



Better Extraction From Text Towards Enhanced Retrieval (BETTER) Proposers' Day

John R. Beieler, PhD
Program Manager

March 29, 2018

Disclaimers

- This presentation is provided solely for information and planning purposes
- The Proposers' Day does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers' Day changes the requirements set forth in a BAA
- A BAA supersedes anything presented or said by IARPA at the Proposers' Day

Goals

- Familiarize participants with IARPA's interest in the BETTER program
- Please ask questions and provide feedback, as this is your chance to alter the course of events.
- Foster discussion of complementary capabilities among potential program participants, AKA teaming. Take a chance, someone might have a missing piece of your puzzle.

Questions

- During this session, questions should be recorded on note cards. They will be answered for everyone's benefit at a later point in the presentation.
- If/when a BAA is released, questions can only be submitted to the e-mail address provided in the BAA and will only be answered in writing on the program website.



Agenda

Time	Topic	Speaker
8:30AM - 9:00AM	Registration	
9:00AM - 9:15AM	Welcome, Logistics, Proposers' Day Goals	Dr. John R. Beieler Program Manager, IARPA
9:15AM - 9:45 AM	IARPA Overview	Dr. Paul Lehner Chief of Testing and Evaluation, IARPA
9:45AM - 10:30AM	BETTER Program Overview	Dr. John R. Beieler Program Manager, IARPA
10:30AM - 11:00AM	Break	
11:00AM - 11:20AM	Doing Business with IARPA	Acquisitions Team, IARPA
11:20AM - 12:00PM	BETTER Questions & Answers	Dr. John R. Beieler Program Manager, IARPA
12:00PM - 1:30PM	No-Host Lunch	
1:30PM - 2:30PM	Offerors' Capabilities Briefings	Attendees (No Government)
2:30PM - 5:00PM	Poster Session, Networking and Teaming Discussions	Attendees (No Government)

IARPA Overview

**Dr. Paul Lehner, Chief of Test and Evaluation
Intelligence Advanced Research Projects Activity**



Office of the Director of National Intelligence

I A R P A
BE THE FUTURE



The United States Intelligence Community



IARPA Mission

IARPA envisions and leads *high-risk, high-payoff research* that delivers innovative technology for future *overwhelming intelligence advantage*

- Our problems are **complex** and **multidisciplinary**
- We emphasize **technical excellence & technical truth**

IARPA Method

Bring the best minds to bear on our problems

- Full and open competition to the greatest possible extent
- World-class, rotational Program Managers

Define and execute research programs that:

- Have goals that are clear, measureable, ambitious and credible
- Employ independent and rigorous Test & Evaluation
- Involve IC partners from start to finish
- Run from three to five years
- Publish peer-reviewed results and data, to the greatest possible extent
- Transition new capabilities to intelligence community partners



4 Core Research Thrusts



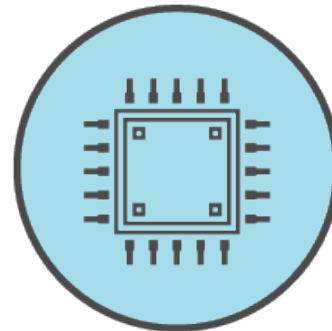
Analysis



Anticipatory Intelligence



Collection



Computing



Analysis R&D



“Maximize insight from the information we collect, in a timely fashion”



LARGE DATA VOLUMES
AND VARIETIES

Provide powerful new sources of information from massive, noisy data



SOCIAL, CULTURAL, AND
LINGUISTIC FACTORS

Analyze language and speech to produce insights into groups and organizations



IMPROVING ANALYTIC
PROCESSES

Enhance analytic process at the individual and group level



Anticipatory Intelligence R&D



“Detect and forecast significant events”



S & T
INTELLIGENCE

Detect and forecast the emergence of new technical capabilities



INDICATIONS &
WARNINGS

Provide early warning of societal crises, disease outbreaks, insider threats, and cyber attacks



STRATEGIC
FORECASTING

Forecast major geopolitical trends and rare events



Collection R&D



“Dramatically improve the value of collected data”



NOVEL ACCESS

Reach hard targets in denied areas



ASSET VALIDATION AND
IDENTITY INTELLIGENCE

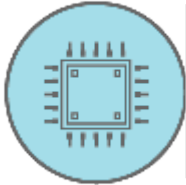
Assess trustworthiness and advance biometrics in real-world conditions



LOCATING, TRACKING
AND DETECTING

Accurately locate and track intelligence interests and detect CBRNE agents

Computing R&D



“Operate effectively in a globally interdependent and networked environment”



COMPUTATIONAL
POWER

Revolutionary advances to solve problems intractable with today's computers



TRUSTWORTHY
COMPONENTS

Gain the benefits of leading-edge hardware and software without compromising security



SAFE AND SECURE
SYSTEMS

Protecting systems against cyber threats



How to Engage with IARPA

Getting Started with IARPA

At IARPA, we take real risks, solve hard problems, and invest in high-risk/high-payoff research that has the potential to provide our nation with an overwhelming intelligence advantage.

Are you interested in partnering with us to advance the state-of-the-art in research and development?

[Read More](#)

iarpa.gov | 301-851-7500

info@iarpa.gov

Reach out to our Program Managers.

Schedule a visit if you are in the DC area or invite us to visit you

Opportunities to Engage:

RFIS AND WORKSHOPS

Opportunities to learn what is coming, and to influence programs.

“SEEDLINGS”

Typically a 9-12 month study; you can submit your research proposal at any time. We strongly encourage informal discussion with a PM before proposal submission.

PRIZE CHALLENGES

No proposals required. Submit solutions to our problems – if your solutions are the best, you receive a cash prize and bragging rights.

RESEARCH PROGRAMS

Multi-year research funding opportunities on specific topics.

Better Extraction From Text Towards Enhanced Retrieval (BETTER) Program Overview



Office of the Director of National Intelligence

I A R P A
BE THE FUTURE

BETTER Overview

- BETTER is anticipated to be a multi-year research and development program
- The program aims to develop enhanced methods for personalized, multilingual semantic extraction and retrieval from text
- BETTER seeks to accomplish this by combining research efforts from information extraction, information retrieval, and active learning to create technologies that enable the extraction and retrieval of fine-grained, personalized knowledge

BETTER

- Challenge
 - Too much to read
 - Multiple languages
 - Finding orthogonal information
- Opportunity
 - Better methods to extract complex semantic information from documents
 - Using extracted semantic information to surface relevant documents
- This program creates technologies that enable the extraction and retrieval of fine-grained, personalized knowledge for an analyst across diverse domains and languages.

Scenario 1

Analyst 1 is interested in a wide range of political events, including civil unrest. Through the course of her daily workflow, she develops a corpus of documents that captures the nature of the events, entities, and links that are of interest to her problem area.



Scenario 1 – How can BETTER help?

- Provide automated suggestions of relevant events, entities, and links highlighted in the analyst's discovered reporting
- Provide suggestions of most relevant documents based on extracted features
- Automatically tag incoming information according to the analyst's pre-determined knowledge
- *Systematize and provide automated support to the knowledge extraction process conducted by an analyst*

Scenario 2

- Analyst 2 has been issued a quick-turn question focusing on civil unrest events by civilian actors in Country Y. Analyst 2 lacks deep domain expertise in this area, but is able to find a handful of reports that capture the essence of the phenomenon on which he is focused.



Scenario 2 – How can BETTER help?

- Leverage information developed by Analyst 1 in Scenario 1 to inform Analyst 2's problem area
 - Avoid developing a massive corpus and relearning a solved problem area
- Broaden the search criteria based on previously extracted entities, events, and relations
 - Find orthogonal information based on extracted events from differing ontological classes
- *Make use of existing knowledge to compress the discovery cycle*

BETTER Approach

- Complex extraction
 - Fuzzy events and actors
- Personalized extraction
 - No more “one ontology to rule them all”
- IE+IR
 - Use semantic features to triage documents
- Multilingual
 - One-to-many models

Program Organization

Phase

Language 1

Domain 1

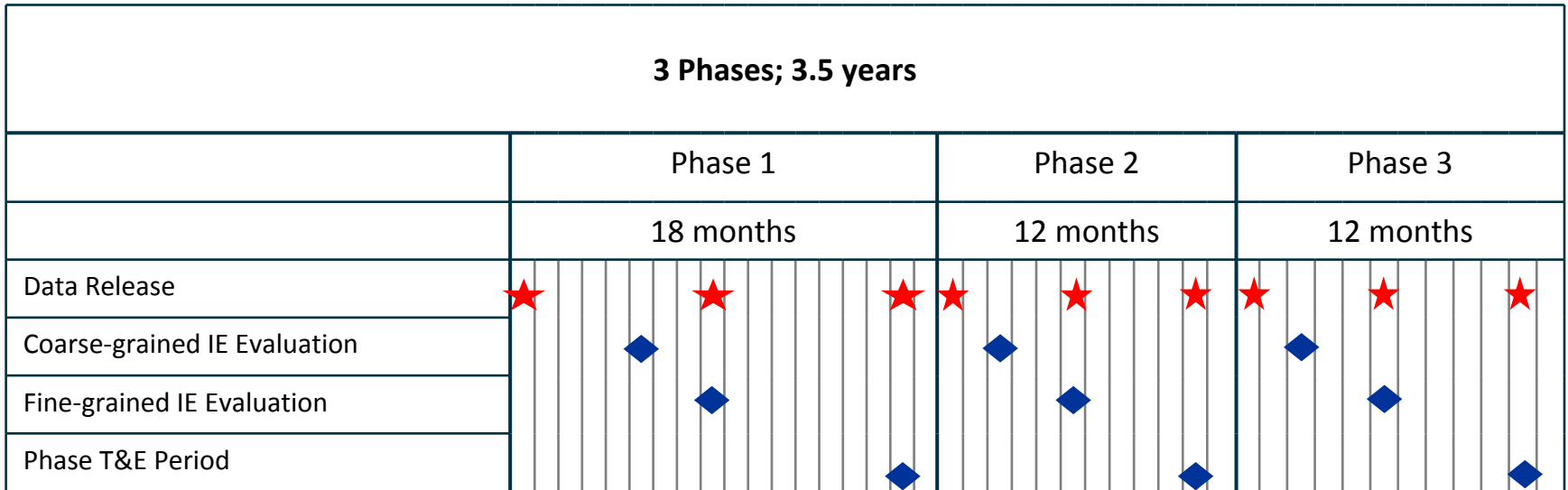
Language 1

Domain 2

Languages 2 & 3

Domain 3

Program Timeline



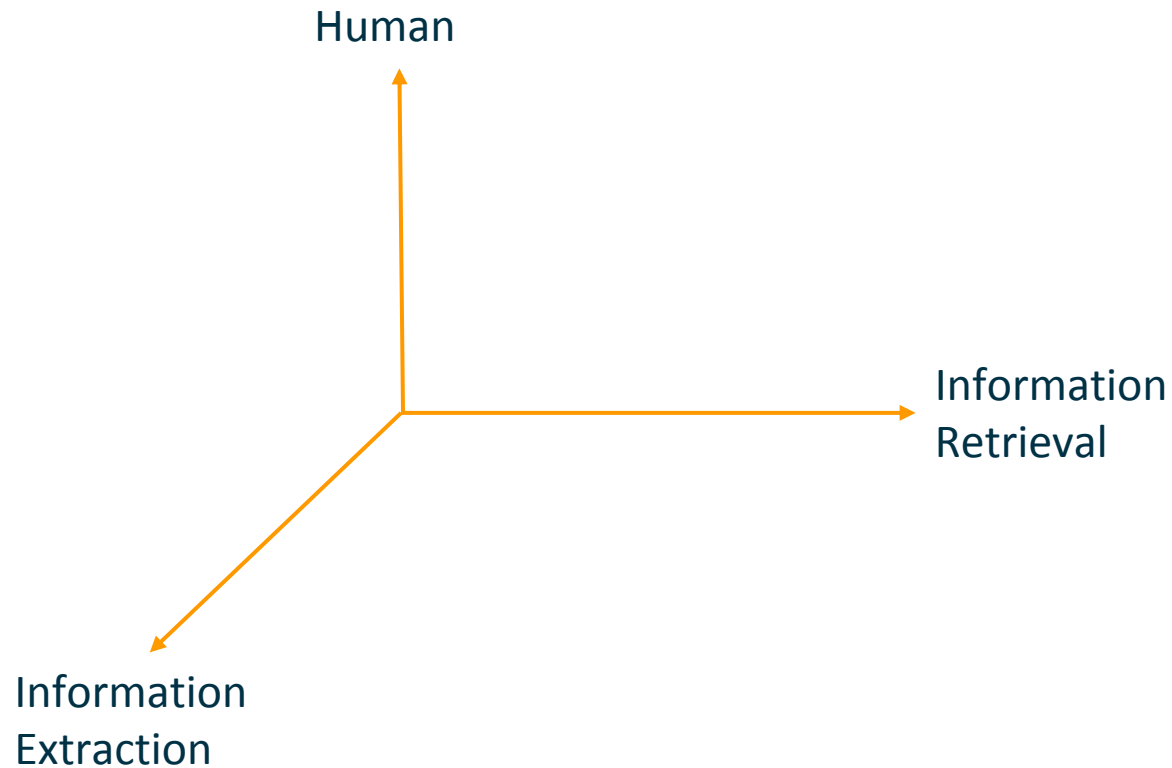


Testing what?

- Ability to extract complex events
 - Sentence-/paragraph-level classification of event type
 - Span+role ID of actors
- Ability to perform cross-lingual extraction
- Ability to apply extracted information towards a semantic information retrieval task
- Ability to incorporate human feedback into developed models



Test and Evaluation Overview



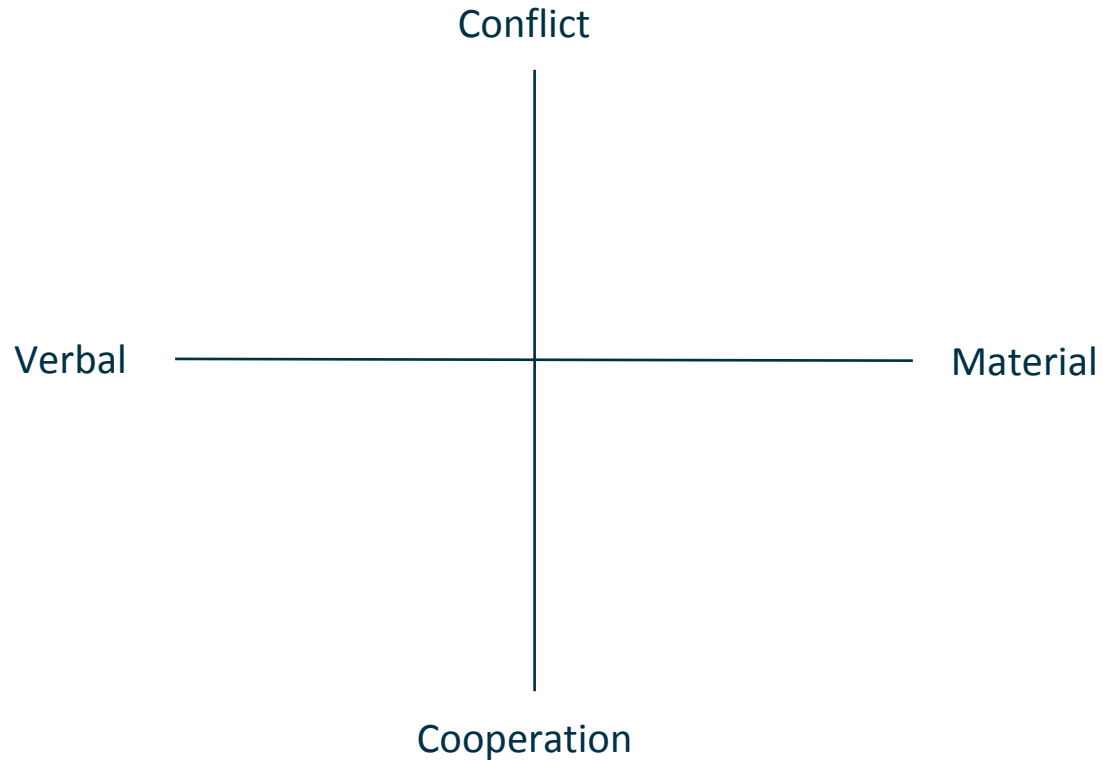


Test and Evaluation Example

- Teams are tasked with extracting political events from a corpus.
 - High level, QuadClass ontology of material/verbal cooperation/conflict
- Performers extract QuadClass events, followed by a fine-grained event (attack), and finally discover new aspects of the fine-grained event (military action)



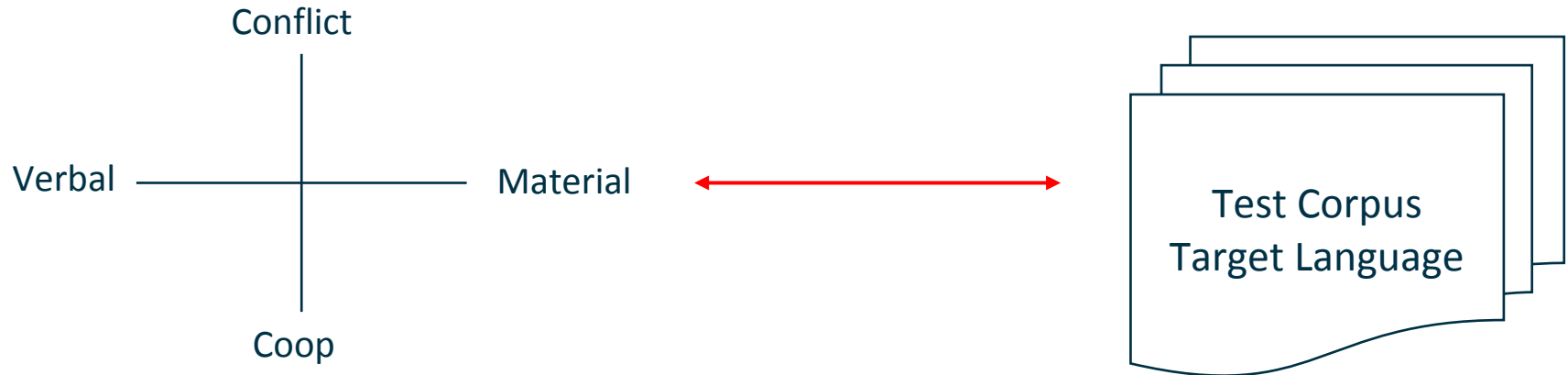
QuadClass



Test and Evaluation Example

Step 1 – Coarse Extraction

Train Corpus



Performers extract who-did-what-to-whom using the coarse **QuadClass labels**. Training examples are provided in English, while testing occurs against a target-language corpus, e.g., French.



Test and Evaluation Example

Step 2 – Fine-grained Extraction

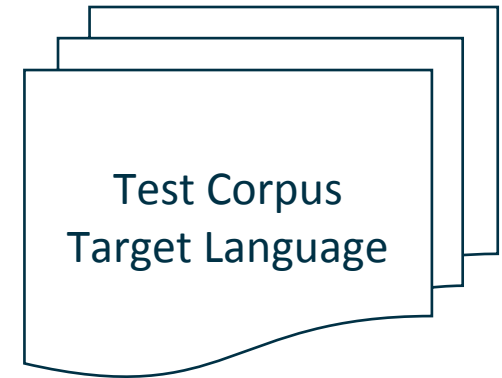
Train Corpus

Material

Rebels attacked the town.

The protestors marched in opposition to recent measures by the government.

Conflict





Test and Evaluation Example

Step 2 – Fine-grained Extraction

Train Corpus

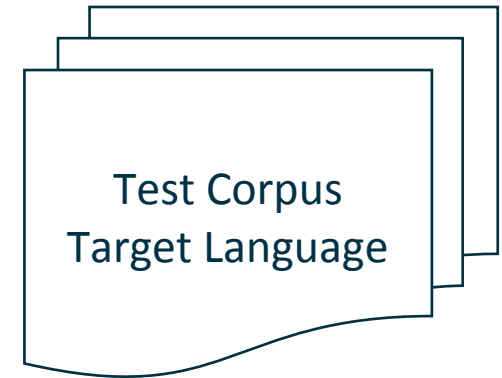
Material

ATTACK

[Rebels attacked the town.]

The protestors marched in opposition to recent measures by the government.

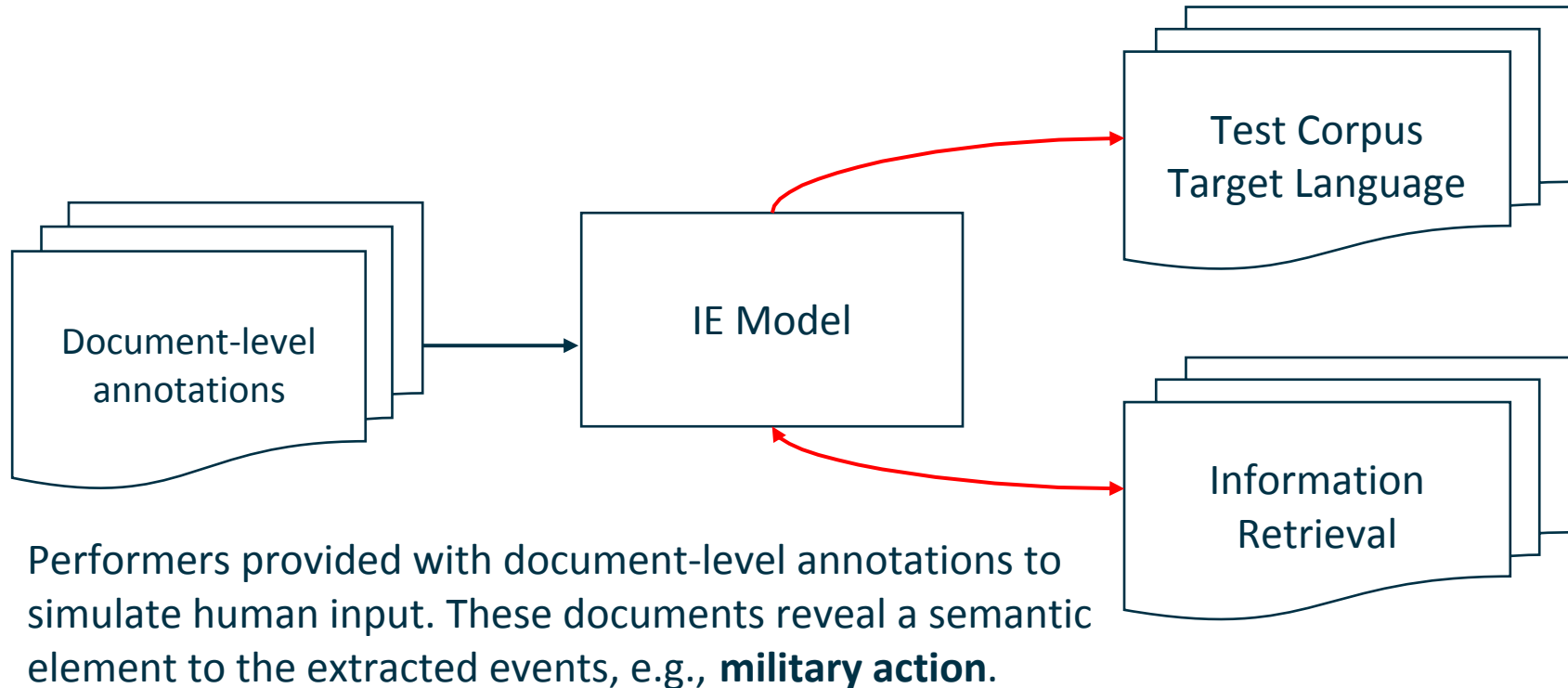
Conflict



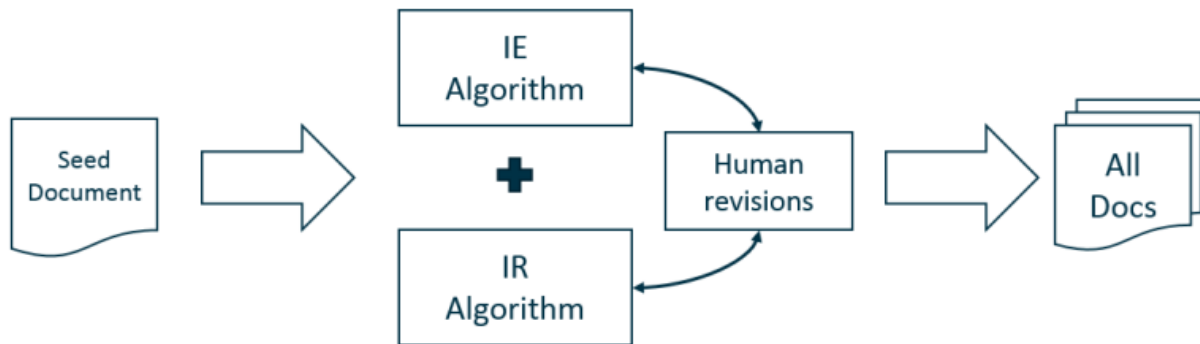
Performers make use of existing, coarse-grained models to extract a fine-grained event, e.g., **attack**. A driving assumption is that the extracted events from Step 1 will be directly applicable to the extraction of attack events, i.e., attack events will be a subset of events extracted in Step 1. Performers will be shown a limited number, n , of annotated examples in the training corpus used in Step 1. This n will be variable throughout the life of the program.

Test and Evaluation Example

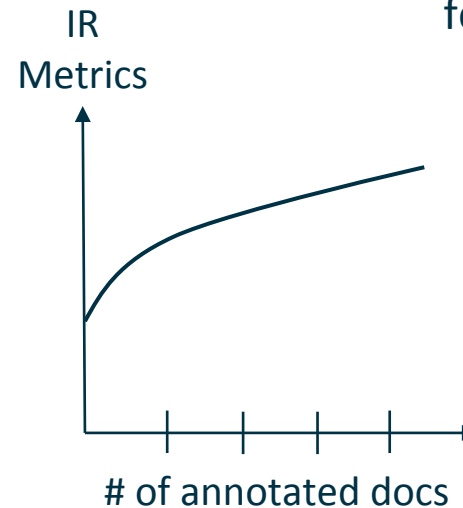
Step 3 – Semantics and IR



Test and Evaluation Example Human Interactions



The IR evaluation will include a human annotation component where performers are allowed to solicit limited feedback on documents.

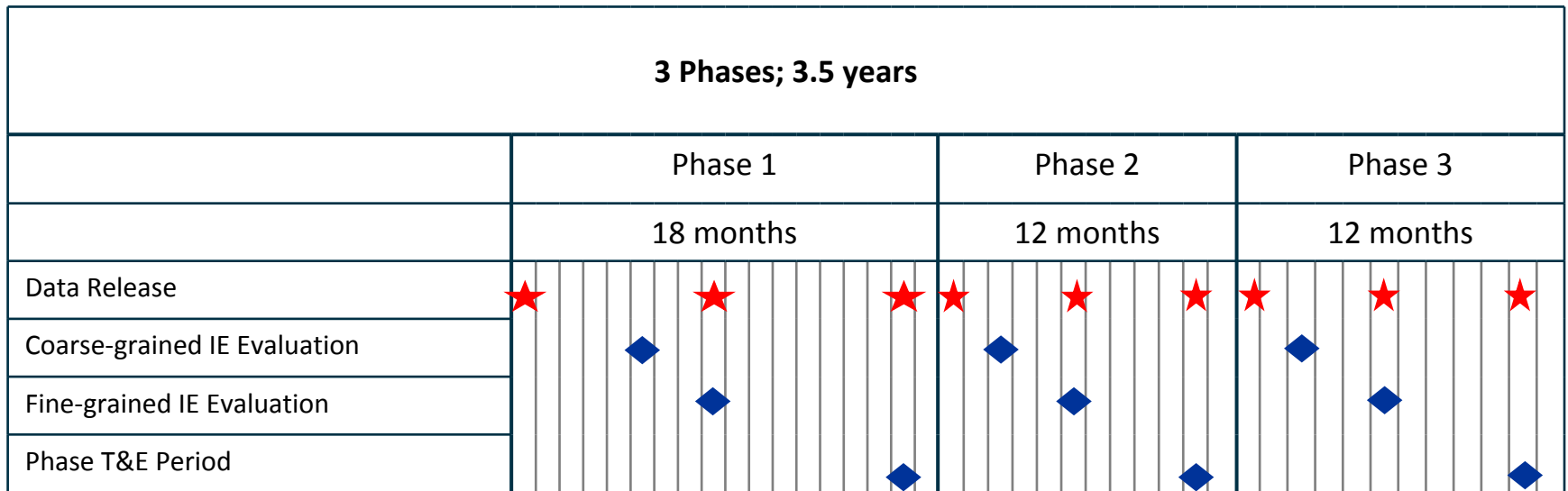


T&E Tasks Recap

- Task 1 – Coarse grained event extraction
 - Sentence-/paragraph-level classification of event type
 - Span+role ID of actors
- Task 2 – Fine grained event extraction
 - Sentence-/paragraph-level classification of event type
 - Span+role ID of actors
- Task 3 – Finer grained event extraction + information retrieval
 - Sentence-/paragraph-level classification of event type
 - Document-level annotations
 - Span+role ID of actors
 - Combine information extraction with information retrieval task
 - Include human-in-the-loop input



Program Timeline





Metrics

- **IE Tasks: Overall F1-score** – The balance between event classification F1 and span+role ID F1.
 - $F1_{event} * F1_{span}$
 - F1-score – Mean between precision and recall.
 - $2 * \frac{precision * recall}{precision + recall}$
- **IR Task: Average precision** – A summary of the precision-recall curve at each cut-point, k , for relevancy determinations. A perfect ranking of all relevant documents would give an average precision of 1.
 - $\frac{\sum_{k=1}^n (P(k) \times rel(k))}{\text{number of relevant documents}}$, where n is the total number of documents
 - $P(k)$ is the precision, as defined in the IE Tasks section, at cut-point k
 - $rel(k)$ is the binary relevancy determination of item k



T&E Milestones Phase 1

Description	Metric	Month 6 Waypoint	Month 9 Waypoint	Month 17 Milestone
Coarse IE		.40		
Fine-grained IE	F1		.40	
Finer-grained IE				.35
IR	AP			.10

*Values in cells are percent reduction in error from baseline model. Formula is $(1 - \text{baseline}) * \text{reduction} + \text{baseline}$.

Parameters

- **Revisions** – The number of human judgments allowed for each performer system
- **Document annotations** – The number of document-level annotations provided for the information retrieval task

Program Parameters

Description	Phase 1	Phase 2	Phase 3
Revisions	50	40	20
Document Annotations	~1000	~100	~10

Data

- Focus on news-like data
 - E.g., New York Times, Washington Post, etc.
- Government team will provide annotated datasets with enough training examples to enable a wide range of techniques

Data Regimes

- Two data regimes: constrained and unconstrained
- Constrained
 - Performers only allowed to use resources identified by the Government team
- Unconstrained
 - Performers allowed to use data harvested from any source
 - May not use proprietary datasets and must provide the Government with an accounting of all resources used and sources from which data is drawn

Deliverables

- Developed models delivered in software containers
 - Docker
- Models must be capable of interacting with an API
 - REST or message queue
- Results must output to a JSON schema
- *Exact API and schema details will be provided upon program kickoff*

Point of Contact

Dr. John R. Beiler

Program Manager

IARPA, Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity
Washington, DC 20511

Phone: (301) 851-7441

Fax: (301) 851-7673

Electronic mail: dni-iarpa-baa-18-05@iarpa.gov
(include IARPA-BAA-18-05 in the Subject Line)

Website: www.iarpa.gov

Questions? Please fill out cards.



Agenda

Time	Topic	Speaker
8:30AM - 9:00AM	Registration	
9:00AM - 9:15AM	Welcome, Logistics, Proposers' Day Goals	Dr. John R. Beieler Program Manager, IARPA
9:15AM - 9:45 AM	IARPA Overview	Dr. Paul Lehner Chief of Testing and Evaluation, IARPA
9:45AM - 10:30AM	BETTER Program Overview	Dr. John R. Beieler Program Manager, IARPA
10:30AM - 11:00AM	Break	
11:00AM - 11:20AM	Doing Business with IARPA	Acquisitions Team, IARPA
11:20AM - 12:00PM	BETTER Questions & Answers	Dr. John R. Beieler Program Manager, IARPA
12:00PM - 1:30PM	No-Host Lunch	
1:30PM - 2:30PM	Offerors' Capabilities Briefings	Attendees (No Government)
2:30PM - 5:00PM	Poster Session, Networking and Teaming Discussions	Attendees (No Government)

Doing Business with IARPA

Acquisition Team



Office of the Director of National Intelligence

I A R P A
BE THE FUTURE



Doing Business with IARPA - Recurring Questions

- Questions and Answers (<http://www.iarpa.gov/index.php/faqs>)
- Eligibility Info
- Intellectual Property
- Pre-Publication Review
- Preparing the Proposal (Broad Agency Announcement (BAA) Section 4)
 - Electronic Proposal Delivery (<https://iarpa-ideas.gov>)
- Organizational Conflicts of Interest
(<http://www.iarpa.gov/index.php/working-with-iarpa/iarpas-approach-to-oci>)
- Other points to consider

RECOMMENDATION: Please read the entire BAA. There may be some changes to our standard BAA template so please take note.



Questions & Answers

- There will be a specified period for questions stated in the BAA. All questions and answers will be posted. (Note: Questions may be submitted not only regarding technical requirements but all other sections of the BAA).
- Send your questions as soon as possible
 - BETTER BAA: **dni-iarpa-baa-18-05@iarpa.gov**
 - Write questions as clearly as possible
 - Do NOT include proprietary information *or mark as proprietary or otherwise confidential.*
- Pay attention to Section 4 (Proposal & Submission Information)
- Frequently Asked Questions can be found on the IARPA website: <http://www.iarpa.gov/index.php/faqs>

Eligible Applicants

- Collaborative efforts/teaming strongly encouraged
 - Content, communications, networking, and team formation are the responsibility of Proposers

- Foreign organizations and/or individuals may participate subject to: Non-Disclosure Agreements, Security Regulations, Export Control Laws, etc., as appropriate. See BAA for further information.

Ineligible Organizations

Other Government Agencies, Federally Funded Research and Development Centers (FFRDCs), University Affiliated Research Centers (UARCs), and any organizations that have a special relationship with the Government, including access to privileged and/or proprietary information, or access to Government equipment or real property, are not eligible to submit proposals under this BAA or participate as team members under proposals submitted by eligible entities.



Intellectual Property (IP)

- Intellectual Property Ownership.
 - The Government generally does not seek to own the intellectual property in technical data and computer software developed under Government contracts; it generally acquires only the right to use the technical data/computer software.
 - Thus, performers may usually freely use their data for their own commercial purposes (unless restricted by U.S. export control laws or security classification).
 - For inventions first conceived or actually reduced to practice under a contract, grant, or cooperative agreement for this effort, IARPA will obtain a nonexclusive, nontransferable, irrevocable, paid-up license to practice, or have practiced for or on its behalf, such invention throughout the world; Offeror may elect to retain title as described in the award instrument.
- Please note that IARPA generally uses the Government Purpose Rights (GPR) approach for data developed with mixed funding, however, please see the BAA for specific program requirements regarding IP.
- *Explicitly state* in the proposal any *asserted* restrictions on deliverables relating to existing materials (data, software, tools, etc.)



Pre-Publication Review

- IARPA encourages publication for peer review of UNCLASSIFIED research
- It is anticipated that research funded under this Program will be unclassified research that shall not require a pre-publication review. However, performers should note that pre-publication approval of certain information may be required if it is determined that its release may result in the disclosure of sensitive intelligence information.
- A courtesy soft copy of any work submitted for publication shall be provided to the IARPA Program Manager and the Contracting Officer Representative (COR) a minimum of 5 days prior to release in any forum.



Preparing the Proposal

- Follow the detailed instructions for preparing the proposal (BAA Section 4)
- Proposal submissions (BAA Section 4):
 - Offerors must follow BAA instructions regarding submittal of proposal
 - Interested Offerors must register electronically IAW instructions on: <https://iarpa-ideas.gov>
 - Interested Offerors are strongly encouraged to register in IDEAS at least 1 week prior to proposal “Due Date”
 - For Classified proposal submissions, the BAA will have further instructions. Not anticipated for BETTER.
- Check FBO & IARPA website for amendments and Q&As
- Read proposal Evaluation Criteria and Method of Evaluation and Selection carefully (BAA Section 5)



Organizational Conflict of Interest (OCI)

- If a prospective offeror, or any of its proposed subcontractor teammates, believes that a potential conflict of interest exists or may exist (whether organizational or otherwise), the offeror should promptly raise the issue with IARPA as instructed in the BAA. (BAA Section 3)
- Potential conflicts include but are not limited to any instance where an offeror, or any of its proposed subcontractor teammates, is providing either scientific, engineering and technical assistance (SETA) or technical consultation to IARPA.
- Without a waiver from the IARPA Director, neither an offeror, nor its proposed subcontractor teammates, can simultaneously provide SETA support or technical consultation to IARPA and compete or perform as a Performer under this solicitation.



Other

- IARPA funds Applied Research for the Intelligence Community (IC)
- IARPA cannot waive the requirements of Export Administrative Regulation (EAR) or International Traffic in Arms Regulation (ITAR)
- Not subject to DoD funding restrictions for R&D related to overhead rates - IARPA is not DoD



Disclaimer

- Content of the Final BAA will be specific to this program
- The information conveyed in this brief and discussion is for planning purposes and is subject to change prior to the release of the Final BAA.

Point of Contact

Dr. John R. Beiler

Program Manager

IARPA, Office of the Director of National Intelligence
Intelligence Advanced Research Projects Activity
Washington, DC 20511

Phone: (301) 851-7441

Fax: (301) 851-7673

Electronic mail: dni-iarpa-baa-18-05@iarpa.gov
(include IARPA-BAA-18-05 in the Subject Line)

Website: www.iarpa.gov

Questions? Please fill out cards.



Agenda

Time	Topic	Speaker
8:30AM - 9:00AM	Registration	
9:00AM - 9:15AM	Welcome, Logistics, Proposers' Day Goals	Dr. John R. Beieler Program Manager, IARPA
9:15AM - 9:45 AM	IARPA Overview	Dr. Paul Lehner Chief of Testing and Evaluation, IARPA
9:45AM - 10:30AM	BETTER Program Overview	Dr. John R. Beieler Program Manager, IARPA
10:30AM - 11:00AM	Break	
11:00AM - 11:20AM	Doing Business with IARPA	Acquisitions Team, IARPA
11:20AM - 12:00PM	BETTER Questions & Answers	Dr. John R. Beieler Program Manager, IARPA
12:00PM - 1:30PM	No-Host Lunch	
1:30PM - 2:30PM	Offerors' Capabilities Briefings	Attendees (No Government)
2:30PM - 5:00PM	Poster Session, Networking and Teaming Discussions	Attendees (No Government)