Advanced Research and Development Activity (ARDA)

Call for 2005 Challenge Workshop Proposals

The Advanced Research and Development Activity (http://www.ic-arda.org) requests Challenge Workshop proposals for 2005.

ARDA is an intelligence community organization whose mission is to sponsor high-risk, high-payoff research designed to leverage leading edge technology to solve some of the most critical problems facing the Intelligence Community (IC). ARDA has established three regional centers to create partnerships between government and industry/academic experts to identify and engage in focused workshops to actively solve complex IC problems. ARDA encourages novel approaches, non-traditional government contractors, and new cross-organizational teams.

This call invites proposals for both “Challenge Workshops” and “Seedling Workshops.” The objectives of the Challenge Workshops are (1) to fund revolutionary research addressing critical IC needs and (2) to accelerate research and development (R&D) to real solutions, prototypes or first proofs-of-principle. Challenge workshops typically support existing ARDA technical thrust areas, are funded at the $1M to $1.5M level, and are conducted over a 12-18 month period.

Seedling Workshops, by comparison, are funded at lower levels ($100K-$500K), and may explore new areas outside of the existing ARDA thrust areas. They are short term, novel investigations that might result in larger workshops or new ARDA research thrusts. Examples of seedling concepts include: a baseline of the current state of the art in an emerging area, definition of new research areas, creation of a research roadmap, or exploration of a novel, high-risk information technology concept. Seedling Workshop proposals should be written using the same format as that required for Challenge Workshops specified in this call.

2005 Challenge Workshop Topic Areas

The 2005 Challenge Problems will focus on the following technical areas of interest, although ARDA may also consider compelling proposals that fall outside these areas:

1. Information Exploitation (Info-X, http://www.ic-arda.org/InfoExploit/index.html) in particular the VACE and AQUAINT programs.
2. Novel Intelligence from Massive Data (NIMD)
3. Advanced Capabilities for Intelligence Analysis (ACIA)
4. Information Assurance (IA)
5. Advanced Imaging (as a seedling workshop)
6. Nanoelectronics for High Performance Computing (as a seedling workshop)

A key criterion for the selection of a Challenge Workshop proposal will be a credible plan for transitioning workshop results and technologies for further development and implementation in the Intelligence Community. To that end, proposals are expected to identify government champions who can facilitate this technology transition. In addition, ARDA has extended the funding level ($1M-$1.5M) and duration (up to 18 months) from previous years to support the development and engineering required to successfully demonstrate the practical use of novel research results, as a precursor to their adoption by the IC.

Challenge Workshops will be executed through ARDA's three regional research centers: The Northeast Regional Research Center (NRRC, http://nrrc.mitre.org), the Northwest Regional Research Center (NWRRC), and the newly-established Southeast Regional Research Center (SRRC). A Challenge project will likely consist of multiple collaborative workshops; and ARDA invites proposers to work with the Regional Research Center directors to design workshop formats/models that would most successfully achieve results and IC impact.

1. Information Exploitation (Info-X) Challenge Problems

ARDA is interested in Information Exploitation (Info-X, http://www.ic-arda.org/InfoExploit/index.html) challenge workshop proposals supporting the VACE and AQUAINT advanced R&D programs.

1.1 VACE Program

The objective of the Video Analysis and Content Extraction (VACE) Program is the research and development of new algorithms and implementations for automated video content extraction, multi-modal fusion and event understanding. In VACE Phase I, the program achieved significant advances in video content analysis. These advances include techniques for the automated detection of scene objects such as human bodies, human faces, vehicles and both overlay and in-scene text. Initial research results have also been obtained for tracking, enhancing and recognition of these objects. Additionally, some very preliminary progress was made on the automated analysis of human activities and interpretation of video sequences.

In VACE Phase II, the program continues to build on the core technologies developed in Phase I by making them more robust and scalable. VACE Phase II builds on this foundation with a focus on developing capabilities that are specific to video - such as processing full motion video and addressing variables such as camera position, focal length, motion, distortion, and atmospheric conditions. Technologies can be applied to five primary video data sources: news broadcast video, meeting/conference video, surveillance video, unmanned aerial vehicle (UAV) motion imagery, and ground reconnaissance video (i.e. video scenes of various indoor and outdoor activities involving people, vehicles and facilities shot from handheld cameras).

Ultimately, the goals of the overall VACE Program are to develop technologies that will provide:

1. Significant improvement in indexing and retrieval performance for video data;
2. Autonomous video understanding;
3. Ancillary improvement for still image processing of key frames extracted from video stream;
4. Enabling the development of advanced applications/processing functions for video querying/retrieval, video browsing, video monitoring, video mining, and content-based routing; and
5. A drastic reduction in volume for video storage and forwarding mechanisms.

VACE supports both the application of technology to achieve important new video capabilities for the analyst and the development of underlying enabling technology that makes those applications possible. Throughout Phase II and into Phase III, the program will continue to aspire towards the goal of investigating innovative algorithms and implementations that represent revolutionary advances in automatic video content extraction, multi-modal fusion and event understanding.

1.2 AQUAINT Program

ARDA's Advanced Question Answering for Intelligence (AQUAINT) Program is pursuing advanced research for scenario-based, advanced question answering in which multiple, inter-related questions are asked in a particular topic area by a skilled, professional information analyst who is attempting to respond to larger, more complex information needs or requirements. This vision goes significantly beyond question and answer capabilities for single, isolated, factually-based questions whose answers can be found as a single string or within a relatively short window of text in a single document. In phase I, the program achieved significant advances in the development of question and answering methodologies. In Phase II, the program will continue to build on the core technologies developed in Phase I by making them more robust and scalable. Research and development will include:

1. Question Understanding and Interpretation (including contextual interpretation, query expansion, query taxonomy),
2. Determining the Answer (including information retrieval and extraction from multiple media/languages and data types, interpretation, synthesis, resolving conflicting information, justification),
3. Formulating and Presenting the Answer (including summarization, synthesis, generation), AND/OR
4. Cross-Cutting/ Enabling/Enhancing Technologies that directly and materially support the goals of the AQUAINT Program and one or more of the areas (1) to (3) listed above (including advanced reasoning, content representation, user interaction, NLP, contextual analysis)

ARDA is also interested in proposals that are technologically cross-cutting and involve enabling technologies that would directly and materially impact the future direction and success of the AQUAINT Program and one or more of the four AQUAINT Challenge areas listed above.

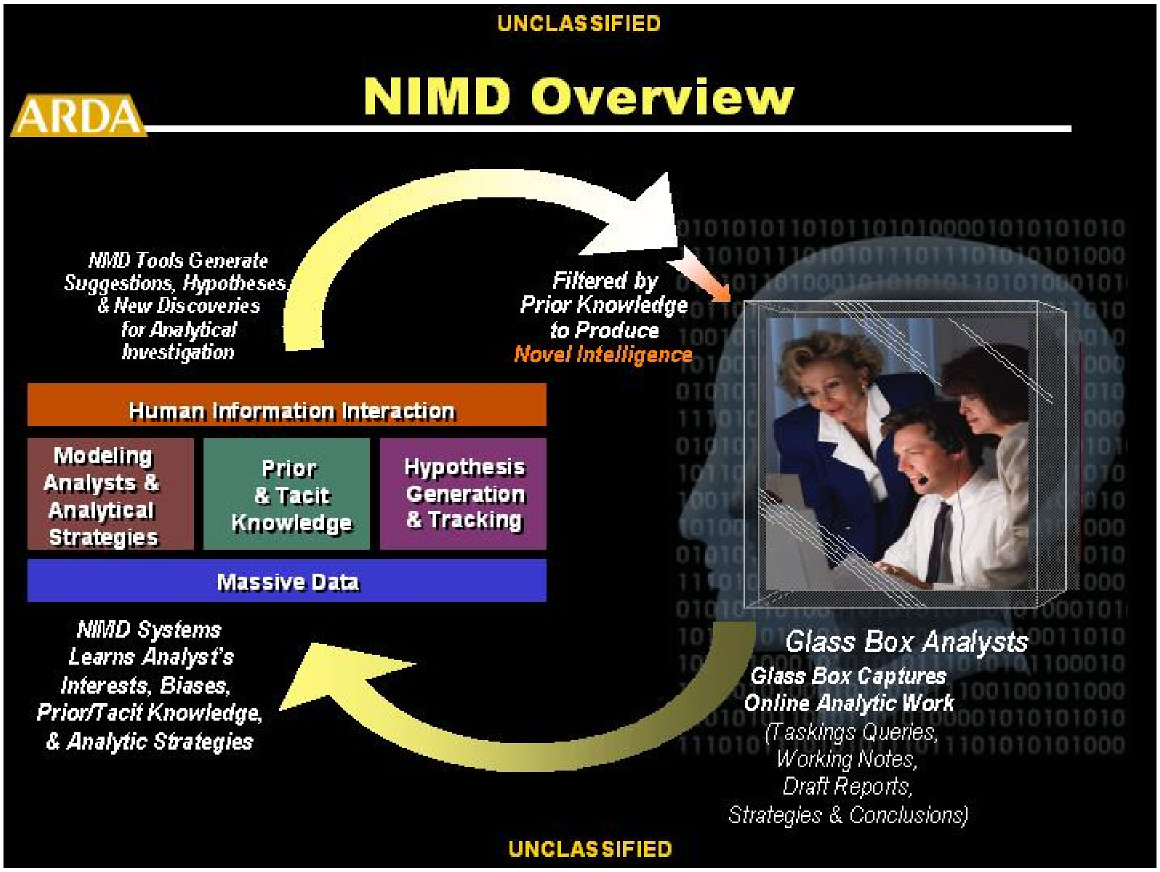
1.3 Other Information Exploitation Areas; Cross-cutting Topics

While VACE and AQUAINT are the primary areas of interest, ARDA will also consider Challenge Problem proposals that address the overall Information Exploitation program in areas not covered by the above, such as data filtering and selection, content data markup, content data transformation, information discovery, information understanding, synthesis and fusion, information retrieval, analytic knowledge, presentation and visualization, assessment and interpretation, and information analysis. Of particular interest are cross cutting areas that address important challenges such as new methods for evaluation, annotated data (along with annotation standards), architectural issues, and advancing the art and science of analyst/system interaction including issues of human computer interaction, tacit knowledge and dialogue-discourse.

2 Novel Intelligence from Massive Data (NIMD)

NIMD aims to preempt strategic surprise by addressing root causes of analytic errors related to bias, assumptions, and premature attachment to a single hypothesis. The program may also assist with capture and reuse of analytic best practices. NIMD Phase I target users are open-source and all-source analysts, while future work will extend to multi-INT and collaborative analysis.

At the heart of NIMD is a piece of software called the Glass Box that resides on an analyst's workstation and captures the parts of the analytic process that happen online – queries and the sources to which they are sent, search results, documents viewed, draft reports, etc. – recording what an analyst actually does online during analysis. NIMD research is developing techniques and tools that infer the state of analysts and the analytic process from Glass Box data, assist analysts in making explicit their analytic (cognitive) state, and use the captured knowledge and analytic models to drive automated organization and exploration of massive data. The majority of NIMD research is focused in five areas: modeling analysts and analytic processes, capturing and reusing prior and tacit knowledge, generating and managing hypotheses, organizing/structuring massive data (mostly unstructured text), and human interaction with information. In short, NIMD funds research to build smart software assistants and devil's advocates that help analysts deal with information overload, detect early indicators of strategic surprise, and avoid analytic errors.



ARDA is interested in challenge proposals in two key areas: data triage and knowledge representation.

2.1 The **Data Triage** challenge is to develop predictive measures of significance and models that enable real-time categorization of streaming petabytes of data to support decisions about which data to store, which to elevate for immediate analysis, and which to delete without further attention. Fundamental questions in this challenge area include:

- How can we anticipate the analytic value of information (as close to the source as possible) to avoid throwing out important data (due to storage/transmission limitations)?
- How can a system anticipate future analytic needs and paradigm shifts?
- Are anomaly detection measures insufficient? What will work?

The product of this activity could be algorithms to meet needs or plans for a future program. A key measure of success would be the accuracy of predicted analytic value.

2.2 The **Knowledge Representation (KR)** challenge is to practically represent knowledge in a form that can enhance machine processing and enable interoperability of technology developed by multiple vendors for multiple ARDA programs of research. The knowledge in question may be about users, external threats or situations, the analytic tasking(s) at hand, analytic processes, or from data being analyzed. Fundamental questions in this challenge area include:

- What are the costs (lost functionality, reduced performance, translation/mediation development costs) and benefits (smaller set of formats to support) of reducing the number of KR formats supported by the IC?
- How do we capture, represent, and use context? Which dimensions of context are critical to reuse of knowledge for decision-making? For subsequent analysis? For understanding analytic audit trails?
- How do we capture and represent key constructs such as assumptions, hypotheses, data, evidence, projections or simulations that extend beyond more conventionally represented elements such as entities, attributes, relations, events, and topics?
- What are the relationships among high-level representations of knowledge (graph representations, structured argumentation, planning, inference nets, logical forms, text, etc.)? Can they be translated?
- How do these relationships affect traditional knowledge representation questions (knowledge representation languages, ontologies, etc.)?

Possible solutions might include limiting the number of KR alternatives, providing a standard interface or markup language (e.g., OWL or beyond), creating knowledge transformations (e.g., KIF or beyond). The resultant product could include a report laying out a cost/benefit analysis of the path ahead (e.g., some limited number of ontologies, KR languages, system interoperability).

3. Advanced Capabilities for Intelligence Analysis (ACIA)

ARDA is interested in challenge workshop proposals that address one or more of the following priority analysis problem areas that would result in advanced technical means to enhance analytic tradecraft:

3.1 Sensemaking: Intelligence analysts are often described as folks who "find needles in haystacks" or "connect the dots" or "put the puzzle pieces together." Perhaps. But in the world of asymmetric threats, we first have to recognize that the objects seen through an intelligence aperture indeed form a haystack, or a puzzle, or a set of connectable dots. We need to derive or discover holistic contexts and scenarios that make the presence and interactions of these objects "sensible". Furthermore, we have to do this without regard to whether any of the observed objects will eventually prove important in answering intelligence questions. Even if we think we see a smoking gun, we can't discover its import without knowledge of the putative crime scene. "Sensemaking" or generating plausible contexts that account for sets of observations and objects, is thus akin to turning a random collection of books into a library (or into multiple, alternative libraries) by indexing, cataloging, and arranging them without regard to how library patrons might eventually value particular volumes. Most intelligence systems are built to drive down the sheer volume of collected information through some sort of value-driven "filtering and selection". Sensemaking requires that we absorb – and welcome – "everything".

3.2 Pathfinding: If "Sensemaking" is akin to a library-building process, then "pathfinding" is the activity pursued by the library's patrons in navigating the stacks. Intelligence analysts refer to the "thread of logic" that leads from one knowledge source to another in the context of "proving an intelligence theorem" (e.g., a proposition that answers an intelligence question). Pathfinding is the generation and pursuit of all such threads, including those that analysts may have inadvertently overlooked.

3.3 Present information in future contexts: Intelligence analysts share a recurring nightmare that begins with some investigative commission asking "What did you know, and when did you know it?" All too often, the conclusions drawn from the analysis of present information are irrelevant at best or incorrect at worst in the context of future events or discoveries (including analytic discoveries). Is there a way to revisit "old" data and conclusions (particularly when those "conclusions" result in bypassing data) to alert analysts that their previous impressions now require revision? Or to be more adventurous, is there a way to construct and use plausible futures in order to provide additional, novel interpretations for today's collection?

3.4 Rapid, massive, multilingual information discovery: A workshop related to the above areas could address the challenge of topically processing massive, multilingual volumes of textual information at high speeds. Processing might need to generate and/or exploit both document bibliographic metadata (e.g., data, time, author, length, format) as well as document metadata content (e.g., language, subject, entities, relationships within the document). Processing challenges include the need for very rapid search, clustering, correlation (with current or previously viewed data and metadata), topic detection and tracking, selection, and dissemination (especially of extracts given large data sizes). ARDA is interested in progress measured by quantitative and qualitative performance measures (e.g., 95% correct detection of a language within 50 words).

4. Information Assurance (IA) Challenge Problem

ARDA has a comprehensive program in information assurance that addresses challenges such as intrusion detection, malicious software, and insider threat. In addition, there is a growing need for significantly greater information sharing within the IC, DoD, and DHS. Advanced information sharing requires information assurance technologies to ensure (1) the ability to securely move data among different levels of security and (2) effective downgrading and declassification. There is a growing concern regarding malicious or inadvertent insider exfiltration of classified information via unauthorized channels.

ARDA is interested in challenge workshop proposals that promise to enhance the data sharing and downgrading process. Current methods typically either search for classified elements via keyword or strings of words or statistically tag a corpus of documents into classified and unclassified subsets with about 70-80% accuracy. These methods are manually intensive and/or require human training and thus are error prone and inherently inefficient. ARDA seeks challenge workshop proposals that will develop an ontological semantic-based approach to provide unambiguous translation of classification guides into machine executable rules and, using these rules, demonstrate increased performance for sanitization, declassification, downgrading, and security control of sensitive textual and data information. ARDA expects proposers to develop evaluation methodologies and metrics to assess the effectiveness and accuracy of the approach as well as experimental results that provide evidence supporting research assertions/claims. Some elements of this workshop will be conducted at the TS//SCI level, therefore, the proposal should address how the workshop and its participants will address this.

5. Advanced Imaging

ARDA is interested in a Seedling Workshop that will advance imaging systems performance in an integrated, lightweight, and compact structure. This imaging system will use diffractive optics and pupil phase diversity to achieve improved spatial and spectral resolution. Of particular interest is a modular periodic architecture with the overall system design optimally exploiting information theoretic methods (ala Fisher) in the context of an interdependent optical/detection/processing/visualization subsystem and system. The imaging system should perform as a hyperspectral and/or ultraspectral camera in contrast to an electro-optic imager.

Primary challenges include development of dispersive diffractive optics and the utilization of partial interferometric beam combination to improve spatial resolution. Proposals should include measurable milestones such as an analytical proof of principle, simulation of module, modular design, prototype development, performance of subsystem and/or laboratory demonstration.

6. Nanoelectronics for High Performance Computing

ARDA is interested in a Seedling Workshop whose aim is to formulate a strategy and roadmap for the application of nanoeletronics to high performance computing. Molecular scale engineering affords us the ability to put molecules or other nanostructures where we want them, by design. We can build entire systems of macroscopic extent (e.g., very powerful new “nanocomputers”); ultra-dense, nose-like sensors; and smaller, more efficient power sources). *Nano*electronics is especially important because of challenges with *micro*electronics miniaturization including fabrication, cost, interconnect and heat dissipation challenges.

ARDA is interested in a challenge workshop that will initiate dialog between developers of high-performance computing and nanoelectronics researchers/engineers. ARDA would like to assess the potential for unique features of nanoelectronics to drive forward high-performance computing, especially low power and high density. ARDA would like to develop a roadmap for applying nanoelectronics to enhance the performance of next-generation supercomputers. The aim is to stimulate out-of-the-box thinking in the high-performance computing industry that is so mission critical to the U.S. Intelligence Community.

Challenge Problem Proposal Content and Format

The Challenge Problem proposal shall not exceed 10 pages and must succinctly address each of the following key elements:

Problem: Succinct definition of problem to be addressed. (1 page or less)

Approach: Method for solving the problem, e.g., collection of data, creation of algorithms, evaluation, integration and test of existing heterogeneous capabilities, study of human processes to inspire new approach. (2 pages or less)

Domain and data sets: Size, required annotation, availability/intellectual property. (less than 1 page)

Evaluation: Measures and methods, qualitative/quantitative. (less than 1 page)

Proposed leader, team and roles Lists names of individuals and their institutions, as well as their primary role, e.g., lead, annotator, developer, statistician, etc. (1 page or less)

Plan: Key tasks/milestones, dates, including pre-workshop preparation, training, lectures/seminars and potential post-workshop activities including technology transition (see next). (3 pages or less)

Government Champion and Technology Transition: Identification of a credible and engaged Government Champion and technology transition plan, to include addressing any special required data sets, testing, integration, and security/classification issues. (less than 1 page)

Impact: Product (e.g., software, algorithms, data, report), performance, process, or other outcomes. (1 page or less)

Resources: Required staff, data, tools, and infrastructure. (less than 1 page)

Issues: Membership, resources, intellectual property, other (less than 1 page)

A proposal template and example of a past successful proposal will soon be available at http://nrrc.mitre.org/.

Proposal Cover Sheet

Each proposal shall have a one-page cover sheet that includes the following information:

(1) Program addressed, i.e., [1]. Info-X (AQUAINT, VACE or Other); [2] NIMD (Data Triage or KR); [3]ACIA (Sensemaking, Pathfinding, Future Contexts, Rapid Discovery); [4] Information Assurance; [5] Advanced Imaging; [6] Nanoelectronics; [7] Other.

(2) Challenge focus (e.g., from the above list or other)

(3) Proposal title
(4) Technical point of contact including: name, telephone number, electronic mail address, fax (if available) and mailing address
(5) Administrative point of contact including: name, telephone number, electronic mail address, fax (if available) and mailing address
(6) Government champion including: name, telephone number, electronic mail address, fax (if available) and mailing address
(7) Summary of the resources of the proposed research, including total level of effort and any resource/cost sharing if relevant. This need not be a detailed cost estimate but rather provide a high level summary of the resources needed.
(8) Contractor's type of business, selected from among the following categories: academic, industrial, non-profit, government, national laboratory.

Challenge Problem Selection Criteria

Challenge Problems will be selected based on fundamental issues such as

- Is the problem being addressed important to the Intelligence Community?
- Is the workshop environment the best way to address the problem?
- Does the proposal have the right leadership and team?
- Is challenge sufficiently provocative to attract the best talent and provide impact on government needs?
- Does the Challenge Problem have an Intelligence Community customer/champion or a plausible approach to engage one?
- Will the activity result in a successfully transitioned capability to analysts?
- Is the proposal aligned with ARDA thrust and Intelligence Community roadmaps?

Successful proposals will include an Intelligence Community champion(s) (proactive individuals who carry results forward to customers), compelling, accurate problem statements, a clear definition of success (impact to the IC or a specific operation), team and center commitment and buy-in, and synergy with Regional Research Center strengths. Proposals will be SMART, that is Stretch (demonstrate a substantial advance in the state of the art that will leave the field permanently changed), Measurable (be focused on clearly defined measurables), Aligned (with IC Challenge problems), Realistic (a practical plan achievable with due diligence execution practical), and Timely (a clear and practical schedule).

Specifically, the following criteria will be applied to select among competing Challenge Problem proposals (relative weighting of criteria is indicated parenthetically after each criterion):

- **Team –** The quality, experience, and skill of the workshop participants and lead(s). Included is participant commitment measured in the agreement to personally attend the entire workshop and (possibly) the contribution of resources (30% of overall score);

- **Technical Quality –** The feasibility (including achievability within the given time frame), innovation, and evaluability of the proposed effort (40% of overall score);
- **Expected Impact and Technology Transfer –** alignment with ARDA thrust and IC roadmaps; identified government champion; likelihood for transition to ARDA sponsors. (30% of overall score)

Awards

We expect to support approximately four large Challenge Problems via this call at approximately the level of $1,000,000 to $1,500,000. Seedling workshops will be funded at a level from between $100,000 to $500,000.

Due Date

Proposals are due by 5:00 p.m. EST, 13 October 2004 via electronic mail to Dr. David Day (day@mitre.org). However, we encourage informing Dr. Day and the relevant regional research center directors of your intent to submit along with your topic of interest as early as possible.

Graduate Students

Following Challenge Problem selection, the NRRC will run a competition to identify extraordinary graduate students to participate in the Challenge Workshops. Applications for graduate students are on line at http://nrrc.mitre.org/.

NRRC, NWRRC and SRRC

ARDA focuses on revolutionary not evolutionary advances in information technology for intelligence community challenge problems. It aims to achieve well-defined goals with measurable results based on sound scientific methodology. The NRRC, NWRRC, and SRRC are essential elements of ARDA's Exploratory Program. The regional research centers focus on the reinforcement of ARDA thrusts by targeting scientific results that have a positive impact on Intelligence Community (IC) problems, engaging regional experts from commercial, academic, government and non-profit organizations, infusing technology into government workforce, and transferring technology to and from industry. The NRRC, NWRRC, and SRRC are sponsored by ARDA, a US Government entity which sponsors and promotes research of import to the IC which includes but is not limited to the CIA, DIA, NSA, NIMA and NRO. Results from 2002, 2003, and 2004 NRRC workshops are available on the web at http://nrrc.mitre.org.

The MITRE Corporation, Pacific Northwest National Laboratory (PNNL), and GTRI/ORNL.

As specified in Challenge Problem proposals, MITRE (www.mitre.org), PNNL (www.pnl.gov), and GTRI/ORNL (www.gtri.gatech.edu, www.ornl.gov) will provide facilities, technical assistance, as well as contractual assistance to the Challenge Problem participants.

Schedule

The schedule for the 2005 Challenge Problems is as follows.

31 August 2004:	Call for Challenge Workshop proposals issued
13 October 2004:	Proposals due (electronically mail to day@mitre.org by 5:00pm EST).
27 October 2004:	ARDA selects proposals for oral presentation to Executive Committee (EC)
8-9 November 2004:	Oral proposals presented to EC; ARDA makes final selection (location: ARDA, Baltimore MD)
15 November 2004:	Notification of selected Challenge Problems issued
15 Nov 2004-Jan 2005:	Project Planning - Final proposals (final membership, costing, infrastructure requirements)
January 2005:	Challenge Problem work begins. Multiple workshops conducted over the next 12-18 months. ARDA and the EC will conduct a mid-term review and a final review during the period of performance.

NWRRC POC:
Dr. Rich Quadrel
Director, Northwest Regional Research Center (NWRRC)
Pacific Northwest National Laboratory
P.O. Box 999, MS K7-10
Richland, WA 99352
Email: Rich.Quadrel@pnl.gov
Tel: 509-375-5933

NRRC POCs:
Dr. Mark Maybury
Executive Director, Northeast Regional Research Center (NRRC)
The MITRE Corporation
Bedford, MA 01730
Email: maybury@mitre.org
Tel: (781) 271-7230

Ms. Penny Chase
Deputy Director, NRRC
The MITRE Corporation
Bedford, MA 01730
Email: pc@mitre.org
Tel: (781) 271-2113

Dr. David Day
Program Manager, NRRC
The MITRE Corporation
Bedford, MA 01730
Email: day@mitre.org
Tel: (781) 271-2854

SRRC POCs:

Dr. John Meadors
Director, Southeast Regional Research Center (SRRC)
Georgia Tech Research Institute (GTRI)
Atlanta, GA 30332-0824
Email: john.meadors@gtri.gatech.edu
Tel: (404) 894-2539

Ms. April Lewis
Deputy Director, SRRC
Oak Ridge National Laboratory
1 Bethel Valley Road
P.O. Box 2008
Oak Ridge, TN 37831
Email: lewisaa@ornl.gov
Tel: (865) 576-2045