DMPonline The DCC Data Management Planning Tool

Cortical map development in a synthetic littermate

Funding body: Sir Henry Wellcome Postdoctoral Fellowship
Lead organisation: The University of Edinburgh
Project dates: 01 Oct 2012 to 30 Sept 2016

1. What data outputs will your research generate and what data will have value to other researchers?
Animal behavioural experiments will record data from sensors (video cameras and microphones) both external to litters of huddling laboratory rodents and internally via a novel synthetic littermate ("surrogate") sensing device. These data will be presented to animals as sensory stimuli in later electrophysiology experiments to test the predictions of neural network simulations. A full behavioural experiment will produce audio and visual data files, e.g WAV, AVI. Lab-book style metadata will be created for each session. Electrophysiological data will consist of neuronal spike trains recorded from multielectrode arrays and may include optical imaging videography data.

Data sets will not be self-explanatory in isolation but behavioural data will be self-explanatory in the context of metadata describing the design of the surrogate device used to collect them. Original data will be stored as uncompressed movie and sound files and will be automatically aligned. Aligned files will be auto-generated at compressed rates for quick inspection and web-based sharing. Electrophysiology data will be managed separately and will be stored with extensive metadata describing, for example, electrode recording position and depth in anatomical coordinates, anaesthetic doses, explicit links to sensory stimulus source data files etc. Contextualizing summary statistics will be stored alongside.

Large-scale neural network simulations will be run using the Topographica Neural Map Simulator[1] and the BRain And Head Modelling System[2]. Simulation results will be stored in compressed formats as matrices of pre- and post-simulation synaptic connection weights, along with a copy of the top-level script used to specify the algorithm, stamped with the subversion repository tags appropriate to the software release.

Downsized audio-visual data files will be shared as they are created with the project research sponsors and collaborators via a password protected online repository (via existing SAMBA front-end to dedicated departmental servers at Sheffield). Access to neural data collected in experiments in Paris will be made available (beyond the project partners and collaborators) at the discretion of Professor Shulz as research sponsor providing specialist equipment and supervision. A database of audio-visual data files from the behavioural experiments will be made generally available via two major online releases at the end of phases 2 and 3 of the project, with the intention that they constitute the first standard reference database of natural multisensory experiences.

In addition to the project sponsors and collaborators, organisations most likely to reuse the dataset include primary developmental psychobiologists, machine-vision, machine-learning and computational neuroscience groups, as well as biomimetic and developmental robotics groups. The dataset of surrogate littermate multisensory experiences aims to do for models of multisensory models what existing standard 'natural image' databases currently provide for biological and computer vision projects.

2. When will you share the data?
A general web-based release of a *database of naturalistic multisensory input patterns* will be scheduled for the end of research phase 2 (April 2015) and research phase 3 (September 2016). Electrophysiology data will be made available at the discretion of Prof Shulz after journal publication. Publications will be targeted to journals complying with the Wellcome Trust open access mandate. The applicant reserves the right to use the data before opening it up to wider use by not releasing video to the general community until the modelling key stage in that phase is complete (no longer than 6 months).

3. Where will you make the data available?
Behavioural audio-visual data routinely maintained at Sheffield will made available to the research sponsors and collaborators via an established web-based interface to the departmental servers. In the longer term, the database of naturalistic multisensory input patterns and electrophysiological data collected in Paris will be made publicly available through the 'Edinburgh DataShare' data repository service[3]. Supplementing journal outputs, simulation metadata scripts will also be made available via ModelDB[4].

[1] http://www.topographica.org
[2] http://www.brahms.sourceforge.net
[3] http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/data-repository

4. How will other researchers be able to access the data?
Data housed on the University of Sheffield departmental servers will be available to the research sponsors and collaborators via a private web-based interface already in place (password and University firewall protection). Discretionary password access may be granted to third parties by the applicant. The database of naturalistic multisensory input patterns and the electrophysiology data will be made publicly available (i.e., searchable via Google) using provisions offered by the Edinburgh Data Library.

5. Are any limits to data sharing required - for example, to either safeguard research participants or to gain appropriate intellectual property protection?
No limits to sharing data are required. No data will be "personal data" in terms of the Data Protection Act (1998) or equivalent legislation outside the UK. There are unlikely to be any embargo periods for political/commercial/patent reasons.

6. How will you ensure that key datasets are preserved to ensure their long-term value?
The Edinburgh Data Library will be responsible for ongoing maintenance after the end of the project (free of charge under their current terms) and will make data and resources available dependent on access. Descriptive metadata and documentation will enable researchers to understand the data and the data itself will be stored in open source file formats where practicable. In most cases data will be available in more than one format.

7. What resources will you require to deliver your plan?
The strategies outlined will not incur direct costs as they will be utilising methods and resources already in place at Sheffield and Edinburgh. A contribution proportionate to the costs of short-term maintenance of data on University hosted servers in Sheffield may be reclaimed under the research costs of the fellowship. Currently data is centrally held on a NetApp FAS3160 fileserver and is chargeable at £400 per Terabyte per year. It is anticipated that a total of X Terabytes of storage will be needed at a cost of £XXX. This will ensure that multiple backups of data exist in several physical locations. The applicant will manage data backup as part of the day-to-day workload and will make use of scripted backup and verification procedures recommended by Sheffield.

Signature		Date	
Print name		Role	

Signature		Date	20/01/2012
Print name	Dr Richard R. Plant	Role	JISC Data Management Planning and Storage for Psychology project officer

[4] http://senselab.med.yale.edu/modeldb