

Emerging Internet Image Archives Visualizing Biological Species and Medical Conditions

This paper discusses Internet sites which enable the user to see images of plants, animals, and human disease conditions on computer monitors. The organization of the electronic archives at these sites commonly mimics bioscience taxonomy. Among the entities experimenting with biological images for the Internet are international and government science agencies, academic and research institutions, businesses, computer labs, interest groups, and innovative individuals.

Digital visualization for the Internet is driven by evolving technology and legacy materials. Most images presented are raster bit-mapped files in the Graphics Interchange Format (GIF, extension: .gif) or use the Joint Photographic Experts Group compression (JPEG/.jpg). GIF and JPEG standards are incorporated in Hypertext Mark-Up Language (HTML) and in software for viewing, editing, and printing images. Platforms, operating systems, and graphics software can almost universally import and export files as GIF and JPEG often though the standard for transporting images (TIFF) between formats. Software "plug-ins" allow additional image file formats to be viewed.

The physical source media for the electronic images of biological subjects include specimen, film and digital photographs, slides, prints, drawings, x-rays, magnetic resonating images, electrocardiograms, and sonograms. Images displayed were originally digitized in dozens of raster and vector image file formats native to proprietary equipment and software used to capture, scan, or edit electronic images.

Human factors figures in visualization. Can people recognize the subject of the image? Recognition involves cognition, culturally determined perception, cultural preferences for style, and the visual experience and training (technical or aesthetic) of the viewer. Composition and clarity critical for human perception may be affected by the view, shape, and indication of scale, details, or color if critical for identification: what about the subject is displayed. Focus, size, tones, background or context, the placement of the subject, and the method used to create a digital image affect human perception: how technically the image is made. Images with high resolution can convey finely detailed content and color tone gradients, can be

by Leslie A. Brownrigg *

magnified (zoomed up), and require a higher density of pixels: the differentiated units of equal size that make up an electronic image. High pixel density implies more bytes of data and larger files. Large files tax the capacity of Internet packet transmission, and serving and receiving computers. "Lossy" principle

compressions (including JPEG) reduce file byte size by literally "losing" visual bits, degrading images irretrievably. Sizes of raster files posted for viewing on the Internet to depict or identify biological species and human disease conditions were observed to range from around five kilobytes for highly compressed photo "thumbnails" and black and white line drawings to hundreds of thousands of kilobytes. The large size of electronic image files is one reason why more image files are posted for downloading than for immediate viewing on the Internet. File size influences the commonly made decision to locate image files in a terminal position, at the end of an archive in an auxiliary Computer Graphics Interface (CGI) bin or CD-ROM that exclusively warehouses image files. On home pages or directories, links to images may be embedded in items of an inventory list of subject or file names, or brief descriptive text, or lossy previews. (1) The terminal or attached position of image files fortuitously permits linking them to and from documents within and among sites on a unique file address.

If directory and file names are semantically meaningful in human language, then their content can be found more easily by Internet search engines which automatically check directory and file names. Using Latin or common names to label the directories and files where information or images about a species are located sets up a smooth interface to browsers. Files named "C:/GIF/Birch.gif" and "flora/Asteraceae/Launeas/L-aborensens/Laborescence.1.jpeg" can be identified as images from their extensions, .gif and .jpeg. One file identifies the subject of the image using a common name, birch, while the other uses a scientific name, L. aborensence. Each Latin Genus and species name is short hand metadata for a unique taxon, defined and characterized in biological data bases and monographs. Bioscience ranks life forms— Kingdom, Phylum, Class, Order, Family, Genus, Species (and their subs and branches). The second file is nested in a cascade of directory names which mark the subject's botanical

classification. Several Internet sites have adapted the taxonomic levels of biological classifications as a structure for their archives of electronic materials about species. materials in hierarchies. Taxonomic levels that group similar life forms gave a jump start to the logical organization of elaborate archives of biological species.

Several Internet enterprises are vying to construct taxonomies or phylogenetic lists to describe and compile data for all the living species on earth— or in a large region. Some of these sites incorporate illustration. Others are choosing technologies incompatible with the special requirements of image files.

The organization of “The Tree of Life: A Phylogenetic Navigation Systems” is instructive. Tree of Life is a federated site distributed on 18 servers in three countries and is linked to additional servers where specialized sites note their on-line “place” in the tree. Tree’s “page” on frogs (Salientia) resides on the University of Texas server along with “Herps of Texas.” Adding a refereed description for a higher order “clade” or “terminal taxon” to the Tree is facilitated by the data entry program MacClade which can read from and to the widely used Phylogenetic Analysis Using Parsimony (PAUP) taxonomic program. Each basic “page” in the Tree of Life covering a taxon or more generalized life form is illustrated with scientific line drawings and compressed photo images. Some images hide the HTML indices to clade Latin names by which Tree’s internal web crawler navigates among its “pages.” Images can be attached or linked at any point to Tree documents and generated from Tree’s random and searchable internal browser facility. Tree’s photographic images range from 11,000 to 328,000+ bytes in size .(2) Tree’s is a collaborative effort among entomologists at the the universities that serve and host this federation. Its content is best developed for insects.

UNESCO’s Man and the Biosphere (MAB), national government scientific agencies, and research or academic institutions also initiated on-line sites with comprehensive ambitions. In a complex of sites sponsored and affiliated with U.S. government agencies, attempts are underway to link sites and to standardize meta-data aspects for the inventory and description of biological species. Agencies in three cabinet-level departments of the United States federal government and collaborators are forging distributed Internet inventory data bases. Next include the US Information Center for the Environment (ICE), National Biological Information Infrastructure (NBII), and “ITIS”- the Inter-agency Taxonomic Information System.(3) Standards, data base layouts, and “meta maker” data entry programs being tested emphasize location and text data and cannot accommodate images. (4) Images on the sites affiliated with this complex are rare, and tend to adorn home pages, or lie in ad hoc galleries, or buried in menus. (5)

Bioscience requires at least one “voucher specimen” be held in a scientific repository such as a botanic garden, herbarium, zoo, museum, university, or research organization to establish and reference the existence and scientific name of a unique species. Many Internet sites experimenting with visualizing originate from these institutions that collect and classify. Some sites cling to traditional presentations of specimen, showing animals confined in aquaria or cages or dried plants laid out on conventional herbarium sheets (6). Others deploy new media to photograph living organisms in fields and forests (7) or through the microscope.

Many institutions post digital peeks into their formidable collections (8) and label images as copyright, hesitant to reveal their research patrimony absent “pay per view” or use fees. Proceeds from the market for digital visuals can potentially reimburse the high costs of processing, archiving, and serving image files. On-line mechanisms include limiting access (on passwords) to subscribers, collecting fees to view or to download files, and receiving a payment for each “hit” to the site as use royalties from subscription services (9) or from advertisers. Commercial tie-ins are possible. Several sites display biological visuals to catalog items for sale and businesses sponsor sites and popular federations to attract customers. Until recently, one herbarium’s Internet presence was hosted and sponsored by a garden supply site. Two popular sites posting images of felines were sponsored respectively by a pet food company and an airline. Access to the bulk of images on-line that visualize human disease conditions is already structured primarily through subscription services. Access to electronic archives of medical imaging tends to be strictly privatized to the archive’s donor/users. An off-line market for digital images exists in CD-ROMs, software, and print ads.

Some of the most visually and technically sophisticated sites featuring numerous images of biological species prepared for immediately viewing on the Internet were designed and initiated by experimenting individuals and computer teams. Henriette Kress issues directories full of image files depicting plants, parsimoniously linking each file name she lists directly to an image file. Kress’ botanical site served from Helsinki, Finland (10) is relayed by “mirror” sites on three continents, a tribute to the site’s outstanding content and popularity. Tim Knight developed and maintains sites dedicated to images and other electronic media to depict the living primates and designed Internet sites for several zoos and scientific associations.(11) Knight’s sites are worth “visiting” for their design and the quality and economy of images. A computer technology group at a Texas university is posting images of crops and wildlife. (12) The Korean Research and Development Information Center’s BIOINFO computer imaging project serves 2300+ optimized animal pictures in its on-line archive which can be searched using English vernacular

names in the Roman alphabet or Korean names and script. (13) Independently, the designers of these four sites pioneered similar technical solutions to problems that Internet visualization projects confront. The exact technology (equipment, software, file format, and post digitalization editing) they used differs. In common, the four sites post attractive color images that can be magnified at least once, in files sized in the modest 20-250 kilobyte range. In common, the four sites created the digital images displayed by scanning film photographs or slides. They select among their own and contributed photographs those that can be successfully scanned at quite low pixel densities (under 100 dots per inch), a technique that limits image file size at the moment of digital creation. Photographs suitable for this technical approach must be in-focus, high contrast, high content, and centered on the subject.

It is striking how many Internet images show the faces of animals and flowers of plants, suggesting face or flower are regarded as the most recognizable or visually interesting feature. Images posted to portray biological species are not standardized as to perspective or view whether the organism is large or microscopic.

By contrast, medical imaging by x-ray, magnetic resonance (MRIs), and other devices to diagnose or monitor patients' disease conditions have strictly standardized views. Protocols exist for positioning subjects in relation to the image capturing device to achieve the correct views. Resulting images require a specialized training to interpret. A principle of file compression can be applied to highly standardized views : to reduce file size by removing extraneous features located in predictable areas of the image rather than to "loss" randomized pixels throughout the image. Fixed view medical images (14) can be stored without the extraneous features, and templates of what typically appears in the removed sections can be reinserted to "decompress" the image for viewing.

Human disease conditions are not registered in a single universal taxonomic hierarchy like that for biological species. There is less consistency in approaches to organizing medical images on the Internet. Each disease considered unique will be identified by its own diagnostic code and by vernacular or scientific names. Diseases are grouped into areas of specialization recognized in the medical community, that is, grouped either as related to a particular organ or system of the body, or to a disease processes, or stage in the life cycle. Thus, medical images cluster on sites for medical specialities. Images of melanoma skin tumors (15) can be found in dermatology and cancer archives.

Additional organizing principles are applied. Medical images on the Internet (and private Intranets) are sometimes organized in archives containing a single

technological type of medical image — just CAT-scans, just MRI, although from different institutions and concerning different patients, perhaps because interpreting each type of image is still another medical specialization. Medical imaging is sometimes grouped across types by source (hospital, clinic, practice) and an identifier like date of deposit. Medical images are organized in the "case" of one patient, either walled off in a confidential archive or broadcast as a didactic history in one of the commercial medical teaching and reference Internet subscription services.

Electronic images depicting human medical conditions produced daily due to their role in diagnosis and treatment are lost to retrieval and research when buried in incompatibly indexed and organized archives.

General observations follow relevant to sites with extensive on-line archives that include image files. Site designs typically redress the intrinsically large size of image files by applying compressions and by exercising selectivity. Giving computer directories and files names that have meaning in natural language makes them easier to search. Archives organized by establishing layered classes that generically includes every particular thing in its class, are easy to navigate.

End Notes (Internet addresses of sites):

(1) *Examples of image to image directories are Missouri Botanic Garden's directory at*

<http://www.mobot.org/mobot/research/gallery.lagallery2.html>

<http://www.turtlebackzoo.org/tbz25.htm>

<http://www.cbs.nl/temp/lmi/ff300rnr.htm>

(2) *Tree of Life's format is explained at*

<http://phylogeny.arizona.edu/tree/home.pages/intro.html> and see

<http://phylogeny.arizona.edu/tree/eukaryotes/animals/chordata/osteostraci/osteostraci.gif>

<http://phylogeny.arizona.edu/tree/eukaryotes/animals/chordata/Ichthyostegalia.gif>

<http://phylogeny.arizona.edu/tree/eukaryotes/animals/invertebrates/hexapoda/coleoptera/colerptera.html>

(3) *For more information about this U.S. government Internet enterprise see*

<http://www.nbs.gov/nbii>

<http://trident.ftc.nrcs.usda.gov/itis/>

<http://ice.ucdavis.edu/>

http://ice.ucdavis.edu/US_National_Park_Service/National_Park_Service.html

http://ice/about_NPS_databases.html

<http://www.fs.fed.us/>

(4) <http://www.itis.usda.gov/images/twbscr.gif>

<http://www.nbs.gov/nbi/current.status.html>

<http://www.fw.vt.edu/fishex/macsis.html>

(5) <http://www.nbs.gov/features/photogallery>

<http://www.itis.usda.gov/images/twbscr>

(6) <http://www.inbio.ac/cr/tipos/herbario/Cgrayumii.jpg>

http://www.mpiz-koeln.mpg.de:80/~stueber/ross/potato/sucense_complete.jpg

(7) <http://bluehen.ags.udel.edu/udgarden.html>

(8) <http://www.nfrcg.gov.htm>

<http://www.nmhn.cgi-in/wdb/fish/catalog/query/catgno==00334647>

(9) <http://members.aol.com/zooweb/pictures>

(10) <http://sunsite.unc.edu/herbmed/pictures>

<http://sunsite.sut.ac.jp/pub/acadmeic/medicinal/alternative-healthcare/herbal-medicine/unc.edu>

(11) <http://www.selu.com/~bio/PrimateGallery/primates/species.html/index.html>

<http://www.selu.com/~bio/>

(12) <http://www.ics.tamu.edu/FLORA/gallery.htm>

<http://leviathan.tamu.edu:701s/slides.ef/>

<http://straylight.tamu.edu/tamu/tracy/chklcon.html>

<http://ranch.tamu.edu/rlem/>

(13) <http://bioinfo.kordic.re.kr/animal/>

(14) [http://www.laurie.umdj.edu/database/CollectedApps.acgi\\$ImageList_jpeg?11843](http://www.laurie.umdj.edu/database/CollectedApps.acgi$ImageList_jpeg?11843)

(15) See the Dermatology Online Atlas from the Erlanger Image Database

http://ultaet/med/kli/derma/biblddb/diagnose/dg_m.htm

and see *MD Challenger sample photo:*

<http://www.embbs.com/img/i0000012.jpg>

* Paper presented at IASSIST/IFDO '97, Odense, Denmark, May 6-9, 1997. Leslie A. Brownrigg, United States Bureau of the Census