

CDI FY16 Statement of Interest**First CDI SSF element:** Data Management**Second CDI SSF element:** Science Data Lifecycle**Third CDI SSF element:** Science Project Support**Project Title:** Modeling integration of diverse biological, chemical, and land-use data: Endocrine Disruption Use Case**USGS cost centers:**

Eastern Geographic Science Center and Columbia Environmental Research Center

Principle Investigators:

Cassandra Ladino (orcid.org/0000-0002-9513-6028)

12201 Sunrise valley Drive MS521

Reston VA, 20192

T: 703-648-6188, F: 703-648-4603

Email: ccladino@usgs.gov

Jessica Leet (orcid.org/0000-0001-8142-6043)

4200 New Haven Road

Columbia, MO 65201

T: 573-441-2784, F: 573-876-1896

Email: jleet@usgs.gov**Additional Principle Investigators:** Kelly Smalling (ksmall@usgs.gov)

Project Description: Multiple science theme projects across multiple Science Centers with in the USGS Endocrine Disrupting Compound (EDC) Effects on Fish and Wildlife in the Chesapeake Bay Watershed project funded through the Contaminants Biology Program will be modeled to develop a strategy for integrating and improving preservation of field and laboratory data not maintained in other bureau databases. The strategy will reflect data management guidance and tools provided by the USGS and design of a database to store long term periods of baseline data. The documentation and methods from this project will inform other Programs and Science Centers in developing comprehensive data management strategies.

Project Deliverables:

- Workflow models for each science theme
- Data Management Plan template (based on existing USGS examples but highly adapted to project workflows)
- Database design model
- Presentations at CDI annual Workshop, Chesapeake Bay Workshop, and Chesapeake Bay Endocrine Disrupting Compound annual project meeting

Estimated Budget:

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
1. PERSONNEL (SALARIES including benefits):		
Kelly Smalling – EDC Project Manager		\$10,000
Cassandra Ladino – Information Specialist	\$10,000	\$20,000
Elizabeth Zinecker – Data Steward	\$5,000	
William Miller – Computer Scientist	\$10,000	
David Strong – Web Applications Developer		\$8,000
Jessica Leet – Scientist		\$9,000
Federal Personnel Total:	\$25,000	\$37,000
Contract/Collaborator Personnel Total:		
Total Salaries:	\$25,000	\$47,000
2. TRAVEL EXPENSES:		
Travel Total (Per Diem, Airfare, Mileage/Shuttle) x 3 Trips:	\$2,400	
Total Travel Expenses:	\$2,400	
3. OTHER DIRECT COSTS:		
Total Other Direct Costs:		

CDI15-LC3635

Total Direct Costs:	\$27,400	
Indirect Costs (24.589%):	\$6737.38	
GRAND TOTAL:	\$34,137.38	

Project Summary:

The USGS EDC Effects on Fish and Wildlife in the Chesapeake Bay Watershed project initiated in 2014 includes scientists from nine USGS science centers and multiple mission areas. This project, funded through the Contaminants Biology Program includes integrative studies with both field and laboratory components focused on identification of relevant EDCs, how they enter the Chesapeake Bay watershed, and how they affect aquatic organisms. To facilitate research related to synthesis products, workflows that allow integration of each project's key data holdings need to be developed. There is also an extensive amount of data that is not easily integrated or not appropriate for inclusion in existing Bureau databases such as NWIS and/or BioData.

Policies, Procedural Guidance, and Technical Infrastructure related to Data Management have been extensively developed by the USGS, however there is still a gap between bureau level guidance and implementation within the science centers. Science centers can adopt existing materials and tools, but often these materials are not tailored to science themes or the structure of projects. A strategy including workflow and database models specific to each science theme developed for the EDC project could be used to guide other Science Centers, Programs, and working groups in adopting more comprehensive data management practices. The lessons learned from this project can serve to improve the flow of data as it moves through its life cycle, leading to more effective data management for projects throughout USGS.

Using the EDC project as the case study, this project will build upon existing information and tools available via USGS Data Management guidelines and the Community for Data Integration to develop a more detailed guide for managing and integrating data for a complex and interdisciplinary project. This process will require the expertise of our information specialist and data steward (Cassandra Ladino and Elizabeth Zinecker), as well as input from the EDC scientists (Jessica Leet and Kelly Smalling) and a technical applications development team (William Miller, David Strong, and Cassandra Ladino). The data management team will use the input from the scientists to establish workflow models and work with the applications development team to model and develop a prototype database including a graphic user interface for data query, download, and spatial display. These applications will be valuable in conjunction with geographic analysis approaches being refined by the EGSC team (Dan Jones, Brianna Hammond, and Chris Wright) working to connect the Chesapeake Bay EDC data to landscape characteristics. The data models related to fish health information from this project will also be useful to help improve coordination with the USGS Aquatic Gap Program for Gap Analysis.

The main elements of this project are:

- 1.) Data Management: Establish an efficient process for managing the data for this complex project that can be applied to other similar landscape-scale projects throughout USGS
- 2.) Science Data Lifecycle: Apply this management to new and historical data from the Chesapeake Bay EDC project (i.e. processing fish health, land-cover, and historical water chemistry data not previously submitted to NWIS to be in a well-documented and accessible format)
- 3.) Science Project Support: Enhance the utility of these data in creating synthesis products and decision making as the Chesapeake Bay EDC project moves forward (particularly in planning complimentary laboratory studies based on field observations to establish mechanisms of action of specific EDCs)

The projected outcomes of this project will be data workflow models and a data management template that represents the data management strategy developed for the Chesapeake Bay EDC project. A database model representing the integration strategy will also be developed. These materials can then be used as a framework for other complex interdisciplinary projects throughout USGS that deal with managing many different types of data. A project summary will also be presented at the CDI annual meeting, the USGS Chesapeake Bay workshop, and the Chesapeake Bay EDC annual meeting.