Effective Strategies For Long-Term Data Preservation And Website Management For The Gelman Site: A Case Study Of 1,4-Dioxane Cleanup Project In Michigan

Xiao Shi
aliciashi2333@gmail.com

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
Shi, Xiao, "Effective Strategies For Long-Term Data Preservation And Website Management For The Gelman Site: A Case Study Of 1,4-Dioxane Cleanup Project In Michigan" (2023). Public Health Theses. 2341.
https://elischolar.library.yale.edu/ysphtdl/2341

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
EFFECTIVE STRATEGIES FOR LONG-TERM DATA PRESERVATION AND WEBSITE MANAGEMENT FOR THE GELMAN SITE: A CASE STUDY OF 1,4-DIOXANE CLEANUP PROJECT IN MICHIGAN

By

Xiao Shi

A Thesis
Submitted to the Department of Environmental Health Sciences
Yale School of Public Health
In Fulfillment of the Requirements for the Degree of Master of Public Health

May 2023

Supervised by Dr. Vasilis Vasiliou & Dr. Ying Chen
Abstract

The Gelman project is a cleanup initiative for 1,4-dioxane contamination caused by Gelman Sciences Inc, which might take years or decades to be finished. Document management becomes a challenge for the Gelman project since a substantial number of documents generated over the years can be lost during organizational changes. Scio Residents for Safe Water (SRSW) is a website that tries to integrate sources from different stakeholders for the Gelman project. Keeping them organized and engaged becomes another question. The thesis project aims to provide website management strategies to enhance user engagement and recommends data preservation strategies for this long-term cleanup project, based on the lessons learned from other initiatives with similar needs.

Acknowledgments

Firstly, I would like to express my thanks to my thesis supervisors, without whom I would not have been able to complete this research. I am deeply grateful to Dr. Vasilis Vasiliiou. Your constant guidance and encouragement gave me the courage and determination to navigate the challenges when I need to change my thesis topic. I am honored to have had the opportunity to learn from you. I will always cherish the lessons I have learned under your mentorship. I want to express my sincere gratitude to Dr. Ying Chen for the positive impact you have on my life. Your mentorship has been a source of unwavering support, much like the warmth and care of a mother. Your high standards and sage advice have helped me develop discipline and resilience.

Secondly, I would like to extend a special thank you to the chair of Coalition for Action on Remediation of Dioxane (CARD), Roger Rayle, for guiding me on how to approach the thesis question and generously dedicating his time to provide me with timely feedback as well as detailed comments on my
thesis out of busy schedule. Thirdly, words cannot express my gratitude to my family - my parents, grandparents, and my furry companion. My courage to explore my life and the world stems from the unwavering love and the freedom you gave me. I am so lucky to have my dear friends Juanyi, Xi’ai, Wenya, Xuan, Yiran, Liuyi, Yingxu, Mingyang, Jiaying, Guicheng, Jiawei, Lexin being there for me through all the ups and downs.

Table of Contents

Abstract ............................................................................................................................................. 2
Acknowledgments .......................................................................................................................... 2
Introduction ...................................................................................................................................... 4
  Emerging Water Contaminant 1,4-Dioxane and Human Health Hazards ........................................... 4
  Background of Gelman Project ....................................................................................................... 5
  Challenge I: Website Management ................................................................................................. 6
  Challenge II: Long-Term Data Preservation .................................................................................... 7
Objective/Methodology .................................................................................................................. 7
  Thesis Objective ............................................................................................................................. 7
  Methodology .................................................................................................................................... 7
Website Management ..................................................................................................................... 8
  Content Analysis ............................................................................................................................ 8
    General Content .......................................................................................................................... 8
    Figure 1 ......................................................................................................................................... 9
    Maps ............................................................................................................................................ 9
    Table 1 ......................................................................................................................................... 11
    Documents .................................................................................................................................. 11
    Table 2 ......................................................................................................................................... 12
  Management Strategies ................................................................................................................ 13
    Lessons from the website of Hinkley Chromium Contamination Cleanup .................................. 13
    Lessons from Superfund Community Advisory Group ............................................................... 13
    Lessons from Web-based Health Care Patient Information ........................................................ 14
Long-term Data Preservation ....................................................................................................... 14
  Lessons from Other Projects ........................................................................................................ 15
    Lessons from EPA: Superfund Sites ......................................................................................... 15
    Lessons from DOE: Restoration Activities ................................................................................ 15
    Lessons from NEA: Radioactive Waste Management ................................................................. 16
Introduction

Emerging Water Contaminant 1,4-Dioxane and Human Health Hazards

1,4-dioxane has been a chemical used for industrial manufacturing of chlorinated solvents as a stabilizer\(^1\), which is highly soluble in water, easily seep into and remain in groundwater at high concentration\(^2\). Moreover, 1,4-dioxane is persistent in water since it is difficult to be eliminated with conventional treatment methods. With respect to its wide use, specific characteristics, and persistence, 1,4-dioxane can lead to considerable environmental release, specifically in the water. For instance, 1,4-dioxane is found at more than 20% of tested sites in an Environmental Protection Agency (EPA) Third Unregulated Contaminant Monitoring Rule (UCMR-3) testing\(^3\).

---


1,4-dioxane is regarded as a water contaminant of concern regarding the potential health hazards it poses to the population. Although there are no human toxicological studies assessing toxicity with oral 1,4-dioxane exposure, clear evidence has shown its liver toxicity in animal studies\(^4\) and it is classified as a likely human carcinogen (Group 2B)\(^5\). Recently, increasing attention has been given to 1,4-dioxane as an emerging contaminant. For example, Agency for Toxic Substances and Disease Registry (ATSDR) listed 1,4-dioxane as one of the Superfund priority substances\(^6\). Regulatory efforts have been made to remove 1,4-dioxane across the United States.

**Background of Gelman Project**

A cleanup project for a persistent pollutant like 1,4-dioxane can take years or decades. The Gelman project, a comprehensive cleanup initiative that aims to remediate the contamination caused by Gelman Sciences Inc., might be an example. Gelman Sciences Inc, a former medical filter manufacturer situated near Ann Arbor, Michigan, used the industrial solvent 1,4-dioxane from 1966 to 1986 and left significant contamination in the groundwater\(^7\). The large amount of 1,4-dioxane used over 20 years caused serious groundwater contamination. In 1988, for example, the detected 1,4-dioxane level is above 200,000 \(\mu\text{g}/\text{L}\)\(^8\). The contaminated area has gradually increased to an area of approximately 1 mile

\(^4\) Chen, Y. et al. (2022) “Oxidative stress and genotoxicity in 1,4-dioxane liver toxicity as evidenced in a mouse model of glutathione deficiency”, *The Science of the Total Environment*, 806(Pt 2), 150703.


\(^7\) Jackson, L. E. et al. (2022) “Evaluation Of 1,4-Dioxane Attenuation Processes at The Gelman Site, Michigan, USA”, *The Science of The Total Environment*, 823, 153634.

wide and 4 miles long\textsuperscript{9}. Moreover, since it is difficult to remove 1,4-dioxane through most water treatment process\textsuperscript{10}, the cleanup efforts last for decades and have not been finished yet.

**Challenge I: Website Management**

This project encompasses different entities, leading to a complex and diverse source of documents. Stakeholders involved in this process include EPA, City of Ann Arbor, Scio Township, Ann Arbor Township, Washtenaw County and its Circuit Court, Gelman Sciences, Huron River Watershed Council, Coalition for Action on Remediation of Dioxane (CARD), the State of Michigan and its Department of Environment, Great Lakes, and Energy (EGLE). Each stakeholder has its website.

Regarding the substantial number of stakeholders, it is significant to keep stakeholders organized. Scio Residents for Safe Water (SRSW), a non-profit organization established by citizens in 1995 that maintained a website since its creation to keep the community informed, intergrading sources from different stakeholders\textsuperscript{11}. It is important to examine the elements covered by the website of each stakeholder. It would also be beneficial when EPA takes over administration of the site and CARD morphs into an EPA citizen engagement group with a centralized website to supplement the website that EPA would provide in the future.


Challenge II: Long-Term Data Preservation

The initiation of cleanup efforts commenced in 1992, as by the consent judgment requiring remediation action\textsuperscript{12}. The most recent millstone is that, in 2021, a letter has been sent to EPA to request the Gelman site to be assessed for the National Priorities List (NPL), which is the list of sites of national importance across the United States and its territories where hazardous substances, pollutants, or contaminants have been released or are at risk of being released\textsuperscript{13}. A substantial number of documents can be generated over the years. Organizational changes are likely to happen during the years and documents might be lost during the process. Therefore, how to maintain records and preserve data for the Gelman project becomes a challenge.

Objective/Methodology

Thesis Objective

1. Provide website management strategies to enhance user engagement with the content

2. Offer recommendations on data preservation for the long-term cleanup project from the perspective of various elements of data

Methodology

I searched Google and Google Scholar using the following key words: “long-term program”, “cleanup project” “data preservation”, “record maintenance”, their synonyms and related terms. Moreover, I

\textsuperscript{12} City of Ann Arbor (n.d.) Water Treatment, Gelman 1,4-Dioxane Litigation [Online] Available from: https://www.a2gov.org/departments/water-treatment/Pages/Gelman-1,4-Dioxane-Litigation.aspx (Accessed from: 28th April 2023)

analyzed the official website of 7 stakeholders of the Gelman project (EPA, City of Ann Arbor, Scio Township, Ann Arbor Township, Washtenaw County and its Circuit Court, Gelman Sciences, Huron River Watershed Council, CARD and EGLE).

Website Management

Content Analysis

General Content

The content covered on the websites of stakeholders related to the Gelman project is shown in Figure 1. Each column denotes a stakeholder while each row represents a specific piece of information, such as brief introductions, event timelines, documents release, maps update, conferences or public hearings, frequently asked questions, or contact resources. Yellow cells mean the information is covered on the website as an original source while green cells suggest that the content is covered by citing from other stakeholders. Blank cells show that the content is not covered on the website. It can be seen from Figure 1 that almost all stakeholders include brief introductions, the timeline of the Gelman project, well-sampling information, contact and resources. Most stakeholders cross-reference each other and provide answers to frequently answer questions. CARD, the partnership of local governments and citizens that aims to address the Gelman contamination14, is mentioned by more than half the stakeholders. Each stakeholder provides their perspective within their respective areas of authority, and these perspectives complement each other. Among all these websites, SRSW is the one that includes all elements.

Among these elements, maps warrant further investigation because it contains key geospatial information, clearly informing the public about the distribution of pollution. Table 1 compares 1,4-Dioxane Contamination Maps of the Gelman Site on the website of EGLE, Washtenaw County, and SRSW. The map of Washtenaw County is a static map on its website while EGLE and SRSW use other geospatial systems for their interactive maps. The map provided by the SRSW represents the most recent and comprehensive depiction of the area of interest, encompassing the longest span of time among the available options. In contrast, the map of Washtenaw County has become outdated, as it only reflects the 85 parts-per-billion (ppb) concentration threshold zone and fails to display the recently expanded Prohibition Zone\(^{15}\).

As for the information in the map, all websites provide a plume map, concentrations of 1,4-dioxane of certain wells (Monitoring Well, Extraction Well, Miscellaneous Well, Residence Well), and surface water. EGLE and Washtenaw County both provide more information about the concentration of 1,4-dioxane in Allen Creek Drain. SRSW presents the concentration change by time and elevation. EGLE offers additional information about the concentration of injection/horizontal wells.

<table>
<thead>
<tr>
<th></th>
<th>EGLE(^{16})</th>
<th>Washtenaw County(^{17})</th>
<th>SRSW(^{18})</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform</strong></td>
<td>ArcGIS</td>
<td>NA</td>
<td>Tableau &amp; Downloadable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Google Earth Mashups(^{19})</td>
</tr>
<tr>
<td><strong>Interactivity</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Updated Time</strong></td>
<td>2020</td>
<td>2019 (out-of-date(^{20}))</td>
<td>2022</td>
</tr>
<tr>
<td><strong>Covered Time</strong></td>
<td>2020</td>
<td>2019</td>
<td>1986-2021</td>
</tr>
<tr>
<td><strong>Key Info</strong></td>
<td>1. Plume Map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Concentration of 1,4-dioxane of Surface Water</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Wells: Monitoring Well, Extraction Well, Miscellaneous Well, Residence Well</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Documents also require further examination because it is one of the most complicated elements. As shown in Table 2, EGLE has a database inside of its website with 1,296 documents\textsuperscript{21}, whereas SRSW\textsuperscript{22} has a Google doc compiling 2529 documents (data accessed: April 17, 2023). Specifically, the 2529 documents were from a previous document list of EGLE before it changed to its current website. These 2529 documents include the 1296 documents on EGLE's current list with links. However, copies of the remaining documents do not have links and can only be obtained from the "Information Repositories" listed on the EGLE Gelman website or from the separate repository of scanned copies of SRSW.

Considering the substantial volume of documents of EGLE and SRSW, efforts have been made to summarize them into categories. Basically, the documents of SRSW and EGLE can be classified as data, legal, communication, and other.

\begin{table}[h]
\begin{tabular}{|c|c|c|}
\hline
Injection Well, & Allen Creek Drain & Concentration by time  \\
Horizontal Well, Allen & Concentration by elevation & \\
Creek Drain & & \\
\hline
\end{tabular}
\caption{Gelman Site Contamination Maps: Stakeholder Websites Comparison}
\end{table}


\textsuperscript{24} Scio Residents for Safe Water (n.d.) \textit{Scio Residents for Safe Water Website} [Online] Available from:
<table>
<thead>
<tr>
<th>Data</th>
<th>Data</th>
<th>Analytical Report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Well Log/Record</td>
</tr>
<tr>
<td>Legal</td>
<td>Legal</td>
<td>Legal (Court Orders, Stipulated Orders, Affidavit, Dispute Resolutions)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Permits and Approvals</td>
</tr>
<tr>
<td>Communication</td>
<td>Periodic E-mail Updates</td>
<td>Letter &amp; E-mail &amp; Communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Comments and Responses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Meeting Agenda/Memo/Notice</td>
</tr>
<tr>
<td>Other</td>
<td>Selected Documents</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recent Documents for Information Repositories</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2: Gelman Site Documents: Stakeholder Websites Comparison**

By aggregating and synthesizing data from diverse sources, SRSW makes the website informative, providing users with a more robust and multifaceted understanding of the subject. However, including too much information on a centralized website can overwhelm visitors, be difficult to maintain and update. Therefore, it is essential to strike a balance between providing sufficient information to users and avoiding overwhelming them with an excessive amount of data. Furthermore, managing a website with copious amounts of data necessitates the employment of effective website management strategies to improve visitor engagement with the content.

---

https://www.srsw.org (Accessed from: 28th April 2023)
Management Strategies

Lessons from the website of Hinkley Chromium Contamination Cleanup

An integrated map searcher, a clean web outlook and a new user guide might improve user loyalty in the long term. For the Hinkley Chromium Contamination Cleanup project, all documents are available at the Lahontan Water Board office in Victorville and on the State Water Board’s database called Geotracker\(^{25}\). Different from SRSW, Geotracker used an integrated map searcher with a clean outlook. SRSW users might need to manually check the map to find the site information while Geotracker users can find specific site information without knowing the exact location. Moreover, the simplified web outlook is more user-friendly in the sense of lessening the cognitive burden and stressing the most important information. The new user guide serves a similar function by reducing the potential confusion and anxiety for new users.

Lessons from Superfund Community Advisory Group

A public discussion board or similar interactive platform for expressing concerns related to this project might facilitate the participation of stakeholders, especially residents living in proximity to the Gelman site. The EPA Superfund has a Community Advisory Group that provides a public forum for community members to present their needs\(^ {26} \). The rationale of an advisory group can be used online as well. Using the discussion board will save time for residents because the discussion board enables them to contact relevant stakeholders with a single post instead of multiple e-mails.


Lessons from Web-based Health Care Patient Information

The Gelman project website should improve its adaptability to various devices and browsers. A mobile app would be optimal regarding the increasing number of phone users globally. While the Gelman project is not directly related to healthcare, the principles identified by Oktay (2021) that would improve patient engagement for websites related to healthcare information, can still be used as principles for the Gelman project website management. Oktay (2021) proposed that adaptability, accessibility, values of content, and credibility as important factors.

Long-term Data Preservation

Notably, the efforts described in the previous session of putting all the information into a website is a significant way of data preservation. The information has been both conserved on the website and widely disseminated to the public. During the dissemination process, the information has been transformed into public knowledge and memory using humans as information carriers. Moreover, keeping stakeholders informed is very likely to encourage their participation.

Lessons from Other Projects

Lessons from EPA: Superfund Sites

Gelman site is not yet on the EPA National Priorities List. However, it can learn from EPA, which has an established and standard data storage process that the Gelman site can learn from, including data catalogs\(^28\), data standards\(^29\), open data access\(^30\) and superfund sites glossary. Given that the Environmental Protection Agency (EPA) manages numerous Superfund sites, the data management process involved is likely to be considerably more complex than that required for a single site such as Gelman. Moreover, EPA includes administrative documents specifically for sites on NPL, such as decision documents, which might not be applicable to the Gelman site\(^31\).

Lessons from DOE: Restoration Activities

As early as 2002, the United States Department of Energy (DOE) proposed the need for long-term data preservation and archiving for environmental restoration activities in locations throughout the United States\(^32\). DOE specifically focuses on spatial data, which is also a prominent component of the data of the Gelman project. Multiple stakeholders use advanced geospatial analysis software tools, such as ArcGIS and Tableau, to create comprehensive and informative maps.

---


Lessons from NEA: Radioactive Waste Management

Similar needs have been raised in the field of radioactive waste management. The Organization for Economic Co-operation and Development (OECD) Nuclear Energy Agency (NEA) initiated a project named Preservation of Records, Knowledge, and Memory (RK&M) across generations, which stresses using systematical strategies and multiple approaches that are complementary to each other.\(^{33}\) The records type and preservation time span might be different for the Gelman project and the RK&M project. Moreover, the preservation time of Gelman project data is shorter than the nuclear waste. Nevertheless, the principle of utilizing diverse media, involving various stakeholders, and maintaining data in multiple locations remains unchanged. This thesis will examine the lessons learned from superfund sites, restoration activities and radioactive waste management, from the aspect of different data elements.

Preservation Strategies

Given that there is overlap in content provided by EPA, DOE, and NEA, this research will not discuss the lessons of long-term data preservation learned from these agencies separately. Instead, the preservation strategies will be examined from the perspective of different elements of data.

Data Management Plan: Ideally, before the start of the project, the Data Management Plan (DMP) should be established, which outlines how the data is going to be maintained throughout the project.\(^{34}\)


The DMP should address all aspects of data storage, including security, access, and retrieval. Maintaining this plan throughout the project life cycle, achieving shared understanding among the internal team members, and easy communication with external stakeholders are essential for the success of data preservation. This plan should be regularly updated and reviewed to ensure its relevance to the organization's changing needs and objectives.

**Data Media**: Data should be stored in multiple secure places through various media. Tangible media, such as stone and paper, were popular in previous times. Modern media includes sound recordings, videotapes, and computer-based documents. The Gelman project comprises a significant volume of documents that lack an electronic version, suggesting that the project is still undergoing the transition from paper to electronic records as the recordkeeping medium. It is important to note that the transition should not be misconstrued as evidence of the superiority of electronic media over physical media. Although digitization enables the storage of larger quantities of data, easier replication of data, and more efficient data search, its fragility, the need for continuous maintenance, and its challengeable data protection can be a problem. Physical media, or tangible media, is sometimes more durable than electronic media. For instance, permanent paper can last more than 600 years. Moreover, the physical record is easier to be read and easier to be preserved from a technical perspective. To reduce the risk of data loss due to media failure, it is recommended to utilize at least two different types of media (e.g., disk and paper). Moreover, data should be stored at separate locations. This redundancy approach can

---

significantly increase the reliability and durability of the data, especially when dealing with critical or sensitive information.

**Digital Obsolescence:** Technology sustainability evaluation and periodic data migration are necessary for mitigating technology obsolescence. The data media, format, or other technologies related to data preservation, might be outdated. Therefore, first, the sustainability of data-preservation technology should be evaluated before usage. Second, data migration, defined as organized tasks to transfer between different hardware/software, or, from one generation of technology to a subsequent generation, should be applied. Data migration is labor-intensive and time-consuming. Moreover, it requires manual verification of data integrity of the data and identification for errors. Hence, the conservation of data is contingent not solely upon technological aspects but also on the human element, emphasizing the significance of personnel involved in the process.

**Data Standards:** For digitalized documents, a common set of standards should be established to ensure the consistency of where and how the data will be stored and organized among team members. Data standards, defined by EPA as “documented agreements on representation, format, definition, structuring, tagging, transmission, manipulation, use, and management of data”, allow for the same code and presentation mechanisms to be used for data access. Additionally, it promotes transparent and clear meanings for data that can be reused, provide consistent results during data retrieval, and allow for the same terms, codes, and data structures to be used for data access. Standard, well-supported and stable

---

file formats should be used, so that it is compatible with other systems or applications. Additionally, the utilization of standardized file formats is crucial for data exchange and cooperation. It is also important to ensure that the file is readable for operating systems and software in the future.

**Metadata:** Ideally, each information object should be self-explanatory and can be understood independently\(^{41}\). Self-contained data enables easy data interpretation, data migration and data use regardless of the context in the future. However, completely independently understandable data may not be feasible under certain circumstances, where documentation accompanying datasets describing the data should be in place. The documentation should be clear, well-organized, and concise. Like the use of standard files, well-written documentation can facilitate scientific research and cooperation\(^{42}\). Other collaborators can understand the data without further explanation. Moreover, these documents are likely to help future users identify errors in the data. To avoid data loss, regular data backups are necessary\(^{43}\).

**Data Access:** It is essential to establish guidelines pertaining to data access. For private data, it is crucial to determine who can access data. For semi-public data, it is a common practice to redirect the data request to certain personnel for approval in the research field\(^{44}\). For publicly available data, a web-based portal should be utilized for users to search, discover, and obtain information about the dataset.

---


Discussion and Conclusions

After analyzing the website content of each stakeholder in the Gelman project, it has been found that, among the 7 stakeholders, SRSW has the most comprehensive website that contains all basic elements. Apart from SRSW, the website of EGLE is also complete in the sense that it includes a documentation database and map, which are two significant website components for the Gelman project. To better manage the SRSW website, strategies including a cleaner web outlook, a more interactive platform and better technical adaptivity are recommended.

Regarding the challenge of data loss over the long remediation course, record preservation strategies, namely a data management plan, careful choice of media, technology sustainability evaluation, periodic data migration, common set of standards, self-contained data, guidelines related to data access, have been proposed.

This thesis projects provide a preliminary review of the websites related to the Gelman project, general suggestions about website management, and long-term digital preservation. Future studies are needed to examine the existing situation of data related to Gelman in more detail, which can be achieved by conducting interviews with stakeholders and data experts who are familiar with the project. Moreover, prospective research should discuss the methods of implementing the strategies issues from a more technical perspective. Subsequent research could investigate the legal and ethical issues associated with ownership, sharing, and safety of data when EPA takes over administration of the site and merges different data sources.