ESIP Lab

SUPPORTING DEVELOPMENT OF EARTH SCIENCES CYBER-INFRASTRUCTURE

Project Summary:

Name of project: Enhancing sUAS Data with Semantic Technologies
Project lead and contact details: Andrea Thomer, athomer@umich.edu
Project partners and contact details: Beth Huffer, beth@lingualogica.net; Lindsay Barbieri, lkbar@uvm.edu; Jane Wyngaard, jwyngaar@nd.edu
Proposed start and end date: 1 November 2017 - July 2018 (Summer Meeting)
Budget Requested: $7000
Budget Summary:

- **Travel:** ***$1500*** in support for PIs to attend VOCamp, ***$3000*** in support for PIs and collaborators to participate in 2018 ESIP Winter and Summer meetings and workshops
- **Student stipends:** ***$2500*** for development of Research Process Models for 3 sUAS-based research projects, and development of an information model, ontology and controlled vocabulary for describing sUAS-generated data products

This project relates to the following key priority areas:

☒ Earth Science Cyber-infrastructure

☒ Semantic Technologies

☐ Socioeconomic value of data

☐ Other (explain)

Project Outline:

Project description: The imagery and sensor data collected by small Unmanned Aircraft Systems (sUAS, also called drones) are highly valuable to a broad range of researchers in many domains. However, if these data are to be fully utilised they need to be augmented with machine-readable, semantically-rich metadata, and annotated in ways that make their provenance (the record of the processes that created the data) explicit. In sUAS-based research this is particularly challenging, as a typical project involves multiple agents (e.g. people, sUAS, controllers, computers, software systems, time-series sensors), and complex processes (e.g. imaging processing algorithms, data processing workflows) with often inexplicit relationships. It is also currently unclear which (if any) ontologies are most appropriate for use with sUAS data.

We will draft a preliminary *drone data ontology* for sUAS data capture, analytics and archival by examining data workflows for 3 representative sUAS-based research projects. For each project we will **(a)** inventory sUAS-generated data products to identify and articulate their provenance and key attributes; **(b)** develop an information model that clearly defines the data provenance, key attributes, and their interrelationships; **(c) use this model as a framework for ontology development, and** identify the vocabulary terms needed to express its attributes and their relationships, either from existing ontologies or by introducing terms and definitions using the YAMZ metadictionary tool, soliciting subject matter expertise to publicly vet definitions. The results will inform the writing of a prototype metadata standard recommendation for sUAS based research, which will be published and shared for further global comment and development.

Project objectives: (1) facilitate usability and reusability of sUAS-generated data through the adoption of metadata standards and semantics and the explicit articulation of key sUAS data classes; **(2)** support semantic interoperability of sUAS-generated data; **(3)** create a collaborative environment for community-driven ontology development to ensure usability and interoperability of sUAS-generated data.

Description of key project steps and timeline:

1. Complete case studies and determine minimum metadata requirements from each *(months 1-3; 10-12. Deliverables: 3 research process models; a first draft of a minimum information model for sUAS-based research.)*

During the 2017 ESIP sUAS Data Management Workshop-Hackathon, we identified 3 distinct cases of sUAS-based science that will serve as exemplars in this project:

- sUAS-based biodiversity monitoring. Contributor: Joe Adams, USGS
- sUAS-based biomass and agricultural runoff monitoring. Contributor: Jane Wyngaard.
- sUAS-based atmospheric greenhouse gas monitoring. Contributor: Lindsay Barbieri

The first two cases' data workflows will be diagrammed and analyzed by Thomer and her student using methods adapted from systems analysis (Thomer et al. (in press); Thomer 2017). This work will result in two diagrams: a UML-based "activity diagram", and a "provenance graph" based on the W3C PROV-O ontology; and two inventories: an "artifact inventory" that lists all of the data products created and used in a research process, and a "process inventory" that lists all the steps in a workflow.

A "minimum information model" (MIM; Palmer et al., 2017) will be developed based on the case studies, and will list the data and metadata attributes (organized into classes) that are critical from the perspective of the target research projects; the MIM also articulates the relationships between those attributes (and their classes). This model will be iteratively revised and used in ontology development work, and will inform our recommendations for metadata standards.

2. Identify potentially appropriate ontologies, map ontology terms to MIM, and identify gaps in existing ontologies *(months 1, 3, 9. Deliverables: mapped terms to MIM; list of gaps in ontologies/areas for further development)*

Working from prior experience and familiarity with sUAS workflows, we will begin identifying candidate ontologies through attendance of the VoCamp event in November 2017 [url]. This workshop is an important opportunity to meet and collaborate with experts in the field of ontology development. We will then host a work session as part of the Annual Geosemantics Symposium (pre-ESIP Winter, 2018) at which we will identify and map additional vocabularies to the MIM. We will note terms and concepts for which ontologies do not presently exist, thereby identifying gaps that must be filled by a drone data ontology.

3. Propose new terms for drone data ontology on YAMZ *(months 1-9. Deliverables: terms for a drone data ontology)*

We will propose terms that are currently absent from ontologies, but necessary for the drone data ontology using YAMZ.net, a tool designed by the California Digital Library for crowd-

sourcing the creation and refinement of metadata terms and definitions. We will use YAMZ to present candidate terms to members of the sUAS science community, where they can suggest term definitions, request clarifications, propose new terms, and make other comments. Terms and definitions that become stable in YAMZ, will provide the raw material for semantic analysis and incorporation into the relevant ontologies. Funding is requested to support student work guiding use of YAMZ (e.g. engaging stakeholders through emails to answer questions and encourage them to vote for terms, entering candidate terms onto the site), and in integrating up-voted terms into a canonical ontology.

4. Test implementation with Barbieri workflow; write report of drone data ontology *(months 6-12. Deliverables: white paper, a case study demonstrating use of ontology)*
Once a draft drone data ontology has been completed, we will apply it to Barbieri's data collection and processing workflow occurring April-June 2017. We will simultaneously analyse Barbieri's data workflow following the methods in step one; this will allow us to document Barbieri's ontology application workflow as well as her data collection workflow while verifying the usefulness of the drone data ontology. We will write a report summarizing this work and presenting the ontology and case studies. Funding is requested to support student work assisting in the application of the ontology and writing of the report.

Month/task	1	2	3	4	5	6	7	8	9	10	11	12
1. Case analysis & information modeling												
2. Ontology identification												
3. Ontology development with YAMZ												
4. Test implementation												
4. Writing & synthesis												

A more detailed Gantt chart is available here

Benefits and advancements of key Earth sciences priority areas:
This project will benefit the sUAS research community by providing a clear path toward the use and adoption of semantics to help ensure sUAS data are (re)useable, shareable and interoperable; further it benefits researchers in semantic technology development by contributed 3 concrete use cases and building critical bridges with a community of "domain" researchers (e.g. the sUAS research community). This work additionally benefits earth science cyberinfrastructure developers by providing them with the terms needed to robustly describe sUAS data in a repository environment, thereby making it easier to retrieve and manage.

Project Partners (as applicable):
Description of project partners and their involvement:

- Andrea Thomer is an assistant professor in the University of Michigan School of Information; she conducts research in digital curation and earth science informatics.
- Beth Huffer is an Information Systems Engineer and the owner of Lingua Logica LLC, a WOSB that specializes in research and development of semantic technologies.

- Lindsay Barbieri, PhD student at the University of Vermont. She will provide a sUAS atmospheric science workflow, incorporating project outcomes in 2018 data collection.
- Jane Wyngaard is a Data Science Technologist at the University of Notre Dame's Centre for Research Computing. Her research focuses on adapting technologies for scientific research purposes, including customising the software and hardware for sUAS data capture platforms.

Involvement of ESIP collaboration areas:
ESIP Drone Cluster (members and contacts), Semantic Technology Committee

Additional Information:
What groups/audiences will be engaged in the project?

- VOCamp community
- sUAS user communities through the 3 Case Studies: USGS, Notre Dame, UVM
- Students hired through the project (supervisor: Thomer, University of Michigan; Huffer)
- Managers and users of relevant external ontologies e.g. SWEET Ontology, Environment Ontology (ENVO), Chemical Entities of Biological Interest Ontology (CheBi)

How will you judge that project has had impact?
An immediate metric of impact will be the quantity of responses and interest we see on YAMZ in response to our targeted invitation to the relevant community. We will also test our drone data ontology with Barbieri's workflow to assess the feasibility and suitability of its use in a real world context. In the longer term, we will judge the success of this project on the uptake of the drone data ontology by other members of the sUAS research community.

How will you share the knowledge generated by the project?

- Engage with subject matter experts (sUAS users, relevant science disciplines, semantic technology community) to participate in the development of term definitions using the YAMZ metadata dictionary. They will receive notifications via YAMZ as terms become stable and, when applicable, are aligned with appropriate external ontologies.
- Present this work at the ESIP 2018 Summer Meeting, coordinating with the ESIP Clusters: Drone, Agriculture and Climate, and the Semantic Technology Committee.
- Present this work at the following correlating Research Data Alliance (RDA) sUAS data IG meeting, thus engaging the international sUAS and research data communities.
- Disseminate results through journal and conference publications aimed at both the earth science and information science communities.