



October 5, 2023

Data Management Plan to Optimize an Industrial Extrusion Process

A Data Management Plan created with DMP Assistant

Data Management Planning Expert Group



Information

Summary

This data management plan was created using the template provided on the DMP Assistant platform. In addition to its creators, the research data management working group of the communauté de pratique des conseillères et conseillers à la recherche [research advisors' community of practice], which is led by the Association pour la recherche au collégial, helped draft the plan. This document will be applied to research led by a centre collégial de transfert de technologie [college technology transfer centre] (CCTT) in Quebec. The research is being conducted in partnership with a manufacturing company. The raw data gathered as a result of this research will be freely available under a Creative Commons licence.

Administrative details

Project name: Optimizing an Industrial Plastics Engineering Process

Creators: Marie-Josée Fraser, David Brodeur and Vincent Thomasset-Laperrière

Institution: CEGEP de Sherbrooke

Funding agency: Quebec's Ministry of Economy and Innovation and Force.IA

Description: A company requested [Productique Québec](#)'s assistance in optimizing its plastics engineering process. The quality of its product varies based on multiple environmental factors. As a result, production parameters must be adjusted to ensure that manufacturing tolerances are respected and that the final product is consistent. It is not easy for operators to manually adjust the production parameters quickly, due to the many physical factors that must be taken into consideration. This activity requires considerable time and wastes large quantities of material at the start of production. To resolve this issue, Productique Québec would like to instrumentalize this process and analyze the data obtained using artificial intelligence techniques, which provide operators with real-time feedback to quickly adjust production parameters as needed. Relevant data includes product specifications, production parameters, final inspection measures and environmental conditions.



Data Collection

What types of data will you gather, create, combine, acquire or log?

A clear understanding of the types of data you will collect, generate and/or acquire during your research will help you think about and plan for effective data management during the project.

Data gathered will include:

- Design data (configurable computer-aided design [CAD] files)
- Technical specifications for parts produced using set criteria
- Production parameters
- Environmental conditions
- Inspection measures

The types of data gathered will be:

- Digital data
- Tabular data
- Graphic data

In which file formats will your data be collected? Will these formats ensure that the data can be re-used, shared and accessed in the long term?

To ensure that the data can be re-used, shared and accessed in the long term, it will be collected in the following non-proprietary formats:

- CSV or JSON (digital data)
- PNG (images)
- STEP (CAD data)





What conventions and procedures will you use to structure, name and manage versions of your files to help you and others better understand how your data is organized?

Data management instructions and procedures will be made available in files in the root directory where they are saved, which serves as the main directory. Those files are as follows:

- A README.txt file to describe writing conventions and backup procedures
- Various dated versions of this data management plan

Secondary directories will be used to back up research data. They will be defined in the README.txt file. Each data file's name will have the following structure: brief description or an acronym that reflects its contents (e.g., SpecPC23a) followed by the associated research project number (e.g., 561114-22), the date the file was created in YYYYMMDD format (e.g., 20220518), the author's initials (e.g., DB) and an identifier that allows different versions to be saved. Full example: SpecPC23a_561114-22_20220518_DB_v01





Documentation and Metadata

What documentation will be required to ensure that the data can be read and interpreted correctly in the future?

The following information will be recorded in an INFO.txt file that will be placed in the root directory of the research project. It will contain the names of project collaborators, their email addresses and their duties. It will also contain a summary of the research project, the objective and the research methodology. In addition, it will contain the following two categories of information:

1. Information about raw data:

- A glossary defining the data
- Types and formats of data gathered
- Data collection methods
- Units of measurement used
- How the data is classified
- Data collection tools
- Information about the quality of the data

2. Information about analyzed data:

- Hypotheses
- Details of the computer code used to process the data
- Data analysis explanations

If applicable, a scientific paper about the research project will be placed in the data root directory once it is published.





How will you ensure that documentation will be created and entered consistently throughout your project?

Reviewing and implementing effective processes for capturing important metadata and documentation throughout your research project will support short- and long-term activities, including third party data analysis, data dissemination and data sharing and re-use as appropriate.

The practice of documenting research projects is already seamlessly integrated into our processes. Productique Québec's editorial board reviews this documentation at the end of each stage of our projects and at the very end. Documentation is stored in a SharePoint (Microsoft 365) collaborative and cloud-based institutional library, in a Wiki that is dedicated to the project and accessible to all members of the research team.

Computer codes used to collect and process research data will be developed using the [GIT open-source version control tool](#) to keep track of work done during programming and testing. Each validated element for which a version is created is accompanied by comments that can be used to track the project's progress.

If you use a metadata standard or tools to document and describe your data, please list them here.

- The ISA-95 standard is used to structure manufacturing data.
- Metadata will be described in accordance with [FRDR recommendations](#), whose metadata profile is based on the [Dublin Core](#) standard.





Storage, Access and Backup

What are your project's expected storage needs (in megabytes, gigabytes, terabytes, etc.) and how long will the data be saved?

Given the multiple file versions of the data that will be gathered and the expected data conservation time, we expect that we will need approximately one gigabyte of storage space.

How and where will your data be stored and backed up during your research project?

The data will be stored in three locations:

1. All data generated by the project will be stored in a SharePoint library managed by Productique Québec's Director of Research and Partnerships. Productique Québec's Head of Information Technology oversees the security of that SharePoint library, which is accessible via secure connection to Productique Québec's Office365 office suite.
2. Daily backups of the library's contents are automatically created on a secure physical server located at the CEGEP de Sherbrooke. They are overseen by the CEGEP's IT department.
3. A weekly copy of the server described in point 2 is kept off-site, not connected to the network, to ensure that the data is preserved in the event of a major incident.

How will the research team and other collaborators access, modify and contribute to the data throughout the project?

Members of the research team will have read access to all raw data in the SharePoint cloud library used to store research project data. Write access rights may be granted by the research coordinator, who owns the library. Each person will have a directory in which to save raw data. Analysis findings will be saved in a secondary directory that they will create using their directory of raw data. The principal investigator will be able to access all files in read and write mode.

Anyone helping to develop computer codes will have access to the data through an account on the software hosting and development management site [GitHub](#).





Preservation

Where will you store your data in the long term and access it at the end of your research project?

All data will be kept in a project SharePoint library dedicated to the project to ensure it complies with the archiving standards of the institution's data retention schedule and its responsible research conduct policy (for at least five years). At the end of this period, we will assess the necessity of extending the data's storage life.

The raw data will be stored in Canada's Federated Research Data Repository ([FRDR](#)), an open repository to efficiently integrate datasets. If needed, research data professionals work with researchers to ensure that data stored in the FRDR is organized and approved.

Describe how you will ensure that your data is ready for retention. Please consider: file formats that are appropriate for conservation, but that maintain data integrity; file anonymization and de-identification, including documentation files.

The files will be stored in non-proprietary formats. They will be accompanied by the INFO.txt files described in the Documentation and Metadata section. No personal data will be gathered during the research project.





Sharing and Reuse

What data will you share and in what format will you share it (e.g., raw, processed, analyzed or final data)?

Only raw data will be shared. This means it can be used by a community of entrepreneurs looking to train artificial intelligence algorithms to optimize their own production processes. It can also be used for academic training activities to provide students with examples of realistic scenarios for exercises and practical assignments.

Have you considered which end-user licence you will include with your data?

The data will be shared under a Creative Commons licence for "Attribution-NonCommercial" ([CC BY-NC](#)) after an agreement is signed with representatives of the company for which the research project was conducted.

What measures will you take to inform the research community of your data's existence?

The findings from the research project will be shared in:

- Reports submitted to the Ministry of Economy and Innovation and the Alliance Force.IA
- Presentations and seminars organized by [Synchronex](#), the Centres collégiaux de transfert de technologie [college technology transfer centre](CCTT) network, of which Productique Québec is a member (industrial partners are invited to these events)
- Presentations at scientific conferences
- Scientific papers resulting from the project
- Transfer activities for CEGEP and university education

The presentation of the findings will reference the data published in the Federated Research Data Repository (FRDR).

A digital unique permanent identifier (DOI) will be assigned to the data via the FRDR, further boosting its visibility.





Responsibilities and Resources

Please indicate who will be responsible for project data management before and after the project and the main data management duties for which this person will be responsible.

Productique Québec's Research and Partnerships Director will ensure that all members of the research team follow the instructions in the data management plan. A member of the team will be tasked with implementing the data management plan. Their duties will be as follows:

- Creating and managing the directory in which the project data is stored
- Creating a README.txt file to describe writing conventions and backup procedures
- Verifying data compliance at each step of the research process
- Training members of the research team to manage project data
- Preparing data and metadata to be shared in the digital repository

Providing concrete examples of team members' duties makes it easier to allocate data management tasks and avoid any oversights.

How will you delegate responsibility for data management activities if there are significant staffing changes in terms of staff members tasked with overseeing project data, particularly a change in principal investigator?

Productique Québec's Research and Partnerships Director is responsible for research data management. If they were unable to carry out their duties, another person would be appointed by Productique Québec's CEO to take on that responsibility. If a smooth transition was not possible, the person taking responsibility for data management could use the data management instructions and procedures available in files saved in the root directory, as described in the first part of this management plan. If needed, the CEGEP de Sherbrooke's Research Office could help facilitate the transition.





What resources will you require to implement your data management plan? What is the approximate overall cost of data management?

The resources that will be used to implement this data management plan are as follows:

- Duties of the person responsible for implementing the management plan: 10 hours of work
- Revision by the editorial board: 10 hours





Ethics and Legal Compliance

If your project involves sensitive data, how will you ensure that it is handled securely and that the data is only accessible to approved members of the project?

Not applicable.

If applicable, what strategies will you use to address the secondary use of data, particularly data of a sensitive nature?

Not applicable.

How will you handle legal, ethical and intellectual property issues?

Before submitting a grant application with industrial partners, Productique Québec makes sure to discuss confidentiality and intellectual property issues. If needed, a contract is signed before work begins. In our case, there is nothing preventing the research data from being freely shared.

