LEADING PIPELINE RESEARCH

Pipeline Research Council International

Data Management Plan

Distribution: Public

Last revised: 7/24/2025 8:49:00 AM

Authors: Gary Choquette

Contents

1. Executive Summary 3
2. Purpose and Background 3
3. Scope 4
4. Requirements 4
Appendix A – DMP Template 7
Appendix B – Data Transmittal Example 8
Revision Log 10

1. Executive Summary

Information becomes knowledge, knowledge becomes power. We have all heard that phrase. But where/how do we get the information? Information comes from data; data being unprocessed facts. It is only in the process of organizing and analyzing data that it becomes information. Software is one tool for converting data into information.

This document is intended to provide an overview of the Pipeline Research Council International's (PRCI) process of developing software development and associated data management. The review includes lessons learned and recommendations to improve the process with the goal of a more efficient transformation of data into power for the pipeline transportation industry.

With the current state of software development tools and network connectivity, the time has come for PRCI to focus on developing cloud-based (hosted) applications. Hosted applications have some advantages over desktop applications: They can be less expensive, simpler to manage, and easier to update and use. Desktop or mobile applications (or hybrid applications) may still be preferable to hosted applications network connectivity is problematic.

PRCI is developing a data storage, reporting, and retrieval platform commonly referred to as the Virtual Technology Development Center (VTDC). The intent of this effort is to offer a repository of key information produced by PRCI member companies and/or research contractors to ease future data analysis and data mining. To facilitate the development of the VTDC, there are many implementation issues that overlap the software development. The approaches outlined here are intended to accommodate PRCI's software applications, the VTDC, and the management of other PRCI data.

2. Purpose and Background

Producing research data can be a major expense of a project. Capturing the research data and including it as a project deliverable is an expectation of PRCI for two main reasons:

- If there are any questions or concerns with the conclusions or recommendations produced as part of the research, they can be independently verified against the source data.
- The dataset can potentially be used for subsequent research and model validation.

Safeguarding the privacy of participants and protecting confidential and proprietary data is important. To help this process, PRCI requires research proposals to include a plan for managing and providing access to final research data or to state why their data cannot be made available.

This plan is based on the principles:

- Protecting confidentiality and personal privacy is a cardinal precept.
- Sharing and preserving data are central to protecting the integrity of research by facilitating independent validation of results.
- Data management should be an integral part of the research planning process.
- The extent of which research data needs to be shared or preserved varies by project; allowances must be made to meet the specific requirements and needs for the corresponding research. Factors to consider include the benefits and costs, including preserving and promoting competitiveness as well as enhancing the safety, integrity, and environmental footprint of pipeline systems.
- Proprietary interests, business confidential information, intellectual property rights, and other relevant rights will continue to be recognized and appropriately protected.
- Capturing the data gathered as part of the project is important such that independent analysis or additional analytics can be performed.
- Effective data management has the potential to increase the pace of discovery and promotes the efficient and effective use of research funding and resources.

3. Scope

PRCI's definition of data follows that of the US Government's Office of Management and Budget:

> *Research data are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples).*
>
> *Research data also do not include:*
>
> *(A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and*
>
> *(B) Personal and medical information and similar information[,] the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.*

Data are understood to include the recorded technical information and the associated metadata (describing the data including but not limited to the associated engineering units, calibration information, uncertainty, date gathered, and other relevant data, such as gas composition if applicable, that is constant throughout the test), descriptions of the software required to read and use the data, and associated software documentation.

4. Requirements

All research ideas, proposals, and project plans submitted to PRCI will be required to include a separate stand-alone document titled [Project Title] Data Management Plan (DMP). The DMP should describe whether and how data generated through the course of the proposed research will be shared and preserved (including the corresponding timeframe) or explain why data sharing and/or preservation are not possible or appropriate. At a minimum, DMPs must describe how the researcher will manage data such that it is persevered for potential sharing for the purposes of allowing for independent validation of published results.

DMPs must provide a plan for making research data that underlie the results and findings in peer-reviewed publications digitally accessible at the time of publication or within three months after publication. This includes data (or how to access data) that are displayed in charts and figures. This does not include preliminary work, work notes, draft work, test plans, peer reviews, emails, or laboratory specimens. This requirement could be met by including the data as an 'other final deliverable' posted to PRCI's website.

DMPs will be reviewed as part of the PRCI proposal/project plan review process. PRCI Program Managers will provide guidance to proposers and awardees, as well as monitor compliance with DMPs.

The DMP will identify any sensitive subsets of data and who should/should not have access to that information.

Whenever possible, a DMP should follow any community-based standards. A sample template can be found in Appendix A – DMP Template.

Access to the data will be determined based on guidance provided in the corresponding PRCI Technical Committee Charter.

In rare cases, exceptions to the requirement to develop a DMP may be allowed. Those exceptions must be justified, reviewed, and approved by PRCI's Vice President of Operations.

Upon the completion of the work, the data will be submitted to PRCI via its website (with the appropriate security visibility set). The data will be zipped into a single file which also includes a transmittal sheet

outlining what data is being provided and its metadata. An example template for this transmittal letter can be found in Appendix B – Data Transmittal Template. The plan should detail what special data security requirements are expected for the data as well as any related protocols for protecting sensitive data.

The plan should outline the type of data that will be gathered as part of the research project, how the data will be stored (e.g., SQL database, Excel spreadsheet, csv, JSON, etc.), and the approximate size of the data to be gathered.

The plan should address data quality management as well as any associated metadata (including the uncertainty/accuracy of the data).

5. Confidentiality

PRCI has confidentiality provisions in its contract with the research contractor. The DMP must not conflict with those provisions.

In many cases, additional data may be provided to a research effort by others (not the research contractor), for example, pipeline operating companies, equipment suppliers, etc. In those cases, the data to be shared is expected to comply with the project DMP. A non-disclosure agreement may be required to obtain that data. Non-disclosure agreements should not be executed directly between the research contractor and these supplying the data. All non-disclosure agreements must be executed between PRCI and the related party. The PRCI researcher is a party to that non-disclosure under the terms of their contract with PRCI. All requests made to a PRCI research contractor for the execution of a non-disclosure agreement should be referred to the PRCI project management support. Some PRCI members have standing (evergreen) non-disclosure agreements with PRCI that only require the completion of an exhibit detailing the project specific data provisions.

In some cases, a formal Equipment Supplier Agreement may also be required if external parties are loaning or providing equipment that will produce data for a research project. Note that Equipment Supplier Agreements are required regardless of PRCI membership status (i.e., PRCI members providing equipment for a project should have an executed Equipment Supplier Agreement).

6. Best Practices for Producing and Maintaining High-Quality Data

Producing and maintaining high-quality data is essential for ensuring the integrity and reliability of research findings. Here are some best practices to follow:

6.1. Data Collection

Use consistent and standardized methods for data collection to ensure uniformity and comparability of data across different studies. Maintain detailed documentation of data collection procedures, including any instruments or tools used, to facilitate reproducibility and transparency.

6.2. Data Quality Management

Establish quality control measures to identify and correct errors or inconsistencies in the data. This may include regular audits, validation checks, and peer reviews. Capture and maintain comprehensive metadata that describes the data, including information on data sources, collection methods, and any transformations applied.

6.3. Data Storage and Preservation

Select suitable storage solutions that ensure data security, accessibility, and scalability. Options are discussed above.

Implement robust security measures to protect sensitive data from unauthorized access or breaches. This includes encryption, access controls, and regular security assessments.

6.4. Data Sharing and Accessibility

Develop a plan for sharing data with the research community, including timelines for making data accessible and any restrictions on access. Adhere to community-based standards and best practices for data sharing and preservation whenever possible to ensure compatibility and interoperability.

As a best practice, establish color codes or add footers to indicate the security requirements for a dataset (e.g., code data as ‘red’ for data protected by a non-disclosure agreement).

When data is blinded, it is best practice to have a cross-reference database such that a blinded record can be traced back to the original unblinded data record.

6.5. Data Documentation

Keep thorough documentation of all aspects of data management, including data collection, processing, storage, and sharing. This documentation should be updated regularly to reflect any changes or updates.

Include clear instructions for accessing and using the data, as well as any special handling or protection requirements.

6.6. Data Blinding

Blinding source data from identifying information may be required for maintaining privacy or to meet the obligations of non-disclosure agreements. One effective method is converting names to generic identifiers, such as replacing proper names, equipment, or location names with codes or pseudonyms. This ensures that individuals cannot be directly identified from the data. Another technique involves scaling data, which can be particularly useful when it does not alter the conclusions or results of the research. For instance, numerical values can be adjusted to a different scale while preserving the relative differences between them. Additionally, removing or translating location information, such as latitude and longitude coordinates, can further anonymize the data. This can be done by either omitting the location details entirely or converting them to broader geographic regions that do not pinpoint exact locations.

6.7. Regular Review and Updates

Implement regular review processes to assess the quality and completeness of the data management plan. This may involve asking for feedback from researchers and incorporating best practices from the field. Revise and update data management plans as needed to ensure they remain current and relevant. This includes addressing any new challenges or opportunities that arise.

Appendix A – DMP Template

PRCI has posted a skeleton of a data management plan that can be used as both a template and a skeleton of a data management plan. It can be found online at: Template - PRCI Data Management Plan Skeleton.

Note that any content in the skeleton between square brackets ([]) inclusive of the brackets should be replaced with content specific to the research project.

Appendix B – Data Transmittal Example

The information provided below serves as a template for the official submittal of data upon the completion of a PRCI research effort. All information in the square brackets ([]), INCLUSIVE OF THE BRACKETS, should be replaced with project specific information.

Pipeline Research Council International Data Management Plan

Project number: [ProjectCode]

Project Title: [ProjectTitle]

By: [Last revised by name, date]

Herein is the data gathered under this research project. Related metadata information can be found [provide details here].

[Provide any additional notes or explanations of the data as needed and any special instructions for data handling/protection/blinding].

Data sensitivity: this data is [confidential and subject to the terms of the XXX non-disclosure agreement, not sensitive, etc.]

Revision Log

Date	By	Summary of changes
5/17/2020	G Choquette	Initial draft for review sans templates/examples
12/21/2020	G Choquette	Added a confidentiality section and the Appendices
7/23/2025	G Choquette	Added more details and general revisions. Deleted content on a draft management plan and replaced with a link to the data management plan skeleton posted on PRIME.