VDI Center for Visual & Decision Informatics

Transforming Data Adaptation Science and Service: An Innovative Visual Ontology Application

PROJECT ID | CS15-6 **TYPE |** [X] New [ ] Continuing **START DATE |** July 2015

PROJECT LEAD/PARTICIPANTS | Jane Greenberg, Chaomei Chen, Xia Lin, Adrian Ogletree, Erin Clary (CCI, Drexel)

DESCRIPTION | Explorations in data reuse and repurposing are crucial, given the **growth in digital assets accessible** in a wide variety of domains. Arguably, the best-documented data are in scientific publishing, motivated by international data sharing requirements, open data initiatives, and software sustainability programs. A **severe limitation** is that research on data reuse is shallow, primarily based upon simple citation counts and impact measures. Scientific endeavors can be significantly transformed by reporting data reuse via rich ontological associations (equivalence, derivative, sequential, etc.) in order to more meaningfully demonstrate adaptation during the lifecycle. An ontology tracking high-impact cases of data and algorithm adaptation will provide a more accurate view of reuse and enable radical, new adaptation combinations to be learned and pursued. The objectives of the proposed work are to: **1)** Develop an ontology for tracking *data and algorithm adaptation* in multiple domains, such as the **biological and earth sciences, astronomy, health data domain**, and an **engineering sub-domain**, drawing from structured and unstructured data sources; **2)** Develop an algorithm that incorporates the ontology and more accurately tracks data and algorithm adaptation; and **3)** Enable data and algorithm producers, owners, and publishers to **transform data adaptation services** and **support better science**.

EXPERIMENTAL PLAN | A team of experts in ontology design/metadata, data reuse, citation analysis, and visualization will work to: **a)** Semantically map and integrate existing terminologies to form a base ontology of adaptation associations, drawing from DataCite, Dublin Core relation taxonomy, and other initiatives that can provide insight. **b)** Identify high-impact data and algorithms for testing and refining the base level ontology. Target domains: **biological and earth sciences**, **health, and an engineering sub-domain**. Example sources include the Global Wood Density Data, with over 10,000 downloads and 100 known reuse cases since its 2009 publication (Dryad repository); and Francisco Herrera's reporting of the top 10 Data Mining Algorithms. **c)** Develop a visual ontology application to record, track, extract, and potentially predict data and algorithm adaptation drawing from targeted corpus. **d)** Work with IAB members to evaluate the effectiveness and efficiency of the prototype ontology application in other real-world domains seeking to transform data adaptation science and services.

RELATED WORK | Research team members have conducted extensive work in ontology design/metadata, linked data, citation analysis, predictive analysis, and visualization working with diverse, large and long tail, structured/unstructured data sources, from real-world applications in the biological and earth sciences, medical informatics, and engineering/materials science.

HOW OURS IS DIFFERENT | The proposed work presents a novel approach for tracking data reuse via an innovative ontology application targeting adaptation. Adaptation encompasses reuse and repurposing, and allows for radical, new reuse combinations to be pursued. We will **advance** scientific practice and services by: (1) developing a visual ontology application for extracting and tracking rich associations of data and algorithm adaptation; and (2) leveraging the ontological application for better data adaptation science and service. Our ontological application will help CVDI partners in strategically planning for high-impact data and algorithms outputs, including novel uses.

MILESTONES FOR YEAR 1

4 months: Develop rich relationships for capturing data and algorithm reuse scenarios.

6 months: Develop an algorithm for tracking, extracting, and modeling adaptation, drawing from the target sample.

12 months: Test the effectiveness and efficiency of the prototype multi-level ontology in real-world settings.

Milestones	4 months	6 months	12 months
Develop the ontology			
Develop tracking algorithm			
Test prototype ontology, transfer knowledge			

DELIVERABLES

1) Prototype visual ontology application capturing data and algorithm adaptation in the targeted test domains, both large and long-tail data. 2) An algorithm for tracking, extracting, and modeling adaptation science and service. 3) An approach for CVDI partners to determine and strategically plan for greater impact of data, application, and algorithm outputs.

BUDGET FOR YEAR 1

Students (2 doc. students)	$80,000
Equipment	$4,000
Travel	$8,000
Overhead	$9,200
Total	**$92,000**

ECONOMICS | Defining and validating ontological associations that demonstrate data and algorithm adaptation will help CVDI partners to: 1.) provide services that support better science; 2.) conduct better science; 3.) yield a greater ROI.

POTENTIAL MEMBER COMPANY BENEFITS | This project will benefit IABs by providing an approach for CVDI partners to transform data adaptation and strategically plan for greater impact of data and algorithm outputs.

PROGRESS TO DATE | We have tested the value of selected ontological associations among a wide range of biological data types and identified a set of data and algorithms with known high use/repurposing activity.

KNOWLEDGE TRANSFER TARGET DATE | 12 months