

## IARPA-RFI-14-04

### Synopsis

#### Request for Information (RFI): Coding Societal Events

The Intelligence Advanced Research Projects Activity (IARPA) is seeking information on methods to extract and code societal events from unstructured data, and information on existing structured databases of such events. For this RFI, a “societal event” is meant in a broad sense and includes, but is not limited to, social, political, epidemiological, cyber, economic, counterintelligence, and science and technology events. This RFI is issued solely for information gathering and planning purposes; this RFI does not constitute a formal solicitation for proposals. The following sections of this announcement contain details on the scope of technical efforts of interest, along with instructions for the submission of responses.

### Background & Scope

IARPA develops technologies to forecast a broad set of well-defined societal events relevant to national security. The test and evaluation process for these technologies requires an objective “ground truth” for events, generated in near-real-time. For many of the events of interest, the ground truth is developed by extracting events from unstructured data, often news text. IARPA is interested in event coding for specific classes of societal events, as well as solutions for retraining an event coder to code new event classes, as they emerge. The purpose of this RFI is to identify existing event coders and event data, and to identify potential approaches to event coding that would advance the state-of-the-art for future programs.

Specifically, the purpose of this RFI is to identify:

- *Existing structured databases of societal events relevant to national security.* Relevant events include those listed above. Databases that are in the public domain or to which the Government has rights are preferable, but commercial databases can also be included. Databases of historical events that are not maintained or up-to-date are not of interest. Detailed descriptions of the event class, the unstructured data source and coding method used, and the fields in the database should be included. Databases that include entries for event type, actor, date/time, and location, are preferred.
- *Existing taxonomies/ontologies of specific event classes.* The taxonomies/ontologies of interest are those for which events have explicit coding rules that a human analyst could follow with expectation of reasonable inter-coder agreement. Simple lists of events, with vague definitions, for an event class are not of interest.
- *Existing methods (e.g., processes, models, or products) that detect a new, emergent event and/or actor class based on the analysis of streaming data and develop a taxonomy/ontology for that new class.* Such methods need not be fully automated, but some automation is required. A discussion of how these methods could also support the development of explicit coding rules for the new class, and the level of efforts it would require to develop such coding rules, is of interest.

- *Existing methods (e.g., processes, models, or products) that extract and code specific events from unstructured data.* Such methods need not be fully automated, but some automation is required. Performance metrics for these methods are essential and should be included. Where possible, metrics should be compared to those published for other event coders, such as SERIF or TABARI. A discussion of how these methods perform across different data sources (e.g., news report versus blogs) is of interest, as are approaches to deduplicating event entries. A discussion of the cost to implement and maintain (e.g., in terms of data, methods, labor, and/or computational resources) is also of interest.
- *Existing methods (e.g., processes, models, or products) to retrain an event coder to code a new class of events.* Such methods need not be fully automated, but the level of effort required to train the event coder on the second class of events should be significantly less than the level of effort to train the event coder on the first class of events. Performance metrics for these methods are essential and should be included.
- *Methods (e.g., processes, models, or products) that could be used as the basis for the development of a “generic” event coder.* Such methods should generate outputs that can be processed to obtain a specific class of events, including the broad set of events envisioned in this RFI. These methods should also be able to integrate existing or new taxonomies/ontologies of specific event classes.
- *Metrics and protocols to assess the performance of an event coder and the performance of the process required to train it to code a new class of events.* We are particularly interested in metrics other than precision, recall, or F-measure.

Responses to this RFI should answer any or all of the following questions:

- 1) What structured databases exist that contain both historical and frequently updated events relevant to national security? These events include, but are not limited to social, political, epidemiological, cyber, economic, counterintelligence, and science and technology events.
- 2) What are existing taxonomies/ontologies of specific event and/or actor classes?
- 3) What are existing methods to detect a new, emergent event and/or actor class based on the analysis of the streaming data?
- 4) What are existing methods to extract and code events from unstructured data?
- 5) What are existing methods to retrain an event coder to code a new class of events?
- 6) What are existing methods that could be used as the basis for the development of a “generic” event coder capable of coding many new classes of events?
- 7) What are metrics and protocols to assess the performance of an event coder and the performance of the process required to train it to code a new class of events?
- 8) What novel approaches to event coding do you propose that would advance the state-of-the-art as described in your answers to questions 1-7?

### **Preparation Instructions to Respondents**

IARPA requests that respondents submit ideas related to this topic for use by the Government in formulating a potential program. IARPA requests that submittals briefly and clearly describe the potential approach or concept, outline critical technical issues/obstacles, describe how the approach may address those issues/obstacles and comment on the expected performance and

robustness of the proposed approach. If appropriate, respondents may also choose to provide a non-proprietary rough order of magnitude (ROM) regarding what such approaches might require in terms of funding and other resources for one or more years. This announcement contains all of the information required to submit a response. No additional forms, kits, or other materials are needed.

IARPA appreciates responses from all capable and qualified sources from within and outside of the US. Because IARPA is interested in an integrated approach, responses from teams with complementary areas of expertise are encouraged.

Responses have the following formatting requirements:

1. A one page cover sheet that identifies the title, organization(s), respondent's technical and administrative points of contact - including names, addresses, phone and fax numbers, and email addresses of all co-authors, and clearly indicating its association with RFI-14-04;
2. A substantive, focused, one-half page executive summary;
3. A description (limited to 5 pages in minimum 12 point Times New Roman font, appropriate for single-sided, single-spaced 8.5 by 11 inch paper, with 1-inch margins) of the technical challenges and suggested approach(es);
4. A list of citations (any significant claims or reports of success must be accompanied by citations, and reference material MUST be attached);
5. Optionally, a single overview briefing chart graphically depicting the key ideas.

### **Submission Instructions to Respondents**

Responses to this RFI are due no later than 4:00pm, Local Time, College Park, MD on March 21, 2014. All submissions must be electronically submitted to [dni-iarpa-rfi-14-04@iarpa.gov](mailto:dni-iarpa-rfi-14-04@iarpa.gov) as a PDF document. Inquiries to this RFI must be submitted to [dni-iarpa-rfi-14-04@iarpa.gov](mailto:dni-iarpa-rfi-14-04@iarpa.gov). Do not send questions with proprietary content. No telephone inquiries will be accepted.

### **DISCLAIMERS AND IMPORTANT NOTES**

This is an RFI issued solely for information and planning purposes and does not constitute a solicitation. Respondents are advised that IARPA is under no obligation to acknowledge receipt of the information received, or provide feedback to respondents with respect to any information submitted under this RFI.

Responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Respondents are solely responsible for all expenses associated with responding to this RFI. IARPA will not provide reimbursement for costs incurred in responding to this RFI. It is the respondent's responsibility to ensure that the submitted material has been approved for public release by the information owner.

The Government does not intend to award a contract on the basis of this RFI or to otherwise pay for the information solicited, nor is the Government obligated to issue a solicitation based on responses received. Neither proprietary nor classified concepts or information should be included

in the submittal. Input on technical aspects of the responses may be solicited by IARPA from non-Government consultants/experts who are bound by appropriate non-disclosure requirements.

**Contracting Office Address:**

Office of the Director of National Intelligence  
Intelligence Advanced Research Projects Activity  
Washington, District of Columbia 20511  
United States

**Primary Point of Contact:**

Jason Matheny  
Intelligence Advanced Research Projects Activity  
dni-iarpa-rfi-14-04@iarpa.gov