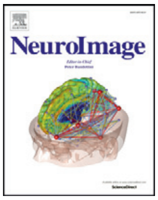
NeuroImage 144 (2017) 270–274

Contents lists available at ScienceDirect

NeuroImage

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



The Brain Analysis Library of Spatial maps and Atlases (BALSA) database

David C. Van Essen *, John Smith, Matthew F. Glasser, Jennifer Elam, Chad J. Donahue, Donna L. Dierker, Erin K. Reid, Timothy Coalson, John Harwell

Neuroscience Department, Washington University School of Medicine, St. Louis, MO, United States

ARTICLE INFO

Article history:
Received 18 December 2015
Revised 29 March 2016
Accepted 1 April 2016
Available online 10 April 2016

Keywords:
Neuroimaging
Neuroanatomy
Connectivity
Human
Nonhuman primate

ABSTRACT

We report on a new neuroimaging database, BALSA, that is a repository for extensively analyzed neuroimaging datasets from humans and nonhuman primates. BALSA is organized into two distinct sections. BALSA Reference is a curated repository of reference data accurately mapped to brain atlas surfaces and volumes, including various types of anatomically and functionally derived spatial maps as well as brain connectivity. BALSA Studies is a repository of extensively analyzed neuroimaging and neuroanatomical datasets associated with specific published studies, as voluntarily submitted by authors. It is particularly well suited for sharing of neuroimaging data as displayed in published figures. Uploading and downloading of data to BALSA involves 'scene' files that replicate how datasets appear in Connectome Workbench visualization software. Altogether, BALSA offers efficient access to richly informative datasets that are related to but transcend the images available in scientific publications.

BALSA (Brain Analysis Library of Spatial maps and Atlases) is a new database dedicated to hosting extensively analyzed neuroimaging and neuroanatomical datasets. BALSA (http://balsa.wustl.edu) serves dual purposes, complementary to one another and to what is provided by existing neuroimaging databases (see *Relationship to Other Resources*). BALSA Reference is a curated repository of reference data accurately mapped to brain atlas surfaces and volumes. It includes various types of anatomically and functionally derived spatial maps, such as cortical and subcortical parcellations, myelin maps, and retinotopic maps. It also includes connectivity data derived from retrograde tracers (macaque), plus tractography (from diffusion imaging) and resting-state functional connectivity (human and macaque) in a well-defined spatial coordinate system for each species. BALSA Studies will enable sharing of extensively analyzed neuroimaging and neuroanatomical datasets associated with published studies, as voluntarily submitted by authors. It is particularly well suited for sharing of neuroimaging data as displayed in published figures but can include many ancillary files and data representations.

1. Human Connectome Project data

A major source of data for both BALSA Reference and BALSA Studies will come from the Human Connectome Project (HCP; http://humanconnectome.org; Van Essen et al. 2013). The HCP has acquired, analyzed, and shared high quality multimodal neuroimaging data from a large number of healthy young adults using cutting-edge preprocessing methods (Glasser et al. 2013) that include improved intersubject alignment based on a new Multimodal Surface Matching registration method (Robinson et al. 2014; Glasser et al., in revision). A growing number of studies are analyzing HCP data using the Connectome Workbench software platform (http://www.humanconnectome.org/software/connectome-workbench.html), which includes many features that capitalize on key aspects of the HCP data. This includes CIFTI 'grayordinate' datasets, which efficiently represent cortical surface vertices and subcortical gray matter voxels in a standardized format. Much of the extensively analyzed HCP data emerging from such studies is well suited for BALSA, in contrast to the unprocessed and minimally preprocessed HCP data from 1200 individual subjects that are better handled by the ConnectomeDB database (http://connectomedb.org; Marcus et al. 2013; Hodge et al. 2016).

The HCP data in BALSA Reference will include extensive group average data, including cortical surfaces and structural MRI volumes; and maps of cortical thickness, myelin, folding, sulcal depth, task-fMRI activations, and fMRI-based resting state networks (RSNs). These are derived from the "S900" datasets released in December, 2015. Another major data component involves a 180-area per hemisphere group average cortical parcellation (Glasser et al., in revision), plus a variety of datasets mapped to this parcellation (e.g., parcellated myelin maps and task-fMRI maps). A third component will include maps of data from hundreds of individual HCP subjects contained in composite files, enabling efficient cross-subject comparisons of individual-subject parcellations, myelin maps, task-fMRI maps, etc.

HCP data to be included in BALSA Studies will include extensively analyzed data from individuals and group averages associated with specific publications, including the aforementioned cortical parcellation

---

* Corresponding author.
E-mail address: vanessen@wustl.edu (D.C. Van Essen).

http://dx.doi.org/10.1016/j.neuroimage.2016.04.002
1053-8119/

study (Glasser et al., in revision). This includes detailed evidence supporting the delineation of cortical areal boundaries using the multi-modal HCP data.

2. Nonhuman primate data

BALSA Reference includes MRI-based group average atlases of macaque — the MY19 (MacaqueYerkes19) atlas (Donahue et al., in review) – and chimpanzee – the CY29 (ChimpYerkes29) atlas. Both were generated in the Van Essen lab using preprocessing pipelines similar to those implemented for the HCP (Glasser et al. 2013, 2014). The MY19 and CY29 atlases include cortical myelin maps and thickness maps that provide useful architectonic landmarks for both species. The macaque atlas also includes a variety of cortical parcellations mapped from the F99 atlas previously used in the Van Essen lab (Van Essen and Dierker 2007), plus quantitative retrograde tracer connectivity data from the Kennedy lab (Markov et al., 2014). BALSA Studies will include tractography data from postmortem Old World monkey brains (Donahue et al., in revision).

3. 'Scene files' for data upload, preview, download, and visualization

Many of the data files available in BALSA Reference and BALSA Studies are maximally informative when displayed in appropriate combinations, such as a cortical area parcellation displayed on an appropriate cortical surface mesh and overlaid on complementary data types (e.g., myelin maps and task-fMRI maps). For many published figures this can involve complex, multilayered, and multipanel displays. For these and other reasons, BALSA is organized around 'scene files' that facilitate efficient data upload, previewing, download, and internal data management. A scene file contains any number of individual scenes; each scene enables exact replication of the spatial configuration and data overlays used in a given data display (e.g., zooming, labeling, thresholding, and color palettes), whether it is simple or highly complex and whether it represents a reference dataset or a precise replication of a published figure.

Scene files were originally implemented for Caret software and are extensively used in the SumsDB database (http://sumsdb.wustl.edu) (Van Essen et al. 2001; Dickson et al. 2001). BALSA capitalizes on numerous improvements in scene file capabilities that have been implemented for Connectome Workbench (version 1.2 and higher), including text annotations and multi-panel displays such as that illustrated in Fig. 1 for four human cortical parcellation schemes mapped to the same HCP atlas surface. Other scenes in the same scene file show left and right hemisphere views of the individual parcellations with each cortical area (or RSN) having an attached label (i.e., that rotates/pans/zooms along with the surface) to facilitate easy identification of areas.

4. Data modalities, formats, and availability

BALSA became publicly accessible in the spring of 2016. Early content for BALSA Reference includes group average volume and surface data from the aforementioned MRI-based atlases of macaque, chimpanzee, and human brains. For BALSA Studies, early content will include studies from the Van Essen lab comparing tractography to tracers in the macaque (Donahue et al., in revision) and the aforementioned human cortical parcellation based on HCP data (Glasser et al., in revision).

Imaging modalities supported by BALSA (i.e., with relevant metadata extracted as described below) include structural MRI (T1w, T2w scans), task-fMRI, resting-state fMRI, and diffusion imaging. Many derived data types will also be supported, including cortical surfaces, myelin maps and thickness maps, resting-state networks, and tractography results. In addition, the macaque datasets include quantitative retrograde tracer data at an area-to-area granularity.

Although BALSA is designed to store and manage neuroimaging data, extensive individual-subject behavioral and demographic data for all of the HCP datasets is available in the ConnectomeDB database (https://db.humanconnectome.org/).

5. Standard and nonstandard neuroimaging data formats

BALSA can handle a variety of standard neuroimaging data formats, including NIFTI volume data, GIFTI surface data, and CIFTI 'grayordinate' data (surface vertices plus volume voxels, Glasser et al. 2013). BALSA also supports other Workbench file formats that have proven useful. Besides the aforementioned scene files, there are 'border' files, 'foci' files, and several tractography-related files that are not currently standardized across neuroimaging platforms.

6. File identifiers and citable URLs

Every dataset, scene, and individual data file uploaded to BALSA will be assigned a unique, permanent BALSA identifier (an "OID") and a related unique URL. For example, each scene has a URL (http://balsa.wustl.edu/[OID]), allowing it to be accessed directly (e.g., by including the URL in a publication, as in Fig. 1 above). In the converse direction, each scene in a scene file, and its corresponding webpage in BALSA, can include a PMID and/or doi that links it directly to a relevant publication. These bidirectional linkages emulate features previously introduced in SumsDB (e.g., Van Essen et al. 2012a, 2012b).

Each file is also associated with file metadata that includes a "file path" (filename along with the directory structure used when the scene was uploaded) along with selected metadata extracted from the file contents (including provenance metadata that Connectome Workbench adds to output files); this information is stored in a unique "file metadata ID", distinct from the file OID. When a file with exactly the same contents (based on hash comparisons) is uploaded as part of a different scene file, BALSA will use the original OID and data file (thus avoiding inefficient data storage and saving space) but will link to that file using the file's metadata ID associated with the new scene file.

7. Searchable metadata tags

Individual files in BALSA are associated with searchable metadata tags, in which each tag has a category and a value (as "category:value" pairs). Tags can be assigned automatically by a scanning process that uses "handles" (prespecified text strings) to scan the file name, file path, and within-file metadata whose fields are customized for each file type. Each scene aggregates the tags of all its supporting files, and each scene file aggregates the tags for its constituent scenes. Importantly, tags for both scenes and files can be edited (by curators or data providers), to handle cases where the scanning process erroneously assigned tags or failed to assign appropriate tags. Currently implemented categories include: species, modality, stereotaxic space (based on the atlas template), volume registration method, surface geographic convention (e.g., HCP's Conte69_fs_LR, FreeSurfer's fsaverage, or macaque MY19 or F99), surface resolution (e.g., HCP's 164k or 32k meshes), parcellation (e.g., HCP_MMP1.0); we anticipate incorporating additional categories as the diversity of uploaded datasets grows and user feedback is obtained regarding what is most useful.

8. Data curation

The BALSA upload process requires that each data submitter set up an account and provide contact information as well as essential information about the uploaded dataset. This includes specifying any constraints on data sharing (e.g., HCP Open or Restricted Access Data Use Terms) and also providing an assurance that any protected health information (PHI) has been redacted from any human data. Once an archive is uploaded and processed in BALSA, an automatic notification will be