Integrating Brain Data Spatially: Spatial Data Infrastructure and Atlas Environment for Online Federation and Analysis of Brain Images

Ilya Zaslavsky[1], Haiyun He[2], Joshua Tran[1], Maryann E. Martone[2], Amarnath Gupta[1]
[1]*San Diego Supercomputer Center, University of California San Diego*
[2]*National Center for Microscopy and Imaging Research, UCSD*
{zaslavsk, hhe, trantj, gupta}@sdsc.edu, mmartone@ucsd.edu

Abstract

Numerous digital atlases of the brain have been developed for different species through the efforts of researchers around the world. For a comprehensive picture of brain morphology and function, it is important to discover and bring together various images, segmentations and markup, generated within different projects, and juxtapose them within a single federated atlas interface. This paper describes strategies and tools for integrating distributed sources of brain data spatially. We focus on the organization of web-enabled spatial data sources which include ArcIMS feature and image services and distributed grid sources, and on the Smart Atlas, a GIS-based atlas environment enabling users to discover, access, visualize and query heterogeneous images and image markup. We demonstrate construction of spatial databases from unorganized vector and raster brain data based on ontological relationships between anatomical features from the Unified Medical Language System, spatial registration mechanisms, and tools and user interfaces for federated brain data visualization and analysis.

1. Introduction

Studies of schizophrenia, Parkinson's disease, Alzheimer's disease and other illnesses caused by disruption of brain functions, are often based on collections of brain images, usually obtained at different resolutions through computer tomography for human subjects, or through surgical procedures for other species. Atlases have been a common way to organize such image series. Multiple examples of such atlases with corresponding image segmentation and 2D and 3D visualization techniques have been developed [e.g., for mouse brain: 3, 4, 8, 12, 13]; several are available online [see 11]. A comprehensive list of available brain data sources and atlases is maintained by [6].

Lack of interoperability across different functional and anatomical data sources has been a common problem with these atlases. It has been difficult and time-consuming to align and correlate data under control of different atlas systems manually, even less so automatically. A notable exception are techniques developed within the LONI pipeline processing environment [10] which is designed to support brain atlas generation workflow composed from distributed processing components, and focuses on seamless exchange of image data across them. Several other projects, focused primarily on primate brain, are designing tools for brain cartography from heterogeneous data [15]. However, on-demand integration of heterogeneous spatial datasets remains a serious challenge. In particular, queries such as "*which atlas sources provide images for user-selected area of the brain at a given resolution*", or "*display image segmentations available in one atlas over a newly obtained image stored in a different location on the grid*", or "*compare protein distributions in a given area of the brain available at several atlas sources*" - could not be answered without a dedicated infrastructure, where atlas data sources can be queried spatially. Such an infrastructure is being developed within the NIH-funded Biomedical Informatics Research Network (BIRN) project. Non-spatial source registration, grid management, mediation and integrated view formulation, and ontology management issues within BIRN, have been addressed elsewhere [eg. 1, 5]. In this paper, we explore principles and techniques that enable spatial data interoperability, including spatial registration, discovery, query, and visualization across brain data sources.

The challenges of spatial integration of brain image data are caused by extreme heterogeneity in image formats, image spatial properties, spatial registration conventions, types and organization of image annotation, access, retrieval and display mechanisms. We address these challenges by (1) organizing standard online spatial data sources and developing spatial source wrappers that handle format and coordinate system conversion, for both image and annotation data, coupled with mechanisms for image cutting and downsampling, and registering and indexing them in a consistent way, using stereotaxic coordinates, and (2) developing a new generation atlas client that can access heterogeneous brain data sources, support query formulation against integrated views over these sources, and visualize and juxtapose images and markup retrieved from them. The paper is structured accordingly: the next section discusses strategies and tools we developed for organizing spatial sources of brain data into a federated system, followed by a section that describes user interface solutions, and a conclusion.

2. Organization of spatial data sources

The brain image data currently used in this project, derive from published atlases [8], or are generated in BIRN-affiliated labs at Harvard, Duke University, UCLA, CalTech, and UCSD. They may represent raster images (coronal, sagittal or horizontal) at different resolutions (from 256x256 matrices available from LONI, to ~2Gb microscopic images from NCMIR, UCSD), and differently registered (in stereotaxic coordinates [7, 8], in image coordinates within an Analyze cube, in ad hoc canvas coordinates of particular drawing/imaging software, etc.). Some of these images may also include vector markup, i.e. delineation of anatomic features, locations of anatomic feature labels and/or other annotations, as well as coordinate grid artifacts. To enable spatial integration of such diverse sources, we organize them into "atlas data sources" by adding explicit spatial registration information, creating image cutting and downsampling wrappers (for raster data) and converting unorganized image markup into queryable vector feature services.

2.1 Converting atlas markup into a spatial database

The most recent published atlases of rat and mouse brain [7, 8] are collections of stereotaxic slices with detailed vector markup, available as Adobe Illustrator drawings. In order to convert the vector markup into a queryable topologically-correct spatial database of anatomical features and serve it online we used the following procedure:

1. Convert Adobe Illustrator atlas files into SVG (Scalable Vector Graphics). SVG was used as the intermediate format because of the ease with which SVG data are served and manipulated on the web (the first version of the Smart Atlas was based on SVG), and the ability to manipulate feature and label coordinates in XML.
2. Parse the resultant SVG files, detecting grid lines (longer horizontal and vertical lines with numerical labels near line ends) and computing transformations from the coordinate system of the drawing into the stereotaxic coordinate system.
3. Using the transformations derived in Step 2, compute stereotaxic coordinates for SVG *path* and *text* elements, and insert them as coordinates of features and labels into two Oracle Spatial tables (*polylines* and *labels*). If SVG path statements contained Bezier curves, they are decomposed into polylines, with the density of points proportional to curve length.
4. Export the two Oracle Spatial tables into polyline and point shapefiles (*shapefile* is a standard format used by many GIS packages, in particular the ArcGIS family [2]).
5. Using ArcObjects-based scripts (the ArcGIS scripting engine), separate the *polylines* shapefle into background lines (including grid lines and tic marks, label pointer lines, and insert graphics) and anatomical features; separate the *points* shapefile into label points associated with background graphics, and with anatomical features.
6. Shift feature label points to positions pointed to by pointer lines, and mirror all label points around the midline on coronal slices (as the source atlas implies feature symmetry and labels features on one side only).
7. Using planar enforcement procedures (CLEAN and BUILD operations in ArcGIS), build a topologically correct polygon coverage from the anatomic feature contours, by adding or removing nodes within user-defined geometric tolerance values.
8. Assign feature labels to polygons that contain label points, by a spatial join of the two coverages.
9. For polygons that receive multiple labels, find paths (chains of relationships) that connect label pairs in the UMLS metathesaurus [14], using the SDSC Knowledge Map Explorer (KnowME) [5]. The relationship chain may point towards one of two possible solutions: either the polygon should be split (which sometimes can be done automatically by changing the tolerance value and repeating steps 7 and 8), or the labels in a single polygon can be represented by one label (if the labels are in particular *Child-Parent* or *Broader-Narrower* relationships). For example, a group of labels DG (*dentate gyrus, UMLS Id C0152314),* PoDG *(polymorph layer of the dentate gyrus),*

CA1 *(field CA1 of hippocampus, Id C0019564)* inside one polygon, have a common parent *hippocampus*, which should be used to label the polygon. In another example, for labels B (*basal nucleus of Meynert*, *Id C0004788*) and LGP (*lateral globus pallidus, Id C0262267*) inside a single polygon, KnowME discovers relationship RN ("narrower") between B and *basal ganglia*, and relationships PAR ("parent") between *basal ganglia* and *globus pallidus*, and between *globus pallidus* and *external (lateral) globus pallidus.* From this we can conclude that the anatomic feature should be labeled LGP (indeed, basal nucleus cells are found within LGP, though their precise locations are not known). While such ontology-based heuristics reduce the number of problematic polygons further, the automatic procedure is ultimately followed by an expert check.

The described procedure uses both geometric and semantic considerations to construct, for each atlas plate, a topologically correct polygon coverage in stereotaxic coordinates, which can now be served online. Markup for all layers, including anatomical feature coverages, labels, and background lines, is served as a single *ArcIMS feature service* [2], so that it can be retrieved, visualized and queried from an atlas client or middleware as shown below.

2.2 Internet sources of brain image data

In addition to serving vector data, the atlas sources are built to serve large (up to 2Gb, in our experience to date) images or image fragments. The data grid environment adopted by BIRN is managed by the SDSC Storage Resource Broker [9]. BIRN users place images under control of SRB, and make them available to selected research collaborators. To retrieve images from the data grid, we designed an ImageMagick-based SRB proxy procedure, which, after being installed at all SRB racks, enables image cutting and downsampling at the source. This approach minimizes data movement in the system and makes it possible to visualize images or fragments at the client at reasonable resolution.

Images that are components of standard atlas collections are organized and served differently. Each image plate is referenced to the stereotaxic coordinate system (with warping as necessary), and served through *ArcIMS image service*, which ensures that the image fragments retrieved with ArcXML GET_IMAGE requests are small in size, have adequate resolution, and are correctly placed vis-à-vis other spatial data.

2.3 Spatial registration sources

Spatial characteristics and access information for both vector and raster atlas sources are referenced in a spatial registration source, built as an Oracle Spatial database. They include:

- Service URI (for ArcIMS sources: URL, service name and layer name; for images within SRB: SRB path)
- Coordinate system of the layer. If the layer is not in stereotaxic coordinates, this information is used to translate client spatial requests into the source coordinate system, and transform responses back into stereotaxic system for display at the client.
- Plane to which the image or markup belongs, described by the four standard plane equation coefficients, with respect to stereotaxic coordinate axes (dorsal/ventral to bregma; anterior/posterior to bregma; lateral to midsagittal, as used in the canonical atlas [8].
- Positioning of the image within the plane. We borrow from standard GIS georeferencing mechanism of "*world files*", i.e. recording 6 parameters of offset and pixel size along both axes, and rotation, for each image. When an image is retrieved from a grid source, these parameters are used to translate requests from stereotaxic coordinates to pixel coordinates used by the SRB proxy, and to dynamically generate a *world file* for an image fragment so that it is correctly positioned in the client's canvas. Note that this information is not used for accessing ArcIMS image services that already serve images in stereotaxic coordinates.
- Image contour or feature extent as a spatial data object (SDO).

Beyond the spatial characteristics, the registry includes layer names and descriptions (including slice type and the associated numeric constant which may simplify search significantly for non-oblique slices), as well as image thumbnail paths - but these standard accessories are outside the scope of this paper.

2.4 A typical spatial query processing scenario

Within the system shown in Figure 1, the queries outlined in the introduction of this paper are executed using one or both of the following requests:

getSources ({slice_plane_params, 2Dshape}, sp_rel)

queries a spatial registration source and returns an XML document with a set of discovered data sources within the user-specified region defined by the equation of the current atlas plane (*slice_plane_params*) and stereotaxic coordinates of a selection on that plane (*2Dshape*). *Sp_rel* specifies an Oracle Spatial predicate (such as *intersects, crosses, contains, within a distance of*) used to select the sources. If the user requests sources that describe an anatomical feature delineated on several atlas slices, then the request includes a set of (*slice_plane_params*, *2Dshape)* pairs.

getFragment(URI,plane_params,2Dshape,image_dims)

(marked (2) in Figure 1) queries an image or feature source and returns an image or geometry stream respectively, directly to atlas client. The inputs for this request are the service URI retrieved from the spatial registry in the previous request, plus the extent or shape of the data area to be returned (this defaults to the current visible extent of the displayed slice). For ArcIMS sources, this request is converted into an ArcXML statement posted to the source URL, while for grid data sources it is converted to a request against SRB proxy described above.

For more complicated queries, both types of requests will be routed through the BIRN mediator, which is currently being implemented (shown as dotted lines in Figure 1).

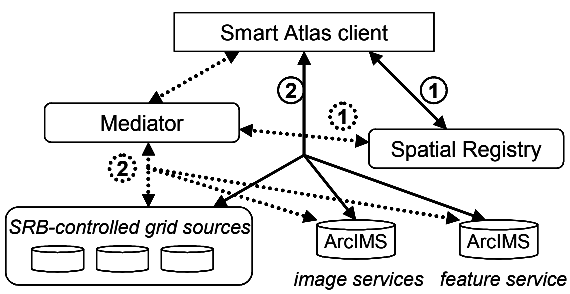

Figure 1. Organization of spatial sources and requests

3. The SMART Atlas

The Smart[1] Atlas client is developed as a Java application that can be invoked via Java WebStart from the BIRN portal at www.nbirn.net. The following key functions are supported by the atlas:

- Navigating slices by panning/zooming, identifying and selecting anatomical features, browsing and querying feature attributes, modifying display styles, along with other tools common for geographic map navigation. In addition to spatial selection tools available in the ESRI MapObjects Java toolkit [2], we developed "spherical selection", which performs selection on neighboring slices as well, within the user defined sphere dimensions (Figure 2.)

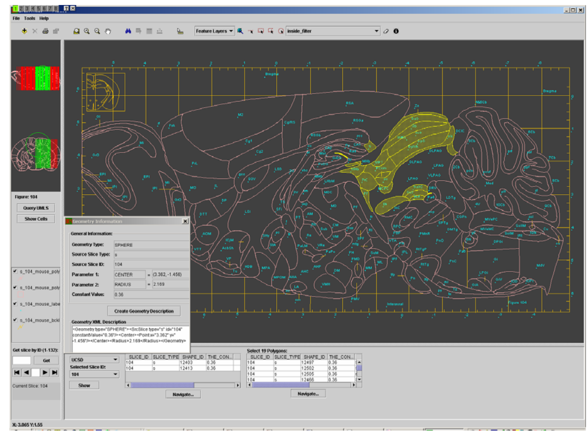

Figure 2. The Smart Atlas user interface: results of spherical selection shown on neighboring slices and displayed both graphically and in linked tables

- Source discovery by querying the spatial registry, followed by importing the discovered vector and raster data from distributed atlas sources and grid sources, as well as from local storage, and overlaying feature and image data (Figure 3.)

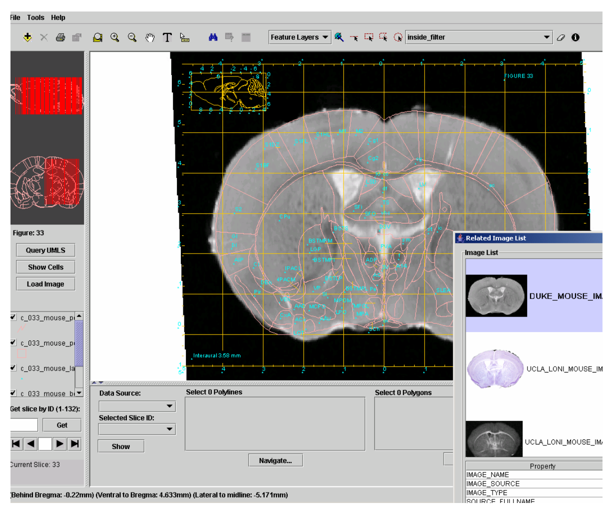

Figure 3. Finding and retrieving remote images aligned with the current vector slice

- Coordinated image visualization. It is often impossible or meaningless to warp images to complete alignment with vector slices (e.g., when comparing different species or development stages.) However, it is possible to visualize such images side-by-side, connecting features based on their relationships in an ontology like the UMLS. As users select features on the left image, the

[1] The name derives from the initial SVG-based version of the atlas called Spatial Markup And Rendering Tool.

Smart Atlas highlights and color-codes those features in the right image, which reflect concepts that are in particular UMLS relationship with the selected concept (Figure 4). This approach allows users to explore topological differences in feature organization across species and/or brain development stages.

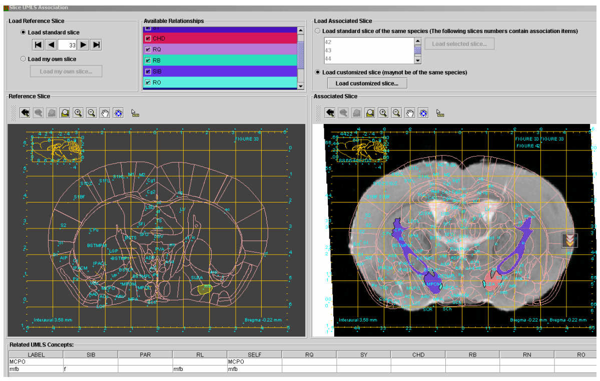

Figure 4. Coordinated image visualization in the atlas

4. Conclusion

In developing the spatial data federation infrastructure for biological images and markup, we addressed the following issues:

(1) an architecture for registering and federating heterogeneous image and markup sources, including grid-based and GIS service-based sources. This architecture emphasizes vector markup sources that can be discovered and queried independently of the images they are derived from. We also outlined the spatial characteristics and core requests needed for discovery and retrieval of large image and vector source fragments in a form that is acceptable to the client.

(2) a procedure for ontology-driven construction of a spatial database of segmented anatomical features from unorganized feature markup. We have demonstrated a successful application of this procedure to developing spatial databases for mouse brain feature markup.

(3) the development of an atlas client to support both semantic and spatial browsing and querying. We have shown specific spatial navigation and query capabilities of the client, as well as a coordinated visualization technique that uses semantic relationships to bridge otherwise spatially non-alignable slices.

The illustrated Atlas client software and data services are in different stages of maturity. Our future work focuses primarily on developing wrappers for additional spatial data types available in the BIRN grid, and designing automatic spatial registration techniques.

References

[1] Gupta A, Ludäscher B, Martone ME, Rajasekar A, Ross E, Qian X, Santini S, He H, Zaslavsky I (2003) BIRN-M: A Semantic Mediator for Solving Real-World Neuroscience Problems ", ACM Conf. on Management of Data (SIGMOD): 678.

[2] ESRI, 2004. ArcIMS, ArcGIS, ArcObjects, MapObjects Java (www.esri.com).

[3] Hof PR, Young WG, Bloom FE, Belichenko PV, Celio MR (2000) Comparative Cytoarchitectonic Atlas of the C57BL6 and 129 Sv Mouse Brains. New York: Elsevier.

[4] MacKenzie-Graham A, Lee EF, Dinov ID, Bota M, Shattuck DW, Ruffins S, Yuan H, Konstantinidis F, Pitiot A, Ding Y, Hu G, Jacobs RE, Toga AW (2004) A multimodal, multidimensional atlas of the C57BL/6J mouse brain. Journal of Anatomy 204:2, 93.

[5] Martone ME, Gupta A, Ludäscher B, Zaslavsky I, Ellisman M (2002). "Federation of Brain Data through Knowledge-guided Mediation", in Neuroscience Databases, A Practical Guide (ed. R Kötter), Kluwer Academic Publishers: 275-292.

[6] Nielsen FA (2004) Bibliography on Neuroinformatics (http://www.imm.dtu.dk/~fn/bib/Nielsen2001BibNeuroinformatics.pdf; current as of 3/8/2004).

[7] Paxinos G, Watson C (1998) The Rat Brain in Stereotaxic Coordinates, San Diego: Academic Press.

[8] Paxinos G, Franklin KBJ (2001) The Mouse Brain in Stereotaxic Coordinates, 2nd ed. San Diego: Academic Press.

[9] Rajasekar A, Wan M, Moore R, Jagatheesan A, Kremenek G (2002), Real-life Experiences with Data Grids: Case Studies in using the SRB, The 6th International Conference on High Performance Computing (HPCAsia-2002), Bangalore, India.

[10] Rex DE, Ma JQ, Toga AW (2003) The LONI Pipeline Processing Environment. Neuroimage 19, 1033-1048.

[11] Rosen GD, Williams AG, Capra JA, Connolly MT, Cruz B, Lu L, et al. (2000) The Mouse Brain Library. (http://www.mbl.org).

[12] Toga AW, Santori EM, Hazani R, Ambach K (1995) A 3D digital map of rat brain. Brain Res. Bull. 38, 77-85.

[13] Toga AW, Thompson PM (1998) Multimodal brain atlases. In Advances in Biomedical Image Databases (ed. Wong S), pp. 53-88. Dordrecht, The Netherlands: Kluwer Academic Press.

[14] Unified Medical Language System (http://www.nlm.nih.gov/research/unls/umlsmain.html)

[15] Van Essen DC et al. (2001) An integrated software suite for surface-based analysis of cerebral cortex. J Am Med Inform Assoc 8, 443-59.

Acknowledgements

The work is supported by NIH grants from NCRR RR04050 and NIDA DA016602 through the Human Brain Project, and NSF grants supporting NPACI NSF-ASC 97-5249 and MCB-9728338. The Biomedical Informatics Research network is supported by NIH grants RR08605-08S1 (BIRN-CC) and RR043050-S1 (Mouse BIRN).