SECTION 1. PROJECT ADMINISTRATIVE INFORMATION

Indicate which element(s) of the CDI Science Support Framework the proposal relates to:
Computational Tools and Services, Data and Information Assets, Communities of Practice

Project title: Integration of National Soil and Wetland Datasets: A Toolkit for Reproducible Calculation and Quality Assessment of Imputed Wetland Soil Properties

Name of lead USGS cost center requesting funding: National Research Program/Eastern Branch

Name of USGS Principal Investigator, ORCID ID, mailing address, telephone, fax, and email:
Eric Sundquist ORCID ID 0000-0002-1449-8802
U.S. Geological Survey, 384 Woods Hole Rd., Woods Hole, MA 02543
Phone: 508-457-2397, Fax: 508-457-2310, Email: esundqui@usgs.gov

Names and contact information for additional Principal Investigators or collaborators:
Norman Bliss, ASRC InuTeq (Contractor to the USGS EROS Center), bliss@usgs.gov
Sharon Waltman, USDA/NRCS, Sharon.Waltman@wv.usda.gov
Rusty Griffin, USFWS/NWI, rusty_griffin@fws.gov
Lisamarie Windham-Myers, USGS NRP/WB, lwindham-myers@usgs.gov

Short description of the project (two or three sentences):
We propose to develop and document a toolkit to impute wetland soil properties by integrating Soil Survey Geographic (SSURGO) data with the National Wetlands Inventory (NWI) and other data sources relevant to the extent and properties of wetlands. Because the spatial distribution of wetland-associated soils is subject to inherent uncertainties in the SSURGO database, complex methods are required to extract the relevant soil information and integrate it with other data sources to determine the best available estimates of wetland soil properties. We will extend our experience with Python data extraction and GIS data integration to produce a set of code-generating code, Python Notebooks, and integrated datasets that will enable other users to reproduce and adapt our methods for a broad range of applications.

List of anticipated deliverables from the project:
Tools for data extraction, integration, and imputation that will be documented and made available in the form of Python Notebooks and metadata.
Integrated GIS datasets of wetland soil properties that will be available to the public.
Presentations at professional meetings and journal publications highlighting imputed wetland soil properties relevant to water resources, carbon storage, and habitat.

SECTION 2. ESTIMATED BUDGET

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
PERSONNEL (SALARIES including benefits):		
Federal Personnel Total (support for Eric Sundquist, 1 month in kind, existing Federal salary):	$0	$18,000 (in kind)
Contract/Collaborator Personnel Total (support for Norman Bliss, 2 months under existing EROS contract):	$40,000	$0
Total Salaries:	$40,000	$0
TRAVEL EXPENSES:		
Travel Total (Per Diem, Airfare, Mileage/Shuttle) **x** # of Trips:	$2000	$2000
Other Expenses (e.g. Registration Fees):	$500	$500
Total Travel Expenses:	$2500	$2500
OTHER DIRECT COSTS: (itemize)		
Equipment (including software, hardware, purchases/rentals):	$0	$0
Publication Costs:	$0	$0
Office Supplies, Training, Other Expenses (specify):	$0	$0
Total Other Direct Costs:	$0	$0
Total Direct Costs:	$42,500	$0
Indirect Costs (%):	$7500	$0
GRAND TOTAL:	$50,000	$20,500

SECTION 3. PROJECT SUMMARY

Wetland soils are vital to the Nation because of their role in sustaining water resources, supporting critical ecosystems, and sequestering significant concentrations of biologically-produced carbon. The U.S. has the world's most detailed continent-scale digital datasets for soils and wetlands, yet scientists and land managers have long struggled with the challenge of integrating these datasets in ways that can be applied to research and to resource assessment and management. We propose to develop and document a set of methods to impute wetland soil properties by integrating Soil Survey Geographic (SSURGO) data with the National Wetlands Inventory (NWI) and other data sources relevant to the extent and properties of wetlands. Our methods and documentation will build on our current research and development of best practices for analysis and application of soil and wetland data.

Our team includes co-investigators who have significant responsibilities for maintaining the SSURGO and NWI datasets. SSURGO is the detailed soil geography dataset provided by the USDA National Resources Conservation Service. SSURGO data are provided as a map layer with associated tables of relational attributes for each soil map unit. Many attributes, including those that are associated with wetlands, are described for soil components that are defined as fractional areas of map units without further spatial specification. Within each soil component, attributes are provided for vertical depth intervals (horizons). Each map unit may occur as multiple polygons, and many quantitative horizon attributes and fractional component areas are provided with high and low values to represent a range of estimates. Data from the NWI are provided by the U.S. Fish and Wildlife Service. Wetland types and riparian habitats are mapped with emphasis on repeated surveys and quality assurance to monitor time-dependent status and trends of wetland type areas.

The spatial distribution of wetland-associated soil components is subject to inherent uncertainties in the SSURGO database, and the time-dependent spatial detail of the NWI is not paralleled in SSURGO repeated surveys. Complex methods are therefore needed to extract the relevant soil information and integrate it with other data sources to determine the best available estimates of wetland soil properties. These methods typically require selection of appropriate soil components, extraction of component horizon data, and aggregation to map layers after correction for gaps and inconsistencies in the data. Because soil map units seldom align with wetland map layers, wetland soil properties must be imputed from proximal soil map units selected using other data sources, such as maps of topography, hydrography, and ecosystems.

We have experience with these procedures using Python code and GIS spatial analysis. The transferability of this work is limited. We propose to extend our work to the development of a toolkit that will enable other users to reproduce and adapt our methods for a broad range of applications. We will provide this toolkit in the form of code-generating Python code, Jupyter (iPython) Notebooks, and integrated geospatial datasets with metadata. These tools should provide complete reproducibility of the integrated datasets, as well as a platform for users to adapt and modify the procedures for their own use.

The development and sharing of this toolkit will leverage our existing capabilities and significantly improve the accessibility of the SSURGO dataset for a variety of applications beyond our immediate interests. Many of the basic procedures (such as extraction of component and horizon properties, aggregation of component horizons, and screening for gaps and inconsistencies) are already described in a draft catalog of recommended SSURGO "best practices," but they are not available in readily usable form. The availability of imputed soil properties will also enhance the value of the NWI in monitoring the status and trends of critical wetland services such as water-holding capacity and carbon storage.

Our proposed work will address several components of the CDI Science Support Framework. We will develop Computational Tools and Services in the form of a toolkit for data integration, extraction, and imputation. We will contribute to Data and Information Assets by developing integrated datasets of imputed wetland soil properties. We will incorporate the expertise and views of important Communities of Practice through involvement of co-investigators who serve as data managers and providers (NRCS, NWI, EROS) and scientist-users (NRCS, NWI, EROS, NRP). Most importantly, we believe that our contribution will be amplified beyond the duration of the proposed work through greater accessibility and broader applications of the Nation's outstanding soil and wetland data resources.