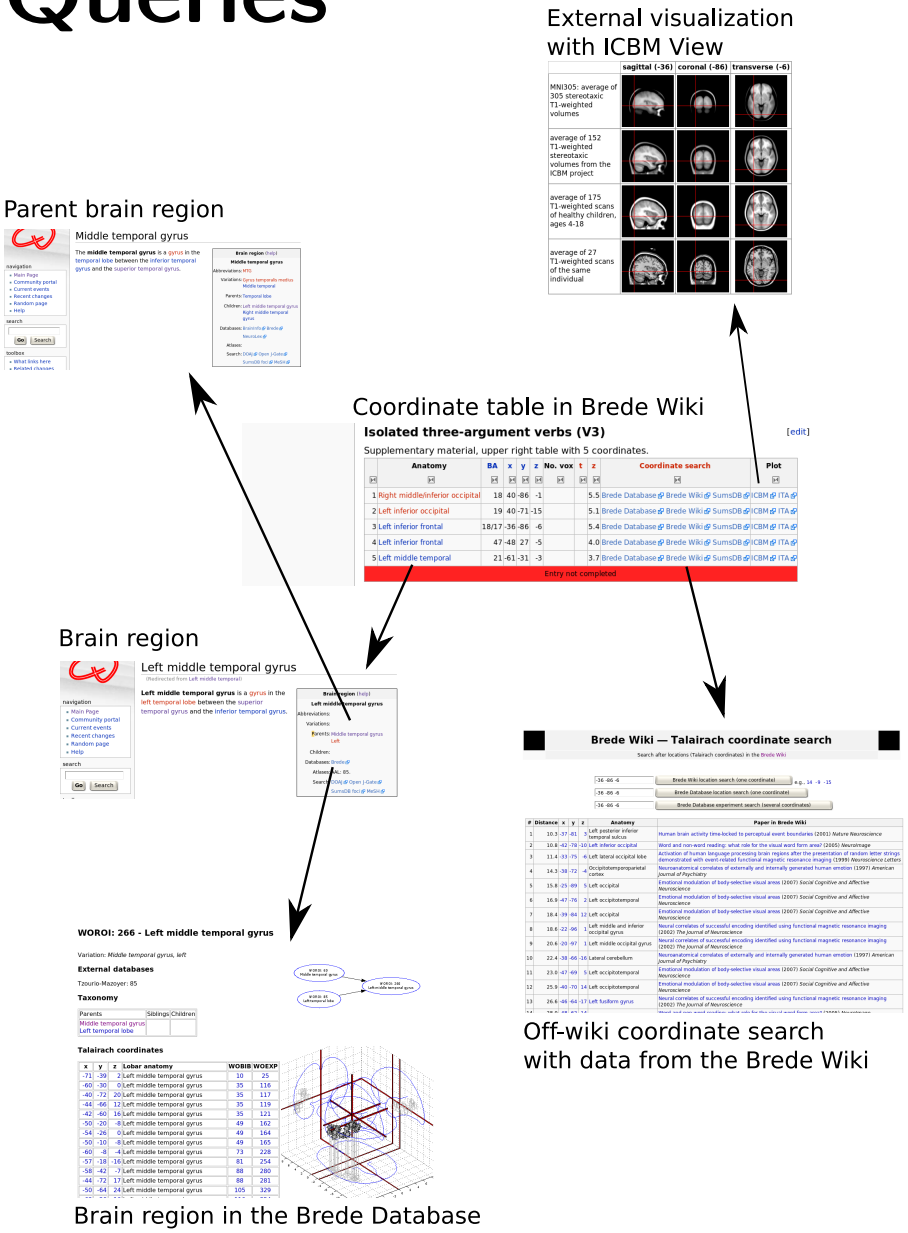
Laboratory: missing *laboratory*

Results

Volume: Contrast image



Queries



Structured content can be extracted (like DBpedia on Wikipedia)

Queries are possible, but not within the wiki

Query on nearby coordinates with an off-wiki script.

So-called “SKOS file” (Miles and Bechhofer, 2009) generated for brain region and topic hierarchies from the structured content.

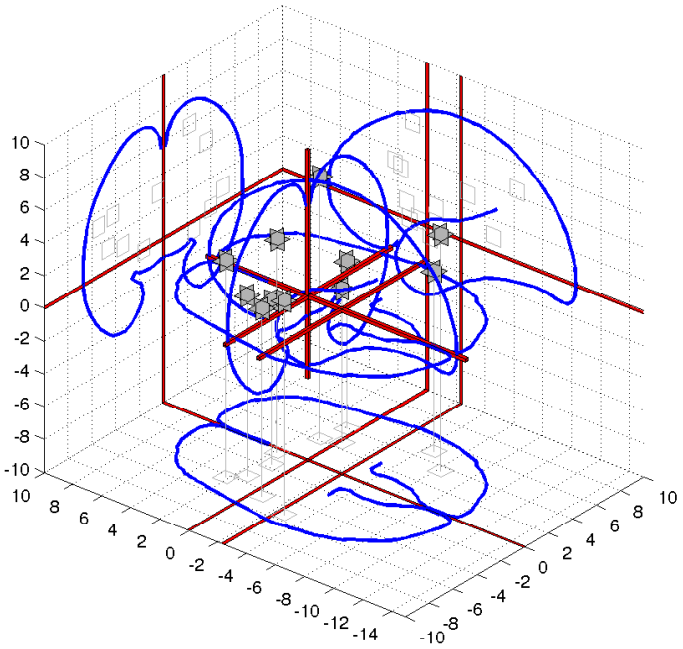
Brede Wiki and Toolbox integration

You can get data from the Brede Wiki into Matlab, e.g., paper in the Brede Wiki (Lin et al., 2008):

```
>> title = 'Brain maps of Iowa gambling task';
>> Ls = brede_web_bw2loc(title);
>> figure, brede_ta3_frame, brede_ta3_loc(Ls)
```

Get the page from the Web site and extract the information within the templates and convert to a structure that fits the Brede Toolbox and Database.

Finally, plot the locations.

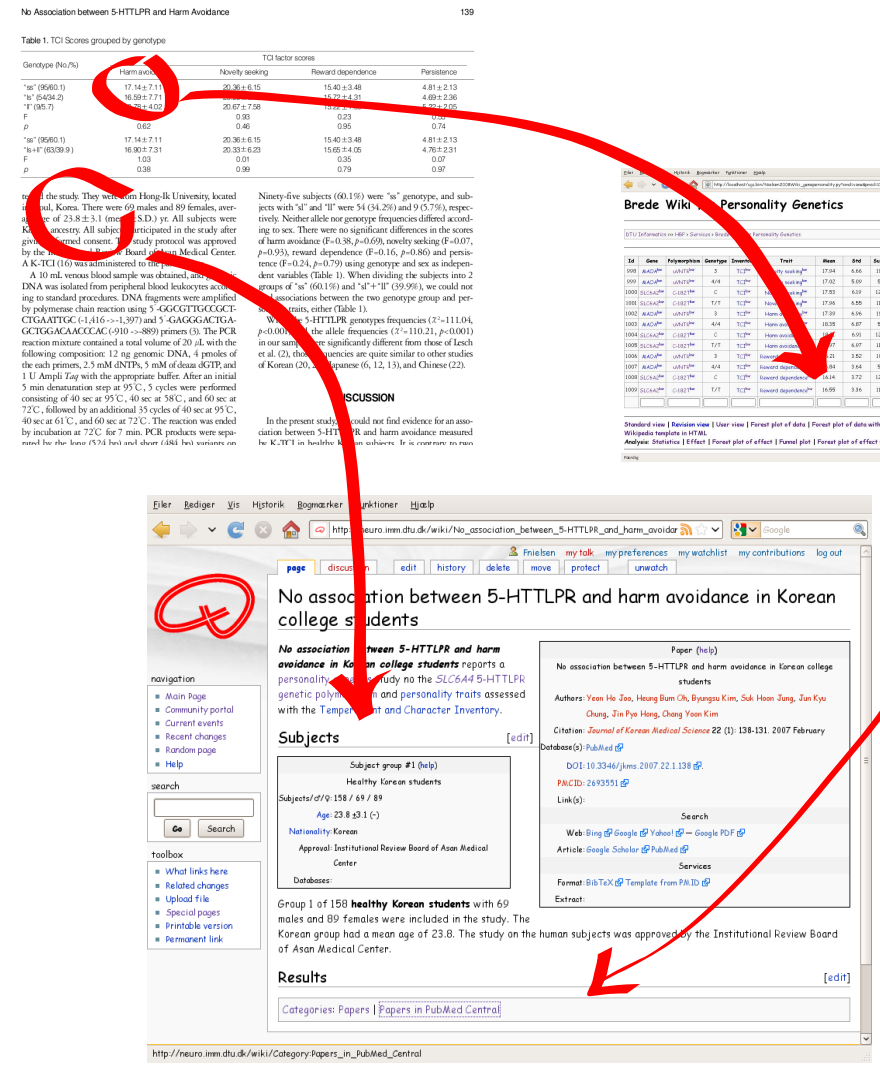


Issues

Contribution is difficult: Presently "raw" data entry ☹

Online interactive meta-analysis is not immediately available ☹

Brede Wiki for personality genetics



Data entry in the wiki in a table-like interface: Gene, polymophism, genotype, inventory, trait, personality scores, subject group, PMID.

"Normal" Brede Wiki keeps track of data entry.

Data can also be exported to the Brede Wiki.

So far typed in data from 87 studies with 2815 personality scores.

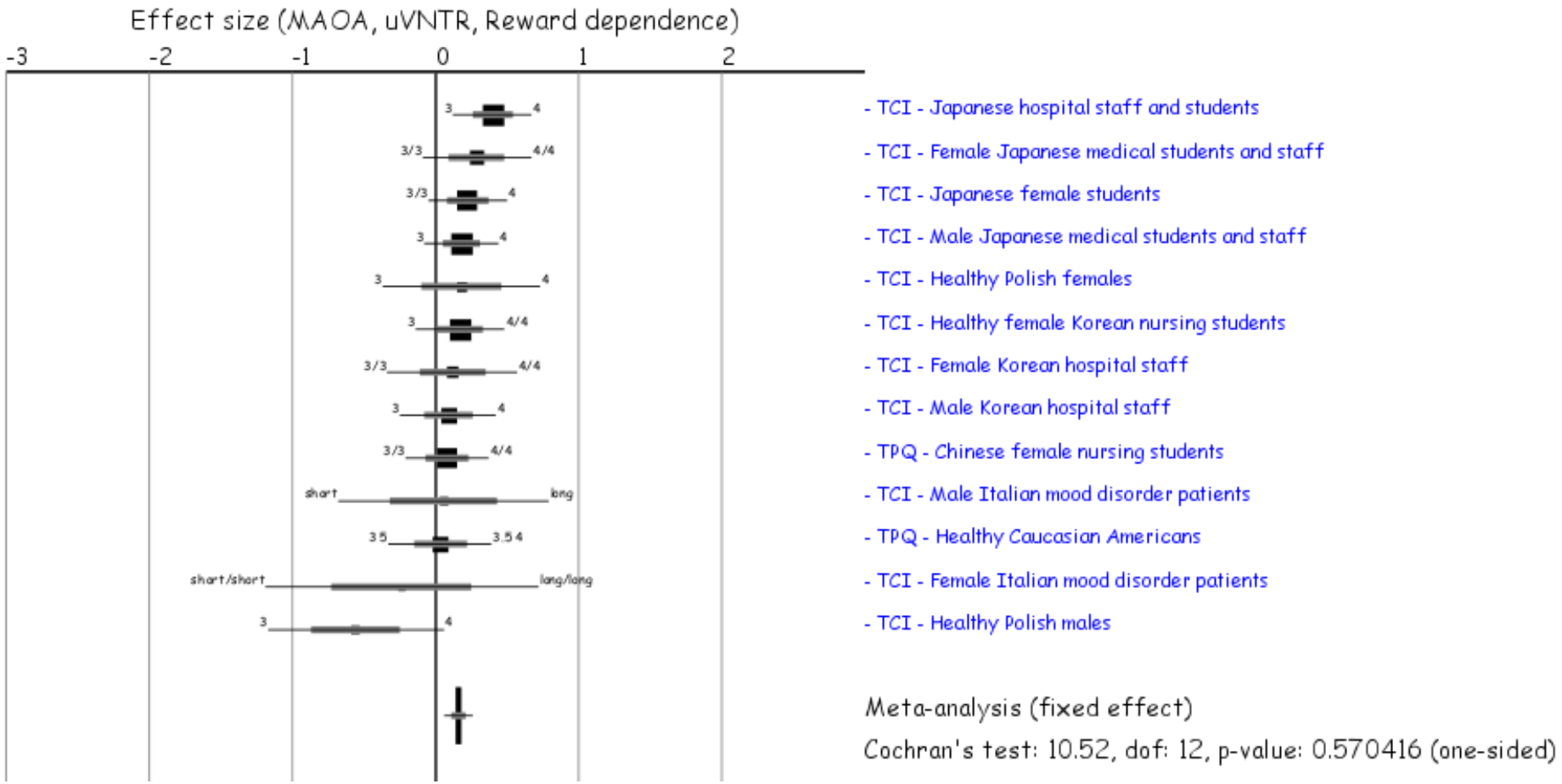
Meta-analysis across traits and polymorphisms

	Effect	Std	P	Studies	Subjects	Gene	Polymorphism	Trait
1	0.854	0.223	0.00013	2	107	ESR1	TA repeat	Harm avoidance
2	-1.102	0.289	0.00014	2	245	HTR3A	C178T	Harm avoidance
3	-0.779	0.220	0.00039	1	90	ESR1	TA repeat	Anxiety
4	-0.445	0.135	0.00098	1	247	TH	TCAT repeat	Extraversion
5	-0.401	0.123	0.00108	1	315	DRD4	Exon 3 VNTR	Positive emotions
6	0.165	0.051	0.00118	13	1747	MAOA	uVNTR	Reward dependence
7	-0.393	0.123	0.00135	1	315	DRD4	Exon 3 VNTR	Extraversion
8	-1.355	0.427	0.00152	1	125	HTR3A	C178T	Nonconformity
9	-0.758	0.240	0.00161	1	122	SLC6A4	5-HTTLPR	Activity
10	-0.174	0.055	0.00163	16	1791	SLC6A4	5-HTTLPR	Agreeableness

Large-scale data mining across all recorded personality traits and polymorphisms and present the result on the wiki.

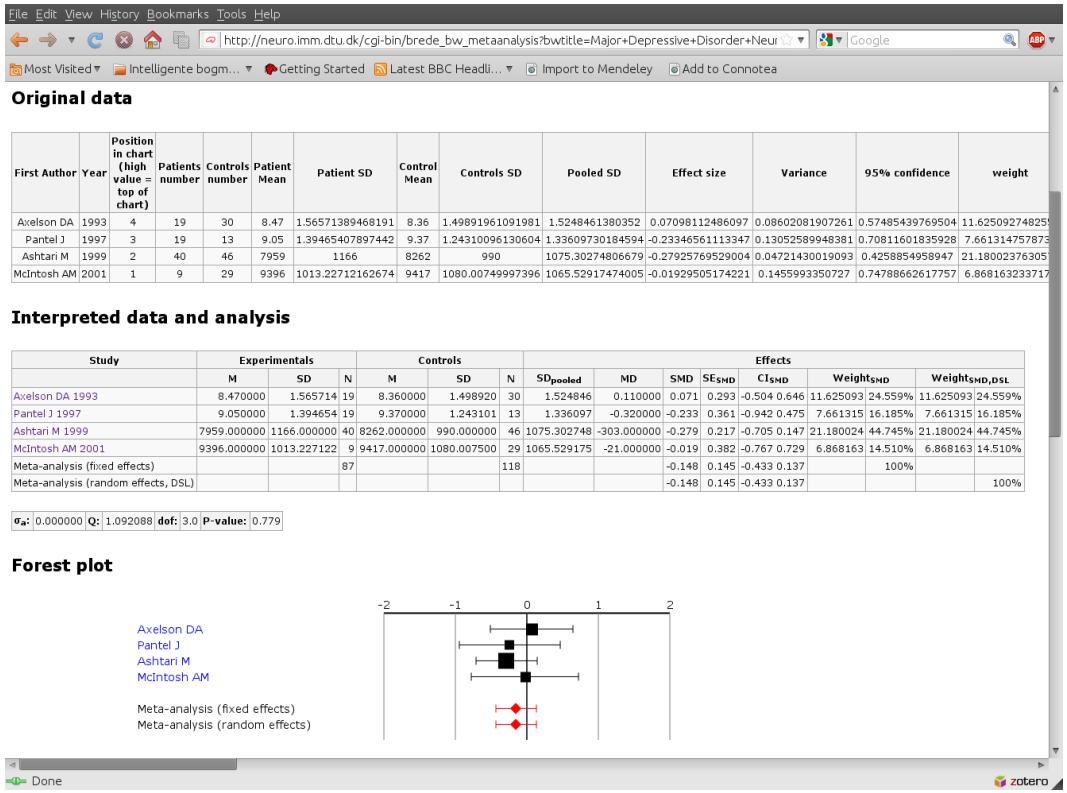
Order meta-analytic results, e.g., with respect to P-value

MAOA uVNTR/reward dependence



Forest plot generated by the wiki for the "warrior gene" and Cloninger's reward dependence with meta-analysis and Cochrane's test.

Meta-analysis with a MediaWiki-based wiki



Data sets from structural neuroimaging meta-analyses across multiple brain regions by Kempton et al. (2008, 2011)

Represent the data a comma-separated values on a separate page.

Web script extracts the data and present a meta-analysis on a Web page.

Online forest plot and funnel plots.

Data quality

Spamming is a serious problem on smaller open wikis.

On an open wiki you need daily surveillance and block vandals and delete pages

MediaWiki has numerous methods deal with spamming

MediaWiki can also restrict who is able to edit and signup

On the Brede Wiki incomplete data entry is indicated with a prominent red color

What data is incomplete is not apparent for users during search

Data quality: Outlier detection

#	Value	x	y	z	term	Lobar anatomy	WOBIB	WOEXP
1	-111.6671	-11	-76	-53	parietal	Left superior parietal lobe	176	539
2	-86.8558	-50	-52	17	right	Right occipitotemporal cortex	4	9
3	-71.9572	57	-22	13	left	Left Heschl's gyrus	164	499
4	-59.8896	-46	-25	13	right	Right superior temporal gyrus	164	499
5	-56.9524	46	40	-4	left	Left middle frontal gyrus	88	283
6	-56.4492	-11	-76	-53	superior	Left superior parietal lobe	176	539
7	-56.1964	-30	59	3	lobule	Left middle frontal lobule	88	283
8	-55.4707	-39	-33	46	right	Right intraparietal sulcus	24	77
9	-53.914	-44	-62	4	anterior	Left anterior middle temporal gyrus	104	324
10	-53.3065	40	10	0	left	Left frontal insula	17	50
11	-52.9443	-8	-32	-4	frontal	Mesolimbic, right inferior frontal gyrus	111	343
12	-51.4662	52	-30	20	cingulate	Right postcentral gyrus/posterior cingulate gyrus	35	119
13	-44.1041	-34	-48	-18	right	Right fusiform gyrus	145	443
14	-42.5693	8	-68	-35	precuneus	Right precuneus	23	73
15	-40.5044	-37	35	36	right	Right middle frontal gyrus	178	561
16	-39.5277	-44	32	9	lobule	Left inferior frontal lobule	88	283
17	-37.978	-15	-53	-2	anterior	Left lingual gyrus/anterior calcarine sulcus	3	7
18	-36.8332	-34	3	15	right	Right inferior frontal cortex	161	493
19	-34.993	-30	-40	-25	anterior	Left anterior cerebellum	180	563
20	-33.018	38	18	-26	fusiform	Fusiform gyrus	31	104

What about data entry errors and other peculiarities?

Data mining for outliers using an automated algorithm that looks at the redundancy between the anatomical label and the 3D coordinate (Nielsen and Hansen, 2002a).

Here "parietal" in "left superior parietal lobe" does not "fit" with $z = -53$ and "right" in "Right occipitotemporal cortex" does not fit with $x = -50$.

Data sharing

Data from the Brede Database distributed as XML, comma-separated values file, JSON, SKOS.

Brede Database aggregated into Antonia Hamilton's AMAT, — an thereby also SumsDB.

Federated into *Neuroscience Information Framework*

Acawiki, http://acawiki.org/, academic summaries.

Issue with license: CC-by or CC-by-sa.

The Brede Wiki available from

http://neuro.imm.dtu.dk/wiki/

Brede Database

http://neuro.imm.dtu.dk/services/brededatabase

Brede Toolbox

http://neuro.imm.dtu.dk/software/brede

Thanks!

References

Blinkenberg, M., Bonde, C., Holm, S., Svarer, C., Andersen, J., Paulson, O. B., and Law, I. (1996). Rate dependence of regional cerebral activation during performance of a repetitive motor task: a PET study. *Journal of Cerebral Blood Flow and Metabolism*, 16(5):794–803. PMID: 878424. WOBIB: 166.

Ingvar, M. (1999). Pain and functional imaging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 354(1387):1347–1358. PMID: 10466155.

Kempton, M. J., Geddes, J. R., Ettinger, U., Williams, S. C. R., and Grasby, P. M. (2008). Meta-analysis, database, and meta-regression of 98 structural imaging studies in bipolar disorder. *Archives of General Psychiatry*, 65(9):1017–1032. http://archpsyc.ama-assn.org/cgi/content/full/65/9/1017.

Kempton, M. J., Salvador, Z., Munafo, M. R., Geddes, J. R., Simmons, A., Frangou, S., and Williams, S. C. R. (2011). Structural neuroimaging studies in major depressive disorder: meta-analysis and comparison with bipolar disorder. *Archives of General Psychiatry*, 68(7):675–690. PMID: 21727252. DOI: 10.1001/archgenpsychiatry.2011.60. WOBIB: .

Lin, C.-H., Chiu, Y.-C., Cheng, C.-M., and Hsieh, J.-C. (2008). Brain maps of Iowa gambling task. *BMC Neuroscience*, 9:72. DOI: 10.1186/1471-2202-9-72.

Miles, A. and Bechhofer, S. (2009). SKOS Simple Knowledge Organization System Reference. W3C candidate recommendation, W3C, MIT. http://www.w3.org/TR/2009/CR-skos-reference-20090317/.

Nielsen, F. Å. (2003). The Brede database: a small database for functional neuroimaging. *NeuroImage*, 19(2). http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/2879/pdf/imm2879.pdf. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.

Nielsen, F. Å. (2005). Mass meta-analysis in Talairach space. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 985–992, Cambridge, MA. MIT Press. http://books.nips.cc/papers/files/nips17/NIPS2004_0511.pdf.

Nielsen, F. Å. (2009). Brede Wiki: Neuroscience data structured in a wiki. In Lange, C., Schaffert, S., Skaf-Molli, H., and Völkel, M., editors, *Proceedings of the Fourth Workshop on Semantic Wikis — The Semantic Wiki Web*, volume 464 of *CEUR Workshop Proceedings*, pages 129–133, Aachen, Germany. RWTH Aachen University. http://ceur-ws.org/Vol-464/paper-09.pdf.

Nielsen, F. Å., Balslev, D., and Hansen, L. K. (2005). Mining the posterior cingulate: Segregation between memory and pain component. *NeuroImage*, 27(3):520–532. DOI: 10.1016/j.neuroimage.2005.04.034. Text mining of PubMed abstracts for detection of topics in neuroimaging studies mentioning posterior cingulate. Subsequent analysis of the spatial distribution of the Talairach coordinates in the clustered papers.

Nielsen, F. Å. and Hansen, L. K. (2002a). Finding related functional neuroimaging volumes. *NeuroImage*, 16(2). http://www.imm.dtu.dk/~fn/ps/Nielsen2002Finding_abstract.ps.gz. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom.

Nielsen, F. Å. and Hansen, L. K. (2002b). Modeling of activation data in the BrainMap^{TM} database: Detection of outliers. *Human Brain Mapping*, 15(3):146–156. DOI: 10.1002/hbm.10012. http://www3.interscience.wiley.com/cgi-bin/abstract/89013001/. CiteSeer: http://citeseer.ist.psu.edu/nielsen02modeling.html.

Nielsen, F. Å. and Hansen, L. K. (2004). Finding related functional neuroimaging volumes. *Artificial Intelligence in Medicine*, 30(2):141–151. PMID: 14992762. http://www.imm.dtu.dk/~fn/Nielsen2002Finding/.

Szewczyk, M. M. (2008). Databases for neuroscience. Master's thesis, Technical University of Denmark, Kongens Lyngby, Denmark. http://orbit.dtu.dk/getResource?recordId=223565&objectId=1&versionId=1. IMM-MSC-2008-92.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, 16(3 part 1):765–780.

PMID: 12169260. DOI: 10.1006/nimg.2002.1131. http://www.sciencedirect.com/science/article/-B6WNP-46HDMPV-N/2/xb87ce95b60732a8f0c917e288efe59004.

Turner, J. A. and Laird, A. R. (2011). The Cognitive Paradigm Ontology: design and application. Neuroinformatics. In press.

Wilkowski, B., Szewczyk, M., Rasmussen, P. M., Hansen, L. K., and Nielsen, F. Å. (2009). Coordinate-based meta-analytic search for the SPM neuroimaging pipeline. In *Proceedings of the Second International Conference on Health Informatics*, pages 11–17. INSTICC Press.