STATEMENT OF INTEREST CDI RFP FOR FY 2016

SECTION 1. PROJECT ADMINISTRATION INFORMATION

- **Elements of CDI SSF**: Monitoring, Assessment and Research; Data; Processing; Analysis; Publishing/Sharing; Semantics; Applications
- **Project title**: Crowd-sourced Earthquake Detections Integrated into Seismic Processing
- **USGS cost center requesting funding**: Geologic Hazards Science Center, Golden CO
- **USGS principal investigator**: Michelle Guy; orcid.org/0000-0003-3450-4656; PO Box 25046 MS 966, Denver Federal Center, Denver CO 80225-0046; office 303-273-8650; fax 303-273-8450; mguy@usgs.gov
- **Additional USGS principal investigators and collaborators**: Paul Earle; PO Box 25046 MS 966, Denver Federal Center, Denver CO 80225-0046; office 303-273-8417; fax 303-273-8450; pearle@usgs.gov; Jessica Turner; PO Box 25046 MS 966, Denver Federal Center, Denver CO 80225-0046; office 303-273-8574; fax 303-273-8450; jturner@usgs.gov
- **Project description**: Improve USGS National Earthquake Information Center's (NEIC's) global and regional earthquake detection capabilities through integration of crowd-sourced earthquake detections into existing seismic processing systems. The crowd-sourced earthquake products will follow internationally recognized formats for seismic data exchange and be distributable via existing standard mechanisms, such that the products can be consumed by numerous systems and many organizations that already adhere to these standards.
- **Anticipated deliverable**s:
  1. Documented and well-formatted real-time crowd-sourced event location data products, extending international standard format, which seamlessly integrates with an existing data distribution mechanism and data consumers.
  2. Documented analysis of crowd-sourced event detections integrated into existing seismic processing system and resulting solutions
  3. Upgraded real-time data analysis and visualization open-source web interface.
  4. Improved methods for earthquake location using crowd-sourced data

SECTION 2. ESTIMATED BUDGET

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
1. SALARIES (including Benefits):		
Personnel Total: Analysis, Design and IT Support	--	$18,000
Contract Personnel Total: Software Implementation and Testing	$40,000	--
Total Salaries:	$40,000	$18,000
2. TRAVEL EXPENSES:		
Travel Total: Travel for Michelle Guy to CDI Event	$1,100	--
Total Travel Expenses:	$1,100	--
3. OTHER DIRECT COSTS: (itemize)		
Equipment (development and production hardware and infrastructure):	--	$10,000
Publishing Fee's? Training		
Total Other Direct Costs:	--	$10,000
Total Direct Costs:	$41,100	
Indirect Costs (15%): overhead assessment	$6,150	--
GRAND TOTAL	$47,250	$28,000

SECTION 3. PROJECT SUMMARY

The extensive spatial coverage and near instantaneous distribution of crowd-sourced data enables the rapid detection of earthquakes often before data is available from sparsely distributed seismic sensors. One of the fundamental drawbacks of using crowd-sourced data is the validation and distillation of the voluminous data. Validation is vital if the derived information is shared with the public by an authoritative source such as the USGS. We propose validating crowd-sourced information by integrating tweet-based earthquake detections into the standard NEIC earthquake processing system thus improving NEIC's response time and increasing the number of detected earthquakes. The project builds on a previous CDI proposal and leverages multiple existing systems and data sets. It will result in information products that follow international standards that can be readily distributed and programmatically consumed.

The project will take automatically derived detections from the operational system called TED, which for over four years now has been rapidly detecting felt earthquakes, globally, from tweets only. We will turn the tweet-based, crowd-sourced detections from TED into well-formatted earthquake location products, following the international standard QuakeML (http://earthquake.usgs.gov/earthquakes/feed/v1.0/quakeml.php). We will then feed the formatted detections into NEIC's real-time, global seismic detection and processing system, called Hydra. Both of these existing systems perform 24x7 monitoring and provide assessment tools for scientists to interact with the datasets in both real-time and with historical data.

By following the QuakeML standard, adding in extensions as needed, we achieve seamless integration (no code changes required) across a large suite of local software systems and with numerous collaborators corresponding to both the Publishing/Sharing and Semantics elements of the CDI SSF. The formatted, crowd-sourced location products will be distributed via an existing distribution mechanism, called Product Distribution Layer (PDL), used by the NEIC and numerous collaborators (https://github.com/usgs/pdl). These rapid detections will then be seeds for seismic algorithmic processing on location, phase data association, and magnitudes (http://pubs.er.usgs.gov/publication/ofr20151120). Analysis of this innovative data integration, between qualitative crowd-sourced information and quantitative scientific processing, will determine if timeliness of solutions from the NEIC can be improved and manual efforts reduced. If so, then this integration will supplement a weakness of the automated global seismic system where it can miss smaller, regional earthquakes, most significantly due to sparse coverage of instruments. The crowd-sourced event detections are generally faster than seismically derived global detections, however, they have significantly larger location uncertainties, so improvements to the crowd-sourced location constraints will need to be explored as well.

Significantly, in support of the data analysis effort, the team looks to upgrade to the latest version of Kibana 4, the open source data analysis and visualization software currently used, https://www.elastic.co/guide/en/kibana/4.0/whats-new.html. This software upgrade supports the project's sophisticated data analysis user interface needs as well as better visual information presentation in mobile friendly webpages and for embedding in distributed information products to users, such as emails. These efforts relate to the Applications element of the CDI SSF.

The project will have four major milestones as listed in the anticipated deliverables section. Milestone one, formatting in QuakeML and distribution into seismic processing systems, will take on the order of only two weeks to complete due to using well-established standards. This will set the stage for the remainder of the milestones to be worked on concurrently allowing the team to meet an overall completion well before the end of the fiscal year.

The project is multidisciplinary and the PIs include experts in computer science (Guy, USGS), seismology (Earle, USGS), and seismic monitoring analysis (Turner, USGS). This team has worked with a contractor out of the USGS Fort Collins office several times over the past five years, where the work follows the Agile software development process. This means the team meets every two weeks and reviews deliverables and refines goals for the following two weeks, which mitigates risk and ensures accurate project deliverables on time.