

EXTRAIT · DATASTRATÈGE

Constellations

Classer, typologiser, modéliser

Guide du chercheur à l'ère des données massives et de l'intelligence artificielle

Dener François

Datastratège | Studio

Une publication Datastratège · Québec

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Page de titre

CONSTELLATIONS

Classer, typologiser, modéliser

Guide du chercheur à l'ère des données massives et de l'intelligence artificielle

Collection Firmament · N° 1

Dr Dener François

Datastratege | Studio · Québec · 2026





EXTRAIT · DATASTRATÈGE

Page de droits

Mention	Valeur
Titre	Constellations : classer, typologiser, modéliser. Guide du chercheur à l'ère des données massives et de l'intelligence artificielle
Auteur	Dener François, Ph. D.
Éditeur	Datastratège, Québec (Québec), Canada
Collection	Firmament, N° 1
Édition	Première édition, 2026
Version antérieure	1.0, *Classification supervisée et non supervisée en recherche scientifique — Guide méthodologique intégré* (2026-10-07)
ISBN	En cours d'attribution par Bibliothèque et Archives nationales du Québec (imprimé, PDF et EPUB)
Dépôt légal	Bibliothèque et Archives nationales du Québec, 2026; Bibliothèque et Archives Canada, 2026
Licence	© 2026 Dener François. Tous droits réservés. Aucune partie de cet ouvrage ne peut être reproduite, stockée ou transmise sans l'autorisation écrite de l'éditeur, sauf pour de courtes citations accompagnées de la référence complète. L'extrait gratuit est diffusé sous licence Creative Commons BY-NC-ND 4.0; les jeux de données Villeneuve sous licence CC BY 4.0 et les scripts sous licence MIT, sur ProGera.
Données et scripts	Zenodo, https://doi.org/10.5281/zenodo.23268790 (données CC BY 4.0, code MIT)
Plateforme compagnon	ProGera — https://progera.datastratege.ca/
Mention de non-responsabilité	Les éléments juridiques (Loi 25, EPTC 2) sont présentés à titre informatif et ne constituent pas un avis juridique

EXTRAIT · DATASTRATÈGE

Citation suggérée (APA 7) : François, D. (2026). *Constellations : classer, typologiser, modéliser. Guide du chercheur à l'ère des données massives et de l'intelligence artificielle* (collection Firmament, n° 1). Datastratège.

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Épigraphe

> *Les étoiles n'ont jamais formé de constellations. Ce sont les humains qui les ont dessinées — et chaque peuple a dessiné les siennes.*

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Dédicace

À Henriette François Bernadin, mon épouse adorée,

et à nos fils Nick Hender Bernadin François, Dener Oliver Bernadin François et Yanis Bernadin François,

qui m'ont inspiré, accompagné et soutenu tout au long de ce chemin.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Remerciements

Ma reconnaissance va d'abord à ma direction de thèse à l'Université Laval : le professeur Alexandre Lebel, Ph. D., directeur, et la professeure Geneviève Cloutier, Ph. D., codirectrice. Leur encadrement, leur disponibilité et la richesse de leurs apports ont profondément marqué mon parcours de chercheur.

Je remercie le Dr Yvens Cheremond, collaborateur de Datastratège dans sa dimension de recherche académique et scientifique, qui a codirigé ma thèse dans le cadre de la cotutelle soutenue par l'Agence universitaire de la Francophonie (AUF), dont l'allocation m'a accompagné durant mes trois premières années de doctorat à l'Université Laval. Ma gratitude va aussi à mon binôme, Habila Adamou Djibo, excellent collègue et ami au programme de doctorat, pour son soutien indéfectible et ses encouragements à persévérer dans mes initiatives innovantes.

Je remercie toute l'équipe de recherche du site de l'Université Laval du programme pancanadien sur la qualité de l'environnement bâti, et en particulier le professeur Michel De Blois, coordonnateur du site, pour ses conseils, son soutien et son encadrement.

Ce livre doit aussi beaucoup aux collaborateurs de Datastratège : les ingénieurs-agronomes Guerry Corvil et Emmanuel Pierre, diplômés comme moi de la Faculté d'agronomie et de médecine vétérinaire (FAMV) de l'Université d'État d'Haïti, et l'économiste-statisticien Fanor Joseph, pour leur relecture attentive, leur soutien et leurs encouragements.

Il doit encore à toutes les personnes, collègues, collaboratrices, collaborateurs et contributeurs, qui ont formé le socle de sa révision : vocabulaire, grammaire, orthographe et style; design et visuel; essai des exercices et des bacs à sable; évaluation et certification.

Il est enfin dédié à ses futures lectrices et à ses futurs lecteurs, qui, au fil de leur quête, s'y arrêteront un instant, peut-être longtemps, et en feront, je l'espère, un compagnon de tous les jours.

EXTRAIT DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Avant-propos — D'où vient ce livre

Ce livre n'est pas né d'un plan de cours. Il est né d'un parcours.

Pendant près de vingt-cinq ans, entre Haïti et le Québec, j'ai conduit des projets agricoles, d'ingénierie, de coopération internationale, de gestion publique et de science des données. En parallèle, j'ai mené plusieurs recherches universitaires : trois maîtrises — en méthodes informatiques appliquées à la gestion des entreprises (Université de Picardie Jules Verne), en bases de données et intégration de systèmes (Université Nice Sophia Antipolis) et en coopération internationale et aide humanitaire —, un doctorat en ingénierie et management de projets soutenu en 2025, puis un second parcours doctoral, en aménagement du territoire et développement régional à l'Université Laval, actuellement en phase de rédaction et de validation des résultats.

D'un terrain à l'autre, la même difficulté revenait sous des formes différentes. Des exploitations agricoles à regrouper pour cibler un appui technique. Des projets à classer selon leur niveau de risque. Des territoires à distinguer selon la qualité de leur environnement bâti et leurs effets sur la santé. Chaque fois, il fallait passer de données dispersées à des catégories qui aient du sens — et qu'on puisse défendre devant un comité, un jury, un bailleur ou un décideur public.

J'ai connu les réussites : des typologies qui ont éclairé des décisions, des articles et des rapports qui ont circulé. J'ai aussi connu les erreurs, et ce sont elles qui m'ont le plus appris : une partition élégante qui s'effondrait au premier rééchantillonnage, des noms de classes choisis trop vite et contestés par les acteurs, une validation oubliée qu'un membre de jury a remarquée avant moi, un modèle performant en laboratoire mais inutilisable sur le terrain. Ces épisodes nourrissent les encadrés « Conseil du praticien » de chaque chapitre et la banque de cas réunie en fin d'ouvrage.

Fort de ces expériences riches et diversifiées, j'ai finalement écrit le livre que j'aurais voulu lire au début de chacun de ces parcours. Un livre qui ne se contente pas d'expliquer les algorithmes, mais qui aide à exercer un jugement : pourquoi classer, comment choisir, comment valider, comment nommer, comment décider — et comment adapter tout cela au contexte réel d'une thèse, d'une équipe, d'une institution ou d'une entreprise.

Il s'adresse à l'apprenti-chercheur qui ouvre ses premières données, au doctorant qui doit justifier chacun de ses choix, à la directrice de laboratoire qui veut harmoniser les pratiques de son équipe, et au praticien qui veut faire d'un résultat de recherche un outil. Il se prolonge

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

sur la plateforme ProGera, parce que l'apprentissage de la recherche ne s'arrête pas à la dernière page : il se poursuit dans l'échange, la confrontation et la pratique.

Québec, 7 octobre 2026

Dener François

Prologue — Un ciel saturé de points

Notre époque

Jamais une génération de chercheurs n'a disposé d'autant de données, ni d'outils aussi puissants pour les ordonner. Capteurs urbains, images satellitaires, registres administratifs, dossiers de santé, traces numériques, archives numérisées et corpus de textes produisent chaque jour plus d'observations qu'un laboratoire n'en traitait en une décennie. Aujourd'hui, les bibliothèques de logiciels libres mettent les algorithmes de classification à la portée d'un ordinateur portable, et l'intelligence artificielle générative sait désormais regrouper, étiqueter, résumer et prédire en quelques secondes.

Nous vivons aussi un moment de l'histoire humaine où les sociétés délèguent une part croissante de leurs décisions à des catégories produites par des machines : profils de risque, segments de clientèle, territoires prioritaires, groupes de patients, types d'apprenants. Ces catégories orientent des ressources publiques, des soins, des investissements, des politiques. Elles ne sont jamais neutres.

Le paradoxe

Cette abondance crée un paradoxe. Plus il devient facile de produire des classes, plus il devient difficile de savoir lesquelles ont du sens. Un algorithme de partitionnement trouvera toujours des groupes, même dans des données qui n'en contiennent pas. Un classifieur prédira toujours une étiquette, même lorsque l'étiquette d'origine reflète un biais. Une IA générative proposera toujours des noms séduisants pour des catégories qu'elle n'a pas validées.

La question n'est donc plus *peut-on classer?* Elle est devenue : *qu'est-ce qui autorise à croire une classification, à la nommer, à en faire une typologie, un modèle ou un cadre — et à fonder des décisions sur elle?*

La problématique de l'ouvrage

Trois constats fondent ce livre.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

1. **Un déficit de jugement plutôt que d'outils.** Les manuels de statistique et d'apprentissage automatique expliquent bien les algorithmes; ils disent moins comment choisir entre eux selon la question, l'épistémologie, les données et le contexte, ni comment passer d'une partition à une typologie publiable.
2. **Une fragmentation des traditions.** La classification est enseignée en silos : analyse des données à la française, statistique inférentielle des classes latentes, apprentissage automatique, méthodes configurationnelles, aide multicritère à la décision, recherche qualitative. Le chercheur en formation rencontre rarement ces traditions côte à côte.
3. **Une sous-estimation du contexte.** Une même méthode n'a pas la même valeur dans une thèse individuelle, une équipe, une institution universitaire, un institut indépendant ou une entreprise. Gouvernance, éthique, protection des renseignements personnels et reproductibilité font partie de la validité, pas de l'administration.

La question directrice

Comment conduire, de la question de recherche jusqu'à la publication, une démarche de classification qui soit à la fois rigoureuse, justifiable, éthique et adaptée à son contexte — et qui permette de construire des typologies, des modèles et des cadres méthodologiques utiles à la science et à la décision?

Ce que ce livre apporte

- **Une démarche intégrée** en dix phases, de la question à la restitution, avec la validation comme boucle centrale.
- **Une cartographie comparée** des méthodes supervisées, non supervisées et de leurs alternatives, qui rend visibles les hypothèses de chacune.
- **Un pont explicite** entre classes, typologies, modèles et cadres, appuyé sur la littérature méthodologique.
- **Une grille de choix contextuelle**, de la recherche individuelle au milieu industriel.
- **Une pédagogie de l'immersion progressive**, avec cas, débats, pièges, conseils de praticien et auto-évaluations prolongées sur ProGera.

Ce que ce livre n'est pas

Ce n'est ni un manuel de programmation, ni un traité mathématique. Les formules sont données lorsqu'elles éclairent un choix; le code est fourni sur ProGera. Le livre suppose une culture statistique de base (moyenne, variance, corrélation, régression) et renvoie aux ouvrages de référence pour l'approfondissement mathématique.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Comment lire ce livre

Le livre se lit de deux façons : en continu, comme un parcours de formation en six étapes, ou de façon sélective, chapitre par chapitre, grâce à des rubriques et des encadrés toujours placés au même endroit.

Une progression en six parties

Chaque partie correspond à un moment de la métaphore des constellations et vise un niveau de maîtrise plus élevé : **comprendre** (Partie I — Regarder le ciel), **appliquer** (II — Préparer l'observation), **analyser** (III — Relier les étoiles), **évaluer** (IV — Nommer les constellations) et **créer** (V — Naviguer et décider), puis transmettre (VI — Le métier de chercheur). Trois notions reviennent en spirale à chaque étape, de plus en plus approfondies : la posture épistémologique, la validation et le contexte. Une spirale en marge signale chaque retour.

Chaque chapitre, quatre mouvements

1. **S'orienter** : page d'ouverture, objectifs, prérequis, situation d'ouverture.
2. **Comprendre** : exposé, posture épistémologique, hypothèses de travail.
3. **Confronter** : étude de cas, débat et changement de paradigme, pièges de l'apprenti-chercheur.
4. **S'approprier** : conseil du praticien, à retenir, auto-évaluation, pour aller plus loin.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Les encadrés

Tableau L.1 — Les encadrés du livre et leur fonction

Encadré	Signe en marge	À quoi il sert
À retenir	Étoile polaire, or	L'essentiel à mémoriser
Piège de l'apprenti-chercheur	Étoile filante barrée, rouge brique	Erreur fréquente, sa conséquence, la bonne pratique
À ne pas dire	Bulle barrée, rouge brique	Formulation fautive et sa correction
Conseil du praticien	Boussole, bleu nuit	Leçon tirée de l'expérience de l'auteur
Posture épistémologique	Télescope, violet	Quelle posture adopter et pourquoi
Débat et paradigme	Deux étoiles en tension, turquoise	Positions contradictoires et état du débat
Étude de cas	Carte du ciel, vert sauge	Application complète sur un terrain
Hypothèse	Point d'interrogation stellaire, gris ardoise	Hypothèse ou proposition à tester
Auto-évaluation	Constellation à compléter, bleu ciel	QCM, exercices, code ProGera
Pour aller plus loin	Orbite fléchée, bleu ciel clair	Lectures commentées et ressources

Source : l'auteur.

Une à trois étoiles en tête de section indiquent le niveau de difficulté.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Cinq parcours selon votre profil

Tableau L.2 — Cinq parcours de lecture selon le profil

Vous êtes	Votre objectif	Lisez en priorité
Étudiant·e à la maîtrise	Réaliser une première typologie défendable	Chapitres 1 à 6, 11 et 15
Doctorant·e	Justifier chaque choix et publier	Le livre entier, dans l'ordre
Direction de recherche ou de laboratoire	Encadrer, harmoniser, évaluer	Chapitres 3, 6 et 12 à 15
Équipe d'institut indépendant	Produire des typologies crédibles publiquement	Chapitres 2, 10, 11, 13 et 14
Relier en milieu industriel ou organisationnel	Déployer et gouverner des classifieurs	Chapitres 3, 4, 7 à 9, 13 et 14

Source : l'auteur.

Le livre continue sur ProGera

Chaque chapitre porte un code (CST-01 à CST-19) et un code QR qui mènent à son espace sur ProGera : corrigés commentés, gabarits, jeux de données et scripts, discussion entre lecteurs, binômes étudiant·e–chercheur·e et équipes de projet. Un test de positionnement vous oriente vers le parcours adapté.

Conventions

Les références suivent la norme APA 7. Les termes techniques sont donnés en français, avec l'équivalent anglais entre parenthèses à la première occurrence.

Les repères visuels du livre

Chaque rubrique récurrente porte un pictogramme et une couleur, attribués une fois pour toutes. Le tableau ci-dessous les présente; l'annexe G en donne la légende complète, avec les couleurs des parties.

Tableau L.3 — Les repères visuels du livre

Repère	Rubrique	Couleur et rôle	Sens du pictogramme
✦	Ouverture	Canal 700 · #0A3D62 — structure	Une étoile qui s'allume : l'entrée dans le chapitre
◎	Objectifs	Canal 700 · #0A3D62 — structure	La cible et la trajectoire visée

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Repère	Rubrique	Couleur et rôle	Sens du pictogramme
	Prérequis	Canal 700 · #0A3D62 — structure	Les marches déjà gravies
	Situation d'ouverture	Canal 700 · #0A3D62 — structure	La longue-vue braquée sur un cas réel
	Posture épistémologique	Canal 500 · #1A5F8F — savoir de référence	Le regard que le chercheur porte sur les classes
	Hypothèse de travail	Laiton 500 · #A9762F — décision, réflexion	La fiole : ce qui reste à éprouver
	Étude de cas	Canal 700 · #0A3D62 — structure	La loupe sur les données
	Débat et paradigme	Laiton 500 · #A9762F — décision, réflexion	La balance des positions
	Pièges de l'apprenti-chercheur	Ocre 500 · #9C5518 — attention	Le signal d'attention, jamais une faute
	Conseil du praticien	Patine 500 · #3E7A6B — acquis, validé	La boussole de l'expérience
	À retenir	Patine 500 · #3E7A6B — acquis, validé	Le signet : ce qu'on garde
	Auto-évaluation	Canal 700 · #0A3D62 — structure	La case à cocher soi-même
	Pour aller plus loin	Canal 700 · #0A3D62 — structure	Le chemin vers l'étoile suivante
	Question de réflexion	Laiton 500 · #A9762F — décision, réflexion	La question ouverte au lecteur
	Mises en situation	Laiton 500 · #A9762F — décision, réflexion	L'embranchement : choisir une voie
	Note pédagogique	Canal 700 · #0A3D62 — structure	L'information de cadrage
	Corrigés	Patine 500 · #3E7A6B — acquis, validé	La validation
	Le coin des experts	Canal 500 · #1A5F8F — savoir de référence	L'astrolabe : l'instrument de précision
	Données Villeneuve (fil rouge)	Canal 700 · #0A3D62 — structure	La carte de la ville simulée

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Repère	Rubrique	Couleur et rôle	Sens du pictogramme
	ProGera	Logo officiel de ProGera (dégradé orange–violet)	La plateforme ProGera
	Annexes	Canal 700 · #0A3D62 — structure	Le livre ouvert
	Banque de cas du praticien	Patine 500 · #3E7A6B — acquis, validé	L'archive des cas vécus
	Connexion recherche et organisations	Canal 500 · #1A5F8F — savoir de référence	Le pont entre recherche et organisations
	Pages liminaires	Canal 700 · #0A3D62 — structure	Les pages qui ouvrent le livre
	Lecture Orchestrum	Canal 500 · #1A5F8F — savoir de référence	L'orbite ouverte du modèle et ses cinq pôles

Source : l'auteur.

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio

Sigles et abréviations

Sigle	Signification
ACP	analyse en composantes principales
AFDM	analyse factorielle de données mixtes
AHP	processus de hiérarchie analytique (analytic hierarchy process)
APA	American Psychological Association (norme de citation)
ARI	indice de Rand ajusté (adjusted Rand index)
BIC	critère d'information bayésien
CAH	classification ascendante hiérarchique
CST	code d'un espace de chapitre sur ProGera
csQCA	analyse qualitative comparée à ensembles nets
EFVP	évaluation des facteurs relatifs à la vie privée
EIS	évaluation d'impact sur la santé
EPTC 2	Énoncé de politique des trois conseils : éthique de la recherche avec des êtres humains
FAIR	principes de données faciles à trouver, accessibles, interopérables et réutilisables
HCPC	classification hiérarchique sur composantes principales
HDBSCAN	classification hiérarchique fondée sur la densité, avec détection du bruit
IA	intelligence artificielle
LCA	analyse en classes latentes (latent class analysis)
MEAL	suivi, évaluation, redevabilité et apprentissage
PDCA	cycle planifier, faire, vérifier, agir
QCA	analyse qualitative comparée (qualitative comparative analysis)
REGREDAR	registre global des risques du modèle Orchestrum
t-SNE	plongement stochastique de voisinage à distribution t
UMAP	approximation et projection uniformes de variétés

EXTRAIT · DATASTRATÈGE

Constellations — extrait de présentation, reproduction et diffusion interdites · Datastratège · Datastratège Studio