Development of XROMM Data Management (XDM) tool for an integrated workflow for big image/video data in orofacial research

Munakami, Milson[1]; Takahashi, Kazutaka[1,2]; Runesha, Hakizumwami Birali[1]; Ross, Callum F.[2]

1. Office of Research and National Laboratories, Research Computing Center (RCC), The University of Chicago, Chicago, IL, 60637, USA
2. Department of Organismal Biology and Anatomy, The University of Chicago, IL, 60637, USA

Abstract

In order to capture orofacial behaviors, most estimates of skeletal kinematics used to almost solely rely on motion capture of external markers attached to skin or tight clothing, but these estimates suffered from poor fidelity of skin movement to underlying bone movement. X-ray Reconstruction of Moving Morphology (XROMM) is a 3D imaging technology, developed originally at Brown University, for visualizing rapid skeletal and maker movements *in vivo*. The premises on which XDM is built are to serve as a powerful tool for organizing, searching, and disseminating a large amount of XROMM datasets: raw data files (calibration files, raw video data, 3D models) and associated metadata (trial date, experimenter, provenance, access restrictions, etc.) that were generated by the capability of fast sequential captures developed at the University of Chicago XROMM facility.

Introduction

The University of Chicago (UC) XROMM Facility is a Shared Research Facility thus it is important to follow these SOPs regarding data management in order for users to perform their research effectively and securely. This XROMM Data Management (XDM) Portal gives an easy to use web-based interface to facilitate the XROMM workflow – from computer controlled X-ray instruments to a secure data storage and database that will be very easy to use and share data among researchers.

In order to accommodate a massive amount of video data that can be recorded at UC XROMM facility, we wanted to extend what Brown University has developed to add:

1. State-of-the-art Data curation tool: A web-based curation tool that allows researchers to tag and submit metadata to characterize wide range of behaviors, especially orofacial behaviors across multiple animal species.
2. Connecting Subject, Study, and Trials: The RESTful API services to save, edit and update different Subject, Study, Trials based on user permission levels.
3. Support secure file transfer and data storage: Supports single or 'batch' file transfer along with associated large datasets/files at once to facilitate long-term data preservation.
4. Interdisciplinary effort integration: User can login with their corresponding institutional logins and use the portal seamlessly using secure authentication.

Methods

The core challenge involves transferring the terabytes of data users accumulate while using XROMM to DaLI (Data Lifecycle Instrument) infrastructure in a safe and secure manner, from where it can then be accessed and searched for further analysis. Using this user-friendly web tool that enables users to create studies, upload trial data, and associate metadata with the uploaded files can overcome this great challenge. This web portal can initiate one at a time or batch transfers of files to DaLI and displays the metadata to the UChicago XMA Portal that allows users to search, access and store securely.
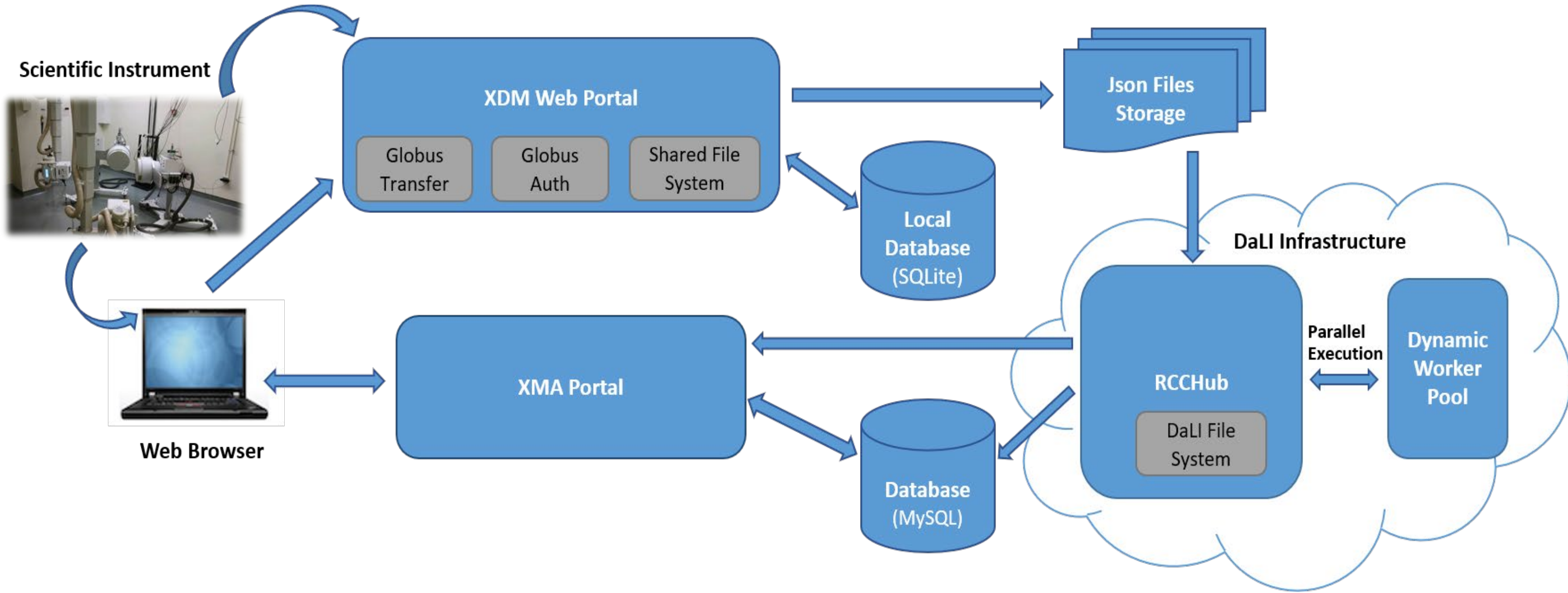
1) XDM Web Portal which gives an interactive web-based interface for researchers with valid credentials. For each XROMM experiment, data is generated from multiple machines within the XROMM facility.
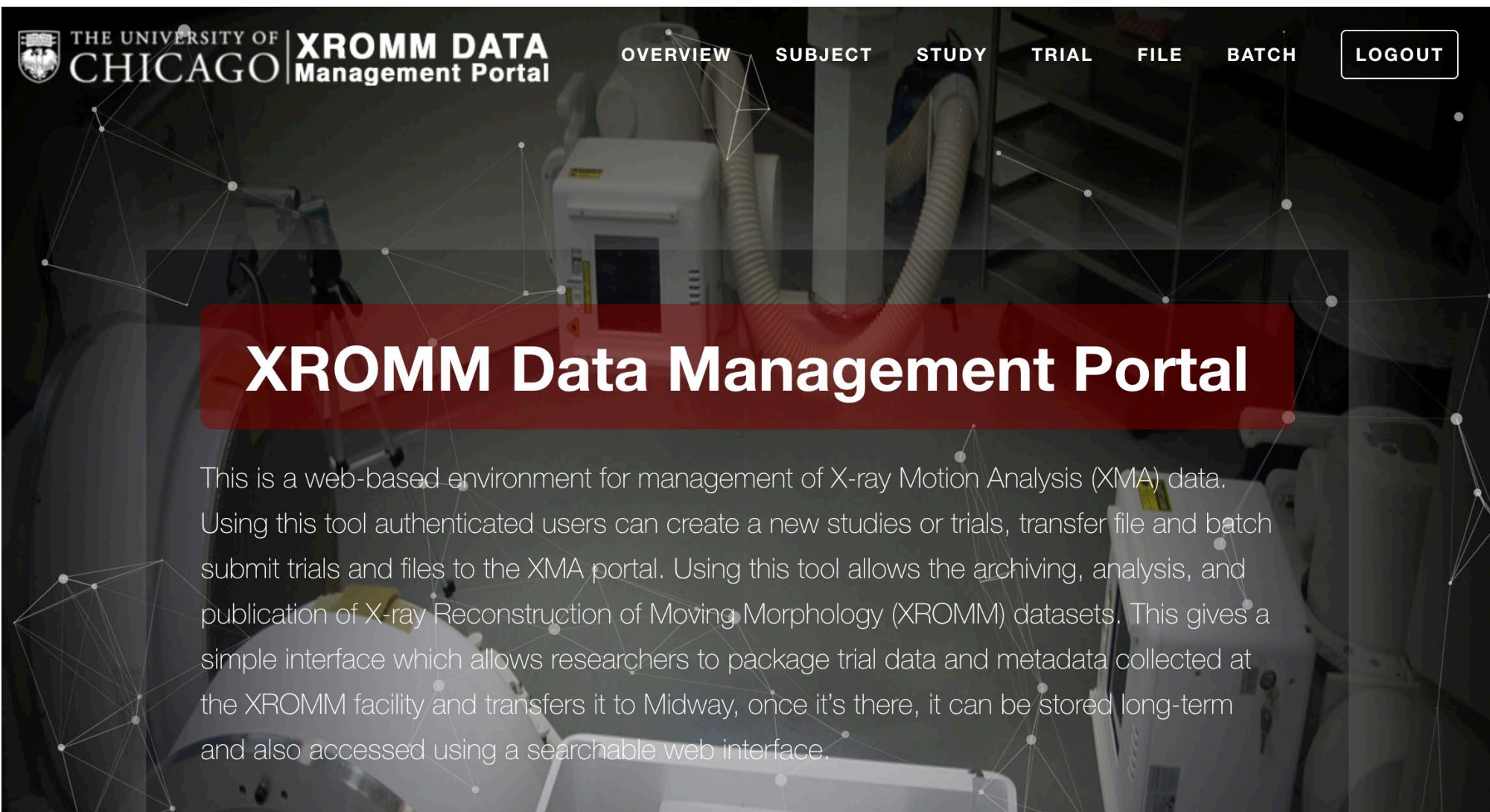
Using "SUBJECT" and "STUDY", a user can define subject and study properties.

Data sources associated with each experiment can include x-ray video, neurophysiology data, EMG data, force recordings, and anatomical models. Complex metadata detailing the nature, calibrations, and provenance of each experimental trial is also collected. Researchers format the standardized Task Parameter Sheet that keeps the experimental logs along with fine-grained metadata records per

Batch Trial / File Upload

data trial and file. Thus, organized trials are batch-uploaded using the interface. It validates and informs any potential user errors. Globus based Auth service and Transfer REST APIs are implemented to ensure the authentication along with a high-speed, reliable and secure data movement to DaLI storage. It validates only access to the authenticated resources and integrity of the transferred files. The massive amount of data and associated metadata asynchronously transferred via the web portal is stored in JSON (JavaScript Object Notation) encoded format.

Results/Systems

Overall architecture - The XDM system is comprised of 4 core components:

1) XDM Web Portal
2) DaLI Infrastructure
3) RCCHub
4) UChicago XMA Portal

2) DaLI (Data Lifecycle Instrument) Infrastructure provides all high performance computing resources allowing rapid analysis and guiding curated data to secure permanent storage. The in-built fine-grained privacy control mechanisms ensure file-level data security by clarifying who will have access to it, how and when based on underlying DaLI file system. (DaLI) enables management and sharing of data from instruments and observations, allowing researchers to:
* acquire, transfer, process, and store data from experiments and observations in a unified workflow.
* manage data collections over their entire lifecycle.
* share and publish data.
* enhance outreach and education opportunities.

3) RCCHub is where the stored JSON files per trial and file are parsed and then pushed to the relational database system for mapping, indexing and storing for quick and faster search and sharing. The uploaded large data along with their associated complex metadata need to be continuously monitored, analyzed, processed and converted to the format suitable for web browsers, long-term data preservation and usable by the 3rd party software. For these purposes, scalable dynamic workers are running in the background that triggers distributed parallel execution of scripted jobs. Some of the generic jobs include database synchronization, regular backups, continuous lossless video conversions, frame-by-frame images generation per video file, file-level fine-grained privacy controls etc.

Examples of XMA Portal interfaces: List of available trials for a given study

- Many of metadata entries can be adjusted for a given study
- Metadata entry changes need to be done in the task parameter sheet that a group of users use for a given study

Examples of XMA Portal interfaces: Multiple camera views of data

Biplanar X-ray data with a behavior data on the browser interface

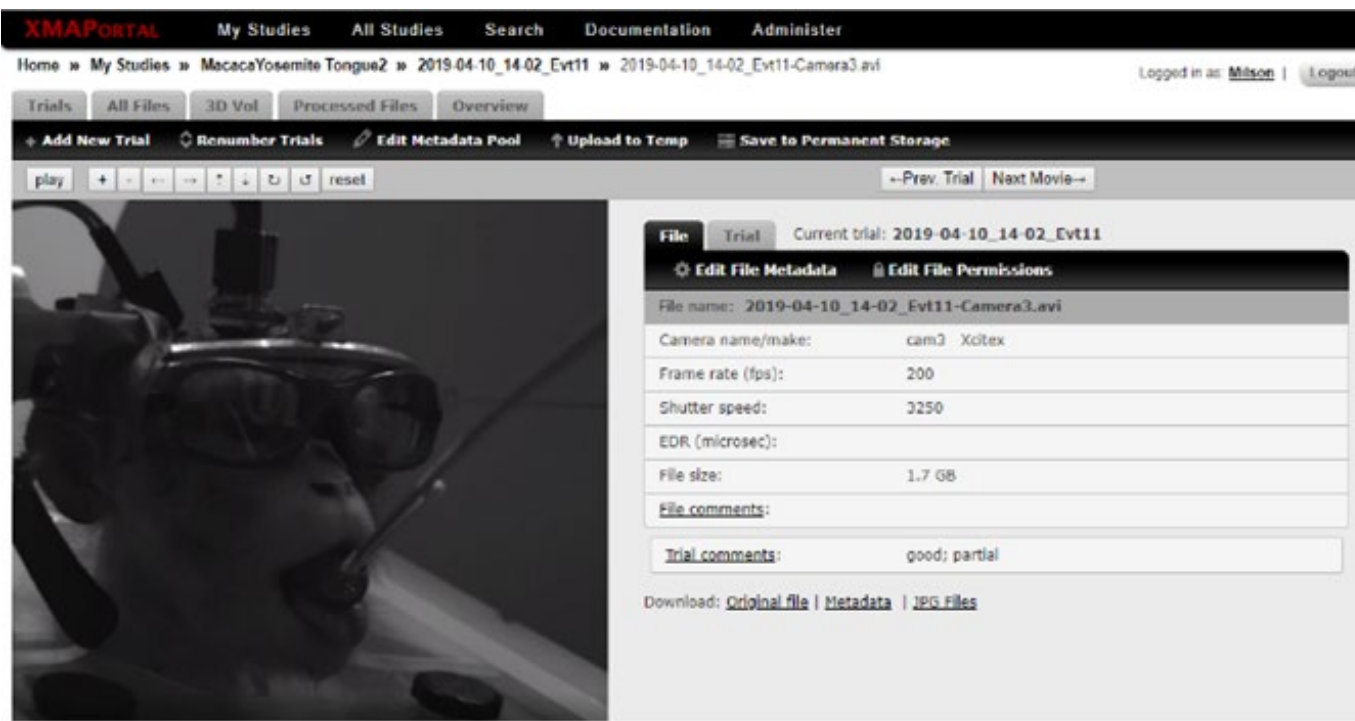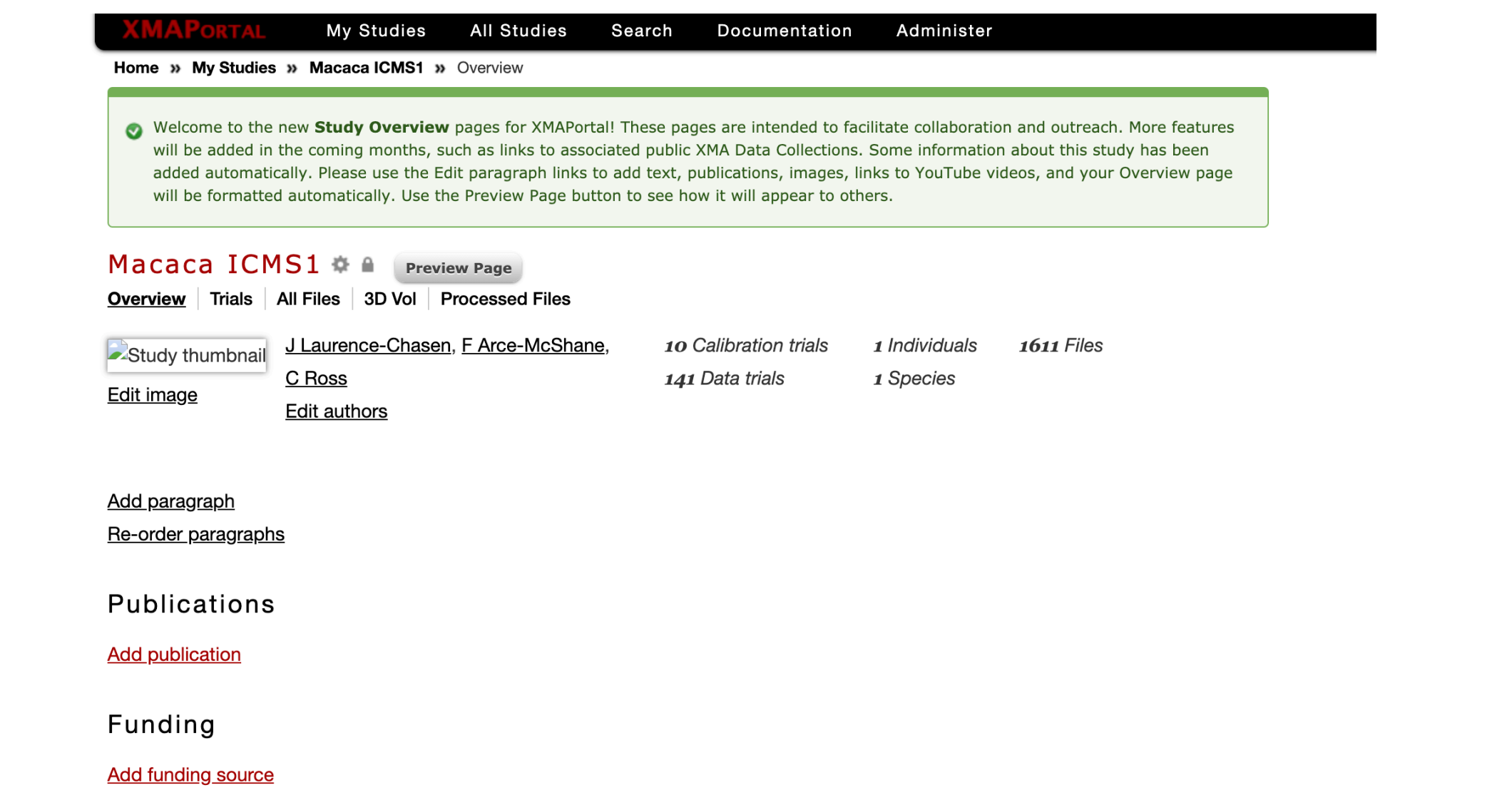- One camera view with metadata associated with the video file

Examples of XMA Portal interfaces: Study summary and Publication/Funding section

Publication database (CrossRef) would be linked for future DOI management for internal (including associated data for publications) and external purposes.

Discussion/Future work

The current XMD system can securely and reliably transfer and organize a massive amount of XROMM data from the facility to the XMA portal. The additional features such as integrating and semi-automating digitization of image files and analyzing kinematics will be useful to streamline the entire XROMM workflow and to reduce the process time to obtain behavioral results.