

Atelier de réflexion « Les données : de qui, pour qui et pour quoi? »

Le vendredi 13 janvier 2023

Résumé des discussions

L'Atelier de réflexion a débuté par un bloc de trois conférences qui visaient à faire ressortir les enjeux des données numériques pour un meilleur partage ou une mutualisation de celles-ci.

Données numériques : mieux comprendre nos enjeux pour un meilleur partage

Selon une étude menée par CIRANO (Royer, A., N. De Marcellis-Warin, I. Peignier, T. Warin, M. Panot, C. Mondin (2020). [Les enjeux du numérique dans le secteur agricole. Défis et opportunités.](#) Rapport CIRANO 2020RP-12, juillet), les grands enjeux identifiés pour le partage des données sont : la propriété, la confidentialité, le consentement et la sécurité. Autres éléments cruciaux : la transparence et la confiance. À cela s'ajoutent des enjeux financiers, car la mise en place d'une plateforme de partage ou de gestion de données est souvent complexe, longue et coûteuse. Il faut donc une compréhension claire des besoins de partage pour que ce dernier soit efficient et éthique, en vue d'atteindre le plein potentiel des données.

Constat : C'est le début d'une révolution qui demande de nouvelles connaissances et amène son lot d'incertitudes. Il y a également des défis collectifs organisationnels qui requièrent donc l'implication de plusieurs personnes.

Les enjeux éthiques et juridiques du partage des données numériques

Les enjeux liés au partage sont de nature éthique et juridique ce qui requiert de mettre en place des pratiques, telles qu'introduites au tableau de la page suivante.

De manière générale, les bonnes pratiques pour un partage de données réussi sont :

- Planification soignée (élaboration d'un plan de gestion des données);
- Pratiques adaptées (sécurité - sauvegarde - transparence - gouvernance);
- Partenaires compétents et sérieux (gestion des accès); et
- Contractualisation du partenariat.

Enjeux éthiques	Pratiques
Propriété et confidentialité	Pour les données personnelles et sensibles, le consentement des fournisseurs des données et les engagements de confidentialité sont essentiels. Le consentement doit être obtenu au moment de la collecte des données et celle-ci doit s'en tenir aux données nécessaires i.e. qui auront une utilité.
Traitement efficient	Examen soucieux des données collectées pour y déceler les problèmes sous-jacents : enregistrements en double, versions incohérentes, dûment annotées, exactitude, etc. Cadre éthique FAIR : <ul style="list-style-type: none"> • (F) Facile à trouver • (A) Accessible • (I) Interopérable • (R) Réutilisable
Sécurité (technique)	Personne responsable dédiée à ces opérations, choix d'un environnement sécurisé approprié, choix des bons partenaires tiers spécialisés, assurer la sécurité des données dans toute la chaîne et assurer les droits des fournisseurs de données.
Gouvernance	Élaboration d'une politique de gestion des données, mise en place d'un comité de gouvernance et identification d'un responsable du déploiement et applications des politiques

Du partage à la mutualisation des données - tendance forte

Mutualisation des données : création de "réservoirs de données numériques sectorielles"

Objectifs :

- Accéder ou générer des données pertinentes ;
- Augmenter le volume et la qualité des données ; et
- Innover par la création de nouvelles connaissances.

La mutualisation rencontre cependant divers obstacles techniques, éthiques et juridiques. Plusieurs lois entrent en ligne de compte comme la LOI 25 qui donnera une protection accrue des renseignements personnels, incluant les données sensibles. D'où l'importance de bien connaître ce que l'on partage pour savoir quelles lois s'appliquent. Pour un partage des données personnelles et sensibles, le consentement des fournisseurs et l'engagement de confidentialité demeurent essentiels. Le droit est dans une logique de protection. On protège la partie qui a généré la donnée, qui l'a collectée. La notion de partage de données est dans une logique individualiste. On peut concevoir une logique plus communautaire des données, comme la **fiducie de données**. C'est un outil de gestion d'un actif selon une finalité très précise. C'est une mise en commun (une mutualisation) dans un cadre transparent, ce qui amène la confiance et le respect des droits et intérêts des diverses parties prenantes. Le fiduciaire administre, il est imputable et

responsable. La mutualisation est portée par une structure juridique indépendante. Elle peut être portée sur une plateforme ou simplement être mise en place pour les modalités de partage. C'est un modèle durable dont l'opérationnalisation requiert des dispositions légales, un inventaire des données et un consensus sur le mode opératoire de la mutualisation des données.

Pour illustrer la diversité lorsque l'on parle des données, **cinq intervenant.e.s** ont présenté un portrait du type de données dont ils disposent, de l'utilisation qu'ils en font, de leurs besoins et de leurs contraintes. Le tableau suivant reprend quelques exemples issus de leurs présentations.

	Données	Utilisations	Besoins	Contraintes
France Brunelle, MAPAQ	Sources et données très diversifiées. Exemples : <ul style="list-style-type: none"> • Exploitations agricoles • Contaminants dans les aliments • Projets de recherche financés • Répertoire des sites agricoles • Fiches d'enregistrements agricoles • Agrométéo Québec • Dépistage du Réseau d'avertissements phytosanitaires (RAP) • Info-Sols • Bilan des ventes de produits de phytoprotection • Etc. 	<ul style="list-style-type: none"> • Publications (ex : indicateurs économiques; profil régional/sectoriel de l'industrie) • Aide à la décision (ex : sécurité des aliments, suivis d'indicateurs) et orientation politique • Soutien aux interventions d'urgence • Données en ligne 	<ul style="list-style-type: none"> • Précision et qualité des renseignements sont essentielles • Suivis d'indicateurs • Connaître les types de données disponibles • Suivis des infrastructures technologiques nécessaires 	<ul style="list-style-type: none"> • Administratives et légales (lois et règlements, ententes et politiques) • Données très diversifiées • Propriété intellectuelle diffuse • Manque de structure • Couverture très large
Catherine Machado, VIA Pôle d'expertise en services-agricoles	<ul style="list-style-type: none"> • Identification des entreprises (NIM, FADQ, ...) • Économiques (revenus, dépenses, budget, quotas, ...) • Agroenvironnementales (rendements des cultures, superficies, nombre d'animaux, ...) • Provenant des travaux réalisés et projections 	<ul style="list-style-type: none"> • Pour producteur : moyenne de groupe, PAEF, bilan phosphore, situation économique des fermes • Pour autres acteurs : vente de données • Futur : portrait par région, cultures engrais, ... • Croisement des données du logiciel de fertilisation avec celui de gestion 	<ul style="list-style-type: none"> • Meilleure circulation de l'information (avec accord producteur) • Bonifier l'utilisation • Éviter de répéter • Que les données soient dynamiques, adaptées au changement que vit l'entreprise pendant l'année • Outil de travail 	<ul style="list-style-type: none"> • Attention aux informations sensibles du producteur • Intérêts divergents entre les parties • Aspect technologique • Nouvelle réglementation peut complexifier

	Données	Utilisations	Besoins	Contraintes
Vincent Godin, <i>Producteur et Président de l'APCQ</i>	<ul style="list-style-type: none"> • Culturelles (température au champ, tension du sol, rendements...) • Gestion (heures pompées, heures travaillées par type de tâche) • Développement ou travaux d'excavation (coordonnées GPS, LIDAR). 	<ul style="list-style-type: none"> • Culturelles (ajustement des durées et fréquences d'arrosage, démarrage des pompes d'irrigation, rétroaction sur applications d'engrais et des pesticides, et suivi de la productivité des champs). Ces données permettent aussi de suivre l'efficacité des applications. • Gestion : planifier et budgétiser les tâches 	<ul style="list-style-type: none"> • Besoin en main d'œuvre spécialisée pour l'analyse et le traitement des données 	<ul style="list-style-type: none"> • Pas de contrainte pour l'accès aux données, mais les données de production doivent être anonymisées • Seule contrainte : Ressources humaines pour le conditionnement et l'analyse des données recueillies
Silvio Gumiere, <i>Chercheur et professeur, Université Laval</i>	<ul style="list-style-type: none"> • Séries temporelles (tension, humidité, pluie, température, débit, concentration, isotopique ...) • Imagerie (drones, satellite, GPR, CT-Scan) • Spatialisées (propriétés hydrodynamiques, rendement, microbiologiques) • Simulations climatiques futures 	<ul style="list-style-type: none"> • Développement des modèles • Calage et validation des modèles • Identification des processus et avancement des connaissances • Recommandations pratiques 	<ul style="list-style-type: none"> • Données fiables et faciles d'accès • Possibilité d'utilisation en temps réel • Uniformisation selon le type • Métadonnées disponibles • Contrôle de qualité 	<ul style="list-style-type: none"> • Données des projets en développement coopératif • Données propriétaires (fabricant des sondes et plateformes) • Données de débit, météo, qualité, climat futur, provenant de différentes plateformes
Victor Morin, <i>Producteur et Président XLKey</i>	<ul style="list-style-type: none"> • Données brutes spatiales et temporelles • Sols et terrain (conductivité électrique, MO, pH, topographie, compaction, humidité, caractéristique physique, biologique et chimique) • Culture (hauteur, couleur RGB, autres bandes spectrales, conditions climatiques) 	<ul style="list-style-type: none"> • Stratégique : Générer des zones, localiser les zones problématiques, identifier et chiffrer chaque problème, prendre action • Au champ : Améliorer le terrain, gestion localisée des intrants, efficacité opérationnelle (GPS), indices de performance (ex : carte de rentabilité, émission de GES) 	<ul style="list-style-type: none"> • Accessibilité aux données d'opération • Diagnostic des pertes immédiates pour apporter des solutions à la prochaine année (parler d'économie (argent)) et la préparation sur le terrain) 	

	Données	Utilisations	Besoins	Contraintes
Victor Morin <i>(suite)</i>	<ul style="list-style-type: none"> • Métadonnées des opérations au champ (type d'intrant, dose, vitesse, profondeur) 			

Plénière

Le coût pour générer les données et le type de données

La plupart des participant.e.s semblent d'accord que les données dont on dispose ne sont pas toujours d'excellente qualité et ne sont pas toujours pertinentes. Par exemple, pour les sols, on se contente trop souvent de mesures simples, faciles à obtenir et peu coûteuses (comme le carbone organique et le pH de surface). L'ensemble des spécialistes s'entendent sur le fait qu'il doit y avoir une évaluation sur l'ensemble du profil du sol si on veut être en mesure d'utiliser les données de manière efficace et précise.

De plus, de nombreux producteurs.rices ne seraient pas conscient.e.s de la valeur des données, et surtout, que cela devient un actif. Ils.elles ont beaucoup de données qui « dorment », car ils.elles ne savent pas quoi en faire. Il serait important de trouver d'abord le moyen de les sensibiliser à la valeur des données...avant d'essayer de les convaincre d'une quelconque action.

Question sur la formation des agronomes

Les algorithmes de calculs étant de plus en plus puissants, le travail des agronomes sera amené à changer. Les agronomes seront appelés à devenir des accompagnateurs pour l'introduction et l'utilisation de la technologie et cet aspect entraînera également des répercussions sur les recommandations. Il y a un besoin de formation pour les agronomes et celle-ci doit évoluer avec le contexte technologique pour les aider dans l'élaboration du service-conseil. À ce sujet, le programme de formation en agronomie de l'Université Laval est en révision et il y aura une modification au programme pour répondre à cette demande et ainsi mieux préparer les futurs agronomes. L'agronome sera en mesure d'optimiser l'utilisation des outils en s'assurant de la qualité des données (entrées et sorties) et par ses recommandations, selon le code d'éthique, de déontologie et de pratique professionnelle.

Anne-Sophie Hulin précise qu'il n'y a pas encore de réglementation de l'intelligence artificielle (IA) au Canada. Même s'il n'y a rien sur le droit, il y a quand même des balises et l'IA ne remplace pas l'humain en ce qui concerne la responsabilité. Il y a toujours une notion de « comportement raisonnable » qui détermine le niveau de responsabilité.

La notion de responsabilité du concepteur est également discutée. Il est primordial qu'il y ait une transparence au niveau de l'algorithme (sur les « données » afin d'être capable d'évaluer la qualité de la donnée). Mais, il y a encore là une problématique au niveau de la propriété des données.

Il y a un grand défi pour bien clarifier les différences entre les données brutes, les données valorisées et le produit final.

Question sur la structure de traitement des données

Il y a tout un travail à faire sur le format des données. L'ensemble des participant.e.s s'entendent sur le fait que les données sont très hétérogènes, qu'il est difficile de les faire communiquer entre elles; quelques fois elles sont même cryptées. Il existe des systèmes pour uniformiser les données, c'est donc une question de volonté de coordonner tout cela.

Besoin d'harmonisation au sein des banques de données

Une fois le contrôle de qualité effectuée par un agronome pour le choix du type de données, l'acceptation de la donnée et le rejet de données aberrantes, il reste un travail d'harmonisation pour pouvoir les intégrer au sein des banques de données. De nombreux témoignages des participant.e.s vont dans le sens d'une structure ouverte et accessible pour permettre à l'agronome, responsable des recommandations, d'acquérir une confiance dans l'outil au lieu d'être un frein à son adoption. Le modèle de **Fiducie** apparaît très prometteur pour mettre en place des structures légales plus faciles d'opération, plus souples, afin de regrouper les équipes de programmeurs et de spécialistes du domaine. Pour certains participant.e.s, c'est peut-être la clé. Et s'il y a un certain volume de producteurs.rices qui adhèrent, cela peut être beaucoup moins coûteux. Ils doivent être au premier plan de cet objectif de création de banques de données et d'outils pour en soutirer toute la connaissance qui s'y trouve cachée.

Un participant a proposé l'implantation d'un logiciel de gouvernance des données. Ses fonctions seraient de décrire les métadonnées, la gouvernance, les responsabilités de chacun, etc. Cela permettrait un lignage des données. Le consommateur de données serait soumis à une entente contractuelle. Il y a un besoin de se concerter, de se faire confiance.

La rétribution du service-conseil

C'est l'enfant pauvre du débat sauf pour plusieurs compagnies qui investissent depuis des années dans ce secteur. Il faut pouvoir démontrer la valeur du service-conseil. Certain.e.s participant.e.s ont suggéré de faire des essais pairés (gestion d'un champ avec et sans service-conseil) et des études de cas pour démontrer la rentabilité de mettre en place des interventions à cet effet. C'est ce qui a été fait avec AGIRRSOL (Logiciel d'Aide à la Gestion de l'IRRigation en SOLs organiques) avec les chercheurs Silvio Gumiere de l'Université Laval et Alain Rousseau de l'INRS. Il est suggéré de continuer dans cette voie pour la conservation et la santé des sols. Cependant, le service-conseil ne peut démarrer et s'installer que si des études indépendantes permettent d'en évaluer réellement la pertinence et la rentabilité. Beaucoup plus d'efforts de recherche doivent porter sur ces aspects et cela interpelle l'Axe 4 du RQRAD. Ces études doivent essayer de démontrer la rentabilité pour une entreprise d'une pleine exploitation des informations rendues disponibles par la technologie et par le choix d'indicateurs appropriés.

Selon certains participant.e.s, il y aurait aussi un besoin au niveau de la mise en place de normes canadiennes. Cela permettrait le développement d'outils de visualisation, ce qui aiderait à convaincre les utilisateurs.rices.

Conclusions

Le RQRAD s'est donné la mission d'amorcer la réflexion sur les enjeux reliés aux données massives. En quoi le RQRAD pourrait venir stimuler ce chantier à la prochaine étape? Rappelons également le mandat du RQRAD : mobiliser les forces...pour se doter d'une agriculture durable.

Plusieurs participant.e.s rappellent qu'il est temps de passer à l'action, mais qu'il est important de travailler avec des organismes qui ont déjà fait des réflexions ou actions comme l'OBVIA, Lactanet, IID, IVADO, CIRANO, etc.

Il ressort de cet atelier de réflexion qu'il serait pertinent de mettre en place des comités/groupes de réflexion sur :

- L'uniformisation des données (inspirée de l'OGC – Open Geospatial Consortium (<https://www.ogc.org/projects/groups/geosciencedwg>) ou le WaterML – modèle de standard technique et d'information modèle utiliser pour représenter des chroniques temporelles de données hydrologiques). Il faudrait impliquer les associations de producteurs.rices qui ont des données pour définir justement un modèle d'uniformisation des données en agriculture;
- La gouvernance des données (incluant la confiance des données). Échanger avec des groupes ayant mis en place une base de données à la disponibilité de ses membres. On peut penser à contacter Isabelle Cayer du CCTT en intelligence numérique concernant l'OGSL (<https://catalogue.ogsl.ca/>), l'observatoire global du Saint-Laurent;
- La cogitation de la définition d'un projet pilote (type preuve de concept) d'une mutualisation de données massives pour illustrer la valeur ajoutée que l'on peut en soutirer avec l'IA pour répondre à des questions de recherche, projet qui pourrait être fait avec une ou deux associations de producteurs.rices ou encore avec la Financière agricole du Québec.

On s'attend que le RQRAD joue son rôle de mobilisateur et concerté les différents acteurs pour initier la mise en place de ces comités/groupes.