

The XMA portal. A web environment for data management

Kia Huffman and Elizabeth Brainerd, Department of Ecology and Evolutionary Biology, Brown University

Jaime E. Combariza Center for Computation and Visualization, Brown University

Many academic institutions are currently deploying resources so that researchers can comply with funding agencies requirements to effectively store, organize, document, and disseminate data sets. This advances science through reuse of data and by fostering collaborations. Sustainable data management best practices and data curation are critical strategies for data intensive research¹ because data sets are becoming the new instruments of science. Reliable access to data, as well as being able to navigate through the data, are critical steps to simplify the user experience and to enable fruitful collaborations, validation and verification of results. At Brown University, we are currently developing centralized resources that should facilitate data management by interfacing file systems with digital repositories as the entry point. There are a few research groups, however, that are developing their own processes and resources with the intent to produce a sustainable solution for their own disciplines.

Our case study was developed in the Brainerd research lab and comprises a file system (24 terabytes), a database (MySQL) and web portal (PHP, Java, JavaScript, Python) for access, management, storage, and sharing of video data and metadata for X-ray Motion Analysis (XMA) studies. The Keck Foundation funded a project to develop the “X-ray Reconstruction of Moving Morphology, XROMM^{2,3}”, multiuser facility (currently users from Brown University, Providence College, Rhode Island Hospital, Tufts Veterinary School, Harvard University, The University of Nevada, Las Vegas, the University of Montana and others). XROMM is a set of 3D x-ray motion analysis methods that combine skeletal movement data from *in vivo* x-ray videos with skeletal morphology from 3D bone scans (e.g. computed tomography (CT) or laser scans).

In the past, when different datasets needed to be shared among users geographically distant, the process was tedious, extremely inefficient and often involved multiple copies of datasets shared physically via ground mail. The Keck Foundation funding was used to establish a central repository housed at the Center for Computation and Visualization (CCV)⁴, which provides the cyberinfrastructure for research at Brown University. This file system is UPS protected and is backed up to tape to preserve data against hardware failure or for a potential disaster recovery situation. Subsequently, local and offsite users were able to upload data via a single user system (anonymous user). Users were expected to document their own datasets, but very few complied. Documentation was often not complete, thus making reuse of the data a difficult process even though all datasets were available to all users. While this simple

repository provided a quick temporary solution, we realized further implementation was needed to include an easy format for documentation/metadata. Furthermore, one of the most important issues in data management is to provide readily secure access to potential users or collaborators without losing ownership of the data. PIs wanted the capability to make data public or private and to grant permission to access their data. Users are now able to reproduce results and use the repository for citation and dissemination. Figure 1, shows the overall data organization within the XMA portal. Repositories contain several studies and these studies may be associated with one or more PIs and the user who is conducting the experiments.

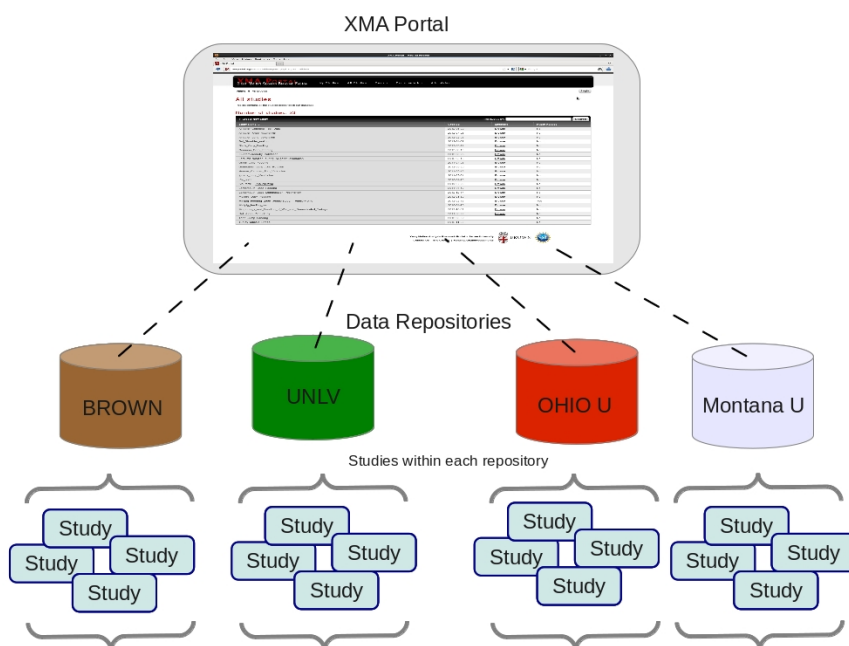


Figure 1. Data organization within the XMA portal

In 2009, funding from NSF was used to establish a Research Coordination Network for X-ray Motion Analysis (RCN-XMA). This project includes offering 1-week summer short courses at Brown and UNLV and also developing an XMA Web Portal, database, and collaboration hotspot for sharing software, tools, information and data. This group has provided valuable feedback to add features and improve the users experience.

A front end or portal was developed to facilitate and control access to the data. We evaluated some powerful existing portal and content/data management solutions such as GridSphere⁵, Liferay⁶, Jetspeed⁷, Django⁸, Drupal⁹, Fedora¹⁰, i-RODS¹¹ and finally decided to establish a data management system for XROMM data because we wanted

to be as flexible as possible to accommodate future needs. We chose to build a custom portal using PHP for front-end, and MySQL database for the backend. In addition, collecting or producing metadata is a process that users tend to avoid unless it is simple and straightforward. We collect minimal information; the menus are self explanatory enough that users add all necessary information. The portal allows the user to easily add information about the data, which is immediately added to the database to transparently create the metadata (see Figure 2).

XMA PORTAL Sandbox
X-RAY MOTION ANALYSIS RESEARCH PORTAL My Studies All Studies Search Document

Home » My Studies » Jaime_test_study_paper » Add New Trial

Add New Trial

Trial Number: 2

Use trial template: Reverse trial list

*Trial Name: (max. 50 char., alphanumeric char., blank s

*Trial Date: Year: Month: Day: OR Unknown

Item marked with * is required

Individ. Research Subject Name: [Add Additional](#)

Genus & Species: [Add Additional](#)

Weight: Unit:

Age: Year: Month: Day: Hour:

Behavior: [Add Additional](#)

Attribute 1: Temp [Add Additional](#)

Attribute 2: [Add Additional](#)

Attribute 3: [Add Additional](#)

Comment (max. 250 char.):

Trial performed with instruments in: Brown/Keck Brown/C-arm UNLV OhioU

Figure 2. Portal menu to create a trial/study and generate metadata

The data collected at this facility is mainly used for X-ray movies, and sizes vary from about 500MB to 7 GB per movie (each camera can record up to 1000 images/second, so the data for a 3 sec movie could be around 6 GB). In addition, we collect auxiliary data such as 3D scans/CTscans/MRIs. A typical study can contain hundreds of movies, so there is a demand for large-scale storage.

A ticketing system (OS-ticket) is used to submit tickets when users need new features or encounter bugs. Documentation is available through a Wiki that shows novice users how to interact with the portal¹². Currently, we have 40 users from 6 different institutions on the xma portal and we are hosting 23 different XROMM studies. We are in process of inviting other institutions, Ohio University and University of Montana to upload their XROMM data into our portal repository.

The XMA Portal solves many fundamental problems by integrating essential calibration files and metadata linked to the raw video data, however accessing XROMM data for analysis remains a bottleneck in the workflow. The XMA Portal could potentially take over these preparation tasks (on a local server), delivering a package of video data, metadata, and calibration images to the analysis platform, and then receiving processed data and information about the analysis process (i.e. provenance) and cataloging them back into the database.

Feedback from users is, in general, very positive, as they are pleased with the way they can easily manage data and metadata in one place and the data sharing mechanism with other researchers is highly appreciated. Users are concerned with the long wait time for loading large size video files, which is exacerbated by low speed internet connectivity. As a solution we plan to make high quality compressed movie versions of the original movies. We are quantitatively evaluating if the compressed movies are comparable in quality, for analysis purposes, with the original uncompressed movies.

References

- 1 - Managing Research Data, Edited by Graham Pryor, 2012
- 2 - www.xromm.org
- 3 - java-source.net/open-source/portals/gridsphere
- 4 - **Brainerd, E.L., D.B. Baier, S.M. Gatesy, T.L. Hedrick, K.A. Metzger, S.L. Gilbert, and J.J. Crisco.** (2010). X-ray Reconstruction of Moving Morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. *Journal of Experimental Zoology, Part A*. **313A**: 262–279. DOI: 10.1002/jez.589
- 5 - www.brown.edu/Departments/CCV
- 6 - www.Liferay.com
- 7 - portals.apache.org/jetspeed-2/
- 8 - www.djangoproject.com/
- 9 - www.drupal.org
- 10 - www.fedoraproject.org
- 11 - www.irods.org
- 12 - wiki.brown.edu/confluence/display/ctx/XMA+Portal+User+Manual