Vision for the Chesapeake Bay Data Enterprise

The 2009 Executive Order (EO) directed the Chesapeake Bay Program (CBP) to improve its management of environmental information by initiating "the design and development of a Chesapeake Bay Data Enterprise system to share scientific data between partners" and gave EPA the lead on that task. USGS and NOAA were directed to work with EPA on this development to ensure that the data enterprise includes relevant data and information to support all CBP goals. The EO also directed EPA, USGS, and NOAA to support the monitoring alliance. In the simplest terms, the Chesapeake Bay Data Enterprise will be: a data storage, access, and management system used to share a broad range of scientific and administrative data between partners to support modeling and analysis tools, as well as decision support and accountability systems. The data enterprise is a continuation and expansion of the existing Chesapeake Information Management System (CIMS) that was adopted by the Chesapeake Executive Council in 1996.

The overarching goals of building the data enterprise are to organize, standardize and maintain all CBP data in one central, unbranded, partnership location. Data flows into and out of this system would be automated wherever possible in order to reduce manual intervention and processing. Data products would be of known (high) quality, and would serve as authoritative datasets for the program. Data would be easily accessible in necessary formats by public, partners, and CBP. Partnership developed decision support systems would leverage the authoritative data managed within the data enterprise to provide consistent, defensible and authoritative data across the various tools. Additional CBP stated goals of this effort are:

- Quality: ensure proper documentation and metadata to better inform users of quality
- Timeliness: ensure submissions meet targeted delivery dates for specified needs/tools
- Accessibility: ensure CBP stakeholders can easily access and obtain necessary data
- Consistency: ensure all CBP data systems are utilizing the same version of a given dataset

Work completed to date

As the leader of the Data Enterprise effort, EPA worked with contractor Booz Allen Hamilton (BAH) to develop a plan for the overall data enterprise effort. Utilizing the BAH standard methodology for evaluating the relative maturity of data flows, the Data Enterprise leadership worked with CBP experts in two indicator areas (non-tidal nitrogen and tidal dissolved oxygen) to review the current data processes and deliver a set of recommendations based upon observed gaps or weaknesses. These recommendations also included ways to improve CBP existing architecture and infrastructure, to further the vision of the data enterprise. Those recommendations were detailed in a findings report that received in October 2011. Future anticipated work is based on those recommendations, and on continued analysis of specific data flows within the program.

Future work to implement Data Enterprise Vision

Based upon the received recommendations, there are a number of projects CBP should implement in order to help achieve its vision. In addition, continued work to analyze additional data themes is critical in order to address needs specific to those data flows. The Data Enterprise leadership has initiated several projects in FY 2012 to begin implementing its vision:

- Development of a quality assurance and data import tool for tidal and nontidal data
- Development of a comprehensive metadata database, catalog, and access system for partnership resources
- Development and implementation of a CBP cloud computing strategy and architecture

TASK 1. Develop quality assurance tool
The scope of this task is to develop a revised data upload and quality assurance tool to enable the submission and review of water quality data for the nontidal and tidal monitoring programs. Data providers will access a user interface to upload their water quality data. The tool will then review the submission to ensure its completeness and accuracy. CBP's current tool is focused only on tidal water quality data, and is built on outdated technology; this task will address those issues through a phased implementation approach.

- *Prototype tool development: due July 31, 2012*
- *Final tool development: due December 31, 2012*
- *Training and outreach with data providers: due January 31, 2013*
- *Data submission via new tool: due March 31, 2013*

TASK 2. Develop metadata catalog and database, and define standards
CBP realizes the value in sharing its data with the broader stakeholder community, through its own website, as well as with external sites such as EPA's Environmental Dataset Gateway and Data.gov. In order to facilitate this data sharing, a comprehensive data catalog of all its assets is required. This work will require the development of web services to support the promotion of data to external catalogs, such as Data.gov. The outcome of this effort is publication of CBP assets to the broadest range of stakeholders through Data.gov.

- *Prototype catalog development: due May 30, 2012*
- *Final catalog development: due September 30, 2012*
- *User training and outreach: due October 24, 2012*
- *Catalog population: TBD*
- *Metadata entry tool development: TBD*

{Tasks under development but not yet initiated}
In addition to cataloging our datasets in order to promote accessibility, we also need to be sure to provide our partners and the interested public a way to obtain user-determined sets of data via a sophisticated yet user-friendly data query and download system. This work, while separate from the data catalog task, would occur in close coordination with the catalog development to ensure a streamlined user experience, from search to access of CBP data resources. The outcome of this effort is a redeveloped DataHub with embedded data visualization tools, advanced search and discovery functions, and publication to multiple formats.

There is also work being done to define data/metadata standards for CBP. In order to have an effective, usable, searchable metadata catalog, and a functional, standardized data submission/quality assurance tool, CBP will have to formalize certain data and metadata standards for its data holdings. Additionally, we will evaluate ways in which various standards can be transformed to accommodate differences among partners. Finally, we will need to determine if and how we can configure our historical data to new data standards in order to standardize our data temporally. The outcome of this effort is a single set of tidal and non-tidal water quality data standards consistent, to the extent possible, with existing published standards.

TASK 3. Develop and implement cloud computing architecture pilot – Chesapeake*Stat* in the cloud
CBP has an interest in developing its cloud computing infrastructure, and has identified a number of areas where the move to the cloud for hosting would be beneficial. As there are many government initiatives positioned to utilize current cloud offerings, CBP would like to develop a strategic approach to development of its cloud infrastructure and procurement. As a pilot effort, we will host ChesapeakeStat in a cloud environment. We are

analyzing various cloud computing solutions available to government organizations; once a solution is selected, we will need to procure it for the duration of this pilot.

- *Final architecture documentation and cloud computing requirements: due March 21, 2012*
- *Draft cloud implementation: due May 24, 2012*
- *Final cloud pilot implementation: due June 12, 2012*

Data Enterprise Workplan - FY12

Task	Status	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12
Booz Allen Hamilton													
Recommendations document	Funded	X	X										
Cloud pilot with ChesapeakeStat	Funded				X	X	X	X	X	X			
Vistronix													
QAT development	Funded						X	X	X	X	X	X	X
Innovate													
Data catalog	Funded				X	X	X	X	X	X	X	X	X
Population & EME	Funded												X
Data hub requirement gathering?	Not funded												
Indus													
CIMS security plan	Funded							X	X	X	X		
Other outstanding tasks													
Metadata database creation	Not funded												
Standards development	Not funded												
Data hub redesign	Not funded												
Cloud procurement	Not funded												
Additional cloud implementation	Not funded												
Additional data theme analysis	Not funded												
Governance for data enterprise	Not funded												