Smart & Connected Communities Outline

## Project Summary (1-page limit)

###### The project summary may not exceed one page in length and must consist of the following three clearly labeled sections: Overview, Intellectual Merit, Broader Impacts

### Overview

###### Include the title of the project, the name of the PI and lead institution, and a summary of how the project integrates social and technological research dimensions, research capacity-building, and community engagement

#### Title: Enhancing and Enabling Local Government Capacity to Re-purpose, Combine, Analyze and Present Local Administrative Data

#### PI: Aaron Schroeder, Virginia Tech

#### Summary

Many localities lack both the analytic and technological capacity to carry out the processes necessary to re-purpose their existing administartive datasets, combine those datasets with addtional locality-relevant data sets from other sources, analyze those datasets, or translate those analyses into information relevant to local decision-makers. Accordingly, it is also difficult for many localities to work cooperatively with policy-area researchers who require access to high-quality datasets.

Working with the latest in open-source data preparation and analysis technologies, we will work with local government partners to enhance their capacity for:

* re-purposing local administrative datasets (i.e. processes for access, management, linkage, and transformation)
* combining prepared datasets with other locality-relevant datasets (e.g. State, Federal, other)
* conducting analyses of these datasets both on their own and in partnership with policy-area researchers
* preparing the results of analysis for presentation and use by decision-makers (e.g. dashboards)

These capacity-building efforts will be focused around a specific policy-area of the locality's choosing (e.g. aging infrastructure, environmental quality, public health) and be supported by a virtual analytic environment composed of a set of technological platforms specifically configured to support each stage of data preparation, analysis and use. Within this environment, government analysts and policy-area researchers can work cooperatively on community-relevant issues using all available locality-relevant datasets. Additionally, as locality capacity increases, if a locality wishes to begin hosting the technologies themselves, all platforms within the virtual environment will be easily transferable.

### Intellectual Merit

###### Provide a brief summary of the intellectual merit of the proposed project detailing the potential fundamental and integrative research advances; and

### Broader Impacts

###### Provide a brief summary of the broader impacts of the proposed project, including potential impacts on the community defined and the field.

## Project Description (4-page limit)

###### In preparing the Project Description, consider the Program Description for IRGs and the Additional Solicitation Specific Review Criteria because the preliminary proposal is used to assess suitability for submitting a full proposal. The Project Description of the preliminary proposal is limited to four pages and must consist of the following six clearly labeled sections:

### Vision & Goals

###### Describe the vision and goals of the proposed research. Briefly describe how the project will contribute to our understanding of S&CC research, to research capacity-building, and to the engaged community.

There is tremendous potential for enhancing the effectiveness of local government decision-making by bringing to bear analyses of all locality-relevant data sources, regardless their location. Local government decision-making based on these data and the associated research are instrumental to improving performance, cost, and function. Local governments themselves often have significant stores of historic and current electronic data at their disposal. However, the insights and patterns that analysis of that data might reveal are often obscured from use by siloed information technology systems that have been developed solely for the purpose of processing isolated transactions. [Innovations in Public Service Delivery Issue No. 4Predictive Analytics:Driving Improvements Using Data Stephen Goldsmith, Susan Crawford, and Benjamin Weinryb Grohsgal].

Unfortunately, many localities lack both the analytic and technological capacity to carry out just the first steps of re-purposing their administrative datasets (i.e. profiling, cleaning, transforming) to make them ready for analysis in combination with other datasets (e.g. State, Federal, other). In turn, this inability to provide datasets suitable for analysis can make it exceedingly difficult to work cooperatively with policy-area researchers who can bring invaluable insights to the decision-making process.

As a natural part of the outreach mission of our institutions, we will work with local governments to enhance their capability for data-driven decision making by providing a virtual analytic venue where government analysts and academic researchers can work cooperatively on community-relevant issues using all available locality-relevant datasets, including locally-derived data sources (e.g administrative data, sensor data), sources derived by neighboring communities, state and federal data sources, and data provided by non-governmental entities (e.g. community-oriented non-profit organizations).

To support this goal:

* Develop, provide, and implement a sustainable comprehensive, loosely federated, end-to-end set of data science processes, including processes of data ingestion and management, data analytics, and analysis presentation that will support local government evidence-based decision making and researchers engaged in community based research. These Data Processes and Platforms will be developed in a manner to make them easily replicated and curated beyond their development stage to create a statewide and ultimately national ecosystem.
* Develop and implement the set of data science processes with a set of actively managed technology platforms providing the latest in open-source database, GIS, data analytic, and data presentaion technologies
* Establish a community-engagement model that keep barriers to participation as low as possible. For example, there will be no expectation of any significant modification to existing local government data systems, such as data standards, as a prerequisite of particiption. Instead, the system host (VT/ISU) takes on the responsibility for:
  + maintaining a comprehesive database of metadata of all data sources being provided by participating localities, including mappings between data sets using different data standards
  + providing to the locality, with support, the requisite technologies needed to securly connect their existing data resources to the venue.
* Create and implement a sustianibilty plan for these Data Processes and Platforms.

### Integrative Research Approach

###### Describe clearly the proposed research including its plan for integrative research and community engagement. Define the community, associated stakeholders, and how the community will engage in meaningful ways with the proposed research. Briefly describe the evaluation approach including initial metrics.

1. Deploy Scalable, Transferable, and Cost-Effective Data Management, Analytics, and Presentation Platforms (see figure 1)

Built using the latest open-source data technologies, individual technology platforms can be hosted locally or in the cloud. A set of Virtual Private Servers (VPSs) is dedicated to each participating locality. As the data management and analysis expertise of a community grows, it may become desirable for that community to have a more direct controlling role over the technology platforms. In fact, such developments are a sure sign of success in capacity-building efforts. The technology plaforms are purposely designed and constructed for ease of transferability for just such a circumstance. The computer code that underlies the platforms is open-source and thus will be made available to any community under GNU Public License along with documentation.

\* Data Management Platform  
\* Data Analytics Platform  
\* Data Lexicon  
\* Data Presentation Platform

1. Begin by working with local government with whom we have existing working relationships
2. Commence a Community Engagement Process including:
   * Issue Identification & Preliminary Hypothesis Generation: develop, via situational analysis, an understanding of critical community issues as well as a set of variables suspected to be causitively related to each issue
   * Initiation of three distict discovery processes:
     + Data Management System Status Discovery
     + Data Analytics Capabilities Assessment
     + Data Discovery & Inventory
   * Draft and agree upon necessary data agreements (e.g. MOAs, MOUs)
   * Draft and agree upon a Data Access Plan
3. Deploy necessary data connection technologies as required by the Data Access Plan
4. Profile data and fill the metadata repository (Lexicon)

At its base, the CLD3 Lexicon serves the function of a metadata repository—a database created to store metadata from various systems. Metadata is information about the structures that contain the actual data. Metadata is often said to be "data about data", but the Lexicon goes far beyond this definition, proving a centralized node of data source information that can be used for provenance tracking and data linkage within a heterogeneous network of data sources (A. D. Schroeder 2013b). Specifically, the Lexicon is an inventory of and history of changes to:

* every available data field in every available data source
* the structure of their storage
* possible values and meanings of the information stored
* possible transformations of each set of field values from one data source to another set of field values from another data source
* methods of data source access
* matching algorithms and how they are to be used in conjunction with possible field value transformations The Lexicon provides fundamental functions for the operation of the CLD3 Framework. It is a requirement that all data partners in the network provide the data information necessary for its operation. The Lexicon is maintained in an RDBMS by CLD3 staff, thus removing the complexity required for high quality data linkage from all data partners (i.e., a standardization scheme enforced among data partners).

1. Establish access by local government analysts and academic researchers to platforms
   * Prove training and support
2. Conduct cooperative analysis
3. Generate reports, dashboards, etc for presentation to local government decision-makers
4. Cooperatively assess the utility of the provided analyses for decision-making purposes (focus groups?)

### Research Capacity-Building

###### Describe activities towards building, developing, and engaging research talent to focus on S&CC integrative research challenges.

### Integration and Multi-Disciplinary Context

###### Characterize the multidisciplinary, holistic nature of the approach by identifying the disciplines involved in the research, and how the proposed research elements are integrated.

Defintiion of Transdisciplinary

Public Administration & Policy Statistics Economics Computer Science Psychology

### Management

###### Describe the participating institution(s) and key personnel briefly (see Project Personnel and Partner Institutions section). Address the roles and responsibilities of the project’s management, and how it will support integration of tasks at a high level and engagement with the community throughout the course of the planned research activities. Describe the approach to collaboration among project members.

### Budget and Subawardees

###### Provide an estimated aggregate budget request with number of years of support requested. Provide a list of subawardee institutions.

## Project Personnel and Partner Institutions (1-page limit)

###### Provide current, accurate information for all personnel and institutions involved in the project. Follow the same format as described for Project Personnel and Partner Institutions in the Full Proposal Preparation Instructions below. Submit as a supplementary document.

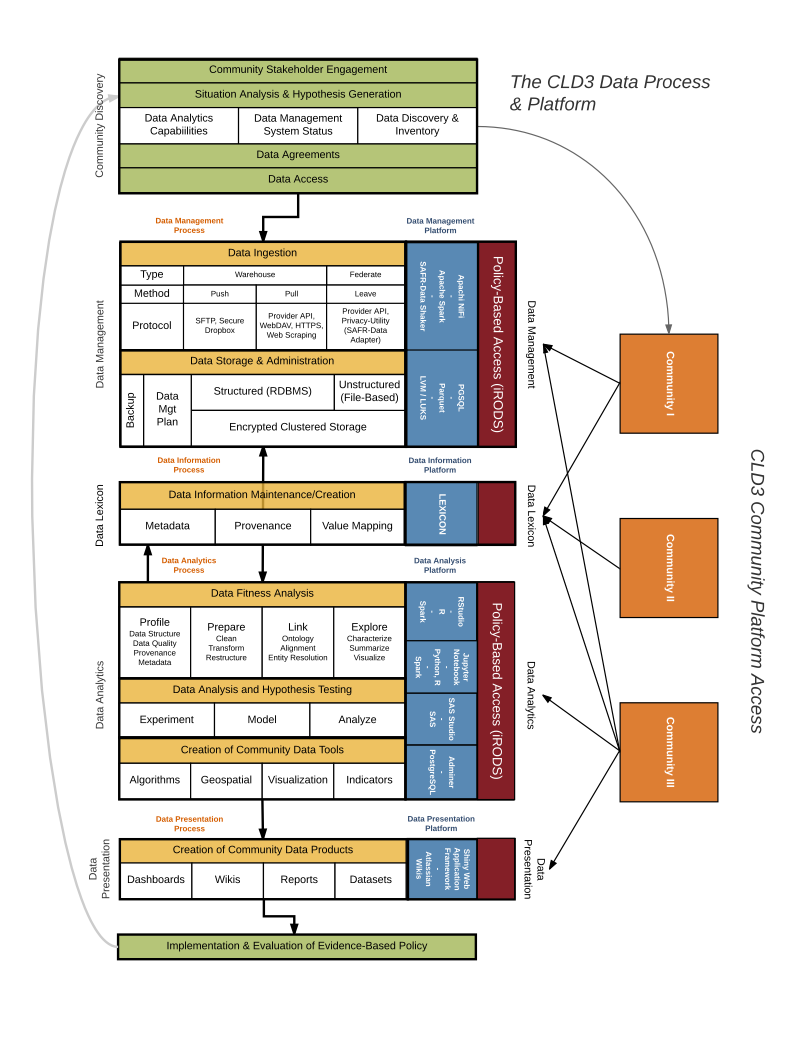


Figure 1: CLD3 Processes and Platforms Framework