Project Description: A Prototype System for Multi-Disciplinary Shared Cyberinfrastructure – Chesapeake Bay Environmental Observatory (CBEO)

I. Results of Prior Research

Prior NSF-sponsored research conducted by project PIs over the past five years is summarized in **Table 1**. These projects are briefly described below.

Ball, **Burns**, **Di Toro**, **Gross**, and **Kemp** (Table 1, Row 1 = T1) have developed a vision statement and conceptual design for a ***Chesapeake Bay Environmental Observatory (CBEO)***. Two focus group meetings canvassed the user community of environmental managers, scientists, engineers, and information infrastructure stakeholders in the Bay region and provided a clear mandate to develop a CBEO that initially focuses on hypoxia. On a related CLEANER planning grant **Piasecki** (T2) investigated the currently installed cyberinfrastructure (CI) components for the NSF-LTER site in Baltimore to guide CI development for Eos. This includes the conceptualization of a data center central clearinghouse for diverse data types. **Piasecki** (T3) organized the 3rd of four CI workshops to help elaborate CI needs for engineering research and education, focused on Cyber-Engineering CI needs and potential NSF-CI/CE funding schemes.

Burns (T4) has worked with multiple collaborators to develop the National Virtual Observatory: a large-scale federation of astronomy databases. Burns' current focus is on an *altruistic* network cache that improves the performance of distributed queries, reducing bandwidth requirements by a factor of five. **Burns** (T5) used clustered database systems to store the results of large-scale turbulence simulations for Internet use, allowing turbulence computation experiments to achieve increased detail and wider distribution. The data federation and SQL query schema developed for the NVO will be applied and extended for use in with the CBEO test bed.

Zaslavsky is a co-PI on a CLEANER planning grant (T6) that explores cyberinfrastructure components for coastal margins, the foci of intense human-environmental interaction. He leads spatial data infrastructure development on major multi-collaborator CI projects at SDSC: GEON (T7) develops services-based architecture and software to support registration, semantic annotation, ontology-enabled query, analysis, visualization and integration of large volumes of distributed community data and services. CUAHSI HIS (T8) has focused on services for extracting hydrologic time series from federal data repositories and integrating them for use in different client applications. The GEON system will host the CBEO node.

Murray (T9) has established a Research Experience for Undergraduates (REU) site that hosts 14 students per year over 12 weeks of summer research. The program provides mentoring and networking for marine science careers through student work with researchers at 3 laboratories. In its 15th year, this MD Sea Grant REU program has supported over 170 students. **Murray** (T10) is a member of a multi-investigator project that established the Center for Ocean Science Education Excellence Mid-Atlantic serving 4 states (VA, MD, NJ, NY). The focus is on K-12 classroom use of Ches. Bay observational data. The 3rd year will provide master teachers to serve as mentor teachers in the ESOS program. This model will be extended for the CBEO proposed here.

Cuker (T11-12) is the PI in the latest in a continuous series of grants (over 19 years) to support a collaboration between Hampton U. and the Am. Soc. Limnol. and Oceanogr. (ASLO) to increase the numbers of underrepresented minorities electing careers in aquatic sciences. This program targets minority students, faculty, and professionals. It has supported 522 students: 72% undergraduates, 28% grad students; 61% African Americans, 27% Hispanic, 8% Native Americans, and 4% Pacific Islanders. Also a collaboration between Hampton U., Old Dominion U., and the Virginia Institute of Marine Science funds the Hall-Bonner (H-B) Doctoral Program, which presently supports 9 doctoral students: 7 African Americans and 2 Hispanics. Four more minority students will enroll in fall of 2006. An H-B scholar will be partially supported from the proposed CBEO project, thus leveraging grant funds.

Kemp (T13-14) and colleagues have focused on how primary production, nutrient cycling, trophic structures and transfer efficiencies are affected by changes in watershed inputs and physical circulation features in the Chesapeake Bay. Two five-year projects have changed the understanding of the role of coastal ecosystems in coupling watersheds with oceans, and their support of secondary production. These projects have generated over 120 publications, including 24 from Kemp's research group.

Ball (T15) investigated characteristics and sorption properties of soot and char that impact the fate of organic contaminants in aquatic environments. The project, with RET supplements, supported 3 doctoral students, 6 undergraduate students, and summer research for 6 high school science teachers, including 5 from under-represented groups.

Table 1. Prior NSF-Sponsored Research of the PIs (Last Five Years)

ID	Project PI(s)	Award	Title	Amt.	Period	Publications
1	**W. Ball**, **R. Burns** (JHU) **D. Di Toro** (U.Del) **T. Gross** (CRC) **W.M. Kemp** (UMCES)	BES0414 -372, -429 -214, -347	Collaborative Large-Scale Engineering Analysis Network for Environmental Research (CLEANER) with Focus on the Chesapeake Bay	$70K	8/01/04 to 7/31/06	Report planned; presented: NSF 12/04; AEESP 7/06; ERF 10/06
2	**M. Piasecki** (Drexel U.) co-PI C. Welty (UMBC)	BES 414204	CyberInfrastructure for a Field Facility in Balt. as part of an Eng'g Analysis Network	$70K	7/01/04 to 6/31/06	Report planned; pres'd NSF 12/04
3	**M. Piasecki** (Drexel U.), PI	CMS 0429002	3rd Workshop on Opportunities for CyberEngineering and CyberInfrastructure	$49K	4/01/04 to 3/31/05	workshop rep. to all NSF-ENG PIs
4	**R. Burns**, (JHU), PI	IIS 0430848	SEI(AST)+II Bypass-Yield Caching for Large-Scale Scientific Database Workloads in the World-Wide Telescope	$632K	10/01/04 to 9/30/07	2 journal publications
5	**R. Burns** (JHU) co-PI Szalay et al. (JHU)	AST 0428325	ITR-ASE-(int+sim) Exploring Complex Flow 100 Terabyte Datasets	$2.17M	9/01/04 to 8/30/09	Report planned; pres'd NSF 10/044
6	**I Zaslavsky** (SDSC), co-PI w/J. Bonner (TAMU), T. Wentling (UIUC), K Jones (Howard U.), N Love (VA Polytech)	BES 0414476	CLEANER: Collab. Research: Collab. Large-Scale Eng'g Analysis Network for Env. Research for the Coastal Margin	$85K	8/1/04 - 7/31/06	Report planned; conf. proc
7	**I. Zaslavsky** (SDSC) C. Baru et al. (SDSC)	EAR 0225673	ITR Collaborative Research: GEON: A Research Project to Create Cyberinfrastructure for the Geosciences	$5.6M	10/1/02 to 9/30/07	Multiple journal pubs., and conf. proc.
8	**I. Zaslavsky** (SDSC) C. Baru et al. (SDSC)	EAR 0413182	Coll. Research: Develop. of Informatics Infrastructure for the Hydrologic Sciences	$1.0M	4/1/04 to 3/31/06	First status report, presented NSF, AGU, ESRI Conf.
9	**L. Murra**y (UMCES) coordinating PI	OCE 0215094	Collaborative Research: Establishment of Mid-Atlantic COSEE	$462K	08/01/02 to 07/3107	numerous students supported
10	**L. Murray** (UMCES) co-PI Leffler et al.	OCE 0139268	Undergraduate Research Experiences in Estuarine Processes	$284K	3/1/96 to 2/29/01	numerous students supported
11	**B.E. Cuker** (HU), PI	OCE 0437461	Expanding Linkages Between Careers in Aquatic Sciences & Under-Rep'd Minorities	$649K ($2M).	1/1/05 - 12/31/09	numerous students supported
12	**B.E. Cuker** (HU) coordinating PI	GEO 0302581	A Linked Initiative for Developing Minority Doctoral Scholars in Ocean Sciences	$1.3M	2002 to 2007	numerous students supported
13	**W.M. Kemp** coordinating PI:	BSR 8814272	LMER: Responses of Land Margin Ecosystem to Changes in Inputs: Nutrient Cycling, Production, Export	$340K ($2.5M)	1988 to-1993	~64 total pubs. 12 from Kemp
14	**W.M. Kemp** (UMCES) co-PI	DEB 9412113	LMER: Trophic Interactions in Estuarine Systems	$400K ($3M)	1994 to 2000	~61 total pubs. 11 from Kemp
15	**W. Bal**l (JHU) PI H. Fairbrother (JHU)	BES 0332160	Surface Characteristics and Sorption Properties of Chars and Soots	$443K	9/01/03 to 8/31/06	5 publications, 11 conf. proc.

II. Motivation and Overview: CEO:P

This proposal is motivated by the NSF call to develop and deploy "prototype cyberinfrastructure for environmental observatories" [CEO:P] and then "demonstrate [its] viability in order to inform the planning of, and development of, an environmental cyber-infrastructure [CI] for large-scale, environmental observing systems." (NSF, 2005) In this context, the Chesapeake Bay is a large-scale human-dominated system that would both benefit from an environmental observatory [EO] and provide an ideal site to deploy and test CI prototype components. We propose to develop a Chesapeake Bay Environmental Observatory (CBEO) as a prototype to demonstrate the potential of newly developed CI components for transforming environmental research, education, and management. The project uses a specific problem of hypoxia that will directly involve users and test the prototype's capabilities.

The CBEO project has the 11 "must-have" characteristics cited in Section II of the solicitation (NSF, 2005):

1. *The "project team includes both environmental researchers and information scientists, with ... researchers from ocean science, ecology, ... and environmental engineering."*

This is a collaborative effort between researchers in environmental engineering (Ball, Di Toro, Piasecki), ocean/marine science (Gross, Cuker), ecology (Kemp, Murray) and computer science (Burns, Zaslavsky), all PIs work across disciplines – e.g., Piasecki develops hydroinformatics, Di Toro, Ball, Piasecki, and Cuker conduct water quality modeling, and Kemp studies marine/estuarine systems (Section IV B). Seven of the PIs were already successfully collaborating on NSF planning grants prior to this proposal (Section A).

2. *The "project includes the development and deployment of a prototype of a component of cyberinfrastructure for environmental observatories."*

This project will deploy new prototypes of CI components to 1) extend the GEON core software stack via well-defined interfaces that enable "plugging in" of new (hydrodynamic and water quality) data types; (2) provide services for CBEO data registration, metadata description, semantic annotation, search and integration; and (3) join disparate and incommensurable data sets through new interpolation and averaging tools that take advantage of multiple variables and types of data (Section IV A,D, E).

3. *This proposal "identifies one or more questions about how cyberinfrastructure for environmental observatories might be designed and function and describes how these will be addressed."*

Challenges and questions about CI design and function are described in Section C**,** and approaches for addressing these are outlined in Section IV**.** Briefly, a CBEO test-bed (*CBEO:T*) will be developed simultaneously with a CBEO network node (*CBEO:N)* that uses foundations established by the existing GEON structure, and these will finally be merged into a fully functional CBEO networked prototype.

4. *The "viability of the [CBEO] approach" is "demonstrat[ed] in [this] proposal by identifying ... compelling environmental research questions that the prototype will be used to address [and] tackle ... within the lifetime of the project;"*

This project is motivated by the prospect that answers to key questions about coastal hypoxia lie at the intersections of existing observational and modeling data sets that have yet to be properly examined because of difficulties in linking, visualizing and analyzing data sets at appropriate time-space scales. Research questions and solution strategies are described in Sections IV-C, D and E. A separate CBEO science team (*CBEO:S*) wills continually existing refine and explore new research questions simultaneously and in close collaboration with activities of the *CBEO:T* and *CBEO:N* teams.

5. *The project "works with real environmental data, includes a mechanism for general users to access the prototype, and demonstrates the utility of its approach by attracting users outside of those researchers directly involved in the project.*

The various *CBEO* teams will work with selected databases, data streams and model-derived data from the extensive body of existing data (III-B3, IV-D,E) and provide an access mechanism for outside users of CBEO:T and CBEO:N (IV-F,G) respectively. Sections IV-F and G "describ[e] the types of users targeted and how the users' experiences with the prototype will be documented." Scientific researchers, engineering managers, and educators are all targeted, with outreach to the latter as described in Section V**.**

6. *The project leverages the products of existing cyberinfrastructure development efforts, for example, those supported through the ... SEIII, [AST, EAR, OCE, CMS and BES] programs;*

All PIs have conducted substantial research of direct relevance to this project (A). The CBEO will leverage ongoing developments made through the BES CL*EAN*ER (now WaTERS) program, EAR programs to develop CUAHSI-HIS (now also part of WaTERs), educational outreach programs through NSF-ENG and OCE, and most importantly, substantial CI development for the National Virtual Observatory (IIS, AST) and GEON (EAR) observatory network programs.

7. *"Where further IT development is needed, the project clearly identifies the gap that this development seeks to fill and its importance to environmental observatory cyberinfrastructure."*

The current challenges of shared CI for EOs and how the CBEO prototype will develop and deploy new means to meet some of these challenges is the subject of this proposal. See also items 2 and 3 above.

8. *"Th[is] project pursues an end-to-end approach to [two] major component[s] of cyberinfrastructure" – i.e. a test-bed observatory and a new type of node for an existing national network. These components will be useful for both regional applications and for shared CI with other researchers, nationwide.*

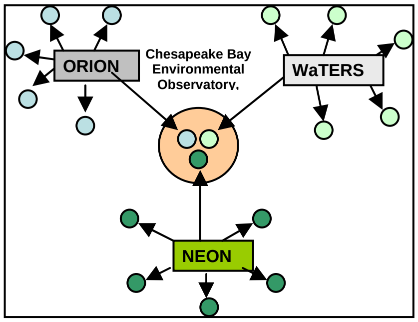

Fig 1. CBEO interactions with other EOs.

This proposal "articulat[es] the types of data involved (III-B3) and the ways in which users might wish to use these data through the careful exploration of use cases (IV-C, F), and then [describes] the deploy[ment of] a prototype that implements these types of uses." The CBEO prototype will be "capable of working with representative data and being operated by representative users performing tasks that are themselves representative of those needed in the pursuit of environmental research or education activities." (IV-C, D)

9. *The project advances the technological capabilities of the environmental research community beyond what is currently possible;*

This project will provide the research community with new means to view and use multiple datasets and data streams, in conjunction with and comparison to each other and with modeling data (IV).

10. *The project will make " extensive use of existing data streams or data sets" and it will be capable of "leverag[ing] data that can be expected to come online as a result of other environmental observation projects during the lifetime of the proposed project;*

Existing observational data, existing data streams, and past (archived) sets of model-generated data (modeling output) will be used to demonstrate the viability of the CBEO. Existing Bay data are extensive (III-B3), and the CBEO will also accommodate a new data stream coming on-line through the MD-DNR (IV-F).

11. *"The project leads to ... flexible cyberinfrastructure ... that is amenable to extension and upgrade."*

By integrating metadata standards developed through CUAHSI-HIS/WaTERS (IV-E), data federation and querying approaches developed for the National Virtual Observatory (IV-D), and software stacks and other approaches developed for the GEON web interface (IV-E), the prototype components developed in this work are assured to be flexible and amenable to extension and upgrade. In this regard, we envision that our system will later be integrated to multiple EO networks, as illustrated in Fig. 1.

III. Background: Science and Cyberinfrastructure Challenges on the Chesapeake Bay

III A. Chesapeake Bay as a Study Site

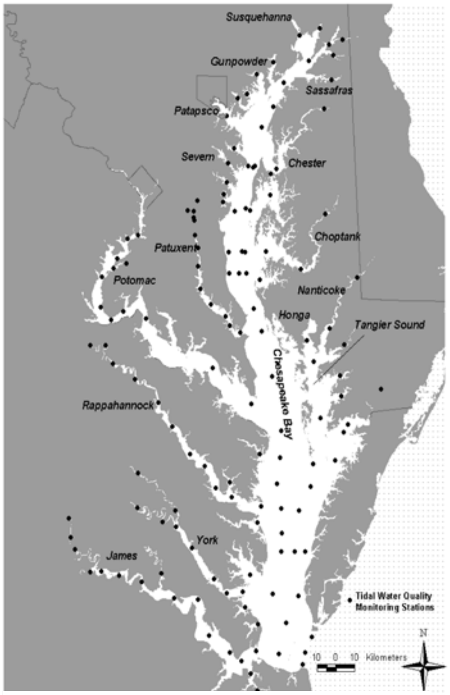

Fig.2. Chesapeake Bay and location of water quality monitoring stations grab-sampled at multiple depths twice-monthly since 1985.

The Chesapeake Bay (Fig.2) is an ideal location for researchers to demonstrate how CI can be used to help answer complex and unresolved science questions, enhance the ability of educators to teach environmental science, and inform management. The Chesapeake is a well-studied and intensely monitored coastal ecosystem with active research and resource management activities. The degradation of water quality, sea grass communities, and benthic animal populations from human activities is well documented (e.g., Kemp et al. 2005). The region has a long-standing history of collaboration between researchers and managers, with an objective of management based on sound scientific understanding of system processes and responses to disturbance (e.g., Malone et al. 1993, Boesch et al.

2001). Federal mandates and multi-jurisdictional initiatives motivate an aggressive Bay restoration program (Boesch et al. 2001). A systematic monitoring program has been in place since 1984, and numerous integrated research programs have assessed various aspects of system dynamics (e.g., Kemp et al. 2004, 2005, Roman et al. 2005). A coupled hydrodynamic and water quality model has been calibrated (Cerco and Cole 1993, Cerco 1995, Johnson et al. 1993), upgraded (e.g., Cerco 2005), and repeatedly run to produce multi-year (1986-2000) simulations of physical circulation and ecological dynamics, at a spatial scale of 10 to 1 km^2 horizontal and 1 to 5 m vertical and a temporal scale of minutes to hours (Cerco 1995). These and other models are available to all researchers and managers, with facilitated by the Chesapeake Bay Community Modeling Program managed by PI Gross (Gross et al., 2005).

III B. Chesapeake Bay Science: Challenges, Questions, and Resources

1. Hypoxia as a Prototypical Science and Research Challenge. The seasonal depletion of dissolved oxygen (O_2) from coastal waters, hypoxia, is a widespread problem of growing proportions that is tied to human influences (Rabalais et al. 2002, Howarth et al. 2000), and of worldwide concern (e.g., Rosenberg 1990, Diaz 2001). Accumulating evidence indicates that the intensity and extent of summer hypoxia in the Ches. Bay have been increasing since the early 1950s (Hagy et al. 2004). Although *inter-annual* variations in extent and duration correlate with fluctuations in freshwater discharge (Fig. 3a), *long-term* increases follow trends of increasing nutrient loading from the watershed (Fig. 3b, Hagy et al. 2004). Thus, hypoxia in coastal waters is a consequence of interactions between watershed land-use, hydrology, climate, and oceanographic processes, the understanding of which require linked EOs in the various domains. Even within the estuarine domain, a regional EO could transform our process understanding, because some fundamental and unanswered questions about hypoxia remain (Nixon 1995, Cloern 2001). Our ability to address these questions would be transformed by better means of finding, integrating, interpolating, and visualizing multiple data sets (including model-derived data) that are complementary but disparate in their characteristics (precision, temporal coverage, spatial coverage, etc.).

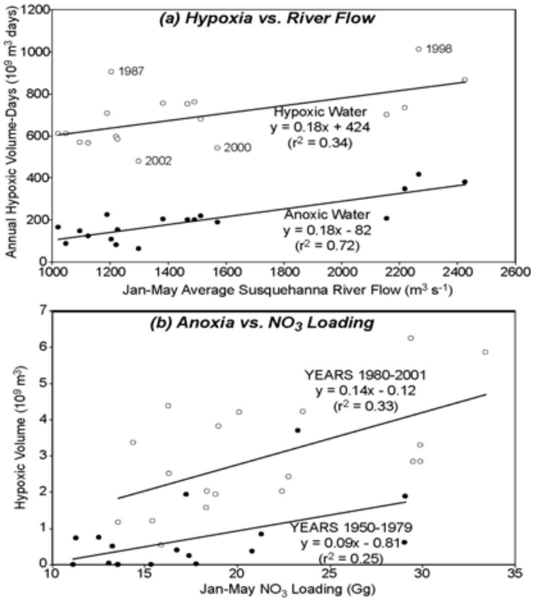

Fig. 3**.** Relationships between Bay hypoxia and river flow or nutrient loading. (a) Time-integrated volume of hypoxic and anoxic (<0.2 mg O_2 l^{-1}) water *vs.* winter-spring river flow. (b) Mid-summer volume of anoxic water *vs.* nitrate loading for earlier (1950-1979) and more recent (1980-2001) time.

2. Chesapeake Bay Science Questions. To provide science focus for the project, we have identified three related questions pertaining to hypoxia. The first derives from Fig. 3b, which indicates that for a given rate of N loading (total N is highly correlated to NO_3), a larger hypoxic volume has generally been observed during the last twenty years in contrast to 1950-1979. The considerable scatter in the relationships is largely attributable to fluctuations in river flow. If proven significant, these differences would imply the existence of highly non-linear underlying mechanisms that are currently not well understood but which have major implications for prediction and management (Kemp et al. 2005). Conclusively validating these trends and elucidating their causes would transform our understanding of the processes causing hypoxia.

The second question pertains to the observed relationships between year-to-year variations in hypoxia and river flow (e.g., Fig. 3a). Close examination of these trends reveals considerable scatter which is poorly understood, and it is unclear why the existing coupled hydrodynamic-water quality model, although calibrated to reproduce seasonal patterns, is incapable of simulating these inter-annual trends (Fig. 4). We suspect that that there are problems both with estimating the time course of hypoxic volume and with modeled processes, both hydrodynamic and biological. Exploring these issues requires better integration of existing observational data sets with each other and with the model-derived output.

The third question involves the regional and seasonal balance between source and sinks of organic matter that regulate O_2 consumption and hypoxia in bottom waters. Although prior analyses suggested that organic matter production by algae in shallow waters may fuel respiration in deep channel waters (Kemp et al. 1997), the general structure of the water quality model being used for management (Cerco & Noel 2004) reflects the more traditional view that phytoplankton production in surface of the deeper regions of the Bay is the major source of organic input fueling hypoxia (e.g., Malone et al. 1988). A recently established shallow water monitoring program should provide the requisite data (MD DNR, 2005) that we will incorporate into the CBEO.

3. Data Resources for the Chesapeake Bay. Varied observatories and monitoring programs generate extensive databases relevant to hypoxia, including water quality conditions, physical structure and circulation, sediment characteristics, biogeochemical processes, abundances of plankton, benthos, and fish populations. On the Bay, an extensive set of data have been and are being collected at a wide range of scales using diverse means including: grab samples, moored and towed sensor systems, satellite and aircraft remote sensing. Additionally data from acoustic methods for sampling water depth, sediment character and fish abundance, sediment cores, biogeochemical rate incubations, and output from numerical simulations of physical circulation, water quality, and ecological rates are available (Table 2). The time and space scales of measurements range from minutes to decades and from centimeters to hundreds of kilometers.

If the various data sources could be accessed and their different spatial and temporal scales properly resolved and interpolated in a user-friendly way by a web-based workbench, it would transform our ability to visualize, analyze and interpret the data. For example, Fig. 5 shows fine-scale distributions of water quality obtained from towed undulating sensor systems over the whole Bay. Data from this monitoring program include measurements of physical, chemical and biotic aspects of water quality, but the complementary large data bases with key ecological rates (e.g., primary productivity, community respiration and nutrient recycling, and organic particle sinking) were obtained under different research programs at different scales and times. Moreover, output of numerical model computations provide “model-derived data” for other physical circulation, water quality, and ecological properties. These model output data are available at fine resolution, as discussed in III-A. To address the scientific questions posed, we must interpolate and integrate the various databases so that computations of flux, transformation, and coherence can be made and patterns and relationships visualized and quantitatively analyzed.

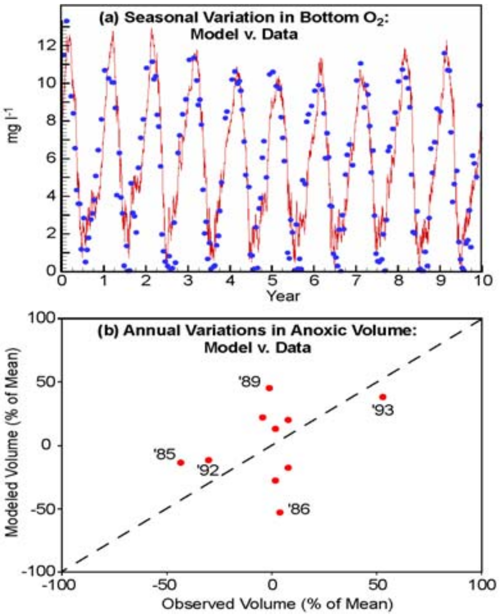

Fig. 4. Simulations of a water quality model for Chesapeake Bay (Cerco 1995) showing (a) close match for bottom O_2 at seasonal scales, but (b) weak match for interannual variations in the integrated volume of summer hypoxic water (C. Cerco, personal communication.).

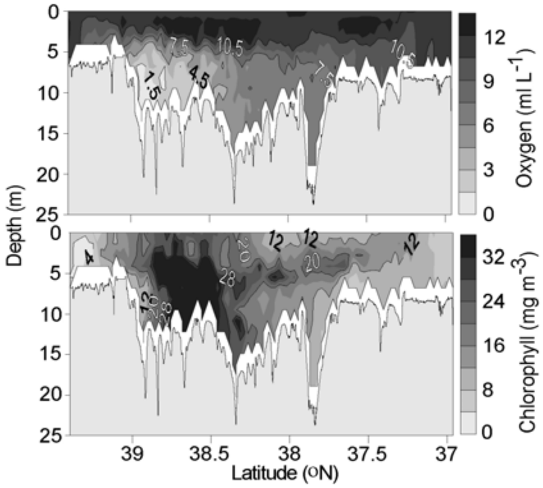

Fig. 5. Dissolved oxygen (upper panel) and chlorophyll-*a* (lower) derived from transects with a vertically undulating sensor system (*Scanfish*) towed along the central axis of Chesapeake Bay (April 2000). Data from the NSF LMER:TIES project (Table 1; see Roman et al. 2005).

III C. Challenges and Objectives for Shared CI on the Chesapeake

Resolving problems of semantic, syntactic, and content heterogeneity across datasets maintained at

Table 2. Space and time scales of Chesapeake Bay data bases relevant to hypoxia questions.

Database	Variables Included	Space Scales[b]		Time Scales		Website Address or Data Source
		Grain	Extent	Grain	Extent	
1) Grab Sample Monitoring	T, S, Chl, N, P C, Si, O_2, k_d	z, 1m x, 10 km	0-30 m 300 km	2-4 wk	20 yr	www.chesapeakebay.net
2) Fixed Sensors						www.hpl.umces.cbos.edu www.chesapeakebay.net www.cbos.org
a. Shallow(DNR)	T ,S, Chl, O_2, k_d	0 km	1-10 km	15 min	2-5 yr	
b. Deep (CBOS)	T, S, u, v, w	10 km	300 km	15 min	5 yr	
3) Underway Sensors						
a. Shallow (Dataflow)	T, S, Chl ,O_2, k_d	1 m	1-10 km	4 wk	1-3 yr	www.eyesonthebay.net
b. Deep (NSF TIES)	T, S, Chl,O_2, zooplankton	z, 1m x, 10 km	10 m 100 km	3 mo	5-6 yr	www.chesapeake.org/ties Roman et al. 2005
4) Remote Sensing						
a. SeaWiFS (satellite)	Chl	1-10 km	200 km	1-10 d	10 yr	Harding et al. 1994
b. ODAS (aircraft)	Chl	5 m x 50 m	300 km	1-4 wk	10 yr	Harding et al. 2005
5) Bathymetry	depth, volume	z, 1m, x,1 km	300 km	na	na	Cronin & Pritchard 1975
6) River inputs	flow	10 km	300 km	1 d	50 yr	www.usgs.gov
7) Nutrient load	N, P, Si	10 km	300 km	1 d	20 yr	www.chesapeakebay.net
8) Climate	T, rainfall, wind, tides	10 km	300 km	1-24 h	50 yr	www.noaa.gov
9) WQ Model	T, S, Chl, N, P C, Si, O_2, k_d	z, 1m, x,1 km	300 km	1 hr	15 yr	C. Cerco (US Army CoE),personal commun...
10) Hydro. Model	T, S, u, v, w, $K_{x,y,}$	z, 1 m, x,1 km	300 km	5 min	15 yr	B. Johnson (US CoE) personal commun.

[a] Abbreviated for variables: T=temperature, S=salinity, Chl=phytoplankton chlorophyll-*a*, O_2=dissolved oxygen, N=nitrogen, P=phosphorus, C=organic carbon, Si=silica, k_d=diffuse light attenuation coefficient, u, v, w = velocity components in x, y, z directions, $K_{x,y,z}$ =turbulent mixing in x, y, z.

[b] Cartesian coordinates: "x" follows land-sea gradient, "y" is horizontal axis perpendicular to x, and "z" is the vertical axis.

different locations for different purposes is a major challenge facing all EOs. In investigating hypoxia we face not only problems of semantic and syntactic heterogeneities and lack of relevant metadata descriptions, but also the challenge of integrating data collected at different spatial and temporal resolutions, under different sampling schemes, with different frequencies and extents of coverage, and with different levels of precision and uncertainty. Formal representation and metadata description of such disparate datasets are required. This is a pre-requisite for developing scalable and robust means for querying, exploring and integrating observation data, and preparing them for further analyses. The CBEO project will address these challenges and develop data registration, query, integration and analysis tools for shared use within the CBEO as well as across multiple networks (Fig 1) as a prototype of how CI can allow multiple EONs to share tools and data.

1. Inconsistent Semantic and Syntactic Representations. The major challenge to storage and federation of disparate datasets is developing and implementing appropriate metadata standards and data registration systems that fully reflect data properties. The key to data interoperability at a basic level (search, viewing, retrieval, and analysis) is resolution of syntactic and semantic differences through a mediation layer (Ludaescher et al. 2005). This layer reconciles the different metadata standards and implements vocabularies that allow connections between otherwise disparate data descriptions, e.g. "Gage Height" = "Stage." Although this example is trivial, similar inconsistencies exist in the way metadata annotations are encoded (plain text *vs.* XML) and published, and the manner in which data are stored (location, format). This is a potentially serious problem for interoperability of EO networks, because the ecological community networks are either already using (LTER) or planning to adopt (NEON) the Ecological Metadata Language (EML, 2003) for metadata descriptions, whereas the ocean community network (ORION/OOI) is slated to use IOOS-developed (FGDC-based) Data Management and Communications (OCEAN.us, 2005) and the WaTERS community (through CUAHSI-HIS) is planning to use an ISO-based standard. Interoperability will therefore require that metadata

tags be mapped, semantics reconciled, and both syntactics and semantics implemented in a way that allows automatic mediation without human intervention (e.g. Bermudez and Piasecki, 2006; MMI, 2005). Designing new services to meet these challenges will be an important aspect of the CBEO.

2. Integration of Disparate Information Systems. Integration of data collected within heterogeneous but overlapping cross-disciplinary observation networks requires that the CBEO rest on a sound but flexible technological foundation (e.g., that established by GEON). Required components for data integration include: i) format, projection and unit conversion; ii) knowledge-based data registration and query rewriting, and iii) spatio-temporal interpolation techniques. The most important challenges are to develop a framework that i) can be modified and extended; ii) is sufficiently flexible, transparent and documented to allow integration into future new systems; iii) is based on well established standards with a high degree of provenance; iv) has an access interface that can be programmed against; and v) is robust and scalable for supported data sources, data types, and interpolation models, EON developers must find conceptual representations that map out both the storage structure (i.e., database and file system) and the descriptive elements (e.g., metadata tags, semantic conventions). This applies to both intra-EON and extra-EON information systems. It is also important to permit basic access to data through web-services that are operating system independent, can be integrated into multiple user applications and workflows, and can use a uniform simple logic. Web services that provide data exposure to the "outside" world are one of the great promises of EONs and progress is being made. For example, SOAP web services that permit access to USGS National Water Information System have been made public at SDSC's server for use in custom desktop or web-based end user applications, such as the Hydrology Data Access System (SDSC 2006b). The project team at Drexel U. and SDSC, which are partners in the CUAHSI Hydrologic Information System project, will integrate these or modified versions into the CBEO.

3. Merging of Data for User Applications. The CBEO will allow the use of disparate collections of observational data and model-derived results. Many Bay-related data sets fall into two classes: 4-dimensional data (4D = 3 spatial + 1 time) and if variable type (e.g. temperature, velocity, concentrations) is also considered as a dimension, then 5D. In this context, a metadata description is an additional variable that is useful for such tasks as "weighting" other types of data for display, interpolation, or averaging. In fact, most single-variable data represent polygons in 4D space-time because they represent averages of some sort, e.g. a concentration derived from a photograph pixel or a measured (model-derived) concentration within a sampled (grid element) volume over the compositing time of the sample (or temporal discretization of the model). The rest of the data universe can be even more complex, such as descriptive data for sediment cores, living resources. Initially we will focus on "normal" 5D hydrodynamic and water quality data, and then store and retrieve more complex data.

4. Configurability of a Virtual Observatory. The CBEO will be a multipurpose and multifaceted observatory. However, each user's individual view should meet their particular needs. Such a custom CBEO view will likely represent a particular spatio-temporal window over a subset of observatory datasets and available processing workflows. Several existing CI projects allow for such personal user areas maintained at grid nodes, such as (myGEON, 2006) in the GEON project, and (myLEAD, 2006) in the LEAD project. In the personal areas, authorized users may reference datasets and other resources found in resource catalogs, and perform supported functions using common software stacks that include semantic annotation, data transformation, and online mapping. In GEON, an ability to develop user-defined processing workflows is also being added. In the proposed system, we will extend the personal workbench concept to enable users to define an integrated view over selected CBEO resources, specify how these resources are linked, and instantiate the symbolic representation as a standards-compliant XML document or a well-known format with accompanying metadata, which in turn can be used as input to analysis programs. The latter step would rely on GEON or SEEK services for resolving semantic mismatches.

5. Challenges for Interpolation. Integrating and analyzing relationships between datasets is difficult because of their spatial and temporal heterogeneity. Unlike most database applications, data cannot be joined by equality or a simple predicate. Rather, data have complex relationships defined by the underlying science and properties of data generation, e.g. sampling discipline and measurement error. Generally, putting the data on a common basis requires interpolation algorithms. For example, a scientific question that examines point data collected hourly at buoys with pixilated data collected daily by satellite requires interpolation in space-time.

There are two important and challenging aspects to data interpolation. First, the distributed nature of the CBEO complicates the evaluation of interpolation functions. Data can brought together at a single site and locally interpolation. In a distributed environment, this strategy requires transfer of unnecessarily large amounts of data, e.g. all readings used in the interpolation kernel. For performance reasons, interpolation should be evaluated at the databases resulting in smaller data transfers (Malik *et al.*, 2005) thus improving performance. To allow scientists to customize retrieval and interpolation functions, we propose to develop interpolated spatial joins using SQL language extensions and user-defined functions. User-defined functions, UDFs, allow complex programs, written in Java, C++ and C#, to be executed at the database (Chamberlain, 1998). They are the fundamental construct enabling database federations to ``bring the computation to the data" (Szalay *et al.*, 2002). The challenge here is to provide class libraries of interpolation functions that can be invoked directly and provide customized interpolation functions to be executed at the database. A successful application has been developed by PI R. Burns for the National Virtual Observatory [Malik *et al.*, 2003].

Second, a common feature of all observational data is its incompleteness in 4D, in contrast to model derived data that is continuous in time and exits in all polyhedral model segments. The problem of commonly used interpolation schemes, from simple weighting schemes, e.g. inverse least squares, to more complex methods, e.g. kriging, is the need for an interpolation kernel to relate observed to interpolated points.

The presence of model data, specifically hydrodynamic flow fields, as part of CBEO suggests that an interpolation scheme can be devised that uses this information. This method has been used with flow fields that are estimated from the data to be interpolated (Chin and Mariano, 1997, Gibson and Spann, 2003, Larson et al., 1998, Ueng and Wang, 2005, Yang and Parvin, 2003, Zhang and Akambhamettu, 2003.) The CBEO derived flow fields are independent of the data and can therefore be used with sparse observational data. This type of interpolation is "intelligent" – it will only use data points for the interpolation that could have flowed to the interpolation point. So barriers to transport, e.g. pycnoclines and land boundaries, are properly taken into account. This is an exciting and novel application of integrated 5D cube data that can greatly expand the analysis capability and resulting insight that can be derived from the CBEO integrated data. We regard this possibility as a potentially transformative innovation for integrating disparate datasets.

6. Challenges from a User's Perspective. From a user's point of view, the CBEO should be the vehicle for performing research using disparate collections of data sets and model outputs that could not previously have been used together. In the context of hypoxia, a user might be most interested in applying merged 5D data of the type discussed above (III-C3). The following are examples in the form of query scenarios.

1. **Queries for existence and extent of data**. For a variable in a particular spatial and temporal domain, with a defined granularity, what is the data coverage? CBEO would return a 5D data cube representing presence/absence of data that for visualization purposes could be sub sampled: 1 variable at a time (4D), grouped into time intervals (3D) and spatially sliced and collapsed into 2D slices.

2. **Queries for a data merge**. Select from an extant 5D subset and produce the merged data set. Select an averaging element in 4D space-time. Populate all the elements with point estimates or multiple entries if overlapping data are present. Preserve the data polygonal structures. Visualize the results using tools described in 1.

3. **Queries for data interpolation**. For filled elements, specify compositing algorithms in time and space (e.g. volumetrically weighted averages that incorporate point values). For unfilled elements, choose an interpolation algorithm, for example, velocity field directed interpolation, or multiple regression estimates.

4. **Queries for analysis**. A number of examples come to mind; i) *Space-time averages*: the volume of the space-time isosurface for $O_2 < 2.0$ mg/L. The volume of the intersection of two isosurfaces. ii) *Areal mass fluxes*: Daily average flux of nitrate across a plane in the bay. Flux of O_2 across the pycnocline, defined as a surface bisecting e.g. a 1.0 g/L-m vertical change in salinity. The flux of POC from shallow regions to deep; iii) *Sources*: The net monthly production of algal biomass in the photosynthetic zone accounting for loss by turbulent mixing and settling to the sediment. The sediment flux of ammonia during oxic and anoxic periods.

5. **Queries for model – data comparison**. A typical annual model run can produce 100 Gbytes of output for the hydrodynamics and the 25 computed biological and chemical variables. During the same year there are perhaps 1 Gbyte of observational data. Compute and visualize statistics at various time and space scales.

6. **Queries for data assimilation**. Produce a "source difference" 5D cubes in which the fifth dimension is the source/sink that must have been present to produce the observed concentration, corrected for transport effects using the velocity and mixing data. Arguably, this is the most useful form of data assimilation since the

output is a direct quantification of the processes responsible for the changes in the state variables.

These six query scenarios are the basis web services and user interfaces development targeting the following three categories of users: 1) project PIs and their research teams, for whom configurability and flexibility of the CBEO environment and the ability to quickly integrate datasets are paramount, 2) undergraduate and graduate students, with the emphasis on observatory data exploration, packaged queries, and curriculum integration, and 3) researchers and system integrators outside the observatory domain, for whom CBEO shall provide a gateway to well-documented Chesapeake Bay observations and model-derived data.

IV. Project Approach – CBEO Development

IV A. CBEO Overview: An End-to-End System Framework

The CBEO end-to-end cyberinfrastructure will deliver *observation data services* (III-C-6) that include: i) registering and annotating observation data resources of different types, ii) searching, accessing and exploring the available metadata and data, iii) configuring and transforming the data into a common spatio-temporal framework, iv) using the transformed data as input to analysis, modeling and visualization systems, and v) providing personalized user interfaces to access the services. Such an end-to-end system would leverage several cyberinfrastructure components developed in other projects including GEON (Baru 2004, Zaslavsky, 2005a,b) and the National Virtual Observatory (NVO; Szalay *et al*., 2002). GEON cyberinfrastructure follows the principles of services-oriented architecture and represents a system of point-of-presence (PoP) data and compute nodes, each with a respective stack of layered software components (GEON pack). In CBEO, we will deploy and extend a GEON data node. Re-using core grid and data management services, including security and authentication management already implemented within the GridSphere portal, will let CBEO:N focus on its core contributions: support for additional types of data in CBEO and services for integrating disparate datasets and interfacing them with common analysis systems. This will include, specifically: (1) modeling and representing the additional types of data being integrated in CBEO (III-B3); (2) developing web services that extend GEON registration and search functionality to the new data types; (3) defining standard interfaces for incorporating the additional data types in the software stack; (4) developing methodology for transforming and packaging a user-configured fragment of virtual observatory as an input for analysis, modeling and visualization software; (5) utilizing powerful visualization software products, like the Integrated Data Viewer, IDV (UNIDATA 2005) or the FERRET system (PMEL, 2006), that has been integrated with the GEON environment (SDSC, 2005a); and (6) defining communication interfaces between the CBEO services and applications, and the GridSphere portal framework used in GEON, which already implements GAMA security and authentication management (GAMA 2006). The latter task is particularly important as a step towards making the emerging cyberinfrastructure components re-useable across different observation networks. In the case of GEON – CBEO:N integration, we intend to re-use and enhance GEON's data registration, semantic annotation and search web services to accommodate additional data types, and develop additional user front ends "wrapping" GEON's GridSphere portal. The developed SOAP web services and applications will be documented and published via the GEON catalog, to make them available for programmatic and web browser-based access from external cyberinfrastructure applications.

IV B. Project Organization and Management

The development of a successful CBEO will require a close cooperation among domain scientists (DSs) and computer scientists (CSs). It is critically important that the needs/wants of the DSs be technically feasible and properly balanced against the time and effort that CSs must expend. Conversely, any developments, abstractions, or applications envisioned by the CSs must be evaluated by DSs as contributory to the solution of pressing science questions. Thus, parallel activities and active dialog are important. To effectively achieve this collaboration, we plan the administrative structure shown in Table 3. This group will meet quarterly either in person or via conference facility to discuss progress and monitor developments.

Four parallel tracks of activity will be simultaneously undertaken. Each track will be led by a separate PI with at least three other PIs as team members. Each PI will be on a total of three teams. Teams will formally share results and review progress quarterly.

(1) ***CBEO: S*** *– Science and Management;* PIs: **Kemp**, Gross, Di Toro, Ball, Murray, Cuker; Tasks: IV-C

(2) ***CBEO: T*** *– Test Bed;* PIs: **Burns**, Ball, DiToro, Gross, Kemp, Piasecki, Murray, Cuker; Tasks: IV-D
(3) ***CBEO: N*** *– Node Development for Networks;* PIs: **Piasecki**, Zaslavsky, Gross, Burns; Tasks: IV-E
(4) ***CBEO: E*** *– Education and Outreach*; PIs: **Murray**, Cuker, Zaslavsky, Gross, Burns; Tasks: V-A

Table 3. Project Administration – Project Team Leaders

Name	Discipline	Programmatic Responsibility	Admin. Responsibility
D. Di Toro	Environ. Engineering	Water Quality Modeling	Director
W. Ball	Environ. Engineering	Integration: Sci. / Models / Eng'g Mgmnt.	Assistant Director
T. Gross	Ocean Science	Hydrodynamic Data, Modeling	Administrative Director
W.M. Kemp	Ecology	Science: Data Sources, Bay Processes	***CBEO:S*** Team Leader
R. Burns	Computer Science	CBEO Test Bed	***CBEO:T*** Team Leader
M. Piasecki / I. Zaslavsky	Eng'g – Hydroinformatics / Computer Science	CBEO Network Node	***CBEO:N*** Team Leaders
L. Murray / B. Cuker	Marine Science / Marine Science	Education and Outreach	***CBEO: E*** Team Leaders

IV C. Chesapeake Bay Science and Management (CBEO:S)

The *CBEO:S* team will develop and address well-posed science questions, first using the local test bed *CBEO:T* and then the CBEO network node *CBEO:N*. The team will continually refine its understanding of both the science questions and the CI tools that require further development. This will be communicated to the *CBEO:T* and *CBEO:N* teams. Full-time effort by the UMCES student and partial effort by the UD student will be devoted under PI guidance, to the following activities:

1. Examine the Hagy et al. finding (Fig. 3) in detail. Use the *CBEO* to explore anoxia development, using velocity-directed and multivariate interpolation schemes to produce O_2 sources and sinks as variables. Explore the statistical significance of the perceived shift in loading: hypoxia relationships. (UMCES).
2. Perform a rigorous comparison of the 15 yrs of modeling output data and the integrated observational data set. This is in contrast to previous analyses that used only the fixed station data (Fig 3a). Focus on calibrations of source and sink terms as well as state variables (UD).
3. More carefully compare the predicted *vs.* observed hypoxic volume (Fig 3b). Use the integrated observational data set to seek reasons why the model fails to describe inter-annual variations. Examine mis-calibration(s) under item 2 as potential cause(s) of problems (UD).
4. Analyze the flux of organic matter and dissolved oxygen from the shallow water regions of the Bay using the integrated data set and the hydrodynamic model transport (velocity and diffusion) data (UMCES).
5. Use the integrated data set to make improved estimates of primary production (Cerco and Noel, 2004), a fundamental component of the models. Compare results with the integrated model-derived data (UD).

IV D. CBEO Test Bed Development (CBEO:T)

The ***CBEO:T*** will be a clustered database system (III-C) including the integration of model-derived data with observations. This will be a computational platform for hypoxia research and will be made a node in the GEON framework (IV-E), thus becoming globally accessible to the scientific community. Led by PI Burns and a post-doc, the *CBEO:T* team will work with doctoral students at JHU (Comp. Sci.) and UD (Environ. Eng'g) to design, test, and implement a local test-bed system. The *CBEO:T* will be constructed from selected existing databases and data-streams (III-C); archived output from Bay-scale model runs (IV-H); a new data-stream that is currently under development by for MD DNR's "Eyes on the Bay" program (IV-H); and existing and newly developed tools for data integration, visualization, interpolation, and analysis. Specific tasks to be undertaken include:

1. Construct a unified schema across existing data sources and model data that resolve structural and semantic heterogeneity within the data in close collaboration with *CBEO:N* (IV-E**)**;
2. Develop SQL language constructs for spatial joins that resolve spatial and temporal heterogeneity among point data, continuous data, and spatially-averaged data;
   - Adapt the middleware SQL parser and optimizer developed for the NVO to translate dataset joins to standard SQL operations and schedule these across multiple databases **(**III-C5**)**;
   - Build a library of user-defined functions (UDFs) that implement interpolation functions for temporal and

spatial joins executed at the databases, including the velocity-directed interpolation (III-C5);

3. Provide Web-services interfaces to the test bed databases so that our identified users (IV-F) may invoke, manage, and visualize scientific questions using the *CBEO:T* across the Internet (IV-D);
4. Assist in integrating the *CBEO:T* databases into *CBEO:N* (IV-E) and
5. Maintain dialogue with the user community through moderated mailing lists and the construction of a CBEO WIKI in which users and PIs share practices and experience (IV-F).

IV E. CBEO Node Development for National Networks (CBEO:N)

The incorporation of the CBEO into a national network will be via a GEON node CBEO:N. We will build upon data federation and query approaches that have been developed for the NVO and we will deploy and extend a GEON data node. Novel CI components that the CBEO will provide include: 1) extension of the GEON core software stack via well-defined interfaces that enable "plugging in" of new (hydrodynamic and water quality) data types; (2) new services for CBEO data registration, metadata description, semantic annotation, search and integration; and (3) new tools to join disparate and incommensurable data sets through interpolation and averaging approaches that take advantage of multiple datasets and variables.

During the first two years of the project, a team led by PIs Piasecki and Zaslavsky will work primarily at Drexel Univ. and the SDSC to implement the CBEO into a national EO Network (EON). During year three, the team will integrate additional functionalities developed by the CBEO:T team to bring those tools and data to the nationally networked environment. One major specific task will be to develop a data access portal environment – that is, a web-based user-configurable environment from which CBEO users will be able to launch data registration, search, data transformation, analysis, modeling and visualization applications.

The Drexel team (M. Piasecki and students) will work on the following tasks:

1. Collect the available metadata for the selected data sources and analyze them for completeness and coverage as well as for keywords used in describing the data. This will include data sets that have been annotated using different metadata frameworks. Missing metadata will have to be developed or researched.
2. Based on results from (1), design a metadata profile for use with model-generated data that is compliant with the ISO 19115 standard. Based on this reference standard, develop conceptualizations of the data sources and metadata profiles and use those to define a mapping framework that will mediate (translate) between the data source descriptions and the project ISO 19115 profile.
3. Design a controlled vocabulary for use with model-generated data that can be mapped against EML and DMAC (FGDC-based) metadata descriptions. This includes the development of cross-walks that map metadata tags, as well as ontologies that map community-specific keywords to each other.
4. Develop a descriptive ontology for the hypoxia problem that can be registered with the GEON based framework. This is of particular importance as it allows the registration of data sets with specific components within the hypoxia ontology representation.

The SDSC team (I. Zaslavsky and a programmer) will focus its efforts on the following four tasks:

1. Develop and deploy a digital observatory node based on GEON software stack. This task includes development of a mechanism for extending core functionality of the GEON software stack to include registration of additional data types common for aqueous EO systems, and design and develop services for search, data access and format conversion with these additional data types and formats.
2. In collaboration with the Drexel team, develop a system for registering available CBEO data sources, associated metadata and semantic annotation web services at the digital observatory node, and orchestrate the registration of heterogeneous datasets available for the study area. Develop a system for registering model-derived datasets to support discovery and integration of such datasets with other network resources.
3. In collaboration with researchers from the *CBEO:T* team, design and deploy web services for spatio-temporal interpolation, and for selecting appropriate interpolation schemes based on registered source metadata. These services will help prepare the EO for integration and provide a prototype for similar CI endeavors where integration between data and model observations is essential.
4. Design and deploy an EO portal to help users discover, access, and manipulate available data resources and services. As part of the portal, develop a personal "myCBEO" area where registered users can maintain discovered resources, define model-specific views, manage these configurations, and prepare the datasets for download and input in analysis, interpolation, visualization (IDV) and modeling software.

IV F. Plan for the Participation of Outside Users

Community feedback is essential for successful design and application of the CBEO. The *CBEO:T* will therefore be made available to outside users very soon after its creation, and feedback will be immediately solicited through the Wiki. The targeted users will include environmental scientists, resource managers, educators and students. We intend to hold annual (yrs 1 and 2) and semi-annual (yr 3) workshops that will be teaching – learning experiences for all parties. Our development group will demonstrate the current CBEO system, provide training and seek advice and criticism. Potential users from academe and regulatory agencies will be invited, including Bay managers and scientists who have collected and maintained the databases and models that will be incorporated into the CBEO, many of whom attended focus-group meetings for the Chesapeake CL*EAN*ER planning grant. Letters of interest and commitment from key individuals are attached to this proposal. The workshops will be followed up with a questionnaire to document the user's experience. The moderated mailing list and CBEO Wiki will be a forum for users and PIs to share practices and experience on a continual basis, with rapid communication and feedback.

Educational specialists and historically under-represented groups in science and engineering will be preferentially included in the training workshops through the ESEP, COSEE-MA and MAST educational programs (V-A). In addition to the design workshop, the science education specialists will hold workshops and educator courses (V-A). Participants from an on-going outreach project at JHU (BIGSTEP) will participate in MAST cruises and workshops in order to bring results of this project into other national programs of outreach to native Americans and others, e.g. through the MATIES program described in V-B.

IV G. Contractual Services

During this project we will be securing resources, services and expertise of a number of groups who are the current custodians and collectors of significant databases or model results. Although all the data we anticipate using is in the public domain, we need individualized help in data downloading, translation, creating metadata and, perhaps, altering the web access methods currently used by the data providers. Subcontracts for these services will be issued to C. Cerco (WES, Army CoE) for the 15 yr model-data, B. Michaels (Md DNR), to facilitate collaboration on analysis of shallow water monitoring data, H. Wang (VIMS) for collaboration on use of existing shallow water model output, M. Li (UMCES) for collaborating on use of existing output of ROMS model that includes alternative vertical turbulence closure schemes, and W. Boicourt (UMCES) for collaboration in use of Chesapeake Bay Observing System. Cooperative agreements with government agencies, which cannot accept NSF originated funds, have been secured from NOAA PORTS, NESDIS and NDBC. The subcontracting will be handled by the CRC, advised by the Project Team Leaders.

IV H. CBEO Development Schedule

The CBEO development schedule is presented in Table 4. The more detailed tasks are listed in Sections IV and V.

Table 4. CBEO Development Schedule

Yr	CBEO:S	CBEO:T	CBEO:N	CBEO:E
1	Data set assembly. Initial testing and analyses: model-data comparison: 15 yrs	Data sets: model and observations. Configure the test bed. Implement joining operations.	Create metadata and ontologies for GEON	Hold 1st user workshop near the end of the year for CBEO:T
2	Flux analyses; explore shallow-deep relationship wrt net productivity	Interpolation: flow field driven. Model data storage as polyhedra	Initial GEON datasets. Integrate tools and workbench	Hold 2nd user and 1st educational work-shop near the end of the year for CBEO:T
3	Conduct tests on the networked CBEO	Transfer test bed to GEON	Integrate CBEO:T tools into CBEO:N	Hold 3rd user and 2nd educational work-shop for CBEO:N

V. Broader Impacts

V A. CBEO:E —Multicultural Student Development and K-12 Outreach

Recent national media reports on the problem of hypoxia or "dead zones" have raised public awareness of this issue, yet few citizens understand its causes and implications. This proposal provides the unique opportunity to translate the science and technology associated with a high visibility research theme to educate the public. We propose a multi-year, multi-tiered approach to meet the broader impact of this research. We will develop partnerships among scientists, science educators, teachers, and minority students to convey the information to a broad and diverse audience. In our efforts to foster the integration of research and education through this research, we will leverage several existing programs.

The UMCES Teacher Research Fellowship Program, currently in its 5th year, is conducted by science education specialists from UMCES (L. Murray), Maryland Sea Grant (MDSG), and the U. of Maryland Biotechnology Institute's Center of Marine Biotechnology (COMB), who form a collaborative working group: *UMCES – Maryland Sea Grant Environmental Science Education Partnership* (***ESEP***). In this program teachers work directly with scientists to enhance their understanding of science concepts and to develop classroom applications built on that work. The program includes an intensive summer experience, extensive follow-up and support through the academic year, and a comprehensive dissemination effort. We will fund a Teacher Research Fellow to work with scientists to integrate the proposed research into education as described below.

We will coordinate with the NSF's sponsored Center for Ocean Science Excellence, Mid-Atlantic (***COSEE-MA***), whose goal is to integrate research and education programs to encourage lifelong learning experiences for everyone. We developed a course for teachers to bring ocean science to education, using a platform of ocean observing systems. We will integrate hypoxia, its causes and implications to estuaries and coastal oceans, into our unit on ecosystem health. This will be facilitated by ***COSEE-MA*** Co-PI's and course co-leads, Laura Murray (UMCES) and Deidre Gibson (HU).

Hampton University's ***MAST*** (Multicultural students At Sea Together) project is run by B. Cuker (Cuker 2003). ***MAST*** combines the study of marine science, policy, the heritage of African Americans and Native Americans on the Chesapeake and under sail, and seamanship. It involves a diverse crew (70 participants since 2000, 60% African American, 28% Hispanic, 10% Native American, 2% Pacific Islander, 1% white) of undergraduate and graduate students from across the nation in a month long cruise of the Chesapeake aboard a 53-foot sailing vessel. ***MAST*** focuses on oxygen depletion in the Bay using a Seabird CTD sensor array, to collect data on oxygen, temperature, salinity, and chlorophyll a in the main-stem at 20 stations. The mid-June sampling data records the declining oxygenation of bottom waters over the last half decade as well as increasing levels of Chlorophyll a in surface waters. The ***MAST*** students will link with CBEO partners by building the ***MAST*** data sets into the greater project during visits with CBEO partners at three ***MAST*** ports of call; Baltimore, SERC and Horn Point Laboratory. One graduate student from the Hall-Bonner Program for Minority Doctoral Scholars in Ocean Sciences will use these data toward his or her dissertation. The Hall-Bonner program (NSF-supported, run by B. Cuker) is a "critical mass" program (Cuker 2001) for minority students seeking degrees Ph.D's in marine science. Presently a MS student at Hampton University, Ms. Jihan Davis is using the ***MAST*** data for her MS thesis, and Ms. Davis will likely continue this work for her Ph.D. under the Hall-Bonner program, partially supported by this project. The Hall-Bonner student will also work closely with the lead-teacher in L. Murray's and D. Gibson's COSEE program to (a) learn effective strategies for interacting with K-12 students and (b) help the lead-teacher master the scientific concepts of the project.

In the first year, we will work with a Teacher Research Fellow (from the ***ESEP*** program) on hypoxia and the associated cyber-infrastructure. The teacher will have computer skills and some experience with observing system science (e.g. ***COSEE-MA*** teacher). The teacher will develop a research project that will also yield a classroom application for K-12, and which will be tested in the teacher's classroom during the school year. The teacher will also help guide the visualization of the data in the overall project to assure that it will be applicable to general public use. The teacher will also work directly with the Hall-Bonner graduate student as described above. At the end of the first year, we this teacher will become a "master teacher", who will partner with the project for the next two years to help disseminate information on the program.

In years 2 and 3 the master teacher will expand and finalize the classroom application and continue to work with the Hall-Bonner graduate student on data visualization and analysis. Formative and summative evaluations of the K-12 students will serve to test the success of the application. We will post the application on the ***ESEP*** and ***CBEO*** websites, and disseminate it through the National ***COSEE*** network. The master teacher will also work with a graduate student and high school teacher from JHU's ***BIGSTEP*** program (described below) to expedite research-to-education by training them in the use of the water quality related classroom applications. The teacher will work with L. Murray, the graduate student, and other teachers to conduct two 2-day workshops per year for peer teachers to train them in the use of the classroom application. We will expand the application to non-formal educators employed by the Chesapeake Bay Foundation through the *Chesapeake Bay Ecology* class, taught annually by L. Murray. The CBF educators, will, in turn, use the application to teach their field trip students about hypoxia in the Chesapeake Bay. We also expect that the ***BIGSTEP*** teacher and graduate student will be able to bring project concepts into their future classes.

At JHU, an existing grant – "Broader Impact from Graduate Students Transferring Engineering Principles (***BIGSTEP***) to K-12 Education" (L. Abts, Exec. Director) will assign one doctoral Fellow for each of the first two years of the CBEO program. (See attached Letter of Support.) The Fellow will work with teachers from schools in Baltimore and Native American reservations in Minnesota toward the objective of developing a "kit" of soil and water quality sensors to address science, technology, engineering and math (STEM) concepts. The proposed CBEO programs would gainfully expand the types of data that could be monitored, and these will be used as prototypical data for some K-12 visualization software that is under development. In the latter regard, H. Boussalis (Elec. & Comp. Eng'g, Cal State LA) is a ***BIGSTEP*** collaborator who is working to develop a portal that will collate and present information from remote and dispersed sensors for display in formats appropriate for intellectually challenging students from a wide range of cultures and with varied learning styles. This tool will be shared with L. Murray, D. Gibson, and the Master Teacher for possible workshop use and feedback. The JHU team also works closely with D. Gourneau, a ***BIGSTEP*** collaborator and Chairman of the Board of Trustees of the Smithsonian National Museum of the American Indian, to develop strategies for incorporating distributed information into programs of the American Indian Science and Engineering Society (AISES). We will explore the integration of CBEO results with this program.

Finally, this project can also interface with the University of Delaware's award-winning Research-Based Education for Undergraduates (REU) project and its two programs that support minority and under represented students: ***RISE*** (Resources to Insure Successful Engineers) and ***NUCLEUS*** (Network of Undergraduate Collaborative Learning Experience for Underrepresented Scholars) programs. These students begin their studies the summer prior to, or upon completion of, their freshman year by taking a Summer Session course and pursuing research. ***RISE*** is among the nation's oldest and most successful minority engineering programs, designed to recruit and retain academically prepared native-born African American, Hispanic American, and Native American students.

V B. Broader Impacts on Science and Management

Programs of resource management (such as the Chesapeake Bay Program) typically evaluate land use and other management options based on modeled responses of the ecosystems to proposed changes. The fundamental processes governing hypoxia are not, however, adequately incorporated into current models for the Bay and, it must be presumed, other estuarine systems. The CBEO will help rectify this by providing better evaluation of past observational data and modeling results, and thus improved understanding of hypoxia. The prototypes and methods of the CBEO will also benefit a wide variety of environmental systems where integration of disparate data sets is needed and where improved CI can shed new insight to long-standing problems. Finally, the impacts of this work on the application of cyberinfrastructure to environmental observatories are enormous, as described in this proposal.

References

Baru, C. (2005) GEON: The GEON Grid Software Architecture. ESRI International User Conference, San Diego, August 2004.

Baru, C., Zaslavsky, I., Wahadj, R. (2005). System Architecture, Chapter 3 in *CUAHSI Hydrologic Information System Status Report,* Version 1, September 15, 2005.

Bermudez, L., Piasecki, M., (to appear 2006) "Community Metadata Profiles for the Semantic Web", *Geoinformatica Journal*, Springer, accepted for publication, March 2005.

Boesch DF, Brinsfield RB, Magnien RE (2001) Chesapeake Bay eutrophication: Scientific understanding, ecosystem restoration, and challenges for agriculture. *J Environ Qual* 30: 303-320

Boicourt WC (1992) Influences of circulation processes on dissolved oxygen in the Chesapeake Bay. In Smith DE, Leffler M, Mackiernan G (eds.) Oxygen dynamics in the Chesapeake Bay. A synthesis of recent research. *Maryland Sea Grant Pub*l, College Park, p 7-59

Brush MJ, Brawley JW, Nixon SW, Kremer JN (2002) Modeling phytoplankton production: problems with the Eppley curve and an empirical alternative. *Mar Ecol Prog Ser* 238:31-45

Cerco CF (1995) Simulation of long-term trends in Chesapeake Bay eutrophication. *J Environ Eng ASCE* 121: 298-310

Cerco CF, Cole T (1993) 3-dimensional eutrophication model of Chesapeake Bay. *J Environ Eng ASCE* 119:1006-1025

Cerco CF, Meyers M (2000) Tributary refinements to Chesapeake Bay model. *J Environ Eng* 126: 164-174

Cerco CF, Noel MR (2004) Process-based primary production modeling in Chesapeake Bay. *Mar Ecol Prog Ser* 282:45-58

Chamberlin, D. (1998). *A complete guide to DB2 Universal Database*. Morgan Kaufmann Publishers, Inc., San Francisco, California..

Chin, T. M., & Mariano, A. J. (1997). Space-time interpolation of oceanic fronts. *IEEE Trans. Geoscience and Remote Sensing*, 35(3), 734-746.

Cloern JE (2001) Our evolving conceptual model of the coastal eutrophication problem. *Mar Ecol Prog Ser* 210: 223-253

Cornwell JC, Sampou PA. (1995) Environmental controls on iron sulfide mineral formation in a coastal plain estuary. In Vairavamurthy MA, Schoonen MAA (eds) *Geochemical transformations of sedimentary sulfur*. American Chemical Society, p 224-242

Cronin WB, Pritchard EW (1975) Additional statistics on the dimensions of the Chesapeake Bay and its tributaries: Cross-section widths and segment volumes per meter depth. *Special Report 42*, Chesapeake Bay Institute, The Johns Hopkins Univ., Ref 75-3, Baltimore, MD

Cuker, B.E. (2001). Steps to increasing minority participation in the aquatic sciences: Catching up with shifting demographics. *Bulletin of the American Society of Limnology and Oceanography*. **10**:17-21.

Cuker, B.E. (2003). Minorities At Sea Together (MAST): A model interdisciplinary program for minority college students. *Current, The Journal of Marine Education*. **18**:45-51.

Diaz RJ (2001) Overview of hypoxia around the world. *J Environ Qual* 30: 275-281

EML (2003), Ecological Metadata Language, accessed January 2006, <http://knb.ecoinformatics.org/software/eml/>

Grid Account Management Architecture (GAMA). Accessed January 2006. <http://grid-devel.sdsc.edu/gridsphere/gridsphere?cid=gama>

Gibson, D., & Spann, M. (2003). Robust optical flow estimation based on a sparse motion trajectory set. *IEEE Trans. Image Processin*g, 12(4), 431-445.

Gross et al. (2005). Chesapeake Community Modeling Program, wwww.chesapeake.or/ChesCommModel.htlm.>

Hagy JD, Boynton WR, Keefe CW (2004) Hypoxia in Chesapeake Bay, 1950-2001: Long-term change in relation to nutrient loading and river flow. *Estuaries* 27:634-658.

Harding LW, Itseire EC, Esaias WE (1994) Estimates of phytoplankton biomass in the Chesapeake Bay from aircraft remote sensing of chlorophyll concentrations, 1989-92. *Remote Sens. Environ*. 49:41-56

Harding LW, Magnuson A, Mallonee ME (2005) SeaWiFS retrievals of chlorophyll in Chesapeake Bay and the mid-Atlantic bight. *Estuar Coast Shelf Sci* 62:75-94

Howarth R, Anderson D, Cloern J, Elfring C, Hopkinson C, Lapointe B, Malone T, Marcus N, McGlathery K, Sarpley A, Walker D (2000) Nutrient pollution of coastal rivers, bays, and seas. *Issues in Ecol* 7:1-15

Johnson BH, Kim KW, Heath RE, Hshieh BB, Butler HL (1993) Validation of a three-dimensional hydrodynamic model of Chesapeake Bay. *J Hydraulic Eng'g* 119:2-20.

Joye SB, Hollibaugh JT (1995). Influence of sulfide inhibition of nitrification on nitrogen regeneration in sediments. *Science* 270: 623-625

Kemp WM, Batiuk R, Bartleson R, Bergstrom P, Carter V, Gallegos G, Hunley W, Karrh L, Koch E, Landwehr J, Moore K, Murray L, Naylor M, Rybicki N, Stevenson JC, Wilcox D (2004) *Habitat requirements for submerged aquatic vegetation in Chesapeake Bay: Water quality, light regime, and physical-chemical factors. Estuaries* 27:363-377.

Kemp WM, Sampou P, Caffrey J, Mayer M, Henriksen K, BoyntonWR (1990) Ammonium recycling versus denitrification in Chesapeake Bay sediments. *Limnol Oceanogr* 35: 1545-1563

Kemp WM, Smith EM, Marvin-DiPasquale M, Boynton WR (1997) Organic carbon-balance and net ecosystem metabolism in Chesapeake Bay. *Mar Ecol Prog Ser* 150: 229-248

Kemp, W.M., W.R. Boynton, J.E. Adolf, D.F. Boesch, W.C. Boicourt, G. Brush, J.C. Cornwell, T.R. Fisher, P.M. Glibert, J.D. Hagy, L.W. Harding, E.D. Houde, D.G. Kimmel, W.D. Miller, R.I.E. Newell, M.R. Roman, E.M. Smith, and J.C. Stevenson. 2005. Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. *Mar. Ecol. Prog. Ser. 303*: 1-29.

Larson, R., Conradsen, K., & Ersboll, B. K. (1998). Estimation of dense image flow fields in fluids. *IEEE Trans. Geoscience and Remote Sensing*, 36(1), 256-264.

LEAD, (2005), Linked Environments for Atmospheric Discovery, accessed January 2006, <http://lead.ou.edu/>

Ludaescher, B., Lin, K., Bowers, S., Jaeger-Frank, E., Brodaric, B., Baru, C., (2006), Managing Scientific Data: From Data Integration to Scientific Workflows, *GSA Today, Special Issue on Geoinformatics*, to appear.

Malik, T., A. S. Szalay, A. S. Budavri, and A. R. Thakar. SkyQuery (2003). A WebService approach to federate databases. In *Proceedings of the Conference on Innovative Data Systems Research*.

Malik, T., R. Burns, and A. Chaudary (2005). Bypass caching: Making scientific databases good network citizens. In *Proceedings of the International Conference on Data Engineering*.

Malone TC, Boynton W, Horton T, and Stevenson C (1993). Nutrient loading to surface waters: Chesapeake case study. In Uman MF (ed) *Keeping pace with science and engineering*. National Academy Press, p 8-38

Malone TC, Crocker LH, Pike SE, Wendler BW (1988) Influences of river flow on the dynamics of Phytoplankton production in a partially stratified estuary. *Mar Ecol Progr Ser 48*: 235-249

MD DNR (2005). Maryland Department of Natural Resources "Eyes on the Bay" web page, <www.eyesonthebay.net>, accessed on January 15, 2006.

Memon, A., C. Baru, I. Zaslavsky, S. Mock, A. Behere (2004). "GEON: Standards-based secure invocation of Arcweb services," ESRI International User Conference, San Diego, August 2004.

MMI (2006), Marine Metadata Initiative at the Monterrey Bay Aquarium research Institute, MBARI, accessed January 2006, <http://marinemetadata.org/>

myGEON (2006), accessed in January 2006, http://www.geongrid.org/mygeon.html

myLEAD (2006), accessed in January 2006, http://www.cs.indiana.edu/~machrist/lead-portal/getting_started.html, or http://lead.ou.edu/

Nixon, S.W. (1995) Coastal marine eutrophication: a definition, social causes, and future concerns. *Ophelia* 41: 199-219

OCEAN.us (2005), Data Management and Communications Plan (DMAC) for Research and Operational Integrated Ocean Observing Systems (IOOS), March 2005, <http://dmac.ocean.us/dacsc/imp_plan.jsp>

Pihl L, Baden SP, Diaz RJ, Schaffner LC (1992) Hypoxia-induced structural changes in the diet of bottom-feeding fish and Crustacea. *Mar Biol* 112: 349-361

PMEL, (2005), FERRET: an interactive computer visualization and analysis environment, developed at the Pacific Marine Environmental Laboratory, accessed January 2006, <http://ferret.pmel.noaa.gov/Ferret/>

Pritchard DW (1967) Observations of circulation in coastal plain estuaries. In Lauff GH (ed) *Estuaries*. American Assoc Advancement of Science, p 37-44

Rabalais, N. N., R. E. Turner and D. Scavia (2002). Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River. *Bioscience*. 52: 129-142.

Ritter, C. and P. A. Montagna (1999). Seasonal hypoxia and models of benthic response in a Texas Bay. *Estuaries*. 22: 7-20.

Roman, M, X. Zhang, C. McGilliard and W. Boicourt (2005). Seasonal and annual variability in the spatial patterns of plankton biomass in Chesapeake Bay. *Limnol. Oceanogr.* 50: 480-492.

Rosenberg R. (1990). Negative oxygen trends in Swedish coastal bottom waters. *Mar Pollut Bull* 21:335-339

Schubel JR, Pritchard DW (1986). Responses of upper Chesapeake Bay to variations in discharge of the Susquehanna River. *Estuaries* 9: 23

SDSC, (2006a), The Geosciences Network, GEON: Building CyberInfrastructure for the GeoSciences, accessed January 2006, <http://www.geongrid.org/>

SDSC, (2006b), http://river.sdsc.edu/NWISTS/nwis.asmx http://river.sdsc.edu/HDAS

Szalay, A. S., J. Gray, A. R. Thakar, P. Z. Kunszt, T. Malik, J. Raddick, C. Stoughton, and J. vandenBerg. (2002). The SDSS Skyserver: Public access to the Sloan Digital Sky Server data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*,.

Ueng, S.-K., & Wang, S.-C. (2005). Interpolation and visualization for advected scalar fields. In *Visualization, IEEE* 2005.

UNIDATA, (2005), Integrated Data Viewe (IDV), <http://www.unidata.ucar.edu/software/metapps/>

Yang, Q., & Parvin, B. (2003). High-resolution reconstruction of sparse data from dense low-resolution spatio-temporal data. *IEEE Trans. Image Processing*, 12(6), 671-677.

Zaslavsky, I., A. Memon, G. Memon (2005a). "Integration across heterogeneous spatial data and applications within a large cyberinfrastructure project", *GIS Development*, May 2005, 9(5).

Zaslavsky, I., C. Baru, K. Bhatia, A. Memon, P. Velikhov, V. Veytser (2005b). "Grid-enabled mediation services for geospatial information". In *Next Generation Geospatial Information. From Digital Image Analysis to Spatiotemporal Databases*, ed. by P. Agouris & A. Croitoru), Taylor & Francis, 2005, pp. 15-24.

Zhang, Y., & Akambhamettu, C. (2003). On 3-D scene flow and structure recovery from multiview image sequences. *IEEE Trans. Sys. Man and Cybernetics B.*, 33(4), 592-606.