

# CDI FY17 Request for Proposals

## USGS Data at Risk: Expanding Legacy Data Inventory and Preservation Strategies

**Submission Title:** USGS Data at Risk: Expanding Legacy Data Inventory and Preservation Strategies

**Lead PI:** Anthony Everette

**Mission Area:** Ecosystems

**Region:** Southwest Region

**Organization:** Fort Collins Science Center

**Orcid:** 0000-0003-2539-9129

**Phone:** 9702269225

**Email:** everettel@usgs.gov

**City:** Fort Collins

**State:** CO

**Co-PIs and Collaborators:**

**Type:** CO-PI

**Name:** John Faundeen

**Mission Area:** Ecosystems

**Region:** Southwest Region

**Organization:** Earth Resources Observation and Science

**Orcid:** na

**Phone:** 6055946092

**Email:** faundeen@usgs.gov

**City:** Sioux Falls

**State:** SD

**Science Support Framework Element 1:** Science Data Lifecycle - Preservation

**Science Support Framework Element 2:** Science Data Lifecycle - Processing

**Science Support Framework Element 3:** Science Data Lifecycle - Publishing/Sharing

**In-Kind Match:** \$20,000.00

**List of anticipated deliverables from the project:** 1. More complete USGS legacy data inventory 2. Legacy Data Inventory and Reporting System (LDIRS) v.2 with improved evaluation, prioritization and reporting features. 3. Case studies describing real legacy data inventory efforts by 3-5 USGS science centers. 4. Preservation of 3-5 legacy data products with significant potential impact to Missions that are at significant risk of damage or loss.

**Lead Cost Center:** Fort Collins Science Center

**Notes, Comments:**

**Project Description:** This FY17 proposal seeks to expand on those DaR16 project successes by 1) increasing the growth rate of the USGS legacy data inventory; improving the legacy data evaluation and prioritization methods based on FY16 DaR project experiences and feedback; conduct a 2017 "request for legacy data" competition to continue to identify USGS Mission-critical legacy data and provide technical support to preserve and publish the data at greatest risk of loss.

**Total Budget:** \$50,000.00

**Title: USGS Data at Risk: Expanding Legacy Data Inventory and Preservation Strategies**

**USGS Lead Principal Investigator: Lance Everett**

**Narrative:** As one of the largest and oldest science organizations in the world, USGS has produced decades of historical science data that is currently unavailable to the greater scientific community yet could be invaluable for extending our historical understanding of the world's natural resources, landscapes and hazards. Recognizing the value of preserving legacy data for new scientific questions, USGS has been investigating methods of inventorying and preserving legacy data through projects such as the USGS Data Rescue Program (2006-2013), the Legacy Data Inventory and Reporting System (LDIRS; CDI 2014), and the FY16 "Data at Risk" project (DaR; CDI 2016).

In particular, the FY16 DaR project represents a confluence of USGS data policy, technology, and passionate determination to provide USGS science centers with legacy data preservation support, tools, and methods. The primary objectives and results of the FY16 DaR project were:

1. *Create a USGS legacy data inventory that catalogs and describes known USGS legacy data sets.*  
We used the Legacy Data Inventory and Reporting System (LDIRS) to conduct a USGS-wide "Request for Legacy Data" competition. We received 43 submissions from 20 USGS science centers with potential impacts across all USGS Missions. An additional 200+ unfunded "Data Rescue Program" requests received from 2006-2013 are being added to the inventory. The continuously growing legacy data inventory is available at:  
<https://www.fort.usgs.gov/ldi/legacy-products>
2. *Develop methods to evaluate and prioritize legacy data sets based on USGS Mission objectives.*  
The FY16 DaR project team has developed a "DaR16" method to score risk and significance factors associated with individual legacy data products, calculate inventory priorities, and report the results. You can learn more about the "DaR16 Risk and Significance Evaluations" at:  
<https://www.fort.usgs.gov/ldi/evaluations-reports/dar16>
3. *Preserve and release select, priority legacy data sets at risk of damage or loss.*  
Based on the DaR16 evaluation priorities, we selected the top 5 legacy data products and are currently on schedule to preserve and publish these as official USGS data releases by the end of 2016. The list of DaR's selected FY16 preservation projects is available at:  
<https://www.fort.usgs.gov/ldi/2016-data-at-risk-project/selected-fy16-products>
4. *Develop time and resource estimates to preserve and release legacy data.*  
For each of the 5 selected preservation projects, we're collecting data on the resources required to complete each step of our data management plans (e.g., acquire, process, analyze, publish). This will better inform future legacy data preservation and release estimates.

This FY17 CDI proposal seeks to expand on the DaR16 project successes by:

1. *Increasing the growth rate of the USGS legacy data inventory.*  
We will conduct complete legacy data inventories of 3-5 USGS science centers and develop case studies based on those efforts. Case studies will be available via LDIRS. We will also present the case studies and the DaR methods at the CDI 2017 Workshop and a CDI community meeting later in FY17.
2. *Improving the legacy data evaluation and prioritization methods based on FY16 DaR project experiences and feedback.*

This will include score refinement of existing LDIRS risk and significance factors; the addition of new risk and significance factors; automation of the per-submission aggregation of evaluation scores and prioritization algorithms; and creation of legacy data dashboards that display multiple LDIRS reports in a single user display for Mission Area and science center reports.

3. *Conducting a 2017 "Request for Legacy Data" competition to continue to identify USGS Mission-critical legacy data products and provide technical support to preserve and publish those data at greatest risk of loss.*

We will conduct another "Request for Legacy Data" competition, select 3-5 preservation projects with the greatest potential impact to USGS Missions and most-at-risk of damage or loss. We will provide preservation expertise and support to preserve and publish those products in 2017 and continue to gather resource requirement data to better inform all USGS preservation efforts.

**BUDGET:**

Budget Category	Funding Requested	Matching Funds
1. PERSONNEL (SALARIES including benefits):		
Federal Personnel Total:	0.00	18,000.00
Contractor Personnel Total:	27,200.00	0.00
Total Salaries:	27,200.00	18,000.00
2. TRAVEL EXPENSES:		
Travel Total:	7500.00	0.00
<ul style="list-style-type: none"> <li>• 2017 CDI Workshop Presentation</li> <li>• Science center inventory/case study visits</li> </ul>		
Total Travel Expenses:	7500.00	0.00
3. OTHER DIRECT COSTS: (itemize)		
Equipment:	2000.00	2000.00
Shipping:	500.00	0.00
Total Other Direct Costs:	2500.00	2000.00
Total Direct Costs:	37,200.00	20,000.00
Indirect Costs (26%):	12,800.00	0.00
<b>GRAND TOTAL:</b>	<b>50,00.00</b>	<b>20,000.00</b>

\*Note: Travel must be included for at least one representative to attend a CDI event. Travel cannot include field data collection.