XDATA Questions and Answers
(As of April 30, 2012)

Q107. When will the common data be available?
A107. The common data are anticipated to be available by the program kick-off (anticipated October 1, 2012) and will be provided as part of the XDATA source repository and technology integration facility (see the "Government-furnished Resources" section in the BAA).

Q106. How will the common data be distributed? How will it be made available at the hosted test facility?
A106. Data will be resident as part of common computer framework, asymmetric resource cloud environment, at the XDATA technology integration facility and will be accessible via the facility or secure VPN.

Q105. Can you provide an example of Big Data?
A105. No.

Q104. Can you give an example of the areas of application and data that will be used?
A104. For the purposes of the program, we will maintain an unclassified data repository. Refer to the BAA -- http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx -- for other details in terms of structured and semi-structured data.

Q103. Are you considering data fusion?
A103. Yes, in the sense that applications are heterogeneous in terms of data types and volume.

Q102. Who is the end user?
A102. End users are the services -- Army, Navy, Air Force, and Marines -- as well as other agencies in the intelligence community and other government entities.

Q101. What will the data look like?
A101. It will look like bytes. Details in BAA -- http://www.darpa.mil/Opportunities/Solicitations/I2O_Solicitations.aspx.

Q100. How much funding is expected per area?
A100. The funding per area will be determined based on the quality of the proposals.

Q99. Should we make assumptions on the government cloud software stack? Will it use HDFS, Accumulo or other common packages?
A99. For proposal purposes the answer is no. At program kick-off, as part of the software and hardware procurement, we will have identified a baseline set of software

and a baseline hardware stack. The kick-off meeting and subsequent PI meetings will include technical exchanges about the strengths and weaknesses of the baseline system.

Q98. Please provide more details on types of the use cases and data types available for testing.

A98. This will be provided at the program kick-off. For now, consider a variety of examples including combinations of spatial-temporal data such as sensor feeds, structured meta-data containing some unstructured text fields, network oriented information, as well as in general just processing different types of information at large scale.

Q97. Is the analysis for wide-area motion imagery data analogs for tracking targets and detection of suspicious activity patterns relevant for XDATA?

A97. We are interested in core capability for evaluating patterns of activity, anomalous behavior, etc. to understand and to discover information in scientifically-collected data, including sensor feeds, text documents, and meta-data. However, core image processing algorithms, core video processing, and speech processing are discouraged as out of scope.

Q96. Is there any data to be provided to TA1 and TA2?

A96. Yes. See the BAA.

Q95. What do you see as the biggest data visualization challenges within XDATA?

A95. The primary challenge is to enable a variety of users to easily explore and understand data using principles of analytical reasoning. A major challenge is the framework itself and how to maintain an extensible open source library. The scale of data is also a challenge. The visualizations and options for interacting with data should be a function of data volume, type of question, time constraints, etc.

Q94. What kind of feedback will submitters receive from an abstract?

A94. Feedback will be provided with the intent of encouraging and discouraging proposals. This will consist of advice on certain aspects from the abstract that should be focused on or an explanation as to why a proposal is not recommended (e.g., the research proposed is out of scope).

Q93. Will LIDAR, HSI/MSI, EO, IR, videos collected by military sensors be considered by the XDATA program?

A93. Meta-data collected with the sensor platforms are within scope and will be considered. Raw sensor information or signal processing algorithms on those raw sensors are not and will not be considered.

Q92. What are the open source data planned for the program?

A92. This has not yet been determined.

Q91. Are all open source data captured from the web?

A91. No.

Q90. Will real-time or transient cyber interactions need to be captured and analyzed in the XDATA program?

A90. Network oriented interactions are within scope.

Q89. Can a white paper address a part of a technical area, not the whole technical area?

A89. Yes. The same holds true for proposals. Small proposals (e.g., those in the range of $100K to $250K) that only address sub components of technical areas are acceptable. Medium to larger size teams that want to address an entire Technical Area, based on technical expertise and emphasis, are also encouraged.

Q88. Will there be an emphasis on any types of data for the program, e.g., financial network data and its analysis and exploitation?

A88. A variety of types of data are available as there will be a variety of government users. Financial network information is within scope, as are communications, sensor information, and meta-data with associated text documents.

Q87. Will the users participating in the technology integration facility be considered "human use" testing?

A87. For Technical Area 4, yes.

Q86. Can you break out the $25 million budget? How much to performer staff? How much toward industry? TA1 thru TA4?

A86. No. The figure of $25 million per year for four years was part of the White House press release on the Big Data initiative. That amount may or may not be available under this BAA. Awards under this BAA are subject to the quality of the proposals received.

Q85. Does the cover sheet count as one of the four pages in the abstract?

A85. No.

Q84. The BAA states a 54-month duration, but the Proposers' Day presentation stated that this should have a quick turnaround time. Could we get more clarification on this?

A84. Quick turnaround refers to the ability for users and computer scientists to develop new analytics and interfaces. The program itself is anticipated to have a 54-month duration as described in the BAA.

Q83. Could you tell us what type of funding is behind XDATA, 6.1 – 6.4?

A83. At this time, DARPA anticipates 6.2 funding.

Q82. Can we submit multiple abstracts on a TA?

A82. Yes.

Q81. How many teams do you expect to fund for TAs 1 and 2?

A81. Per the BAA, multiple awards are anticipated.

Q80. Will TA1 performers be paired with TA2 performers who will develop UI components with their algorithms in mind, or should TA1 performers partner with a TA2 performer to ensure that their algorithms can be driven via a graphical interface?

A80. It will depend on the proposals. TA1 and TA2 performers' works will be closely related. The pairing will be done once the program starts. As the BAA states, all performers will be expected to work cooperatively with one another to develop, integrate, implement, test and evaluate XDATA capabilities.

Q79. Are you committed to Map Reduce vs. other cloud-based frameworks?

A79. No. MapReduce is an excellent mechanism for indexing large amounts of text documents, but has not been sufficiently tested for scientific discovery, for iterating over data, etc. Other cloud-based frameworks are encouraged as are proposals that address asymmetric hardware including standard cloud racks, GPU racks, large RAM machines, etc. However, MapReduce and its cousins will likely be the baseline because those are best practice.

Q78. Can you give specific CONOP/scenario for data exploitation that illustrates uncertainty propagation issues for TA1?

A78. Yes. For example, we query the results from text search or an area of interest that go across multiple data sources and timeframes, and we would like both results and error bars that are a function of information processing steps used to quantify or quantize the information.

Q77. What are the expected roles of university teams vs. industrial?

A77. It will depend on the merit of the proposals.

Q76. Do you see TA 1 more as 6.1 or 6.2?

A76. At this time, the funding is expected to be 6.2. We expect a set of functional implementations; however we also expect to address certain core challenges per the BAA.

Q75. Do you prefer to see a large number of small teams or a small number of large teams?

A75. Neither. It will depend on the quality of the proposals.

Q74. For TA1, would it make sense to propose for a data type, e.g., imagery, text, full motion video?

A74. It depends on the quality of the proposal. Clear and unique expertise in a particular data type and limited in scope and small in size, then yes that's appropriate. However, proposals that focus on core DSP, e.g., image processing, are discouraged.

Q73. What is the thinking behind separation of TA1 and TA2 and having proposals target only one? In particular, there is a lot of discussion around query optimization and formulation in TA2 that might be applicable to TA1. How should this be called out in the proposal?

A73. This is primarily a function of implementation and convenience. While addressing one TA, you can highlight the aspects of the other TA.

Q72. Will use cases be available before awards? If so, they could help focus research direction.

A72. No, not beyond what we are discussing here, but feedback will be provided in the abstract phase.

Q71. Do TA2 tools support all interaction aspects, e.g., query and visualization, or are visualization-only solutions acceptable?

A71. Proposals should be in accordance with strength in the area of work. Domain-specific query languages, human computer interaction, techniques that adapt to users' behavior are valid as are visualization-only solutions that address screen layout, animation, etc.

Q70. Regarding TA1 and TA2, can we use commercial cloud computing for experiments/ development?

A70. The entire framework is to be open source. If experimentation on existing software enables development of better open source software, then yes. However, performers will be measured on the output of the software and its integration across TAs.

Q69. You reference "cloud implementation experience" for TA3. Do you anticipate the need for a new cloud environment to be set-up to support the work being done for TA1 and TA2? If so, will this cloud be built by DARPA with government furnished equipment, or will you consider building and managing of this cloud by the TA3 awardee?

A69. DARPA will furnish the cloud for the testing purposes. We will not fund the construction of a new cloud facility.

Q68. What is the level of analysis that is in scope for TA2? Specifically, would models for adaptive displays be part of TA2, or should TA2 vs. TA1 algorithms drive such adaption?

A68. Yes, adaptive displays are within the scope for TA2. Relationships to information processing techniques as part of TA1 are also in scope.

Q67. TA4 is developing (evaluation) principles on the kinds of interfaces that address different problems. Do you see TA1 participating in that process?

A67. Yes, the TA4 evaluator will have a close relationship with TA1 and TA2.

Q66. Are streaming data expected to be at a low enough rate to be processed by the CPU of a standard machine?

A66. Yes. For the time being we will not consider high performance computing specialized processors, other than to the extent that they are part of a non-uniform cloud that may have some components like GPUs.

Q65. What technology-readiness levels (TRLs) are you looking for TA1? For TA2?

A65. We expect a progression from TRL 2 and TRL 3 to beginning to higher levels as components and systems mature.

Q64. Does XDATA aspire to advance a broad range of big-data analytics or advance a specific kind of analytic on big data?

A64. Primary aim is to advance a flexible framework to allow rapid development and adaption of analytics and interfaces to address a variety of problems, including specific analytics as appropriate, as well as the range and function of asymmetric resources.

Q63. Are meta-tools, e.g., that select which among alternative algorithms to use for given data or a given problem, better binned in TA1 or TA2?

A63. It will be a function of the proposal (e.g., if it's a part of the interface, then TA2; otherwise, TA1).

Q62. The user interface design in TA2 and computational methods/implementation in TA1 have the potential to greatly influence each other. What ways will these teams communicate needs other than through the TA3 team?

A62. As mentioned before, communication across TAs will occur through the XDATA technology integration facility, through summer workshops, through common architecture, through common core repositories, and user feedback and evaluation.

Q61. To what extent does open source correspond to "free" here?

A61. Open source means licensing for software will be determined, equivalent to BSD or Apache, where commercialization of derivative products is allowed. HCI testing is expected to be done at the XDATA technology integration facility.

Q60. Is the intent to select more complimentary or more competitive performers?

A60. There is no set of "go/no go" or phase metrics. The intent is to get the best set of performers to work in the program. Consider competitiveness as best people, and "complementary-ness" as team building.

Q59. Large data is often used in connection with problem-specific models which drive queries. Who will define such models in the program?

A59. Users, available data, and proxy problems that get established will help drive some of that design process, as will best practice and what's possible in terms of implementation for certain algorithms.

Q58. The scope of TA1 and TA2 interest areas are relatively general and generic. Are there any specific operational challenges that can be made available to the potential proposers?

A58. Yes, there are example problems of interest. We will also define proxy problems across open and unclassified data, as well as provide user feedback from government participants. Those in TA3 and TA4 may have access to classified information and may have the opportunity to provide implementation operational cases.

Q57. Are you specifically interested in a particular kind of analysis, e.g., graph analysis, spatio-temporal analysis, image analysis, and video analytics?

A57. Yes, this is part of the challenge of extensibility. However, we are not interested in areas that focus on a particular type of signal, e.g., speech processing, video processing. We are interested in building a toolkit that can do time-sensitive processing, spatially-sensitive processing, etc.

Q56. For TA1 and TA2, for academic partners, will a post-doc satisfy requirements for the 8-10 week camp with shorter visits from professors?

A56. Yes, for the 8 -10 week summer workshops, a participant per proposer is likely to be sufficient to qualify (see the "Government-furnished Resources" section in the BAA). To the degree that the best people can be provided, it will make the experience more enjoyable and successful.

Q55. Are there any specific parts of the open source stack to work with, e.g., Hadrop, HBase, Accumulo, etc.?

A55. Yes, there will be a baseline software stack.

Q54. What scale do you see evaluation taking place at, 100s/1000s of machines? Terabytes or petabytes of data?

A54. Initially, 100s of machines, terabytes of data; later stages will have 1000s of machines, petabytes of data.

Q53. How many awards to you expect across the technical areas?

A53. It will depend on the quality of the proposals.

Q52. Are you looking for new algorithms/frameworks or integration of existing approaches that scales? If both, which do you believe is more important/relevant/needed?

A52. Both. For some performers, we may expect them to do core implementation for baseline algorithms in analytics and visualization.

Q51. How do we ensure there is relevant data available?

A51. Data will be furnished as part of the government supplied resources; however, proposals that suggest available data are also welcome.

Q50. Is the XDATA technology integration facility for integration and testing only or also for development by performers of TA1 and TA2?

A50. Per the BAA, the technology integration facility is for all performers and is to facilitate agile and collaborative software development, integration, and testing/evaluation. User

interaction, use-case development, and integration, test, and evaluation are intended to take place at this facility.

Q49. Are there existing programs/systems that you see as a transition target, e.g., DCGS/DIB?

A49. Yes, we plan to transition the capabilities and technologies to the services and other government enterprises.

Q48. Please elaborate on your open-source vision. Do you mean public open-source or can it include open APIs, but a proprietary platform with government purpose rights?

A48. It depends on the proposal. Proprietary platforms with APIs will be considered in exceptional circumstances; however, in order to facilitate transition and use across enterprise platform for the government, unlimited rights and public open source is strongly encouraged.

Q47. For fundamental research, can deliverables be reports or must there always be software deliverables?

A47. There will be a variety of deliverables to include reports and software and will depend on the type of award instrument. See the BAA.

Q46. Approximately how many projects would you consider desirable for TA1 and TA2?

A46. It depends on the quality of the proposals.

Q45. Is all output from TA1 and TA2 expected to be open source? This appears to be released as above and beyond government use rights.

A45. Yes, all output from TA1 and TA2 is expected to be open source.

Q44. What was the previous research performed in this area? Who performed the "seedling" effort and what were the outcomes?

A44. No additional information is given at this time.

Q43. What is DARPA's hard problem that DARPA is trying to solve with this BAA?

A43. The DARPA hard problems are stated in the BAA.

Q42. Can DARPA share information about transition partners?

A42. No.

Q41. Can DARPA share the SETA contractors that will support this effort?

A41. No.

Q40. What type end product is DARPA seeking?

A40. DARPA is seeking software, new ideas, and new innovations.

Q39. The BAA specifies a final option period of 6 months. Do you have a special purpose in mind for this short period?

A39. We don't have anything specific in mind. It will be used as the end of the transition phase and for closing out the program.

Q38. Can options be included in the proposal?

A38. Yes.

Q37. Can you say more about the summer integration events?

A37. The model for the summer integration events to be held at the XDATA technology integration facility is taken from other workshops that government, academia and industry have put on. Core challenges will be identified in the fall by PIs, and then focused work will take place in the summer with everyone around.

Q36. Given that the resultant software will be open source, will DARPA host the source repository, issue tracking system, mailing lists, etc., or will this be developed in an existing open source community?

A36. Yes, DARPA will host all of the above during the program; proposals that suggest open source frameworks are encouraged.

Q35. Can an organization submit multiple separate projects as separate proposals under a single TA?

A35. Yes.

Q34. What is the value of the abstract in the selection process?

A34. Abstracts have no value in the selection process. Their value is in the potential to reduce effort in proposal preparation.

Q33. How will the abstract be compared to the final proposal?

A33. Abstracts will not be compared to final proposals.

Q32. Does the subcontractor proposal count against the 20 page limit?

A32. Yes, subcontractor Statements of Work count against the 20 page limit for the technical volume.

Q31. Is the subcontractor proposal part of the cost?

A31. Subcontractor Statements of Work are not part of the cost volume however, subcontractor cost proposals are.

Q30. What role do you see for NIH?

A30. The XDATA program is part of the White House's Big Data initiative, and NIH is participating through a NSF solicitation. We expect interaction between DARPA and NSF in the form of reviews and feedback.

Q29. Will copies of the briefing materials be available?

A29. Yes, on the program's site -- http://www.darpa.mil/Our_Work/I2O/Programs/XDATA.aspx.

Q28. Will TS-SCI require polygraphs?

A28. TS-SCI clearances do not require polygraphs.

Q27. What is the relationship of XDATA to Nexus 7 and GUARD DOG?

A27. There is no anticipated relationship. XDATA is an independent effort with separate funding, BAA, proposals, etc.

Q26. Who are the intended users of the system, e.g., scientists, military personnel, all of the above?

A26. All of the above.

Q25. Will the abstracts and proposals be reviewed by the same people?

A25. Possibly. Abstract reviewers may also participate in the proposal scientific review process.

Q24. Will the whole program be unclassified or will there be an avenue for classified work in TA1 or TA2?

A24. The entire XDATA program will be unclassified. Classified work will be considered, especially for TA3 and TA4 performers with relationship to T1 and T2 performers.

Q23. When will responses to the abstracts be expected?

A23. We anticipate providing responses and feedback to abstracts within 10-15 days.

Q22. Is the full proposal only contingent upon acceptance of the abstract?

A22. No, proposals are accepted regardless of feedback.

Q21. Is there a minimum/maximum number of PIs (team members) for TA1?

A21. No, it depends on the quality of the proposal.

Q20. What is the minimum scope (cost-wise) of a proposal that will be entertained from a university team?

A20. There is no minimum; it depends on the proposal.

Q19. Do you have or are you looking for someone to set up and run the XDATA technology integration facility?

A19. No. We expect TA3, in coordination with TA4 and TA1 performers, to manage the library but not the facility.

Q18. How does TA3 differ from co-design?

A18. TA 3 performers may propose a co-design approach, or alternatives for enabling agile software development, oversight, and testing with user feedback.

Q17. Will some non-public datasets be accepted for evaluation purposes?

A17. It will be considered if it's part of the proposal.

Q16. In which area of innovation is expected? Software/tools/interfaces? Data analytics algorithms? High-performance computing? All of the above?

A16. All of the above. However, not high performance computing in the sense of a Cray computer, but yes in the sense of distributed computing platforms.

Q15. Is there an upper limit on the amount of funds proposers can request for? Or does it entirely depend on the proposer's assessment of the proposed work?

A15. There is no limit on the amount of funds proposed for, but funding will depend on the quality of the proposal and tasks, and cost realism.

Q14. Is the focus more on new methods or providing scalable analytics for existing methods on big data?

A14. Both, depending on what the method is.

Q13. Are there some application domains more preferred than others?

A13. It will depend on the proposal and expertise.

Q12. How does cost sharing work and who approaches who?

A12. Cost sharing is not a requirement unless you are proposing to do an OTA (Other Transaction Agreement). See Section IV.C.2.e of the BAA for further information.

Q11. Will consideration be given for university proposers used to grant mechanism?

A11. Procurement contracts are used to buy or procure a supply or service and if you are supplying something (deliverable), it will generally be a procurement contract. However, final award instrument determination will be made by the contracting officer.

Q10. What will the abstract feedback look like?

A10. The abstrack feedback will be a written statement provided via email.

Q9. Are the abstracts held confidential?

A9. Per the BAA, abstracts are protected as source selection information.

Q8. Is part of TA2 helping figure out and selecting the right tool?

A8. Yes.

Q7. One of the objectives is to allow users quickly to develop and deploy new analytics. Are we expected to bring an example of our own analytics to demonstrate it, or will DARPA provide some sample analytic methods as part of the proxy questions to be posed to the system? If we are going to build a framework for analytics, where will the specific analytics that are testing on that framework come from?

A7. Yes, you should propose your own analytics/metrics and the way to extend them to the environment that we are trying to operate in, but understand that there will be proxy problems with data much like the statistics, economics, or the machine learning communities currently operate in.

Q6. We talked about extending methods and getting them to work in a scalable framework in a distributed fashion. Are brand new methods or patterns in evolving data or streaming data within scope?

A6. Yes.

Q5. Is it OK to propose methods that rely on web APIs that we may have paid to get high volume?

A5. Proposals that require proprietary licensing or costing that are not part of an open source public infrastructure will be considered in exceptional circumstances. However, the focus should be in the development of algorithms and implementation.

Q4. In TA2 area of the solicitation, you talk about not wanting purely widgets. Can you elaborate?

A4. We are trying to avoid highly specialized, brittle, small solutions. We would like a variety of analytic and visual techniques and to be able to construct them quickly.

Q3. Is there any aspect of the overall program that will deal with the dissemination of the open source software beyond the performer team?

A3. We expect this to be truly open source, such as a BSD license, and that the integration facility will maintain a repository that at some point, through the transition to other government partners, will be made available to the wider public. In particular, any organization that is involved in the process will have a copy to use and distribute as they please.

Q2. We are a company that is trying to make code base open source but the U.S. government is prohibiting us from doing that. It is in the area of high performance computing. Will this be an issue with this solicitation? If awarded, will DARPA help us make those solutions open source in the purest sense?

A2. It will be addressed on a case by case basis.

Q1. Is there going to be a human computer based lab environment made available for the center or is that something that should be proposed under TA3?

A1. This will be done through the XDATA technology integration facility primarily, but the Technical Area (TA) 4 evaluator may have some of the equipment in their own facility, in which case this will not be replicated by TA3 performers.