APPLICATION OF SYSTEMS ENGINEERING METHODS TO THE DEVELOPMENT OF INFORMATION RESOURCE MANAGEMENT PROGRAMS

Robert W. Hofer
Effective Solutions Inc.

ABSTRACT

The Civilian Radioactive Waste Management program has presented challenges to information professionals which have never been confronted before. This is the first major research program to require full public access to a vast range of scientific and engineering data. Estimated information system life cycle requirements commensurate with the repository's 10,000 year life are unprecedented, especially when compared with the needs of the commercial nuclear industry. The range of information system users and the differences in the ability of various classes of users to access sophisticated information resources has likewise never been approached in this area before. Consequently, application of previous experience associated with the commercial nuclear power industry information system design falls far short of meeting the information management and dissemination requirements of this program. This paper describes the overall information requirements governing the data and documents associated with the site characterization phase of the geologic repository program. It further outlines a design methodology for the development of an information resource management program which responds to these criteria, emphasizing the use of a systems engineering approach. The paper concludes with a projection of future requirements for information management associated with the geologic repository program and suggested approaches to planning for these needs.

Geologic Repository Information Environment

The geologic repository effort is characterized by a unique information environment. The engineering, construction and field and laboratory research activities are generating enormous volumes and diverse forms of information, many of which have been identified as quality-related. The range of such information readily surpasses that associated with the commercial nuclear power industry. For example, while commercial power plants must complete exhaustive site studies to determine the suitability of a particular location, the samples and types of documentation are far less than those being accumulated to determine the physical characteristics of a geologic repository. Here multiple sites are being subjected to detailed characterization studies using all manner of scientific instrumentation and site exploration activities. The types of documentation and information generated are in all media (including documents, magnetic tapes and other computer storage media, physical samples, photographs, radiographs, and instrumentation charts). The requirements for the collection, administration and dissemination of these various types of information are extensive and also surpass any commercial project's commitment to public disclosure. The information environment has the following major organizational elements:

A. Office of Civilian Radioactive Waste Management

   The thorough implementation of the provisions of the National Waste Policy Act Energy has led to the following general criteria which govern the information environment of the OCRWM:

   - there will be a controlled program established for the collection, administration and dissemination of information generated or collected during the conduct of OCRWM work activities.
   - multiple jurisdictions and organizations have the right to access site characterization and preliminary engineering documents and data as this information is generated or collected. The wide range of such users and the differences in the skills of their

personnel, available information resources, and differing views and use of the data accessed all combine to complicate the information system design process. Regardless of such complexities, the OCRWM is committed to this public access and its information management systems must be fully responsive to the attendant design considerations.

- the complex project organizations which are developing this information resource management programs to facilitate access and dissemination. While the precise information environments of each entity differ in their physical, operational and organizational structure, the information management programs instituted by each must allow for the ready exchange of information; the careful tracking of information origination, destination, and related quality assurance corroboration of information validity and verification; and development of archival retention programs to ensure the ongoing availability of information resources.

B. Nuclear Regulatory Commission

The impact of licensing requirements and application of quality assurance program provisions to information management are a vital consideration. Issues confronting the geologic repository information management program in this area include software quality assurance; data validation, verification, and traceability; imposition of program controls on shared information resources (for example international data collection and management systems relating to the waste repository program), and the retention program to ensure ready access to information throughout the life of the repository.

At an operational level, also important in this area is the impact of the licensing process itself upon the information systems used by the geologic projects. The needs for system resources to support the expedited NRC review process include both a full text storage system to expedite data and document retrieval and distributed information indexing systems which permit to the extent possible full access and discovery of data unconstrained by organizational interpretation or categorization. These distributed systems may require the utilization of some shared resources to fully ensure for all parties information security, flexibility, access, data consistency and data change control. In this latter regard, access to one-of-a-kind items (e.g. core samples) will become a special condern where use by one project participant permanently precludes usage by others.

In terms of software quality assurance and traceability, the impact of licensing upon "closed" or proprietary systems is a vital concern. The system access limitations imposed by vendors in the area of basic input/output software and the impact of this upon utilization of such systems in the licensing area by interested parties which were not directly involved in the initial procurement must be assured.

C. OGR Project Organizations

The requirements confronting the OGR itself within the U.S. Department of Energy, and amoung the OGR and its constituent projects creata a multifaceted information environment. The most immediate concern focuses on the hardware and software utilized by these various participants (and their suppliers, subcon-tractors and consultants), but of equal importance at a more detailed level are the following areas of information management concern:

1. Determination of responsibility for data identification. At what point in an information system life cycle is each data type identified and defined? One critical issue here focuses on the determination of responsibility for the definition of the data types which will be utilized to compare the various candidate sites. Within each project, responsibility for the definition of data types utilized by more than one project participant must be determined as early as possible within the information life cycle to ensure consistent interpretation, usage, reporting transportability of data among the participants. Additional specific concerns in this area are determining reporting levels of measurement, accuracy requirements, number of samples needed to establish a value, reliance upon standard or widely accepted formulas to reduce or convert data, etc. Within each OCRWM project, data definition acceptable in one project area (e.g. site characterization), may not be acceptable in others (e.g. container design or engineeered barriers), and where conflicts exist consistency must be determined to ensure overall project integrity. As indicated above, the same concern exists among the OCRWM projects and between the OCRWM and other involved organizations.

2. Application of management controls to data and documentation during the initial data collection or document preparation activities. The point at which data quality assurance is ensured and data disseminated for use must be clearly understood. To meet the urgent needs of various participants, raw data may be distributed prior to ensuring that all the pre-requisite calibration tests, procedures review, etc. have been reviewed. If this is the case, then deficiency reporting procedures must be in place should subsequent reviews indicate that data was incorrectly gathered or some other irregularity exists concerning data collection. However, to delay in distributing data until after all possible quality reviews have been completed would seriously impair project schedules and appear to be non-responsive to project commitments.

3. Dissemination of work-in-progress to permit parallel or concurrent evaluation. Many of the complex research and design acitivities of the project will extend for many years and

final results will require enourmous amounts of data. Various scientific and engineering disciplines resist premature release of data where all activities have not been completed.

For both professional ethics and public interest, the early release of incomplete data sets can have serious ramifications. Alternatively, to require the public to wait until all tests in a sequence have been completed may significantly inhibit alternative and concurrent analyses from fully contributing to the program. This dilemma must be resolved on a case-by-case basis, and the information systems associated with instances must be responsibe to supporting the data reporting conclusions which are reached.

4. Maintenance of data integrity and traceability in a distributed environment. Where the usage of data is associated with aprallel testing or calculations, the ability of users to "refresh" data sets is mandatory. The "approved values" which are being used by various project participants must be readily available, and where usage may impact the integrity of such data sets, the capability must exist to repeatedly distribute such sets. Moreover, as the project reaches conclusions regarding various data parameters, such "approved values" must be readily disseminated so that all work can proceed in an integrated manner.

5. Distribution of quality assurance requirements and administration of quality control provisions among various organizations whose services have been secured through a variety of procurement processes. Regardless of the level of the project originating or collecting data (e.g. the prime contractors, their suppliers, sub-contractors, or consultants, and in turn the support staff and suppliers for these entities) rigorous quality standards must be imposed to ensure complete data traceability and consistent ability to independently validate all data being utilized.

Application of Systems Techniques to Information Management System Design

Having defined the information management environment, this section outlines how to apply system engineering techniques to develop an information management architecture responsive to the needs of the various users. The focus is upon the development of a methodology which can consistently be applied at both the OGR summary level, and at the project/project participant level.

In keeping with the emphasis of the session, the development of the information system definition is discussed in terms of three inter-related systems necessary in the standard application of the system engineering process:

A. The product (i.e. the information management) system

A fundamental issue in information management system design is to fully appreciate the scope and breadth of the system. Narrowly focusing on the various consistituent parts of an overall information management program, such as "records management", "quality assurance records", "document control", "data administration", "licensing support," etc., while each vitally important in and of itselt, minimizes the overall system and leads to the twin evils of redundancy and omission. A fully understood product here is the totality of all information, regardless of its collection or collection or storage medium. The integration of the various resources is far greater than the sum of the individual parts.

In this regard, the information system cannot be viewed separately from the various scientific- or engineering-related data collection, analysis, conversion, or manipulation systems. There is not a separate information system, standing apart from the OCRWM's many and diverse work activities. The information system is the overall framework in which all these activities are conducted.

B. The organization (the system of people, resources and institutions using the product)

Heretofore in the conduct of OCRWM work activities individual work units have defined their particular information support needs and designed, developed and implemented localized systems. It has been assumed that through the development of an integrated records management program that the outputs of these various systems would be brought together and at some future time form the basis of a total paper-oriented system which would meet the various needs of all other participants. Increasingly this approach has been found to be inadequate. It is mandatory that at the point of initially identifying the need for information that the relationship between the need and all other inter-related project activities be determined. The broadest sample of organizations is needed to ensure that information is available to all users in the form required. This should be one of the primary focuses of the individual project's management integration efforts on the one hand, and the OCRWM's oversight role on the other. No organizational unit can presume to "own" any information. It must be assumed that others will require access and must be included in the initial specification of the information requirement.

C. The environment of the system.

While the previous two considerations were heavily oriented towards establishing the scope of the information system, here the concern is the process of identifying the functions that the information management system must perform, the functional analysis of these actions, the requirements of the system, and the specifications

to implement the system will be described and documented. This includes the steps taken to ensure traceability, verification and baseline control of the information system. The broadest view of the functionality required and the various uses of information must be taken to ensure full support of project activities and preclude the need for future information system design or modifications.

Future Trends in Information Management

The technology available to respond to the information management needs of this program is undergoing rapid and continuous evolution. Those trends which will be most influential upon the program and their impact are summarized below:

A. Optical disk storage for texts and documents.

This technology has the great advantages of high capacity, high density and ready dissemination of data and documents. In keeping with the overall requirements of the program, these characteristics are paramount in evaluating overall information architecture. While definitive results regarding the archival nature of these technologies have not yet been determined, nontheless it can be assumed that these technologies will be a vital part of future program information system design.

However, while the ability to efficiently and economically disseminate vast volumes of information is an obvious benefit of these technologies, the responsibility to provide neutral acces (where the data or documents are accessible by all parties without one party pre-determining relationships or relative importance of individual items-- either implicitly or explicitly) and effective retrieval (precision/recall measurements applied to theoretically unbounded searches) are burdens which must be considered prior to any commitment to this technology. Merely placing n-millions of pages of information in the public domain without giving a viable means of accessing such information does not meet the spirit or intent of the DOE's commitments.

B. Distribution of data and data base capabilities.

Economies in the ability to disseminate project data permit and wide distribution of digital information. However, the maintenance of such dispersed collections and the tradeoffs between their stand-alone access and control, as contrasted to the problems of resource compatibility, configuration management and change control which are better managed through the use of shared resources, are vital concerns in evaluating the best means of supporting the information access needs of various participants.