1.1 INTRODUCTION

Several billion dollars are spent in developed countries every year to investigate and remediate contaminated sites. Investigation/remediation managers seek to understand the technical aspects of their sites under short timelines, and to minimize costs. Regulators impose demanding standards and often complex requirements in an effort to minimize existing and future impacts to the environment, and protect public health. Within this arena, the environmental data collected in support of investigation and remediation efforts is central to the decision making process, and therefore to a successful project outcome. The efficient management and flow of this data is a critical factor in supporting aggressive timelines, complex decisions, and long term monitoring of trends. Moreover, the data is a significant asset of its owners and it is often not recognized as such, thus the value it represents is often neglected or forgotten.

1.2 DATA MANAGEMENT STRATEGIES CONCEPT

Site investigation, remediation, and monitoring projects typically generate data in large volumes and in varying types or formats. In order to efficiently organize and evaluate this data, some form of data management must occur, whether designed specifically for the project, or part of an organization-wide policy.

In previous decades, most data management activities were focused around archival – collating the data for long term storage – and little attention was given to strategies for efficient analysis and reuse. Over the last decade, however, advancements in web technologies, hardware, software, hand-held or other mobile devices for rapid data collection, have transformed the traditional investigation workflow and raised the expectations of consultants and clients alike with respect to turnaround time, 24/7 access to data, visualization options, and integration with ancillary datasets. Moreover, in this era of ‘Big Data’, clients understand that the data generated during a typical project has potential value beyond the initial purpose for which it was collected, and the responsible stewardship of this data as an ‘asset’, and not a by-product is a core requirement of the project, and increasingly, of the client relationship.

Successful data management strategies in today's consulting arena, therefore, involve some if not all of the following key objectives:

- Stewardship of the data throughout the lifecycle of the project, and beyond if required,
- obtaining maximum value from the data (not simply storing bytes in a database),
- open, 'data-centric' policy with respect to access by clients, partner consultants, and maintaining safe options for database transfer on project completion or change of hands,
- integrating information from other sources for critical analysis and decision making, and
- Ensuring that data quality objectives are met and maintained.

Adopting a strategy focused on the stewardship, integration, and exploration of data allows the organization to transform the data management activities that they are required to do anyway, from a tactical, project-specific resource to a strategic asset. This asset in turn provides greater value back to the business by avoiding or reducing some of the common problems that beset the consulting industry (Figure 2.1) such as:

- slow response to change/operational issues,
- sub-optimal decisions based on poor or out of date information,
- degraded data quality resulting in inefficient, incorrect, or duplicated work, and,
- external reputation issues that result from the above.

1.3 DATA MANAGEMENT APPROACH AT ERM

Over the last decade, ERM has worked with clients and industry experts in order to embed a data management strategy within our operations. This strategy includes the following key objectives:

- Collecting and safely maintaining electronic project data across global geographies and technical disciplines,
- Maximizing data integrity and reducing costs,
- Supporting and enhancing decision making processes (i.e. making more informed decisions, faster),
- Enhancing communication of results with visual, easy-to-use, interfaces,
- Sharing data for collaboration with clients, stakeholders, contractors, regulators and partners, etc.,

- Integrating with existing ERM Information Technology processes and entities.

An overview of the resulting infrastructure and workflow is provided below.

Software and Data Flow

ERM uses a standardized workflow for data management activities that allows us to track the flow of data from field to report to archiving, and employs trained data management personnel to oversee activities (Figure 2.2). The approach centers on Relational Database Management Systems (RDBMS) technologies as the core infrastructure, and specifically one of the main tools used is EarthSoft's Environmental Quality Information System (EQuIS), though there are others in use depending on specific geographies and client requirements.

EarthSoft's EQuIS is a relational database management system that can use SQL Server or ORACLE as an RDBMS platform. In a typical configuration it is comprised of desktop software – EQuIS Professional – which is used by trained staff for data upload, modification, and reporting, and of server/web software for hosting the relational database and providing access to reporting and visualization tools for specialized project staff or casual users.

The use of an established RDBMS infrastructure like SQL server or ORACLE means that EQuIS is extendable – outside functions can be used for custom modelling, display, reporting, or calculations. Moreover, EQuIS is a data management system in which data storage, data management and data access are the key purposes of the system, allowing users to use specific visualization and analysis systems to interact with the data, such as:

- ESRI's ArcGIS
- C-Tech's MVS

This allows experts to use the tools they are accustomed to, while accessing data from a central source to avoid duplication and quality issues such as copy/paste or typing errors. Using one central source increases user confidence in the information they are consuming, as both quality controlled and up to date.

At ERM, data management using EQuIS has been widely adopted in the Americas where there is a great incentive via the widespread adoption of electronic data deliverables (EDD) among laboratories and regulatory bodies. Most US-based laboratories can create EQuIS EDD formats and typically

provide them as part of a normal deliverable at no extra charge. A similar standard is also rising in Europe, and several laboratories are providing EQuIS deliverables on request. Easily processed EDD formats of high quality are a key requirement to meeting efficiency goals and providing clear custody information to consumers of the data in the long and short term.

Automation

One of the main advantages of ERM's existing data management workflow is the use of EDD automation, where laboratories email or FTP completed EDDs to a mailbox on the database server for automatic processing. This allows for seamless upload of data from the laboratory to the database, without the intervention of project or data management staff. The data submitted through this automated workflow (Figure 2.3) are subject to quality and completeness checks (QA/QC process) that are not only faster, but are of higher quality than those done manually. Any EDDs that fail the automated quality checks (a failure can be due to a missing field, a non-matching valid value, or an incorrect field type) are rejected and sent back to the laboratory with a report identifying the error. This places the responsibility of error-free EDD compilation directly in the hands of the laboratory, reducing consultant hours, and therefore the ultimate cost to clients. As an indicator of the maturity of EQuIS electronic data deliverables - several regulatory bodies in the US use EQuIS as their own primary data management system, and thus also require that formatted EDDs be submitted directly to them. Examples include US Environmental Protection Agency (EPA) Regions 4 and 5, both of which distribute slightly modified EQuIS formats for submission to their databases. New York State Department of Environmental Conservation (NYS DEC) has a similar system in place.

Types of Data

While data management is commonly seen to be focused on management of laboratory analytical data, the EQuIS approach encompasses most data types collected during a site assessment, including:

- Water quality data,
- geological data, soil types, etc.,
- geotechnical parameters such as density, drilling parameters, and
- electronic data logger information, such as continuous water level or temperature, etc.

A variety of formats exist to incorporate data, including manual input via generic EDD, and automated connections are available for uploading data directly from the field, such as the EQuIS Data Gathering Engine (EDGE) and

data loggers (EQuIS LIVE agents), etc. These automated connections for data storage are part of the EQuIS infrastructure and can be configured in advance to facilitate rapid, error-free recording of most data types collected in the field, whether for assessment or monitoring.

Quality Control and Planning

Most organizations have workflows implemented that govern the flow of data from field to report, and the best will have rigorous quality control procedures in place. The migration of analytical or field data from its place of generation to its final resting place in the EQuIS data model is the period during which it is most vulnerable to error, human or otherwise. For example, in many cases, field chain of custody (COC) forms are on paper, are completed by hand in the field, and are then manually transcribed at the laboratory. Misinterpretations due to legibility or ambiguity (i.e. the letter ‘s’ versus the number ‘5’) can lead to the creation of inaccurate sample or location names or depths in the resulting EDD deliverable. The use of electronic COC and field-based data collection devices reduces the impact of transcription error, but does not eliminate it. Rigorous verification of EDDs, whether analytical data or other, is a key part of ERM’s workflow both during the upload process and afterwards. This QA program results in the identification and correction of errors before reporting and analysis tasks are underway. Data that has been through the verification process and released for reporting is noted as ‘final’ in the database, and may be subject to additional quality actions including validation by a project chemist.

Data Evaluation and Outputs

Once data is finalized per project requirements, it is available for analysis and reporting. One of the strengths of EQuIS is that connectivity to analysis and visualization software is built in, with either direct connection, as in the case of ESRI’s ArcGIS, or formatted output that feeds directly into other specialized software, such as:

- Golden Software’s Surfer,
- C-Tech EVS (Figure 2.4),
- Rockware’s Rockworks,
- and many office, modeling, and statistical analysis suites.

While this kind of interaction is useful to a certain group of users such as geologists and subject matter experts who will interact with the database on their desktop computers, there are other classes of user that, for example, are interested only in summary information, status updates, or derived

visualizations. For this class of 'casual' user, EarthSoft have developed EQuIS Enterprise – a dashboard-based website (Figure 2.5) customizable for the user or for the individual project site.

Dashboards are comprised of configurable widgets – each serving a distinct purpose such as running preconfigured reports or having access to a complete report library (Figure 2.6), listing the currently uploading EDDs, automatically graphing the latest data (Figure 2.7), or showing the site on an interactive map. EQuIS administrators have control over visibility of each object in Enterprise, meaning that certain users can be restricted to certain projects, widgets, or dashboards, and that the degree of freedom to customize for each individual can be controlled based on their needs and skill levels.

1.4 CONCLUSIONS

In order to be competitive, make smart decisions, and realize the value of the substantial costs spent on collecting environmental data - businesses engaged in site investigation, remediation, and/or monitoring activities should implement a robust, flexible, and comprehensive data management strategy. This should be the business "norm", as the benefits are measurable in both the short and long terms.

The strategy should involve trained personnel, hardware, RDBMS, technical software, analytical laboratories, QA/QC processes, and data sharing tools within a comprehensive infrastructure, governed by a standardized workflow. These components and the process that control them are critical in order to minimize error, conduct high volume analysis in a timely manner, and facilitate data and information sharing among the project team including the subject matter experts, clients, contractors, and regional and local authorities.

Archival of data, while still a key component of a successful data management strategy, is expanded in scope to allow for data retrieval, analysis, or transition to other parties with minimal effort. Given the frequent consultant transitions that take place on some of the largest of sites, this data-centric strategy resonates with clients who are often reticent in committing to a long term plan for their data resource.

To take advantage of the many benefits of using a data management system, ERM recommends an approach that includes the following elements:

- Electronic Data Deliverable (EDD) quality, completeness, and format standardization,
- Automated data collection, wherever practical
- Standardized, comprehensive QA/QC processes supporting all received data,
- Standardized reports and format options (for example, .xlsx, ArcGIS .shp),
- Document management capability, web accessible,
- Data Accessibility: Secure 24 hours/ 7 days per week access to data, reports and maps,
- Modelling and spatial/statistical analysis software integration, including direct connection or facilitated export so that skilled personnel can use the ‘best tool for the task.

As web and desktop technologies continue to evolve, the consultant with the above items in operation is well placed to compete in the global market, where individual sites are expanding to large portfolios, and clients want a similar level of service throughout their global footprint.

Looking forward to the next decade, the field data collection and automation facets of the data management workflow will take center stage, along with an increasing interest from clients in accessing, visualizing, and assessing their data in real time. Investing in the ‘back office’ technology and personnel in order to support these advances should be a priority task for any consultants engaged in site investigation, remediation, or monitoring activities.

Figure 2.1 Environment is complex by definition

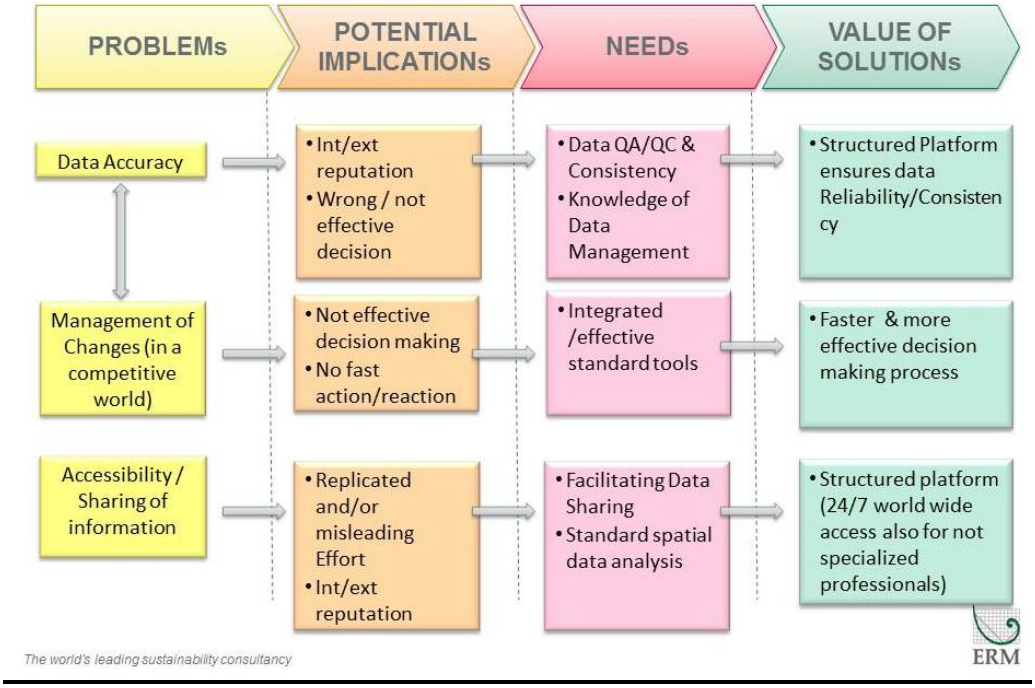

Figure 2.2 Site Contamination, Remediation; monitoring Optimised Workflow

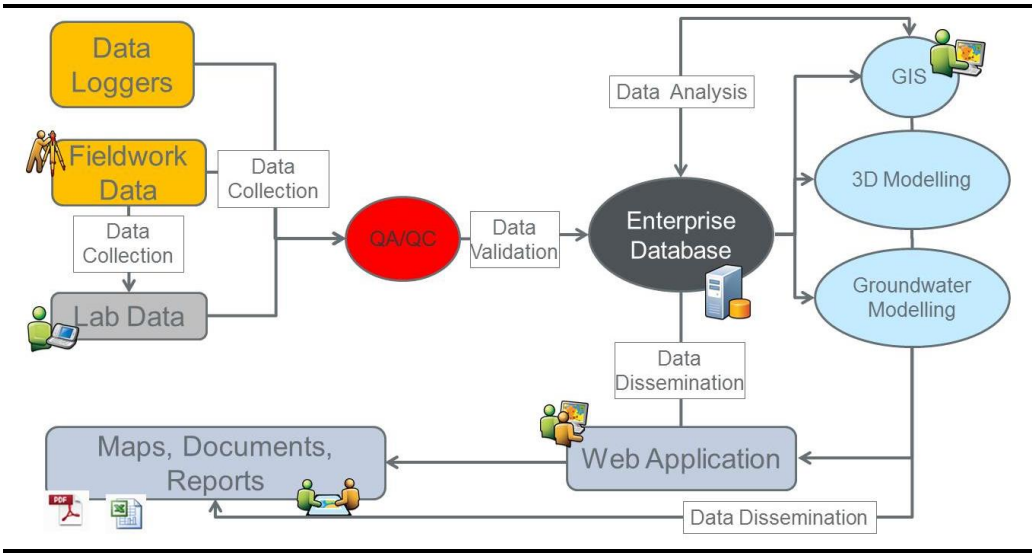

Figure 2.3 Automatic Control on Laboratory Data Upload

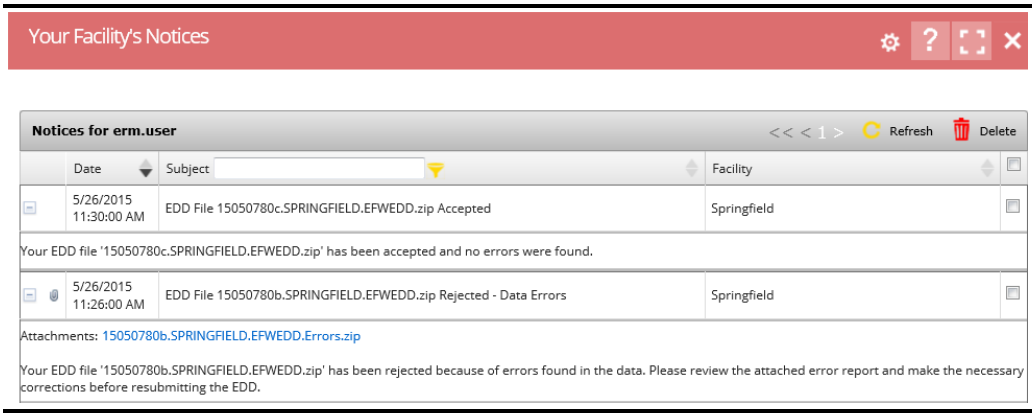

Figure 2.4 Effective Implementation of 3D Modelling through integration with MVS

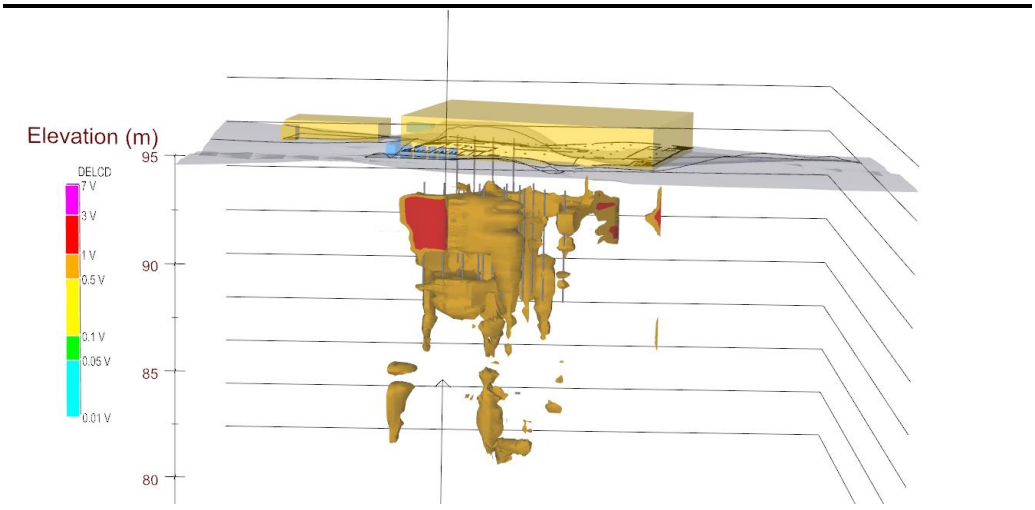

Figure 2.5 Site Management through a user dashboard

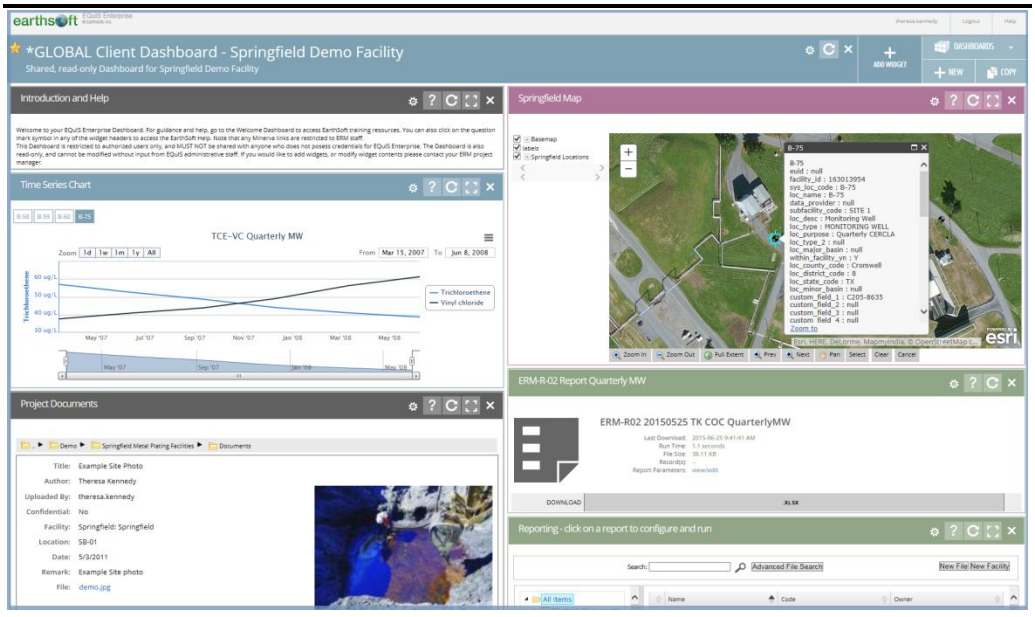

Figure 2.6 Extensive and Customizable Reporting Lybrary

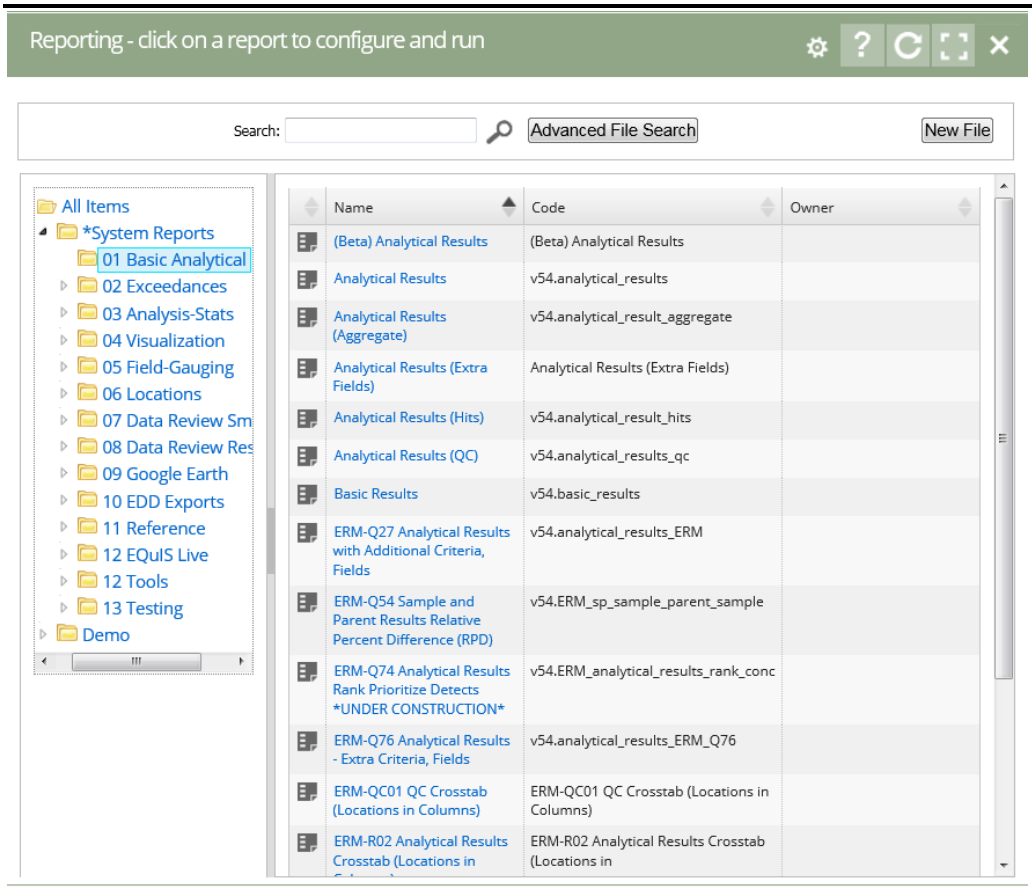

Figure 2.7 Automatic Trend Analysis