

HR001117S0003
Synergistic Discovery and Design (SD2)
Frequently Asked Questions

As of December 12, 2016

Q42: Will Proposers day slides be made available?

A42: Yes. Proposers Day slides are available on the DARPA opportunity webpage.
<http://www.darpa.mil/work-with-us/opportunities>

Q41: How many years is the program anticipated to run?

A41: 4 years, with 3 phases (Phase 1/Phase 2: 18 months; Phase 3: 12 months)

Q40: When will program start?

A40: We anticipate a Kickoff meeting in July and an extended working PI meeting in August.

Q39: How many TA1 teams do you anticipate?

A39: We anticipate multiple teams. The number of awards will depend on the quality of the proposals received.

Q38: How many TA2 teams do you anticipate?

A38: We anticipate multiple teams. The number of awards will depend on the quality of the proposals received.

Q37: Anticipated funding was included in the Proposer's Day brief. Is this funding across teams or per team?

A37: Actual funding will depend upon the proposals received. Anticipated funding varies for each TA as follows:

TA1: We anticipate multiple teams and the funding posed is on a per team basis.

TA2: We anticipate multiple teams and the funding posed is on a per team basis.

TA3 (High throughput team): We anticipate a single team or multiple teams with a teaming agreement. The funding posed will be distributed across the entire TA3 high-throughput team.

Synergistic Discovery and Design (SD2)

Frequently Asked Questions

TA3 (Mid-low throughput team): We anticipate a single team or pairs of teams with teaming agreements. The funding posed will be distributed across each set of teams that proposes to collect data within a particular domain.

TA4: We anticipate a single team or multiple teams with a teaming agreement. The funding posed will be distributed across the entire TA4 team.

TA5: We anticipate a single team and funding will be distributed across this one team.

Q36: Will a team be able to propose other combined TA's (e.g. TA1 + TA4 or TA3 + TA4)

A36: No. We will only accept joint TA1/TA2 proposals.

Q35: We have expertise that could fit into multiple TAs, can we submit to all TAs?

A35: Yes, but please read the BAA carefully and prioritize proposal efforts based on compatibility between your capabilities and TA requirements. If you have expertise across multiple TAs, we recommend that you submit an abstract to each TA you are considering to receive specific feedback with regards to the technical approach. Each abstract must address a single TA. When submitting abstracts to multiple TAs, please include statements in each abstract that emphasize any relationship between the technical approaches.

Q34: Can proposers from the same organization submit separate proposals to TA1, TA2, TA3, TA4?

A34: Yes.

Q33: Is there a firewall for the abstract phase between TA1 and TA4?

A33: No, there is no firewall between TA1 and TA4 submissions.

Q32: Is teaming encouraged or required across the experimental and computational efforts (vertical teams)?

A32: Teaming is not required within or across technical areas. Teaming, however, is encouraged within TA3 and within TA4. One prime or multiple primes with a teaming

Synergistic Discovery and Design (SD2)

Frequently Asked Questions

agreement is envisioned for TA3 and for TA4. Teaming is encouraged for submitting proposals jointly to TA1/TA2.

Q31: I could collaborate with a number of different people and expertise. If I provide an abstract with the skills I can offer, can you provide feedback on what teaming would be most useful to the program?

A31: We cannot provide feedback on who to team with for the SD2 program. We recommend making use of the Teaming website to inform potential performers of your skill set and interest in teaming.

<https://www.schafertmd.com/darpa/i2o/sd2/teaming/tiki-index.php>

Q30: How will proposal teams be formed?

A30: Teams must be formed prior to proposal submission. Anticipated teaming should be mentioned in abstracts. We anticipate teams submit joint TA3 solutions and joint TA4 solutions. Teams across technical areas will not be accepted.

We recommend making use of the Teaming website to inform potential performers of your skill set and interest in teaming.

<https://www.schafertmd.com/darpa/i2o/sd2/teaming/tiki-index.php>

Q29: Can you provide more guidance on the types of challenge problems anticipated?

A29: To help form your proposal, the BAA includes examples of challenge problems in Section VIII.F. We are interested in real world design challenges that today's scientific community does not yet know how to solve. Challenge problems must come from domains that lack complete models (e.g. synthetic biology, neuro-computation, and polymer chemistry), and they must not involve domains in which researchers rely heavily on simulations (e.g., integrated circuit design, aeronautics design, or vehicle design). It is desired that challenge problems satisfy the following guidelines:

- lead to discovery and design hypotheses that can be experimentally tested in days, weeks, or months,
- do not pose IP or data sharing restrictions, and
- explore topics relevant to DoD interests, such as reproducibility and robustness.

Q28: Should TA1-TA4 proposals be problem-domain agnostic in anticipation of having problems defined by TA5?

A28: We anticipate leveraging domain expertise from performers in TA1-3 to generate compelling challenge problems that will drive algorithm development. TA5 will facilitate collection, refinement, and prioritization of challenge problems. TA1-TA3 proposals should provide examples of the types of challenges that their approach can address. We expect TA3 proposals to include substantial detail about the types of challenge problems that can be addressed using TA3 experimental techniques.

Q27: How much of a role will TA5 play in defining challenge problems? Will performers in other technical areas have input as to the selection of challenge problems?

A27: TA5 will elicit knowledge from domain experts across the SD2 program to collect, refine, and prioritize compelling challenge problems for TA3 datasets. They will also work with government representatives to develop and refine challenge problems, particularly for government provided datasets. Final selection of challenge problems will be made in coordination with the program manager and SD2 SETA team. We anticipate that the challenge problems will come from number of sources. The data provided by TA3 will form the foundation for certain challenges. As a result, TA3 should define a breadth of challenges that might be addressed with data they can provide. The development of TA2 design capabilities will require a complimentary set challenges that exploit TA3 experimental capabilities. TA5 will facilitate the collection of information from all sources and help with resource allocation should demand exceed experimental resources.

Q26: How do TA1 and TA2 know what kind of data will be available and what types of systems to model when writing the proposal?

A26: We anticipate TA3 performers will provide high-throughput omics data collection from experiments that involve single celled organisms. TA1/TA2 proposals that specify an alternative domain are in scope, and such proposals should include a notional TA3 performer that can provide data within the alternative domain.

Q25: Can data from outside the program be used extensively (if made available to the shared site)?

Synergistic Discovery and Design (SD2)

Frequently Asked Questions

A25: Data from outside sources is of interest as long as (1) there are no restrictions to sharing the data with the broader scientific community and (2) there is a mechanism to obtain experimental updates in order to test hypotheses, validate models generated by TA1 algorithms, and/or evaluate designs generated by TA2 algorithms. Please refer to the TA3 section of the BAA (page 13) that discusses the types of data we envision for the SD2 program.

Q24: Do you expect the algorithms and techniques developed by TA1/TA2, to be validated and tested in domains other than biology, neuroscience, and polymer science?

A24: We are excited about a variety of domains and encourage additional application domains for which high fidelity models do not exist. If you have questions about the applicability of alternate domains, please submit an abstract for specific feedback.

Q23: Do proposals to TA1 require expertise in biology?

A23: Expertise in biology is not required. Depending on the nature of the TA1 technical approach, it may be recommended. TA1 algorithms will be demonstrated in a variety of domains, and we expect application domains other than biology to be accepted. We invite you to submit an abstract to receive feedback on the proposed application domain. We request that domains other than biology point to a potential TA3 source for data collection.

Please note: Human subjects research is out of scope for SD2. Experiments must be completed in days, weeks, or possibly months. Application domains that already have reasonably accurate simulation models are not in scope. Data must be generated experimentally. Any data generated via simulations are not in scope.

Q22: What is the anticipated skill set of end users for TA1 algorithms?

A22: We are agnostic as to which skill sets are required to apply TA1 techniques. The goal of SD2 is to accelerate data-driven model discovery and design in domains that lack complete models. TA1 techniques may require data scientist and domain expert intervention. However, the overall goal of TA1 is to develop computational techniques that reduce the amount of human intervention required to make novel discoveries or accelerate the pace of novel discovery.

Q21: Based on the expertise requirements (CS, data science), TA1 seems to be focused on descriptive statistics/statistical learning models not grounded in fundamental biophysics/biochemistry. Is that out of scope? Or are purely empirical models the focus?

A21: Within TA1, we anticipate receiving proposals for a breadth of complementary TA1 approaches. We view domain-agnostic approaches, such as neural networks, to be complementary with knowledge-based approaches that are grounded in biophysics and/or biochemistry. Throughout the program, we will be interested in studying which techniques excel within various challenge problems. We discourage methods that require humans to extract excessive amounts of information from literature via manual processes. We do anticipate that certain TA1 techniques will embed a spectrum of domain knowledge into the algorithms used to generate scientific models. We hope to empirically test approaches throughout the course of the program via TA3 experiments. If you have questions about a specific approach please submit an abstract.

Q20: How quickly do TA1 algorithms need to scale in the program if we have a new algorithm that still needs to be proven?

A20: For novel methods, we would like to see demonstrations that the algorithms provide value on small datasets before teams invest resources in scaling the algorithms. Proposals should identify schedules and milestones commensurate with the technical readiness level of their approach. Proposals should clearly describe the approach, provide evidence regarding the amount of research that has been conducted on the method to date, and provide a technical plan for scaling the algorithm should initial demonstrations prove to be compelling.

Cost must be commensurate with the technical readiness of each approach. We expect more investment in software development for algorithms that have been tested and validated. We expect a small investment in untested approaches to determine whether the technique yields valuable results prior to making significant investments into software development and scalability.

Q19: Are you looking for empirical models?

A19: We want useful models. Model utility will be evaluated in one of two ways. Models should generate accurate hypotheses that can be validated experimentally. Models should also accelerate design of novel capabilities that would not be possible given today's techniques.

Q18: Does TA2 team selection depend on SME alignment with TA3/4 teams?

A18: We anticipate TA3 performers will provide high-throughput omics data collection from experimentation with single celled organisms. Proposals with a TA2 approach compatible with single celled organism experiments are in scope. Teams without biological experience can partner to provide such expertise.

Proposals to TA2 that involve an alternative application domain are also in scope as long as the proposal specifies a notional experimental laboratory that could provide relevant data. TA2 proposals that involve an alternative application domain and do not reference a notional experimental laboratory will be considered non-compliant. If the TA3 laboratory referenced indicated in the TA2 proposal is not selected, this will not affect the TA2 proposal selectability.

Q17: It seems that the domain areas of the data supplied by TA3 will constrain domain areas of TA2. Is this correct? Is it expected that domains will be biological in nature?

A17: The TA2 methods will be expected to use TA3 supplied data or government supplied data. The TA3 high-throughput domains are anticipated to involve single celled organisms, however the TA3 mid-low throughput domains are not constrained to single celled organisms (e.g. chemistry, neuroscience, physical phenomenon that lack high fidelity models).

Q16: Should TA2 approaches be domain specific to certain classes of challenge problems or be domain agnostic?

A16: TA2 algorithms for targeted design of system components should be domain specific. However, it is desirable that the TA2 experimental planning algorithms be domain agnostic. Note that TA2-only proposals must specify the intended application domain for targeted design and outline a notional design challenge within the specified domain. Please refer to the industry day slides for an example of notional challenge problems for the primary and alternative application domains. <http://www.darpa.mil/work-with-us/opportunities>.

Q15: For TA3 proposals, what is the best way to communicate cost for experimental capabilities? Please explain what is meant by “base” and “options” in the Proposer’s Day Brief.

A15: Clearly communicate the breadth and depth of experimental capabilities available. Provide a menu that outlines sets of experiments and indicates a natural breakdown of cost for those offerings. Items on the menu might include novel experiments, expensive experiments, or data collection efforts tailored to an interesting challenge problem. Please refer to the chart in the TA3 section (page 14) of the BAA for information about the desired level of detail.

Due to the dynamic nature of the program, it is impossible to predict exactly what experiments will need to be conducted throughout the SD2 program. Instead, strong proposals will describe a breadth and depth of experimental capabilities and provide a menu of experiment sets that could be achieved at comparable price points.

With regards to cost, one way to frame a proposal would be to describe a base period that focuses on core capabilities. The base would provide experimental capabilities required to successfully achieve the goals and objectives specified in the BAA. Option periods would describe additional data collection efforts and/or alternative experiments that might provide value to SD2 but are not strictly necessary to achieve program goals. Such a proposal format would allow the program manager to decide whether or not to execute options based on lessons learned throughout the program.

Q14: Are you looking for data from single cells or single celled organisms?

A14: For TA3 high-throughput teams, we are looking for data from experiments on singled celled organisms.

Q13: What volume of experiments will be run on the TA3 platform?

A13: In aggregate, we anticipate generating an average of 500GB to 1TB per day of raw data. The aggregate data volumes will be computed across all TA3 performers. In practice, the volume of experiments will depend on challenge problem requirements and the nature of TA1 / TA2 hypotheses and designs. We anticipate that some challenge problems will require TA3 performers to reproduce the same experiment in multiple geospatially distributed experimental laboratories.

Q12: Will SD2 invest in significant infrastructure for TA3 capabilities?

A12: No, additional TA3 infrastructure is out of scope for TA3. SD2 will invest in the consumable resources required to run experiments on existing TA3 infrastructure.

Q11: Is it ok to satisfy experimental capabilities by sub-contracting a 3rd party vendor?

A11: Subcontracting a 3rd party vendor is ok. However, we recommend submitting an abstract that specifies the types of experimental capabilities that the prime and sub-contractors can provide in order to receive feedback on the combined capabilities.

Q10: For TA3 capabilities, are you interested in one single technique with greater depth or a complete pipeline of techniques?

A10: For high-throughput proposals, we are interested in a comprehensive solution that provides both breadth and depth. If your platform is applicable for multiple domains or types of experiments, please indicate this. For low-mid throughput proposals, we are interested in novel experimental capabilities that could support compelling challenge problems.

Q9: What types of data conventions and schemas are you interested in?

A9: We would like data stored in such a way to facilitate easy computation. We encourage techniques that enable new data types to be loaded rapidly. We discourage brittle techniques or approaches that require significant reprogramming every time data types change or a new type of data is introduced. Throughout the program, we anticipate that SD2 performers will converge on data formatting convention that are useful for computation. Existing schemas that already support flexible computation are in scope. Proposals that rely heavily on inflexible, fixed schema will be viewed unfavorably.

Q8: Will performers that provide shared resources (e.g. TA4) be allowed/expected to assert user agreements with other collaborators, like expecting user indemnifications, warranties, release of IP rights not part of the program, etc? Preferably the government will eliminate the need for these types of team-to-team legalities.

A8: To promote collaborative sharing of information, each contract will include an Associate Contractor Agreement clause. The Associate Contractor Agreement clause will address the data sharing requirements between teams needed for the SD2 program. However,

each performer will need to determine whatever legal protections are necessary for their individual companies when sharing data with other SD2 performers.

We are very interested in solutions that align with open source paradigms. We will look favorably upon approaches that are compatible with open source/architecture regimes. We anticipate having open-source releases of the software and data throughout the course of the program.

Q7: Can you speak to the importance of visualization and user interfaces for collaboration as they relate to the TA4 environment?

A7: We would like TA4 to rapidly set up a core set of visualization tools that support trend analysis and aggregate statistics on newly ingested data. Throughout the program, TA4 will provide tools that support data visualization and exploration of algorithmic results. We expect TA4 proposals to outline a baseline set of visualization capabilities and identify any novel extensions that would facilitate collaborative research. The visualization capabilities selected for implementation / installation will depend on performer needs demonstrated throughout the program.

Q6: Are you envisioning significant development for the Data and Analysis Hub?

A6: We are expecting TA4 performers to leverage commercial software and/or open source tools and develop novel extensions when required. To be successful, the SD2 program will require core capabilities to be operational prior to the kick-off. These capabilities will likely require use of tools and techniques that are readily available. Slide 19 in the Proposer's Day brief includes a figure that specifies the range of novel versus off the shelf technologies envisioned for SD2. Strong proposals will outline a vision that supports new ways to support data-driven science. Strong proposals will prioritize development of novel techniques based on appraisals of how crucial such extensions are to the SD2 performer base and how much impact novel techniques might have on collaborative data-driven science.

Q5: Does the Data and Analysis Hub need to function at the TB scale right away?

A5: The Data and Analysis Hub should minimally operate at the TB scale by the program kick-off. The Hub can expand to PB scale at a rate commensurate with experimental data collection.

Q4: How much domain knowledge is required for TA4?

A4: Domain knowledge will be helpful for data ingest. Otherwise, we are not requiring TA4 performers to have specific domain knowledge.

Q3: Both TA4/TA5 are gateways to other TAs and APIs for ingest and retrieval will be critical to designs for other TAs to succeed. Is there any guidance/expectations on what those APIs/schemas might look like? How soon is it envisioned that notional APIs/schemas might become available? This is also a consideration for all interfaces between TAs.

A3: The program will not have a centralized system integrator. TA5 will not provide APIs. Conversations about APIs will commence at the kick-off and continue at the first working PI meeting. Throughout, TA2, TA3, and TA4 performers are encouraged to rapidly prototype APIs and regularly collect feedback from performers who use the APIs. The TA5 team will ensure that key API decisions are made at working meetings.

Q2: If you are in TA5 and have your own AWS server, will that be funded through the program?

A2: No.

Q1: What will TA5 be responsible for versus the SD2 SETA team?

A1: There will be a natural synergy between TA5 performers and the SD2 SETA team. TA5 performers will coordinate program wide evaluation. TA5 performers must elicit knowledge from domain experts across the SD2 program to establish quarterly challenge problems that drive the development of tools for data-driven design in domains that lack models. TA5 performers will work with government representatives each quarter to finalize the list of challenge problems and validation experiments. In addition, TA5 performers will act as an evaluator to assess whether SD2 methods support discovery and design beyond what is otherwise possible today. TA5 performers are also responsible for organizing the technical agenda for working PI meetings in coordination with government representatives.