Data, data everywhere, nor an image to read - Finding open image databases.

Release 0.00

David R. Holmes, III and Richard A. Robb

July 15, 2007

Biomedical Imaging Resource, Mayo Clinic College of Medicine, Rochester, MN

Abstract

Open Science includes access to both open source software/methodologies and open data. While there has been progress in open image databases, the results of these efforts are under-reported. As such, imaging scientists are unaware of the available data. In addition, many researchers are interested in providing their data to the greater research community, but may be unaware of the process to release the data. The purpose of this paper is to describe our efforts in developing an open website which includes information on accessible medical image databases as well as some of the logistics for providing an open image database. Most importantly, the authors are requesting participation from the community to contribute to the Medical Image Database Repository (http://midr.org) in order to consolidate the collect knowledge of the community.

Contents

At the MICCIA 2005 Open Source Workshop, the keynote speaker, Dr. Michael Vannier, challenged the open source community to develop new image processing methodologies which are clinically translatable. During his talk, Dr.Vannier observed that there are massive amounts of medical image data but the research community is not utilizing it. This led the authors to consider a quote from “The Rime of the Ancient Mariner” – *Water, water everywhere, nor a drop to drink.* In the context of Dr Vannier’s comments, we would say ”Data, data everywhere, nor an image to read.” The primary challenge is that most acquired image data is clinical and often unavailable to the imaging scientist. In addition, the data that is available

through open databases is often under-promoted and overlooked by researchers. The purpose of this paper is to describe the development of an open website where the image analysis community can contribute and consolidate knowledge on open image databases as well as approaches to providing open data in the future.

A recent article in Biomedical Computation Review [4] provides a nice overview of some of the issues relating to image databases. Appropriately, the article begins with a description of the Visible Human Project [2]. The VHP has been a very successful image collection with wide-spread acceptance throughout the field. In additional, it is spawned the development of several other VHP-like image collections[12, 7]. The VHP project has been successful for many reasons; however, we suggest that the particular strength of the VHP is that the project was specifically targeted as an open image database with clear specifications for the outcome. There have been several other image databases [6, 8, 10] which have shown similar success. In most cases, the intended use of the databases is unique, therefore providing an important and novel contribution to the community. These examples have all been well-accepted and utilized by the research community. In order to best inform the imaging community, we suggest that the consolidation of the communities' knowledge of medical image databases will help to further disseminate both existing and future image databases.

1 Implementation

Our approach to consolidating and disseminating information on medical image databases is to develop a wiki-based website which will store meta-information about existing medical image databases. The wiki-based approach was chosen over other, more traditional, approaches such as a conventional website or newsgroup, because it provides structured access to all of the data while maintaining an open process to add/edit content by the community.

2 Medical Image Database Repository (http://www.midr.org)

The website has been built with conventional web development tools. The front matter is standard html with links into the wiki(See Figure 1). MediaWiki [5] , a popular wiki engine, was used in the development of the wiki pages. The wiki content is split into two major categories - "databases" and "logistics." The Database Wiki is organized with a page per database. Each database page includes database contact and content information. The Logistic Wiki is a more general wiki will included other information related to image databases such as HIPPA requirements [9], suggestions for text on open databases to be included in IRB or Animal Care Protocols, and links to relevant software.

3 The Database Wiki

The intent of the Database Wiki is to provide basic meta-information about various open image databases available online; at this time, neither the wiki nor the website will provide the data from an image database. In order to provide a common look-and-feel across the database descriptions, a template has been built which contains all of the proposed content headers. The content is divided into four major categories - introduction and text description of the database, contact information, database details, and reference material. A screenshot of a wiki page is shown in Figure 2. Details on the categorical data are shown in Table 1.

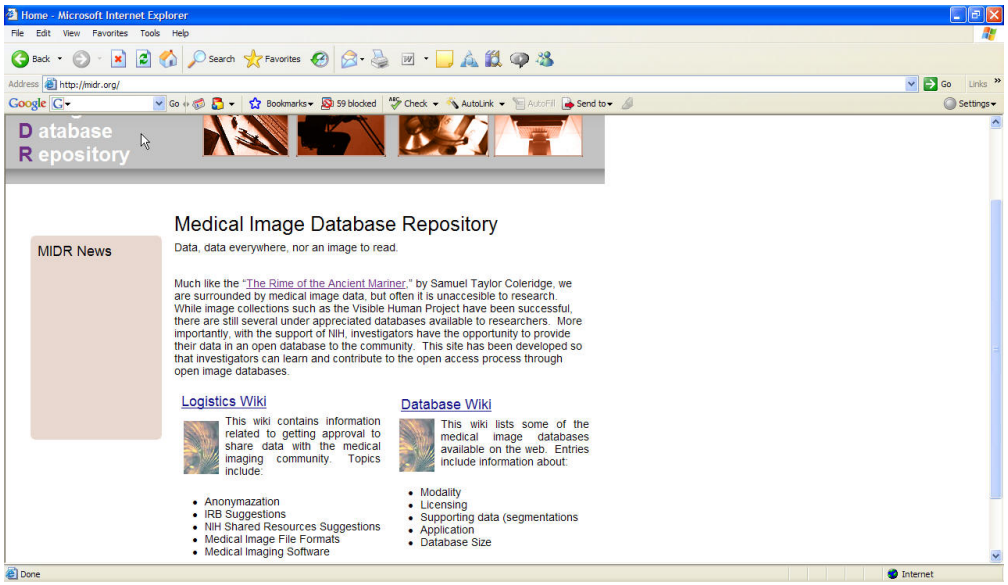

Figure 1: Snapshot from the Medical Image Database Repository website (http://midr.org)

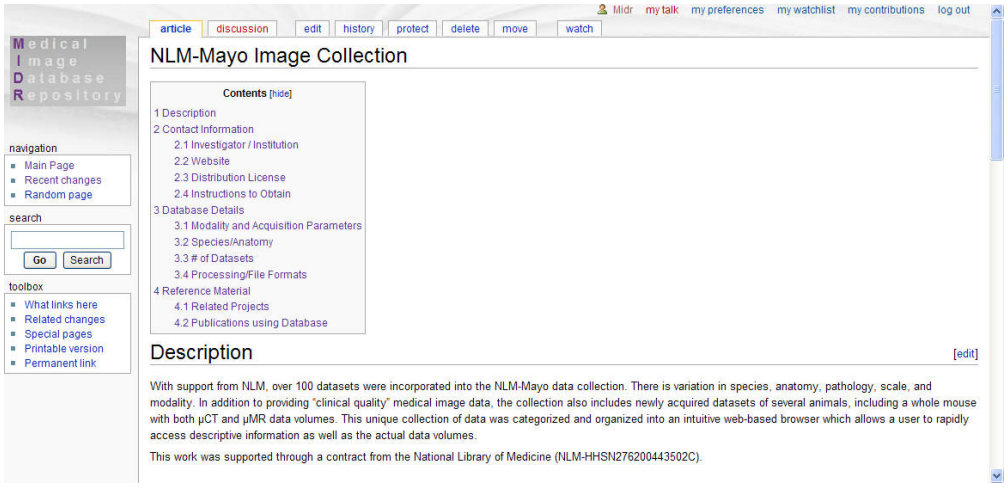

Figure 2: Snapshot from the Database Wiki using [3] as an example.

Table 1: Proposed content for image database entries.

Abstract
Contact Information
Investigator / Institution
Website
Distribution License
Instructions to Obtain
Database Details
Modality and Acquisition Parameters
Species/Anatomy
of Datasets
Processing / File Formats
Reference Material
Related Projects
Publications using Database

4 The Logistics Wiki

The Logistics Wiki serves as the primary repository of information relating to the creation, management, distribution, and processing of medical image databases. Articles on the regulatory aspects of image data dissemination will be particularly important for researchers interested in disseminating data. It is the hope of the authors that boilerplate text can be developed for insertion into IRB protocols based on examples of previously successful IRB protocols. A small section of the wiki will be devoted to medical image software for anonymization, file format conversion, and image processing; however, it is understood that there are already successful examples of wiki-style software lists on the web [1].

5 Discussion

All too often, research is hindered by inadequate tools and resources. Open science, if embraced, promises to enable the research community through the sharing of knowledge, tools, and resources. There have been very successful examples of both open source software [11] and open data collections [2]; however, there are many other resources available to researcher in the community. While search engines such as Google, provide a cursory approach to find out about open image databases, the time required to parse through false hits or unmaintained websites is substantial.

We are hopeful that a community-maintained resource, such as midr.org, will provide a common repository of information on open medical image databases. The continuing challenge, however, will be to active engage members of the research community to effectively contribute. While the authors will be actively updating and supporting this website, it is essential that other provide content as well to ensure broad recognition and coverage of available databases. Unmaintained or incomplete data will hinder future efforts.

References

[1] A Crabb. idoimaging.com: Free medical imaging software. http://idoimaging.com. 4

[2] MJ Ackerman. Accessing the visible human project. *D-lib Magazine*, Oct, 1995. (document), 5

[3] DR Holmes III, EL Workman, and RA Robb. The nlm-mayo image collection: Common access to uncommon data. *ISC/NA-MIC/MICCAI Workshop on Open Source Software*, 2005. 2

[4] MA Kunz. Imaging collections: How they're stacking up. *Biomedical Computation Reivew*, Summer, 2007. (document)

[5] MediaWiki. Mediawiki. http://mediawiki.org. 2

[6] Montreal Neuologic Institute. Brainweb. http://www.bic.mni.mcgill.ca/brainweb/. (document)

[7] JS Park, MS Chung, and et al. Visible korean human: Improved serially sectioned images of the entire body. *IEEE Trans Med Imaging*, 24(3):352–360, 2005. (document)

[8] UC Davis. The brain atlas project. http://nir.cs.ucdavis.edu/. (document)

[9] US Department of Health and Human Services. Medical privacy - national standards to protect the privacy of personal health information. http://www.hhs.gov/ocr/hippa. 2

[10] J West, JM Fitzpatrick, MY Wang, and et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, 1997. (document)

[11] TS Yoo, MJ Ackerman, WE Lorensen, W Schroeder, V Chalana, S Aylward, D Metaxes, and R Whitaker. Engineering and algorithm design for an image processing api: A technical report on itk - the insight toolkit. *In Proc. of Medicine Meets Virtual Reality*, pages 586–592, 2002. 5

[12] SX Zhang, PA Heng, and et al. Creation of the chinese visible human data set. *The Anatomical Record*, 275B:190–195, 2003. (document)