Data Integration, Management, Analysis and Visualization

Request for Information

HHS-CDC-RFI-2015-DCIPHER

Date:	September 29, 2014

1 INTRODUCTION

THIS IS A REQUEST FOR INFORMATION (RFI) ONLY to identify a single platform that can integrate, analyze, visualize, and report on key surveillance, epidemiologic, laboratory, environmental and other types and sources of data during emergency or routine investigations in an efficient, and timely manner. The information provided in the RFI is subject to change and is not binding on the Government. The Centers for Disease Control and Prevention (CDC) has not made a commitment to procure any of the items discussed, and release of this RFI should not be construed as such a commitment or as authorization to incur cost for which reimbursement would be required or sought. All submissions become Government property and will not be returned.

The objectives of this RFI are to:

1. Collect information regarding the capabilities of existing data integration, management, analysis and visualization tools and platforms.
2. Evaluate the ability of existing platforms to support the provided functional and non-functional requirements.
3. Optional and at the discretion of the Government: invite vendors to demonstrate the use of their product(s) in-person, onsite at CDC.

TABLE OF CONTENTS

2 EXECUTIVE SUMMARY

CDC plays a critical role in the detection, response, mitigation, and recovery during public health emergencies and infectious disease outbreak responses. CDC is also the primary federal agency conducting public health surveillance. In this role, the CDC collects, generates, and analyzes a plethora of data in support of disease surveillance and epidemiologic investigation activities, laboratory testing, scenario modeling, intelligence gathering, environmental investigation, and medical countermeasures deployment. CDC is currently able to meet data integration needs to manage public health emergencies. However, CDC faces a number of challenges to meet these needs. Currently CDC programs, during routine and emergency investigations, deal with many process-driven and technical challenges in their capacity to collect, integrate and analyze numerous data types and sources. CDC's Data Collation and Integration for Public Health Event Responses (DCIPHER) project staff seeks information on products that can electronically integrate, manage, analyze, visualize, report on and share key surveillance, epidemiologic, laboratory, environmental and other findings during public health investigations and responses.

3 AGENCY OVERVIEW

CDC plays a critical role in the detection, response, and recovery during public health emergencies, infectious disease outbreaks, and public health surveillance. Management of such events and effective allocation of medical countermeasures require situational awareness of disease risk, validated analytics, pathogen characteristics, spread, and impacts to society and critical infrastructure. During large-scale events, CDC is required to make rapid public health decisions based on multiple data sources (e.g., laboratory, epidemiological, environmental sampling, climate, and situational awareness), and needs to rapidly collate, analyze and share these data with internal and external partners.

Further, as charged by the CDC Director, there is a broad Surveillance Strategy now in place that aims to improve CDC's overall surveillance capabilities (http://www.cdc.gov/ophss/docs/CDC-Surveillance-Strategy-Final.pdf). The Surveillance Strategy guides efforts to make systems more adaptable and versatile, make data more timely, and population- and geographically-specific, as well as consolidate systems, and reduce the reporting burden and reporting redundancies for CDC partners. The DCIPHER project supports all three goals outlined in the Surveillance Strategy.

4 PROJECT OVERVIEW

4.1 BACKGROUND

CDC is currently able to meet data integration needs to manage public health emergencies. However, CDC faces a number of challenges to meet these needs, as outlined in Figure 1. These challenges represent examples that some CDC programs face. Laboratory testing for evidence of a pathogen in both clinical specimens and environmental samples may be performed in multiple CDC laboratories involving multiple Centers/Institute/Offices (CIOs) and throughout the Laboratory Response Network (LRN). The results from laboratory testing are contained within and reported through a variety of Information Technology (IT) systems. Epidemiology-related systems, data elements collected, and data collection methods have evolved independently of each other and are often disease-specific. Additionally, these systems have developed with different functionality and some with limited ability to 1) interface with laboratory data from the CIOs and LRN, 2) integrate laboratory, environmental and epidemiological data, and 3) link data with the CDC Emergency Operations Center (EOC) situational awareness program. Existing workflows for epidemiologic, environmental, and laboratory data collation vary amongst CIOs, and can include labor-intensive manual data aggregation and manual quality control reviews. Such processes can contribute to a delayed and potentially error-prone understanding of the scope and severity of a public health emergency and hinder CDC's ability to maximally leverage available data and information for improving outcomes.

Example Program-Level Challenges	Example Enterprise-Level Challenges
➢ Absence of an established electronic system for data management during an outbreak for many CDC programs	➢ Few of the existing IT systems capture data elements in a manner that is consistent across the different systems and programs, making integration of data difficult
➢ Inefficient electronic surveillance systems for data capture, particularly during an outbreak or EOC activation (e.g., poor timeliness, not scalable)	➢ Visualization, analysis, and data sharing is handled differently by each system
➢ IT challenges that prevent automated integration of data from multiple sources (e.g., epidemiologic and laboratory data)	➢ Data transmission methods vary widely across CDC programs (e.g., via fax, email, telephone, file transfer protocols, electronic messaging)
➢ Inability to quickly evolve and incorporate new/modified data collection mechanisms	➢ Many processes are labor-intensive and depend on manual data aggregation and manual quality control reviews
➢ Limited or inefficient visualization, data management, and data sharing tools	➢ Limited reuse of data collected during outbreak events and limited reuse of analysis and visualization routines among programs and across outbreak investigations

Figure 1. Challenges to Data Integration at the Program and Enterprise Level.

CDC needs a single electronic platform that integrates, manages, analyzes, visualizes, reports on and shares key surveillance, epidemiologic, laboratory and environmental findings during public health investigations and responses in an efficient, and timely manner (See Figure 2). The envisioned platform will enable CDC to standardize a core set of data elements across multiple surveillance

programs and event responses to capture data in a consistent manner, as well as integrate new data types and unstructured data. It will have an intuitive user interface that will enable infrequent and new users as well as experienced users to successfully operate the system and conduct advanced analytics with limited training. The system will provide CDC's external partners with near real-time access to event data through a secure interface. The platform will be designed and implemented as an enterprise data integration platform for agency-wide use which can extend to the EOC and all major public health event responses.

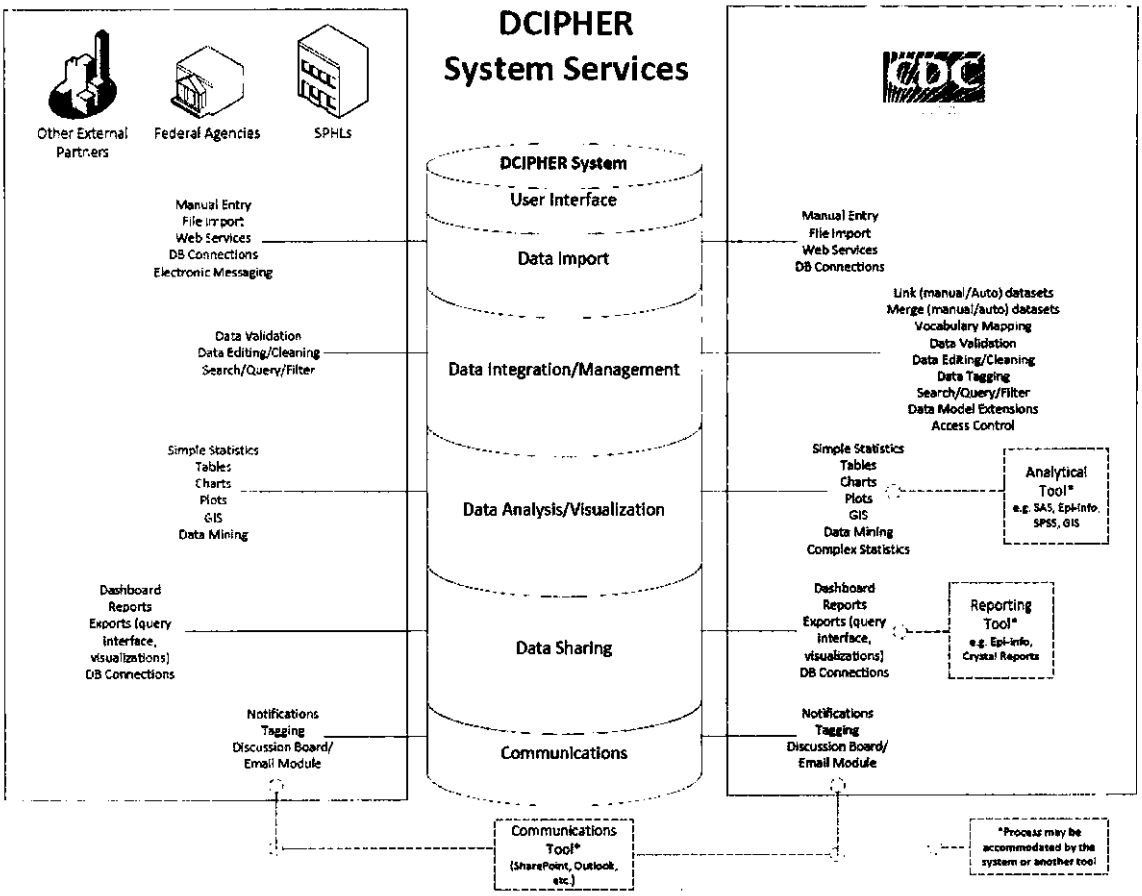

Figure 2. DCIPHER System Services.

CDC expects to integrate the following types of data:

1. Structured data, such as databases and spreadsheets that will come in many different formats;
2. Semi-structured data, such as Adobe PDF files and emails; and
3. Unstructured data, such as free text data fields, maps, photographs, and notes.

However, expected data transmission pathways will vary widely and will range from direct continual system-to-system data feeds and electronic messaging (i.e., HL7) to manual data entry directly into the platform and integration of email attachments, with many variations in between. The types, sources and formats of the data will also continually evolve.

The initial data streams that CDC is focused on integrating are epidemiologic, laboratory and environmental data streams. These data will come from a variety of sources and through multiple data collection pathways, and will include personally identifiable information. Epidemiologic data will be ingested from case investigation and other supplemental forms from internal CDC programs and from external state health departments. Laboratory test orders and results data will be ingested from internal CDC laboratories via a laboratory information management system (LIMS) database, and from state public health laboratories via the LRN. Environmental data from site investigations and management activities will be ingested from internal CDC programs, and various state and federal entities. Future data streams can include medical countermeasure, strategic national stockpile, clinical data from electronic health records and/or health information exchanges, and data from other federal agencies.

4.2 GOAL

DCIPHER is a multi-year project to develop, test and deploy an extensible rapid data integration, collaboration, management, analysis and visualization platform for public health event responses and surveillance activities across the CDC programs.

4.3 OBJECTIVES

The objectives of DCIPHER are to implement an enterprise data integration, management, analysis, and visualization platform utilizing best-in-breed technology to:

1. Integrate critical epidemiological, laboratory, environmental, situational awareness, countermeasure and response, and other known and *ad hoc* data collections and systems.
2. Build a platform that is:
    a. Designed with standardized processes;
    b. Able to facilitate stakeholder collaboration and real-time data sharing;
    c. Accessible through a single graphical user interface on mobile and PC platforms;
    d. In compliance with CDC security requirements;
    e. Driven by analytic requirements;
    f. Scalable;
    g. Adaptable to all public health events; and
    h. Interoperable with existing IT systems.
3. Create a flexible and powerful analytic interface.
4. Visualize data from multiple data streams in a single unified view.
5. Develop user-friendly visualization tools to appropriately display data to support event response activities.
6. Support timely processing and provisioning of new data sources for platform users.
7. Securely share data with external partners.

4.4 SCOPE

The scope of the DCIPHER project includes:

- Develop and operationalize a full scale platform for data integration, management, analysis, and visualization
- Integrate laboratory, epidemiological, environmental and other types of information, including legacy and future data
- Standardize integrated data elements
- Develop an enterprise data model to aid rapid data integration
- Develop processes and linkages that reduce the need for manual manipulation of data
- Manage internal and external access to data

4.5 STAKEHOLDERS

The below list describes stakeholders that are envisioned for the DCIPHER project. This list is not exhaustive and is not in any particular order.

- CDC Office of Infectious Diseases, and Centers therein
- CDC Office of Public Health Preparedness and Emergency Response
- CDC Center for Global Health
- Other federal agencies (e.g., Environmental Protection Agency, Department of Homeland Security, Federal Bureau of Investigation, Food and Drug Administration, U.S. Department of Agriculture, Assistant Secretary for Preparedness and Response)
- State Health Departments including LRN member laboratories

5 RFI REQUIREMENTS PROCESS

See section 7 for detailed RFI instructions and excel attachment.

5.1 STRUCTURE OF THE DOCUMENT

Below is the Table of Contents from the Excel spreadsheet that lists all of the requirements that the DCIPHER team seeks in a data integration and analytics platform. Excel spreadsheet is attached to this RFI:

Nonfunctional Requirements
- Objective 1: Architecture Requirements
- Objective 2: Performance Requirements
- Objective 3: System Security Requirements
- Objective 4: Data Security Requirements

Functional Requirements
- Objective 1: Data Integration
- Objective 2.1: Data Visualization: Search and Query
- Objective 2.2: Data Visualization: User Interface
- Objective 2.3: Data Visualization: Analysis
- Objective 3: Data Sharing
- Objective 4: Data Management
- Objective 5: Communications

Vendors should provide responses to each requirement according to the instructions outlined in Section 7 of this document.

5.2 PARTICIPATION TO RFI

Vendors willing to participate should confirm to CDC no later than October 06, 2014 their **Intent to Respond.** All Vendors confirming their participation should send the Intent to Respond to the attention of:

Kim Morris
Email: ycy1@cdc.gov
Phone: 770-488-2621

5.3 RFI SCHEDULE

Please response by 11:59 pm EST in accordance with RFI key dates as indicated below:

October 06, 2014	Deadline for confirming Intent to Respond
October 15, 2014	Deadline for submitting questions
October 22, 2014	Response to all questions provided back to Vendors
October 30, 2014	Deadline for receiving RFI

5.4 RFI RELATED QUESTIONS / CLARIFICATIONS / SUBMISSION

All questions related to this RFI should be directed to

Kim Morris
Email: ycy1@cdc.gov
Phone: 770-488-2621

Vendors must ensure that the response to this RFI is received via email no later than 11:59 pm on closing date October 30, 2014.

5.5 RFI TERMS & CONDITIONS

5.5.1 Disclaimer

This RFI is only a request for information about potential products / services and no contractual obligation on behalf of Agency whatsoever shall arise from the RFI process. This RFI does not commit Agency to pay any cost incurred in the preparation or submission of any response to the RFI.

5.5.2 Confidentiality & RFI Ownership

This RFI is both confidential and proprietary to Agency, and Agency reserves the right to recall the RFI in its entirety or in part. Contractors cannot and agree that they will not duplicate, distribute or otherwise disseminate or make available this document or the information contained in it without the express written consent of Ministry/Agency.

Vendors shall not include or reference this RFI in any publicity without prior written approval from the client, which, if granted, shall be granted by the individual named above. Vendors must accept all of the foregoing terms and conditions without exception. All responses to the RFI will become the property of Agency and will not be returned.

6 HIGH LEVEL BUSINESS REQUIREMENTS

This section requests information about your product(s)' ability to support the Government's platform requirements. Responses to each requirement shall be entered into two tabs within the Microsoft Excel workbook, titled "Nonfunctional Requirements" and "Functional Requirements", that is included as part of this RFI attachments. The requirements are further grouped within each tab according to the Government's high-level objectives. The Nonfunctional Requirements contains four sections: Architecture, Performance, System Security, and Data Security. The Functional Requirements contains five sections: Data Integration, Data Visualization, Data Sharing, Data Management, and Communications. The Data Visualization section contains three sub-sections, Search and Query, User Interface, and Analysis. Specific responses to each requirement shall be entered into columns C ("Base / Part / None") and D ("Explanation").

In the "Base / Part / None" column the vendor shall enter:

- "Base" if the current product version provides the required functionality in the base installation of the product ("out-of-the-box")
- "Part" if the current product version provides part of the required functionality
- "None" if the current product version provides none of the required functionality

In the "Explanation" column the vendor shall enter:

- If "Base", describe how the product(s) performs or supports the requirement
- If "Part", describe what the product(s) performs and what actions are necessary to satisfy the rest of the requirement
- If "None", describe what actions are necessary to satisfy the requirement in the current product version

7 RESPONSE FORMAT

Vendors' responses should consist of two files, a Microsoft Word or Adobe Acrobat PDF document that addresses all questions in Section 7.1, and a Microsoft Excel workbook that addresses all of the requirements capabilities referred to in Section 6. The Word/PDF document shall not exceed 16 pages double spaced at 12 point font excluding spreadsheet, Times New Roman. Vendors shall read the information requests in each worksheet and enter their response in the Response column.

As this is not a request for proposals, supporting cost data is not required nor requested. Any pricing data provided, including estimates, is informational only and must be contained within the allowed 16 pages.

7.1 CAPABILITIES STATEMENT

Provide answers to the following questions:

1	Provide a brief profile and history of your company, including if you are owned by a parent company.
2	Provide a description of your partnership agreements with companion companies, system integrators, and/or support organizations.
3	Describe your customer base (e.g., health care, public health, federal government, etc.).
4	Provide a brief history and evolution of your product(s).
5	Describe your technology roadmap for the next 3 to 5 years.
6	What are the most important advances in product capability and architecture expected in the next two years?
7	What are the available user and technical support options for your product(s)? What is the process and timeframe for resolving issues?
8	Describe the platform and software technologies required to run all of the modules of your product(s), including hardware requirements. What software, if any, is required for external users who will not be operating within the CDC firewall?
9	Describe your virtual machine policy.
10	Describe how system security is defined, including options to define and control user actions and access to data.
11	Describe how your product(s) utilizes cloud computing.
12	Describe the capabilities and level to which users can customize the product(s).
13	Describe the process for installing patches and upgrading your product(s) to the next major version. How do you address change management between product updates and customizations in order to make patches and future upgrades? How are patches and product upgrades communicated, released and applied? Can data, configurations, and existing user

	business rules be migrated? Does installing patches and upgrading require any product/platform down time?
14	Describe the types of user customization that typically make upgrades to future product releases difficult or unachievable.
15	Briefly describe what functions are provided with your product(s) (e.g., data integration, data management, data analysis, data visualization, reporting, etc.).
16	Does your product(s) support capabilities to manage controlled vocabularies, dictionaries, additional language grammar or ontologies/thesauri?
17	Describe the user interface/s. If there is no single graphical user interface by which users can access all aspects/modules of your product(s), describe how the interfaces are connected.
18	Describe the usability of the product(s) and how quickly users are able to learn to use the product(s) capabilities.
19	Describe the product's capability to produce dashboards that show data from multiple sources (e.g., laboratory, epidemiological, environmental, medical countermeasure programs, etc.).
20	Describe available training for product(s) administration and advanced configuration, if available, as well as for end-users. What delivery methods and training locations are available?
21	Does the product(s) have an active online community open to all product users sharing experiences and answering questions from peers?
22	Describe your process for gathering user feedback and prioritizing future enhancements.
23	Describe the interoperability features of your product for direct (machine-machine API) and indirect (customization modules) integration within your product(s) and third-party tools and systems.
24	Describe your site license model, including how licensing is administered and monitored, and any services required for licensing.
25	Provide a brief rationale of why your product should be regarded as better than the competition.