Environmental Informatics

David R. Richards [1], Cary D. Butler [2], Robert M. Wallace [3], and Jeffery P. Holland [4]

[1] Technical Director for Computational Science and Engineering, Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS, David.R.Richards@erdc.usace.army.mil

[2] Technical Director for Information, Assurance, and Interoperability, Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS, Cary.D.Butler@erdc.usace.army.mil

[3] Research Hydraulic Engineer, Coastal and Hydraulics Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS, Robert.M.Wallace@erdc.usace.army.mil

[4] Laboratory Director, Information Technology Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS, Jeffery.P.Holland@erdc.usace.army.mil

Abstract

This paper defines the term "informatics" as it pertains to the environmental science and engineering activities of the U.S. Army Engineer Research and Development Center (ERDC). It dissects informatics into its component parts to encourage understanding of the terminology and to describe how ERDC will use informatics ideas in the research, development, and fielding of science and engineering products in the future. The informatics components discussed include: 1) integrated frameworks; 2) data, data fusion, aggregation, management, and mining; 3) modeling and assessment; 4) decision support; 5) knowledge management; and 6) GIS/CADD. A discussion is provided that shows how a well-designed informatics plan can increase the relevance and sustainability of our environmental products within ERDC and the Corps of Engineers.

Background

The word "informatics" is a new invention in the English language like many others that come from the information science field. It is typically used to describe the application of information technology to solve societal problems. Typical examples of the societal problems come from the biological, sociological, hydrological, and geotechnical sciences. Practitioners in these respective fields have used the terms "bio-informatics," "socio-informatics," "hydro-informatics," and "geo-informatics" to describe their specific interest in informatics. Equivalently, the terminology "environmental informatics" is an invention that describes the informatics technologies and tools that are necessary for supporting our environmental mission in the Corps of Engineers. However, in the

interest of brevity, informatics will be used hereafter to describe the entire environmental informatics area.

Within the context of the U.S. Army Corps of Engineers, the concept of informatics can be applied in a number of areas. First, like most business operations, informatics is involved with and has revolutionized the day-to-day operation or practices of the Corps. For instance, we communicate using e-mail rather than writing letters, and we use a database management system (DBMS) system to allocate and appropriate funds. This trend will only increase and is not engineering specific in nature.

A second and more intriguing concept of informatics, with regard to the primary purposes of the U.S. Army Engineer Research and Development Center (ERDC), is the use of computers to model and simulate physical phenomenon. A primary mission of the ERDC laboratories over the last 25 years has been to create computer programs that simulate hydraulic, hydrologic, geotechnical, structural, and environmental processes so that better planning, engineering, and operations decisions can be made for Corps of Engineers projects. ERDC has made significant informatics contributions in these technical fields as evidenced by the large numbers of computational programs and practices used by the Corps and commercialized for the private sector.

A third aspect of informatics, one that is critical to the Corps, is the area of engineering management decision making. This area of informatics includes the research areas of visualization, automation, and decision support. This area is particularly important to the Corps because it is the primary method for analyzing and disseminating information developed by simulation-based informatics tools. The Corps has been a leader in this area and has developed a number of visualization and user support systems that are in wide use throughout the United States and the rest of the world. This aspect of informatics is also the most laden with opportunities because the science behind the decision support and automation tools is not as mature as in the area of simulation.

Different groups within the Corps have been quite successful in developing and fielding informatics products even before this terminology was invented. Rapidly evolving information technologies such as the Web and modern object-oriented programming languages and practices make it possible to achieve new levels of productivity in our technical and business activities. However, we must be mindful to pursue synergistic informatics developments among Corps development groups if we hope to develop the rich and efficient informatics toolset that is essential to perform our Corps mission. Having this dialogue on the definition and structure of informatics is an early step in this development process. The informatics development process is rapidly maturing within the ERDC research and development program and it is our hope that reporting on ERDC informatics developments will be useful to the Corps as a whole.

Components of Informatics

In an effort to manage informatics developments at ERDC, informatics has been organized into the following components: 1) Integrated Frameworks; 2) Data, Data Fusion, Aggregation, Management, and Mining; 3) Modeling and Assessment; 4)

Decision Support; 5) Knowledge Management; and 6) GIS/CADD. Figure 1 provides a symbolic description of how the different components of informatics are related. Integrated frameworks, at the base of the pyramid, is shown providing the foundation upon which all of the other components are based. Data occupies the first level of the pyramid followed by more highly evolved components ranging from modeling and assessment to knowledge management. The arrow depicts the flow of informatics activities from gathering raw data to assembling information to acquiring knowledge. It is important to note that GIS and CADD are not separate layers but rather are pervasive throughout the components of informatics. They are so pervasive in our work that they deserve special consideration and treatment if we are to manage these development activities effectively.

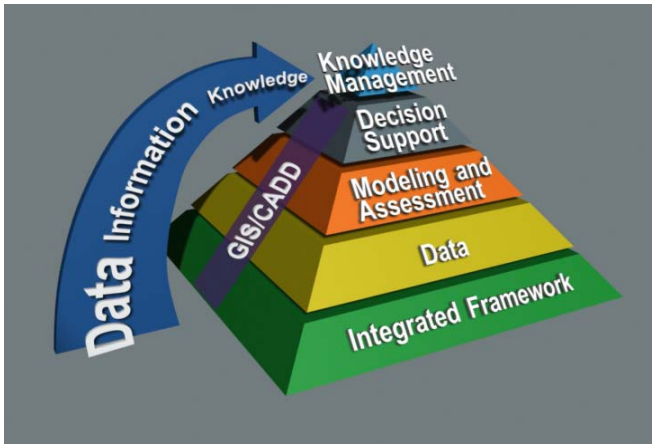

Figure 1: Components of Informatics

Definitions for the informatics components are provided below. ERDC examples are provided for illustration, but there are undoubtedly other examples from other groups in the Corps and elsewhere.

Integrated Frameworks

Integrated frameworks provide a common technology structure to deliver information and technology. Every ERDC computational science and engineering product (data, tools, and models) requires a technical architecture that supports the packaging, delivery, and infusion of ERDC research. Establishing commonalities in the technical architecture will improve usability and interoperability as well as reduce the total cost of the product.

Several frameworks are being used throughout ERDC to produce specific capabilities. Currently, one of our military environmental quality efforts is being built on a technical architecture called Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) developed by the Department of Energy (DOE). The Battlespace

Environment program has fully embraced a commercial GIS framework developed by ESRI. Developments in ERDC's Installation Management Program (Fort Future) are built, in part, upon another DOE framework called the Dynamic Information Architecture System (DIAS). Meanwhile, we are investing in a cooperative technical architecture through the Civil Works Geospatial Technologies Program called the Common Delivery Framework, or CDF. While it would be ideal to have only one framework for informatics, it would likely be as effective as having only one motor vehicle type for every purpose regardless of the specific needs of commuters, farmers, and truckers. However, having multiple frameworks does provide technical and institutional challenges, so the purpose of CDF is to provide a common framework that will encourage sharing of components and connectivity between multiple frameworks and tools.

Data, Data Fusion, Aggregation, Management, and Mining

Data are a key ingredient to our decision processes, and the organization, analysis, presentation, and effective reuse of data may be the most important of informatics challenges. This component focuses on a common set of methodologies that locate, collect, manipulate, and use data in support of ERDC product lines. The effective use of data requires establishing a formal database structure and the consolidation of disparate information sources for the purpose of discovering useful information and ultimately for driving higher-level informatics tools.

Improvements in how ERDC, the Corps, and other agencies corporately deal with information increase our efficiency in data assembly and processing. Data-related activities make up a substantial part of most modeling efforts. Efficiencies gained here have the potential for improving the accuracy of our modeling efforts and reducing the costs of studies by over fifty percent. Dealing with data in a corporate way will require the collaboration of multiple government agencies and require synergistic developments in both our military and civil works mission areas. Promising areas for collaboration include developing standards and common tools for analyzing terrain, hydrographic, weather, soils, water quality, and biological data.

Modeling and Assessment

Modeling is the development of physical, analytical, or numerical analogues of observable phenomena. Physical models are smaller scale replicas of full-scale problems. Analytical models are mathematically solvable depictions of physical phenomena. Numerical models are used to solve complex mathematical equations not easily solvable by analytical methods. All of these model types are used to simulate observable phenomena at various time and space scales and to assess the impacts of projects on the environment. While modeling and assessment can be viewed as decision-support activities, they typically employ the highest-level types of analysis in the decision- support world. Modeling and assessment are typically process driven activities using complicated mathematical depictions of physical processes that require numerical analyses and computers.

Developing and applying models and assessment tools are essential steps in capturing and advancing our understanding of processes (or systems) in scientific domains. Models and assessment tools are important in forecasting (and hind casting) potential outcomes of proposed actions, and as aids to inform various decision processes. While specific properties captured in models (whether physical, biological, or mathematical) differ with each application area, there are many technical issues that span our business areas, such as 1) methods for representing processes and process interactions; 2) methodologies for code development, partitioning, and reuse; 3) computational procedures and resources; and 4) interoperations among related models. Savings of time and funding can be gained through improved coordination of model development activities within and across application areas.

Specific informatics needs for modeling and simulation include Web-accessible boundary and initial condition data, graphical user interfaces to run the models and make assessments, and publishing capabilities (preferably on the Web) to share model information with stakeholders.

Decision Support

Decision support includes the processes and associated tools that are used to assist decision-making at all levels of analysis. This includes multiple communities of practice that includes but is not limited to science and engineering, planning, operations, financial, legal, and administrative. Decision support, in each community, can be categorized by the technical level of analysis varying from lower levels of analysis that are useful for screening level decisions to higher levels of analysis that are required for final designs and specifications. In the science and engineering community, low levels of analysis could include crude index level models using GIS while higher levels could include multi-dimensional finite-element analysis of dynamic systems. Although decision support includes a wide range of tools, it is commonly thought of as the lower levels of analysis and supporting informatics tools as differentiated from sophisticated modeling and simulation activities. Decision-support tools are typically data driven using empiricism or simple analytical solution techniques.

One of the most important aspects of this component is the ability to optimize our ability to collect and synthesize information from both the data and the modeling and assessment components. The technology required to provide decision-support activities to decision makers should be minimized to decrease the training burden on the user base. Ideally, decision-support capabilities would be distributed via the Web (or an Intranet), thereby requiring no more than a simple Web browser to access our decision-support capabilities.

Knowledge Management

ERDC products exist in many forms, such as reports, tools, models, databases, Web pages, etc. Knowledge management (KM) provides the technical infostructure that builds a collaborative information-rich environment among ERDC customers, employees,

and business partners. KM is about Web-enabling strategic processes and applications to increase business responsiveness and innovation. KM enhances the value of ERDC to its customers by: (a) empowering users by providing fast access to relevant information; (b) increasing overall productivity through faster-better decision making; (c) facilitating collaboration, work flow, mentoring and training through knowledge-based communities of practice; and, (d) maintaining and/or increasing organizational competitiveness by leveraging internal and external information more effectively. While KM is important for our organization to pursue (especially when we need to capture the institutional knowledge of retiring employees), we have not expended significant developmental resources in this area. However, the Army's Online Knowledge Center (OKC) provides an example of how KM can be implemented.

GIS/CADD

The pervasiveness of spatial data used throughout the science and engineering community motivates the need to corporately address GIS/CADD. There are numerous GIS/CADD applications in both open source and proprietary forms and, as in the integrated frameworks example, there is no one solution that is ideal for all users. However, there is a need to limit the number of GIS/CADD applications and to have them communicate effectively. Standards emplaced within the GIS/CADD area allow ERDC programs to share and reuse GIS/CADD data and the supporting functionality to visualize, manipulate, analyze, and display geospatial information. The ability to expand modeling and decision support into 2-, 3-, and 4-dimensional space/time, as provided by geospatial technologies, will tremendously enhance the products available to our customers.

Environmental Drivers for Improved Informatics Capabilities

The Corps of Engineers is attempting to implement the Chief's environmental operating principles in our business practices. This has often necessitated a holistic approach to environmental resource management where previously we might have taken a more narrow view of the environmental landscape. Given the size and importance of typical Corps projects, a holistic view could mean modeling the entire Mississippi River if problems with eutrophication are to be solved in the Gulf of Mexico. Equivalently, from a process point of view, attempting to restore the Everglades requires modeling both surface water and groundwater in a coupled fashion for all of South Florida as opposed to just the individual projects that together comprise South Florida restoration. Marsh restoration in coastal Louisiana and improving conditions for salmon on the Columbia River are additional examples of regional scale problems that are important to the Corps.

The complexity and size of these and other restoration efforts require fresh thinking as to how we are going to manage the environmental resources and what informatics tools will be required. It is clear that we will be handling large data sets and performing analyses for large geographic regions. We will be analyzing and publishing a variety of physical, biological, and chemical data over and on the Web using a variety of geospatial tools. We will also be using high-performance computers to solve the largest and most complex

problems but will need to perform visualization and decision-support activities on personal computers. It is also clear that our stakeholders will want access to our tools and data over the Web. These requirements necessitate the development of an informatics strategy for our research and development activities.

ERDC Informatics Research and Development Activities

ERDC is currently pursuing a corporate informatics strategy across our research and development business areas and in some cases our larger reimbursable studies. Participation in the informatics process has been voluntary and as a result has resulted in a slowly growing but committed group of informatics collaborators. There are advantages to this evolutionary, as opposed to a revolutionary, approach. While every program is not yet completely committed to a corporate informatics development process, the advantages to participation are becoming obvious to all, and the numbers of collaborators are destined to increase dramatically. The ERDC research programs that are actively involved in the informatics strategy (with their primary areas of contribution) are listed below:

- Geospatial Technologies Program (GIS and Decision Support)
- Common Delivery Framework Program (Frameworks, Databases, Decision Support, Modeling and Assessment, and Knowledge Management)
- Regional Sediment Management Program (Frameworks, Databases, Modeling and Assessment, and Decision Support)
- Coastal Inlets Research Program (Frameworks, Databases, Modeling and Assessment, and Decision Support)
- Civil Works Project Security Program (Frameworks, Databases, Modeling and Assessment, and Decision Support)
- Installation Restoration Research Program (Frameworks, Databases, Modeling and Assessment, and Decision Support)
- Common High Performance Computing Software Support Initiative (Integrated Regional Scale High Performance Computing Modeling)
- System-wide Modeling, Assessment, and Restoration Technologies Program (Regional Scale Watershed Modeling)
- Dredging Operations and Environmental Research Program (Dredged Material Fate Models)

Each of these participating programs is actively involved in joint developments and will share in the synergism that such collaboration brings. For example, the CDF will be used to connect all of the data, modeling, assessment, and decision-support tools developed across programs and provide a common way to access, analyze, and publish data on the Web. Web services developed in the CDF program can be easily reused by all programs with little effort. Parts of the Army Risk Assessment Modeling System (ARAMS) developed in the Installation Restoration Research Program can be reused to provide risk assessments to civil works programs as well as military programs. Regional scale hydrology, hydraulics, and sediment transport models (and their graphical user interfaces) developed in the Regional Sediment Management Program can be used by several of the

other programs both military and civil works. There are limitless examples of how such collaboration in informatics development can increase the capabilities of our tools and actually decrease their development and maintenance cost.

While we have primarily been discussing research and development program involvement in informatics development, it is important to recognize the role that reimbursable projects from Corps Districts and other government agencies have played in fielding many of our informatics products. The Army Environmental Center, Department of Energy, Department of Transportation, Environmental Protection Agency, and Nuclear Regulatory Commission have been significant informatics contributors. Their involvement in our developments is very well received today when we are encouraged by the Office of Management and Budget to improve the efficiencies with which we develop automated and other types of information systems.

Vision for Future Informatics Development

Perhaps the most important need for ERDC and the Corps in the future is to think corporately as opposed to individually when designing and developing informatics products. That is, we should use a collaborative process to develop products that solve multiple problems for multiple communities of practice as opposed to using a one-off individual development strategy. This can be difficult to do when funds are earmarked for complete products as opposed to parts of products that depend in turn on the success of other efforts. It is also difficult for individuals to base their success on the ability of others to perform. However, we must reach some level of corporate trust in order to succeed as a whole. The following lists some key elements in our vision for future informatics development.

- Embrace the Web as a means of delivering data, tools, and knowledge to our sponsors and stakeholders.
- Develop tools for multiple communities of practice as compared to one-customer solutions.
- Encourage the development of dual-use technologies that can support both military and civil works missions at the same time.
- Design computer-based products to be interoperable from the beginning as compared to retrofitting them later.
- Use the CDF as the basis for connecting the components of our informatics structure.
- Embrace the use of commercial-off-the-shelf products when they clearly dominate the marketplace and are an essential part of our corporate strategy.
- Standardize Web technologies to encourage reuse and ease of integration.
- Reuse enabling technologies such as visualization, data management, etc., in multiple business areas to reduce the software maintenance footprint and to further promote interoperability.
- Make informatics tools usable across multiple computing platforms from personal computers to high-performance computers.

- Employ life-cycle management processes in developing, maintaining, and sustaining our informatics products.

Conclusions

Environmental informatics and its components have been defined and their importance to the Corps of Engineers explained. A strategy for implementing corporate informatics development has been outlined. If implemented, the synergistic strategy could yield products of high quality and relevance to the Corps and its cost-sharing sponsors. The development costs for using this strategy would be less than funding disconnected individual informatics efforts. An evolutionary process for corporate informatics implementation is currently in place at ERDC and is offered to the larger Corps family in this paper as a discussion topic for future development.