The key set of data management issues for MDDS include:

Developing architectures for managing massive databases
Utilizing data models for representing the complex data structures
Formulating and optimizing queries
Developing techniques for concurrency control and recovery
Integrating heterogeneous schemas
Meeting timing constraints for queries and transactions
Indexing multimedia data
Maintaining caches and minimizing secondary storage access and communications costs
Enforcing integrity constraints.

1.2 Background
The IC provides analysis on current intelligence priorities for policy makers based upon new and historical data collected from intelligence sources and open sources (e.g., news wire services, magazines). Not only are activities becoming more complex, but changing demands require that the IC process different types as well as larger volumes of data. Factors contributing to the increase in volume include continuing improvements in collection capabilities, more worldwide information, and open sources. At the same time, the IC is faced with decreasing resources, less time to respond, shifting priorities, and wider variety of interests. Consequently, the IC is taking a proactive role in stimulating research in the efficient management of massive databases and

ensuring that IC requirements can be incorporated or adapted into commercial products. Because the challenges are not unique to any one agency, the Community Management Staff (CMS) has commissioned a Massive Digital Data Systems Working Group to address the needs and to identify and evaluate possible solutions.

1.3 Assumptions and Project Requirements
Future intelligence systems must provide a full suite of services for gathering, storing, processing, integrating, retrieving, distributing, manipulating, sharing and presenting intelligence data. The information to be shared is massive including multimedia data such as documents, graphics, video, and audio. It is desired that the systems be adapted to handle new data types.

The goal is to be able to retain the data for potential future analysis in a cost effective manner. The more relevant data would remain on-line, say for 5 years, organized with the most relevant data accessible in the least amount of time. It is expected that 2 to 5 terabytes of new data has to be processed each day. Thus, the total size of the database (both on-line and off-line) could be as large as 20 petabyes with about 300 terabytes of data stored on-line. It is assumed that storage devices (primary, secondary, and even tertiary) for the large multimedia databases as well as data pathways with the required capacity will exist. The access times are about 5 seconds for the data less than a week