This is a Sources Sought Notice. This is NOT a solicitation for proposals, proposal abstracts, or quotations. The Department of Defense (DoD), Washington Headquarters Services (WHS), Acquisition Directorate (AD) is issuing this notice seeking sources capable of providing a Master Data Repository tool that meets the requirements below. DTIC’s goal is to consolidate, unify, manage, control, search, analyze, and disseminate scientific and technical data using a single tool. Consolidation of the data into a single tool will ease and facilitate analysis, management, and control. A single tool is desired to replace the numerous legacy technologies; reduce the number of required skill sets; reduce risk by eliminating many processes, interfaces, and points of integration; and gain efficiencies by applying policies, features, validation rules, and data quality improvements once rather than in multiple data flows. The tool must also provide better functionality than some of DTIC’s existing technologies, namely:

• Storage that can quickly ingest diverse structured and unstructured data without time-consuming preliminary data model standardization.

• Access control, integrated with the data store that will ensure the security of its information but can be more responsive to changing policies and new data sets.

• Search that supports simple and complex queries across structured and unstructured data, delivers fast, accurate, and precise results, can quickly respond to requests for new features, and is integrated with semantic technologies.

• Semantic capabilities for storing and querying RDF triples.

• Analytic and visualization capabilities that can be quickly implemented across minimally normalized data to offer new insights on Department of Defense science and technology research.

• Input systems that can be easily spun up for new data calls and whose data can be integrated, searched, controlled, and analyzed in the same system.

**PRODUCT**

The Contractor shall provide a Master Data Repository (MDR) Product that possesses the following capabilities to meet DTIC’s requirements:

**Functional**

A. Data Ingest and Storage

1. Quickly ingest structured data with diverse schemas without having to establish a common data model and without losing the structure of the ingested data. Ingested data shall be immediately searchable. NoSQL solution preferred.

2. System will be able to pull data directly from a variety of databases (Oracle, MySQL, and MS SQL Server) and from data dumps in a variety of potential formats (XML, CSV, JSON).

3. Quickly ingest unstructured data (e.g. full text documents) in a variety of formats (PDF, Word, PPT, etc.) and make it immediately searchable.

4. Quickly ingest RDF triples.

5. Establish connectors to continually harvest data from repositories that must remain independent.

6. Index content from website crawls.

7. Easily modify data mappings, fields, models, and schemas, both after ingest and optionally during ingest, to enhance functionality.

8. Easily scale up to incorporate new data sets

9. Replace some of the existing databases and unstructured data repositories (file systems) by becoming the primary store of that data.

10. Serve as the database for legacy, custom-built, in-house input systems (support SQL queries and inserts).

B. Search

1. Treat the full text of a document and the metadata describing the document as a single item in search results.

2. Offer simple search of structured and unstructured data, separately and together.

3. Search specific metadata elements/fields. Easily and quickly add new search fields across diverse collections.

4. Support Boolean operators, proximity operators, phrase searching, and nesting in the general index and when limiting to specific fields.

5. Automatically search stems of words to finding singular and plural forms.

6. Support truncation and wildcard searching.

7. Search date and number ranges.

8. Enable building complex queries combining Boolean operators, multiple levels of nesting, field searching, full text searching, phrase searching and truncation using a simple search language.

9. Enable scrolling or paging through all results from a search, not limiting to only the first few thousand results.

10. Return exact counts of search results, not rough estimates, regardless of scale of results.

11. Enable sorting by relevance, date, or alphabetically by fields like Title or Organization.

12. Recommend similar results.

13. Enable browsing/selecting search terms from an ingested authority file/ontology (e.g. list of organizations).

14. Ingest a thesaurus or taxonomy to expand searches to synonyms and related terms.

15. Autocomplete/suggest search terms as the user types in a search query.

16. Enable system administrators to tune the relevance algorithm (e.g. by frequency, occurrence in specific metadata fields, etc.).

17. Offer users option to limit a search to only structured data (excluding indexes of full-text/unstructured data).

18. Allow users to save search strategies to rerun ad hoc or periodically.

19. Save and export metadata search results in a variety of formats (PDF, CSV, and BibTex).

20. Offer facets/navigators/clusters, based on specific metadata or extracted entities, for browsing results.

21. Return a sub-second response for simple search queries.

22. Enable saving of user search preferences.

23. Offers APIs for other applications to search the data

24. Search by geographic location.

25. Highlight search terms in results.

26. Store logs of user search queries.

27. Incorporate results from other search engines.

C. Access Control.

1. Apply role-based access control of search and input users at collection, document, and metadata element levels.

2. Apply role-based access control of search and input users to query, analysis, input, and edit functionality.

3. Apply role-based access control of developers’ access to structured and unstructured data by collection.

4. Apply complex access control rules based on multiple markings within the metadata such as distribution limitation, distribution reason, and classification.

5. Query LDAP for user attributes and access rights.

6. Store logs of users’ data accesses.

7. Offer APIs for other applications to inquire specific user’s access rights for specific data.

8. Control access to the data by other agencies and applications (i.e. include access control in a search API).

9. Easily modify rules to accommodate new policies and new collections.

D. Data Linkage.

1. Establish relationships/links across data sets with diverse schemas.

2. Enable exposing data gaps and inconsistencies across content and implementing strategies to mitigate those gaps.

3. Enable inferring new or missing data from relationships in diverse data sets.

4. Easily add new fields/element types to individual collections to fill data gaps.

5. Enable construction of new composite entities using specific elements from multiple data sets with diverse schemas. For example, integrate user registration data, access rights, search/access/input history, authorship, saved searches, and bibliographies into a single profile for each user.

E. Analysis and Visualization.

1. Support Analytics requests for statistical queries within collections and across linked data.

2. Respond to analytic queries across millions of metadata records in seconds.

3. Support visualization of data through charts, graphs, and maps.

4. Support third party Business Intelligence tools.

5. Store logs of analytics requests.

6. Generate metrics on use of the input, search, and analytics modules of the repository.

F. Semantic Search

1. Provide for storage of ontologies and RDF triples.

2. Offer search of RDF integrated with search of metadata and full text.

3. Support SPARQL queries.

4. Support visualization of semantic relations of data.

G. Document Management

1. Enable input, editing, and management of data by users and internal staff.

2. Ensure data is immediately exposed in search after input/edit/approval.

3. Support workflows for document management.

4. Support document versioning.

5. Allow end users to tag data or enhance data.

6. Enable schema and content validation of data feeds.

7. Enable field validation in manual input forms.

8. Enable users to select input values from values in an existing ingested taxonomy or collection (e.g. select Organization from ingested taxonomy of organizations).

9. Enable users to load batches of records (e.g. as XML, JSON, CSV).

10. Provide ACID transactions to ensure data integrity.

11. Support simultaneous read and write (Multiversion concurrency control)

H. Data Governance and Quality Improvement.

1. Enable applying universal and collection specific data governance across ingested and inputted content.

2. Enable validation of manual inputs and XML uploads.

. Enable quality analysis and improvement of ingested legacy data.

I. Publishing

1. Export metadata on demand and on periodic schedule in a number of formats (XML, HTML, Sitemap, CSV, OAI, MARC, etc.) along with its associated unstructured full text data (PDF, Word, PPT, etc.).

2. Offer APIs for trusted sources to export data on demand.

J. Application Development

1. Easily spin-up front-ends for search, analysis, or input of data for existing or new collections.

**Non-functional:**

K. License

1. Offer on premise and cloud solutions

2. Easily transition licenses between on premise and cloud solutions.

L. Security

1. Meet requirements of DoD Security Technical Implementation Guides (STIG) and DoD Risk Management Framework (RMF).

M. Availability

1. Support clustered architecture on staging and production environments for high availability and disaster recovery.

2. Provide option for database replication.

3. Provide installations on NIPR Development, Staging, and Production environments with full data sets on both Staging and Production.

4. Provide installations on SIPR Staging, and Production environments with full data sets on both Staging and Production.

N. Technical

1. Support SQL and SPARQL.

2. Offer REST APIs.

3. Support Java and JavaScript application development.

4. Run on virtual machines with network storage.

5. Run on Linux, Solaris, and Windows.

4. SUPPORT SERVICES TASKS

The Contractor shall be directly responsible for ensuring the accuracy, timeliness, and completion of all requirements under this effort, none of which are considered inherently governmental functions as defined in FAR 2.101 or Subpart 7.5. In accordance with FAR 37.104, Personal Services are prohibited and will not be performed under this contract. A key concept throughout is for the Contractor to assist in shaping the technology base for DTIC’s future. Development will occur concurrent to DTIC's agile development and testing team.

The Contractor shall perform the following tasks which include, but are not limited to:

A. Installation

1. Install, configure, and optimize the product in NIPRnet environments on premise at DTIC or in DTIC’s chosen cloud provider

2. Install, configure, and optimize the product in SIPRnet environments on premise at DTIC

B. Data Ingest and Storage

1. Ingest structured and unstructured data from DTIC’s primary collections (see Data Details below) without extensive upfront data normalization effort.

2. Ingest data from DTIC’s database of saved searches and bibliographies.

3. Ingest application and web log data.

4. Establish connectors and processes to periodically ingest updates from legacy repositories that DTIC chooses to persist outside of the MDR.

5. Ingest associated RDF triples generated from DTIC’s semantic tools.

6. Modify and transform data as needed to meet specific use cases.

C. Search

1. Implement search service for DTIC’s front-end user interfaces to call with queries and return with search results. Ensure it supports Boolean operators, proximity operators, phrase searching, and complex nesting in the general index and when limiting to specific fields.

2. Implement display service for DTIC’s frontend user interfaces to call with requests for specific metadata records and documents.

3. Implement full text searching.

4. Implement option to search only metadata (excluding the full text index).

5. Implement default option to search both full text (unstructured data) and metadata (structured data).

6. Enable field searching of each single field within each collection and for a limited set of common fields across collections.

7. Implement combined search fields that may search single fields in some databases but multiple fields in others (e.g. Authors, Organizations, Subjects, and Identifying Numbers).

8. Optimize performance for commonly searched fields (e.g. Title, Author, Organization, Report Date, Creation Date, Program Element Number, etc.).

9. Ensure that the metadata record describing a full text document, and the full text document itself, are treated as a single item in search results but as separate items for access control.

10. Implement stemming, truncation and wildcard searching.

11. Implement sort of search results by relevance, date, and alphabetically by title and organization.

12. Ingest a subject thesaurus or taxonomy to expand searches to synonyms and related terms.

13. Store logs of user search queries.

14. Implement auto-completion of search terms based on logs of previous searches.

15. Create service for web front end to store search queries, to retrieve and display saved search strategies to the user, and to rerun searches on demand or on schedule.

16. Create service for web front end to store specific user selected search results to review at a later date, add to, delete, and export in a variety of formats.

17. Implement facets/navigators/clusters of 3-5 metadata elements in search results for users to filter search results.

18. Return sub-second response for simple search queries.

19. Create web service for web front end to store, reference, and modify users’ UI preferences.

D. Access Control.

1. Implement role-based access control of search and analysis queries at collection, document, and metadata element levels. Integrate into service for search queries and display of data. Access control will be based on rules for controlling Scientific and Technical Information, markings of data in the MDR, and user attributes.

2. Implement controlling developers’ access to code and data stored in MDR.

3. Implement query of DTIC’s LDAP for user attributes and access rights.

4. Implement storage of logs of users’ requests for structured and unstructured data and whether or not access was granted.

E. Analysis and Visualization

1. Implement service for analytics requests within collections and across linked data to retrieve statistics on content, user, and log data.

2. Assist implementation of visualization of results from analytics requests through charts, graphs, and maps.

3. Implement storing logs of analytics requests.

4. Implement service to query access logs for the number of times each item in a search results list has been viewed but previous users.

F. Semantic Search

1. Implement search of RDF triples integrated with search of metadata and full text.

2. Assist with creation of SPARQL queries for new analytics requests.

G. Document Management

1. Implement service for document management staff to edit ingested data.

H. Data Governance and Quality Improvement.

1. Implement scripts to correct/improve/standardize data in select elements.

I. Publishing

1. Implement service to export metadata in XML with associated full text files for records selected through search queries.

2. Implement service to publish metadata in XML or HTML with associated full text and Sitemap indexes.

**Non-functional:**

J. Cloud

1. Assist with transitioning on premise implementations to a cloud environment.

K. Security

1. Integrate the system with DTIC’s SSO environment.

2. Assist with documenting system security plan for MDR.

L. Availability

1. Implement clustered environments.

M. Support

1. Provide annual maintenance support.

2. Provide training classes.

3. Fully document code written to fulfill above requirements.

4. Document technical architecture of implementation.

**PERIOD OF PERFORMANCE:** The anticipated period of performance for this requirement is 6 months with (2)-6-month option periods.

**LOCATION:** DTIC expects 95% of the work to be performed at the Government site: Defense Technical Information Center located at the Andrew T. McNamara Building at 875 John J. Kingman Road, Fort Belvoir, VA 22060-6218.

To this end, the WHS Acquisition Directorate is seeking the following information from interested companies:

1. Recommendations to the Draft Performance Work Statement
2. Recommended NAICS code(s)
3. Relevant Corporate Experience: Provide information that demonstrates the extent of your corporate experience performing the same or similar work during the last three (3) years
4. Demonstrated understanding of requirement
5. Unique or innovative ideas/concepts
6. Rough order of magnitude which details relevant labor categories and fully burdened labor rates during the base period and throughout the two (2) option periods
7. Potential problems, challenges or concerns during contract performance

INFORMATION SUBMISSION INSTRUCTIONS:  
The response to this notice is limited to ten (10) single sided pages including all attachments, charts, etc. (single spaced, 12-point font, excluding charts and graphics). The first page of your response shall include the following information: 1) eligible business concern's name, 2) point of contact, 3) address, 4) cage code, 5) current business size, and 6) GSA MOBIS status.

All responses to this Sources Sought Notice must be submitted electronically (via e-mail) to Cameron Earle, Contract Specialist, at cameron.j.earle.civ@mail.mil. Facsimile responses will not be accepted.

RESPONSE TIME/DATE:

Capability statements with respect to the information above are due no later than 1:00 p.m. (EST) on Tuesday, 30 December 2014. Capability statements received after this date and time may not be considered.

DISCLAIMER AND IMPORTANT NOTES:

This Sources Sought Notice is for information and planning purposes only and shall not be construed as a solicitation or as an obligation on the part of the Government. This notice does not obligate the Government to award a contract or otherwise pay for the information provided in the response. The Government reserves the right to use information provided by respondents for any purpose deemed necessary and legally appropriate. Any concern responding to this notice should ensure that its response is complete and sufficiently detailed to allow the Government to determine the concern's qualifications to perform the work.

Respondents are advised that the Government is under no obligation to acknowledge receipt of the information received or provide feedback to respondents with respect to any information submitted. Responses to this notice will not be considered adequate responses to a solicitation.

The Government will NOT accept requests for meetings or briefings. Information and materials submitted in response to this notice WILL NOT be returned.

The Government does not intend to award a contract on the basis of responses received nor otherwise pay for the preparation of any information submitted.

As a result of this notice, the Government MAY issue a solicitation. There is no solicitation available at this time. However, should such a requirement materialize, no basis for claims against the Government shall arise as a result of a response to this Sources Sought Notice or the Government's use of such information as either part of our evaluation process or in developing specifications for any subsequent requirement.

CONFIDENTIALITY:

No proprietary, classified, confidential, or sensitive information should be included in your response. The Government reserves the right to use any non-proprietary technical information in any resultant solicitation(s).