# National Database for Water Infrastructure System

Sunil Sinha[1], and Walter Graf[2]

[1]Associate Professor, Department of Civil and Environmental Engineering, Virginia Tech, Room No.117C, Patton Hall, Blacksburg, 24061; e-mail: ssinha@vt.edu
[2]Program Manager, Water Environmental Research Foundation (WERF), 635 Slaters Lane, Alexandria, VA 22314
Phone: (571) 384-2102; Fax: (703) 299-0742; e-mail: wgraf@werf.org

## ABSTRACT

This paper presents on-going research work titled "A Web-based Interactive National Database of Condition Assessment and Renewal Engineering Technologies for Water and Wastewater Pipelines." This project is funded by the Water Environmental Research Foundation (WERF) through the Environmental Protection Agency's Aging Water Infrastructure Research Program, a research agenda that supports efforts to put the nation's aging infrastructure on a pathway toward sustainability. The development of this research program stems from EPA's Sustainable Water Infrastructure Initiative. This research project will evaluate existing, new, and emerging technologies that address the nation's aging water and wastewater infrastructure.

## INTRODUCTION

Today, municipal governments are facing an infrastructure crisis requiring costly repair, rehabilitation and replacement beyond their capacities (WIN, 2000). There has been a gradual decline in the state of our infrastructure over the past two decades and a growing concern is that these facilities may be inadequate both for current necessities and for projected future growth (ASCE, 2000). Funding for carrying out the repair/ rehabilitation/ replacement activities are limited and needs to be allocated to assets in an optimal way. Also, the accidental damages to the underground pipes caused due to removing materials, driving piles, posts and more has been aggravated by the addition of utilities and new technologies in the underground spaces. This kind of accidental damages to the pipes causes additional burden on the already strained resources. Thus utilities manager needs to take quick informed decisions for implementing the technologies which are appropriate for their situation.

According to the EPA, US water and wastewater infrastructure includes an estimated 16,000 wastewater and 52,000 drinking water industries, which will require a substantial investment of $334 billion and $274 billion respectively for system rehabilitation over the next 20 years (U.S. EPA, 2005 and 2007). In addition the utility engineers responsible for fixing the pipeline infrastructure are governed by various laws such as the Safe Drinking Water Act of 1974, the Clear Water Act of 1977, the Water Quality Act of 1987, and American Society for Testing and Materials Standards, which require standards of compliance that must be maintained. In order to ensure that these current and future needs are met, condition assessment technologies and practices are a critical component of managing our water and wastewater infrastructure. As we contemplate "The Dawn of the Replacement Era," where an

unprecedented investment in pipeline assets will be required [American Water Works Association (AWWA), 2001. it is critical that we develop the knowledge necessary to make sound decisions, and build legacy assets that will sustain future generations.

## CONDITION ASSESSMENT AND PIPE LOCATION TECHNOLOGIES

Condition assessment is the collection of data and information through direct and/or indirect methods, followed by analysis of the data and information to make a determination of the current and/or future structural, water quality, and hydraulic condition of the pipeline. Once the current condition of the pipeline is known funds can be better allocated to the assets which need urgent attention. Buried Pipe Location or Subsurface Utility Engineering (SUE), is the combination of civil engineering, survey, mapping and geophysics when working with utility infrastructure to reduce cost, increase safety, and prevent unplanned disruption of services. Once the existing conditions are known the designer can incorporate this into new projects to avert conflict with utility crossings and determine if relocation is the most feasible alternative versus modification of a projects' scope. At present, the utility managers and decision makers are struggling with easy access to the comprehensive information about the technologies and experiences of other utilities in dealing with different situations. The information about the condition assessment and buried pipe location technologies, best practices, experience with technologies (Positive and Negative), cost of technology, cost models, list of vendors, contractors and case studies etc. that can be used as the knowledge by the various utilities is available but is not shared and is not readily available at present. The proposed database will ensure a single point information center for the utilities where they can find all the relevant information that will help in expediting the decision making process for the selection of appropriate condition assessment and buried pipe location technologies.

## REPAIR, REHABILITATION, AND REPLACEMENT TECHNOLOGIES

System Renewal includes a wide range of repair, rehabilitation, and replacement techniques that bring the pipeline system at acceptable levels of performance within budgets. Various commercially available products present the current state of the art in pipe repair, rehabilitation and replacement. The decision-making process for the proper balance of repair, rehabilitation and replacement is a function of the condition assessment of the pipe, the life-cycle cost of the various repair, rehabilitation, and replacement options, and the related risk reductions. The unstructured process presents a challenge in educating our utility engineers/managers on pipeline system renewal. An interactive web portal will address this challenge by bringing information from all the stakeholders in one database. It will relate guidelines, design options and installation requirements like cost and performance to help the utility make an informed decision. The project will focus on the transfer of performance and cost information on repair, rehabilitation, and replacement technologies for water and wastewater pipes. The web portal will include primary information about individual renewal technologies' cost and performance, case studies for their real world applications, and the list of vendors, consultants, utilities, research association and contractors available for particular technologies.

## DATA COLLECTION PLAN AND GUIDING QUESTIONS

Various utilities have management practice and have used various technologies for CA and RE related to drinking water and wastewater pipeline infrastructure system. However, the gap currently exists between CA and RE technologies available and those used by utilities. In addition, the communication between utilities is very limited and thus sharing "Lessons Learned" is difficult in current environment. The content of the interactive web database will allow the comparison of benefits and drawbacks among alternative CA and RE technologies. Furthermore, access to a collection of data in a centralized location (national database) will allow more effective use of asset management tools. One of the goals for national database is to share the technology, performance, management practice, and cost information between utilities nationwide to help decision making and to avoid the mistakes. There is a need for real world data regarding the cost and performance of technologies or management practice and data mining from real world application must be conducted in prior to develop the database for CA and RE. In the current research, data such as technology, management practice, applications, and experience of utilities will be collected and the performance and cost information on CA and RE for drinking water and wastewater will be transfer to various utilities, vendors, consultants, contractors and research associations. In addition, the research team will provide a comprehensive review and evaluation of collected data for CA and RE. Data collection can compile the CA and RE experience of utilities as described within utility report, case histories, and manuals. It can also verify actual capabilities and limitations of the available technologies and management practice. Sharing data among various utilities will allow the development of a more comprehensive database of product information and "Lessons Learned" through past experiences than any one utility is likely to be able to compile. In addition, encouraging communication between utilities nationwide will allow infrastructure owners to more efficiently address the degradation of drinking water and wastewater infrastructure.

The research team is data mining to academically gather the industry knowledge, and is also identifying gaps in the industry knowledge. The research team will gather the data from various stakeholders such as venders, consultants, utilities, contractors, and research associations. The research team expected different types of information from each stakeholder. For example, lab performance, manufacturing cost, and technology specifications can be collected from venders, and actual site performance, CA/RE cost, and CA/RE product performance data can be obtained from utilities. To obtain the most research relevant data from stakeholders, the guiding questions were developed by the research team and reviewed by Data User Group (DUG). The guiding questions are intended to help research team to focus on the main goal of the research and extract as much data as possible, and to guide stakeholders to collect appropriate information. The guiding questions will help the stakeholders to gather the general information and raw data related with drinking water and wastewater infrastructure which the asset manager can easily obtain. Please note that the guiding questions are reference questions and are not a survey for stakeholders. The stakeholders are not required to answer any questions or fill out any forms. Information about performance and cost from stakeholders are shown in Table 1.

Table 1. Performance, Cost, and Expected Data from Stakeholders

| Stakeholders | Performance Indicator | Cost Indicator | Data / Information |
|---|---|---|---|
| **Vendors** | Lab Performance | Manufacturing Cost | Collection of Technology Specifications |
| **Consultants** | Expected Performance | Engineering Cost | Collection of Project Reports and Technology Analysis |
| **Utilities** | Actual Site Performance | CA & RE Cost | Collection of CA and RE Product Performance Data |
| **Contractors** | Assumed (Predicted) Performance | Contractual Cost | Collection of Contract Costs and Risks |
| **Research Associations** | Reported Performance | Assumed (Predicted) Cost | Collection of Research Reports and Test Reports |

The real world applications as described within provided data from stakeholders will be used to verify the requirements, capabilities and limitations of the available technologies and management practices as documented in the Technology Summary Sheets and Management Practice Summary Sheets. The Technology Summary Sheets should at a minimum include the following contents:

- The requirements for the technology.
- The actual cost to apply the technology.
- Advantages and disadvantages of the technology.
- "Lessons Learned" based on the experience.
- The results and reliability of the technology.

The Management Practice Summary Sheets should at a minimum include the following contents:

- How easily the management practice can be applied.
- Special circumstances and limitations.
- The actual capabilities.
- Advantages and disadvantages.
- "Lessons Learned" based on the experience.

Figure 1 shows the overall plan for data mining, data analyses, and digital library. First, the research team collects relevant data from various stakeholders using the guiding questions. The data from stakeholders may have analogue and digital format. The analogue data such hard copy of report and field hand writing record data will be converted into a digital format for easy access and for analyses. Please note that providing a data is purely voluntary agreement by stakeholders and the research team ask minimum of stakeholders' effort. To minimize the effort of stakeholders, the research team collects a raw data or unprocessed data directly from stakeholders and the research team process the data for analyses. To determine the metadata, the research team will use Dublin Core, which provides a small and fundamental group of text elements through which most data can be described and catalogued (Harvey and Hider, 2004), and domain specific metadata from national database. Once the research team analyzed the data, the results will be used to develop digital library.

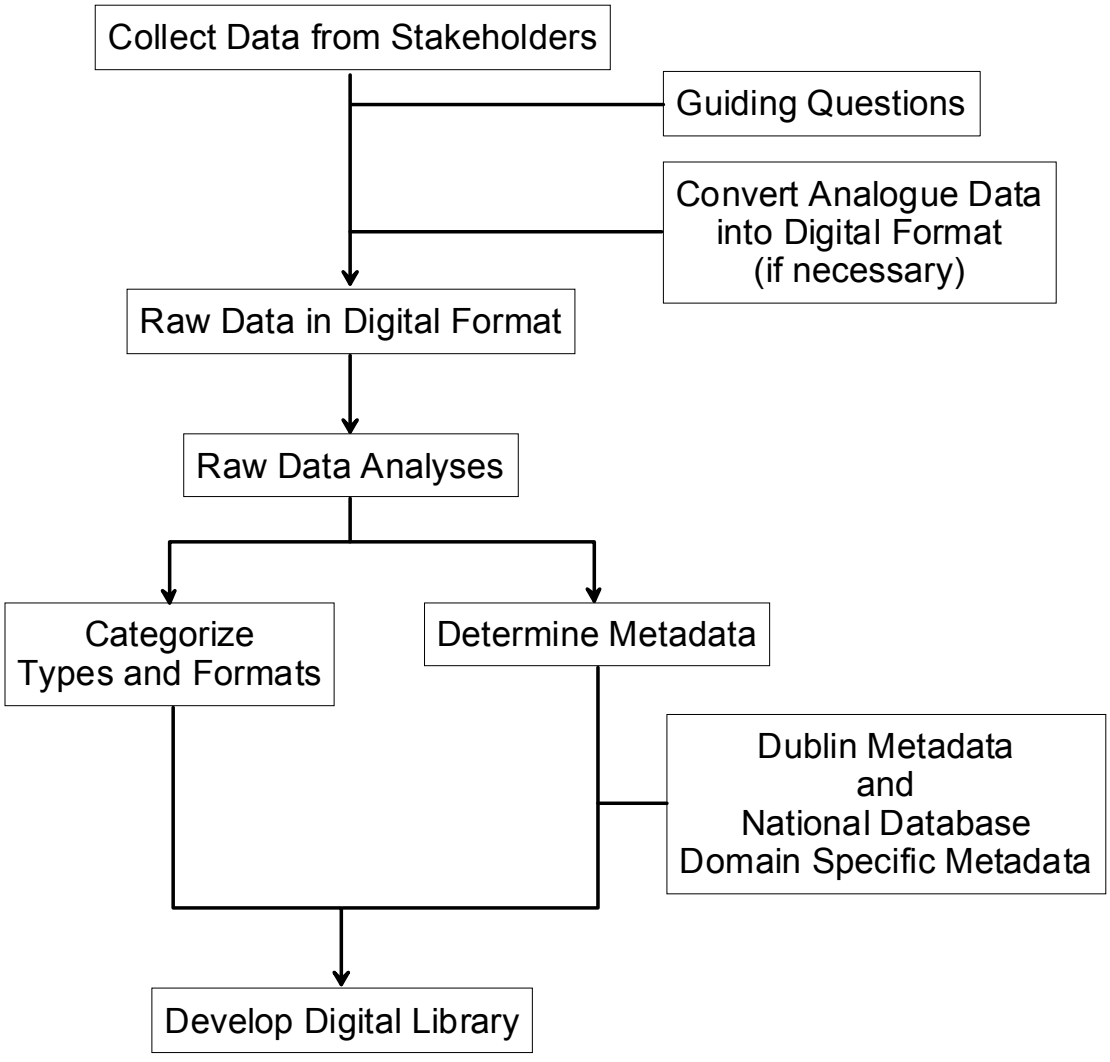


Figure 1. Overall Plan for Data Mining, Data Analyses, and Digital Library

## DEVELOPMENT OF DIGITAL LIBRARY AND NATIONAL DATABASE

A digital library is a type of information retrieval system where resources (data) are stored in digital formats and made accessible through the internet. The national database will be established in the form of a digital library which stores many types of heterogeneous data collected from water and waste water utilities across the United States and also make them available to the stakeholders and the research team through the internet. Figure 2 shows the main user groups and the overall basic architecture of the National Database Digital Library (NDDL). The architecture consists of two major components; the front-end website and the back-end digital library. The details pertaining to the user groups and the two major components of the NDDL are described in the following subsections. Figure 2 shows the tentative structure of the NDDL website which shows its basic contents and functionality. Please note that Figure 3 shows only a 'Home-Page' and additional pages to navigate the data will be developed. The website structure will be updated or modified later according to user study results or any requests or suggestions made by the research group or participating utilities. Figure 4 illustrates the basic architecture of the two preliminary websites. The user can choose a link to access either water or wastewater information from the main page. Each page is then divided into three categories – transmission mains, distribution lines, and service lines for water, and force mains, gravity sewers, and sewer laterals for wastewater. By choosing one of the links in the categories, the user can reach the list of products or technologies and can retrieve the information of interest such as technical data sheets and relevant case studies.

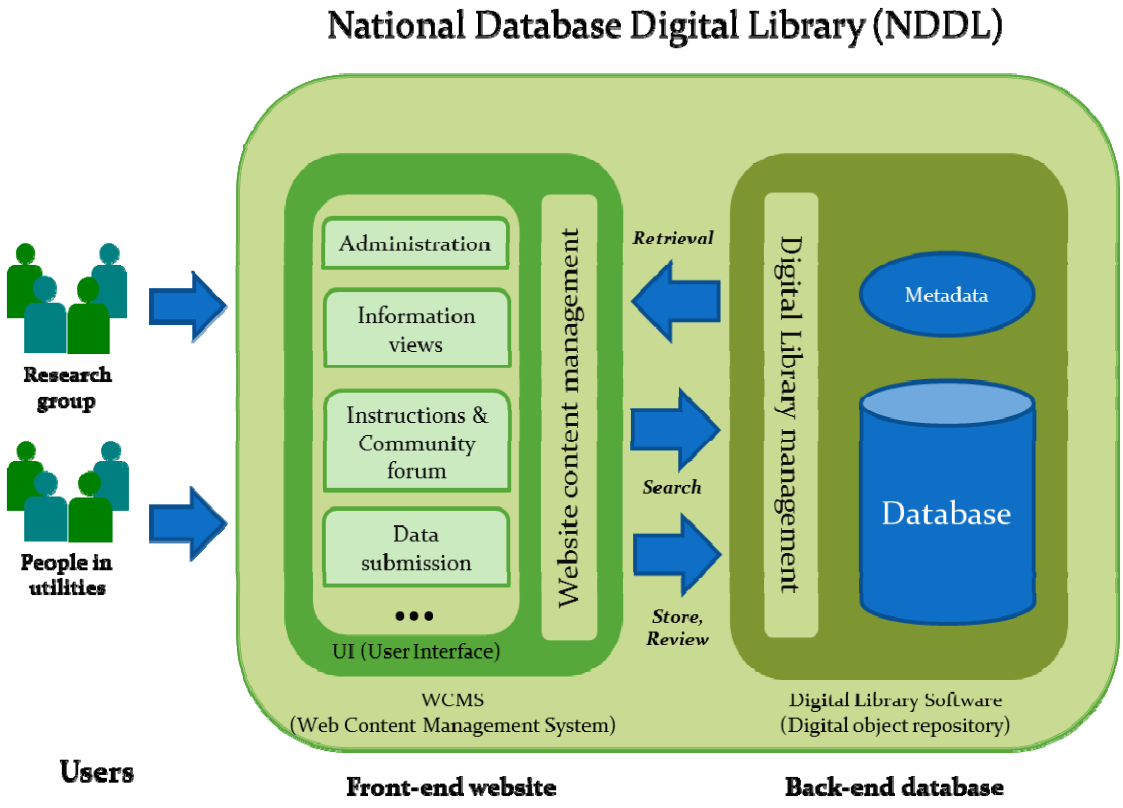


Figure 2. Overall Architecture of the National Database Digital Library

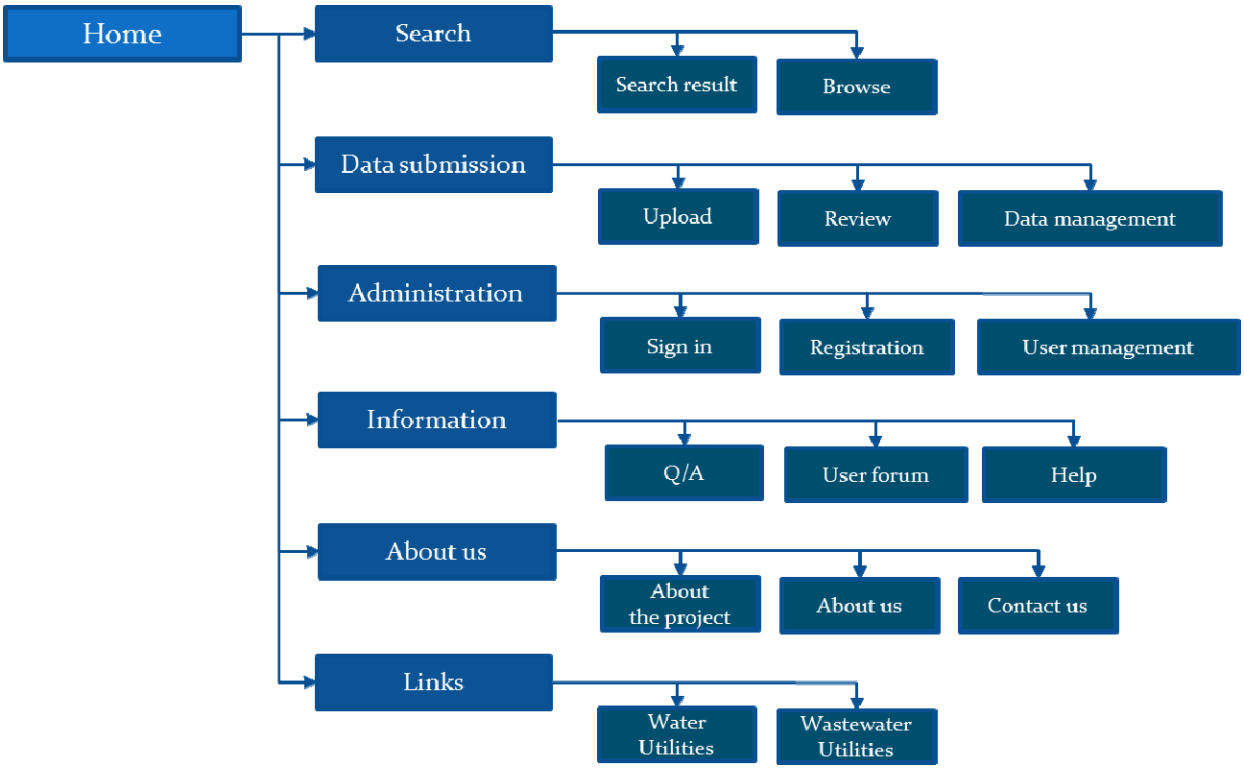


Figure 3. Website Hierarchy of the NDDL

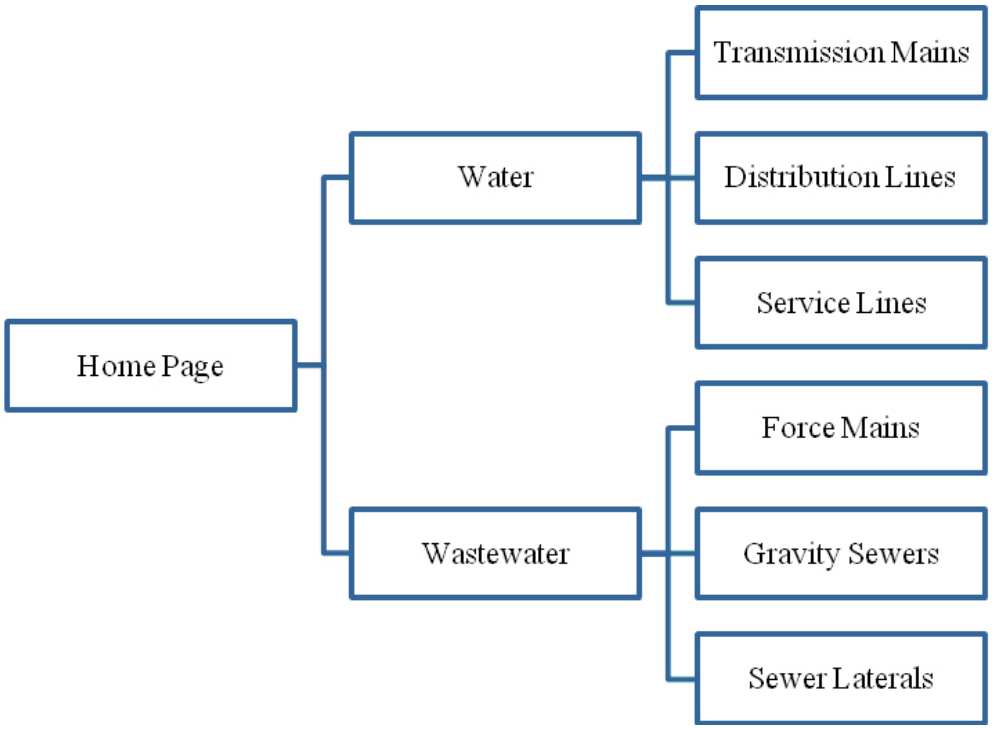


Figure 4. Basic Architecture of the Preliminary Websites

## VALUE OF THE PROPOSED NATIONAL DATABASE

Drinking water and wastewater infrastructure is one of the single largest public investments in the developed world. We have now entered the "dawn of the replacement era" and utilities are rapidly facing new challenges for which there is only limited authoritative guidance available. The technologies available for coping with these new challenges are themselves developing and changing rapidly. Publications have out-of-date information often by the time they are published and initially distributed. Trillions of dollars are scheduled to be spent on system expansion and on system renewal and replacement and up-to-date authoritative guidance regarding available Condition Assessment and Renewal Engineering technologies is needed. The pace of change demands a more efficient means of information transfer to assist agencies with optimizing their infrastructure investments. Virginia Tech's data mining efforts are compiling the experiences of utilities nationwide regarding their applications of Condition Assessment (CA) and Renewal Engineering (RE) technologies. Real world applications as described within Case Histories gleaned from the utilities will be used to verify actual capabilities and limitations of the available technologies as documented in the Technology Summaries and Data Sheets, and will be dynamically linked from such National Database documents. Likewise, Management Practice summaries will also link to Case Histories showing real world implementation of the concepts. The relational database will allow relevant database files to be dynamically associated with the appropriate Web Pages, Technology Summaries and Data Sheets, Management Practice Summaries, Case Histories, etc., all of which will be organized within the Website for the National Database. At this most basic level, the National Database will provide a snapshot of the combined current CA and RE programs of utilities across the nation, including their direct experiences with the costs, capabilities, and limitations associated with the various management practices and technologies which are employed. Entries into the National Database will not only be editorially reviewed, but the capabilities of technologies as described in the Technical Data Sheets will be supported by Case Histories which document the utilities' own experiences, and not just vendor claims. The goal is to provide a single, reliable platform for all industry information pertaining to CA and RE that can be successfully leveraged to augment (not replace) traditional information distribution channels, in order to improve timely access to the most current and reliable CA and RE information.

### Establishing a National Data Standard

The Virginia Tech research team recognized an opportunity to achieve far more than basic, static, and stand-alone functionality. The research team identified the need for establishing data standards to reduce the effort and expense required to update this project in the future and to simplify all future data compilation and analysis pertaining to CA and RE. Without standards for data compatibility, researchers currently spend most of their time, effort, and available budgets gaining access to and then compiling information from multiple sources into compatible formats. The research team has envisioned how the National Database can ultimately provide near instantaneous access to CA and RE information of academic interest, by having utilities elect to self-comply with a national data standard that will greatly simplify

data compilation and that will allow for more of the available time, effort, and budgets across the entire industry to be focused upon data analysis. Hundreds of utilities across North America and even more internationally have come to understand the efficiency and the value of following a national data standard through their adoption and use of National Association of Sewer Service Companies (NASSCO) Pipeline Assessment and Certification Program (PACP) system (or Water Research Centre (WRc) Codes internationally) for collecting visual condition indexing information on their gravity sewer pipes. The national data standard incorporated into the National Database can readily be perceived and promoted as a significant extension of such an already internationally adopted standard. With additional standard field definitions and codes, all data and files pertaining to CA and RE projects can be uploaded into a Geographic Information System (GIS) for data management purposes in a compatible format that permits ease of file retrieval and data analysis. In addition to PACP compatibility, by having the national data standard for water and wastewater asset management also be compatible with the Dublin Core (a standard for entering basic library reference data), all utility files following this national data standard will integrate with all of the major digital library systems which follow the Dublin Core.

**Enhanced File Management for Utilities**

As a value-added contribution of the National Database project, the standard data format of the National Database will be adapted to the file management needs of utilities, and will enable utilities to associate all of their files with the appropriate assets within their GIS databases according to their latitude and longitude coordinates or their unique asset identifier codes. The ability to pull all historical files into a GIS data layer from which the files can be linked and accessed will assist utilities in better data management. Currently, utilities have a difficult time managing the large volume of data pertaining to their assets. Current employees frequently do not know what data has been generated by past employees, or how to access the files that contain them. In many cases, valuable historical reports, relevant conference papers, and case histories are completely lost to the knowledge of current staff. Even the personnel in separate departments are generally unaware of valuable data contributions by each other. Agencies will be encouraged to adopt the standard data structure of the National Database for their in-house files in order to gain the ability to use their own GIS database as a file management system to pull up any and all historical documents pertaining to specific assets, including all documents generated by all departments pertaining to the design, procurement, construction, operation and maintenance, and retirement phases of their assets' histories. If utilities choose to follow the standard data structure of the National Database for all of their asset-related files, utilities will not only be able to readily upload and donate their own "public domain" files to the National Database, but they will also be able to more easily leverage the National Database to query for comparable "public domain" files developed and donated by other utilities. Additionally, by following the standard data structure of the National Database, all files created by the utility will be more accessible for scientific analysis in conjunction with files generated by other utilities.

**A Dynamic Information Compilation and Distribution System**

The National Database will also provide an entirely new platform for more efficient capture and distribution of the combined experiences of utilities nationwide. Unlike many other similar databases in other industries, the digital library of the National Database is structured to function as much more than merely the back-end of the related Website; this National Database will be structured as a dynamic information compilation and distribution system. Whereas current publications such as industry Manuals of Practice and Decision Support Tools have limited supporting case history documentation and the associated costs and technology experiences and capabilities are rapidly out-dated, this National Database will be structured to provide the ability to incorporate dynamically linked, and thereby constantly up-to-date, information from the National Database directly into industry publications and software.

Developers of digital manuals and software tools will be able to program to dynamically identify and pull up-to-date Case Histories directly into the interface of their digital publications. Additionally, with the standard field definitions and standard coding of the National Database, developers of digital manuals and software tools will be able to program to dynamically link to specific fields within specific documents contained in the National Database and thereby dynamically pull up-to-date Technology Data directly into their own publications. This inherent capability is similar to how travel websites are able to dynamically pull flight information from the standard databases of the airlines. As a result of the ability to create such dynamic interactions between industry publication and the National Database, when a new Case History is added to the National Database, all linked publications can dynamically include them as well. When a technology's applicability for use changes as confirmed with additional utility experience (ex.: A technology previously limited to 24" is now routinely installing up to 30"), up-dates within the National Database can also keep dynamically linked publications current as well. One can immediately envision how with such integration to the National Database, Manuals of Practice will remain more relevant, more current, and require fewer major revisions, thereby freeing industry resources to focus on the development of completely new resources. Inevitably utilities, industry associations, and consultants will devise numerous ways to leverage the power of up-to-date information access from the National Database within their publications and software tools.

**Prospective Expansion of the National Database**

Prospectively, the National Database can eventually be expanded to include other forms of information as well, for example:

- Standard coded image files of defects and imperfections uploaded from utilities that can be integrated into CA training programs
- Access to public domain animations and PowerPoint presentations which can be utilized as educational tools during the development of CA and RE projects
- Access to public domain public relations publications such as Town Hall meeting presentation templates and door knockers

The ultimate goal of the National Database is to not only make current CA and RE information readily accessible, but to also permanently reduce duplication of effort for the compilation, distribution, and presentation of such industry knowledge.

## CONCLUSION

The research team expects to develop a prototype website for data collection by end of March, 2011. Lists of metadata and brief description for each metadata field will be included in the website design. The website will be deployed in the preliminary independent hardware for future development. By the end of May, 2011, the research team will move all the initially collected data into a robust web-based digital library system with refined metadata and make the National Water Infrastructure Digital Library (Water-IDL) accessible to all participating stakeholders.

## ACKNOWLEDGEMENT

The authors would like to acknowledge the help and support from various participating Utilities across the nation. The authors would also like to sincerely thank WERF and U.S. EPA for funding this very important research project.

## REFERENCES

American Society of Civil Engineers (ASCE), 2000, "Clean Water Infrastructure Financing," Policy Statement 480, Reston, VA.

American Water Works Association (AWWA), 2001, "Dawn of the Replacement Era: Reinvesting in Drinking Water Infrastructure,"

National Association of Sewer Service Companies NASSCO, 2003, "Pipeline Assessment and Certification Program (PACP) Reference Manual," Pikeville, MD.

United States Environmental Protection Agency (U.S. EPA), 2007, "Innovation and Research for Water Infrastructure for the 21st Century – Research Plan," Report.

United States Environmental Protection Agency (USEPA), 2005, "Managing for Excellence: Analysis of Water and Wastewater Utility Management Systems," Washington, D.C.

Water Infrastructure Network (WIN), 2000, "Clean and Safe Water for the 21st Century," Washington, D.C.

Water Research Center (WRc) 2001, "Sewer Rehabilitation Manual," U.K