



Ressources naturelles
Canada

Natural Resources
Canada

**COMMISSION GÉOLOGIQUE DU CANADA
DOSSIER PUBLIC 9254**

Compilation de cartes géologiques du Canada

**C.J.M. Lawley, P. Giddy, L. Katz, N. Chu, A. Francis, J. Carvajal, M. Pinheiro,
K. Tirona, A. Dettman, H. Chen, F. Aucoin et J.E. Bourdeau**

2025

Canada 

COMMISSION GÉOLOGIQUE DU CANADA
DOSSIER PUBLIC 9254

Compilation de cartes géologiques du Canada

**C.J.M. Lawley¹, P. Giddy², L. Katz², N. Chu², A. Francis², J. Carvajal²,
M. Pinheiro², K. Tirona², A. Dettman², H. Chen², F. Aucoin³ et J.E. Bourdeau¹**

¹Ressources naturelles Canada, Commission géologique du Canada, 601, rue Booth, Ottawa (Ontario) K1A 0E8

²Orix Geoscience Inc., 1400-25, rue Adelaide Est, Toronto (Ontario) M5C 3A1

³Ressources naturelles Canada, Commission géologique du Canada, 490, rue de la Couronne, Québec (Québec) G1K 9A9

2025

© Sa Majesté le Roi du chef du Canada, représenté par le ministre des Ressources naturelles, 2025

Le contenu de cette publication ou de ce produit peut être reproduit en tout ou en partie, et par quelque moyen que ce soit, sous réserve que la reproduction soit effectuée uniquement à des fins personnelles ou publiques mais non commerciales, sans frais ni autre permission, à moins d'avis contraire.

On demande seulement :

- de faire preuve de diligence raisonnable en assurant l'exactitude du matériel reproduit;
- d'indiquer le titre complet du matériel reproduit et le nom de l'organisation qui en est l'auteur;
- d'indiquer que la reproduction est une copie d'un document officiel publié par Ressources naturelles Canada (RNCa) et que la reproduction n'a pas été faite en association avec RNCa ni avec l'appui de celui-ci.

La reproduction et la distribution à des fins commerciales sont interdites, sauf avec la permission écrite de RNCa. Pour de plus amples renseignements, veuillez communiquer avec RNCa à copyright-droitdauteur@nrcan-rncan.gc.ca.

On peut télécharger cette publication gratuitement à partir du Dépôt ouvert des sciences et technologie de RNCa (<https://ostrnrcan-dostrncan.canada.ca/>).

Notation bibliographique conseillée

Lawley, C.J.M., Giddy, P., Katz, L., Chu, N., Francis, A., Carvajal, J., Pinheiro, M., Tirona, K., Dettman, A., Chen, H., Aucoin, F. et Bourdeau, J.E. 2025. Compilation de cartes géologiques du Canada; Commission géologique du Canada, Dossier public 9254, 1 fichier .zip. <https://doi.org/10.4095/p9atxf8aec>

Les publications de cette série ne sont pas révisées; elles sont publiées telles que soumises par l'auteur.

ISSN 2816-7163
ISBN 978-0-660-76593-8
N° de catalogue M183-2/9254F-PDF
<https://doi.org/10.4095/p9atxf8aec>

Compilation de cartes géologiques du Canada

Christopher J.M. Lawley¹, Paige Giddy², Laura Katz², Natalie Chu², Aaron Francis², Juan Carvajal², Matchellon Pinheiro², Keane Tirona², Alexa Dettman², Howard Chen² Francis Aucoin³ et Julie E. Bourdeau¹

¹ Ressources naturelles Canada, Commission géologique du Canada, 601, rue Booth, Ottawa, Ontario, K1A 0E8, Canada

² Orix Geoscience Inc., 1400-25, rue Adelaide Est, Toronto, Ontario, M5C 3A1, Canada

³ Ressources naturelles Canada, Commission géologique du Canada, 490 rue de la Couronne, Québec, G1K 9A9, Canada

Déclaration relative aux auteurs (CRedit) : Conceptualisation (CJML), curation des données (PG; LK; NC; AF; JC; MP; KT; AD; HC; CJML), analyse formelle (CJML; PG; LK; NC; AF; JC; MP; KT; AD; HC), recherche (PG; LK; NC; AF; JC; MP; KT; AD; HC; CJML), méthode (CJML; PG; LK; NC; AF; JC; MP; KT; AD; HC), gestion du projet (CJML; PG), ressources (CJML; PG), logiciel (CJML; AF; JC; HC), supervision (CJML; PG), validation (tous), vérification (tous), visualisation (CJML), rédaction – préparation de l'ébauche initiale (CJML), rédaction – révision et correction (tous)

Résumé

La Compilation de cartes géologiques du Canada (CCGC) est une base de données sur les cartes géologiques du substratum rocheux publiées par les commissions géologiques provinciales, territoriales et autres. L'information substratum géoscientifique contenue dans ces cartes sources a été normalisée, traduite en anglais (subséquentement en français) et regroupée pour fournir une couverture complète du Canada et faciliter le recours ultérieur à diverses applications d'apprentissage machine. Il n'existait auparavant aucun produit unique à l'échelle nationale présentant des renseignements automatiques lithologiques, minéralogiques, métamorphiques, lithostratigraphiques et lithodémiques détaillés. Des améliorations ont également été apportées aux données sources : les erreurs de géométrie ont été corrigées et un nouveau système de classification hiérarchique de la lithologie générale a été appliqué pour subdiviser les types de roches originaux en 35 classes. Chaque lithologie générale est associée à une mesure semi-quantitative de l'incertitude de classification. Autant que possible, des correspondances ont été établies entre les noms lithostatigraphiques et lithodémiques des cartes sources et le Lexique en ligne des noms géologiques canadiens (Weblex). De plus, le traitement du langage naturel a été utilisé pour transformer toute l'information textuelle en unités lexicales (jetons). Les entités polygonales superposées des jeux de données surficielles et les artéfacts de délimitation traversant les frontières politiques n'ont pas été retravaillés dans le cadre de cet l'étude. De ce fait, la CCGC est une mosaïque de cartes géologiques du substratum rocheux qui se chevauchent et dont les échelles (1/30 000 à 1/5 000 000), années de publication (1996 à 2023) et leur fiabilité varient. Les cartes géologiques et géochronologiques du Canada, qui reposent sur les

meilleures données nouvelles géoscientifiques extraites des polygones superposés pour chaque pixel des cartes, sont fournies sous forme de cartes géospatiales matricielles. D'autres cartes géologiques à résolution supérieure seront ajoutées au fil du temps pour combler données manquantes et mettre à jour l'information géoscientifique aux fins de futures applications de la CCGC.

1.0 Introduction

Les cartes géologiques du substratum rocheux sont des représentations visuelles de la roche solide qui se trouve à la surface ou près de celle-ci. Des informations précises et détaillées sur le substratum rocheux sont essentielles pour diverses applications, se rapportant notamment au potentiel minier, aux eaux souterraines, aux géorisques, à la géothermie, à l'ingénierie et à la construction, aux effets environnementaux, à la recherche géoscientifique et à l'éducation. La Commission géologique du Canada (CGC) a publié sa première « Carte géologique du Canada » avant la Confédération. Cette carte comprenait des parties du nord-est américain et les provinces du sud-est du Canada (Logan, 1864). De nombreuses compilations de cartes géologiques du substratum rocheux ont été publiées depuis. Elles offrent une couverture complète de la masse substratum continentale et du domaine extracôtier du pays à des niveaux de détail et de fiabilité variables (Service de géologie et de topographie du Canada, 1945; Douglas, 1969; Commission géologique du Canada, 1884, 1955, 1962; Logan, 1869; Reed et al., 2005; Rice, 1955; Van Hise, 1896; Wheeler et al., 1996; Willis, 1906; Young, 1913, 1924). Avec le temps, les cartes papier ont évolué pour devenir des bases de données numériques dans lesquelles l'information géoscientifique est stockée sous forme d'attributs liés à des polygones géospatiaux. Les données géoscientifiques contenues dans ces jeux de données numériques sont essentielles aux applications en aval qui font appel aux systèmes d'information géographique (SIG) et à d'autres modes d'analyse géospatiale (Harrison et al., 2011; Reed et al., 2005; Wheeler et al., 1996).

La plus récente carte géologique du Canada a été publiée par Wheeler et al. (1996), et les données numériques qui accompagnent cette carte papier à échelle de 1/5 000 000 sont devenues les produits les plus téléchargés du dépôt ouvert de la science et technologie (DOST), la base de données des publications de Ressources naturelles Canada. Les attributs normalisés fournis comprennent les provinces et domaines géologiques, la géologie générale, les types de roche primaires et secondaires, les âges géologiques et le degré d'intensité du métamorphisme. Chacun d'eux est indiqué en français et en anglais, pour une meilleure accessibilité. La carte de Wheeler et al. (1996) est l'un des éléments sur lesquels se fonde la Carte géologique de l'Amérique du Nord à l'échelle de 1/5 000 000 (Reed et al., 2005). Hoffman (2015) présente un compte rendu historique de la vie de John O. Wheeler et des recherches géoscientifiques fondamentales qui sous-tendent ces deux contributions remarquables.

Depuis la publication de la carte de Wheeler et al. (1996), les commissions géologiques provinciales, territoriales et autres ont beaucoup amélioré la qualité et le niveau de détail de leurs bases de données et de leurs cartes géologiques du substratum rocheux. La plupart de ces bases de données publiées plus récemment ont des échelles supérieures à celle de la compilation de portée nationale (1/5 000 000), et révèlent de 5 à 100 fois plus de détails géologiques. Ces cartes détaillées montrent les contacts géologiques avec une précision et une exactitude accrue, un avantage important pour les vastes régions du pays qui sont caractérisées par une géologie du substratum rocheux complexe, et pour l'inclusion de petites intrusions ignées qui sont difficiles à visualiser sur les cartes papier à l'échelle

nationale. De grands progrès ont également été réalisés dans les bases de données récentes quant à l'indication de renseignements géoscientifiques détaillés comme les types de roche, les âges, la minéralogie, les données métamorphiques, lithostratigraphiques et lithodémiques, et les descriptions des affleurements qui appuient l'interprétation privilégiée. Aucun de ces renseignements n'était auparavant disponible à l'échelle nationale. De plus, les bases de données géospatiales qui accompagnent ces plus récentes cartes géologiques du substratum rocheux suivent leur propre modèle de données, et l'information géoscientifique associée à chaque polygone géospatial n'avait jamais été normalisée et combinée en un seul produit numérique.

La Compilation de cartes géologiques du Canada (CCGC) est un produit qui combine l'information géoscientifique traduite, révisée et normalisée provenant de 32 géodatabases publiées, dans le but de d'adresser le manque d'une nouvelle compilation (tableau 1). Les données sont présentées dans une série de dossiers nommés d'après leur format de fichier. L'information géoscientifique normalisée associée aux polygones qui se chevauchent est contenue dans une base de données, en format GeoPackage (dossier « ccgc_gpkg »). La même information est également fournie en format CSV (données non spatiales), aux fins des applications d'apprentissage (dossier « ccgc_csv »). Les fichiers CSV comprennent les coordonnées originales et modifiées des polygones en format Well-known text (WKT), qui peuvent être reconverties en polygones géospatiaux au automatiques besoin. Sept tableaux électroniques contenant les définitions des données géoscientifiques normalisées ainsi que des lexiques, assortis de renvois à des normes de données internationales comme GeoSciML (v2016.01), sont fournis en format Microsoft Excel XLSX dans le dossier « ccgc ». Les progrès continus réalisés en ce qui concerne ce type de normes de données internationales (p. ex. North America Data Model, GeoSciML et INSPIRE) sont susceptibles de grandement améliorer l'interopérabilité des bases de données géospatiales (Laxton, 2017; Sen et Duffy, 2005). De plus, l'indexation spatiale a été utilisée pour établir de nouvelles cartes géologiques et géochronologiques du Canada en fonction des meilleures données géoscientifiques disponibles pour chaque pixel des cartes. Les cartes géologiques et géochronologiques choisie sont fournies sous forme de cartes géospatiales matricielles en format GeoTIFF assorties d'un code de couleur (dossier « ccgc_geotiff ») et sous forme d'images JPEG en couleur prêtes à être publiées (dossier « ccgc_jpeg »).

2.0 Données

2.1 Données provenant des cartes sources

L'information de référence pour chaque carte source figure dans le tableau 1. Les utilisateurs de la CCGC sont invités à consulter ces sources pour obtenir de plus amples renseignements et les données actualisées. Les utilisateurs peuvent trouver la source des données cartographiques contenues dans la CCGC à l'aide du numéro de l'entité polygonale originale (SOURCE_FID). Le numéro d'identification de l'élément source est propre à chaque carte source. Un identifiant de référence unique (SOURCE_REFERENCE_ID) a également été attribué à chaque carte source en fonction de l'année de publication, de l'abréviation du nom de l'éditeur, et du nom de la publication (tableau 1). Un contrôle de ces identifiants sera effectué pour assurer qu'ils demeurent uniques à mesure que de nouvelles cartes géologiques seront ajoutées à la CCGC. En outre, les attributs SOURCE_FID et SOURCE_REFERENCE_ID ont été combinés pour produire un identifiant permanent

(permanent identifier data; PID) unique à chaque polygone compris dans la CCGC (tableau 2). Le PID calculé restera unique et ne changera pas, même si de nouvelles cartes géologiques sont ajoutées à la CCGC au fil du temps. Au besoin, les utilisateurs peuvent se servir de clés lors des jointures de bases de données. Les attributs géospatiaux SOURCE_FID, SOURCE_REFERENCE_ID et PID sont obligatoires (tableau 2). Les autres attributs obligatoires qui fournissent de clés l'information sur les sources sont l'année de publication, les échelles cartographiques (maximum et minimum), la géométrie originale et la référence bibliographique complète (tableau 2). Les renseignements facultatifs sur la source comprennent l'identifiant d'objet numérique (digital object identifier; DOI), qui n'était pas disponible dans certains cas (tableau 2). Les renseignements manquants sur les cartes sources sont indiqués comme étant « inconnus ». Tous les polygones géospatiaux ont été convertis au système de référence géodésique de Lambert de l'Atlas du Canada (EPSG : 3978). Des cartes-index, montrant l'étendue spatiale de chaque carte géologique source sont présentées à la figure 1.

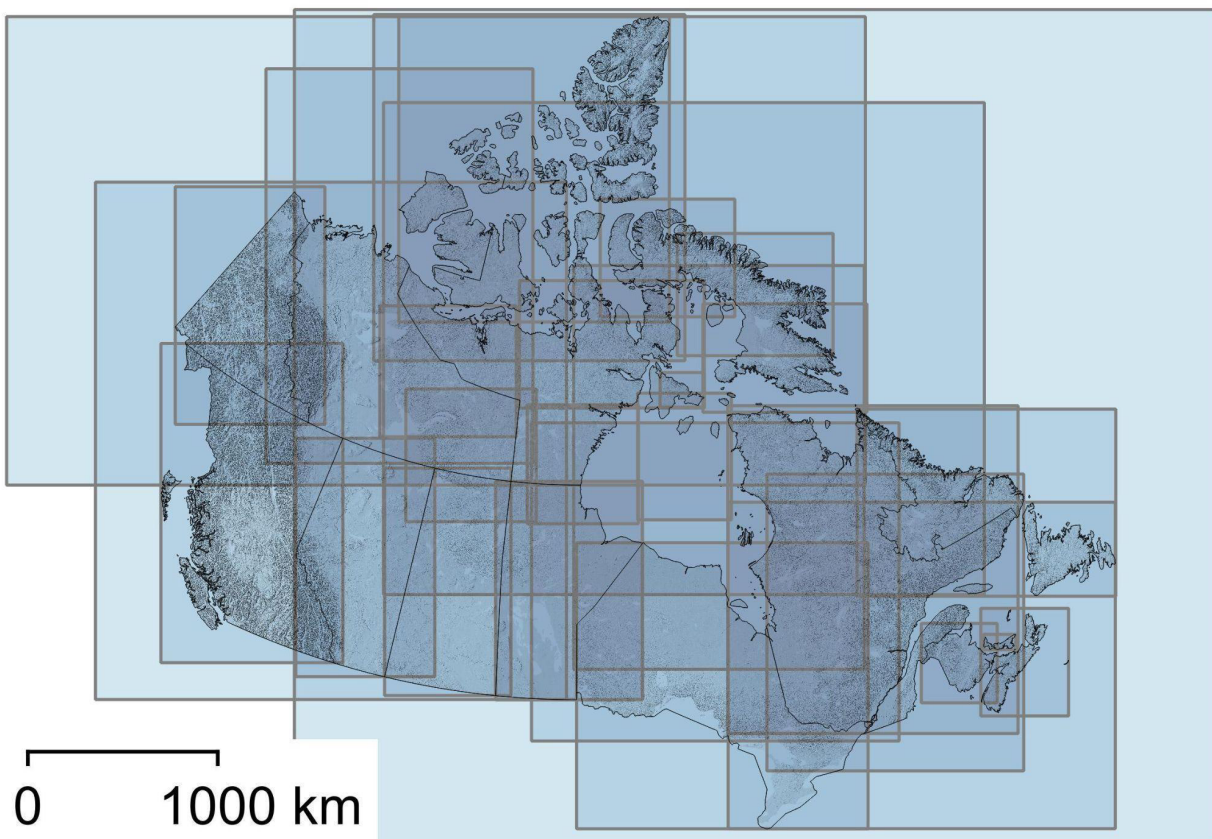


Figure 1 : Étendue spatiale des cartes géologiques sources comprises dans la CCGC ($n = 32$).

2.2 Données lithologiques

Trois méthodes servent à normaliser les données lithologiques des cartes sources. Tout d'abord, l'ensemble des données géoscientifiques disponibles est utilisé pour reclasser les polygones selon 35 lithologies générales (LITHOLOGIE_CLASSIFICATION; tableau 3). Ces lithologies générales sont subdivisées en trois classes de roches : 1) « ignées », 2) « sédimentaires », et 3) « métamorphiques » (tableau 3). Les polygones qui n'entrent pas dans ces grandes classes sont

décrits comme « autres ». Les lithologies générales reposent sur le protolithe qui a été interprété à partir des attributs sédimentaires ou ignées primaires reconnus dans les données textuelles disponibles. De ce fait, la CCGC est une carte géologique interprétative des protolithes rocheux. La classe « métamorphique », quant à elle, est réservée aux polygones dépourvus d'attributs sédimentaires ou ignées primaires. La fiabilité des descriptions de roches utilisées pour interpréter ces lithologies générales est indiquée au moyen de l'échelle d'incertitude semi-quantitative de Tikoff et al. (2023), dans l'attribut LITHOLOGIE_CLASSIFICATION_INCERTITUDE (certain = 100 %; convaincant = 75 %, présomptif = 50 à 75 %, suggestif = 25 à 50 %, permissif = < 25 %, aucune preuve = 0 %; tableau 2). Les attributs LITHOLOGIE_CLASSIFICATION et LITHOLOGIE_CLASSIFICATION_INCERTITUDE sont obligatoires.

Ensuite, les types de roches présents dans les cartes sources sont listés en ordre décroissant d'abondance lorsque cette information a été fournie, sinon en ordre d'apparition (LITHOLOGIE_01 à LITHOLOGIE_10). Les lithologies ignées, sédimentaires et métamorphiques ont autant que possible été reclassées sous des noms officiels pour faciliter les recherches (voir la description plus loin). Une liste complète des lithologies uniques et des termes GeoSciML avec la meilleure correspondance sont présentés dans le tableau 4. Les attributs LITHOLOGIE_01 à LITHOLOGIE_10 sont facultatifs. Presque tous les polygones des cartes sources contenaient moins de 10 types de roches, mais il se pourrait que le nombre d'attributs LITHOLOGIE seront augmenté dans les versions futures de la CCGC. Les types de roches manquants sont indiqués comme « inconnus ».

Enfin, toutes les données textuelles disponibles, y compris les types de roches originaux, sont converties en jetons de mots à l'aide du traitement du langage naturel (voir la description à section 3.0). Un point-virgule sert à séparer les jetons. L'orthographe de tous les champs de description a été vérifiée avant le traitement du langage naturel. Toutes les données lithologiques textuelles sont inscrites en minuscules.

2.2.1 Roches ignées

Les roches ignées sont subdivisées en 13 classes de lithologie principale en fonction de leur mode de mise en place (roche intrusive ou extrusive) et de leur composition (tableau 3). Les roches ignées extrusives à grains fin se subdivisent dans les classes « ignée_extrusive_alcaline » (p. ex. phonolite, téphrite, basanite), « ignée_extrusive_felsique » (p. ex. rhyolite), « ignée_extrusive_intermédiaire » (p. ex. andésite) et « ignée_extrusive_mafique » (p. ex. basalte). Les roches ignées extrusives indifférenciées de composition incertaine ou mixte se voient attribuer la classe « ignée_extrusive » aux fins de la classification lithologique générale. Les types de roches ignées extrusives (LITHOLOGIE_01 à LITHOLOGIE_10) compris dans les cartes sources sont reclassifiés selon le diagramme alcalins-silice (TAS) de l'Union internationale des sciences géologiques (Le Maitre et al., 2002), dans la mesure du possible. Par exemple, des roches décrites à l'origine comme étant « volcaniques mafiques » sont reclassées comme « basaltes » dans la CCGC (tableau 4). L'attribut LITHOLOGIE_CLASSIFICATION des roches volcanoclastiques indique la composition de roche extrusive correspondante; la classe « ignée_extrusive » est utilisée pour les roches volcanoclastiques indifférenciées de composition incertaine. Les systèmes de classification des roches volcanoclastiques publiés antérieurement n'ont pas été utilisés dans le cadre de cette étude, car les noms plus précis qu'ils emploient étaient souvent absents des ensembles de données sources et que bon nombre des noms proposés ont plusieurs définitions (Cas et Wright, 1988; Fisher, 1961,

1966; Frolova, 2008; White et Houghton, 2006). De ce fait, tous les types de roches volcanoclastiques (p. ex. tuf, cendres, lapilli, lapillistone, tuf à lapilli, brèche de tuf, brèche pyroclastique, brèche volcanique, volcanoclastite, agglomérat) mentionnés dans les ensembles sources sont rangés dans la classe « roche volcanoclastique » dans la CCGC (tableau 4). Les types de roches volcanoclastiques d'origine qui reposent sur un ou plusieurs systèmes de classification formels de ces roches ont été conservés dans l'attribut des jetons. Les qualificatifs ont servi à interpréter l'attribut LITHOLOGIE_CLASSIFICATION, mais pas à déterminer les attributs LITHOLOGIE_01 à LITHOLOGIE_10, à moins qu'ils n'indiquent un type de roche précis (p. ex. les roches volcaniques mafiques, intermédiaires et felsiques sont respectivement le basalte, l'andésite et la rhyolite; tableau 4).

Les roches ignées intrusives sont toutes à grain grossier et se subdivisent dans divers sous-groupes en fonction de leur composition, notamment « ignée_intrusive_felsique » (p. ex. granite), « ignée_intrusive_intermédiaire » (p. ex. diorite), « ignée_intrusive_mafique » (p. ex. gabbro) et « ignée_intrusive_ultramafique » (p. ex. péridotite). La classe « ignée_intrusive » sert aux roches ignées intrusives indifférenciées d'origine incertaine (tableau 3). Bon nombre de ces intrusions ignées mal documentées ont été déduites à partir de levés géophysiques. Deux autres classes de roches ignées intrusives sont utilisées pour certains types de roches en raison de leur importance historique pour la description de la texture des terrains (ignée_intrusive_felsique_pegmatite) ou parce qu'ils ne correspondaient pas bien aux quatre principaux groupes de composition (ignée_intrusive_anorthosite; tableau 3). Les types de roches ignées intrusives originaux sont reclassifiés selon le diagramme de l'Union internationale des sciences géologiques quartz-feldspaths alcalins-plagioclases-feldspathoïdes (QAPF) et le diagramme mafique-ultramafique (Streckeisen, 1976), dans la mesure du possible; ce sont ces classes qui figurent dans les attributs LITHOLOGIE_01 à LITHOLOGIE_10. Par exemple, des roches décrites à l'origine comme des « granitoïdes » sont reclassées comme des « granites » dans la CCGC (tableau 4). Les qualificatifs ont servi à interpréter l'attribut LITHOLOGIE_CLASSIFICATION, mais pas à déterminer la présence de lithologies particulières de roches ignées intrusives (p. ex. les termes « granitique », « granité » et « de type granite » n'ont pas été utilisés pour établir le granite en tant que lithologie pour les attributs LITHOLOGIE_01 à LITHOLOGIE_10; tableau 4).

2.2.2 Roches sédimentaires

Les roches sédimentaires sont subdivisées en six classes de lithologie générale en fonction de leur composition globale ou de la taille de leurs grains (sédimentaire_siliciclastique, sédimentaire_siliciclastique_grossière, sédimentaire_siliciclastique_fin, sédimentaire_chimique, sédimentaire_chimique_carbonate, et sédimentaire_chimique_évacuorite; tableau 3). Les roches siliciclastiques comprennent les clastes à base de silice qui sont généralement dépourvus de minéraux carbonatés. Des dizaines de systèmes ont été proposés pour classer ces roches en fonction de la taille des grains, de la composition, de la texture, du milieu de déposition ou des mécanismes de transport (Crook, 1960; Dickinson, 1970; Dott, 1964; Folk, 1954, 1980; Garzanti, 2019; Klein, 1963; Lazar et al., 2015; Lundegard et Samuels, 1980; Macquaker et Adams, 2003; Okada, 1971; Pettijohn, 1954; Picard, 1971; Shepard, 1954; Spears, 1980; Stow, 1981). Les utilisations correctes et incorrectes de ces systèmes posent des difficultés considérables lors de la création de compilations de portée nationale qui comprennent des observations de centaines de géoscientifiques faites au cours de nombreuses décennies. Par conséquent, cette étude se concentre sur la reclassification des dizaines de types de roches sur la division des dizaines de types de roches siliciclastiques possibles en quatre lithologies fondées sur la taille des grains, soit 1) brèche, 2) conglomérat, 3) grès (p. ex. arénite, arkose, psammite,

grès, wacke, gravillon, orthoquartzite et localement quartzite) et 4) roche argile (p. ex. grès limoneux, mudstone, argilite, schiste, ardoise et pélite; tableau 4). Les qualificatifs (p. ex. sableux, silteux, vaseux) n'ont pas été pris en compte pendant le processus de reclassification, pour concentrer la réinterprétation des types de roches siliciclastiques sur le mot racine aux fins de la détermination des attributs LITHOLOGIE_01 à LITHOLOGIE_10. Le nombre réduit de lithologies siliciclastiques simplifie les recherches, mais tous les types de roches originaux sont indiqués dans l'attribut des jetons pour offrir une nomenclature plus précise, quoique mixte.

Les roches sédimentaires chimiques (allochimiques et orthochimiques) se forment par des processus biologiques et chimiques, et englobent divers types de roches qui ont été redivisés en trois lithologies générales. Le calcaire est la roche sédimentaire chimique la plus commune et comprend principalement la calcite ou l'aragonite (CaCO_3). Le calcaire contenant du calcium a été différencié autant que possible de la dolomie contenant du magnésium et du calcium [$\text{CaMg}(\text{CO}_3)_2$] en fonction des données textuelles disponibles pour les attributs LITHOLOGIE_01 à LITHOLOGIE_10 (tableau 4). Les roches dont la classe était « *dolomite* » ont été reclassées comme des « dolomie » par souci d'uniformité (tableau 4). Cependant, le calcaire et la dolomie sont combinés dans la classe « sédimentaire_chimique_carbonate » aux fins de la classification lithologique générale (tableau 3). De nombreux systèmes ont été proposés pour classer le calcaire et les autres types de roches carbonatées en fonction de la taille des grains, du type de claste, de la texture, des processus diagénétiques et des milieux de dépôt (Bissell et Chilingar, 1967; Dunham, 1962; Folk, 1959; Wolf, 1960; Wright, 1992). Pour faciliter les recherches, les types de roches compris dans ces systèmes (p. ex. mudstone calcaire, mudstone, wackestone, packstone, grainstone, boundstone, bindstone, carbonate cristallin, floatstone, rudstone, bafflestone, framestone) ont été reclassifiés comme du calcaire dans les attributs LITHOLOGIE_01 à LITHOLOGIE_10 (tableau 4), dans le cadre de cette étude. Les roches siliciclastiques décrites à l'aide de qualificatifs laissant entendre une proportion importante de minéraux carbonatés sont aussi classifiées comme du calcaire, pour limiter le nombre de lithologies dans les recherches (p. ex. les termes « micrite », « marne » et « chaux » servent tous à inférer qu'il s'agit de calcaire plutôt que d'une roche siliciclastique; tableau 4).

Les évaporites se forment dans les milieux marins ou lacustres et constituent la deuxième subdivision lithologique générale des roches sédimentaires chimiques (tableau 3). La calcite, le gypse, l'anhydrite, l'halite et la potasse sont les principaux minéraux qui précipitent des solutions et se concentrent pour former des dépôts d'évaporite. Dans de nombreux cas, il a fallu déduire la présence d'exhalites et de dépôts d'évaporite à partir des noms des minéraux plutôt que des types de roches. Les qualificatifs ont été utilisés pour déterminer l'attribut LITHOLOGIE_CLASSIFICATION, mais pas pour déduire les lithologies particulières (p. ex. le terme « évaporitique » ne sert pas à donner la valeur « évaporite » aux attributs LITHOLOGIE_01 à LITHOLOGIE_10).

Les autres classes de la lithologie générale « sédimentaire_chimique » comprennent divers types de roches chimiques ou biologiques, notamment le chert, les formations de fer, la phosphorite, la barytine et le charbon (tableau 3). Ces roches ont tendance à être interstratifiées avec des roches ignées extrusives ou siliciclastiques fines dans des milieux de dépôt pauvres en sédiments, mais sont formées par un éventail de processus biologiques et chimiques différents. À l'exception des formations de fer (p. ex. taconite, formation de fer rubanée), la plupart de ces types de lithologie ont été directement tirés des descriptions contenues dans les sources, et n'ont pas été modifiés (tableau 4). Les qualificatifs ont été utilisés pour déterminer l'attribut LITHOLOGIE_CLASSIFICATION, mais non

pour déduire les lithologies particulières (p. ex. le terme « cherteux » ne sert pas à donner la valeur « chert » aux attributs LITHOLOGIE_01 à LITHOLOGIE_10).

2.2.3 Roches métamorphiques

Les roches métamorphiques sont subdivisées en 11 classes de lithologie générale, et se limitent aux types de roches dépourvues d'attributs ignées ou sédimentaires primaires (tableau 3). Les classes générales de roches métamorphiques reposent sur la combinaison des caractéristiques texturales (roche granoblastique ou foliée) et minéralogiques. Chacune est fondée sur des noms établis dans la littérature scientifique, bien que la définition précise des types métamorphiques varie selon le système de classification informel utilisé. La classe « métamorphique_charnockite » se rapporte à un type de granofels comprenant diverses roches qui contiennent de l'orthopyroxène, du quartz et du feldspath et se forment dans les conditions du faciès métamorphique des granulites (p. ex. charnockite, enderbite, mangérite et opdalite; tableau 3). La classe de lithologie générale « métamorphique_migmatite » désigne aussi des roches qui se forment dans les conditions du faciès métamorphique des granulites, mais contrairement à la classe des charnockites, cette classe repose sur la texture plutôt que la minéralogie (tableau 3). Les autres granofels comprennent élogite (c.-à-d. les roches composées de grenat et de pyroxène), l'amphibolite (c.-à-d. les roches surtout composées d'amphibole), marbre (c.-à-d. les roches surtout composées de calcite) et quartzite (c.-à-d. les roches surtout composées de quartz), qui sont classés en fonction de leur minéralogie observée (tableau 3). Ces types de granofels peuvent se trouver dans un large éventail de conditions de pression et de température. En revanche, les roches métamorphiques foliées se divisent dans les classes « métamorphique_gneiss » et « métamorphique_schiste » en fonction de leur texture (tableau 3). La classe « métamorphique_gneiss » se subdivise en orthogneiss et en paragneiss d'après le protolithe le plus probable, respectivement igné ou sédimentaire. Toutes les autres roches métamorphiques fortement foliées sont placées dans la classe « métamorphique_schiste », quelle que soit la taille de leurs grains (p. ex. schiste, phyllite; tableau 3). Les noms historiques des roches métamorphiques et les roches métamorphiques appartenant à un même groupe ont été ramenés à des lithologies standard aux fins des attributs LITHOLOGIE_01 à LITHOLOGIE_10 (tableau 4).

2.2.4 Autres

Les sédiments non consolidés et les roches qui ne sont pas clairement décrites comme étant ignées, métamorphiques ou sédimentaires sont reclassifiés sous cinq classes supplémentaires : « autre_faille » (p. ex. cataclasite, mylonite, pseudotachylite), « autre_hydrothermal » (p. ex. veines, roches métagénétiques), « autre_mélange » (p. ex. mélange), « autre_superficiel » (p. ex. sédiments non consolidés à semi-consolidés) et « autre_inconnu » (p. ex. zones non cartographiées, glace, eau et roches d'origine incertaine; tableau 3). Les classes de type « autre » regroupent le plus vaste éventail de noms originaux ayant dû être reclassifiés pour faciliter les recherches. Par exemple, la lithologie « roche de faille » sert dans les attributs LITHOLOGIE_01 à LITHOLOGIE_10 à reclasser des dizaines de noms originaux de roches de faille qui ne correspondaient pas nécessairement aux multiples classifications élaborées au cours des années pour ce type de roches (Higgins, 1971; Killick, 2003; Sibson, 1977; Woodcock et Mort, 2008). De même, les veines et les autres formes d'altération hydrothermale qui masquent complètement la lithologie du protolithe sont regroupées sous la lithologie « roche métagénétique ». Considérant qu'il est difficile de procéder à l'interprétation de l'origine de ces types de roches altérées par des processus hydrothermaux avec les renseignements disponibles, le

protolithe est alors utilisé pour la classification lithologique généralisée dans la mesure du possible. La classe des mélanges comprend des brèches à matrice contenant des blocs de lithologies mixtes, qui tendent à être associées avec les marges continentales. Le processus diapirique, structural ou sédimentaire à l'origine des mélanges est suffisamment distinct des failles (et des autres classes de lithologie générale) pour mériter sa propre classe (Festa et coll., 2010). Cependant, il faut noter que le terme « mélange » a été utilisé à la fois comme terme descriptif et comme lithologie dans les cartes sources et que le concept a également évolué au fil du temps. La classe « autre_superficiel » comprend les dépôts glaciaires et les autres dépôts sédimentaires superficiels. Tous les noms d'origine utilisés pour décrire ces dépôts superficiels sont reclassés en lithologie « dépôt superficiel » pour les attributs LITHOLOGIE_01 à LITHOLOGIE_10, car la CCGC est destinée à être utilisée comme une compilation de données géoscientifiques sur le substratum rocheux (tableau 4). Enfin, la classe des inconnus sert à reclasser toutes les zones non cartographiées et les roches d'origine incertaine. Tous les termes géoscientifiques d'origine utilisés pour décrire les polygones sont préservés en tant que jetons, pour donner des renseignements contextuels supplémentaires.

2.3 Données géochronologiques

Les données géochronologiques sont normalisées et subdivisées selon la charte chronostratigraphique internationale (CCI) de la commission internationale de stratigraphie (ICS, version de septembre 2023; Cohen et al., 2013). Aucune tentative n'a été faite pour corriger les cartes sources en fonction de l'évolution des définitions employées dans la CCI au fil du temps. Les âges maximums et minimums sont indiqués en utilisant les subdivisions temporelles les plus précises disponibles dans les cartes sources, avec les subdivisions temporelles plus générales assignées en conséquence. Par exemple, une roche sédimentaire déposée au cours du « Cambrien » est attribuée par défaut à l'ère « Paléozoïque » et à l'éon « Phanérozoïque », même si les subdivisions plus générales n'étaient pas comprises dans les cartes sources. Cependant, l'époque et l'âge de cette roche sont classés comme « inconnus ». Dans le cas où un seul âge était indiqué pour un polygone, les données d'âges maximum et minimum sont identiques. L'intégration des données des cartes sources dans la CCI a également nécessité de renommer les polygones de l'ère « Tertiaire » en périodes « Paléogène » et « Néogène ». De même, les âges indiqués comme « Précambrien » dans les cartes sources sont traduits par les valeurs « Archéen » et « Protérozoïque » dans les attributs d'âges maximum et minimum, respectivement. Les âges maximums et minimums absolus et leurs incertitudes sont rapportés tels quels, et peuvent correspondre à des intrusions et des dykes plus jeunes que leurs roches hôtes. Par conséquent, les âges absolus peuvent être incompatibles avec les autres informations sur l'âge fournies dans le format ICC. Toutes les informations géochronologiques manquantes sont signalées comme « inconnues ». Les informations géochronologiques textuelles sont indiquées en majuscules.

2.4 Données minéralogiques

Les données minéralogiques sont tirées directement des cartes sources, et la liste des noms officiels des minéraux publiée par la Commission on New Minerals, Nomenclature and Classification de l'International Mineralogical Association (CNMNC; Janvier 2023) a été utilisée autant que possible. Cependant, il a souvent fallu dévier de ces noms officiels en fonction du contenu des cartes sources, notamment pour employer des groupes de minéraux courants (p. ex. amphibole, grenat, pyroxène et

serpentine). Les noms historiques ont été convertis en leur groupe minéral correspondant (c.-à-d. hornblende = amphibole; hypersthène = orthopyroxène) pour faciliter les recherches. Les qualificatifs spéciaux servant à décrire les altérations métamorphiques ou hydrothermales ont été convertis en groupes de minéraux dans la mesure du possible. Toutes les données minéralogiques manquantes sont indiquées comme « inconnues ».

2.5 Données métamorphiques

Les faciès métamorphiques maximums et minimums reposent sur les données des cartes sources et se subdivisent en faciès normalisés, notamment les faciès des cornéennes, zéolites, prehnite-pumpellyite, schistes verts, amphibolites, granulites, schistes bleus et élogites (Schmid et coll., 2007). Lorsque les polygones indiquent qu'un seul gradient métamorphique, les attributs METAMORPHIQUE_MAXIMUM et METAMORPHIQUE_MINIMUM contiennent la même information. Par conséquent, l'attribut METAMORPHIQUE_MAXIMUM fournit probablement l'information la plus utile pour un grand nombre de polygones ne comprenant pas de données sur les variations des conditions de pression et de température. Les roches dites granulites et élogites sont utilisées pour inférer un métamorphisme au faciès des granulites et des élogites, respectivement, même si le degré d'intensité du métamorphisme n'était pas inscrit sur les cartes sources. Cependant, les amphibolites n'ont pas mené à la détermination d'un métamorphisme au faciès des amphibolites, parce que de multiples processus géologiques peuvent produire des roches riches en amphibole et que leur mode d'origine était rarement présent dans les données sources. De plus, les qualificatifs servant à décrire les faciès métamorphiques ont été exclus des renseignements métamorphiques normalisés, parce qu'il est possible qu'ils aient été employés de différentes manières au fil du temps (p. ex. supérieur, milieu, inférieur, faible, moyen, élevé). Toutes les données métamorphiques manquantes sont indiquées comme « inconnues ». L'information métamorphique textuelle est inscrite en minuscules.

2.6 Données lithostratigraphiques

Les données lithostratigraphiques sont normalisées dans les unités suivantes, d'après les lignes directrices du Guide stratigraphique international (GSI) et du Code stratigraphique nord-américain (CSNA) : « supergroupe » (deux groupes ou plus), « groupe » (deux formations ou plus), « formation » (unité fondamentale de la lithostratigraphie), « membre » (subdivision lithologique nommée d'une formation), « lit » (couche distinctive nommée dans un membre ou une formation) et « coulée » (plus petite couche distinctive d'une séquence volcanique; ISSC, 2000; Murphy et Salvador, 1999; NASC, 1983, 2005, 2021; Salvador, 1994). Chacun de ces termes est réservé aux roches ignées ou sédimentaires qui respectent le principe de superposition comme décrit dans le GSI et le CSNA. Tous les autres noms lithostratigraphiques non conformes à la nomenclature de ces ouvrages de référence sont indiqués dans les attributs LITHOSTRATIGRAPHIQUE_AUTRE. Pour certaines cartes géologiques, les renseignements lithostratigraphiques ont dû être tirés de tableaux électroniques distincts et associés aux polygones à l'aide d'un code alphanumérique. Les noms lithostratigraphiques et les liens entre les différents niveaux d'unités n'ont pas été réinterprétés dans le cadre de cette étude. Il est donc déconseillé de tenter de déterminer de tels liens à partir de l'information lithostratigraphique normalisée de la CCGC. Par exemple, les multiples noms d'une formation ne correspondent pas nécessairement au même nom de groupe, et ce même à l'intérieur d'un seul polygone. Les données biostratigraphiques, allostratigraphiques et pédostratigraphiques n'ont pas été

normalisées, mais toute l'information textuelle extraite des cartes sources est préservée dans l'attribut des jetons de mots. Toutes les données lithostratigraphiques manquantes dans les attributs normalisés sont indiquées comme « inconnues ». L'information lithostratigraphique textuelle est inscrite en lettres majuscules.

De plus, les données lithostratigraphiques normalisées ont été mises en correspondance avec le Lexique en ligne des noms géologiques canadiens (Weblex; <https://weblex.canada.ca>) lorsqu'il était possible de le faire à l'aide du nom normalisé et de l'endroit (province ou territoire). Cette base de données fournit beaucoup plus d'information sur le statut, l'emplacement, l'âge, la lithologie et l'histoire des noms lithostratigraphiques au Canada. Les descriptions détaillées contenues dans Weblex sont particulièrement importantes pour les applications de traitement du langage naturel utilisées en aval (Lawley et al., 2023). Au fil du temps, les utilisateurs ajoutent de nouveaux noms à Weblex, alors les liens avec les noms lithostratigraphiques de la CCGC devront être mis à jour dans les futures publications. Tous les renseignements de Weblex sont reportés tels quels et dans la même casse, de sorte que la forme longue des noms anglais puisse servir à lier la CCGC et Weblex aux fins des applications en aval. Les renseignements absents de Weblex sont indiqués comme « inconnus ».

2.7 Données lithodémiques

Les données lithodémiques ne sont fournies que pour les roches qui ne respectent pas le principe de superposition (Gillespie et Leslie, 2021; Hattin, 1991; NASC, 1983, 2005, 2021). Les cartes sources fournissent ce type de données en indiquant les assemblages cartographiables de roches métamorphiques, ignées ou hautement déformées présentant une importance particulière. Cependant, les systèmes de dénomination lithodémique n'ont pas une structure aussi formelle que la nomenclature lithostratigraphique, et les cartes sources contiennent beaucoup de noms qui ne se conforment à aucune norme connue (Gillespie et Leslie, 2021; NASC, 1983, 2005, 2021). Par conséquent, les noms lithodémiques ont été traités tels quels et regroupés en supercomplexes (deux complexes ou plus), complexes (association d'au moins deux roches intrusives et très déformées ou très métamorphisées), supersuites (deux suites ou plus) et suites (association d'au moins deux roches intrusives et très déformées ou très métamorphisées). Tous les autres noms lithodémiques informels (p. ex. séries, zones, essaims, plutons, sills, dykes) ont été combinés dans les attributs LITHODEMIQUE_AUTRE. Le rang et l'importance de ces autres noms lithodémiques sont très variables. Des recherches supplémentaires sont nécessaires pour classer de façon formelle l'information lithodémique à une échelle nationale (Gillespie et Leslie, 2021). Comme pour les informations lithostratigraphiques décrites ci-dessus, les polygones cartographiques individuels peuvent contenir plusieurs noms de lithodèmes qui ne sont pas liés entre eux. L'information lithodémique textuelle est indiquée en majuscules. L'information lithodémique normalisée est liée à la base de données Weblex dans la mesure du possible. Tous les renseignements de Weblex sont reportés tels quels. Les données lithologiques manquantes sont indiquées comme « inconnues ».

3.0 Méthodes de traitement du langage naturel

L'ensemble du traitement de texte a été fait à l'aide des paquets « tidytext » (Silge et Robinson, 2016), « deeplr » (<https://github.com/zumbov2/deeplr>) et « tidyverse » (Wickham et al., 2019), utilisant

R (R Core Team, 2024). Tout d'abord, les données textuelles de chaque carte source ont été concaténées en un seul attribut. Elles ont ensuite été transformées en jetons de mots écrits en lettres minuscules (Silge et Robinson, 2016). Au cours de ce processus de création des jetons, les nombres, la ponctuation, les espaces et les jetons comptant moins de trois caractères ont été éliminés (Lawley et al., 2023). Les jetons uniques ont alors été extraits et traduits en anglais à l'aide de l'interface de programmation d'applications de DeepL (<https://www.deepl.com>). Les textes ainsi produits ont été manuellement inspectés pour trouver les erreurs introduites pendant le processus de traduction et créer une liste des jetons non informatifs (mots vides). De nombreux mots vides correspondaient à des codes de cartographie des roches employés dans les ensembles de données sources. Les noms de personnes, de lieux et d'animaux ont été ajoutés à la liste des mots vides pour orienter la recherche subséquente par traitement du langage naturel sur l'information géoscientifique. Une liste standard de mots vides en anglais ($n = 1\,149$) a également été utilisée pour éliminer les jetons non informatifs des analyses ultérieures (p. ex. *he*, *she*, *they*; Silge et Robinson, 2016). Les jetons traduits, corrigés et épurés ont ensuite été concaténés pour chaque PID (« JETON_ANGLAIS »). En tout, le corpus de la CCGC comprend 5 141 215 jetons, dont 4 430 sont uniques. Les données textuelles manquantes sont indiquées comme « inconnues ». Les jetons de mots sont inscrits en minuscules et sont séparés par un point-virgule.

4.0 Cartes géologique et géochronologique du Canada

Toutes les opérations spatiales ont été faites utilisant R à l'aide des paquets « sf » (Pebesma, 2018), « terra » (Hijmans et al., 2022), « fasterize » (<https://github.com/ecohealthalliance/fasterize>) et « tidyverse » (Wickham et al., 2019). Des machines virtuelles ont été utilisées pour les opérations requérant beaucoup de mémoire, généralement à l'aide des instances EC2 M5 disponibles auprès d'Amazon Web Services (<https://aws.amazon.com>) et les plateformes Posit Workbench et SageMaker (Joshi, 2020). L'indexation spatiale est basée sur le système hiérarchique *S2 discrete global grid system* (DGGS) de Google, qui divise la planète en quadrilatères avec résolutions spatiales variables (<http://s2geometry.io>). Le CGMC utilise la résolution 15 du S2 DGGS, qui correspond à six milliards de quadrilatères au niveau mondial. À cette résolution, chaque quadrilatère correspond approximativement à une longueur d'arête de 290 m et à une superficie de 79 173 m². (http://s2geometry.io/resources/s2cell_statistics). La division du Canada en cellules S2 a été réalisée à partir du polygone simplifié du littoral national à échelle de 1/10 000 000 téléchargé depuis Natural Earth (<https://www.naturalearthdata.com>). Environ 121 194 444 cellules S2 couvrent la masse continentale canadienne à la résolution 15, bien que ce nombre puisse varier selon le niveau de détail du littoral ayant servi à calculer les limites terrestres de la zone cartographiée. Le processus d'indexation spatiale pourrait être répété à des résolutions supérieures (quelques millimètres) ou inférieures (des milliers de kilomètres) en fonction de l'utilisation visée et des ressources informatiques disponibles. La même méthode a été employée pour indexer spatialement les 32 cartes géologiques du substratum rocheux utilisées comme données sources. La « meilleure » carte source spatialement indexée a ensuite été extraite pour chaque cellule S2 en suivant la démarche heuristique suivante : 1) élimination des polygones pour lesquels l'incertitude de la classification lithologique était de type « permissif ». Cette étape a permis de supprimer les polygones dont la lithologie a été inférée à partir de données géophysiques et se trouvant enfouis sous couverture phanérozoïque au Manitoba; 2) classement en ordre décroissant du rang lithologique (autre_inconnu = 1; autre_superficiel = 1; autre_faille = 2, autre_hydrothermal = 2, autre_mélange = 2; tous les autres = 3). Cette étape de tri

diminue la probabilité que des zones non cartographiées soient comprises dans la carte géologique choisie; 3) classement en ordre décroissant des attributs LITHOLOGIE_CLASSIFICATION_INCERTITUDE afin de prioriser les polygones présentant un faible niveau d'incertitude; 4) classement en ordre croissant des attributs SOURCE_ECHELLE_MAXIMUM pour donner priorité aux cartes sources présentant le plus de détails géologiques; 5) classement en ordre décroissant des attributs SOURCE_ANNÉE afin de favoriser l'information la plus récente; et 6) classement en ordre alphabétique des attributs LITHOLOGIE_CLASSIFICATION. Cette démarche visait à trouver la meilleure information géoscientifique disponible pour chaque cellule S2 et a servi à dresser la carte géologique assemblée du Canada (figure 2). La même méthode a été utilisée pour créer la nouvelle carte géochronologique du Canada, sauf qu'une étape supplémentaire a été nécessaire pour éliminer les données sources qui n'avaient aucune information d'âge avant de trier les cellules S2 (figures 2 et 3). Aux fins de la visualisation, les cartes géologique et géochronologique du Canada ont été rastérisées en une grille carrée de 200 m sur 200 m et exportées dans un format de fichier GeoTIFF. Les attributs de couleur de chacun de ces fichiers sont fournis dans le dossier « ccgc_geotiff ».

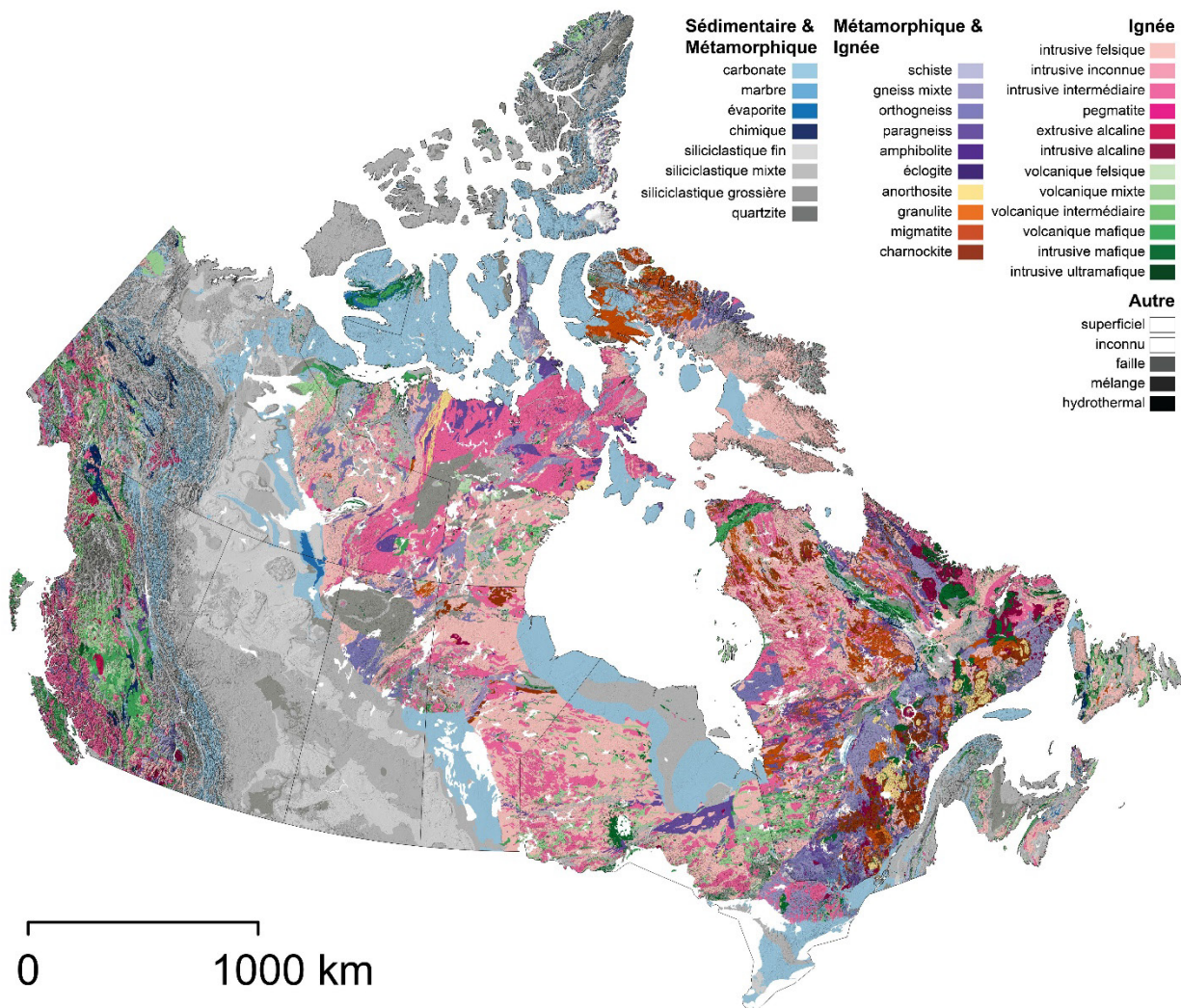
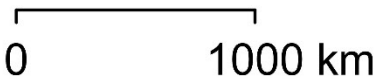


Figure 2 : Nouvelle carte géologique du Canada basée sur la meilleure information géoscientifique extraite de la CCGC et superposée à une carte matricielle du relief par ombres portées.



5.0 Conclusion

15

ponctuelles et linéaires, documents supplémentaires et images). Les commissions géologiques provinciales et territoriales fournissent les renseignements géoscientifiques les plus récents sur le substratum rocheux de leur région respective. La CCGC est un produit scientifique destiné à l'évaluation du substratum du potentiel minier, des eaux souterraines, des géorisques, de la géothermie et de l'environnement à l'échelle nationale. Au fil du temps, de nouvelles cartes géologiques du substratum rocheux à plus haute résolution pour des endroits prioritaires seront ajoutées à la CCGC afin d'améliorer la performance des applications d'apprentissage machine subséquentes.

6.0 Remerciements

Ces travaux ont été réalisés dans le cadre de l'initiative Géoscience et données pour les minéraux critiques (GDMC). Toutes les données comprises dans cette compilation ont été tirées de cartes géologiques publiées antérieurement par les commissions géologiques provinciales, territoriales et autres organismes. Les lecteurs sont priés de consulter les données sources pour obtenir de plus amples renseignements. La révision et les commentaires de Steven Zhang, d'Alex Zagorevski, Rob Rainbird et de Valérie Bécu ont amélioré ce rapport. Nous remercions particulièrement l'équipe d'Orix Geoscience pour sa collaboration et son soutien. Nous remercions également le Bureau de la traduction.

7.0 Références

- Bissell, H. J., & Chilingar, G. V. (1967). Chapter 4 Classification of Sedimentary Carbonate Rocks. In G. V. Chilingar, H. J. Bissell, & R. W. Fairbridge (Eds.), *Developments in Sedimentology* (Vol. 9, pp. 87–168). Elsevier. [https://doi.org/10.1016/S0070-4571\(08\)71112-9](https://doi.org/10.1016/S0070-4571(08)71112-9)
- Canada Bureau of Geology and Topography. (1945). Geological Map of the Dominion of Canada. Geological Survey of Canada, "A" Series Map 820A, 2. <https://doi.org/10.4095/106719>
- Cas, R. A. F., & Wright, J. V. (1988). Classification of modern and ancient volcanoclastic rocks of pyroclastic and epiclastic origins. In R. A. F. Cas & J. V. Wright (Eds.), *Volcanic Successions Modern and Ancient: A geological approach to processes, products and successions* (pp. 348–361). Springer Netherlands. https://doi.org/10.1007/978-94-009-3167-1_12
- Cohen, K. M., Finney, S. C., Gibbard, P. L., & Fan, J.-X. (2013). The ICS international chronostratigraphic chart. *Episodes Journal of International Geoscience*, 36(3), 199–204.
- Crook, K. A. W. (1960). Classification of arenites. *American Journal of Science*, 258(6), 419–428. <https://doi.org/10.2475/ajs.258.6.419>
- Dickinson, W. R. (1970). Interpreting detrital modes of graywacke and arkose. *Journal of Sedimentary Research*, 40(2), 695–707. <https://doi.org/10.1306/74D72018-2B21-11D7-8648000102C1865D>
- Dott, R. H. (1964). Wacke, graywacke and matrix; what approach to immature sandstone classification? *Journal of Sedimentary Research*, 34(3), 625–632. <https://doi.org/10.1306/74D71109-2B21-11D7-8648000102C1865D>

- Douglas, R. (1969). Geological map of Canada. Geological Survey of Canada "A" Series Map 1250A, 1. <https://doi.org/10.4095/108937>
- Dunham, R. J. (1962). Classification of Carbonate Rocks According to Depositional Texture. In W. E. Ham (Ed.), *Classification of Carbonate Rocks—A Symposium* (Vol. 1, pp. 108–121). American Association of Petroleum Geologists. <https://doi.org/10.1306/M1357>
- Festa, A., Pini, G.A., Dilek, Y., & Codegone, G. (2010). Mélanges and mélange-forming processes: A historical overview and new concepts. *International Geology Review*, 52, 1040–1105. <https://doi.org/10.1080/00206810903557704>
- Fisher, R. V. (1961). Proposed classification of volcanoclastic sediments and rocks. *Geological Society of America Bulletin*, 72(9), 1409–1414.
- Fisher, R. V. (1966). Rocks composed of volcanic fragments and their classification. *Earth-Science Reviews*, 1(4), 287–298.
- Folk, R. L. (1954). The Distinction between Grain Size and Mineral Composition in Sedimentary-Rock Nomenclature. *The Journal of Geology*, 62(4), 344–359. <https://doi.org/10.1086/626171>
- Folk, R. L. (1959). Practical Petrographic Classification of Limestones1. *American Association of Petroleum Geologists Bulletin*, 43(1), 1–38. <https://doi.org/10.1306/0BDA5C36-16BD-11D7-8645000102C1865D>
- Folk, R. L. (1980). *Petrology of sedimentary rocks*. Hemphill Publishing Company. <http://hdl.handle.net/2152/22930>
- Frolova, Yu. V. (2008). Specific features in the composition, structure, and properties of volcanoclastic rocks. *Moscow University Geology Bulletin*, 63(1), 28–37. <https://doi.org/10.1007/s11969-008-1004-z>
- Garzanti, E. (2019). Petrographic classification of sand and sandstone. *Earth-Science Reviews*, 192, 545–563. <https://doi.org/10.1016/j.earscirev.2018.12.014>
- Geological Survey of Canada. (1884). Map of the Dominion of Canada, geologically colored from surveys made by the Geological Corps, 1842 to 1882. Geological Survey of Canada, Multicoloured Geological Map 411, 2.
- Geological Survey of Canada. (1955). Geological map of Canada. Geological Survey of Canada, "A" Series Map 1045A, 1. <https://doi.org/10.4095/216740>
- Geological Survey of Canada. (1962). Geological map of Canada. Geological Survey of Canada, "A" Series Map 1045A, (Ed. Rev.), 1. <https://doi.org/10.4095/106724>
- Gillespie, M. R., & Leslie, A. G. (2021). BRUCS: A new system for classifying and naming mappable rock units. *Journal of the Geological Society*, 178(4), 18. <https://doi.org/10.1144/jgs2020-212>
- Harrison, J., St-Onge, M., Petrov, O., Strelnikov, Lopatin, B., Wilson, F., Tella, S., Paul, D., Lynds, T., Shokalsky, S., Hults, C., Bergman, S., Jepsen, H., & Solli, A. (2011). Geological map of the Arctic / Carte géologique de l'Arctique. Geological Survey of Canada, "A" Series Map 2159A, 9. <https://doi.org/10.4095/287868>
- Hattin, D. E. (1991). Lithodemes, suites, supersuites, and complexes: Intrusive, metamorphic, and genetically mixed assemblages of rocks now embraced by North American Code of Stratigraphic Nomenclature. *Precambrian Research*, 50(3), 355–357. [https://doi.org/10.1016/0301-9268\(91\)90029-A](https://doi.org/10.1016/0301-9268(91)90029-A)

- Higgins, M. W. (1971). Cataclastic rocks. U.S. Geological Survey Professional Paper 687, 687, 97. <https://doi.org/10.3133/pp687>
- Hijmans, R. J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., & Sumner, M. D. (2022). Package 'terra.' The Comprehensive R Archive Network, 340.
- Hoffman, P. F. (2015). The Tooth of Time: JO Wheeler. *Geoscience Canada*, 42(4), 373–382.
- ISSC. (2000). International subcommission on stratigraphic classification of IUGS International Commission on Stratigraphy: International stratigraphic guide—an abridged version. *GeoArabia*, 5(2), 231–266. <https://doi.org/10.2113/geoarabia0502231>
- Joshi, A. V. (2020). Amazon's machine learning toolkit: Sagemaker. In *Machine learning and artificial intelligence* (pp. 233–243). Springer. https://doi.org/10.1007/978-3-030-26622-6_24
- Killick, A. M. (2003). Fault Rock Classification: An aid to structural interpretation in mine and exploration geology. *South African Journal of Geology*, 106(4), 395–402. <https://doi.org/10.2113/106.4.395>
- Klein, G. de V. (1963). Analysis and review of sandstone classifications in the North American geological literature, 1940–1960. *Geological Society of America Bulletin*, 74(5), 555–576. [https://doi.org/10.1130/0016-7606\(1963\)74\[555:AAROSC\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1963)74[555:AAROSC]2.0.CO;2)
- Lawley, C. J. M., Gadd, M. G., Parsa, M., Lederer, G. W., Graham, G. E., & Ford, A. (2023). Applications of Natural Language Processing to Geoscience Text Data and Prospectivity Modeling. *Natural Resources Research*, 32(4), 1503–1527. <https://doi.org/10.1007/s11053-023-10216-1>
- Laxton, J. L. (2017). Geological map fusion: OneGeology-Europe and INSPIRE. Geological Society, London, Special Publications, 408(1), 147–160. <https://doi.org/10.1144/SP408.16>
- Lazar, O. R., Bohacs, K. M., Macquaker, J. H. S., Schieber, J., & Demko, T. M. (2015). Capturing Key Attributes of Fine-Grained Sedimentary Rocks In Outcrops, Cores, and Thin Sections: Nomenclature and Description Guidelines. *Journal of Sedimentary Research*, 85(3), 230–246. <https://doi.org/10.2110/jsr.2015.11>
- Le Maitre, R. W., Streckeisen, A., Zanettin, B., Le Bas, M. J., Bonin, B., & Bateman, P. (Eds.). (2002). *Igneous Rocks: A Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511535581>
- Logan, W. (1864). Geological map of Canada and the adjacent regions including parts of other British provinces and of the United States. Geological Survey of Canada, Multicoloured Geological Map 53, 1. <https://doi.org/10.4095/133901>
- Logan, W. (1869). Geological map of Canada and the adjacent regions, including parts of other British provinces and of the United States. Geological Survey of Canada, Multicoloured Geological Map No. 65, 2. <https://doi.org/10.4095/220748>
- Lundegard, P. D., & Samuels, N. D. (1980). Field classification of fine-grained sedimentary rocks. *Journal of Sedimentary Research*, 50(3), 781–786. <https://doi.org/10.1306/212F7AE5-2B24-11D7-8648000102C1865D>
- Macquaker, J. H. S., & Adams, A. E. (2003). Maximizing Information from Fine-Grained Sedimentary Rocks: An Inclusive Nomenclature for Mudstones. *Journal of Sedimentary Research*, 73(5), 735–744. <https://doi.org/10.1306/012203730735>

- Murphy, M. A., & Salvador, A. (1999). International stratigraphic guide—An abridged version. *Episodes Journal of International Geoscience*, 22(4), 255–271. <https://doi.org/10.18814/epiiugs/1999/v22i4/002>
- NASC. (1983). North American Stratigraphic Code. *American Association of Petroleum Geologists Bulletin*, 67, 841–875.
- NASC. (2005). North American Stratigraphic Code. *American Association of Petroleum Geologists Bulletin*, 89(11), 1547–1591. <https://doi.org/10.1306/07050504129>
- NASC. (2021). North American Stratigraphic Code. *Stratigraphy*, 18, 153–204. <https://doi.org/10.29041/strat.18.3.01>
- Okada, H. (1971). Classification of Sandstone: Analysis and Proposal. *The Journal of Geology*, 79(5), 509–525. <https://doi.org/10.1086/627673>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pettijohn, F. J. (1954). Classification of Sandstones. *The Journal of Geology*, 62(4), 360–365. <https://doi.org/10.1086/626172>
- Picard, M. D. (1971). Classification of fine-grained sedimentary rocks. *Journal of Sedimentary Research*, 41(1), 179–195. <https://doi.org/10.1306/74D7221B-2B21-11D7-8648000102C1865D>
- R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reed, J. C., Jr., Wheeler, J. O., Tucholke, B. E., Stettner, W. R., & Soller, D. R. (2005). Decade of North American Geology Geologic Map of North America—Perspectives and explanation. In J. C. Reed Jr., J. O. Wheeler, B. E. Tucholke, W. R. Stettner, & D. R. Soller (Eds.), *Decade of North American Geology Geologic Map of North America—Perspectives and explanation* (Vol. 1, pp. 1–28). Geological Society of America. <https://doi.org/10.1130/DNAG-CSMS-v1.1>
- Rice, H. (1955). Geological Map of Canada. Geological Survey of Canada Map 1045A, Scale 1:7,603,200 (1 in to 120 Mi), 1.
- Salvador, A. (1994). *International Stratigraphic Guide: A Guide to Stratigraphic Classification, Terminology, and Procedure*. Geological Society of America.
- Schmid, R., Fettes, D., Harte, B., Davis, E., & Desmons, J. (2007). How to name a metamorphic rock. *Metamorphic Rocks: A Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Metamorphic Rocks*. Cambridge University Press, Cambridge, 3–15.
- Sen, M., & Duffy, T. (2005). GeoSciML: Development of a generic GeoScience Markup Language. *Computers & Geosciences*, 31(9), 1095–1103. <https://doi.org/10.1016/j.cageo.2004.12.003>
- Shepard, F. P. (1954). Nomenclature based on sand-silt-clay ratios. *Journal of Sedimentary Research*, 24(3), 151–158. <https://doi.org/10.1306/D4269774-2B26-11D7-8648000102C1865D>
- Sibson, R. H. (1977). Fault rocks and fault mechanisms. *Journal of the Geological Society*, 133(3), 191–213. <https://doi.org/10.1144/gsjgs.133.3.0191>
- Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, 1(3), 37. <https://doi.org/doi:10.21105/joss.00037>

- Spears, D. A. (1980). Towards a classification of shales. *Journal of the Geological Society*, 137(2), 125–129. <https://doi.org/10.1144/gsjgs.137.2.0125>
- Stow, D. A. (1981). Fine-grained sediments: Terminology. *Quarterly Journal of Engineering Geology*, 14(4), 243–244. <https://doi.org/10.1144/GSL.QJEG.1981.014.04.02>
- Streckeisen, A. (1976). To each plutonic rock its proper name. *Earth-Science Reviews*, 12(1), 1–33. [https://doi.org/10.1016/0012-8252\(76\)90052-0](https://doi.org/10.1016/0012-8252(76)90052-0)
- Tikoff, B. (2023). Improving the Practice of Geology through Explicit Inclusion of Scientific Uncertainty for Data and Models. *GSA Today*, 33(7). <https://doi.org/10.1130/GSATG560A.1>
- Van Hise, C. (1896). Principles of North American geology. U.S. Geological Survey Annual Report, 16, 571–843.
- Wheeler, J., Hoffman, P., Card, K., Davidson, A., Sanford, B., Okulitch, A., & Roest, W. (1996). Geological map of Canada / Carte géologique du Canada. Geological Survey of Canada, “A” Series Map 1860A, 3; 1 CD-ROM. <https://doi.org/10.4095/208175>
- White, J. D. L., & Houghton, B. F. (2006). Primary volcanoclastic rocks. *Geology*, 34(8), 677–680. <https://doi.org/10.1130/G22346.1>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Willis, B. (1906). Geologic Map of North America. 10th International Geological Congress, Mexico City, Mexico, Scale 1:5,000,000, 1.
- Wolf, K. H. (1960). Simplified Limestone Classification. *American Association of Petroleum Geologists Bulletin*, 44(8), 1414–1416. <https://doi.org/10.1306/0BDA61B5-16BD-11D7-8645000102C1865D>
- Woodcock, N. H., & Mort, K. (2008). Classification of fault breccias and related fault rocks. *Geological Magazine*, 145(3), 435–440. <https://doi.org/10.1017/S0016756808004883>
- Wright, V. P. (1992). A revised classification of limestones. *Sedimentary Geology*, 76(3), 177–185. [https://doi.org/10.1016/0037-0738\(92\)90082-3](https://doi.org/10.1016/0037-0738(92)90082-3)
- Young, G. (1913). Geological Map of the Dominion of Canada, and Newfoundland. Geological Survey of Canada Map 91A, 1. <https://doi.org/10.4095/106720>
- Young, G. (1924). Geological map of the Dominion of Canada and Newfoundland. Geological Survey of Canada, Multicoloured Geological Map No. 1277, 1. <https://doi.org/10.4095/223833>