The Virtual Skeleton Database: An open access repository for biomedical research and collaboration

Michael Kistler *, Marcel Pfahrer †, Roman Niklaus †, Viktor Tomas† and Philippe Buechler *
*Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland
†Engineering and Information Technology, Bern University of Applied Sciences, Switzerland
contact: michael.kistler@istb.unibe.ch

***Abstract*—The access to the largest source, the images stored by clinical institutions in their PACS, is for scientific researches rarely possible due to technical and legal restrictions. Even after this time consuming process, collections of datasets are often lost or mishandled resulting in replication of work, frequently within the same institution or network. To solve these problems, we proposed a centralized storage system called the Virtual Skeleton Database (VSD).The VSD provides a system tailored to the needs of the medical image analysis community. The VSD offers generic tools to store, exchange and collaborate on virtually any digital format. The hosted data is accessible for the community while collaboration and access tools catalyse their productivity.**

I. Introduction

Parallel to the digital innovations on consumer level, the medical research community is in progress to adapt to new technology for working with medial image data. They move away from local device storage like Hard-Drives and Compact-Disk to collaborative centralised applications. In the medical field, the picture archiving and communication system (PACS) were introduced and are now widely accepted in clinical institutions. However, while ideal as a medical images repository, PACS are patient centric and remain not flexible enough for the needs of biomedical researchers. Inspired by PACS, several projects were realized for the biomedical research community. As an example, the Extensible Neuroimaging Archive Toolkit (XNAT) [1] has developed an open source solution written in JAVA for the for the Neuroimaging research community where collaborators can store and share datasets in their projects. Several other projects are building infrastructure to host data collections in their respective research domain. For example, the Living Human Digital Library [2] aims to provide a large collection of raw and processed data on the anatomo-functional characteristics of the human musculo-skeletal system at dimensional scales spanning from the whole body down to the molecules. The Internet Brain Segmentation Repository [3] provides manually-guided expert segmentation results along with magnetic resonance brain image data. Although these projects provide extremely important data to the scientific community. However, they either investigate a limited number of subjects, are specific to a certain organ or pathology, or present a mix of patients and pathologies. Recently, the MIDAS media archiving platform [4] has emerged. MIDAS extends the functionality of DSpace [5] to fit the needs of medical and scientific meta-data and provides a collaborative platform.

The objective of the Virtual Skeleton Database (VSD) is to provide and maintain an open access repository for scientific researchers to find, preserve, trace, and share data resources. The goal of the VSD is to establish and support a research community which collaborates at the data processing and content acquisition.

II. Methods

The general features of an online repository include security, storage and data integrity, ease of use, quality of the content and data persistence. For medical image data, the legal, the privacy and ethical guidelines are of key importance. However, if the datasets are used for research purposes only, these aspects can basically be reduced to privacy protection. Therefore, a registration procedure which forces the VSD to exist as a research community is implemented. The individual research unit is administering their members to control the access to the research data.

High quality datasets, searchability, and consistency are vital for a repository. Therefore, the VSD is collecting and hosting 100 post-mortem full body CT datasets, calibrated by the European Spine Phantom [6] to allow for proper mapping of bone mineral content to mechanical property. To ensure searchability and consistent meta-data, the VSD features a reviewing concept similar to Blog software like WordPress. Before publishing to the community, the owner is asked to review his data and to fill missing information. This process includes dataset identification, the creative-commons license is specified, relations to existing content can be established, and for the meta-data annotation of anatomical structures, the Foundational Model of Anatomy (FMA) [7] ontology is used. Standard file permission sets can be used to selectively distribute objects and organized into folders. Using the folder sharing options, users can easily achieve a collaborative resource management or can share entire collections. For each object, user can set a rating, they can add a personal comment or they can initiate an objects discussion.

To ensure the system's flexibility, the data model of the VSD is centered around the concept of generic objects. An object can be any kind of medical image format, labeled images, 3D model or more complex object structure to store statistical shape models. Up to date, the VSD can read classical DICOM file through ClearCanvas framework, image types recognized by ITK library (MetaImage, Niftii and Analyze) and the HDF5 file structure for statistical model utilized by the statismo framework (www.statismo.org) [8]. During the upload process, meta-data is extracted from the images and the datasets are anonymized. The anonymization consist of filetype specific header manipulation and replacing filenames

by a generic filename, preserving the information for the uploader. To keep the system storage efficient, the uploading functionality has mechanisms to avoid duplicated objects on the disk and supports object versioning.

III. Results

Currently approximately 2'500 objects were uploaded by about 100 users distribute into ten research units are registered on the VSD, including 40 full body post-mortem CT datasets. Additional 60 full body sets are scanned and are being prepared for upload. In addition, over 600 head dataset with mandibular and orbital segmentation, 150 CBCT and more than 800 CT and segmentation for the lower extremities are or will be available as the upload is progressing.

A. Multimodal Brain Tumor Segmentation Challennge

The VSD was used to host a Challenge on Multimodal Brain Tumor Segmentation (BRATS) [9] which was part of the MICCAI conference 2012 in Nice. The participants registered on the website to the BRATS research unit group and download the designated datasets which were collected, uploaded and organized on the VSD website by the organizers. Participants uploaded their result to the VSD and the system automatically recognized the challenge segmentation using the VSD naming convention. The evaluation of the Brain Segmentations was initiate by the user from the Web interface and performed by an integrated ITK script. Individual results are presented on the website and detailed scores can be exported for further analysis.

IV. Conclusion

The VSD is a flexible system equipped with an ontology powered search, compelling collection of full body datasets. The system proved its value and concept of being a generic repository by hosting an Tumor Segmentation Challenge online. The registration, data management and sharing functionality were used without any change to the VSD and the system provides the required background processing. The concept drew attention from different research field and is supported by three Swiss research institution. This drives further development and will support the community. Therefore, the VSD's potential to serve as a research repository for an interdispiplinary audience will be further exploited.

References

[1] D. S. Marcus, T. R. Olsen, M. Ramaratnam, and R. L. Buckner, "The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data." *Neuroinformatics*, vol. 5, no. 1, pp. 11–34, Jan. 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17426351

[2] "Living human project." [Online]. Available: http://www.livinghuman.org/

[3] "Internet Brain Segmentation Repository." [Online]. Available: http://www.cma.mgh.harvard.edu/ibsr/

[4] J. Jomier, S. R. Aylward, C. Marion, J. Lee, and M. Styner, "A digital archiving system and distributed server-side processing of large datasets," K. M. Siddiqui and B. J. Liu, Eds., Feb. 2009, pp. 726 413–726 413–8. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.812227

[5] "DSpace, an open-source solution for accessing, managing and preserving scholarly works." [Online]. Available: http://www.dspace.org/

[6] W. A. Kalender, D. Felsenberg, H. K. Genant, M. Fischer, J. Dequeker, and J. Reeve, "The European Spine Phantom–a tool for standardization and quality control in spinal bone mineral measurements by DXA and QCT." *European journal of radiology*, vol. 20, no. 2, pp. 83–92, Jul. 1995. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/7588873

[7] C. Rosse and L. V. M. Jr, "A reference ontology for biomedical informatics : the Foundational Model of Anatomy," vol. 36, pp. 478–500, 2003.

[8] M. Lüthi, T. Albrecht, T. Gass, O. Goksel, M. Kistler, H. Bousleiman, M. Reyes, B. Philippe, P. C. Cattin, and T. Vetter, "Statismo - A framework for PCA based statistical models ." *The Insight Journal*, pp. 1–18, 2012. [Online]. Available: http://www.insight-journal.org/browse/publication/871

[9] B. Menze, A. Jakab, S. Bauer, M. Reyes, M. Prastawa, and K. van Leemput, "BRATS 2012 - MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation," 2012. [Online]. Available: http://www.imm.dtu.dk/projects/BRATS2012