Award Number: W81XWH-11-2-0226

TITLE: Integrative Lifecourse and Genetic Analysis of Military Working Dogs

PRINCIPAL INVESTIGATOR: Kun Huang, Ph.D.

CONTRACTING ORGANIZATION: The Ohio State University
Columbus, OH 43210-1016

REPORT DATE: October 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2012	Annual	25 September 2011–24 September 2012

4. TITLE AND SUBTITLE

Integrative Lifecourse and Genetic Analysis of Military Working Dogs

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-2-0226

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Kun Huang

E-Mail: kun.huang@osumc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Ohio State University
Columbus, OH 43210-1016

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
During the first year of this three year project, we have made strong progress on the proposed aims. The major accomplishments have been in the following areas: creation of our military dog registry database, determination of the optimal protocols to convert clinical records from paper to electronic versions, establishing molecular genetic methodologies, and development of statistical methodologies for genomic analysis. Regarding the military dogs to be studied, we established a close working relationship with collaborators at the Daniel E. Holland Military Working Dog Hospital at Lackland AFB. We hired and placed in that location a veterinary technician who is responsible for acquisition of on-site data and biological materials. We developed multiple new methodologies capable of handling the diverse data types seen in the military working dog program. We acquired information on the breeding program that will allow us to focus in on the most informative dogs for our molecular studies. Most importantly, we conceived of and validated a novel statistical methodology to identify signal from noise in high throughput genetic or genomic data (which we call GIA). This success with GIA goes beyond our expectation and we anticipate it will result in a highly cited article and a widely used methodology; it will also improve our power to detect cancer-associated variation in our study. In addition to those planned activities, we identified other opportunities to acquire high value data and resources by establishing collaborations with other groups - mainly the US Joint Pathology Center (JPC), the Transportation Security Administration (TSA) dog program, and the biomedical informatician and dog imaging specialist Dr. David Gutman at Emory University. The JPC collaboration provides us with cancer phenotype data. The greatest challenge has been in making progress while the institutional agreements were under negotiation (now completed, awaiting signatures). That delay was solved by 1) developing the molecular methods and analysis through use of another cohort of dogs (Greyhounds from Co-PI Dr. Couto), and 2) establishing the standardized clinical data acquisition and analysis through use of a third cohort of dogs from the TSA, which provides us with canine health records from the same system at the Daniel E. Holland Military Working Dog Hospital. We have thus made excellent progress on our aims. Our continued work will lead to novel findings regarding the initiation, promotion, and progression of cancer.

15. SUBJECT TERMS none provided

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU		19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

Introduction

The purpose of this proposal is to provide insight into gene environment interactions. It leverages the simplified genetics and detailed records of the military working dog population. There are several critical aspects to meeting the aims of this proposal. 1) development of data driven selection criteria, 2) biological sampling of representative dogs, and 3) generation of mathematical methodologies capable of handling heterogenous data and statistical tests in consistent manner and providing clear and understandable results that are biologically valid. Each of these criteria posses their own challenges that must be overcome before the project will be successful. Here we will provide a breakdown of the previous year's work that has occurred and document our progress towards achieving the specific aims we proposed. Specifically we focus on the description of progress related to the work in Dr. Kun Huang's group. While Dr. Huang (Partnering PI) is involved in all tasks and aims listed in the proposal, here we focus on Tasks 1, 2, 6 and 8. The description of all tasks and related achievements and outcome are described in the separate annual report submitted by Dr. Carlos Alveraz (leading PI).

Body

Task 1- Regulatory Approval: The regulatory approval has been successfully negotiated and is currently awaiting signatures from DoD. We have received IACUC animal protocol Exempt status for our data protocol and our biological sampling protocol has been revised as requested by the Lackland AFB IACUC Committee and is currently under review. We also have received approval to acquire the relevant cancer pathology data for the military dogs from the Joint Pathology Center, Silver Spring, MD. To that end, JPC has sent to our DoD collaborators at Lackland AFB copies of all pathology reports associated with the DoD puppy program.

Task 2- Data Capture of Veterinary Records: We have received data from the Transportation Security Administration (TSA). This data was used to establish the preliminary database. In addition we have identified software that will be highly efficacious to capture data from the pathology records. We have also established the necessary computational infrastructure at Lackland AFB. We have a high-speed scanner in place and a high-performance computer to process all of the data collected. We have also conducted regular conference with Lackland AFB, or technician on site Mrs. Michelle Perez. Our site visits have yielded additional collaborations.

Task 6- Adaptation of existing resources, data storage and hosting: There have been two areas of progress:

1.) Development of a Canine Medical Record System
2.) Design of a Workflow for digitizing paper medical records

Development of a Canine Medical Record System (CMRS)

A prototype of a research CMRS has been completed. Our research CMRS allows the creation, search, and modification canine medical records based on a group of standard medical record forms used in the military like the Form 1829, Immunization Form, Death Certificate Form, and Master Problem List. The software is meant as a research tool, rather than an operational tool, for viewing, searching, and analyzing medical records that are captured as part of our digitization workflow described below. Our goal is to populate the CMRS database with data generated from our automated digitization process described in the next section.

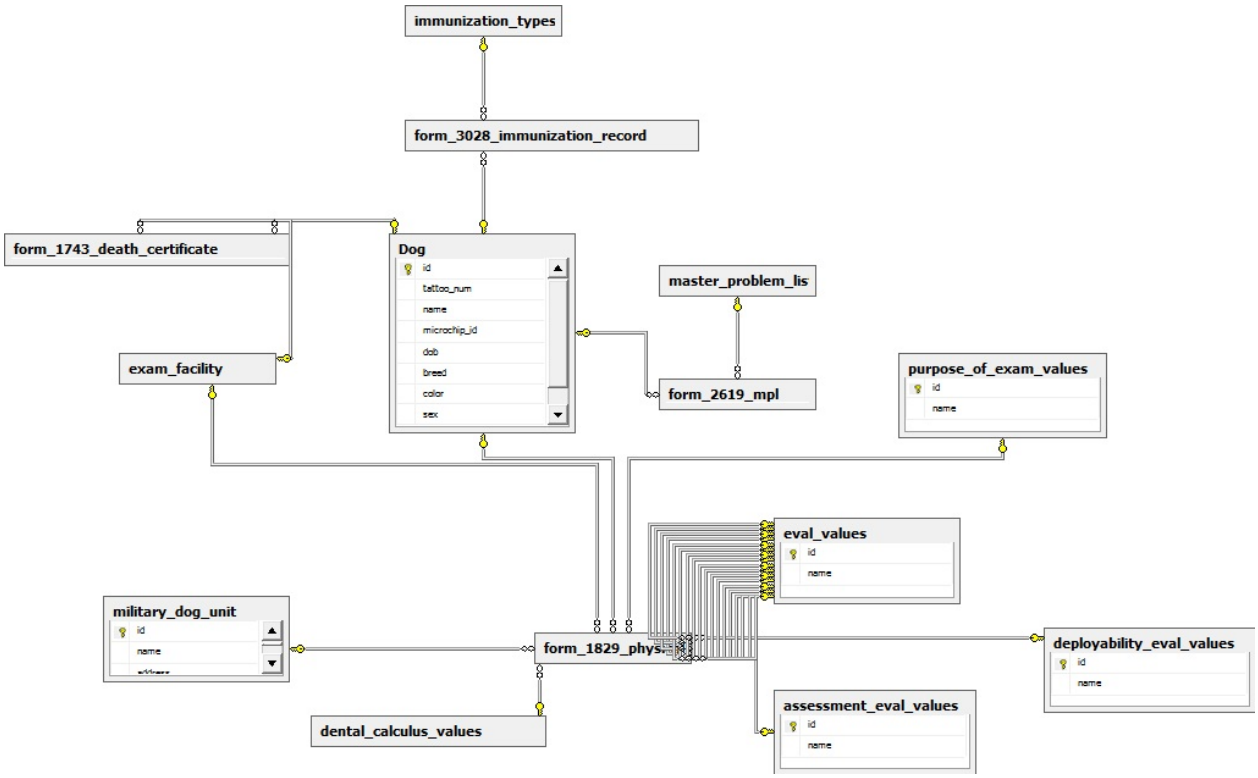

Figure 2 – The database schema for CMRS

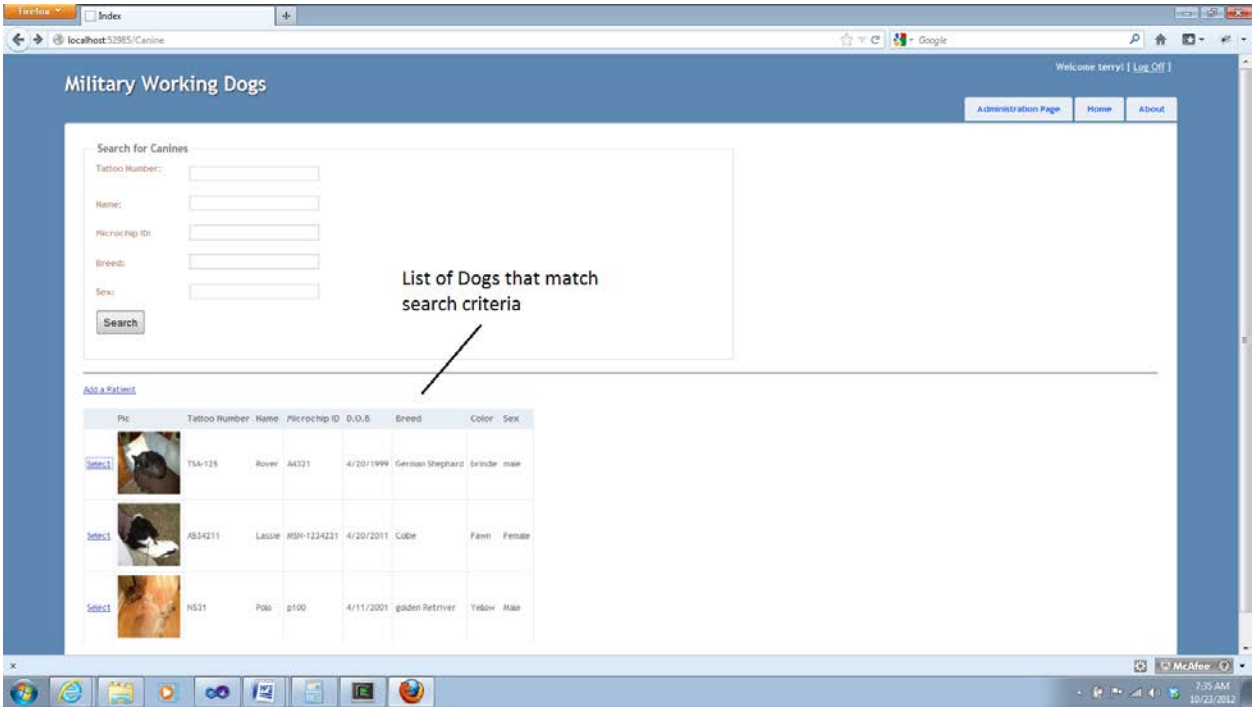

Figure 3 - CMRS Main Search Page (or Start Page)

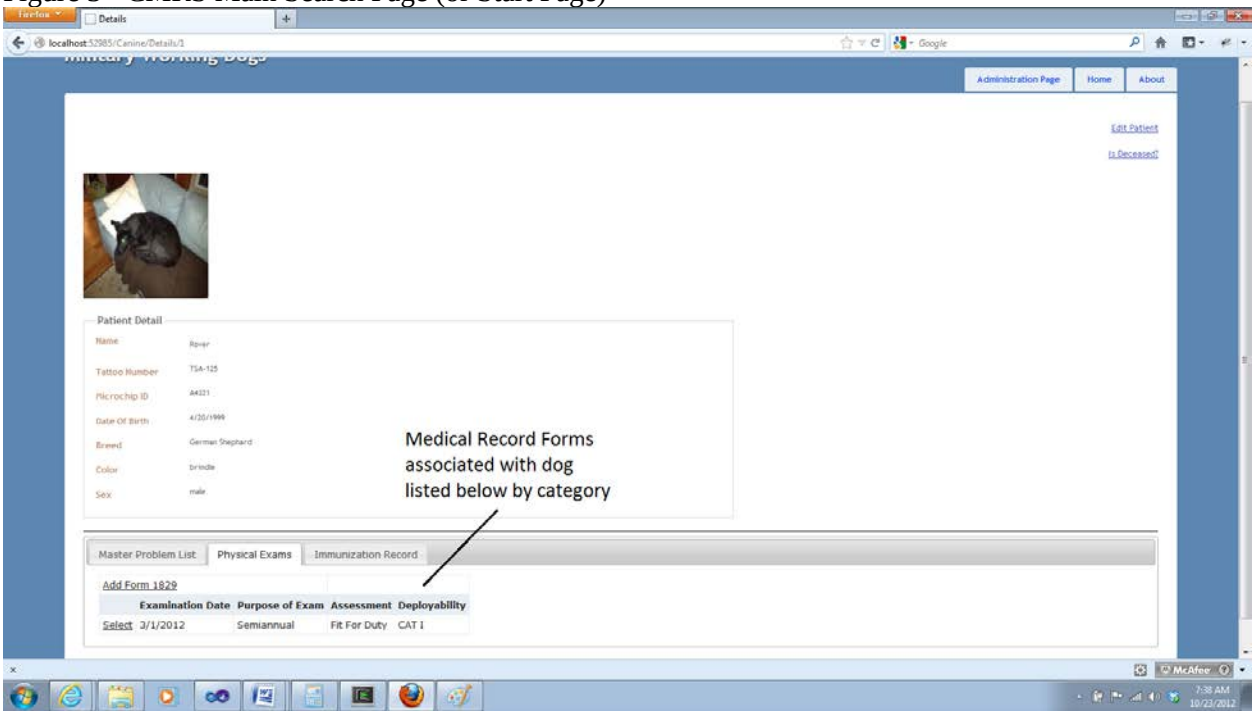

Figure 4 - Canine Detail Page shows general information about a particular canine with a list of its medical record forms (i.e. Immunization, Form 1829, etc.), organized by date and form type, and listed below in a tab strip.

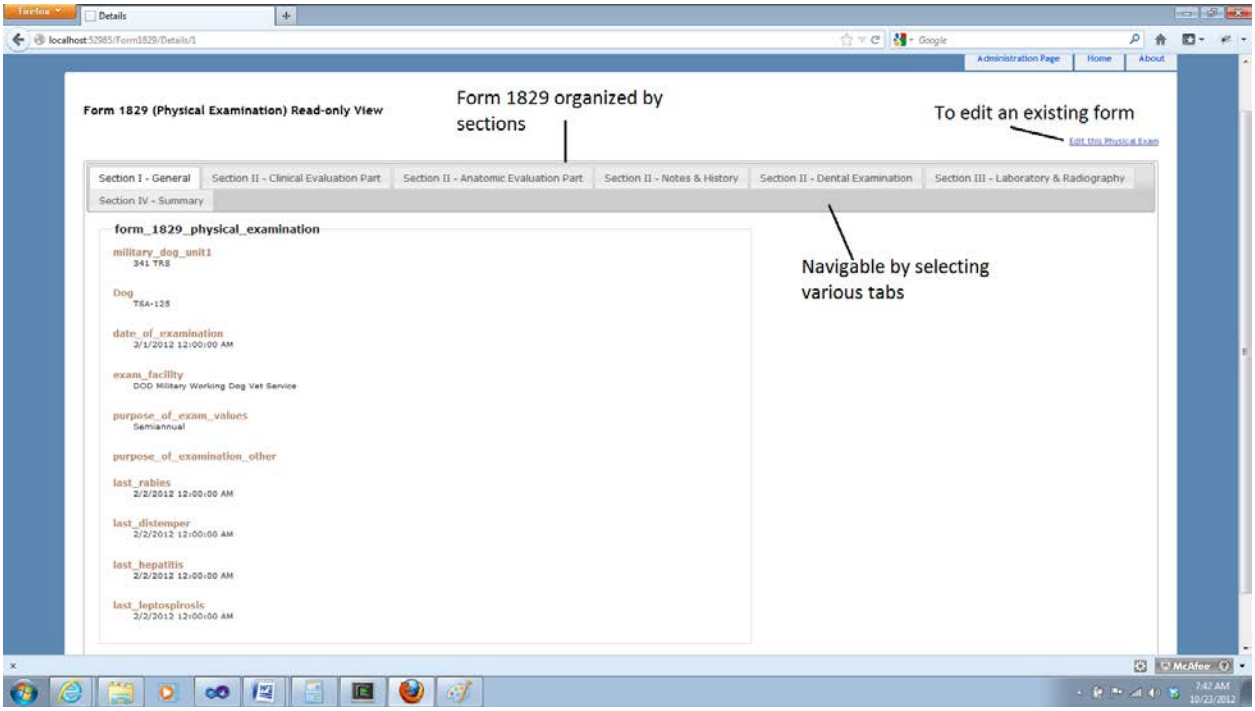

Figure 5 – A Form 1829 retrieved from a database search and displayed (read-only) to the researcher. The Researcher can navigate the form's various sections by selecting on the tabs corresponding to the section from the form.

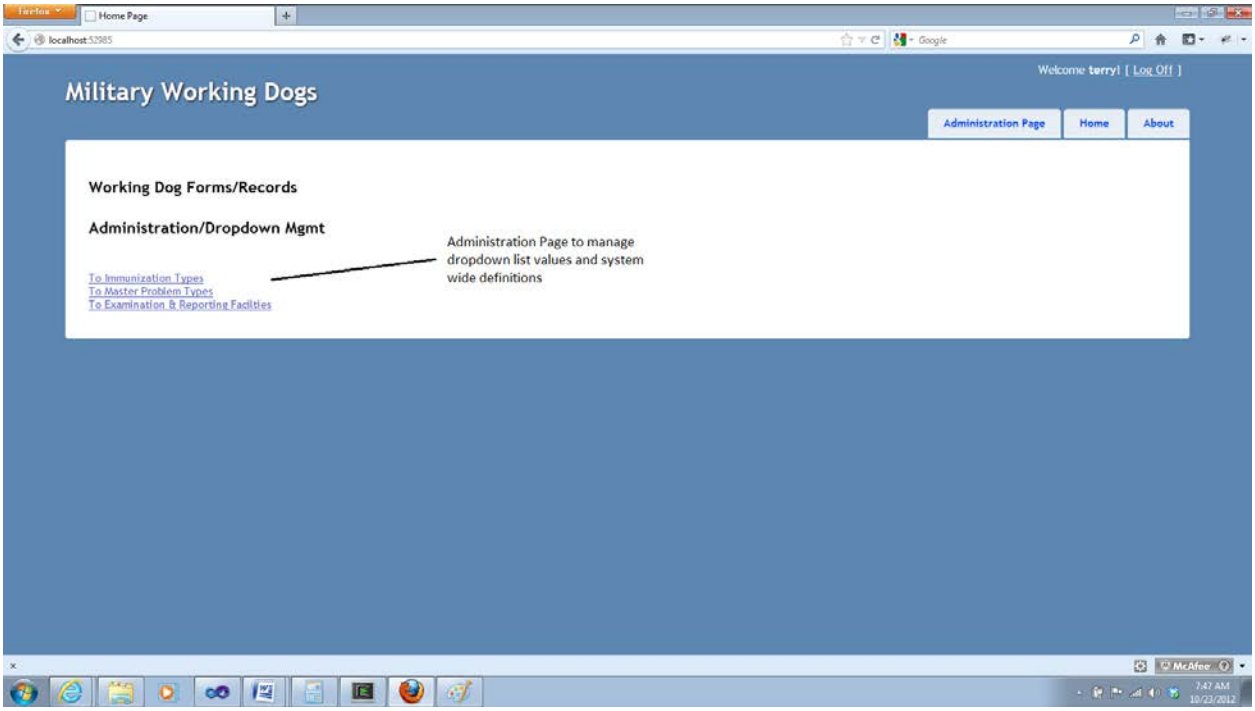

Figure 6 – A listing of the current administration tabs to control various types, definitions, and dropdown lists encountered in the application.

The CMRS has support for ad-hoc query in addition to traditional query support. This is part of the data-warehousing aim of the project. As data related to the cohorts is generated during statistical and bioinformatics analysis (e.g. referred to as derived data), it will be added to our data repository and available for search. Since the structure of derived data is not known in advance, and the best way to search such data is not known either, the search must be more flexible and driven by researcher needs. Ad-hoc queries allow a researcher to navigate a data type and construct a unique query. Generated queries can be saved for future sessions or shared amongst researchers.

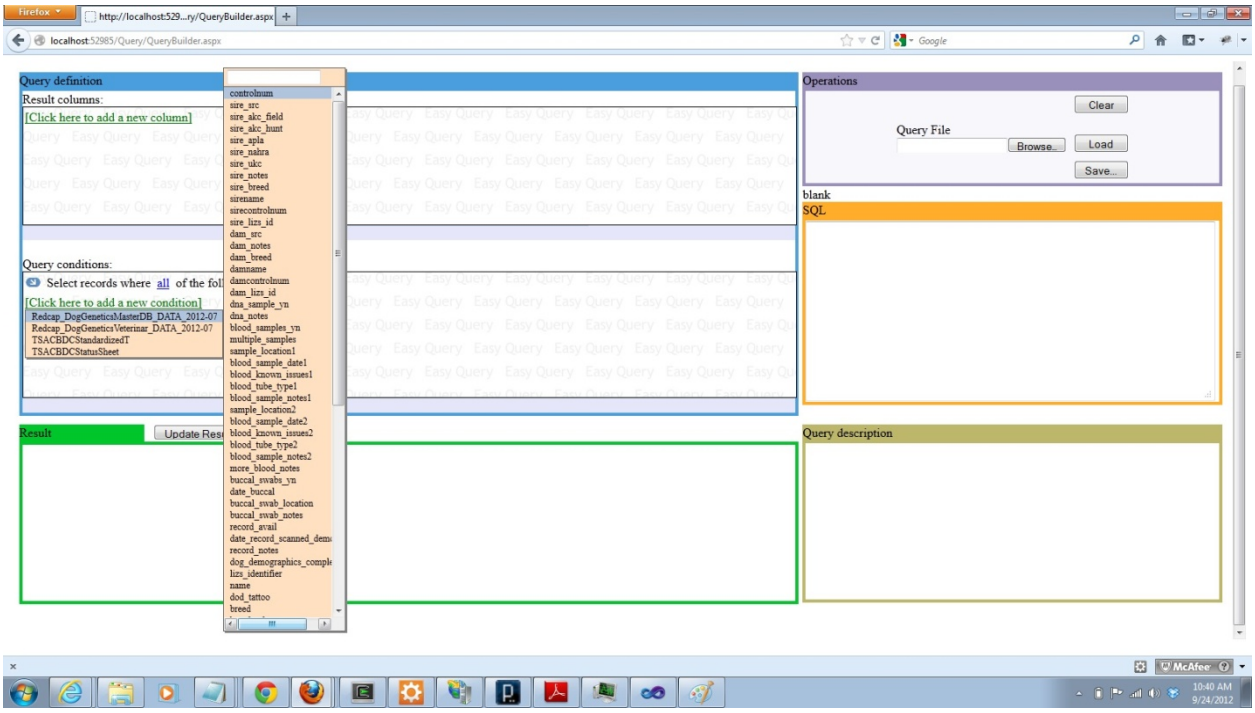

Figure 7 – Ad-hoc query page where a researcher can navigate a data set and construct a query based upon the various data types contained within. The list above shows the various attributes a particular data type has in this case. This example was taken from Liz Hare's Dog Genetics data we used to build the ad-hoc prototype.

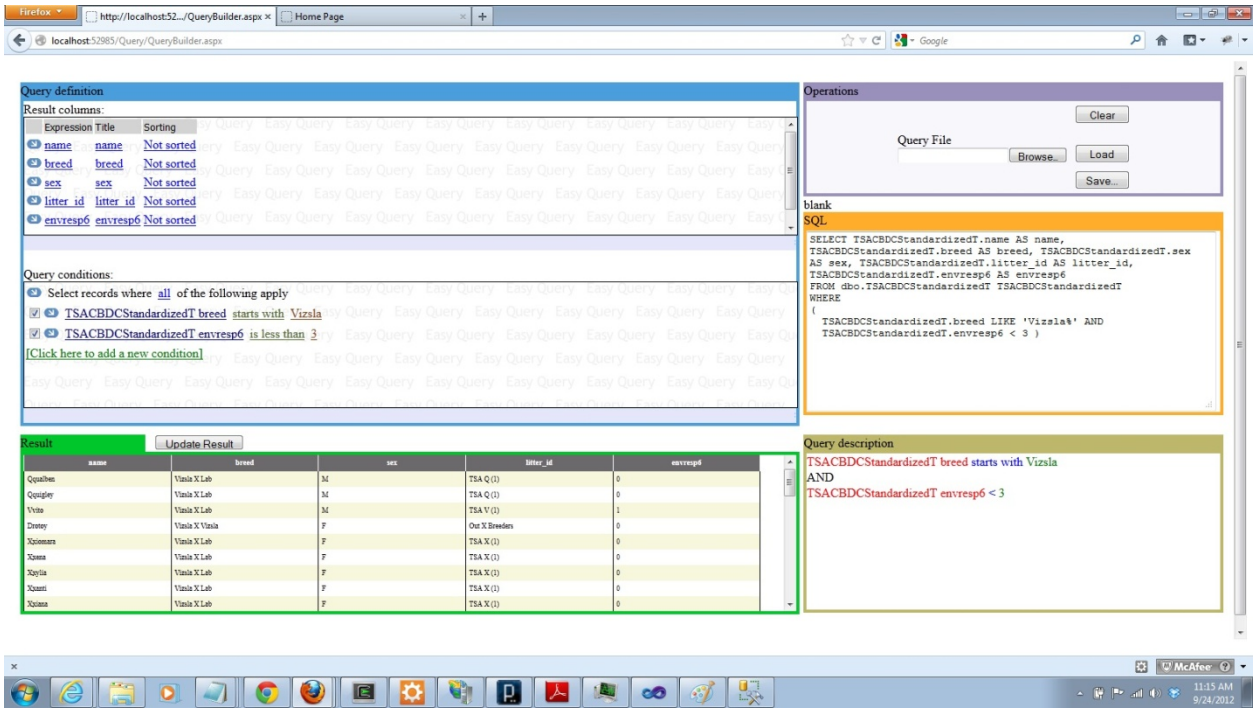

Figure 8 – A completed query that has been executed. The bottom left hand column shows the results while the panel on the right displays details about how the search was constructed. The query panels at the top allow for customization and tweaking of the query details.

Digitization Workflow

A preliminary design of a workflow to digitize (i.e. make database searchable) an archive of canine medical records at Lackland. The workflow is designed to be automated and scalable. A critical part of this workflow is incorporating third-party software for OCR (Optical Character Recognition), ICR (Intelligent Character Recognition) and HWR (Handwriting Recognition). A number of various software packages were evaluated and we have settled on a particular software vendor, ABBY fixed-form. We looked carefully at Form 1829, and the Chronological Medical Record

form to ensure that the third-party software is able to recognize check-boxes and columns and do a reasonable job with handwriting – which is does.

This workflow will also incorporate the Joint Pathology Center Pathology Reports (and Consultation reports). The text will be entered into a data repository that will be further processed by a different pipeline to evaluate the quality of the OCR, ICR, and HWR. The raw data will then be cross-referenced with a controlled vocabulary (e.g. SNOMED, VeNom, ICD- 9) and the matches, or occurrences of controlled terms in the documents, will be stored in an inverted index or specialized database table. The terms can then be correlated against specific canines and stored in a document database. We have not yet evaluated any of the various Controlled Vocabularies.

Task 8- Project management, Quality control and assurance, and Security: The development of a data acquisition pipeline is continuing at the expected pace. This has been achieved by collaborating with the TSA dog program and acquiring records which they have digitized. However, no DoD data has been released because the CRADA has not been executed. As soon as that is achieved, we have the personnel and computational system in place to proceed rapidly.

Key Research Accomplishments

- Establishment of close working relationship with LTC's Cyle Richard and T. Joy Atkin, Chief of Epidemiology and Chief of Pathology respectively at the LTC. Daniel E. Holland Military Working Dog Hospital at Joint Base San Antonio.
- Hiring of a Registered Veterinary Technician, Mrs. Michelle Perez. Mrs. Perez was a former technician with the Veterinary Service and worked as a yard handler prior to her current position with us.
- Establishment of informatics infra-structure at Lackland AFB
- Creation of a highly flexible data-infrastructure robust enough to handle military working dog records and queries of said records.
- Development of collaboration with Dr. David Gutman of Emory University allowing enhance processing of pathological samples and increased informatics support
- Development of collaboration with the TSA breeding program, enabling advanced prototyping of data-structures prior to release of DoD records.
- Development of collaboration with JPC, adding cancer pathology data for improved phenotyping
- Cataloging of the DoD breeding program "puppy program" to guide the identification of the most informative dogs for cancer studies

Reportable Outcomes

- DAPER development
- Application for BAA 12-1

Conclusion

Thus far the project has made excellent progress given the obstacles such as lack of a CRADA. We have identified alternative data sources and have made excellent progress on completion of the database. We are currently working with Lackland AFB to expedite signing of the finalized CRADA. We anticipate that it should be executed shortly. We have also made excellent progress on publications and methodology development. We have developed a highly flexible infrastructure capable of handling the diverse data-types from the working dog program. Continued funding will lead to publications and completions of the aims proposed.