THE ROYAL
SOCIETY

doi 10.1098/rstb.2001.0913

'The surface management system' (SuMS) database: a surface-based database to aid cortical surface reconstruction, visualization and analysis

James Dickson, Heather Drury and David C. Van Essen*

Department of Anatomy and Neurobiology, Washington University School of Medicine, St Louis, MO 63110, USA

Surface reconstructions of the cerebral cortex are increasingly widely used in the analysis and visualization of cortical structure, function and connectivity. From a neuroinformatics perspective, dealing with surface-related data poses a number of challenges. These include the multiplicity of configurations in which surfaces are routinely viewed (e.g. inflated maps, spheres and flat maps), plus the diversity of experimental data that can be represented on any given surface. To address these challenges, we have developed a surface management system (SuMS) that allows automated storage and retrieval of complex surface-related datasets.

SuMS provides a systematic framework for the classification, storage and retrieval of many types of surface-related data and associated volume data. Within this classification framework, it serves as a version-control system capable of handling large numbers of surface and volume datasets. With built-in database management system support, SuMS provides rapid search and retrieval capabilities across all the datasets, while also incorporating multiple security levels to regulate access. SuMS is implemented in Java and can be accessed via a Web interface (WebSuMS) or using downloaded client software. Thus, SuMS is well positioned to act as a multiplatform, multi-user 'surface request broker' for the neuroscience community.

Keywords: cortex; surface; database; neuroanatomy; magnetic resonance imaging

1. INTRODUCTION

Over the past decade there has been an explosion of experimental data pertaining to the structure, function and development of the central nervous system. This has led to an acute need for better methods of organizing, storing and accessing the ever-increasing flood of information (Pechura & Martin 1991; Shepherd *et al.* 1998; Chicurel 2000). Designing and implementing the requisite neuroinformatics methods has proven to be a major challenge for a variety of technical reasons. Among the most important of these are the tremendous complexity of nervous system structure, its variability from one individual to the next, and the great diversity in the types of experimental data and in the spatial and temporal resolution with which they are acquired.

One basic need is to have an appropriate spatial framework around which to organize experimental data. For the cerebral cortex, which is the dominant structure of the mammalian brain and the focus of the present report, it is important that this spatial framework should respect three underlying facts about cortical structure and organization: (i) the cortex is a sheet-like structure that is folded into complex and irregular convolutions in many species, especially in humans; (ii) the pattern of cortical convolutions varies markedly from one individual to the next; and (iii) there is also variability in the mosaic of functionally specialized cortical areas arrayed across the cortical sheet; each cortical area varies substantially in size, shape and geographical location from one individual to the next (Maunsell & Van Essen 1987; Rademacher *et al.* 1993; Lewis & Van Essen 2000; Amunts *et al.* 1999, 2000; Geyer *et al.* 2000).

These fundamental characteristics of cerebral cortex provide a compelling motivation for the use of explicit surface reconstructions that capture the shape of the convolutions in each experimental hemisphere under investigation. Until recently, though, the widespread use of cortical surface reconstructions in neuroscience was impeded by the lack of automated methods for generating accurate surface reconstructions and manipulating these surfaces into different configurations. This situation has changed dramatically in the past few years, and several software packages are now available that allow automated cortical surface reconstruction and flattening. These include MRGRAY (Wandell *et al.* 2000), FREESURFER (Dale *et al.* 1999; Fischl *et al.* 1999*a*), BRAIN VOYAGER (Goebel *et al.* 1998; Goebel 2000) and the SUREFIT–CARET combination from our laboratory (Van Essen & Drury 1997; Van Essen *et al.* 1998, 2001*a*). Consequently, it is probable that surface-based analyses will become a standard approach for neuroimaging and neuroanatomical studies of the cerebral cortex. The output of new surface reconstructions emanating from numerous laboratories

---

*Author for correspondence (vanessen@v1.wustl.edu).

© 2001 The Royal Society

around the world could approach or exceed 100 000 per year in the current decade (see § 4).

Surface representations are particularly well-suited for comparing experimental results obtained from different individuals. One promising way to make such comparisons uses surface-based warping to bring one hemisphere into register with another (Joshi 1997; Drury *et al.* 1999; Van Essen *et al.* 1997, 1998; Fischl *et al.* 1999*b*). Because this approach inherently respects the topology of the cortical sheet, surface-based warping has a major advantage over the many volume-based approaches that have been used to address the registration problem (see Toga 1999). For any pair of surfaces (a source and a target map), there may be any number of surface-based deformations that differ in the deformation algorithm used or in the geographical and/or functional constraints used in aiming for good registration. From a neuroinformatics perspective, dealing with surface-based deformations poses special challenges because of the need to track a complex body of information that includes key characteristics of the source, target and deformed surfaces.

Surface-based warping can be used to map structural and/or functional data from many individuals onto a common substrate, or atlas. By expressing results on the atlas probabilistically rather than deterministically, a surface-based atlas can preserve information about individual variability of different structural and functional subdivisions and about uncertainties associated with deformations from individual maps to the atlas (Drury *et al.* 1999; Fischl *et al.* 1999*b*; Van Essen *et al.* 2001*b*). While much remains to be done in developing a rigorous framework for this approach, it is probable that probabilistic surface-based atlases will become increasingly valuable repositories for many different types of information about cortical organization, function and development.

To help cope with the impending flood of experimental data represented on individual surfaces and on surface-based atlases, we have developed a surface management system (SuMS). A number of performance requirements guided the design and implementation of SuMS. Among our major objectives were that SuMS should:

(i) store surfaces and surface-related data in a sensibly organized data structure;
(ii) store volume data and other spatial data relevant to the surface reconstruction and analysis process;
(iii) store data related to surface-based deformations and surface-based atlases;
(iv) include efficient strategies for identifying and retrieving data files of interest based on a variety of search criteria;
(v) incorporate an efficient method for automated data entry;
(vi) interface with readily available software for surface reconstruction and visualization, and incorporate multiple levels of security and access control;
(vii) be accessible via either a Web interface or previously downloaded client software;
(viii) be readily extendable to multiple database sites.

SuMS meets these design objectives with an architecture that comprises four main components: (i) a database structure designed specifically for cortical surface representations; (ii) a high-capacity, secure data cache capable of storing the rapidly growing amounts of data (iii) a database manager (SYBASE structured query language (SQL) server) for coordinating the file cache and database access; and (iv) search and retrieval of stored data via a Web browser (WebSuMS) or by downloadable SuMS Client software.

2. DESIGN CONSIDERATIONS

SuMS is primarily a database for the storage and retrieval of surfaces and surface-based experimental data, but it can also store a variety of associated volume-based data. Data entry and retrieval from SuMS requires familiarity with several basic aspects of surface and volume representations. Accordingly, it is useful to briefly review the fundamentals of surface reconstruction, representation, transformation and visualization.

(a) *Surface and reconstruction substrates*

Surface reconstruction begins with the acquisition of image data suitable for visualizing the shape of the cortical convolutions in the particular hemisphere under consideration (figure 1*a*). The structural image data may be obtained from structural magnetic resonance imaging (MRI), cryosection images of the cut brain surface or stained histological sections. Structural MRI and cryosection data are particularly well-suited for automated surface reconstruction and are the main focus of our discussion.

Numerous methods are now available for cortical surface reconstruction. The methods currently in widest use (see § 1) involve an initial sequence of image-processing steps applied to the volumetric image data. This yields a segmentation (black–white volume) whose boundary demarcates the shape of the cortex. For example, figure 1*b* shows a slice through a segmented volume generated by the SUREFIT method, in which the segmentation boundary runs midway through the cortical thickness. (This choice of target layer accurately represents the associated cortical volume in both gyral and sulcal regions.) The segmented volume, which we identify as the 'reconstruction substrate', provides the template for a subsequent step of explicit surface reconstruction (figure 1*d*). This entails generating a wire-frame mesh (tessellation) whose nodes lie along the boundary of the reconstruction substrate, using a method such as the Marching Cubes algorithm (Lorensen & Cline 1987).

An alternative approach to surface reconstruction involves tracing contours along the cortical extent in a sequence of images (figure 1*c*). This is generally done by manual tracing, although automated tracing is feasible if image quality is high enough. Once the contours are brought into register, nodes on adjacent contours can be linked together to form a wire-frame tessellation using a method such as the NUAGES algorithm (Geiger 1993). Thus, the reconstruction substrate for the contour-tracing method is a sequence of contours rather than a segmented volume, and the surface is reconstructed by a different method, but the end product is fundamentally the same: a wire-frame surface reconstruction.

Surface reconstructions have many advantages for visualization and data analysis, but they are imperfect and incomplete representations of cortical shape. The

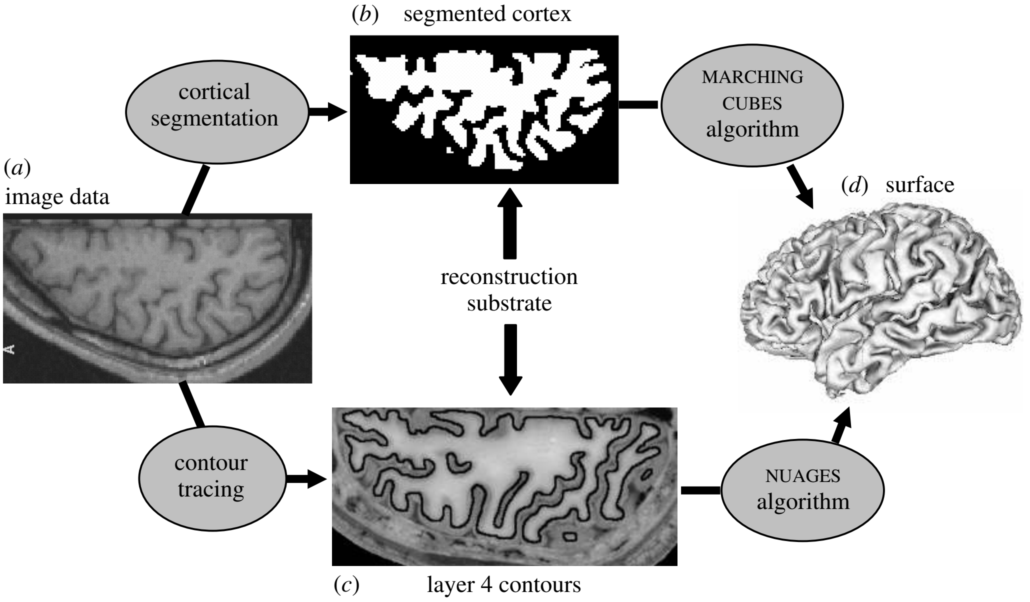


Figure 1. Generation of reconstruction substrates and explicit surface representations from image data. (*a*) A structural MRI image from the left hemisphere of a human brain (a horizontal slice through dorsal cerebral cortex of Case Demo.L). (*b*) The process of cortical segmentation leads to a segmented (black–white) volume whose boundary represents the shape of the cerebral cortex. In this case, the SUREFIT method was used to segment the volume, and the boundary of the reconstruction substrate runs approximately along the mid-thickness (layer 4) of the cortex, i.e. halfway between the inner and outer cortical boundaries. (*c*) An alternative method for generating a reconstruction substrate is to draw contours along the desired trajectory on each of a series of sections through the cortex. In this case, the contours are drawn along layer 4 on a different hemisphere (cf. Van Essen & Drury 1997) in which the image data came from cryosection photographs of the cut brain surface. (*d*) An explicit surface reconstruction can be generated from a segmented volume using an algorithm such as MARCHING CUBES, and an analogous surface reconstruction can be generated from section contours using the NUAGES algorithm. This particular surface was generated from the volume segmentation.

imperfections arise because noise, distortions and other irregularities in the image data inevitably imply some degree of uncertainty and inaccuracy in each surface reconstruction. These inaccuracies sometimes lead to topological errors ('handles') in a surface reconstruction, but even knowing that a surface is topologically correct provides no assurance of whether its trajectory deviates substantially (i.e. millimetres) from that of the desired target layer in various places. The incompleteness arises because a surface is a two-dimensional representation, whereas the cortical sheet has substantial thickness that varies by about a factor of two from one region to another (*ca.* 2–4 mm in humans (Von Economo & Koskinas 1925; Sisodiya *et al.* 1999)). Thus, any given surface can at best represent the trajectory of a particular layer, or boundary, within the cortical sheet. On the other hand, methods for automatically estimating cortical thickness have recently been developed (Fischl & Dale 2000; MacDonald *et al.* 2000). These allow the full three-dimensional shape of the cortex to be represented explicitly in a surface-based framework (e.g. as a metric of thickness at each location in the cortex, or as paired surfaces representing inner and outer cortical boundaries).

Errors and uncertainties in surface reconstructions and in accompanying estimates of cortical thickness can have a large impact on subsequent stages of data analysis, yet they are difficult to quantify and may vary widely from one region to another. Accordingly, SuMS accommodates storage of the original image volume data as well as the reconstruction substrate and other key intermediate volume representations. This allows for *post hoc* evaluation of the quality of particular reconstructions, plus the prospect of regenerating higher-quality surfaces and thickness estimates using more refined methods that may emerge.

(b) *Surface configuration and topology*

Once a surface has been reconstructed, it can be modified in various ways to facilitate visualization and/or analysis of results. The most common way of representing surface geometry involves a set of discrete nodes that are linked to neighbouring nodes to form a triangulated mesh (figure 2*a*). The positions of the nodes in space defines the surface configuration (figure 2*b*), while the pattern of links defines the surface topology (figure 2*c*). The distinction between configuration and topology is important because a surface reconstruction (figure 2*d*) can be modified in shape without changing its topology (figure 2*e*), and it can be modified in topology without changing its shape (figure 2*f*). In practice, it is desirable to have rapid access to many combinations of surface configurations and topologies for any given surface reconstruction.

(c) *Volume-based and surface-based coordinate systems*

Neuroimaging data are typically acquired in a three-dimensional coordinate system whose origin lies in one corner of the image volume. For some purposes, it is advantageous to retain the original ('native') coordinate space; for other purposes it is preferable to convert (by translation and/or rotation) to a coordinate space based on standard anatomical landmarks—usually the anterior

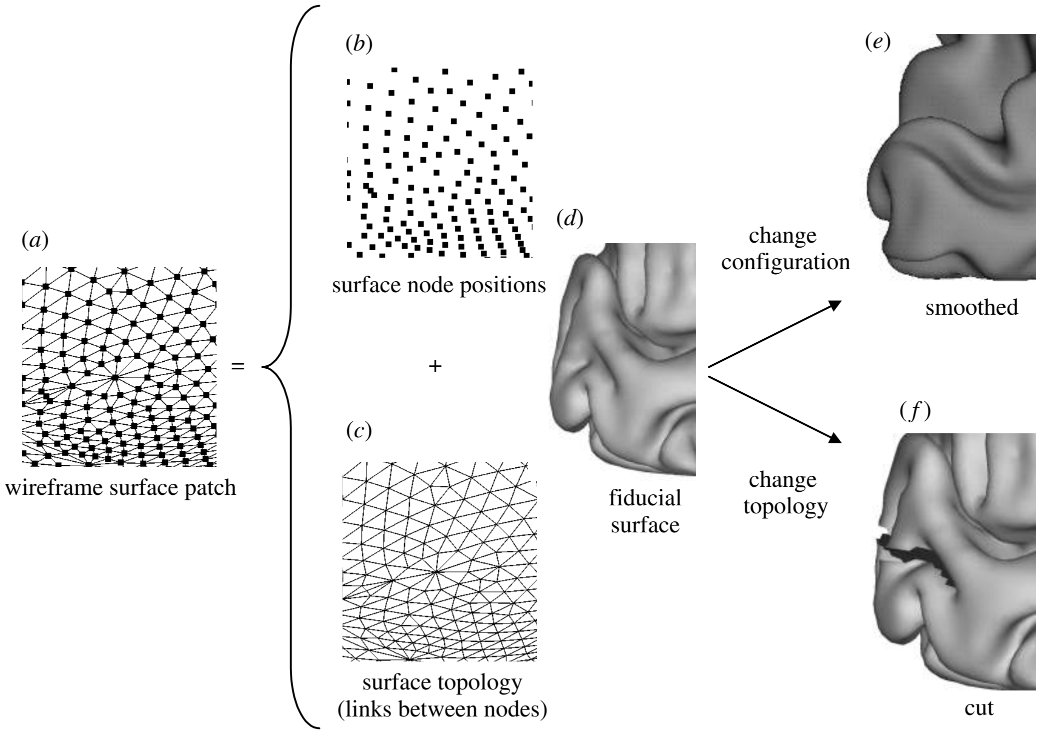


Figure 2. Surface geometry. (*a*) A schematic wire-frame surface patch. (*b*) Surface configuration is defined by the positions of the surface nodes. (*c*) Surface topology is defined by the pattern of links between nodes. (*d*) A rendered three-dimensional view of a portion of human cortex. (*e*) The configuration can be changed without altering the topology (e.g. by smoothing). (*f*) The topology can be changes without altering the configuration (e.g. by introducing cuts).

commissure (AC) and posterior commissure (PC). Once a surface has been reconstructed, it is desirable to establish surface-based coordinates that respect the topology of the cortical sheet. For many purposes, surface-based coordinates based on a spherical map (i.e. latitude and longitude) are especially attractive (Sereno *et al.* 1996; Fischl *et al.* 1999*b*; Drury *et al.* 1999). Cartesian coordinates on a flat map can sometimes be useful as well (Drury *et al.* 1996; Van Essen & Drury 1997). Accordingly, SuMS can track multiple volumetric and surface-based coordinate frames for all datasets.

(d) *Surface and volume families*

SuMS is organized around natural groupings of data, in which surfaces and/or volumes that share a common lineage are handled collectively to facilitate data entry, retrieval and visualization. This is illustrated in figure 3, which shows a related group of volume and surface representations from an exemplar human left hemisphere used as a demonstration dataset (Case Demo.L; data from the study of Corbetta *et al.* (1998)). This case includes both structural MRI data and functional MRI (fMRI) data from a particular behavioural task (involving eye movements). Each panel illustrates a commonly used display format or data type, and they are representative of the even greater diversity of data that in fact are contained within and readily accessible from SuMS.

The ordering of panels in figure 3 reflects a natural progression of processing steps involved in surface reconstruction and analysis. Application of the SureFit method to the structural image data in figure 3*a* yielded the reconstruction substrate (segmented volume) shown in figure 3*b*. The reconstruction substrate in turn was used to generate a fiducial surface (figure 3*d*) that represents the shape of the cortical sheet. The fiducial surface has been painted to reveal the distinction between exposed (gyral) cortex in light grey and buried (sulcal) cortex in dark grey. In addition, several sulci, including the central sulcus (CeS), were automatically identified and shaded in red. The fMRI data from this case can be seen in slices through the volume (e.g. figure 3*c*) and after projection onto the fiducial surface (figure 3*e*). Both formats reveal a number of activation foci but are not well-suited for seeing the overall pattern, which is highly distributed and largely buried in various sulci.

The bottom two rows of figure 3 show maps of cortical geography and fMRI activation on three reconfigured surfaces that were derived from the fiducial surface in figure 3*d*,*e* and accordingly are part of the same surface family. These include an inflated surface (figure 3*f*,*g*), a sphere (figure 3*h*,*i*) and a flat map (figure 3*j*,*k*). Inflation of the cortex smoothes out all but the deepest convolutions, while preserving a representation of cortical geography (i.e. gyri and sulci) by the shading pattern. Inflated maps have the advantage of revealing a much larger fraction of the cortical surface while retaining the shape of a lissencephalic brain, which aids in recognizing locations in the cortex. Spherical maps derive their principal advantage from their geometrical regularity, which is used to define surface-based coordinates of latitude and longitude (Sereno *et al.* 1996; Fischl *et al.* 1999*b*). By convention, the spherical map is orientated so that the ventral tip of the CeS is at the lateral pole, and the CeS has the same orientation as in the fiducial configuration (Drury *et al.* 1999). The flat map has the major advantage of showing the entire pattern of gyri and sulci (or an entire fMRI activation pattern) in a single view, with only modest distortions in surface area relative to the fiducial surface.

The relationships among the different volume and surface representations can be tracked using a marker point placed at corresponding locations in each panel. In figure 3, the marker point (yellow dot and/or white arrow) is centred on one of the fMRI activation foci in the anterior intraparietal sulcus. The marker point has coordinates (29, 67, 96) in the 'native' volume space and $(-44, -43, 47)$ when expressed in standard stereotaxic (AC–PC) coordinates. Its spherical coordinates (latitude $+30°$; longitude 160°) can be ascertained using latitude and longitude contours projected onto both the sphere and the flat map, just as with maps of the Earth's surface. Surface-based coordinates in general (and spherical coordinates in particular) provide a concise, precise and objective way to state the exact locations using a metric that respects the topology of the cortical surface. This is particularly important in regions where the folding pattern varies markedly between individuals, as is the case for most of the human cerebral cortex. In addition, the third dimension within the cortex can be represented in terms of distance along the radial axis (normal to the cortical surface).

For a database to be useful and widely used, it is important that there be no major impediments to rapid data entry, retrieval and visualization. To that end, SuMS is compatible with multiple file formats for both surface and volume data. In addition, SuMS interfaces readily with two complementary software packages developed within our laboratory: SUREFIT, which is used for volume visualization, surface reconstruction and initial surface visualization (figure 3*a*–*c*); and CARET, which is used for surface visualization, inflation and flattening (figure 3*d*–*k*). The common interface involves the use of 'specification files' that provide standardized formats for automated entry and retrieval of volume data (blue arrows) and surface data (red arrows).

3. SuMS METADATA ORGANIZATION

SuMS uses a tripartite strategy for the efficient encoding, storage and retrieval of surface-related and volume-related data from individual cases: (i) information about surfaces can include a variety of file types, each representing a qualitatively distinct set of surface properties. A primary distinction is between surface geometry (i.e. shape and position) and various types of ancillary experimental data. Similarly, volume information can include a variety of file types, representing one or another format for the reconstruction substrate and also key intermediate files, including the image data on which the segmentation is based. (ii) Important characteristics associated with each data file are stored in concise file headers (readily accessed within SuMS, SUREFIT and CARET) and in separate 'metadata' tables that are used for searching the database. (iii) Volume and surface specification files are used to organize and convey the metadata needed for automated entry and retrieval of datasets sharing a common lineage.

(a) *Case and volume metadata*

Figure 4 illustrates the hierarchical strategy used to characterize structural metadata. At the top level (figure 4*a*) are individual cases. Required attributes for each case include the individual, hemisphere, species and investigator ('owner'). Each hemisphere of any individual brain is considered a different case (because cerebral cortex in the two hemispheres consists of two physically separate domains) and is assigned a unique case identifier (Case Demo.L in this example).

The second level in figure 4 illustrates several attributes relating to particular reconstruction substrates and particular surface families. The attributes for reconstruction substrates (figure 4*b*) include information about the target cortical layer (e.g. the middle layer), the reconstruction method (e.g. SUREFIT), and the region reconstructed (e.g. the entire hemisphere). Each reconstruction substrate is assigned a version number, which is combined with the case identifier to obtain a unique identifier for the particular reconstruction (e.g. reconstruction_id = Demo. L.RECON1). The identities of the segmentation file, the structural image file, and other key intermediate files can also be included, if available. Specific file type identifiers such as those listed under 'volume data' in table 1 are used to indicate the type of data contained in each volume.

(b) *Surface-related metadata*

Surface-related data are organized around the concept of a surface family (figure 4*c*). In general, a surface family includes all surfaces that are derived from a common starting surface and contain exactly the same number of surface nodes. The node number is used to assign a unique identity for each family (surfacefamily_id, 65950 in figure 4). Within each surface family there are three primary types of surface configuration. (i) The raw surface is the initially generated surface that conforms to the precise shape of the reconstruction substrate. It typically has a rough appearance (because of the cubical shape of individual voxels if the surface is derived from a segmented volume or because of small distortions or misalignments between sections if it is derived from contour data). (ii) A fiducial surface is an improved estimate of the actual shape of the cortical sheet generated by slight smoothing of the raw surface (cf. figure 3*e* and figure 4*d*). (iii) Families of reconfigured surfaces are derived from a parent fiducial surface by extensive smoothing or more complex manipulations such as cutting and flattening the surface (cf. figure 3*h*–*l* and figure 4*d*).

Some surface characteristics are common to all members of a surface family (e.g. the identity of the reconstruction substrate from which the surface originated) (figure 4*c*). Other characteristics (including 'sampling' and 'resolution' in figure 4*c*) arise from the fact that a surface family can be derived by resampling of an existing surface. by directly generating the surface from a segmentation or other reconstruction substrate. (Resampling is used extensively, for example during cortical flattening in CARET.)

Characteristics that are specific to individual surfaces within a family include the type of topology, the coordinate frame and the configuration. For example, figure 4*d* lists several characteristics of a fiducial surface from Case Demo.L. Its topology is categorized as 'OPEN' (signifying that the perimeter runs along the natural edge of cortex on the medial wall of the hemisphere), its coordinate frame as 'AC' (signifying that the origin is at the AC) and

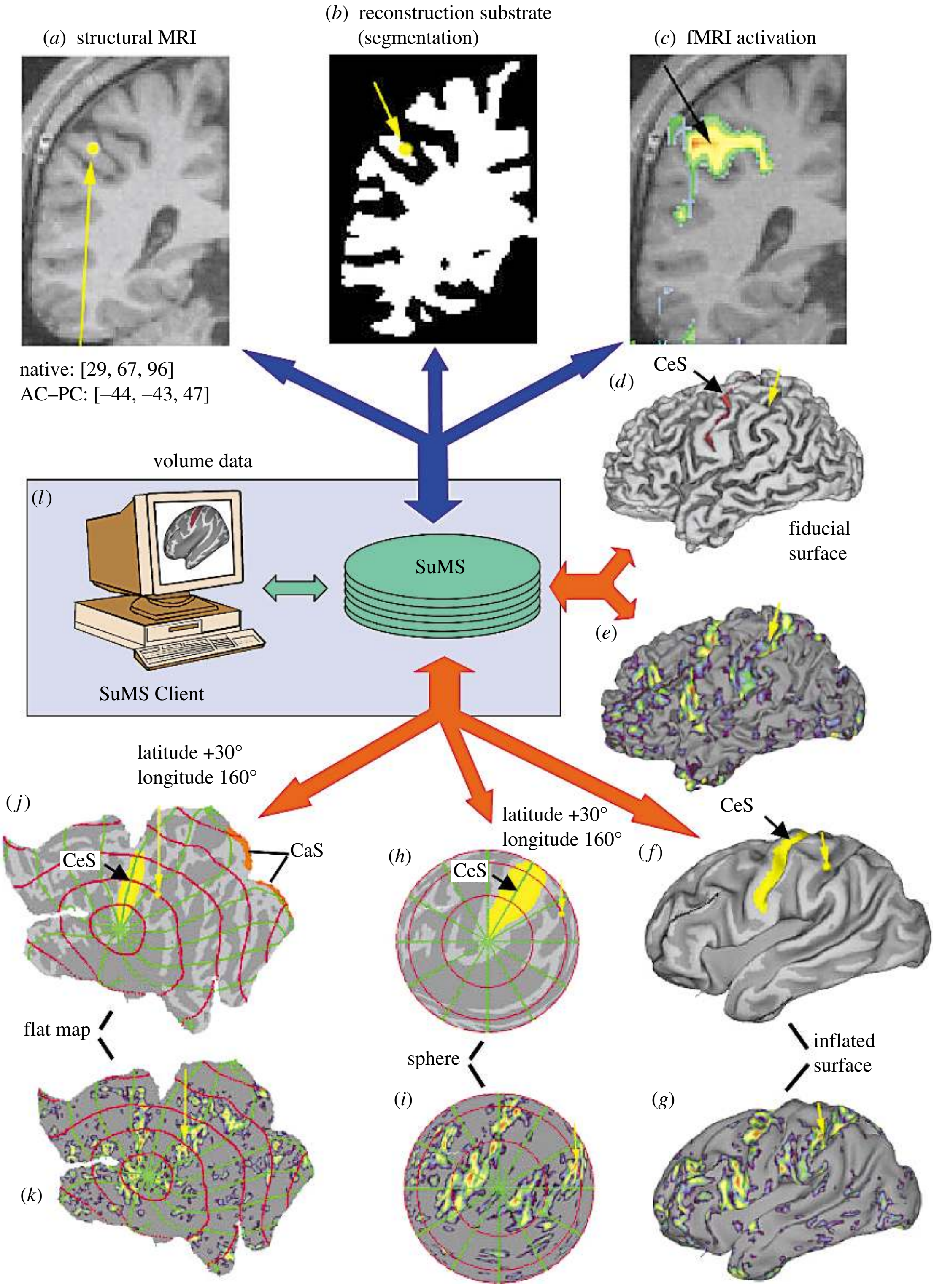
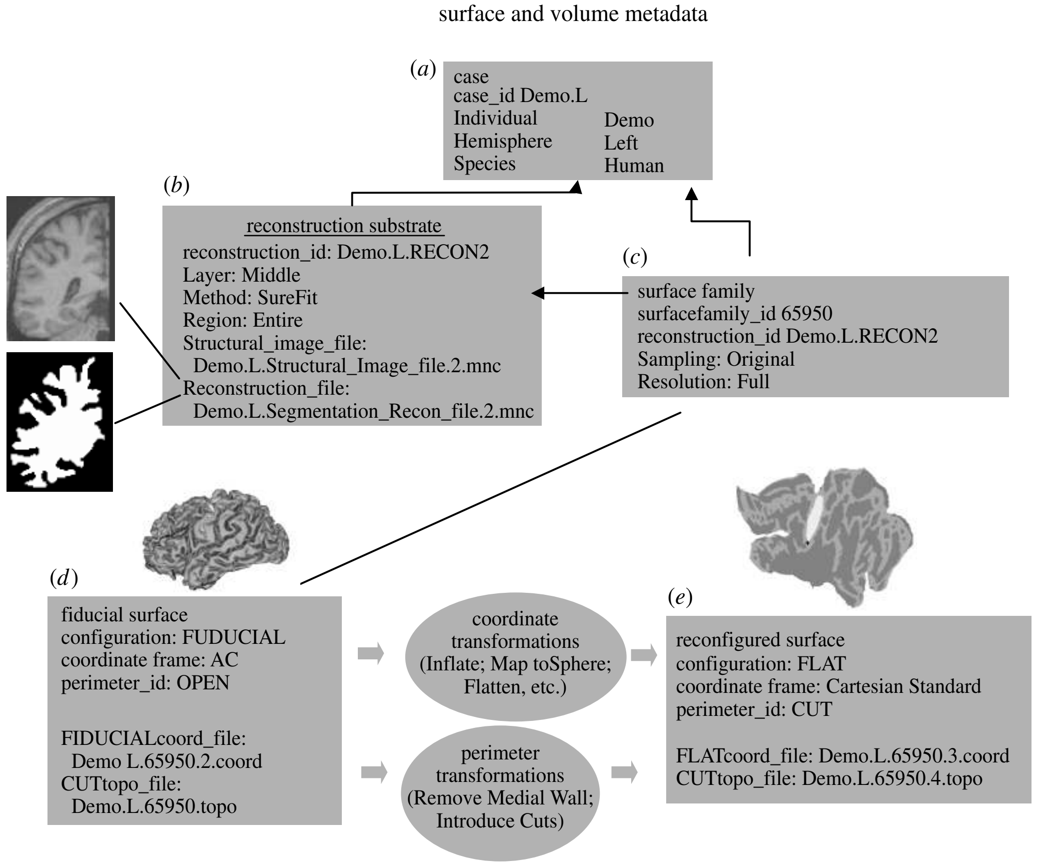


Figure 3. Volume and surface families. Each panel shows a different volume or surface configuration or dataset that is useful for surface-based analyses and which can be automatically entered into SuMS and quickly retrieved from SuMS using appropriate volume and surface specification files. (*a*) A coronal slice through a structural MRI volume (Case Demo.L). The yellow point and arrow indicate a location in the anterior intraparietal sulcus whose three-dimensional coordinates are indicated at the bottom of the panel and which is also marked in subsequent panels. (*b*) The corresponding slice through the reconstruction substrate (a segmented volume generated using the SureFit method). (*c*) The fMRI activation pattern from a paradigm involving shifting visual attention (Corbetta *et al.* 1998). (*d*) The fiducial surface generated from the reconstruction substrate shown in (*a*). The CeS is indicated in red. (*e*) The fMRI activation pattern displayed on the fiducial surface. (*f*) The pattern of cortical geography shown on an inflated map of the cortex. (*g*) The fMRI activation pattern displayed on the inflated surface. (*h*) The pattern of geography displayed on a spherical map in which distortions were reduced by multiresolution morphing (Drury *et al.* 1996, 1999) and the lateral pole was then centred on the ventral tip of the CeS. By convention, zero longitude is designated as the most anterior iso-longitude line and increasing longitude is represented by clockwise rotation for the left hemisphere (counterclockwise for the right hemisphere). The equator is designated as zero latitude, with positive latitudes representing cortex lateral to the equator and negative latitudes representing cortex medial to the equator. By these conventions,



case
case_id Demo.L
Individual Demo
Hemisphere Left
Species Human

reconstruction substrate
reconstruction_id: Demo.L.RECON2
Layer: Middle
Method: SureFit
Region: Entire
Structural_image_file:
Demo.L.Structural_Image_file.2.mnc
Reconstruction_file:
Demo.L.Segmentation_Recon_file.2.mnc

surface family
surfacefamily_id 65950
reconstruction_id Demo.L.RECON2
Sampling: Original
Resolution: Full

fiducial surface
configuration: FUDUCIAL
coordinate frame: AC
perimeter_id: OPEN

FIDUCIALcoord_file:
Demo L.65950.2.coord
CUTtopo_file:
Demo.L.65950.topo

reconfigured surface
configuration: FLAT
coordinate frame: Cartesian Standard
perimeter_id: CUT

FLATcoord_file: Demo.L.65950.3.coord
CUTtopo_file: Demo.L.65950.4.topo

Figure 4. Structural and volume metadata used for data entry and retrieval from SuMS. (*a*) Characteristics that apply to the individual case. (*b*) Characteristics of particular reconstruction substrates for a given case. (*c*) Characteristics of a particular surface family. (*d*) Characteristics of a particular fiducial surface. (*e*) Characteristics of a particular reconfigured surface (a flat map).

its configuration as 'FIDUCIAL'. Figure 4*e* lists characteristics of a flat map, which includes its topology (CUT, signifying the perimeter includes artificial discontinuities), its configuration (FLAT) and its coordinate frame (Cartesian standard, signifying that the origin is at the ventral tip of the CeS).

Many different surfaces can be generated from a combination of topologies and configurations. The matrix in figure 5 represents five standard surface configurations along the vertical axis (raw, fiducial and inflated surfaces, plus spheres and flat maps) and three types of topology along the horizontal axis (closed, open and cut topology). For each topology and configuration category, the database may contain multiple entries derived from each fiducial surface. For example, figure 5 lists two fiducial configurations—one in its native three-dimensional coordinates (useful for visualizing the surface embedded within the original MRI volume) and one in stereotaxic (AC−PC) coordinates (useful because locations are expressed relative to standard anatomical landmarks). It also lists two flat map configurations that differ in the location of cuts (one in 'Cartesian standard' coordinates, like that in figure 3*j*, and one with 'lobar cuts' in which the frontal lobe is uncut).

The four surface topologies and seven configuration or coordinate-frame types represented in figure 5 allow in principle for 28 combinatorial possibilities of surface geometry. Some combinations (six in this case, shown in open circles) are invalid, in that they do not make physical sense and would appear tangled if viewed (e.g. a flat map having a closed topology or a lobar cut topology combined with a Cartesian standard configuration). On the other hand, fiducial, inflated and spherical surfaces can generally be viewed with any topology. Altogether, 22 valid surface geometries can be readily visualized by selection of the desired combination of coordinate and topology file.

Table 1 indicates two distinct options for representing surface geometry. One is to represent surface coordinates

the marker point is at latitude +30°, longitude 160°. (*i*) The fMRI activation pattern displayed on the sphere. (*j*) A flat map generated after applying cuts in standard locations in the occipital, frontal and temporal regions (cf. Van Essen & Drury 1997). The CeS (yellow) and the calcarine sulcus (CaS) (orange) were automatically identified and are shaded on the flat map. The coordinates of the marker point can be expressed in terms of latitude and longitude mapped from the sphere onto the flat map. In addition, locations on the flat map can be expressed by Cartesian coordinates, using the ventral tip of the CeS as the origin. (*k*) The fMRI activation pattern displayed on the flat map. (*l*) The SuMS database can play a central role using volume specification files to automatically enter and retrieve all of the volume data (blue arrows) and surface specification files to enter and retrieve all of the surface data (red arrows). To download these data, connect to: [http://stp.wustl.edu/sums/sums.cgi?specfile =2001-04-23.Demo.L.DicksonEtAl_Figure.3_SURFACES.spec.

Table 1. Volume and surface file types.

(a) volume data

file category	data type
Structural_Image_file	structural MRI or cryosection image volume
Segmentation_Recon_file	reconstruction substrate used for surface generation
Inner_oundary_Recon_file	probabilistic map of inner cortical boundary
Outer_Boundary_Recon_file	probabilistic map of outer cortical boundary
Functional_activation_file	fMRI image volume

(b) surface data

(i) surface geometry

data type	file type	exemplars
node position	coordinate file	Caret coord_file
node topology	topology file	Caret topo_file
node position + topology	geometry file	VTK file

(ii) surface node attributes

data type	file type	exemplars
categorical	paint file	sulci; cortical areas
real-valued	metric file	folding; distortion
pseudo-colour	RGB_ paint file	fMRI activation
probabilistic	atlas file	probabilistic map

(iii) three-dimensional point data

data type	represented by coordinates	represented relative to tiles	exemplars
contours	border file	border projection file	areal boundaries
points	cell file	cell projection file	labelled neurons
	foci file	foci projection file	fMRI activation foci

and surface topology as separate files (Caret format). The other is to represent or to combine surface coordinates and topology into a single surface geometry file, such as a VTK (Visualization ToolKit) format. Each option has advantages and disadvantages; the Caret format is more flexible and efficient for representing a wide variety of surface geometries. SuMS can store any of these file types, as long as the relevant metadata are properly incorporated into the file header(s). It also allows particular files to be tagged as 'milestone' files, signifying that they are the best or most appropriate of that type (e.g. the least distorted spherical map). The milestone option facilitates the rapid retrieval of the most useful files for each surface family.

(c) *Ancillary spatial data*

Besides the surface geometry and volume data described in relation to figures 4 and 5, SuMS handles many types of surface-related experimental data. These are grouped into two main data types according to their spatial characteristics and how they are linked to surfaces (table 1). One major distinction is between surface node data (expressed as attributes assignable to each surface node), and three-dimensional point data (i.e. real-valued data points whose coordinates do not coincide with integral voxel values in the volume or with nodes in the surface).

Surface node data are useful because many types of experimental data can be efficiently represented by assigning attributes directly to nodes on the surface. The data can then be visualized by colouring the nodes (and the associated surface tiles) according to the node attributes, as exemplified in figure 3*d*−*k*. Node-based representation of data has the advantage of automatically preserving node attributes when the surface is modified in configuration or topology. SuMS distinguishes among four types of surface node data (surface node attributes in table 1). Categorical data (identified as 'paint' files) involve up to five discrete categories assignable to each node (e.g. functional and/or geographical subdivisions). Metric data involve real-valued attributes (e.g. cortical folding or cortical thickness) that can be visualized on grey-scale or pseudo-colour maps. Pseudo-colour data include fMRI activation patterns or other appropriately scaled measures that can be visualized using appropriate colour look-up tables. Probabilistic data involve information expressed by assigning each node a probability of

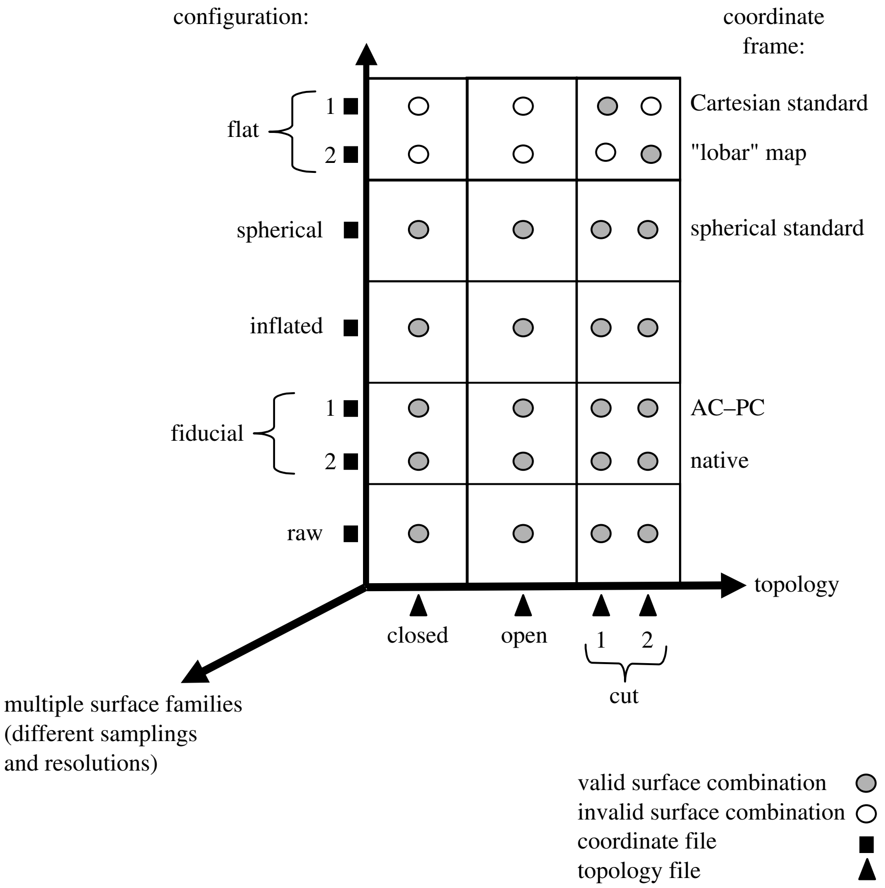


Figure 5. A surface family matrix resulting from a hypothetical combination of multiple topology files and coordinate files belonging to a common surface family and suitable for incorporation into a surface specification file. The topology files include closed, open and cut topologies. The coordinate files include the raw, fiducial, inflated, spherical and flat configurations. Multiple entries have permitted for each type of topology and coordinate file. Most combinations of coordinate and topology file are valid, insofar as they generate geometrically sensible surfaces, but a few are invalid (e.g. a flat map with a closed topology).

belonging to one or another category (e.g. cortical area), based on information derived from a population of cases.

For some types of data, it is desirable to retain spatial information at a resolution finer than the separation between surface nodes and also to encode the distance from the surface (if data points do not lie directly on the surface). This occurs, for example, when charting the distribution of individual neurons labelled by anatomical tracers or immunocytochemistry ('cell' data), when charting the centres of activation foci from neuroimaging data ('foci' data), and when drawing contours along the surface to delineate boundaries between anatomical or functional subdivisions ('border' data). SuMS can preserve this information using either of two file formats: by storing the precise spatial coordinates of each data point directly ('cell', 'foci' and 'border' files); or by storing the location of each data point relative to the nearest tile in the surface (e.g. border-projection files and cell-projection files).

(d) *Metadata for surface-based transformations*

As noted in § 1, a valuable strategy for making comparisons among different individuals involves surface-based warping to transform data from each individual (source) map onto a common target map, such as a surface-based atlas. SuMS is currently designed to handle deformation data generated by landmark-based fluid deformation methods, in which the mapping from source to target is constrained by user-defined contours (Joshi 1997; Bakircioglu *et al.* 1999; Van Essen *et al.* 1998, 2000, 2001*b*). This entails tracking a number of additional types of information besides that contained in a standard surface specification file, such as the identity of the source and target surfaces, predeformation and postdeformation data (landmarks plus passively carried data). This information can be incorporated in a deformation specification file that allows automated entry of the relevant information from each instantiation of the deformation.

(e) *SuMS architecture*

The architecture of SuMS involves several components (figure 6). At its heart are two components involved in data storage: (i) a Sybase database management system (DBMS), which contains the database of SuMS metadata; and (2) the SuMS data cache, which holds the surface and volume data files and is the only system component demanding large data capacity. SuMS provides two distinct modes of communication between users and the database: WebSuMS, which runs on standard Web browsers, and the downloadable SuMS Client, designed to run on any computer that supports JAVA. WebSuMS includes a set of forms and a Web server CGI (common gateway interface) script which dynamically generates database queries and results. This mechanism

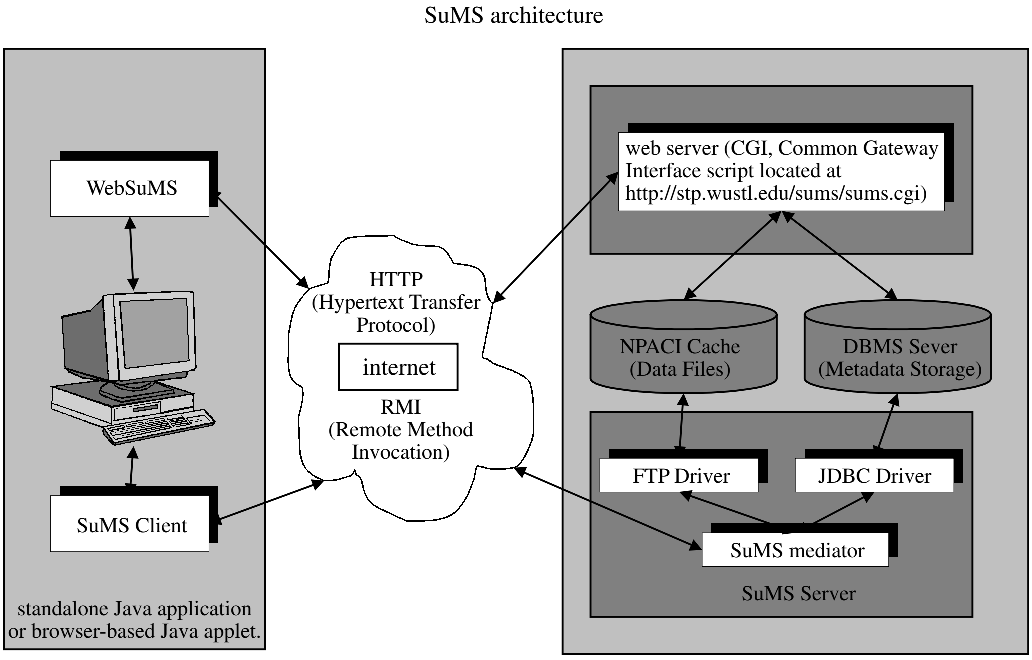


Figure 6. Schematization of the major components of SuMS architecture. When using WebSuMS, the user communicates via the Internet using HTTP and CGI scripts. When using the SuMS Client, the Internet links to the SuMS server are by RMI. Within the SuMS server, the SuMS mediator communicates with the DBMS server via the JDBC driver in order to enter or extract metadata. These metadata are used to guide the entry or retrieval of files from the data cache via the FTP driver.

has the advantage of providing access with no client installations. The SuMS Client is a standard Java application that communicates with the SuMS server via remote method invocation (RMI) and allows for data insertion, retrieval and editing. WebSuMS and the SuMS Client provide similar overall functionality (although not all of the functionality of the SuMS Client has yet been implemented for WebSuMS).

(f) *Data entry, retrieval and search options*

(i) *Specification files*

SuMS is designed for efficient entry, search and retrieval of experimental data using natural groupings of files (identified as 'specification files') that share a common lineage. A volume specification file includes a listing of volume-related files associated with a particular reconstruction substrate. A surface specification file includes a listing of surface-related files associated with a particular surface family. Specification files may also be hybrids that include both volume and surface data.

(ii) *Data entry*

Data entry in SuMS is a simple process in which the name(s) of the volume and/or surface specification files intended for insertion are entered into the appropriate dialogue box in the SuMS Client. The requirements for successful data entry are: (i) that all files are actually present in the directory locations listed in the specification file, and (ii) that the requisite metadata are contained in the appropriate file headers. Files generated using SureFit and Caret (v. 4.3 or higher) automatically include this information in the individual files or in the specification files. At the outset of the data entry process, the metadata for all files are checked for completeness, and prompts are given if required information is missing. A cross-checking process avoids insertion of identical files and ensures that only new datasets are added to SuMS.

(iii) *Data search and retrieval*

SuMS offers several convenient ways to download files of interest. The simplest is to make a direct hyperlink connection to a particular specification file in the database, starting from a separate application such as a journal article or email. For example, all of the data illustrated in figure 3 can be immediately accessed by a hyperlink connection to the appropriate specification file (http://stp.wustl.edu/sums/sums.cgi?specfile = 2001-04-23.Demo.L.DicksonEtAl_figure.3_Surfaces.spec). Selection of this hyperlink brings up a listing of all the files contained in the specification file. Any or all of the listed files can be selected for downloading by HTTP to the desired directory in a single step. Similarly, surface-based atlases of human and macaque cortex can be accessed by hyperlinks in the figure legends of the study where they are described (Van Essen *et al.* 2001*b*). Alternatively, if the name of the desired specification file is known, it can be accessed directly using an appropriate dialogue box in the main SuMS window (see figure 8).

When a search is needed to identify the files of interest, the three-stage process schematized in figure 7 allows easy selection and downloading of datasets that meet the user's search criteria. In brief, the first stage is to choose a combination of search criteria for identifying files of potential interest (step 1a) and then to initiate the search process (step 1b). The second stage is to view the results of the initial search (step 2a) and to select any or all of these files for placement in a search repository (step 2b). The search repository is a listing of files provisionally targeted for downloading. The contents of the repository (akin to a

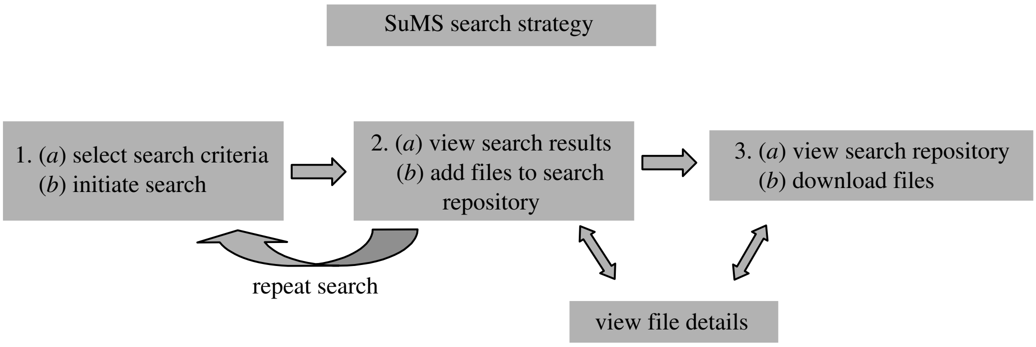


Figure 7. A flow chart for the search and retrieval process using either WebSuMS or the SuMS Client. In each case, the three main stages of the process include (i) initial selection of search criteria, (ii) placing search results in a temporary repository ('shopping basket'), and (iii) downloading of files selected from the search repository.

'shopping cart') can be expanded by repeating the search process using a different set of criteria. The third stage is to select the final set of files from the search repository (step 3a) and then to initiate the download process (step 3b).

To initiate a search, the user can either go to the WebSuMS home page using an Internet browser or launch the previously downloaded SuMS Client. The main SuMS window appears like that in figure 8 when using the SuMS Client. The upper rows include options for data insertion, and for switching to a different database or user without needing to restart SuMS. The lower two-thirds of the window are occupied by the 'Search Criteria' tab, which allows the user to set a number of criteria concurrently. When using WebSuMS, the screen layout is somewhat different, but the functionality is very similar. If the name of the desired file (a specification file or an individual data file) is known, it can be accessed by pasting or typing into the specification file window (A in figure 8).

If the search option is used, a number of criteria can be set concurrently. The first six search options (B1–A6 in figure 8) are related to specific types of metadata that are singled out because of their general utility. These include the species (B1), the case (B2), the reconstruction substrate (B3), two types of keyword search (B4, B5) and the data file type (B6). In the example search illustrated here, the species selected (B1) is 'human', which was selected from a pull-down menu of all species currently in the database. The case name (B2) is 'VH.R' (the right hemisphere of the visible human). This was selected using the adjacent 'Browse' button, which brings up a listing of all individual cases in the database belonging to the species selected. The reconstruction substrate (B3) is selected automatically if the database contains only a single entry for that case, as in this example (VH.R.RECON1). (If there are multiple entries, the choice is made using the adjacent Browse button to show the available reconstruction substrates.) Option B4 searches for key words in the comments section of each file header. This can be very useful when searching for particular datasets that are tagged by suitable wording inserted into the comment field at any stage of data analysis. Option B5 searches for keywords within surface data files that contain textual information outside the file header. This includes border files (which contain the names of particular borders), paint files (which contain the names of different cortical areas, etc.) and other file types listed as 'spatial data' in table 1. In figure 8, the comment field was left blank, and the surface data keyword was set as 'V4', in order to identify datasets that include information about visual area V4.

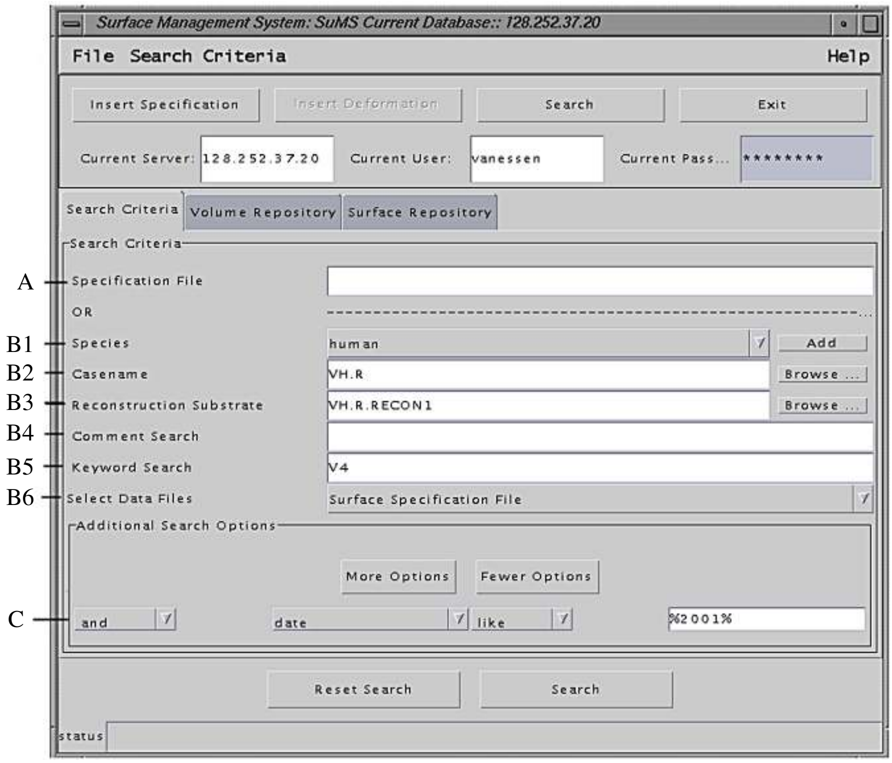
Search option B6 selects the particular file type to be viewed. In the example shown, the selection is for a Surface Specification file. Since specification files typically contain an extended list of files associated with a given case, this is an efficient way to obtain a variety of related surface files whose commonality was established at the time the specification file was inserted. The analogous process for volume data is to select 'Volume Specification File'. Alternatively, this search option allows selection of a single data file type if the data of interest are of only a particular type (e.g. topology files or paint files).

Besides the specific metadata categories just described, general searches can be made on all metadata in the database, using the 'Additional Search Options' section at the bottom of figure 8. In this example, the additional criterion is that the date of data entry was 2001. (This criterion must be met for the specification file itself or any of the individual data files.)

Once the search is initiated (by pressing the 'Search' button) a 'Search Results' window appears. At the top of this window is a list of all files that meet the search criteria. For example, the top section in figure 9*a* lists two surface specification files meeting the criteria listed in figure 8 at the time the database was searched. When one of these specification files is highlighted, its contents are listed (middle section of figure 9*a*). Additional information about individual files can be obtained using the 'Show Detail' option, which brings up a display of the information in the header of the currently highlighted file (figure 9*b*). In short, many different types of information about each file can be rapidly accessed to facilitate the search selection process.

Any or all of the file names in the Search Results window can be transferred to a search repository, using the highlighting option, then pressing the 'Select' button. In general, volume data are automatically placed in the 'Volume Repository' (second tab in the main SuMS window), and surface data are placed in the 'Surface Repository' (third tab in the main SuMS window). Files listed in the search repository (figure 10) can be downloaded individually or as a group (via HTTP, after file

SuMS main window (Search Criteria tab)



Current Server: 128.252.37.20
Current User: vanessen
Current Pass... ********

Search Criteria
Specification File:
OR
Species: human
Casename: VH.R
Reconstruction Substrate: VH.R.RECON1
Comment Search:
Keyword Search: V4
Select Data Files: Surface Specification File
and | date | like | %2001%

Figure 8. The SuMS main window when setting search criteria. The top portion includes options for data insertion, initiating a search (once criteria are set), and changing the database server. The Search and Retrieval tabs are used to select search criteria, or to view results that have been transferred to the Volume Repository or the Surface Repository. The currently selected tab is the Search Criteria, which provides a combination of search options described in the text.

compression) by selecting 'Download All Highlighted Files'. Downloaded files are listed in a customized specification file that comes automatically with the download. SuMS checks whether any of the files have already been downloaded to the current directory in order to avoid unnecessary duplicate downloading.

(iv) *Download times*

Download times depend on several factors, including average file size, total number of files and network transmission rates. The entire dataset illustrated in figure 3 (10 surface files, 3 volume files) totalled 9 MB after file compression (24 MB uncompressed). The data were downloaded in *ca.* 15 s over a local Internet connection. Even though transmission will be slower for distant sites, it should remain within a reasonable range for most situations, especially as the bandwidth of the Internet increases.

(v) *Access control*

SuMS uses a security model based on the designation of owners, readers and writers for each file in the database. This ensures that private data are protected, public data are freely available, and any data can be made available to some investigators (e.g. collaborators) but not others. An owner is the investigator who submitted the dataset. Readers or groups of readers can be added to an access list by the owner to allow viewing of the dataset. Finally, writers can be added by the owner to allow others to edit the dataset record, perhaps for annotation purposes. This security scheme is flexible enough for any owner to allow access to restricted datasets on an individual-user or group basis.

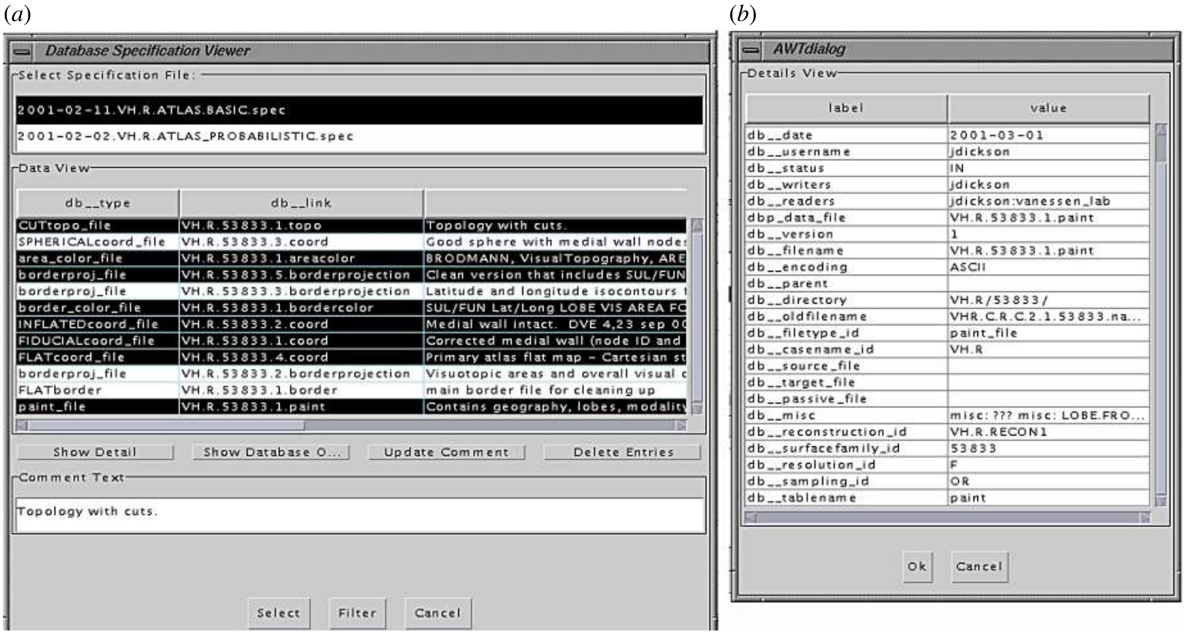
(g) *Database implementation issues*

JAVA was chosen for the implementation of SuMS Client server system, primarily for its platform independence and object-orientated characteristics. Designing a database around the complex nature of cortical surface and volume data benefited from the object-orientated design techniques possible within a JAVA environment. The database access is performed using the JDBC (JAVA database connectivity). This allowed the client application to be written without any database proprietary code. It can be used with any database vendor that implements a driver for the JDBC interface.

(i) *JAVA and relational mapping and data types*

SuMS uses a technique that maps objects created in an application to data stored in a relational database. In this case, the objects are JAVA classes, which form a wrapping over the database tables in the SuMS database. This general technique allowed for the creation of JAVA classes corresponding to the different data types in the SuMS database. Along with classes that match the data in the database, SuMS uses methods that allow manipulation of database objects. This avoids the use of database-specific 'stored procedures' and maintains the independence of the database.

A typical new data type in the database application extends a single JAVA class, which implements standard select, update and delete methods on the database server



2001-02-11.VH.R.ATLAS.BASIC.spec
2001-02-02.VH.R.ATLAS_PROBABILISTIC.spec

db__type	db__link	
CUTtopo_file	VH.R.53833.1.topo	Topology with cuts.
SPHERICALcoord_file	VH.R.53833.3.coord	Good sphere with medial wall node
area_color_file	VH.R.53833.1.areacolor	BRODMANN, VisualTopography, ARE
borderproj_file	VH.R.53833.5.borderprojection	Clean version that includes SUL/FUN
borderproj_file	VH.R.53833.3.borderprojection	Latitude and longitude isocontours
border_color_file	VH.R.53833.1.bordercolor	SUL/FUN Lat/Long LOBE VIS AREA FC
INFLATEDcoord_file	VH.R.53833.2.coord	Medial wall intact. DVE 4,23 sep 0
FIDUCIALcoord_file	VH.R.53833.1.coord	Corrected medial wall (node ID and
FLATcoord_file	VH.R.53833.4.coord	Primary atlas flat map – Cartesian st
borderproj_file	VH.R.53833.2.borderprojection	Visuotopic areas and overall visual
FLATborder	VH.R.53833.1.border	main border file for cleaning up
paint_file	VH.R.53833.1.paint	Contains geography, lobes, modality

Topology with cuts.

label	value
db__date	2001-03-01
db__username	jdickson
db__status	IN
db__writers	jdickson
db__readers	jdickson:vanessen_lab
dbp_data_file	VH.R.53833.1.paint
db__version	1
db__filename	VH.R.53833.1.paint
db__encoding	ASCII
db__parent	
db__directory	VH.R/53833/
db__oldfilename	VHR.C.R.C.2.1.53833.na...
db__filetype_id	paint_file
db__casename_id	VH.R
db__source_file	
db__target_file	
db__passive_file	
db__misc	misc: ??? misc: LOBE.FRO...
db__reconstruction_id	VH.R.RECON1
db__surfacefamily_id	53833
db__resolution_id	F
db__sampling_id	OR
db__tablename	paint

Figure 9. Initial search results. (*a*) 'Search Results Window'. Search results as seen after selection of a surface specification file as one of the search criteria. The top section lists all specification files that meet the search criteria. Specification files are selected if any of their listed data files include the comment search and keyword search criteria. The middle panel lists the data file contained in the currently highlighted specification file. The lower panel shows the comments section of the currently highlighted data file. (*b*) 'File Header Information'. An exemplar of the file header information obtained by selecting the 'Show Detail' option.

application. This method allows for easy extension of the underlying table structure, as may occur if additional surface-related data types need to be stored in SuMS.

Each table in the database conforms to a standard structure. First, a single identifier ('primary key' in database terminology) identifies each entry in the database table. This is followed by a set of attributes used for classification. These are generally a set of predefined attribute values that assist in providing a standardized classification of surface data. These attributes are also used to enforce file naming conventions for the surface file types inserted into the database. Additionally, some data files (e.g. border and cell data), require specification of particular surface data files (coordinate and/or topology files) to which they are linked.

(ii) *RMI*

This is an interface developed by Sun Microsystems and is a JAVA standard for distributed object communication. JAVA RMI allows the programming of distributed applications across the Internet. One JAVA application or applet (the SuMS Client in this context) can call the methods of another JAVA application (the SuMS server) running on a different host machine.

(iii) *Graphical user interface*

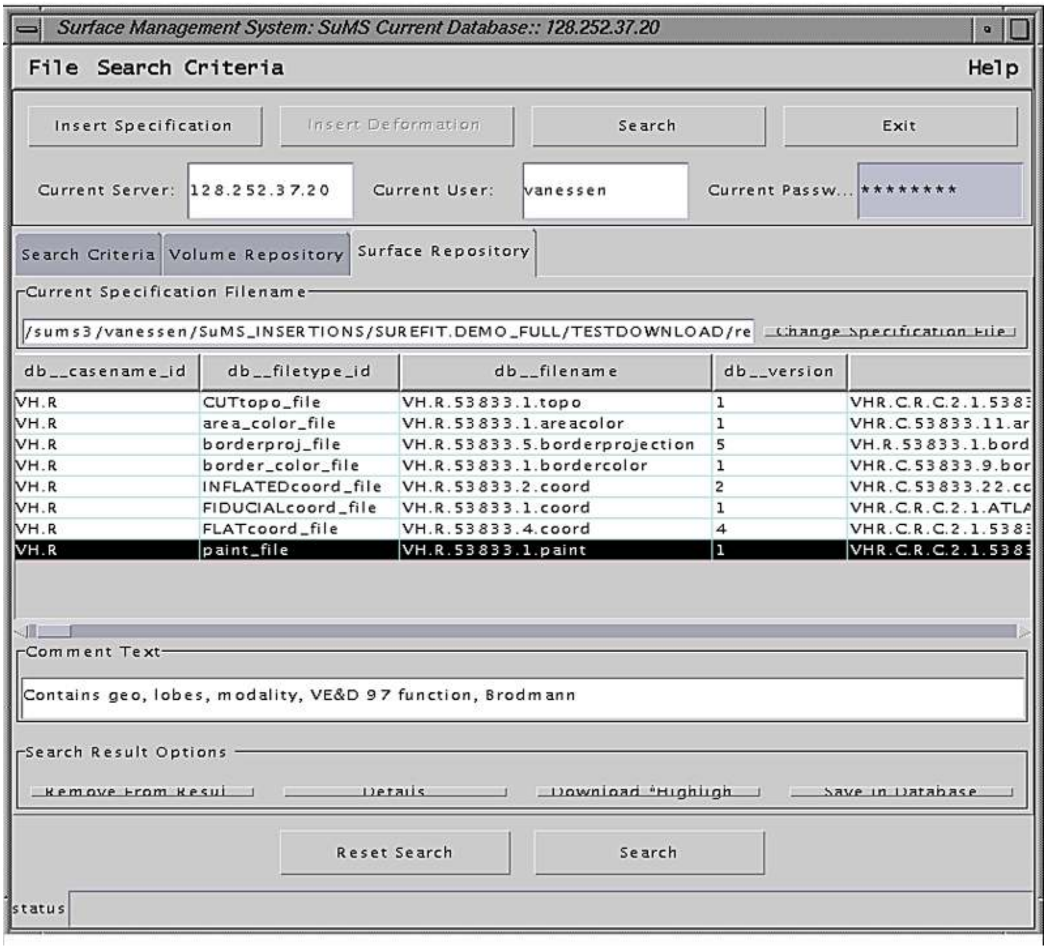
Design of an effective graphical user interface was one of many challenging aspects of SuMS development. Fundamentally, this is because of the inherent complexity of the many different data types and the need to allow users to make a number of context-sensitive search choices in order to extract the desired data. The basic objectives are to provide the user with clear options at each decision stage, along with sufficient information on which to base a choice or decision in a format that is intuitive and quickly digestible.

For the SuMS Client, the JAVA Swing Toolkit was used to develop an interface based on the underlying JAVA table data classes described earlier as the data model. At each interaction stage in the user interface, the user is protected from the underlying SQL queries being generated by the user mouse clicks. This abstraction of the database queries aims to make the database interaction as intuitive as possible. For WebSuMS, analogous functions were implemented using standard HTML forms and CGI-based scripts.

(iv) *Data storage*

The volume and surface data associated with cortical surface-based analyses may include dozens or even hundreds of files per case, given the structural volume data associated with surface reconstruction, the functional volume data associated with complex functional imaging protocols, and the many different types of surface-related data. Individual files range over a wide range of sizes, and for human cortex are typically in the range of 1–10 MB for surface geometry and volume data. The total data storage required per case can easily exceed 100 MB for surface and volume data combined and may approach the gigabyte range for complex datasets such as those associated with event-related fMRI.

Currently, the data in SuMS are stored on a 2 TB data cache at Washington University that is supported by The National Partnership for Advanced Computational Infrastructure (NPACI). We anticipate that the SuMS database will soon take advantage of the Storage Request Broker (Baru *et al.* 1998) developed by the San Diego Supercomputer Center (SDSC). This system allows access to the



Current Server: 128.252.37.20

Current User: vanessen

Current Passw... ********

Current Specification Filename: /sums3/vanessen/SuMS_INSERTIONS/SUREFIT.DEMO_FULL/TESTDOWNLOAD/re

db__casename_id	db__filetype_id	db__filename	db__version	
VH.R	CUTtopo_file	VH.R.53833.1.topo	1	VHR.C.R.C.2.1.538
VH.R	area_color_file	VH.R.53833.1.areacolor	1	VHR.C.53833.11.ar
VH.R	borderproj_file	VH.R.53833.5.borderprojection	5	VH.R.53833.1.bord
VH.R	border_color_file	VH.R.53833.1.bordercolor	1	VHR.C.53833.9.bor
VH.R	INFLATEDcoord_file	VH.R.53833.2.coord	2	VHR.C.53833.22.cc
VH.R	FIDUCIALcoord_file	VH.R.53833.1.coord	1	VHR.C.R.C.2.1.ATL
VH.R	FLATcoord_file	VH.R.53833.4.coord	4	VHR.C.R.C.2.1.538
VH.R	paint_file	VH.R.53833.1.paint	1	VHR.C.R.C.2.1.538

Comment Text: Contains geo, lobes, modality, VE&D 97 function, Brodmann

Figure 10. The main SuMS window on selection of the Surface Repository tab. The middle portion shows all data files currently listed in the Surface Repository. The downloading options in the lower section allow any or all of these files to be downloaded in a single step, along with an appropriate specification file.

combined storage resources of the SDSC, which has a current capacity of 120 TB.

4. DISCUSSION

(a) *Current status and prospects for use*

As with any database or other software application, the degree to which SuMS is used by the neuroscience community will depend upon many factors. Among the most important for SuMS are: (i) the evolving needs of the community; (ii) the robustness and user-friendliness of the software; and (iii) its degree of platform independence and compatibility with available surface reconstruction and visualization software.

(b) *Anticipated need*

Until recently, only a handful of laboratories generated computerized cortical surface reconstructions. With the emergence of software for automated surface reconstruction and manipulation, surface-based analyses should increase dramatically over the next few years. The number of laboratories routinely generating cortical surface reconstructions will probably expand to hundreds or even thousands (including clinical researchers as well as basic neuroscientists). If each laboratory typically studies 10–100 subjects annually, and if most studies involve multiple analyses (e.g. different reconstruction substrates or regional analyses), the number of new surfaces appropriate for a surface-based database could soon exceed 100 000 per year. As the amount of surface-related data continues to expand, the advantages of bringing the data under the umbrella of a database will become increasingly evident, just as has occurred in other scientific arenas such as genomics and protein structure (Berman *et al.* 2000; Marshall 2000; Pennisi 2000).

(c) *User-friendliness and software robustness*

Our overarching objective in designing SuMS has been to establish a platform in which users can easily enter and retrieve surface reconstructions with high reliability as to the origin and nature of each surface and with a minimum of user expertise required in technical aspects of database structure and organization. We believe that the initial version of SuMS comes reasonably close to this objective. Nonetheless, there are numerous ways in which SuMS can be refined and enhanced. User feedback regarding the most useful enhancements will aid in the development of future versions of SuMS.

(d) *Platform independence and interoperability*

WebSuMS, which runs on standard Web browsers, and the SuMS Client, which uses the widely supported JAVA language, are designed to operate on nearly all computer platforms in current use. Once the data are downloaded, users can view surface and volume data using SUREFIT and CARET software, which are freely available and run on multiple platforms (Irix, Solaris and Linux for CARET). Additional steps are planned that will improve the interface between SuMS and various other surface reconstruction and visualization packages in common use. One will be to incorporate file conversion capabilities that convert between the SuMS internal data formats and alternative volume and surface visualization formats used by other software packages. Another will be to facilitate automated data entry from different surface reconstruction packages and to allow reading of SuMS surface specification files by a broader range of surface visualization applications.

Currently, SuMS operates using only a single SuMS server and a single data cache. As usage grows, there will probably be a need for mirror sites and/or local repositories at distributed locations. This is particularly important for investigators in Europe and Asia. Expanding the number of servers should be straightforward, because the SuMS server (like the SuMS Client) runs on any computer that supports JAVA. The Sybase DBMS also runs on a variety of platforms, and in any event SuMS can be linked to alternative DBMSs (e.g. Oracle). It should also be feasible to modify SuMS so that it queries a user-specified group of databases at different locations, thereby facilitating the exploration of data that are stored in a distributed network.

(e) *Security and public versus private access*

Databases offer great opportunities for widespread access to data that are of broad use to the scientific community. Perhaps the most notable success story in this regard is the public availability of sequence data for the genomes of humans and other species (Marshall 2000; Pennisi 2000). This arena has also vividly demonstrated the interplay between competing pressures—one that strives to keep data initially private for the benefit of those who discover or acquire the data, and another that strives to make data available to the scientific community as fast as possible (e.g. daily updating of genome sequence data).

While there are clear parallels between neuroimaging data on the one hand and genome or protein sequence data on the other hand, there are also important differences with regard to privacy and security issues. Perhaps the largest difference is that in neuroimaging studies the amount and nature of the data needed to analyse and interpret any given experimental observation or finding tends to be very large and complex.

Because it includes multiple security levels built directly into the database design, SuMS is inherently flexible with regard to issues of access (public versus private) and can adapt to evolving standards in the neuroimaging community. It is designed to handle large amounts of prepublication data in which access is appropriately limited to individual investigators, laboratories or specific collaborative groups, plus whatever data are available for unrestricted access. Of particular import in the public arena will be surface-based atlases that provide an up-to-date compendium of current knowledge about cortical organization and function in health and disease. The prospective user base for such atlas data is large, as it will include clinicians, educators and students as well as the neuroscience research community.

(f) *Database federations*

SuMS is designed to handle spatial data along with the metadata that describe spatial characteristics of these data. Most experimental studies also include many other types of information that are inherently non-spatial in nature or are otherwise not readily incorporated into the SuMS organizational and metadata framework. For example, functional neuroimaging studies typically include a plethora of information (e.g. behavioural paradigm, data acquisition parameters, individual subject characteristics) that may be crucial for interpreting some aspects of the experimental findings.

Practical considerations argue against a monolithic approach to database design, in which a database would become arbitrarily complex in order to handle all of the data relevant to a given type of experiment (e.g. fMRI studies). An attractive alternative is to establish a federation of databases, each specialized for restricted subsets of the overall domain of experimentation and knowledge (Koslow 2000). The challenge of federation is to achieve effective linkages, so that users can readily extract complex ensembles of information derived from multiple databases. One can envisage, for example, productive federations between SuMS and other databases such as BrainMap (Fox *et al.* 1994) and the recently established National fMRI Data Center (www.fmridc.org), which include more extensive descriptions of non-spatial aspects of functional neuroimaging studies. Efforts towards federation, which is a relatively new area in the database field, can draw from recent successes such as the Protein Data Bank (Berman *et al.* 2000). Successful federation of brain mapping databases will be of great benefit to the neuroscience community.

This work was supported by The National Partnership for Advanced Computational Infrastructure (NPACI) and by a Human Brain Project/Neuroinformatics research grant funded jointly by the US National Institute of Mental Health, National Science Foundation, National Cancer Institute, National Library of Medicine and the National Aeronautics and Space Administration (R01 MH60974-06).

REFERENCES

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B. & Zilles, K. 1999 Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* **412**, 319–341.

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T. & Zilles, K. 2000 Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? *NeuroImage* **11**, 66–84.

Bakircioglu, M., Joshi, S. & Miller, M. I. 1999 Landmark matching on the sphere via large deformation diffeomorphisms. *Proc. SPIE Med. Imaging Image Process.* **3661**, 710–715.

Baru, C., Moore, R., Rajasekar, A. & Wan, M. 1998 The SDSC storage resource broker. In *Proceedings of the CASCON Conference, 1998, Toronto, Canada*.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. 2000 The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.

Chicurel, M. 2000 Databasing the brain. *Nature* **406**, 822–825.

Corbetta, M. (and 10 others) 1998 A common network of functional areas for attention and eye movements. *Neuron* **21**, 761–773.

Dale, A. M., Fischl, B. & Sereno, M. I. 1999 Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9**, 179–194.

Drury, H. A., Van Essen, D. C., Anderson, C. H., Lee, C. W., Coogan, T. A. & Lewis, J. W. 1996 Computerized mappings of the cerebral cortex. A multiresolution flattening method and a surface-based coordinate system. *J. Cogn. Neurosci.* **8**, 1–28.

Drury, H. A., Van Essen, D. C., Corbetta, M. & Snyder, A. Z. 1999 Surface-based analyses of human cerebral cortex. In *Brain warping* (ed. A. Toga), pp. 337–363. New York: Academic Press.

Fischl, B. & Dale, A. M. 2000 Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl Acad. Sci. USA* **97**, 11 050–11 055.

Fischl, B., Sereno, M. I. & Dale, A. M. 1999*a* Cortical surface-based analysis. II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage* **9**, 195–207.

Fischl, B., Sereno, M. I., Tootell, R. B. & Dale, A. M. 1999*b* High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8**, 272–284.

Fox, P. T., Mitiken, S., Davis, G. & Lancaster, J. L. 1994 A database of human functional brainmapping. In *Advances in functional neuroimaging: technical foundations* (ed. R. W. Thatcher, M. Hallett, T. Zeffiro, E. R. John & J. M. Huerta), pp. 95–105. Orlando, FL: Academic Press.

Geiger, B. 1993. *Three-dimensional modeling of human organs and its application to diagnosis and surgical planning*, Technical Report 2105. Institute National de Recherche Informatique et Automatique.

Geyer, S., Schormann, T., Mohlberg, H. & Zilles, K. 2000 Areas 3a, 3b and 1 of human primary somatosensory cortex. *NeuroImage* **11**, 684–696.

Goebel, R. 2000 A fast automated method for flattening cortical surfaces. *NeuroImage* **11**, S680.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H. & Singer, W. 1998 Functional imaging of mirror and inverse reading reveals separate coactivated networks for oculumotion and spatial transformations. *NeuroReport* **9**, 713–719.

Joshi, S. C. 1997 Large deformation landmark based differeomorphic for image matching. PhD thesis, Sever Institute, Washington University, St Louis, MO, USA.

Koslow, S. H. 2000 Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neurosci.* **3**, 863–865.

Lewis, J. W. & Van Essen, D. C. 2000 Cortico-cortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137.

Lorensen, W. & Cline, H. 1987 Marching cubes: a high resolution 3-D surface reconstruction algorithm. *Comp. Graph.* **21**, 163–169.

MacDonald, D., Kabani, N., Avis, D. & Evans, A. C. 2000 Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage* **12**, 340–356.

Marshall, E. 2000 Rival genome sequencers celebrate a milestone together. *Science* **228**, 2294–2295.

Maunsell, J. H. R. & Van Essen, D. C. 1987 The topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchi–tectonic boundaries. *J. Comp. Neurol.* **266**, 535–555.

Pechura, C. M. & Martin, J. B. 1991 *Mapping the brain and its functions*. Washington, DC: National Academy Press.

Pennisi, E. 2000 Finally, the book of life and instructions for navigating it. *Science* **228**, 2304–2307.

Rademacher, J., Caviness Jr, V. S., Steinmetz, H. & Galaburda, A. M. 1993 Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cereb. Cortex* **3**, 313–329.

Sereno, M. I., Dale, A. M., Liu, A. & Tootell, R. B. H. 1996 A surface-based coordinate system for a canonical cortex. *NeuroImage* **3**, S252.

Shepherd, G. M., Mirsky, J. S., Healy, M. D., Singer, M. S., Skoufos, E., Hines, M. S., Nadkarni, P. M. & Miller, P. L. 1998 The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends Neurosci.* **21**, 460–468.

Sisodiya, S. M. N., Free, S., Fish, D. R. & Shorvon, S. D. 1999 Novel magnetic resonance imaging methods for quantifying changes in the cortical ribbon in patients with epilepsy. *Adv. Neurol.* 81–87.

Toga, A. W. (ed.) 1999 *Brain warping*. New York: Academic Press.

Van Essen, D. C. & Drury, H. A 1997 Structural and functional analyses of human cerebral cortex using a surface-based atlas. *J. Neurosci.* **17**, 7079–7102.

Van Essen, D. C., Drury, H. A. & Anderson, C. H. 1990 An automated method for accurately reconstructing the cortical surface. *NeuroImage* **9**, 173.

Van Essen, D. C., Drury, H. A., Joshi, S. C. & Miller, M. I. 1997 Comparing human and monkey cortex by warping of cortical flat maps. *Soc. Neurosci. Abstr.* **23**, 846.

Van Essen, D. C., Drury, H. A., Joshi, S. & Miller, M. I. 1998 Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proc. Natl Acad. Sci. USA* **95**, 788–795.

Van Essen, D. C., Drury, H. A., Hadjikhani, N. & Tootell, R. 2000 A probabilistic surface-based atlas of human visual cortex. *NeuroImage* **10**, 533.

Van Essen, D. C., Drury, H. A., Hanlon, D. & Harwell, J. 2001*a* *User's guide to SureFit and Caret: cortical segmentation, surface reconstruction, and flattening software*. (See http://stp.wustl.edu; software currently in beta-testing mode.)

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B. H., Bakircioglu, M. & Miller, M. I. 2001*b* Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* **41**, 1359–1378.

Von Economo, C. & Koskinas, G. N. 1925 *Die cytoarchitektonik der hirnrinde des erwachsennen menschen*. Berlin: Springer.

Wandell, B. A., Chial, S. & Backus, B. T. 2000 Visualization and measurement of the cortical surface. *J. Cogn. Neurosci.* **12**, 739–752.