1.0 Project Summary

Name of project: Enabling the encoding and visualization of provenance metadata for better discovery and understanding of climate resilience strategies for agriculture-related decision-making

Project lead and contact details:

Brian Wee, Massive Connections LLC, bwee@massiveconnections.com

Project partners and contact details:

William Teng, ADNET Systems, Inc. at NASA GES DISC, william.l.teng@nasa.gov

Nancy Hoebelheinrich, Knowledge Motifs LLC, nhoebel@kmotifs.com

Tom Narock, Assistant Professor at Notre Dame of Maryland University, tnarock@ndm.edu

Proposed start and end date: July 1, 2018 to January 31, 2019

Budget Requested: $7500

Budget Summary: (please refer to Section 2.4 for details on deliverables)

- Develop Vision and Statement of Need (Deliverable #1, #2 | Lead: Brian Wee) $2000
- Gap analysis to determine adequacy of PROV to meet requirements stipulated in Vision and Statement of Need, and determine scope of experimental implementation (Deliverable #3a | Lead: Nancy Hoebelheinrich) $2500
- Gap analysis of machine learning technologies that can facilitate the use of provenance for discovery (Deliverable #3b | Lead: William Teng) $1000
- Extension / development of PROV schema to support documentation of data-to-decisions conceptual workflow; implementation of PROV schema leveraging Provisium and other technologies (Deliverable #4 | Lead: Tom Narock) $2000

2.0 Project Outline

2.1 Abstract

There is an evident need for climate resilience decision makers and planners to more effectively search for and understand existing climate resilience strategies. The proposed project will define a long-term vision and a mid-term statement-of-needs (SON) for achieving an efficient transformation of data into decisions with input from key ESIP stakeholders. Using the vision and SON, the project team shall identify specific goals and objectives for encoding the provenance of at least one agriculture and climate case study that demonstrates how data informed decisions in that situation, and how that provenance can be visualized. The goals and objectives will be tested by initiating an experimental implementation using a community developed provenance schema (e.g., one of the W3C's PROV dialects) and open source provenance expression and visualization tools. We anticipate leveraging the recently funded and completed ESIP Lab Project "Provisium". The case study that the project team envisions using is one that

was identified by the ESIP Agriculture and Climate Cluster and subsequently successfully incorporated into the U.S. Climate Resilience Toolkit (CRT).

2.2 Background

Challenge. Climate resilience challenges are inherently multi-disciplinary problems that require the integration of work products from a broad spectrum of users. The provenance of the transformation of data to decisions along this value chain ultimately informs the value of socio-economic data and information. Capturing that provenance at a conceptual level that is useful for both humans and machine reasoners is a challenge that we do not believe has been adequately addressed.

Vision. Successful capture of the data to decisions provenance would enable capabilities that could leverage existing and /or developing frameworks for different sectors. For example, a municipal manager for a small coastal city could search existing climate resilience projects for case studies that bear similarities to the challenges that the city is experiencing--and retrieve a recommended optimal roadmap as a starting point for decision-making. The query mechanism would ideally be semi-structured, allowing for limited natural language search terms to be entered into a search template that allows a machine reasoner to perform an ontology-powered search against existing climate resilience case studies whose data-to-decisions provenance have been captured in a machine-parsable manner. The result of the query would be a visualization of case studies that have each been structured using that provenance model, and can therefore each be visually represented using a common framework. A candidate framework is the existing US Climate Resilience Toolkit (CRT) "Steps to Resilience" framework. The municipal manager would be able to visually compare the degree of similarity between existing case studies and the local resilience challenge, thereby providing a basis for prioritizing how such case studies might be adapted for a local challenge.

Benefits. Decision makers and planners should have better ways to search for resilience strategies and adopt the entire data-to-decisions workflow for other contexts. Having a canonical representation of the entire data-to-decisions workflow may also help in discovery and usability of data and information. The above vision, if ultimately implemented as a service layer that resides as an application on a server that is separate from the one maintained by NOAA for the CRT, is consistent with public-private partnerships, where value-added service-providers build additional capabilities - like the one described in "Vision" above - that provide additional capabilities to the community.

2.3 Objectives

1. To elicit a stakeholder-driven vision of fully traceable data-to-decisions workflows for climate resilience.
2. To determine the current inadequacies in provenance representation for fully traceable data-to-decisions workflows.
3. To assess the experimental limited-functionality web application, Provisium, to demonstrate the utility of traceable data-to-decisions workflows.

2.4 Key project tasks and timeline

The deliverables from the proposed project are:

1. **Deliverable #1: Vision 2025.** The vision briefly described in Section 2.2's "Vision" will be used to develop a narrative to answer the question "In 7 years' time in the year 2025, what is the ideal technology-enabled climate adaptation planning workflow for a regional-scale climate challenge?" That narrative is termed "Vision 2025" for the purposes of this proposal. We propose an initial focus on agriculture and climate adaptation.

2. **Deliverable #2: Statement of Needs (SON) 2022.** Vision 2025 shall be used to derive a statement-of-needs (SON) for a version of the CRT that is envisioned to be built in four years' time in the year 2022. That version of the CRT is called "CRT-Plus." The SON describing CRT-Plus shall be referred to as SON 2022.

3. **Deliverable #3a/b: Gap analysis. #3a:** Given the 2025 Vision and SON 2022, the team will address the following question: Are the current PROV dialects sufficient to support those future needs? Existing PROV dialects will be evaluated against their ability to represent selected case studies in the US Climate Resilience Toolkit (CRT) that focus on adaptation of agriculture practices in response to the impacts of climate change. Ideally, existing PROV dialects will sufficiently capture the five step protocol stipulated by the CRT framework. **#3b:** We will develop a conceptual diagram of a machine learning infrastructure, including any gaps in technology, that will support the query mechanism of CRT-Plus.

4. **Deliverable #4: Experimental implementation.** Given the information from the gap analysis, the team shall extend or define, as appropriate, the ontology for a PROV dialect consistent with Vision 2025 and SON 2022. We propose to leverage the intellectual contributions of the Provisium team (T. Narock and D. Fills) which has explored ways to visualize provenance metadata encoded using selected PROV dialects.

Tasks associated with the above deliverables are:

1. **Task #1, Vision 2025 stakeholder meetings, Month #1.** Two web-based meetings, each no more than two hours in duration, will be held two weeks apart. At the end of these meetings, we aim to derive the Vision 2025 narrative (Deliverable #1) gleaned from the input of the stakeholders.
2. **Task #2, SON 2022 stakeholder meetings, Month #2, in sequence after Task #1.** Two web-based meetings, each no more than two hours in duration, will be held two weeks apart to derive the SON 2022 (Deliverable #2). These meetings will be based on Vision 2025. However, unlike the Vision 2025 meetings, these meetings will be more technology focused.
3. **Task #3, Gap analyses by project team, Months #3 and #4, in sequence after Task #2.** The project team leads shall assimilate Vision 2025 and SON 2022 to refine Task #4. We strongly suspect that a PROV variant needs to be drafted that captures the provenance of climate adaptation at the level of abstraction stipulated by the NOAA resilience planning protocol. We shall also undertake a gap analysis of machine learning technologies with regards to utilizing provenance for discovery in a manner consistent with Section 2.2's "Vision".
4. **Task #4, Experimental implementation by project team, Months #4, #5, #6, in partial overlap with Task #3.** We intend to use an ESIP agriculture and climate cluster case study that was successfully incorporated into the CRT as the basis for testing the PROV variant described in Task #3. We may extend this activity to one or two additional CRT case studies as resources

allow. Provisium has already implemented a prototype visualization tool using PROV-O-VIZ, and we intend to leverage those existing capabilities.

2.5 Impact

Deliverables #1 and #2 follow from Section 2.2's "Motivation". These deliverables provide an essential "north-star" to the community on how public-private partnerships may enhance the value and usability of the CRT. Deliverables #1 and #2, which form the bases for Deliverables #3 and #4, are stakeholder-driven products that inform future technology development for traceable climate decision-support tools.

Deliverable #3 will inform the provenance ontology community of the higher-level workflow concepts that need to be captured so that higher-level concepts associated with climate resilience planning can be captured and used in machine-enabled discovery processes as described in Section 2.2's "Vision". Enabling this capability will further structure resilience solutions in a common framework for comparisons using a visualization tool. We anticipate that Deliverable #4 will result in a product that can be eventually be used by another project to iterate towards a more complete extension / development of a PROV dialect consistent with Vision 2025 and SON 2022.

3.0 Outreach

For Tasks #1 and #2, we anticipate inviting participants from NOAA, USDA Climate Hubs, USGS, and NASA to incorporate the respective agencies' perspectives, given that these agencies fund programs that are either directly or indirectly relevant to climate resilience planning. For Tasks #3 and #4, the bulk of the work will be undertaken by the Project team leads.

We anticipate providing webinars at key points during the project to update participants and solicit feedback on project deliverables with invitations to the ESIP Semantic Technologies Committee and the RDA Interest Group on Agricultural Data (IGAD). In addition, we will present our progress and findings at ESIP meetings and other venues as appropriate.

4.0 Project Partners

- Brian Wee (Founder and Managing Director, Massive Connections LLC), Principal Investigator
- William Teng (Principal Scientist, NASA GES DISC (ADNET)), Project team
- Nancy Hoebelheinrich (Principal / Information Analyst, Knowledge Motifs LLC), Project team
- Tom Narock (Professor, University of Notre Dame), Technical advisor
- Sky Bristol (Supervisory Biologist, Biogeographic Characterization, USGS) Technical advisor on USGS Provenance use
- Edward Gardiner, NOAA, Technical advisor on US CRT framework and supporting infrastructure