

## Plan de gestion des données pour Optimisation d'un procédé industriel d'extrusion

### RÉSUMÉ

Le présent plan de gestion des données a été créé d'après le modèle de base proposé dans la plateforme de l'Assistant PGD. Outre ses créateurs, le groupe de travail sur la gestion des données de recherche de la communauté de pratique des conseillères et conseillers à la recherche animée par l'Association pour la recherche au collégial a contribué à sa rédaction. Ce document sera mis en œuvre dans le contexte d'une recherche pilotée par un centre collégial de transfert de technologie (CCTT) québécois. La recherche est réalisée en partenariat avec une entreprise manufacturière. Les données brutes recueillies lors de la réalisation de la recherche seront librement diffusées sous licence Creative Common.

### DÉTAILS ADMINISTRATIFS

**Nom du projet :** Optimisation d'un procédé industriel de plasturgie

**Créateurs :** Marie-Josée Fraser, David Brodeur et Vincent Thomasset-Laperrière

**Établissement :** Cégep de Sherbrooke

**Organisme subventionnaire :** Ministère de l'Économie et de l'Innovation du Québec et Force IA

### Description :

Une entreprise a sollicité la collaboration de [Productique Québec](#) pour optimiser son procédé de plasturgie. La qualité de son produit varie en fonction de facteurs environnementaux multiples. Il est donc nécessaire d'ajuster les paramètres de production pour garantir le respect des tolérances exigées par le contexte de fabrication et la constance de la qualité du produit final. L'ajustement manuel rapide des paramètres de production n'est pas facile pour les opérateurs, étant donné que plusieurs facteurs physiques doivent être considérés. Cette activité, qui requiert un temps considérable, gaspille de grandes quantités de matériel au moment du démarrage de la production. Pour résoudre ce problème, Productique Québec propose d'instrumenter le procédé et d'analyser ensuite les données obtenues à l'aide de techniques utilisant l'intelligence artificielle, qui fournissent à l'opérateur des rétroactions en temps réel lui permettant de corriger adéquatement et rapidement les paramètres de production. Les données utiles comportent les spécifications du produit, les paramètres de production, les mesures d'inspection finale et les conditions environnementales.

### COLLECTE DE DONNÉES

**Quels types de données recueillerez-vous, créerez-vous, couplerez-vous, acquérez-vous ou consignerez-vous?**

Les données recueillies comporteront :

- des données de conception (fichiers configurables de conception assistée par ordinateur – CAO);
- des spécifications techniques des pièces produites selon les critères définis;
- des paramètres de production;
- des conditions environnementales;
- des mesures d'inspection.

**Version « ébauche »** (2 novembre 2022). Diffusion autorisée par les ayants droits et l'Alliance de recherche numérique du Canada (Alliance). La version définitive sera disponible sur le site de Alliance au printemps 2023. Dès que la version définitive est disponible, la version « ébauche » sera retirée d'EDUQ.info

Les types de données recueillies seront donc les suivants :

- données numériques;
- données tabulaires;
- données graphiques.

### **Dans quels formats de fichiers vos données seront-elles collectées? Ces formats permettront-ils la réutilisation, le partage et l'accès à long terme aux données?**

Afin de permettre la réutilisation, le partage et l'accès à long terme aux données, celles-ci seront collectées dans les formats non propriétaires suivants :

- CSV ou JSON (données numériques);
- PNG (images);
- STEP (données de CAO).

### **Quelles conventions et procédures utiliserez-vous pour structurer, nommer et gérer les versions de vos fichiers pour vous aider et aider les autres à mieux comprendre comment vos données sont organisées?**

Les consignes et les procédures concernant la gestion des données seront rendues disponibles dans des fichiers à la racine du répertoire où elles sont sauvegardées et qui en constitue le répertoire principal. Ces fichiers sont les suivants :

- un fichier README.txt pour décrire les conventions d'écriture et les procédures de sauvegarde;
- les différentes versions datées du présent plan de gestion des données.

Des répertoires secondaires seront dédiés à la sauvegarde des données de recherche. Ils seront définis dans le fichier README.txt. Le nom de chaque fichier de données suivra les règles suivantes : courte description ou un acronyme qui reflète son contenu (ex. : SpecPC23a), puis le numéro du projet de recherche associé (ex. : 561114-22), la date de création du fichier sous la forme AAAAMMJJ (ex. : 20220518), les initiales l'auteure ou l'auteur (ex. : DB) et un identifiant qui permet d'enregistrer différentes versions. Exemple complet : SpecPC23a\_561114-22\_20220518\_DB\_v01

## **DOCUMENTATION ET MÉTADONNÉES**

### **Quelle documentation sera nécessaire pour que les données soient lues et interprétées correctement à l'avenir?**

Les informations suivantes seront inscrites dans un fichier INFO.txt qui sera déposé à la racine du répertoire du projet de recherche. On y trouvera d'abord les noms des personnes ayant collaboré au projet, leur adresse de courriel et les tâches qu'elles ont réalisées. On y trouvera également un résumé du projet de recherche ainsi que l'objectif et la méthodologie de la recherche. Par la suite, on y trouvera les deux catégories d'informations suivantes :

1. Des informations sur les données brutes :

- un lexique qui définit les données;
- les types et les formats des données recueillies;
- les méthodes de collecte des données;
- les unités de mesure utilisées;
- la manière dont les données sont classées;
- les outils de collecte de données;

**Version « ébauche »** (2 novembre 2022). Diffusion autorisée par les ayants droits et l'Alliance de recherche numérique du Canada (Alliance). La version définitive sera disponible sur le site de Alliance au printemps 2023. Dès que la version définitive est disponible, la version « ébauche » sera retirée d'EDUQ.info

- les informations concernant la qualité des données.

## 2. Des informations sur les données analysées :

- les hypothèses formulées;
- les indications permettant d'accéder au code informatique utilisé pour traiter les données;
- des explications concernant l'analyse des données.

S'il y a lieu, un article scientifique portant sur le projet de recherche sera déposé à la racine du répertoire de données au moment de la publication.

### **Comment ferez-vous en sorte que la documentation soit créée ou saisie de manière cohérente tout au long de votre projet?**

La documentation des projets de recherche est déjà une pratique harmonieusement intégrée à nos processus. Elle est révisée par le comité de rédaction de Productique Québec à la fin de chacune des étapes de nos projets ainsi qu'à la toute fin. La documentation est conservée dans une bibliothèque institutionnelle collaborative et nuagique SharePoint (Microsoft 365), dans un wiki consacré au projet et accessible à l'ensemble des membres de l'équipe de recherche.

Le développement des codes informatiques utilisés pour recueillir et traiter les données de recherche est réalisé avec l'[outil libre de gestion de versions GIT](#) pour conserver des traces du travail effectué lors de la programmation et de ses tests. Chaque élément validé et pour lequel une version est créée est accompagné de commentaires permettant de suivre l'évolution du projet.

### **Si vous utilisez une norme de métadonnées ou des outils pour documenter et décrire vos données, veuillez les énumérer ici.**

- La norme ISA-95 est utilisée pour structurer les données de production manufacturière.
- Les métadonnées seront décrites en suivant les [recommandations du DFDR](#), dont le profil de découverte est fondé sur la norme [Dublin Core](#).

## **STOCKAGE ET SAUVEGARDE**

### **Quels sont les besoins prévus en matière de stockage pour votre projet (en mégaoctets, gigaoctets, téraoctets, etc.) et quelle sera la durée de sauvegarde?**

En tenant compte des multiples versions des fichiers de données qui seront récoltées ainsi que de la durée prévue de conservation, nous estimons que les besoins prévus en espace de stockage seront de l'ordre d'un gigaoctet.

### **De quelle manière et à quel endroit vos données seront-elles stockées et sauvegardées pendant votre projet de recherche?**

Les données seront entreposées à trois endroits :

1. Toutes les données générées par le projet seront enregistrées dans une bibliothèque SharePoint. Celle-ci est gérée par le directeur de la recherche et des partenariats de Productique Québec, et sa sécurité est assurée par la personne responsable des technologies de l'information de Productique Québec. Cette bibliothèque est accessible par une connexion sécurisée à la suite bureautique Office365 de Productique Québec.

**Version « ébauche »** (2 novembre 2022). Diffusion autorisée par les ayants droits et l'Alliance de recherche numérique du Canada (Alliance). La version définitive sera disponible sur le site de Alliance au printemps 2023. Dès que la version définitive est disponible, la version « ébauche » sera retirée d'EDUQ.info

2. Des sauvegardes du contenu de la bibliothèque seront automatiquement effectuées quotidiennement sur un serveur physique sécurisé situé au cégep de Sherbrooke. Elles sont assurées par la Direction des technologies de l'information (DTI) du cégep.
3. Une copie hebdomadaire du serveur décrit au point 2 est conservée hors site et sans connexion au réseau afin d'assurer la préservation des données en cas d'incident majeur.

### **De quelle manière l'équipe de recherche et les autres collaborateurs vont-ils accéder aux données, les modifier et y contribuer tout au long du projet?**

Les membres de l'équipe de recherche accéderont à toutes les données brutes en mode lecture dans la bibliothèque nuagique SharePoint dédiée à l'entreposage des données du projet de recherche; des droits d'accès en mode écriture pourront être octroyés par la coordonnatrice ou le coordonnateur de la recherche, qui est propriétaire de la bibliothèque. Chaque personne aura un répertoire dans lequel elle pourra sauvegarder ses données brutes. Les résultats de ses travaux d'analyse seront sauvegardés dans un répertoire secondaire dédié qu'elle créera à partir de son répertoire de données brutes. La chercheuse principale ou le chercheur principal aura un droit d'accès à tous les fichiers, en mode lecture et en mode écriture.

Toutes les personnes contribuant à l'élaboration des codes informatiques auront accès aux données par un compte sur le site d'hébergement et de gestion de développement de logiciels [GitHub](#).

## **PRÉSERVATION**

### **Où déposerez-vous vos données pour les conserver à long terme et y accéder à la fin de votre projet de recherche?**

Toutes les données seront conservées dans un la bibliothèque SharePoint associée au projet en assurant le respect des normes d'archivistique du calendrier de conservation des données de l'établissement et de la politique de conduite responsable en recherche de ce dernier (durée minimale de 5 ans). À la fin de cette période, une évaluation sera faite quant à la nécessité de prolonger la durée de conservation des données.

Les données brutes seront versées dans le Dépôt fédéré de données de recherche ([DFDR](#)) du Canada, qui est un dépôt ouvert permettant d'intégrer efficacement des jeux de données. En cas de besoin, des personnes professionnelles des données de recherche travaillent avec les chercheuses et les chercheurs pour assurer l'organisation et l'approbation des objets déposés dans le DFDR.

**Indiquez comment vous vous assurerez que vos données sont prêtes pour la conservation. À prendre en considération: formats de fichier appropriés pour la conservation, mais qui préservent l'intégrité des données; anonymisation et dépersonnalisation des fichiers, y compris les fichiers de documentation.**

Les fichiers seront entreposés dans des formats non propriétaires. Ils seront accompagnés des fichiers INFO.txt décrits dans la section Documentation et métadonnées. Aucune donnée personnelle ne sera récoltée dans le projet de recherche.

## **PARTAGE ET RÉUTILISATION**

**Quelles données partagerez-vous et sous quelle forme les partagerez-vous (par exemple, données brutes, traitées, analysées, finales)?**

**Version « ébauche »** (2 novembre 2022). Diffusion autorisée par les ayants droits et l'Alliance de recherche numérique du Canada (Alliance). La version définitive sera disponible sur le site de Alliance au printemps 2023. Dès que la version définitive est disponible, la version « ébauche » sera retirée d'EDUQ.info

Seules les données brutes seront partagées. Elles pourront ainsi être utilisées par une communauté d'entrepreneurs qui désireront entraîner des algorithmes d'intelligence artificielle pour optimiser leurs propres processus de production. Elles pourront également être utilisées dans des activités de formation académique afin de présenter des scénarios réalistes dans les exemples fournis aux étudiantes et étudiants, dans les exercices et les travaux pratiques.

### **Avez-vous songé au type de licence d'utilisateur final à inclure avec vos données?**

Les données seront diffusées sous licence Creative Common pour « attribution sans utilisation commerciale » ([CC BY-NC](#)) à la suite d'une entente signée avec les personnes représentant l'entreprise pour laquelle le projet de recherche a été réalisé.

### **Quelles mesures prendrez-vous pour faire savoir à la communauté des chercheurs que vos données existent?**

Les résultats du projet de recherche seront diffusés :

- dans les rapports remis au ministère de l'Économie et de l'Innovation ainsi qu'au regroupement Force IA;
- lors de présentations et de séminaires organisés par Synchronex, le réseau des centres collégiaux de transfert de technologie (CCTT), dont Productique Québec est membre (des partenaires industriels sont invités à ces événements);
- lors de présentations dans des congrès scientifiques;
- dans les articles scientifiques qui émergeront du projet;
- dans des activités de transfert vers la formation collégiale et universitaire.

La présentation des résultats fera référence aux données publiées dans le Dépôt fédéré de données de recherche (DFDR).

Par le biais du DFDR, un identifiant permanent unique (DOI) est attribué aux données, ce qui augmentera également leur visibilité.

## **RESPONSABILITÉS ET RESSOURCES**

### **Désignez qui sera responsable de la gestion des données de ce projet pendant et après le projet et les principales tâches de gestion des données dont cette personne sera responsable.**

Le directeur de la recherche et des partenariats de Productique Québec s'assurera que tous les membres de l'équipe de recherche suivent les instructions qui apparaissent au plan de gestion de données. Une personne membre de l'équipe sera désignée comme étant responsable de la mise en œuvre du plan de gestion des données. Ses tâches seront les suivantes :

- Création et gestion du répertoire dans lequel les données du projet seront entreposées;
- Création du fichier README.txt décrivant les conventions d'écriture et les procédures de sauvegarde;
- Vérification de la conformité des données à toutes les étapes de la recherche;
- Formation des membres de l'équipe de recherche sur la gestion des données du projet;
- Préparation des données et des métadonnées pour la diffusion dans le dépôt numérique.

**Donner des exemples concrets des tâches qui seront effectuées par les différents membres de l'équipe permet de mieux répartir le travail de gestion des données et évite les oublis.**

**De quelle façon allez-vous administrer les responsabilités quant aux activités de gestion des données s'il y a des changements importants de personnel chargé de superviser les données du projet, notamment un changement de chercheur principal?**

La gestion des données de recherche est sous la responsabilité du directeur de la recherche et des partenariats de Productique Québec. S'il advenait qu'il ne soit plus en mesure de s'acquitter de ses fonctions, une autre personne serait désignée par le directeur général de Productique Québec pour assumer cette responsabilité. Dans la mesure où il ne serait pas possible d'assurer une transition, la personne qui prendrait la responsabilité de la gestion des données pourrait s'appuyer sur les consignes et les procédures concernant la gestion des données disponibles dans des fichiers à la racine du répertoire où elles sont sauvegardées, comme décrit dans la première partie du présent plan de gestion. En cas de besoin, le Bureau de la recherche du cégep de Sherbrooke pourrait soutenir la transition.

**De quelles ressources aurez-vous besoin pour mettre en œuvre votre plan de gestion des données? À combien estimez-vous le coût global de la gestion des données?**

Les ressources qui seront mobilisées pour mettre en œuvre ce plan de gestion de données sont les suivantes :

- Tâches de la personne responsable de la mise en œuvre du plan de gestion : 10 heures de travail
- Révision par le comité de rédaction : 10 heures

## **CONFORMITÉ ÉTHIQUE ET JURIDIQUE**

**Si votre projet comprend des données sensibles, comment vous assurez-vous qu'il est géré de manière sécuritaire et que les données sont accessibles uniquement aux membres approuvés du projet?**

Ne s'applique pas.

**Quelles stratégies allez-vous mettre en œuvre pour gérer les utilisations secondaires des données sensibles, le cas échéant?**

Ne s'applique pas.

**Comment allez-vous gérer les questions juridiques, éthiques et de propriété intellectuelle?**

Avant de présenter une demande de subvention avec des partenaires industriels, Productique Québec s'assure de discuter des enjeux liés à la confidentialité et à la propriété intellectuelle. Le cas échéant, un contrat est signé avant le début des travaux. Dans le cas qui nous concerne, aucun enjeu n'empêche la diffusion libre des données de recherche.

**Dernière modification :** 2022-06-21