DEVELOPMENT OF ARCHIVING AND WEB DISSEMINATION OF GEOTECHNICAL DATA

Jennifer N. Swift

Department of Civil Engineering
University of Southern California

J. Carl Stepp
Consortium of Organizations for Strong-Motion Operating Systems (COSMOS)

Abstract

This paper is a summary of findings and recommendations of a Consortium of Organizations for Strong-Motion Operating Systems (COSMOS)/ Pacific Earthquake Engineering Center (PEER) Lifelines Program workshop on archiving and web dissemination of geotechnical data held on October 4 and 5, 2001. The concept that emerged from the workshop is a central hub that would function as a virtual data center through which data providers share as well as disseminate their data. Presentations during the first day showed that the development of a virtual center for web dissemination of linked geotechnical databases can be largely accomplished using existing technologies to link the organizational elements of the system. The primary needs are to: 1) define the functional requirements of a virtual center, 2) define data formats, data dictionaries, indexes, and exchange standards, and 3) define and link the organizational components of the overall system. The principal consensus recommendation for future development was that, initially, a pilot implementation of a virtual data center should be developed. The pilot system should involve several of largest data providers including California Department of Transportation (Caltrans), Pacific Gas and Electric Company (PG&E), California Geological Survey (CGS), and the U. S. Geological Survey (USGS). Building on this pilot system, the links could be expanded to include other data providers and the general user community.

Introduction

Designing and constructing major parts of the built environment rely on expensive-to-collect geotechnical data on subsurface conditions (Bardet, 1998). Geotechnical investigations are routinely required for design and to obtain approval to construct significant buildings, highway bridges, dams and embankments, and other structures. Such investigations also support and are necessary for improving building codes and engineering practices, for developing and refining ground motion site response modeling, and for numerous research purposes. The data are generally collected following current professional practices. Consistent standards and quality practices are not generally followed, however. Much of the data collected have potentially general application and significant value for broad geotechnical engineering community and construction practice and for university research. The data typically reside in working files and archives of local, state, and federal agencies and private sector organizations. Limited efforts are being made by some data providers to archive data in searchable electronic databases. There are, however, significant barriers to broadly accessing these databases because they are held in multiple data formats, data archiving and dissemination methods are incompatible, and the cost of retrieving data from paper records is high.

Newly available and emerging computer and communications technologies are making possible economical storage and sharing of valuable data to better serve society's needs. COSMOS and the PEER Lifelines Program held a workshop on October 4 and 5, 2001 to document current archiving and web dissemination technologies and to define future developmental needs in order to archive web dissemination of linked geotechnical databases. The workshop was motivated by the recognized economic importance of having geotechnical data readily accessible to the broad community of users. The objectives of the workshop were to develop consensus recommendations for classifying, archiving, and web dissemination of various types of geotechnical data and to develop a plan of action leading to development of a web-based virtual database system linked to multiple databases. Discussion papers together with the complete conclusions and recommendations of the workshop are published in a proceeding (Stepp, et, al., editors, 2002).

A generalized overall concept of how a web-based virtual data dissemination center could be set up is illustrated in Figure 1. The virtual data center hub would not house the data, but could house metadata and/or data indexes and translators that allow data to be accessed through the hub from various linked databases. The concept is that the data sources or providers would also be users and the general user community could access data from all databases through the system hub. Elements of a virtual center are elaborated and development needs are discussed in the remaining sections of this paper.

Archiving and Web Dissemination Architectures

Web Dissemination Architectures

Perhaps the most important finding of the workshop is that architectures for web dissemination of data from multiple databases are readily available. Developments needed for their implementation primarily involve the details of linking databases and accessing data. Two examples of applicable data dissemination architectural schemes identified as being most promising are "federation" and "harvesting". Diagrams displaying the basic elements of federation and harvesting architectures are shown in Figures 2 and 3 (Futrelle, 2002). In the federation architecture, the databases (data providers) themselves provide the search services directly to the end users (Figure 2). In the harvesting architecture, the query capabilities are maintained at the central hub of the system (Figure 3).

In the federation architecture each of the databases must develop and maintain the capability to execute user queries (Figure 2). The user performs the search by searching each of the databases in the system and the queries are done "live". This requires a standard query interface, a standard query language, a standard query protocol, and a standard set of fields to search that must be common for all participating databases. Each linked database must have its own index so that it can perform searches and queries in a reasonable amount of time. Consequently, there must be substantial agreement among the linked databases in terms of what query services are provided. This architecture is advantageous if the data are time critical. Also, there is no duplication of metadata within an index and no need for a central hub to maintain copies of metadata. In emergency response applications, for instance, federation might be the best architecture.

The harvesting architecture, depicted in Figure 3, provides a uniform user interface for querying multiple databases using common indexes, and for obtaining uniform communication of data. In order to implement this architecture the data providers must export or link the metadata from their databases into the harvester hub, which links users to multiple databases. The method to accomplish this must be designed and developed. Figure 4 illustrates in schematic one way a harvesting architecture could be designed to implement a web-based virtual geotechnical database system. In this example, there are four data providers (i.e. data sets), each generating metadata from their own original data sources. The generation and maintenance of the metadata must be the responsibility of the data providers. A defined set of fields or parameters would be exported to the index in order for data to be queried by the harvester. (A data index is a system such as an RDBMS or perhaps versioning software, containing maintained linked lists of the data and metadata that may be retrieved from the databases using the harvester hub.) The data providers must at least have those fields, which must be defined in a common standard for all databases, in their data dictionary. There should be no restriction on fields or parameters that can be exported to the index to be queried by the harvester hub. That is, the harvester hub should allow a user to search on all of the fields in the individual databases, as well as the common overlap described by the standard format. Extensibility can thus be built into this system architecture.

Figure 5 is an expanded schematic showing the elements of a web-based virtual geotechnical database system based on the harvesting architecture. Elements consist of multiple distributed databases maintained by data providers, metadata, data indexes, data translators, and a harvesting hub, which interfaces with users. Metadata are defined as data that describe data, encode relevant semantics, and are optimized for exchange (Futrelle, 2002). Retrieval is actually done at the indexes, which provide random access. Thus the harvester keeps track of the metadata from all of the different data sources. There is one subset or intersection that the harvester knows about and uses to link the user queries to the data. A data index works with the end users' queries by pointing to the data of interest. Though indexes could be maintained either by the data providers or the harvester hub, the data providers could best maintain their indexes, since they understand their data and how the data are organized. Each data provider's data dictionary and thus metadata should conform to a translator. The data translators, which could be maintained either at the harvester hub or by the data providers, filter the retrieved data or metadata into a standard format for dissemination through the virtual center hub to users. Dashed lines and dotted lines indicating the two possible system designs illustrate these alternatives. Different indexes can be developed to harvest different information. The end users should be able to access indexes as well as any sub-indexes within the virtual system that satisfy the requirements of their applications. Other applications besides metadata generation can be attached to a given database. Software already exists that can be used to set up a harvesting scheme such as Figure 5. This is only one scenario; there are other possible variations in terms of the number of data providers and indexes.

The harvesting architecture depicted in Figure 3 and expanded in Figures 4 and 5, is considered to have significant advantages for linking distributed databases through a central hub with a common user interface. For this reason this architecture is recommended for development and implementation of a web-based virtual geotechnical database system.

User Scenarios

The definition of user scenarios was identified in the workshop discussions as a priority need in order to establish functional requirements for a virtual database system. User scenarios identify patterns of geotechnical data use and users. User scenario-based design of the virtual system establishes what virtual system architecture will be required, what kinds of services and software need to be developed, and possibly other needs. The broadest range of user scenarios should be developed in order to establish the basic needs for the design of the virtual system. Focus should be on scenarios that critically depend on integrated data use across several different domains or sub-domains, data formats or instrumentation, and geographic or institutional locations, and should cover most data uses. It is critical that data domain specialists representing the principal data suppliers participate in developing user scenarios, and information technology (IT) specialists should participate in the design of the virtual system.

Data Dictionary and Formatting Standard

Explicit user scenarios also form the basis for development of a data dictionary and formatting standard, a critical element of the virtual system. The development of a strawman data dictionary from existing dictionaries would be a good way to begin this process while concentrating on content, or parameters of interest. Exiting dictionaries include the AGS (Association of Geotechnical and Geoenvironmental Specialists, UK) NGES (National Geotechnical Experimental Sites, UNH) geotechnical data standards, and LAS (Log ASCII Standard, CWLS). The AGS standard represents the practical, applied field side of geotechnical engineering, while the NGES standard represents the more academic or research side. A strawman data dictionary was developed from existing dictionaries and presented as a structured parameter list for this workshop (Turner and Roblee, 2002).

The recommended approach for developing a data dictionary and formatting standard that emerged from the workshop is schematically depicted in Figure 6. The approach is illustrated by example with a virtual system having two databases, database A and database B. (Ultimately the envisioned system would include a number of databases.) The need is to define the common overlap in data dictionaries or parameters of interest. This would require agreement on a common definition of the parameters in the standard. The overlap is the information that would actually be transferred and exchanged. Care should be taken not to try to encompass all parameters included in every existing standard studied, as the resulting standard would become unmanageable. Although the need is to create a standard data format, the data exchange system or format itself should be flexible enough so that end users are able to use the data however they want. The main purpose is to make the data easy to exchange and in a format that participants actually want it for specific uses. Thus both content and semantics need to be defined to produce a usable format.

Data Exchange Standards

Data exchange standards (metadata, data indexes, and data translators) are other major components of a virtual database system that must be given priority development. Data exchange is accomplished by use of translators, which are filters between the data at a database participating in the virtual system and data retrieved through the virtual geotechnical system hub. In physical terms, translators are basically software applications. The definition of the translators will depend on:

1) The requirements for the data producers to generate metadata and convert their data to the chosen standard format for exchange; and
2) The overall information architectural scheme chosen for the virtual system.

Another important need is to establish where the translators and indexes will be located and maintained. These tools could be part of the providers' database systems or of the virtual system hub. These alternatives are further illustrated in Figure 5. The dashed lines and dotted lines indicate the two possible system designs in terms of responsibility. In the dashed line design, the virtual system hub is responsible for maintenance of the data provider's indexes, metadata and translators. Whereas in the dotted line design, the data providers are responsible for their own indexes, metadata, and translator applications. The most appropriate design is for the data providers each to maintain their own translators and indexes with guidance from the virtual system hub. The virtual system hub could develop the indexes and translators, then turn them over to the data providers. In the alternative design the data would end up being centrally managed, which is more problematic. In order to include participation of smaller agencies or professional firms who may want to participate, a combination design would be appropriate where the harvester points to data from the bigger agencies and also stores data from the smaller agencies or companies centrally.

It is important to note that translation of original data into metadata and data in a standard format inevitably involves data loss. A combined format and index will normally contain less information than all of the individual databases that were integrated. Otherwise the functions of those databases would have to be restricted, which is not desirable since the databases could also have other applications attached to them.

An important, though less critical, issue is the format for communicating the data to the end users. This could be via XML, ASCII or some other format. A distinction must be made between the syntax of the data standard (the structure of a line string in some language), and the semantics (the meaning of a line string in some language) of the data standard. The semantics are what each element in that data standard means. As an exchange format ASCII is the most commonly used, and is employed in the AGS, NGES and LAS formats. The recommendation of the workshop is to use ASCII as the exchange format, since it is universally accepted, and to use XML for metadata. Notwithstanding, all format options should be investigated as part of the implementation process. As for communication of the data, for instance by developing viewer applications to display on-the-fly graphics, internet formats such as XML, XSL, and VML could be used, or even image formats such as .gif files could simply be generated by invoking an existing software tool so end users could view data from the central hub. The development and maintenance of such viewing applications would be the responsibility of the hub.

Implementation Actions

This section describes a plan of actions needed to implement a virtual geotechnical data center. The actions focus on development of applications for already existing technologies. That is, the archiving and web dissemination system should use existing database technologies and web dissemination architectures and develop applications needed to implement these in a web-based dissemination system. The harvesting architecture emerged as the most promising for implementing a web-based virtual system. The elements of a web-based virtual database system based on the harvesting architecture are depicted in progressive detail in Figures 3, 4, and 5. Based upon the harvesting architecture, the implementation actions address elements of the system including, 1) definition of the functional requirements of the system, 2) development of a data dictionary and data formatting standard, 3) development of data exchange requirements and standards (handling of metadata, data indexes and data translators). As an implementation strategy initially, a pilot system should be developed linking the databases of two or three data providers. Once the pilot system is operational, implementation could be expanded to include the broadest participation.

Definition of the functional requirements of a web-based system

The central concept and functional objective is a web-based data system that facilitates data dissemination from participating distributed data sets, functioning together as a virtual data system. Definition of the functional requirements for the virtual system is a priority primary need. The important first step is the identification of the data users and data user scenarios, which in turn determine details of the functional requirements of the system such as the scope of the standard data dictionary, the method of handling metadata, and the scope and method for handling data indexes. The identification of data users and data use scenarios must involve the participation of data providers. Geotechnical data providers normally also maintain databases which may be linked to a web-based dissemination hub. Once the functional requirements of the system have been established, the most appropriate implementation of the harvesting architecture can be determined.

A work group constituted of data providers and users can best define the functional requirements of the system. The work group should identify data users and data use scenarios, and use this information to establish the functional requirements of the web-based system, such as depicted schematically in Figures 3, 4 and 5. The work group should include an IT specialist who has experience with the harvesting architecture. As a starting point for developing a comprehensive set of user scenarios, a catalog of the types of available existing data and where the data can be accessed should be developed.

Development of a Data Dictionary Standard

A data dictionary standard is a fundamental need for implementation of any database, and a fundamental requirement for linking distributed data sets in a web-based virtual system. Data providers develop diverse types of data including geological, geophysical, geotechnical, and geoenvironmental. The data are collected for a multitude of purposes ranging from geotechnical characterization of specific sites, to hazard mapping, to general geological mapping, to specific research projects. In the absence of any standard, data are held in diverse formats and critical information that is needed to inform the general user about the data and permit any evaluation of its quality is most often not documented (Bardet, 1998).

Issues relative to developing a standard data dictionary vary depending on whether the standard is for the electronic capture of legacy data held in paper form, data in an existing electronic database, or data to be collected in future projects. Legacy data can be captured in electronic format following a data dictionary standard, but significant metadata needed to evaluate its quality normally will be missing and cannot be recovered. Existing electronic databases have their own data dictionaries. These dictionaries may intersect as depicted in Figure 6, permitting a limited range of common access through a central hub. On the other hand, a common data dictionary standard covering the full range of data and data use scenarios could be developed and used by all data providers for data capture in the future. Such a standard could facilitate the direct capture of field data in electronic format and thereby insure that the appropriate metadata are collected at the same time.

For the purpose of initial implementation a critical need is to develop a data dictionary standard that includes the largest intersection of the dictionaries of existing electronic databases and has built in flexibility for expansion. The standard could then be used for purposes of capturing legacy data in electronic format and as the basis for implementing a web-based virtual system for geotechnical data dissemination linking existing geotechnical databases. Two data dictionary standards are available: the AGS Standard and the LAS Standard. The NGES Geotechnical Database used the AGS Standard as a starting point for the purpose of developing the NGES data dictionary (Satyanarayana and Benoit, 2002), and a strawman data dictionary adapted from existing data dictionaries has been developed by Turner and Roblee (2002).

The NGES data dictionary should be adopted as the starting point for development of a data dictionary that incorporates as broadly as achievable the intersection of existing data dictionaries. The strawman data dictionary developed by Turner and Roblee (2002) could be considered as an alternative starting point for this action. The data dictionary should be flexible to permit expansion as the virtual system is linked to additional databases.

Because of the broad range of types of geotechnical data, the task of developing a data dictionary must necessarily involve a broad range of data specialists. Issues involved with geologic descriptive data, boring log data, and laboratory test data, to cite a few examples, differ significantly and the collection of these data involves specialists with different backgrounds and experience. In order to address these issues in the development of a data dictionary standard, an appropriately wide range of data specialists must be engaged. These should include for example, specialists in geologic non-parametric data, *in situ* field test specialists for geophysical and geotechnical measurements, specialists in obtaining field samples for laboratory testing, and others. In addition, the data of interest generally are referenced to a geographic location. Consequently, the effort must involve specialists in spatial mapping and location technologies; e.g., geographic information systems (GIS).

A work group is currently being established for the purpose of developing an expansible data dictionary. The work group is constituted of representatives of data providers, such as Caltrans and other state DOTs, CGS, U. S. Geological Survey (USGS), Pacific Gas and Electric Company (PG&E), Federal Highway Administration (FHWA). The work group should have participation from professional associations such as American Society for Testing and Materials (ASTM), AAS, non-profit organizations such as Petrotechnical Open Software Corporation (POSC), and should include specialists in GIS technology. The work group should be organized around subgroups necessary to address the range of specific types of geotechnical data to be contained in databases that are to be linked to the web-based virtual system. Subgroups should include geologic subjective non-parametric data, *in situ* geotechnical test data, *in situ* geophysical test data, field sampling and laboratory test data, and spatial and site location information. In order to insure that adequate communication takes place across data types it is recommended that the subgroups all be part of the larger work group.

Data Quality Issues

Data quality is a key issue that must be addressed as a part of the curation of data. Responsibility for data quality must rest with data providers and those who collect data that may become part of a data provider's database. Quality assurance is viewed not so much achieved through process in the form of QA procedures or guidelines as through instilling a culture of quality data reporting. Some assurance of quality can be attained through data checking procedures that can be easily implemented by data providers. Even greater quality assurance can be attained by implementing procedures for data entry that avoid handling of the data recording by multiple persons along the pathway to archiving. The most effective procedure would be to implement mechanisms to enter data into the archive directly as it is obtained in the field. The overriding need is to stimulate those who collect data to think about quality reporting as part of routine data acquisition.

Mechanisms and procedures for assuring data quality necessarily are different for new data and for legacy data (Nigbor et al., 2002). For new data, quality assurance is highly related to the scope and content of the data dictionary as well as to the mechanisms and procedures employed for entering data into a database. For legacy data little can be done along these lines, as the data already exist and are held in a particular format. For these data the primary quality assurance need is to avoid data transfer error in the process of capturing the data in electronic format. It is believed to be possible however, to provide some measure of the quality of legacy data by assessing the degree to which the information available meets the data dictionary requirements for new data entry. By implementing this process, legacy data could be rated according to a measure of quality.

Data quality should be an integral consideration for the development of a comprehensive data dictionary standard. The NGES data dictionary takes data quality into account and is considered a good starting point for the development of a data dictionary standard suitable for the envisioned web-based virtual system. It is believed likely that once a data dictionary standard is in place it will be generally accepted and used by most organizations and contractors. Nevertheless, when data are acquired through specific contractual agreements use of the standard should be required. To the extent achievable procedures should be developed and used to enter data electronically as they are obtained in the field.

In order to flag the quality of legacy data a procedure should be developed for rating the data based on the degree to which the available documentation meets the data dictionary standard.

Development of a Pilot Virtual Geotechnical Data Center System for Caltrans and CGS Geotechnical Databases

This section describes actions for development of a Pilot Virtual Geotechnical Data Center System for CGS, USGS, PG&E and Caltrans geotechnical data sets, which could be expanded in the future to incorporate the broad range of geotechnical data from other agencies, academia, and industry. A schematic of how the CGS or Caltrans data might be captured and structured as an element of the pilot virtual system is shown in Figure 7.

A work group is currently being formed to identify and define optimal information architecture (including archiving) for the virtual system hub, considering the primary agencies' needs, currently available technologies, probable future technological developments, future expansion to include other geotechnical databases, and potential integration with Network for Earthquake Engineering Simulation (NEES) Program. The focus is on implementing the information harvesting architecture recommended by the workshop participants. In addition to defining the basic system architecture that meets the functional requirements of a web-based system, specific responsibilities for maintaining the data dictionaries, the structures for holding and disseminating metadata, the process for developing and structure for holding data indexes, and the responsibility for developing and maintaining data translators should be identified. It is anticipated that these elements would be to some degree unique to each data provider.

One of the primary steps involved in creating the pilot system is to develop a basic system integration plan for each agency that could be expandable to a larger web-based system serving multiple data sets. Of equal importance, data indexes and translator methods and technologies required by each of the primary data providers must be clearly defined. Metadata requirements for data providers to participate in the virtual system would also be defined. A phased implementation plan and organizational structure for each agency to participate in the virtual system should be developed.

Structure and Method for Handling Metadata and Data Exchange

In order to create a virtual system, the structure and method for handling metadata, non-metadata and data exchange needs to be addressed. The structure and methods will be dependent on the user scenarios. The translators and indexes tend to be application-specific, and may be unique to each data provider, though a number of different providers could share a given index.

The structure of the metadata and the data disseminated through the hub must be less specialized, in order to facilitate the implementation of the virtual system. The more portable the technology developed, the easier it would be to expand the system in the future to include other participants.

In order to define steps involved in determining the structure and methods for handling metadata and data exchange standards, it is necessary to define protocols and architectures for accessing metadata and disseminating non-metadata produced by each data provider. Protocol and architecture refers to data processing and platform-specific functionality to be used for data translation and exchange. These protocols and architectures might be unique to each agency, depending on the agencies' needs, currently available technologies, probable future technological developments, and available maintenance resources. For instance, government agencies are often restricted in terms of allowable software, use of newer markup languages, and server-side programming (such as Active Server Pages, ASP) in regards to what can be deployed over the internet. Applicable solutions for data exchange could be determined based on available resources, personnel, and protocols appropriate for the project.

Setting up metadata and non-metadata generation schemes at each agency can be handled in a number of ways. Software programs may be written specifically to translate and extract metadata and data from the agencies' native databases or data archives into a form which can be queried by the virtual system. The recommended amount of data, or rather non-metadata that would be contained within each metadata file depends primarily on how large a given data set is. If the data set for a single borehole or geotechnical investigation site, for instance, is comprised of text information only, then the metadata files would be very small and could contain all of the non-metadata within a given file.

The metadata and non-metadata files should be made accessible through an index of the agencies' currently available files connected to or residing on the virtual system. Recommendations for where the metadata, non-metadata and indexes should reside can be resolved during development of the pilot system. The translators could reside with and be maintained by data providers or they could reside at the system hub and be maintained there depending on their function.

A small work group is being formed to develop metadata structures and translators specific to each agency, aiming toward less specialized data exchange methods for participation in the system hub. The work group consists of persons from the data provider organizations and an IT specialists who have experience with metadata and data exchange standards development. In addition to the above-mentioned technical needs, the work group will also address possible solutions to allow geotechnical data to be viewed on-the-fly as graphics from the system hub.

Acknowledgements

We want to acknowledge the PEER Lifelines Project for its funding support and for providing expert talent that made the workshop a strong success. In particular we want to recognize the contributions of Cliff Roblee and Loren Turner of Caltrans, Chuck Real of the CGS, and W. U. Savage of USGS/PG&E for their leadership in organizing, managing, and reporting the results of breakout discussion groups from which this paper was developed. We thank PEER for providing the workshop facilities and for continuing support of COSMOS.

References

Bardet, J.P., 1998, Proceedings of the planning meeting for Geotechnical Data Management Center (GDMC), Report to the National Science Foundation, Civil Engineering Department, University of Southern California, Los Angeles, California, p. 80.

Futrelle, J., 2002. "Current Database Structures and Future Directions of Technology Development". In Archiving and Web Dissemination of Geotechnical Data, Proceedings of a Workshop held on October 4 and 5, 2001, COSMOS/PEER Lifelines Program.

Nigbor, R. L., Swift. J., and Hu, J., 2002. "Quality Assessment Criteria Experience with Existing Geotechnical Databases". In Archiving and Web Dissemination of Geotechnical Data, Proceedings of a Workshop held on October 4 and 5, 2001, COSMOS/PEER Lifelines Program.

Satyanarayana, R., and Benoit, J., 2002. "National Geotechnical Experimental Sites Geotechnical Database". In Archiving and Web Dissemination of Geotechnical Data, Proceedings of a Workshop held on October 4 and 5, 2001, COSMOS/PEER Lifelines Program.

Stepp, J. C., Swift, J., Turner, L., Real, C., and Savage, W., eds., 2002. Archiving and Web Dissemination of Geotechnical Data. Proceedings of a Workshop held on October 4 and 5, 2001, COSMOS/PEER Lifelines Program.

Turner, L., and Roblee, C., 2002. "Borehole Data Dictionary, Formatting, and Data Capture Requirements". In Archiving and Web Dissemination of Geotechnical Data, Proceedings of a Workshop held on October 4 and 5, 2001, COSMOS/PEER Lifelines Program.

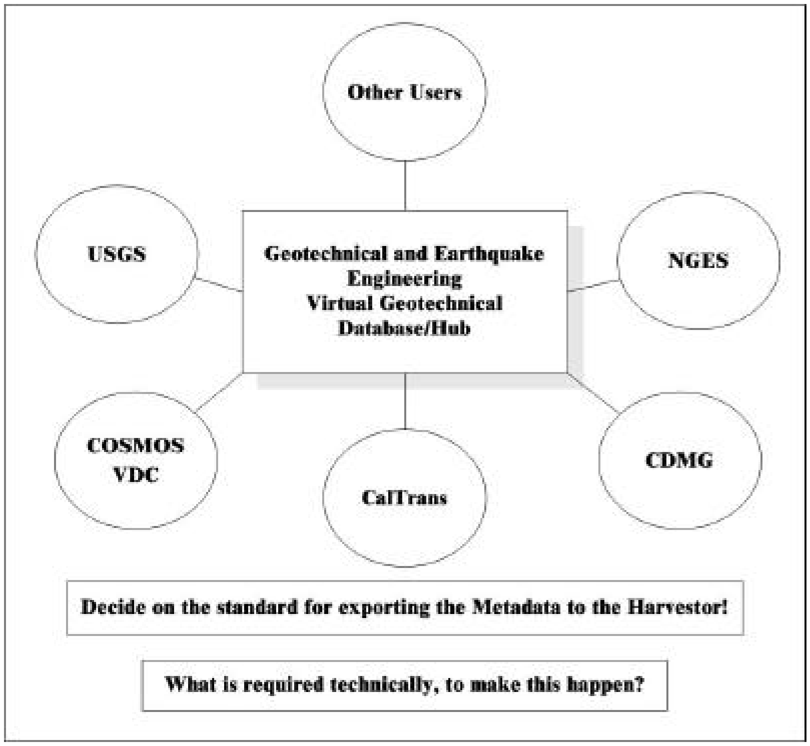

Figure 1. Generalized concept of a web-based virtual data dissemination center.

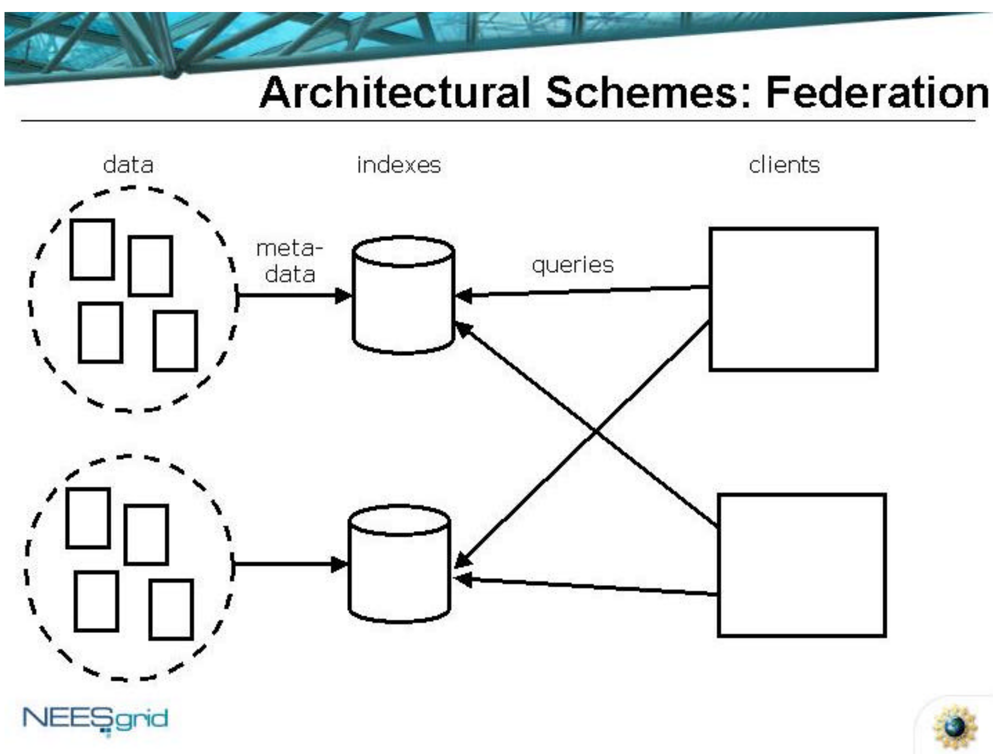

Figure 2. Information architectural scheme: “Federation” (Joe Futrelle, Stepp et al., 2002).

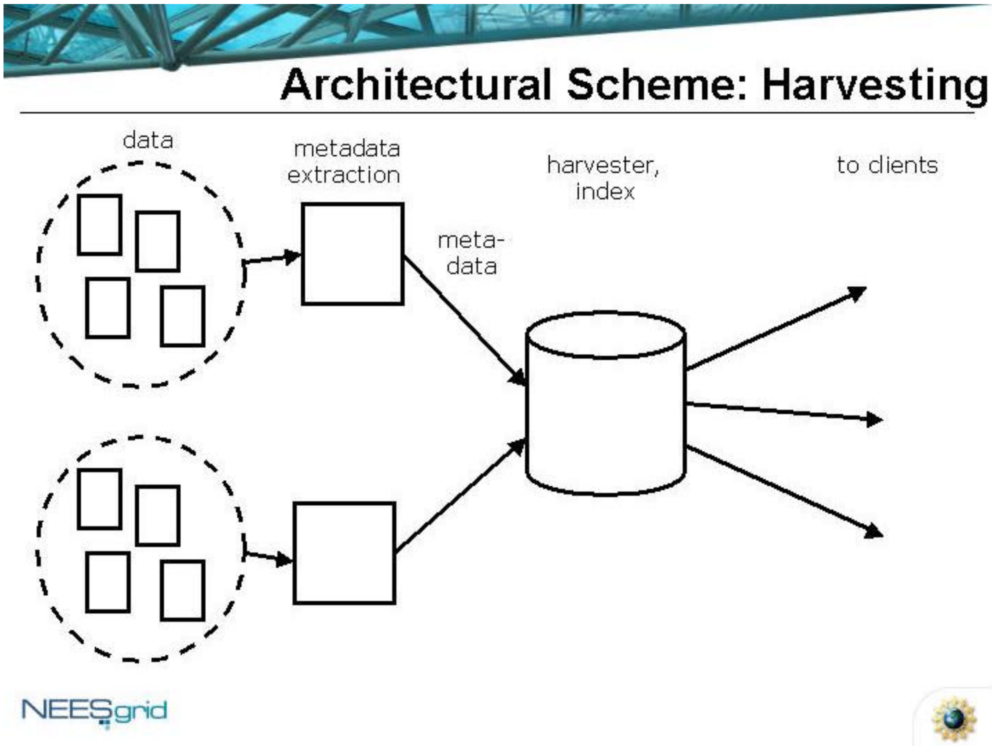

Figure 3. Information architectural scheme: "Harvesting" (Joe Futrelle, Stepp et al., 2002).

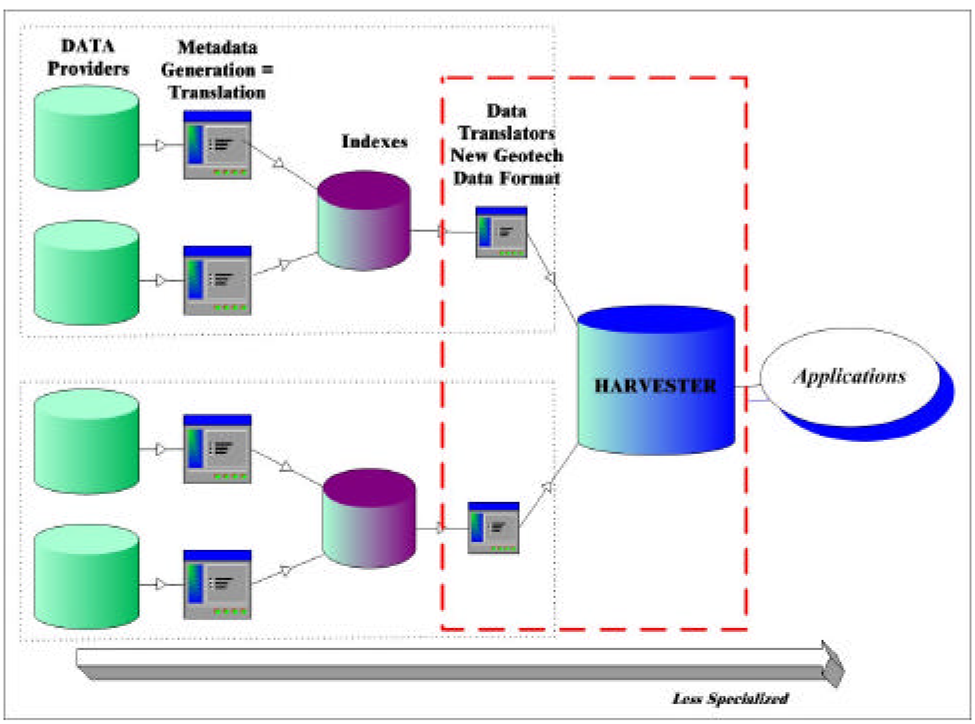

Figure 4. Illustration of a harvesting architecture for a web-based virtual geotechnical data center for four data providers.

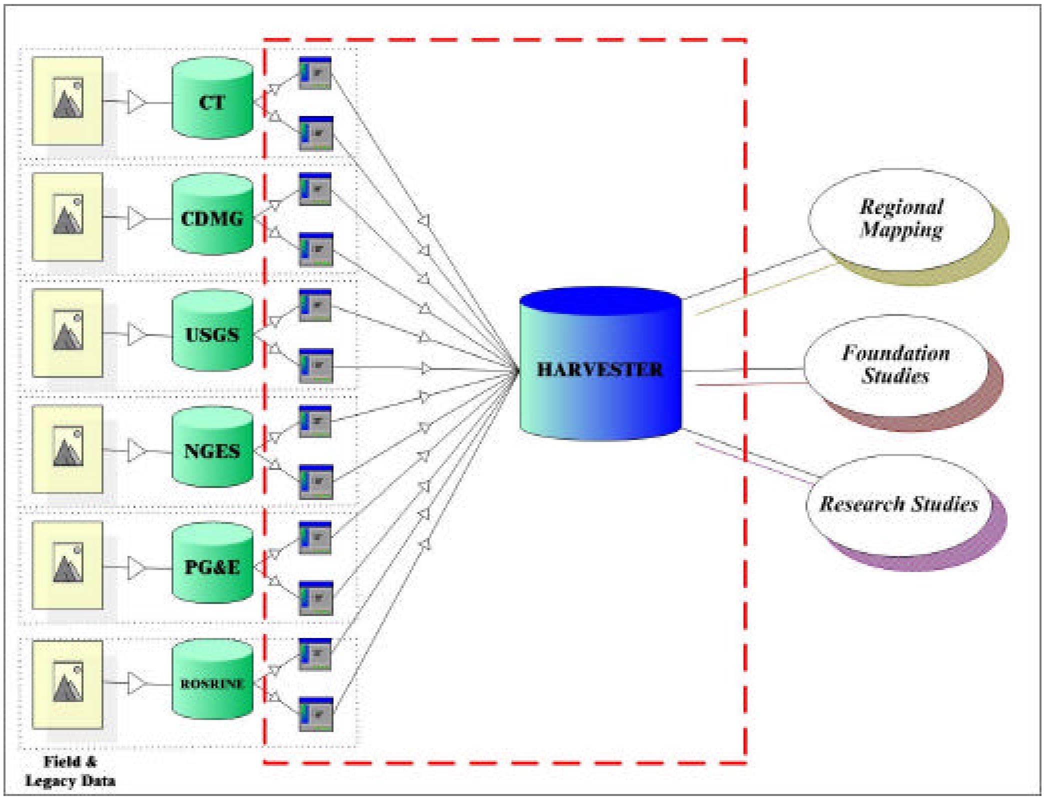

Figure 5. Basic elements of an extensible web-based virtual geotechnical data center.

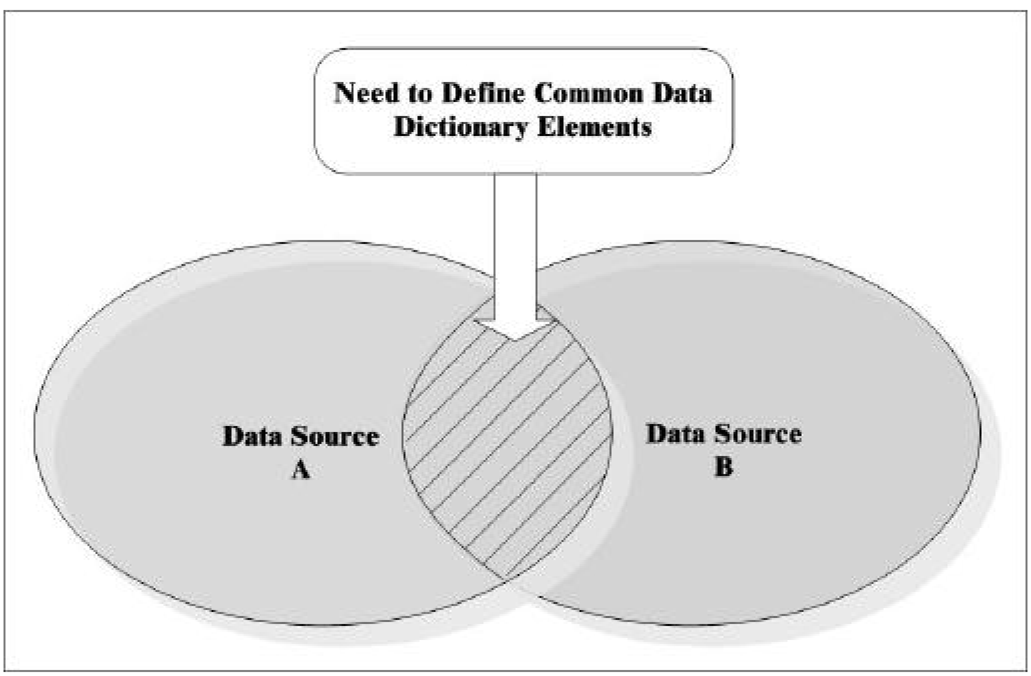

Figure 6. Schematic depiction of an approach for developing a data dictionary standard.

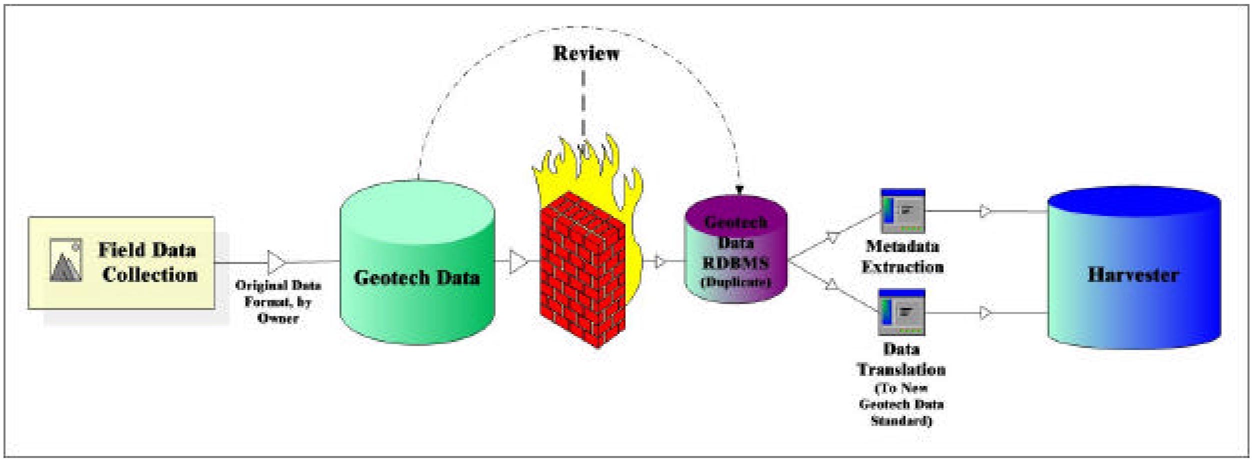

Figure 7. Schematic showing how CGS or Caltrans data could be captured and a structure for participation in a virtual geotechnical database system.