

TITLE: Intelligence Community Initiative in Massive Digital Data Systems

Authors: Hal Curran (NSA), Robert Kluttz (CMS), Dr. Claudia Pierce (NSA),
Dr. Rick Steinheiser (ORD) and Dr. Bhavani Thuraisingham (MITRE)

Related Application Area: Integrated and Transparent Access to Multiple Data Sources

Sponsor: Community Management Staff, Office of Research and Development,
National Security Agency

Description: The Massive Digital Data Systems (MDDS) Initiative is developing high-risk high-pay-off data management technologies for Intelligence Systems of the future. Future Intelligence Systems must effectively manage massive amounts of digital data (i.e., multi-terabytes or greater). Issues such as scalability, design, and integration need to be addressed to realize a wide spectrum of Intelligence Systems ranging from centralized terabyte and petabyte systems comprising many large objects (e.g., images) to distributed heterogeneous databases that contain many small and large objects (e.g., text). Consequently, Massive Digital Data Systems (MDDSs) are needed to store, retrieve, and manage the complex data for the Intelligence Community (IC). While several advances have been made in data management technologies, the complexity and the size of the database, as well as, the needs of the IC, such as multimedia data management, require the development of novel approaches for querying and updating the massive databases. Therefore, the Community Management Staff (CMS) of the IC has started the MDDS Initiative to identify the challenges for massive database management and to develop approaches to meet the challenges. The ultimate vision is to provide for the seamless access and fusion of massive amounts of data, information, and knowledge in a heterogeneous and real-time environment to carry out the functions of the IC with diminishing resources.

To develop massive digital data systems for the IC, several technologies have to be integrated. These include the following: Data management, Knowledge management, Information management, Information visualization, Mass storage management and Distributed processing. The technical focus of the MDDS Initiative is to develop *data management* technologies. These include developing techniques for querying, browsing, and filtering; transaction processing; accesses methods and indexing; metadata management and data modeling; and integrating heterogeneous databases; as well as developing appropriate architectures. Approaches to handle massive databases will be a major focus for this Initiative. Note that secure database management, as well as, other technologies, such as information management and visualization, are beyond the scope of this Initiative. The organizational focus of the MDDS Initiative is to provide seed money to develop data management technologies which are of high-risk and high-pay-off. Further development of the technologies produced under the MDDS Initiative for eventual use by the IC and related organizations is expected to be carried out by other agencies such as the line organizations, as well as, industrial organizations.

The MDDS initiative has awarded fifteen research contracts. These contracts are being executed through the line organizations, such as NSA and CIA. Specific topics being addressed by the research include data mining, data warehousing and heterogeneous database integration, real-time transaction processing in a heterogeneous environment, quality of service management in multimedia databases, indexing multimedia data, scalable architectures, and geographic information systems. The MDDS Team is also coordinating the research with other government agencies by conducting technology exchange meetings.

Scope: Approx: 3 - 4 million dollars per year for 3 - 4 years

Customer: DOD, IC, and other government organizations