**Généralisation cartographique des étendues et des cours d’eau du 1:50000 au 1:250000**

**Michel Pothier**

**Septembre 2018**

[***https://github.com/MichelPothier/GeneralisationCartographique***](https://github.com/MichelPothier/GeneralisationCartographique)

Table des matières

[Introduction 3](#_Toc523917963)

[Données spatiales à Ressources Naturelles Canada (RNCAN) 4](#_Toc523917964)

[Découpage cartographique par entité 4](#_Toc523917965)

[Cartes topographiques 6](#_Toc523917966)

[Catalogues des données spatiales 7](#_Toc523917967)

[Base de Données Géographique (BDG) 8](#_Toc523917968)

[Représentation cartographique 9](#_Toc523917969)

[Classe d’entité 9](#_Toc523917970)

[Élément 10](#_Toc523917971)

[Identifiant unique 10](#_Toc523917972)

[Code spécifique 10](#_Toc523917973)

[Nom du jeu de données 10](#_Toc523917974)

[Attribut 10](#_Toc523917975)

[Géométrie 11](#_Toc523917976)

[Statistiques des données spatiales 15](#_Toc523917977)

[Modèle de données 16](#_Toc523917978)

[Gestion des versions 18](#_Toc523917979)

[Échelle de représentation des données spatiales 19](#_Toc523917980)

[Référence spatiale 19](#_Toc523917981)

[Précision et résolution de la référence spatiale 20](#_Toc523917982)

[Référence spatiale géographique 20](#_Toc523917983)

[Référence spatiale projetée 21](#_Toc523917984)

[Données spatiales généralisées 22](#_Toc523917985)

[Étendues d’eau 22](#_Toc523917986)

[Cours d’eau 23](#_Toc523917987)

[Contrainte d’intégrité 23](#_Toc523917988)

[Contrainte d’intégrité attributive 24](#_Toc523917989)

[Table des contraintes attributives 24](#_Toc523917990)

[Outil de validation des contraintes attributives 25](#_Toc523917991)

[Requêtes attributives 26](#_Toc523917992)

[Méthodes de traitement 28](#_Toc523917993)

[Rapport d’erreurs 29](#_Toc523917994)

[Contrainte d’intégrité spatiale 31](#_Toc523917995)

[Table des contraintes spatiales 31](#_Toc523917996)

[Outil de validation des contraintes spatiales 32](#_Toc523917997)

[Requêtes spatiales 33](#_Toc523917998)

[Méthodes de traitement 34](#_Toc523917999)

[Rapport d’erreurs 35](#_Toc523918000)

[Table des statistiques d’erreurs 36](#_Toc523918001)

[Cartes des statistiques d’erreurs 37](#_Toc523918002)

[Topologie 39](#_Toc523918003)

[Tolérance minimale des données spatiales (Précision) 39](#_Toc523918004)

[Nœud (Node) 40](#_Toc523918005)

[Limite (Edge) 40](#_Toc523918006)

[Face (Face) 41](#_Toc523918007)

[Relations spatiales 42](#_Toc523918008)

[Précision de la référence spatiale 42](#_Toc523918009)

[Relations spatiales prédéfinies 42](#_Toc523918010)

[Relations spatiales non définies / Masque à neuf intersections (Egenhofer et/ou Clementini) 42](#_Toc523918011)

[Opérateurs spatiaux (Topologique) 43](#_Toc523918012)

[Précision de la référence spatiale 44](#_Toc523918013)

[Opérateurs spatiaux prédéfinis (implique deux géométries) 44](#_Toc523918014)

[Autres opérateurs spatiaux (implique une seule géométrie) 44](#_Toc523918015)

[Dimensions minimales 44](#_Toc523918016)

[Proximité des données spatiales (Topologie) 45](#_Toc523918017)

[Distance latérale minimale entre les sommets (Douglass-Peuker) 46](#_Toc523918018)

[Longueur minimale d’une droite 47](#_Toc523918019)

[Longueur maximale d’une droite (Distance pour densifier) 48](#_Toc523918020)

[Longueur minimale d’une ligne 48](#_Toc523918021)

[Longueur minimale d’une Polyligne fermée 49](#_Toc523918022)

[Largeur minimale de généralisation 50](#_Toc523918023)

[Longueur minimale de généralisation 51](#_Toc523918024)

[Superficie minimale d’un anneau extérieur 52](#_Toc523918025)

[Superficie minimale d’un anneau intérieur 52](#_Toc523918026)

[Squelettisation 53](#_Toc523918027)

[Géométries de type ligne (Polyline) et Surface (Polygon) 54](#_Toc523918028)

[Points de connexion avec les éléments en relation 54](#_Toc523918029)

[Enveloppe des géométries et densification 55](#_Toc523918030)

[Triangulation de Delaunay 56](#_Toc523918031)

[Squelettisation à droite d’une Polyligne et à l’intérieur d’un Polygone 57](#_Toc523918032)

[Sommets manquants 57](#_Toc523918033)

[Droites des triangles de Delaunay 58](#_Toc523918034)

[Droites significatives des triangles de Delaunay 59](#_Toc523918035)

[Squelette primaire 59](#_Toc523918036)

[Squelette de base 60](#_Toc523918037)

[Squelette de base significatif 60](#_Toc523918038)

[Squelette minimal 61](#_Toc523918039)

[Squelette minimal avec points de connexion 62](#_Toc523918040)

[Squelettisation à gauche d’une Polyligne et à l’extérieure d’un Polygone 62](#_Toc523918041)

[Sommets manquants 63](#_Toc523918042)

[Droites des triangles de Delaunay 63](#_Toc523918043)

[Droites significatives des triangles de Delaunay 64](#_Toc523918044)

[Squelette primaire 64](#_Toc523918045)

[Squelette de base 65](#_Toc523918046)

[Squelette de base significatif 66](#_Toc523918047)

[Squelette minimal 66](#_Toc523918048)

[Généralisation 67](#_Toc523918049)

[Largeur et longueur de généralisation d’un Polygone 67](#_Toc523918050)

[Prétraitement 68](#_Toc523918051)

[Traitement pour l’Intérieur et l’extérieur d’un Polygone 68](#_Toc523918052)

[Points de connexion 69](#_Toc523918053)

[Droites des triangles de Delaunay 69](#_Toc523918054)

[Largeur de généralisation 70](#_Toc523918055)

[Longueur de généralisation 70](#_Toc523918056)

[Détection des erreurs 71](#_Toc523918057)

[Filtrage des anneaux 72](#_Toc523918058)

[Polygone généralisé 73](#_Toc523918059)

[Squelette du Polygone généralisé 73](#_Toc523918060)

[Lignes d’erreurs de largeur et longueur de généralisation 74](#_Toc523918061)

[Traitements complémentaires 75](#_Toc523918062)

[Largeur et longueur de généralisation entre les Polygones 75](#_Toc523918063)

[Largeur et longueur de généralisation d’une Polyligne 76](#_Toc523918064)

[Prétraitement 76](#_Toc523918065)

[Traitement à droite et à gauche d’une Polyligne 77](#_Toc523918066)

[Points de connexion 77](#_Toc523918067)

[Droites des triangles de Delaunay 78](#_Toc523918068)

[Largeur de généralisation 78](#_Toc523918069)

[Longueur de généralisation 79](#_Toc523918070)

[Détection des erreurs 80](#_Toc523918071)

[Filtrage des lignes 81](#_Toc523918072)

[Polyligne généralisée 81](#_Toc523918073)

[Squelette de la Polyligne généralisée 82](#_Toc523918074)

[Lignes d’erreurs de largeur et longueur de généralisation 83](#_Toc523918075)

[Traitements complémentaires 83](#_Toc523918076)

[Largeur et longueur de généralisation entre les Polylignes 84](#_Toc523918077)

[Différences dans les résultats 85](#_Toc523918078)

[Largeur et longueur de généralisation d’une Polyligne fractionnée 85](#_Toc523918079)

[Fractionner une Polyligne 85](#_Toc523918080)

[Droites des triangles de Delaunay 86](#_Toc523918081)

[Lignes d’erreurs 86](#_Toc523918082)

[Polyligne fractionnée et généralisée 87](#_Toc523918083)

[Processus de généralisation des étendues et des cours d’eau 88](#_Toc523918084)

[1. Extraction des données spatiales 88](#_Toc523918085)

[2. Projection des données spatiales 88](#_Toc523918086)

[3. Corriger la proximité des sommets 89](#_Toc523918087)

[4. Éliminer les lignes superposées 89](#_Toc523918088)

[5. Filtrer les sommets des lignes et des surfaces 89](#_Toc523918089)

[6. Filtrer les droites des lignes et des surfaces 89](#_Toc523918090)

[7. Éliminer les squelettes existants dans les surfaces 89](#_Toc523918091)

[8. Fusion des lignes et des surfaces adjacentes 89](#_Toc523918092)

[9. Filtrer les anneaux intérieurs 90](#_Toc523918093)

[10. Adoucir les lignes et les limites des surfaces 90](#_Toc523918094)

[11. Généraliser l’intérieure et l’extérieure des surfaces 90](#_Toc523918095)

[12. Filtrer les lignes 90](#_Toc523918096)

[13. Filtrer les lignes fermées 90](#_Toc523918097)

[14. Généraliser la partie gauche et la partie droite des lignes 90](#_Toc523918098)

[15. Créer les squelettes des surfaces 90](#_Toc523918099)

[16. Séparer les géométries multiples 91](#_Toc523918100)

[17. Valider les géométries des éléments 91](#_Toc523918101)

[18. Transfert et destruction des attributs d’origine 91](#_Toc523918102)

[Conclusion 92](#_Toc523918103)

# Introduction

La généralisation des données spatiales est un sujet connu en géomatique mais pas encore contrôlée parfaitement. Beaucoup d’organisations possédant et gérant des données spatiales s’y intéresse puisqu’elle permettrait de diminuer de façon importante les coûts d’acquisition des données. Plusieurs techniques et méthodes comme celles présentées dans ce document sont utilisées pour y parvenir. De façon générale, la plupart s’entende pour dire que ce sont des traitements qui permettent de diminuer le niveau de détails des données spatiales. Un processus de généralisation contenant une séquence de ces traitements doit être défini et exécuté selon les besoins recherchés.

Chaque organisation possède des besoins particuliers pour représenter leurs phénomènes cartographiques dans leurs données spatiales. Ces données sont caractérisées par plusieurs paramètres distincts comme la référence spatiale, l’échelle de représentation, les types d’attributs et de géométries, les contraintes spatiales et attributives, etc. Ressources Naturelles Canada (RNCAN) est une organisation publique qui possède plusieurs types de données spatiales afin de satisfaire les besoins de tous leurs clients. Comme on le sait, en raison des coupures budgétaires, des coûts d’acquisition et de mise à jour de ces données, plusieurs stratégies comme celle de la généralisation peuvent être utilisées. La Base de Données Géographique (BDG) est un des produits de RNCAN. Elle contient des données spatiales pour l’ensemble du Canada représentées normalement selon une échelle du 1:50000. Cependant à certains endroits, elle peut contenir des données à l’échelle du 1:20000. On connait assez bien l’état des données de la BDG puisque plusieurs contraintes d’intégrité spatiales et attributives sont utilisées à tous les jours pour les valider. Puisque les étendues et cours d’eau de la BDG sont utilisés pour expliquer les traitements de généralisation, une description détaillée de ces deux classes est présente dans ce document.

La généralisation étant déjà assez complexe à expliquer et à comprendre, seules les traitements pour généraliser les données de la BDG correspondant aux étendues et aux cours d’eau seront détaillés dans ce document. On fera référence à certains concepts comme la topologie, les relations spatiales, les opérateurs spatiaux, les dimensions minimales, la triangulation de Delaunay et à la squelettisation. Tous ces concepts sont nécessaires dans les différents traitements utilisés pour effectuer la généralisation des étendues et des cours d’eau. Chacun de ces concepts sera expliqué afin de bien comprendre son utilité dans l’exécution des traitements de généralisation. Afin de minimiser les coûts et les temps de développement, la technologie ESRI (ArcGis 10.5) a été utilisé comme infrastructure pour les différents outils de traitement de la généralisation.

L’emphase des explications contenues dans ce document se situe surtout pour deux traitements importants, soit celui de squelettisation des géométries de type ligne et surface et celui de largeur et longueur de généralisation des géométries de type ligne et surface. D’autres traitements sont présents et importants mais sont moins difficiles à expliquer. Évidemment, on terminera en proposant et en expliquant une séquence des traitements nécessaire dans un processus de généralisation. À noter que les explications contenues dans ce document pourraient tout aussi bien être valident pour effectuer une généralisation d’une échelle plus grande à une échelle plus petite autre que celle du 1:50000 au 1:250000.

# Données spatiales à Ressources Naturelles Canada (RNCAN)

Un des mandats du Ministères des Ressources Naturelles du Canada (**RNCAN**) est la gestion des données spatiales pour l’ensemble du Territoire Canadien. Plusieurs produits cartographiques sont disponibles sous différentes formes, soit papier, numérique, matricielle, vectorielle et attributive et sous différents découpages cartographiques par entité.

Les cartes topographiques papiers au 1:50000 et au 1:250000 ont été très longtemps les produits les plus importants. Ils sont encore assez en demande aujourd’hui. Avec l’arrivée des nouvelles technologies, les données numériques ont pris beaucoup de place. Les cartes topographiques papiers ont été numérisées pour former le premier produit matriciel nommé « **CanMatrix** » dans des fichiers en format **GeoTiff** et **PDF**.

La **BNDT** (Base Nationale des Données Topographiques) au 1:50000 et au 1:250000 a été le premier produit numérique vectoriel d’importance. Il est disponible dans des fichiers en format **CCOG** et **IFF**. L’utilisation des **BD** (bases de données) et des **SIG** (Système d’information géographique) ont apporté beaucoup de changements dans notre façon de faire. On peut maintenant emmagasiner les données d’un produit pour l’ensemble du Canada dans une **BD** et effectuer différents types d’analyses. Trois bases de données vectorielles sont présentes à Sherbrooke, soit la **BDG** (Base de données Géographique), la **BCG** (Base de Connaissance Géospatiale) et la **Géobase**. Une nouvelle base de données nommé **GeoSys** est en cours de création afin de remplacer la **BDG**, **LBK** et **Géobase**. Différents produits sont générés à partir de ces bases de données. Le produit **CanTopo** est généré à partir de la **BDG** dans des fichiers sous différents formats matriciels. Les produits **CanVec, RRN** (Réseau Routier National) et **RFN** (Réseau Ferroviaire National) sont générés à partir de la **BCG** dans des fichiers sous différents formats vectoriels. Le produit **RHN** (Réseau Hydrographique National) est généré à partir de la **Géobase** dans des fichiers sous différents formats vectoriels.

Les **MNE** (Modèle Numérique d’Élévation) et les images satellites sont de plus en plus en demande surtout avec une grande précision. Ils sont emmagasinés dans des fichiers sous différents formats matriciels.

## Découpage cartographique par entité

Souvent nommé **zone de travail** ou **zone de traitement**, le **découpage** cartographique par entité est utilisé dans toutes les bases de données et dans tous les produits. Bien que restrictif mais nécessaire, le **découpage** permet de régler certains problèmes de limites et de performances. Par exemple, en raison du nombre de sommet, il est techniquement impossible de conserver un élément comme le fleuve St-Laurent dans une BD. Il est également impossible de fournir un fichier contenant toutes les cartes topographiques matricielles au 1:50000 car trop gros à traiter et à transférer. Le découpage apporte cependant d’autres problèmes comme la gestion de ce découpage lors des analyses, des validations et des corrections de données. On peut résumer en disant que plus il y de découpage, plus il y a de problèmes et moins il y de découpage, moins il y a de problèmes.

Les produits **CanMatrix**, **CanTopo** et **BNDT** sont découpés selon le **SNRC** (Système National de Référence Cartographique). La plupart des classes d’entité contenues dans la **BDG** et la **BCG** sont découpées aussi selon le **SNRC** au 1:50000.

|  |  |
| --- | --- |
| Découpage cartographique par entité –  Selon le SNRC au 1:50000 | Découpage cartographique par entité –  Selon le SNRC au 1:250000 |
|  |  |

Les données contenues dans la **Géobase** et le produit **RHN** sont découpées en Bassins/Versants. Ce découpage permet de limiter les problèmes qui sont liés au découpage.

|  |
| --- |
| Découpage cartographique par entité – Selon les limites des Bassins/Versants |
|  |

Par définition, certaines classes d’entités contenues dans la BDG sont découpées par province et selon le Canada. On peut aussi dire qu’il n’y a pas de découpage lorsque le découpage est le Canada.

|  |  |
| --- | --- |
| Découpage cartographique et par entité –  selon les limites des provinces | Découpage cartographique et par entité –  Selon la limite du Canada |
|  |  |

## Cartes topographiques

On peut obtenir deux types de carte topographique à RNCAN, soit celles à l’échelle du 1:50000 et celles à l’échelle du 1:250000. Chaque type de carte possède leurs propres spécifications et ont été créé indépendamment. Aucun traitement de généralisation n’a été effectué, seul le traitement de captage initial ou de mise à jour à partir d’une source de données (ex : Photos aériennes) différente a été utilisé pour chaque type de carte.

L’utilisation d’outils et de traitements de généralisation des données serait un atout appréciable afin de minimiser les coûts et les temps de traitement d’une carte topographique. La carte topographique du 1:250000 pourrait être générée automatiquement à partir des données de la carte topographique du 1:50000 et des outils et traitement de généralisation.

|  |  |
| --- | --- |
| Image de la carte topographique à l’échelle du 1:50000. | Image de la carte topographique à l’échelle du 1:250000. |
|  |  |

On peut appliquer le même principe pour les données contenues dans une base de données au 1:50000 et au 1:250000. La **BDG** contient normalement des données à l’échelle du 1:50000 puisqu’elle est dérivée initialement à partir des données de la **BNDT** au 1:50000. Cependant pour des raisons de coût d’acquisition, de traitements et de mise à jour, il n’y a pas de données au 1:250000 contenues dans une BD. Même que dans la **BDG**, beaucoup de données ont été acquise à partir des données des provinces qui sont environ à l’échelle du 1:20000. Aucun traitement de généralisation des données n’a été effectué dans la **BDG**. Une surabondance d’éléments et de sommets sont donc présents.

## Catalogues des données spatiales

La description des données spatiales contenues dans les bases de données et dans les fichiers des produits est conservée dans un catalogue de données spatiales. Les catalogues des données spatiales sont conservés dans une base de données des catalogues. On peut ainsi extraire l’information sur les classes d’entité, les attributs et certaines contraintes pour chaque catalogue correspondant à chaque base de données spatiales ou produits.

Selon les besoins, on peut extraire l’information des catalogues à partir d’outils ou applications spécifiques comme dans l’exemple ci-dessous. Puisque l’information est dans un base de données, des liens croisés peuvent aussi être effectués entre un catalogue et une base de données dans un SIG pour afficher l’information. Des outils pour générer les domaines des codes de valeurs à partir de la BD des catalogues peuvent également être utilisés.

|  |
| --- |
| Exemple d’un outil pour afficher l’information contenue dans un catalogue de données spatiales. |
|  |

|  |
| --- |
| Exemple d’un outil pour créer les domaines d’attributs à partir d’un catalogue de données spatiales. |
|  |

# Base de Données Géographique (BDG)

Au fil du temps, la **BDG** a été renommé **BDRS** (Base de Données à Référence Spatiale) mais son nom est encore utilisé aujourd’hui. Initialement seules les classes d’entité de type « **BDG** » étaient présentes mais d’autres classes d’entité ont été ajoutées comme celle de **type « Patrimoine, RHN, RRN, RFN et limites municipales »**. Comme mentionné précédemment, les données de la BDG sont conservées dans une base de données Oracle avec la cartouche spatiale **SDE** de **ESRI** (Géodatabase) et en utilisant des géométries binaires. Différentes échelles de représentation des données sont utilisées. Les classes d’entités de type « **Patrimoine** » proviennent essentiellement des données de la **BNDT**, elles sont donc à l’échelle du 1:50000. Les données de types « **BDG »** proviennent surtout des données des provinces qui sont environ à l’échelle du 1:20000. Les données de type « **RHN »** proviennent soit de la **BNDT** ou des provinces mais sont mises à jour à partir des images satellites. Aucun traitement de généralisation n’est effectué dans la **BDG** ce qui cause une surabondance d’éléments et de sommets en plus de rendre les données non homogènes.

À part les contraintes de base contenues dans une **BD**, aucune contrainte d’intégrité spatiale ou attributive n’était présente pour contrôler la qualité des données dans la **BDG**, la **BCG** ou la **Géobase**. Les contrôles de qualité se font normalement dans les différents processus de production ou de distribution des données. Depuis peu de temps et en raison de plusieurs problèmes répétitifs de non-conformité, une validation de plusieurs contraintes d’intégrités spatiales et attributives est maintenant effectuée quotidiennement dans la **BDG**. On peut ainsi connaître l’état des données de la **BDG** en tout temps ce qui n’est pas le cas pour la **BCG** et la **Géobase**.

## Représentation cartographique

Dans la BDG, aucune donnée matricielle n’est présente. On retrouve seulement des données spatiales et attributives dites vectorielles. Il y a évidemment un lien entre les données spatiales et attributives. Le but étant de représenter certains phénomènes terrestres, comme les étendues et les cours d’eau, sous forme d’entité cartographique (**classe d’entité**) dans une base de données spatiales afin de pouvoir visualiser, gérer et manipuler ces données. Ces données contiennent donc une intelligence que l’on veut bien leur donner selon l’information conservée dans la base de données.

### Classe d’entité

Chaque classe d’entité contenue dans une base de données représente un phénomène terrestre spécifique. Dans ArcGis, une classe d’entité est représentée par ce qu’il appelle « **FeatureClass** ». Par exemple, le phénomène des étendues d’eau dans la BDG est représenté par la classe d’entité nommée « **NHN\_ HHYD\_WATERBODY\_2** » de type surface. Dans ces classes d’entité, on retrouve différentes informations conservées sous forme d’attribut pour chaque occurrence (**élément**) de la classe d’entité. Dans ArcGis, chaque classe d’entité contient au moins une référence spatiale, un attribut de type **identifiant** unique et un attribut de type **géométrie**. Elle peut contenir aucun ou plusieurs **éléments** et aucun ou plusieurs autres **attributs** selon l’intelligence que l’on veut bien donner à ces classes d’entité. Sans être obligatoire dans ArcGis, on retrouve un attribut de type « **code spécifique** » permettant de différencier les différents type **d’élément**.

|  |
| --- |
| Information de la classe d’entité des étendues d’eau dans ArcGis (NHN\_HHYD\_WATERBODY\_2). |
|  |

### Élément

Un élément est une occurrence d’une classe d’entité nommé « **Feature** » dans ArcGis. Dans la BDG, chaque élément contient un attribut de type « **identifiant unique** », un attribut de type « **code spécifique** », un attribut de type « **géométrie** » et un attribut de type « **nom du jeu de données** ». Un nom de jeu de données correspond à un découpage d’entité. D’autres **attributs** de différents types sont présents selon les besoins pour chaque classe d’entité.

### Identifiant unique

L’attribut de type « **identifiant unique** » est utilisé, comme son nom le dit, pour identifier de façon unique un élément spécifique dans la classe d’entité. Dans ArcGis, cet attribut de type « **LONG INTEGER**» est normalement nommé « **OBJECTID** ».

### Code spécifique

L’attribut de type « **code spécifique** » est utilisé pour différencier différents type d’élément. Par exemple dans la BDG, un élément de la classe des étendues d’eau représentant un lac permanent contient la valeur **1481312 (Lac, permanent)** dans l’attribut nommé « **CODE\_SPEC** ». Dans ArcGis, cet attribut de type « **LONG INTEGER**» est normalement nommé « **SUBTYPE** ».

### Nom du jeu de données

L’attribut de type « **nom du jeu de données** » est utilisé pour le découpage des éléments des classes d’entité. Par exemple dans la BDG, tous les éléments de la classe des étendues et des cours d’eau sont découpés selon le SNRC à l’échelle du 1:50000. Ainsi lorsqu’un élément est situé à l’intérieur du jeu de données nommé ‘**021L14’**, on retrouvera cette valeur dans l’attribut nommé « **DATASET\_NAME** ».

### Attribut

Les autres attributs présents dans chaque classe d’entité dépendent des besoins pour ces dernières. Ils peuvent être de type « entier, réelle, texte ou autres ». Par exemple dans la BDG, on retrouve des attributs pour les informations suivantes : date de validité, précision des données, provenance des données, etc.

|  |  |
| --- | --- |
| Affichage des attributs d’un élément de la classe d’entité des étendues d’eau : NHN\_HHYD\_WATERBODY\_2. | Affichage des attributs d’un élément de la classe d’entité des cours d’eau : NHN\_HNET\_NETWORK\_LINEAR\_FLOW\_1. |
|  |  |

### Géométrie

L’attribut de type « **géométrie** » est utilisé pour visualisé la représentation spatiale des éléments. Elle contient des coordonnées liées à la référence spatiale et du type de géométrie de la classe d’entité. Par exemple dans la BDG, tous les éléments de la classe des étendues d’eau contiennent des surfaces dont la géométrie est de type « **Polygon** » et la référence spatiale est de type géographique. On retrouve seulement trois type de géométrie dans la BDG, soit les points, lignes et les surfaces. Dans ArcGis, il existe plusieurs types de géométries qui peuvent être simples ou complexes. Pour la BDG, seulement quatre types de géométrie sont utilisés, soit le « **Polygon** » pour les surfaces, la « **Polyline** » pour les lignes et le « **MultiPoint** » ou le « **Point** » pour les points. À noter qu’il existe dans ArcObject un conteneur de géométrie nommé « **GeometryBag** » qui est beaucoup utilisé dans nos outils. Cette géométrie peut contenir plusieurs géométries de type « **Polygon**, **Polyline**, **MultiPoint** et/ou **Point** ». Une classe d’entité ne peut pas contenir une géométrie de type « **GeometryBag** » dans ArcGis. Elle peut seulement être utilisé en mémoire dans un outil.

#### Polygone

La géométrie de type « **Polygon** » est utilisée pour visualisé la représentation spatiale des éléments de type surface comme par exemple les étendues d’eau. Elle peut contenir plusieurs anneaux extérieurs (**Exterior ring**). Chaque anneau extérieur peut contenir plusieurs anneaux intérieurs (**Interior ring**). Un anneau (**Ring**) est en fait une ligne (**Path**) fermée qui possède au moins quatre sommets (**Point**). Dans la BDG, à part les classes d’incohérences, tous les éléments de type surface doivent contenir seulement un anneau extérieur.

|  |  |
| --- | --- |
| **GeometryBag** contenant une géométrie de type **Polygon.** Ce Polygone contient un anneau extérieur et un anneau intérieur pour un total de 35 sommets. | Anneau intérieur contenant 10 sommets dont les coordonnées (X=Latitude, Y=Longitude) sont en géographique. La géométrie ne contient pas d’élévation (Z) et ni de mesure (M). |
|  |  |

#### Polyline

La géométrie de type « **Polyline** » est utilisée pour visualisé la représentation spatiale des éléments de type ligne comme par exemple les cours d’eau. Elle peut contenir plusieurs lignes (**Path**). Une ligne (**Path**) peut être fermée ou non et possède au moins une droite (**Segment**) ou deux sommets (**Point**). Dans la BDG, à part les classes d’incohérences, tous les éléments de type ligne doivent contenir seulement une ligne.

|  |  |
| --- | --- |
| **GeometryBag** contenant six géométries de type **Polyline.** Ces Polylignes contiennent chacune une seule ligne pour un total de 44 sommets. | Ligne (**PATH**) contenant 2 sommets dont les coordonnées (X=Latitude, Y=Longitude) sont en géographique. Les géométries ne contiennent pas d’élévation (Z) et ni de mesure (M). |
|  |  |

#### MultiPoint

La géométrie de type « **MultiPoint** » est utilisée pour visualisé la représentation spatiale des éléments de type point. Seules les classes d’incohérences peuvent contenir ce type de géométrie dans la BDG. Elle peut contenir plusieurs sommets (**Point**). Un sommet (**Point**) correspond à une coordonnée **X** et **Y** et peut contenir une élévation (**Z**) et une mesure (**M**).

|  |  |
| --- | --- |
| **GeometryBag** contenant une seule géométrie de type **MultiPoint.** Ce MultiPoint contient un total de 2 sommets. | **Point** contenant 1 sommet dont les coordonnées (X=Latitude, Y=Longitude) sont en géographique. Les géométries ne contiennent pas d’élévation (Z) et ni de mesure (M). |
|  |  |

#### Point

La géométrie de type « **Point** » est utilisée pour visualisé la représentation spatiale des éléments de type point comme les points d’élévation. Elle contient seulement un sommet. Un sommet (**Point**) correspond à une coordonnée **X** et **Y** et peut contenir une élévation (**Z**) et une mesure (**M**).

|  |  |
| --- | --- |
| **GeometryBag** contenant deux géométries de type **Point.** Chaque Point contient un sommet chacun pour un total de 2 sommets. | **Point** contenant 1 sommet dont les coordonnées (X=Latitude, Y=Longitude) sont en géographique. Les géométries ne contiennent pas d’élévation (Z) et ni de mesure (M). |
|  |  |

## Statistiques des données spatiales

Comme on peut le constater ci-dessous, la **BDG** contient 115 classes d’entité spatiale de type point, ligne ou surface. On distingue les classes d’entité de type point, ligne et surface par la présence des deux derniers caractères « **\_0, \_1 et \_2** » dans le nom de la table. Les autres tables présentes dans la **BDG** sont des tables attributives qui sont non spatiales. L’espace totale de la **BDG** est de **260.33 GB** et celle d’une copie des données de la **BDG** dans une **GDB** (File Geodatabase) compressé correspond à **32 GB**. Le temps nécessaire pour créer cette copie est d’environ 15 heures de traitement. La classe d’entité contenant le plus d’espace, d’éléments et de sommets est celle des régions boisées. Cette dernière utilise environ la moitié de l’espace total, contient 55 209 764 éléments et 3 366 560 171 de sommets ce qui représente 37.5% du nombre total d’éléments total et 50.7% du nombre total de sommets.

Parmi ces 115 classes d’entité, 5 classes d’entité spatiales contiennent plus de 95% de l’espace total de la BDG, soit les régions boisées, les courbes de niveau métriques et impériales, les étendues et les cours d’eau. En effectuant un simple filtrage des sommets pour ces classes d’entité, on pourrait réduire de moitié le nombre total de sommets. En effectuant les différents traitements de généralisation pour ces 5 classes, on pourrait réduire encore de façon significative la quantité d’éléments et de sommets. Les temps de traitement seraient aussi réduits de façon très importante.

Puisque les deux classes d’entité spatiales les plus importantes parmi les plus grosses sont les étendues et les cours d’eau, il va de soi d’utiliser ces dernières pour démontrer et expliquer la faisabilité des traitements de squelettisation et de généralisation du 1:50000 au 1:250000. On pourrait aussi facilement expliquer et démontrer la faisabilité de squelettisation et généralisation des données du 1:20000 au 1:50000 présentent dans la BDG.

|  |
| --- |
| Statistiques d’éléments et de sommets pour chaque classe d’entité de la BDG |
|  |

## Modèle de données

Les classes d’entité spatiales de la **BDG** contiennent plusieurs attributs dont la majorité sont identiques entre elles comme par exemple **OBJECTID, SHAPE, CODE\_SPEC, DATASET\_NAME**, etc. Certains attributs sont utilisés par différents outils contenus dans **ArcGis**.

* L’attribut **OBJECTID** est utilisé par **ArcGis** pour identifier de façon unique chaque élément contenu dans une classe d’entité.
* L’attribut **SHAPE** est utilisé par **ArcGis** pour contenir la géométrie des éléments d’une classe d’entité.
* L’attribut **CODE\_SPEC** peut être utilisé par **ArcGis** comme **SUBTYPE**. On peut appliquer des domaines de valeurs d’attributs par classe ou par **SUBTYPE** avec **ArcGis**. Cet attribut est souvent utilisé pour valider et corriger les contraintes d’intégrités spatiales et attributives.
* L’attribut **DATASET\_NAME** contient le nom du jeu de donnée. Il correspond au découpage de la classe d’entité. Cet attribut est utilisé pour gérer les découpages d’entité lors des traitements de validation et de correction des données spatiales et attributives.

Ci-dessous, on peut constater les différences d’attributs entre les classes d’entité étendues et cours d’eau.

**NHN\_HNET\_NETWORK\_LINEAR\_FLOW\_1: Cours d’eau**

OBJECTID : Identifiant de l’objet

CODE\_SPEC : Code spécifique

ACQUISITION\_TECHNIQUE : Technique d'acquisition

BDG\_ID : Identifiant de la BDG

CD\_ELEM\_TOPO : Code d’élément topographique

COMPLETELY\_COVER : Couvre complètement

DATASET\_NAME : Nom du jeu de données

FLOW\_DIRECTION : Direction de l'écoulement

GEONAMEDB : Base de données toponymiques

IDDATE : Date de l'identifiant de la toponymie

LEVEL\_PRIORITY : Niveau de priorité

MD\_ID : Identifiant de métadonnées

MEP\_ID : Identifiant de mise en place

NAMEID\_1 : Identifiant du nom 1 de la toponymie

NAMEID\_2 : Identifiant du nom 2 de la toponymie

NETWORK\_FLOW\_TYPE : Type de filamentaire

PERMANENCY : Permanence

PLANIMETRIC\_ACCURACY : Précision planimétrique

PROVIDER : Fournisseur

VALIDITY\_DATE : Date de validité

WATER\_DEFINITION : Définition de région hydrique

ZT\_ID : Identifiant de zone de transaction

SHAPE : Géométrie de type ligne

**NHN\_HHYD\_WATERBODY\_2  : Étendue d’eau**

OBJECTID : Identifiant de l’objet

CODE\_SPEC : Code spécifique

ACQUISITION\_TECHNIQUE : Technique d'acquisition

BDG\_ID : Identifiant de la BDG

CD\_ELEM\_TOPO : Code d’élément topographique

COMPLETELY\_COVER : Couvre complètement

DATASET\_NAME : Nom du jeu de données

FLOW\_DIRECTION : Direction de l'écoulement

GEONAMEDB : Base de données toponymiques

IDDATE : Date de l'identifiant de la toponymie

ISOLATED : Isolé

LAKEID\_1 : Identifiant du lac 1 de la toponymie

LAKEID\_2 : Identifiant du lac 2 de la toponymie

MD\_ID : Identifiant de métadonnées

MEP\_ID : Identifiant de mise en place

PERMANENCY : Permanence

PLANIMETRIC\_ACCURACY : Précision planimétrique

PROVIDER : Fournisseur

RIVID\_1 : Identifiant du lac 1 de la toponymie

RIVID\_2 : Identifiant du lac 2 de la toponymie

SHORELINE\_WATER\_LEVEL : Niveau d'eau du rivage

VALIDITY\_DATE : Date de validité

WATER\_DEFINITION : Définition de région hydrique

ZT\_ID : Identifiant de zone de transaction

SHAPE : Géométrie de type surface

|  |  |
| --- | --- |
| Modèle de données pour les cours d’eau.  Classe : **nhn\_hnet\_Network\_Linear\_Flow\_1** | Modèle de données pour les étendues d’eau.  Classe : **nhn\_hhyd\_Waterbody\_** |
|  |  |

## Gestion des versions

Afin de conserver un historique des modifications effectuées sur les données de la **BDG**, une gestion des versions est nécessaire. On n’utilise pas le système de gestion des versions de **ESRI** car trop lourd à gérer pour l’ensemble des données du Canada. On utilise plutôt notre propre système de gestion des versions qui fait référence à un découpage et une zone de travail fixe soit le **SNRC**. En fait, une **MDB** (Personnelle Géodatabase) est créée et conservée dans un répertoire d’archivage pour chaque jeu de données **SNRC** modifiée selon un numéro d’édition et de version. Le suivi de ces versions et des modifications effectuées sont conservées dans une base de données de métadonnées nommée **SIB** (Système d’Information sur les base de données). Cela est vrai pour l’ensemble de nos bases de données et de nos produits.

## Échelle de représentation des données spatiales

Afin d’avoir une homogénéité et une bonne intégration entre les éléments spatiaux contenus dans une base de données ou d’un produit, il est important d’avoir une seule échelle de représentation des données. Pour diverses raisons et comme expliqué précédemment, la BDG contient plusieurs échelles de représentation puisqu’aucune contrainte de correction des dimensions minimales n’est réalisée. Bien qu’il en existe beaucoup d’autres, on peut voir dans les dessins ci-dessous certains des problèmes occasionnés (surabondance, illogisme, etc.).

|  |  |
| --- | --- |
| Échelle de représentation différente entre les provinces. Exemple : On retrouve un cours d’eau (ligne) en Ontario versus une étendue d’eau (surface) au Québec pour le même phénomène. | Échelle de représentation différente entre les provinces. Exemple : On retrouve beaucoup plus d’entités représentées en Colombie Britannique par rapport à l’Alberta. |
|  |  |

## Référence spatiale

La référence spatiale est une information très importante pour les données spatiales puisqu’elle détermine dans quel système de coordonnées les données spatiales sont conservées physiquement dans les bases de données et dans quel système de coordonnées elles sont traitées. Il y a une multitude de références spatiales qui offre chacune leurs avantages et inconvénients. Elles sont regroupées sous deux types, les références spatiales géographiques et les projetées. Toutes les références spatiales possèdent des paramètres communs importants comme la précision et la résolution.

Dans un monde idéal, tous les traitements des données spatiales devraient être réalisés dans le même système de coordonnées que celui dans lequel les données spatiales se retrouvent physiquement dans les bases de données. Pour diverses raisons techniques et autres, ce n’est pas toujours le cas. Par exemple, il est impossible techniquement de conserver les données spatiales du Canada dans une base de données avec la projection UTM (Universelle Transverse Mercator) puisque par définition la projection UTM contient plusieurs projections différentes pour l’ensemble du Canada, soit une pour chaque zone UTM. On ne peut non plus traiter l’ensemble des données du Canada dans une seule zone UTM en raison de la trop grande imprécision et distorsion. On peut seulement traiter l’ensemble des données du Canada par zone UTM.

Pour avoir une intégration parfaite entre les données spatiales, toutes les classes spatiales doivent posséder les mêmes références spatiales, les mêmes étendues et les mêmes résolutions.

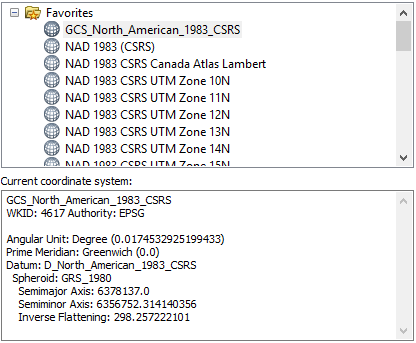
### Précision et résolution de la référence spatiale

Pour diverses raisons techniques, la précision est une caractéristique importante dans une référence spatiale car elle permet d’indiquer quand une valeur A est la même qu’une valeur B même si physiquement elles sont différentes (Exemples : 0.1234=0.1236 si la précision est 0.001 et 0.1234<>0.1236 si la précision est 0.0001).

Il est facile de confondre la précision de la référence spatiale, la résolution de la référence spatiale et la précision des données spatiales. La résolution spatiale est la capacité physique de conserver une valeur selon l’étendue couverte et le type de valeur (entier, réelle, simple ou double). Par défaut, la précision de la référence spatiale est la même celle de la résolution de la référence spatiale mais elle peut être plus grande. La précision des données fait référence à la qualité des données. On calcul ou on estime la précision des données spatiales. On peut dire par exemple qu’on estime à 95% du temps qu’une position X, Y est précis à 5 mètres.

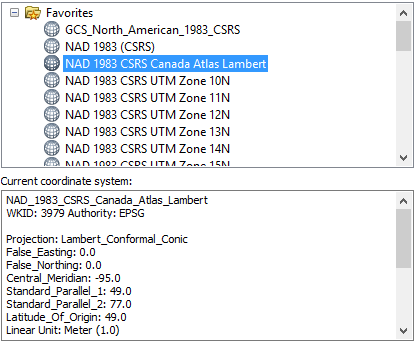
### Référence spatiale géographique

Les références spatiales géographiques sont les plus fidèles façons de représenter la terre puisqu’elles utilisent des angles dans un monde 3D. La particularité des références spatiales géographiques est que le système de coordonnée est présenté sous la forme latitude, longitude dont l’unité de mesure est en degrés. Dans la BDG, puisque les données spatiales sont découpées selon le SNRC (Système National de Référence Cartographique) au 50K et que toutes les données du Canda y sont présentées, la référence spatiale géographique **4617 : GCS\_North\_American\_1983\_CSRS** est utilisée avec la technologie ESRI. La résolution de cette référence spatiale est 0.0000001 degré. Si on convertit cette résolution en mètre, elle sera donc plus grande au nord qu’au sud.



### Référence spatiale projetée

Les références spatiales projetées sont les plus fidèles façons de représenter la terre dans un monde 2D. La particularité des références spatiales projetées est que le système de coordonnée est présenté sous la forme X, Y dont l’unité de mesure est surtout le mètre. Certains traitements de données spatiales peuvent être utilisés avec une référence spatiale géographique mais la plupart doivent utilisés une référence spatiale projetée afin de faciliter les calculs. Par exemple, beaucoup de traitements de validation et correction utilisent des distances et tolérances fixes. La référence spatiale projetée **3979 : NAD 1983 CSRS Canada Atlas Lambert** est donc utilisée avec la technologie ESRI dans beaucoup de traitements de validation et correction. La résolution de cette référence spatiale est 0.001 mètre.



## Données spatiales généralisées

Afin de démontrer la faisabilité des traitements de généralisation des données d’une grande échelle à une autre plus petite, un prototype de traitements a été réalisé interactivement pour deux classes spatiales et deux jeux de données de la BDG, soit les étendues et les cours d’eau des jeux de données 021L et 031P. Dans l’exemple ci-dessous, on peut voir une partie d’un résultat obtenu. On estime à environ 24 heures le temps de traitement pour généraliser 16 jeux de données du 1:50000 à un jeu de données au 1:250000. Une extension des données d’au moins deux fois la largeur de généralisation est nécessaire pour le bon fonctionnement des traitements. Dans l’exemple ci-dessous, on peut constater que le résultat obtenu est tout à fait acceptable.

|  |  |
| --- | --- |
| Données initiales de la BDG à l’échelle d’environ 1:50000. | Données de la BDG généralisées à l’échelle du 1:250000. |
|  |  |

### Étendues d’eau

La généralisation effectuée sur les étendues d’eau dépend des valeurs de dimensions minimales utilisées. Plus elles sont élevées, plus d’éléments ou parties d’éléments seront détruites ou remplacées et plus le traitement sera long à exécuter. Dans le cas des étendues d’eau, lorsqu’une étendue d’eau est généralisée, elle est remplacée par cours d’eau correspondant au squelette de l’étendue d’eau calculé. Les connexions entre les étendues d’eau et les cours d’eau sont conservées. Dans l’exemple ci-dessous, on peut voir que plusieurs étendues d’eau et parties d’étendues d’eau ont été détruites.

|  |  |
| --- | --- |
| Étendue d’eau de la BDG initiale à l’échelle d’environ 1:50000. | Étendue d’eau de la BDG généralisée à l’échelle du 1:250000. |
|  |  |

### Cours d’eau

Comme les étendues d’eau, la généralisation effectuée sur les cours d’eau dépend des valeurs de dimensions minimales utilisées. Plus elles sont élevées, plus d’éléments ou parties d’éléments seront détruites ou remplacées et plus le traitement sera long à exécuter. Dans le cas des cours d’eau, lorsqu’un cours d’eau est généralisé, il est remplacé par un autre cours d’eau correspondant au squelette du cours d’eau calculé. Cependant dans la plupart du temps, les cours d’eau générés à partir des squelettes sont détruits car ils sont trop petits. Les connexions entre les cours d’eau sont conservées. Dans l’exemple ci-dessous, on peut voir que plusieurs cours d’eau et parties de cours d’eau ont été détruits.

|  |  |
| --- | --- |
| Cours d’eau de la BDG initial à l’échelle d’environ 1:50000. | Cours d’eau de la BDG généralisé à l’échelle du 1:250000. |
|  |  |

# Contrainte d’intégrité

Comme son nom le dit, une contrainte d’intégrité est une règle qui sert à s’assurer de l’intégrité des données qu’on acquière, qu’on gère et qu’on distribue. Il existe plusieurs types de contraintes d’intégrité qui peuvent être nommés, caractérisés ou regroupés de plusieurs manières. Elles peuvent être implantées de différentes façons et à plusieurs endroits et niveaux dans nos processus selon les besoins, les capacités et les contraintes techniques et opérationnelles.

Les contraintes d’intégrité ont une importance capitale dans les traitements de généralisation. Elles permettent de prendre pour acquis certaines informations, d’avoir des données homogènes et/ou de connaître l’état des données à traiter. Le fait de savoir que certaines informations dans nos données sont toujours exempt d’erreur, nous permet de simplifier nos traitements et nos processus car on n’a pas à les valider ou à les corriger. Le fait de savoir que nos données sont homogènes nous permet aussi de simplifier nos traitements et nos processus car on n’a pas de traitement additionnel à effectuer pour les rendre homogènes. Le fait de connaître l’état de nos données nous permet d’éviter d’effectuer des traitements inutiles et ainsi minimiser les temps de traitement.

Puisque qu’on parle de généralisation des données de la BDG du 1:50000 au 1:250000 dans ce document, on va expliquer les différentes contraintes d’intégrité utilisées pour valider et corriger les données. Pour différentes raisons, bonnes et/ou mauvaises, le manque de ressources humaines et financières, les difficultés techniques et autres, les données de la BDG ne sont pas exempt d’erreur. On va expliquer les deux grands regroupements de contraintes d’intégrité utilisées, soit les contraintes d’intégrité **attributives** et **spatiales**.

Il est important aussi de mentionné que les contraintes d’intégrité utilisées dans les données pour les données de production et celles utilisées pour les données de la BDG sont différentes puisqu’ils n’utilisent pas le même modèle et le même découpage et zone de travail. On va se concentrer ici seulement sur les contraintes d’intégrité pour les données de la BDG.

## Contrainte d’intégrité attributive

Puisque les données de la BDG sont conservées et gérées dans une base de données Oracle 12c en utilisant une Géodatabase SDE (10.5) de ESRI, certaines contraintes d’intégrité attributives de base sont déjà en place et sont utilisées de façon active, c’est-à-dire que les données ne pourront pas être chargées dans la base de données si ces contraintes ne sont pas respectées. On parle ici par exemple du modèle de données, de classe d’entité, d’attribut, de type d’attribut (Entier, réel, texte, etc.) et de leurs caractéristiques (Valeur nulle, valeur numérique, longueur de texte, valeur unique, etc.).

La plupart des contraintes d’intégrité attributives utilisées dans la BDG sont passives, c’est-à-dire que les données pourront être chargées dans la base de données même si ces contraintes ne sont pas respectées. Elles seront validées et peut être corrigées dans les traitements et processus de production des données de la BDG. Même si des contraintes de domaines de valeurs d’attribut sont possibles dans une Géodatabase de ESRI, on ne les utilise pas dans la BDG. On extrait directement l’information de nos catalogues lorsqu’on veut connaître les détails de la définition des valeurs de nos données.

### Table des contraintes attributives

Pour valider et corriger les erreurs contenues dans les données, on utilise une table de contraintes d’intégrité attributives. Cette table contient toutes les contraintes d’intégrité attributives pour toutes les tables et tous les attributs de nos données. Elle est composée d’un identifiant, d’une étampe, d’une date de création, d’une date de modification, d’un nom de groupe, d’une description, d’un message de correction, d’une requête SQL, d’un nom de table et d’un nom d’attribut.

|  |
| --- |
| Table des contraintes attributives : BDG\_DBA.CONTRAINTE\_INTEGRITE\_SQL |
|  |

### Outil de validation des contraintes attributives

La validation des contraintes d’intégrité attributives se fait en exécutant l’ensemble des requêtes SQL contenues dans cette table. Pour diverses raisons, les corrections ne sont pas toujours réalisées. Elles le sont seulement lorsqu’une correction de non-conformité est exigée et elle se fait normalement de façon « in situ », c’est à dire directement dans la base de données en utilisant une commande SQL.

Un outil a été développé dans le langage Python sous forme de géotraitement pour valider les contraintes d’intégrité attributives en utilisant cette table de contraintes comme un de ses paramètres.

|  |
| --- |
| Outil Python de géotraitement utilisé pour démarrer le traitement de validation des contraintes attributives. |
|  |

### Requêtes attributives

Présentement pour l’ensemble des données de la BDG, on retrouve trois types de contraintes d’intégrité attributives dans la table des contraintes d’intégrité attributives. Celle pour les **domaines de valeurs**, les **expressions régulières** et les **tables des valeurs**.

#### Domaines de valeurs

Les contraintes pour les domaines de valeurs sont utilisées pour valider les valeurs d’attributs selon une liste prédéfinie de valeurs. Cette liste de valeurs prédéfinie doit pouvoir être insérée dans une requête SQL. Elle doit donc être assez limitée.

Bien entendu, on aurait pu utiliser les domaines de ESRI contenus dans une Géodatabase pour effectuer la validation. Puisque d’autres types de contraintes existent comme les expressions régulières et que la technologie ERSI ne les possèdent pas, on voulait utiliser la même méthode de validation pour toutes les contraintes attributives.

|  |
| --- |
| Exemple de contraintes attributives utilisées pour valider les domaines des valeur des attributs. |
|  |

#### Expressions régulières

Les contraintes pour les expressions régulières sont utilisées pour valider les valeurs d’attributs selon un [ensemble](https://fr.wikipedia.org/wiki/Ensemble) de chaînes de caractères possibles en utilisant une syntaxe précise. Cette méthode ne permet pas nécessaire de valider de façon exacte une valeur. Elle permet surtout de valider certains aspects connus des valeurs possibles (Exemple : un SNRC composé de 3 chiffres, deux lettres et 2 chiffres). Cet ensemble de chaine de caractères possibles doit pouvoir être inséré dans une requête SQL. Elle doit donc être assez limitée. La fonction SQL REGEXP\_LIKE disponible dans Oracle est utilisée pour valider ces expressions régulières. Bien que cette méthode est standard, elle a souvent le désavantage d’être difficile à comprendre pour certaines personnes.

|  |
| --- |
| Exemple de contraintes attributives utilisées pour valider les expressions régulières des valeur des attributs. |
|  |

#### Tables des valeurs

Les contraintes pour les tables de valeurs sont utilisées pour valider les valeurs d’attributs qui sont présentes dans une autre table. On utilise cette méthode plutôt qu’une une liste prédéfinie de valeurs lorsque la quantité des valeurs permises est trop grande. Une jointure est utilisée dans la commande SQL pour faire le lien avec les valeurs permises de la table.

|  |
| --- |
| Exemple de contraintes attributives utilisées pour valider les valeurs des attributs qui sont contenues dans une autre table. |
|  |

### Méthodes de traitement

Puisque la quantité d’information contenue dans la BDG est énorme, on devait trouver un moyen performant et complet pour valider toutes les contraintes d’intégrité attributives. Plusieurs méthodes et outils ont été testés et l’utilisation des requêtes SQL dans Oracle nous a permis d’atteindre la complétude et les performances désirées. Ainsi pour valider l’ensemble des contraintes d’intégrité attributives de la BDG (environ 3000 requêtes SQL) via l’outil Python, le temps de traitement est d’environ 3 heures et 30 minutes ce qui est tout à fait acceptable.

Puisque le temps de traitement de la validation de l’ensemble des contraintes d’intégrité attributives est assez court, il n’est pas nécessaire d’effectuer la validation seulement pour l’information qui a été modifiée. Ainsi le traitement de validation pour l’ensemble des données s’exécute à toutes les nuits de façon à obtenir un rapport global d’erreurs des valeurs d’attributs.

### Rapport d’erreurs

Comme on peut le constater ci-dessous, on retrouve dans le rapport d’erreurs la liste des tables modifiées, la liste des identifiants livrés dans la BDG par type de travail, le nombre total d’erreurs, le nombre de requête SQL contenant des erreurs et la liste des contraintes contenant des erreurs incluant les statistiques du traitement de cette contrainte.

Dans les statistiques d’une contrainte, on peut voir la description, le message de correction, la requête SQL utilisée, le nombre de valeurs d’attribut en erreur et le nombre d’erreurs par valeur d’attribut en erreur.

|  |
| --- |
| Exemple d’un rapport d’erreurs obtenu après l’exécution de l’outil pour valider les contraintes d’intégrité attributives de l’ensemble des données de la BDG. |
|  |

Aucune table et carte de statistiques d’erreurs ne sont présente étant donnée la facilité à effectuer l’analyse de ces erreurs.

Avant d’être nommé administrateur des données de la BDG, il n’y avait aucune contrainte d’intégrité pour valider les données de la BDG. Il y avait seulement ceux de la production. De nombreuses corrections de non-conformité, très souvent les mêmes, étaient demandées. Beaucoup d’efforts ont été réalisés depuis afin d’éliminer complètement les erreurs de valeurs d’attributs dans la BDG. Une mauvaise incompréhension des normes, un modèle de données trop flou, une absence de plainte, de volonté et de décision sont de multiples raisons qui font qu’encore aujourd’hui on retrouve encore des erreurs pourtant relativement faciles à corriger.

## Contrainte d’intégrité spatiale

Contrairement aux contraintes d’intégrité attributives, les contraintes d’intégrité spatiales sont beaucoup plus complexes et nombreuses à utiliser et à expliquer. Elles contiennent beaucoup plus de caractéristiques et sont très lourd en traitement et en temps d’exécution.

Puisque les données de la BDG sont conservées et gérées dans une base de données Oracle 12c en utilisant une Géodatabase SDE (10.5) de ESRI, certaines contraintes d’intégrité spatiales de base sont déjà en place et sont utilisées de façon active, c’est-à-dire que les données ne pourront pas être chargées dans la base de données si ces contraintes ne sont pas respectées. On parle ici par exemple du type de géométrie d’une classe d’entité. Un point peut être vide ou contiendra seulement une coordonnée (X,Y ou Lat,Long) avec ou sans élévation (Z) ou mesure (M). Une ligne peut être vide ou contiendra au moins deux coordonnées (X,Y ou Lat,Long) avec ou sans élévation (Z) ou mesure (M). Une surface peut être vide ou contiendra au moins quatre coordonnées (X,Y ou Lat,Long) avec ou sans élévation (Z) ou mesure (M). À moins d’avoir été chargé dans la Géodatabase SDE avec une vieille version, les géométries invalident ne sont plus permises dans SDE. De plus, les géométries se retrouveront toujours à l’intérieure de l’étendue de la référence spatiale de la classe d’entité.

La plupart des contraintes d’intégrité spatiales utilisées dans la BDG sont passives, c’est-à-dire que les données pourront être chargées dans la base de données même si ces contraintes ne sont pas respectées. Elles seront validées et peut être corrigées dans les traitements et processus de production des données de la BDG. Même si des contraintes spatiales sont possibles dans une Géodatabase SDE, on ne les utilise pas dans la BDG car ce serait trop lourd et complexe à gérer.

### Table des contraintes spatiales

Pour valider et corriger les erreurs contenues dans les données, on utilise une table de contraintes d’intégrité spatiales. Cette table contient toutes les contraintes d’intégrité spatiales pour toutes les classes d’entité. Elle est composée d’un identifiant, d’une étampe, d’une date de création, d’une date de modification, d’un nom de groupe, d’une description, d’un message de correction, d’une ou plusieurs requêtes maison, d’un nom de requête et d’un nom de table.

|  |
| --- |
| Table des contraintes spatiales : BDG\_DBA.CONTRAINTE\_INTEGRITE\_SPATIALE |
|  |

### Outil de validation des contraintes spatiales

La validation des contraintes d’intégrité spatiales se fait en exécutant l’ensemble des requêtes spatiales contenues dans cette table. Pour diverses raisons, les corrections ne sont pas toujours réalisées. Elles le sont seulement lorsqu’une correction de non-conformité est exigée et elle se fait normalement de façon « in situ », c’est à dire directement dans la base de données en utilisant diverses commandes d’ArcMap ou divers outils maison.

Un outil interactif a été développé dans le langage Vb.Net sous forme de Add-in dans ArcMap. Un autre outil automatique sous forme de programme exécutable indépendant a aussi été développé pour valider les contraintes d’intégrité spatiales. Ces deux outils utilisent la table des contraintes spatiales comme paramètre d’entrée. Le programme exécutable indépendant peut être démarré de plusieurs façons mais l’utilisation d’un outil Python de géotraitement est privilégiée.

|  |
| --- |
| Outil Python de géotraitement utilisé pour démarrer un traitement de validation des contraintes spatiales. |
|  |

### Requêtes spatiales

Présentement pour l’ensemble des données de la BDG, on retrouve au moins quinze types de requêtes spatiales dans la table des contraintes d’intégrité spatiales. Elles ne sont pas tous présentent car le travail est encore en cours d’exécution. D’autres types de requête pourraient être ajoutés. Une contrainte spatiale contient une ou plusieurs requêtes spatiales. Malgré certains standards existants mais incomplets pour nos besoins, les requêtes spatiales sont des requêtes non standards et sont faits maison. Seule la requête de **Relation spatiale** se rapproche des standards internationaux. On utilise aucune requête spatiale de type SQL ou fonction spatiale Oracle.

Étant donnée la complexité et la quantité des contraintes et requêtes spatiales possibles, elles ne seront pas tous expliquées dans ce document. Un document aussi complet que celui-ci serait nécessaire pour expliquer toutes les possibilités. Seules les requêtes en lien avec les dimensions minimales seront expliquées un peu plus loin dans ce document.

Bien que plusieurs contraintes spatiales ESRI sont possibles dans une Géodatabase pour effectuer la validation, la lourdeur et la gestion de ces dernières était trop importante. De plus, d’autres types de contraintes étaient nécessaires mais la technologie ERSI ne les possèdent pas. On voulait utiliser la même méthode de validation pour toutes les contraintes spatiales.

|  |
| --- |
| Exemple des contraintes d’intégrité spatiales utilisées pour la validation. |
|  |

### Méthodes de traitement

Puisque la quantité d’information contenue dans la BDG est énorme, on devait trouver un moyen performant et complet pour valider toutes les contraintes d’intégrité spatiales. Plusieurs méthodes et outils ont été testés. L’utilisation de diverses fonctionnalités spatiales ESRI sur des données chargées en mémoire nous a permis de se rapprocher le plus de la complétude et les performances désirées.

Contrairement au traitement de validation des contraintes attributives, il est impossible de valider l’ensemble des contraintes d’intégrité spatiales pour l’ensemble des données de la BDG en un seul traitement. Le temps de traitement serait trop grand et il est présentement impossible de charger et traiter l’ensemble des données de la BDG en mémoire. Les fonctionnalités spatiales ESRI sont utilisables seulement dans le système Windows 32 bits ce qui limite la mémoire à environ 3 GB en l’extensionnant au maximum.

Ainsi deux méthodes de traitement sont possibles pour valider l’ensemble des contraintes d’intégrité spatiales pour l’ensemble des données de la BDG (environ 800 contraintes spatiales et il en manque). La première consiste à valider une ou plusieurs contraintes spatiales pour l’ensemble des données d’une seule classe d’entité. C’est la plus performante mais elle est possible seulement pour les classes d’entité contenant peu d’éléments et de sommets. Malheureusement, peu de classes d’entité peuvent utilisées cette méthode en raison de la limite de la mémoire. La deuxième consiste à valider toutes les contraintes spatiales pour une ou plusieurs classes d’entité pour un ou plusieurs identifiants de découpage. Cette méthode est moins performante mais risque moins d’atteindre la limite de mémoire possible.

Plusieurs traitements de validation des contraintes spatiales sont donc nécessaires en utilisant le programme exécutable indépendant. Les deux méthodes ont été utilisées pour valider l’ensemble des contraintes spatiales pour l’ensemble des données de la BDG. Environ trois mois a été nécessaire pour tout valider en utilisant un seul poste de travail (4 CPU, 3.4GHz et 16 GB). Puisque des mises à jour sont effectuées presqu’à tous les jours par identifiant de découpage SNRC et par type de travail sur les données de la BDG, un outil Python de géotraitement a été développé pour lancer les traitements de validation des contraintes spatiales des données livrées dans la BDG par identifiant de découpage SNRC et par type de travail. Cet outil est exécuté à toute les nuits de façon à obtenir un rapport global d’erreurs des contraintes spatiales par identifiant de découpage SNRC et par type de travail.

|  |
| --- |
| Outil Python de géotraitement utilisé pour lancer les traitements de validation des contraintes spatiales des données livrées dans la BDG par identifiant de découpage SNRC et par type de travail. |
|  |

### Rapport d’erreurs

Comme on peut le constater ci-dessous, on retrouve dans le rapport les paramètres du programme de validation des contraintes spatiales utilisées, la liste des contraintes d’intégrité spatiales contenant des erreurs par identifiant de découpage SNRC et ses statistiques de traitement ainsi que les statistiques sur le traitement global exécuté.

Pour chaque contrainte spatiale d’un identifiant de découpage SNRC contenant des erreurs, on peut voir sur la première ligne, le numéro de séquence d’exécution, l’identifiant de découpage SNRC, l’identifiant de la contrainte et le nom de groupe de la contrainte. Pour le reste, on peut voir la description de la contrainte, le message de correction de la contrainte, les requêtes spatiales exécutées, la date de début du traitement, le nombre d’éléments traités, le nombre d’éléments sélectionnés, l’espace mémoire utilisée, la date de fin de traitement et son temps d’exécution, le nom de la Géodatabase d’erreurs créées, le nom de la classe d’erreurs créées, le temps total d’exécution de la contrainte spatiale et l’espace mémoire maximale utilisée ainsi que le nombre d’erreurs trouvées par rapport au nombre d’éléments traités.

Les statistiques sur le traitement global exécuté contiennent le nombre total d’identifiants de découpage SNRC traités, le nombre total de contraintes spatiales traitées, le nombre total d’éléments traités, le nombre total d’erreurs trouvées et le temps total d’exécution du traitement de validation.

|  |
| --- |
| Exemple d’un rapport d’erreurs pour l’identifiant de découpage SNRC **016D11** pour le type de travail **RHN\_ANOM2\_BDG.** |
|  |

### Table des statistiques d’erreurs

Contrairement au contraintes attributives, il n’est pas possible de connaître l’état global des données en un seul traitement pour les contraintes spatiales puisque plusieurs traitements sont nécessaires et une énorme quantité d’erreurs sont présentes. La seule façon d’avoir une vue globale de l’état des données pour les contraintes spatiales est de conserver les résultats dans une table de statistiques d’erreurs par identifiant de découpage SNRC, par classe d’entité et par contrainte spatiale. Ainsi lors de l’exécution d’un traitement de validation, si une table de statistiques est spécifiée, les statistiques seront conservées dans cette dernière. À partir de cette table de statistiques d’erreurs de contraintes spatiales, une vue contenant le nombre et le pourcentage d’erreurs total par contrainte est créée à la volée.

|  |  |
| --- | --- |
| Table contenant les statistiques pour le nombre d’erreurs par identifiant, par table et par contrainte spatiale triées en ordre décroissant. | Vue calculée à la volée contenant les statistiques pour le nombre et pourcentage d’erreurs total. par contrainte spatiale triées en ordre décroissant. |
|  |  |

### Cartes des statistiques d’erreurs

Bien entendu, si on veut une information globale, on peut extraire l’information désirée directement à partir de la table des statistiques des contraintes spatiales via une requête SQL. Mais si on veut une vue graphique de l’état des données spatiales, une carte contenant l’information désirée doit être créée. Un programme Python sous forme de géotraitement a été développé afin de créer une ou plusieurs cartes par classe d’entité et par contrainte spatiale pour l’ensemble des identifiants de découpage SNRC du Canada. L’ensemble des cartes générées sont insérées dans un seul fichier en format PDF sous le nom spécifié suivi de la date de création de ce dernier.

|  |
| --- |
| Outil Python de géotraitement qui permet de créer une ou plusieurs cartes par classe d’entité et par contrainte spatiale pour l’ensemble des identifiants de découpage SNRC du Canada. |
|  |

|  |  |
| --- | --- |
| Carte contenant tous les identifiants de découpage SNRC du Canada et représentant les statistiques d’erreurs de surabondance des sommets pour les cours d’eau. | Carte contenant tous les identifiants de découpage SNRC du Canada et représentant les statistiques d’erreurs d’ajustement des données entre les découpages pour les cours d’eau. |
|  |  |

|  |  |
| --- | --- |
| Carte contenant tous les identifiants de découpage SNRC du Canada et représentant les statistiques d’erreurs de surabondance des sommets pour les étendues d’eau. | Carte contenant tous les identifiants de découpage SNRC du Canada et représentant les statistiques d’erreurs d’ajustement des données entre les découpages pour les étendues d’eau. |
|  |  |

# Topologie

Beaucoup de traitements de validation et de correction des données spatiales nécessite l’utilisation de la topologie car elle permet de connaître les liens existants entre les géométries des classes d’éléments. La connaissance de ces liens est utile afin de voir et questionner l’état des données pour prendre certaines décisions. Elle doit également permettre d’effectuer la correction des imperfections géométriques entre les géométries des éléments en relation selon une tolérance minimale correspondant à la précision des données spatiales désirée. L’insertion et la modification de certains sommets sont donc obligatoires afin de faire correspondre parfaitement les parties de géométries en relation. Une priorité d’insertion et de modification est recommandée mais non obligatoire en raison de sa lourdeur d’utilisation.

Le traitement de topologie est très lourd, prend beaucoup d’espace et doit être performant pour être viable. La topologie peut être créée et conservée physiquement dans une base de données ou peut être calculée à la volée selon les besoins. Pour diverses raisons techniques, les traitements de validation et de correction utilisent la topologie créée à la volée selon les besoins sans les priorités entre les classes d’éléments. Une structure de Nœud (Node), Limite (Edge) et Face (Face) est créée comme résultat du traitement de topologie.

La technologie ESRI possède ces fonctionnalités et est très performante mais ne possède pas l’utilisation de priorité et de Face lors de la création de la topologie à la volée (Voir « IMapTopology et ITopologyGraph » dans ArcObject).

## Tolérance minimale des données spatiales (Précision)

En plus des classes d’éléments et leurs priorités optionnelles comme paramètres d’entrés, la tolérance minimale correspondant à la précision de référence spatiale est demandée comme paramètre d’entrée. Cette valeur est directement liée à ce qui doit être corrigé et considéré comme parfaitement connecté.

|  |
| --- |
| Trois éléments sont présents dans cette vue, 2 cours d’eau (O1 et O3) et une étendue d’eau (O2). |
|  |

## Nœud (Node)

Le **Node** est une composante de la structure de la topologie résultante. Il contient une géométrie de type point et plusieurs fonctionnalités comme celles pour indiquer à quelles et à combien de **Edges** de géométrie d’éléments il est connecté (Voir « ITopologyNode » dans ArcObject de ESRI).

|  |
| --- |
| Pour les trois éléments présents (2 cours d’eau et 1 étendue d’eau), quatre Nœuds (Nodes) sont présents dans cette vue (N1, N2, N3 et N4). |
|  |

## Limite (Edge)

Le **Edge** est une composante de la structure de la topologie résultante. Il contient une géométrie de type ligne et plusieurs fonctionnalités comme celles pour indiquer à quels et à combien de **Nodes** de géométrie d’éléments il est connecté (Voir « ITopologyEdge » dans ArcObject de ESRI).

|  |
| --- |
| Pour les trois éléments présents (2 cours d’eau et 1 étendue d’eau), cinq Limites (Edges) sont présentes dans cette vue (E1, E2, E3, E4 et E5). |
|  |

## Face (Face)

La **Face** est une composante de la structure de la topologie résultante. Il contient une géométrie de type surface et plusieurs fonctionnalités comme celles pour indiquer à quels et à combien de **Nodes** et de **Edges** de géométrie d’éléments il est connecté (Non disponible dans la technologie ESRI à la volée).

|  |
| --- |
| Pour les trois éléments présents (2 cours d’eau et 1 étendue d’eau), Trois Faces (Faces) sont présentes dans cette vue (F1, F2 et F3). |
|  |

# Relations spatiales

Beaucoup de traitements de validation et de correction nécessite aussi l’utilisation des relations spatiales. Les relations spatiales permettent d’indiquer si un état de connexion prédéfini ou non est respecté entre deux géométries ou entre plusieurs géométries. Un nom de relation spatiale est associé à chaque état de connexion **prédéfini**. Un masque à neuf intersections est utilisé pour chaque état de connexion **non défini**.

Le résultat obtenu du traitement de relation spatiale pour deux géométries est un booléen qui indique vrai ou faux. Le résultat obtenu du traitement de relation spatiale pour plusieurs géométries est une structure de liens entre les numéros séquentiels des géométries traitées qui respecte l’état spécifié. Pour être viable, la performance de ce dernier est très importante étant donné la possibilité de traité une très grande quantité de données.

La technologie ESRI possède ces fonctionnalités et est très performante (Voir « IRelationalOperator, IRelationalOperatorNxM et IRelationResult » dans ArcObject).

## Précision de la référence spatiale

Puisque les traitements des relations spatiales sont effectués sur des géométries, une référence spatiale commune doit être utilisée. On peut donc changer la valeur de précision de cette référence spatiale afin d’indiquer ce qui est considéré comme connecté même si physiquement, ils ne le sont pas (Voir la précision des références spatiales pour plus d’information).

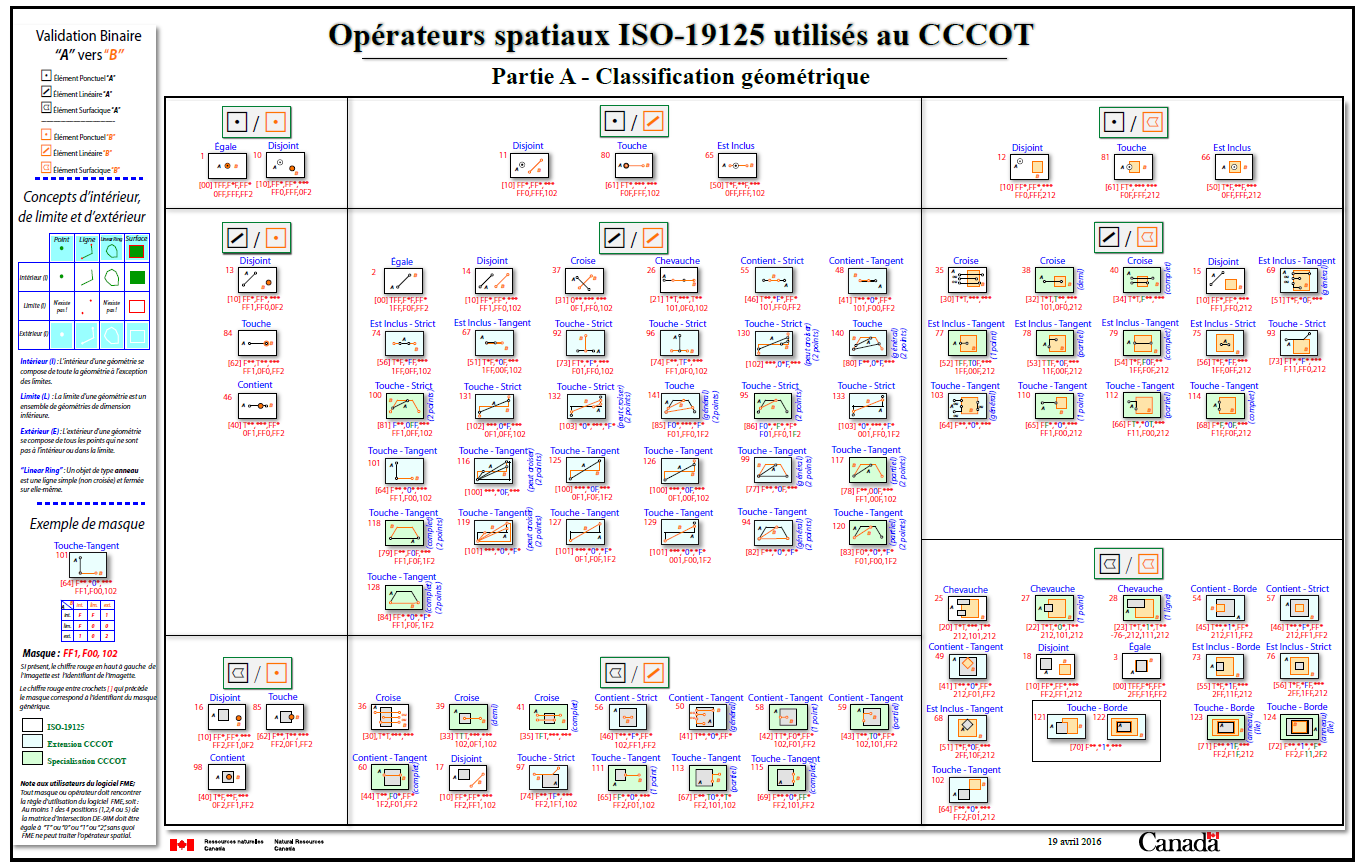
## Relations spatiales prédéfinies

Des standards internationaux ont été établi pour ces relations spatiales prédéfinies. L’avantage majeure d’utiliser les relations spatiales prédéfinies est normalement leur performance car ils ont été conçu pour traiter leur propre cas.

* **Disjoint (Disjoint)** : Aucune intersection n’est présente.
* **Intersecté (Intersect)** : Au moins une intersection de type point, ligne ou surface est présente.
* **Égale (Equals)** : Les géométries sont identiques.
* **Touche (Touches)** : Il y a au moins une intersection entre les limites des géométries.
* **Contient (Contains)** : La première géométrie contient l’autre.
* **Est inclus (Within)** : La première géométrie est incluse dans l’autre.
* **Croise (Crosses)** : La dimension de la géométrie d’intersection est plus petite.
* **Chevauche (Overlaps)** : La dimension de la géométrie d’intersection est la même.

## Relations spatiales non définies / Masque à neuf intersections (Egenhofer et/ou Clementini)

Des standards internationaux ont été établis pour ces relations spatiales non définies. On utilise un masque à neuf intersections. On décompose chaque géométrie en 3 sous géométries, soit l’intérieur (I), la limite (L) et l’extérieur (E) de la géométrie. On compare l’état d’intersection entre ces 3 sous géométries pour les 2 géométries à traiter. Le but étant de fournir un masque en entrée et de retourner un booléen vrai ou faux selon le respect ou non de ce masque. Selon la méthode Egenhofer et le système utilisé, le masque peut contenir les valeurs suivantes (0,F:Faux, 1,V,T:Vrai ou \*:Vrai ou Faux). Selon la méthode Clementini, le masque peut contenir les valeurs selon Egenhofer en plus des valeurs de dimension des géométries d’intersections soient (0:Point, 1:Ligne ou 2:Surface). Le document ci-dessous explique les masques à neuf intersections selon Egenhofer et/ou Clementini cependant ce ne sont pas des opérateurs spatiaux mais des relations spatiales. Comme on peut le constater, il y a de la confusion dans l’utilisation des termes. Le désavantage majeur à utiliser ces relations est que c’est beaucoup plus lent d’exécution.



# Opérateurs spatiaux (Topologique)

Beaucoup de traitements de validation et de correction nécessite aussi l’utilisation des opérateurs spatiaux. Ces derniers sont surtout utilisés pour corriger les données ou pour construire les géométries correspondantes aux erreurs détectées. Contrairement aux relations spatiales, elles permettent de retourner une géométrie. Un nom d’opérateur spatial est associé à chaque traitement **prédéfini**.

Le résultat obtenu du traitement des opérateurs spatiaux pour deux géométries est une géométrie. La géométrie résultante peut être vide ou correspond à la définition même de l’opérateur spatial prédéfini demandé. Pour être viable, la performance de ce dernier est très importante étant donné la possibilité de traité une très grande quantité de données.

La technologie ESRI possède ces fonctionnalités (Voir « ITopologicalOperator » dans ArcObject).

## Précision de la référence spatiale

Tout comme les relations spatiales, les traitements des opérateurs spatiaux sont effectués sur des géométries, une référence spatiale commune doit être utilisée. On peut donc changer la valeur de précision de cette référence spatiale afin d’indiquer ce qui est considéré comme connecté même si physiquement, ils ne le sont pas (Voir la précision des références spatiales pour plus d’information).

## Opérateurs spatiaux prédéfinis (implique deux géométries)

Ces opérateurs spatiaux impliquent toujours deux géométries. Le résultat de ce traitement est une géométrie qui peut être vide ou correspond au traitement demandé. Les principaux utilisés sont les suivants :

* **Intersection (Intersect)** : Retourne la géométrie d’intersection entre deux géométries A et B.
* **Différence (Difference)** : Retourne la partie de la géométrie A qui n’intersecte pas la géométrie B.
* **DifférenceInverse (InverseDifference)** : Retourne la partie de la géométrie B qui n’intersecte pas la géométrie A.
* **Symétrie (SymmetricDifference)** : Retourne les parties des géométries A et B qui ne s’intersectent pas.
* **Union (Union)** : Retourne l’union des deux géométries A et B.

## Autres opérateurs spatiaux (implique une seule géométrie)

D’autres opérateurs spatiaux sont utilisés mais elles impliquent seulement une géométrie. Ils sont nécessaires pour certaines occasions spécifiques. Les principaux utilisés sont les suivants :

* **Simplifier (Simplify)** : Permet de corriger une géométrie erronée (Topologiquement invalide).
* **Limite (Boundary)** : Permet de retourner la limite d’une géométrie.
* **Tampon (Buffer)** : Permet de retourner un tampon d’une géométrie selon une distance.

# Dimensions minimales

Il existe plusieurs contraintes d’intégrité spatiales en lien avec différentes valeurs de dimension minimale comme la précision, la distance latérale, la longueur minimale d’une droite ou d’une ligne, la superficie minimale, etc. En fait, les contraintes de dimension minimale servent à valider et à corriger l’intégrité des représentations géométriques des différents phénomènes cartographiques présentent dans nos données. Chaque phénomène cartographique est représenté par un type de géométrie (point, ligne et/ou surface) qui est normalement directement dépendant d’une seule échelle de représentation désirée (Exemple : 1:50000).

Comme dans les données de la BDG, lorsque plusieurs échelles de représentation sont présentes pour un même phénomène et même type de géométrie, on dit qu’elles sont inconsistantes, c’est-à-dire qu’il y a un manque de cohérence entre les données. On le voit souvent à la limite des jeux de données et des provinces lorsqu’un même phénomène est représenté d’un coté en ligne et de l’autre en surface ou qu’il y a surabondance d’information d’un jeu de données par rapport à un autre. Autre raison de la pertinence des contraintes de dimension minimale est de valider et corriger les illogismes comme des lacs dont la superficie est impossible ou des rivières dont la longueur est impossible en fonction de la source de captage utilisée.

Le terme généralisation est très général et peut être nébuleux pour certains, car il est utilisé à différents niveaux et dans différents contextes. Par exemple, plusieurs personnes disent qu’ils font de la généralisation lorsqu’un traitement selon Douglass-Peuker est réalisé afin d’éliminer la surabondance de sommets. D’autres personnes vont dire qu’ils ont généralisé les bâtiments lorsqu’ils ont transformé ceux-ci en agglomération. Tout ceci n’est pas faut mais on aurait plutôt tendance à dire que la généralisation est plutôt le résultat de la correction de toutes les contraintes d’intégrité qui nécessitent l’utilisation des dimensions minimales en lien avec des échelles de représentation géométrique. En d’autres termes, on peut dire qu’une généralisation a été réalisée lorsque la correction de toutes les contraintes de dimension minimale a été effectuée afin de respecter l’échelle de représentation choisie pour un phénomène et un type de géométrie.

## Proximité des données spatiales (Topologie)

Le traitement de proximité des données spatiales est utilisé pour comparer les sommets entre eux et identifier la connexion entre les éléments (Topologie). Deux sommets d’un même élément ou d’éléments différents dont la distance (précision) est inférieure ou égale à cette dernière sont considérés comme équivalents et sont déplacés selon le centre de ces deux sommets. Après le déplacement, si les deux sommets d’un même élément sont superposés, un seul est conservé. Lorsque deux géométries se retrouvent à l’intérieure de la distance de précision et qu’il y a absence de sommet, des sommets sont insérés dans les géométries afin qu’il y ait une correspondance parfaite. En fait, ce traitement est utilisé pour valider et corriger les erreurs de proximité entre les géométries d’éléments et entre les sommets d’une même géométrie d’un élément. Ce traitement peut être utilisé de façon indépendante ou comme sous traitement d’un autre traitement comme celui de distance latérale minimale entre les sommets.

|  |  |
| --- | --- |
| **Avant :** Plusieurs erreurs de proximité entre les géométries des éléments. | **Après :** Absence d’erreurs de proximité entre les géométries des éléments. |
|  |  |

## Distance latérale minimale entre les sommets (Douglass-Peuker)

L’algorithme de **Douglas-Peuker** est assez bien connu et est souvent utilisé pour valider et corriger la surabondance de sommets. Ce traitement permet de conserver sensiblement l’allure générale d’une géométrie puisqu’il utilise une distance latérale. Afin de ne pas déconnecter les éléments en relation aux éléments traités, la topologie doit être absolument utilisée lors de ce traitement. En fait, le traitement de filtrage des sommets selon **Douglas-Peuker** doit être effectué sur les **Edges** de la topologie des géométries d’éléments et non sur la géométrie des éléments afin de conserver la connexion entre les géométries des éléments. Ce traitement peut être utilisé de façon indépendante ou comme sous traitement d’un autre traitement comme celui de largeur et longueur de généralisation.

|  |  |
| --- | --- |
| **Avant :** Un polygone contenant 25 sommets et deux connexions. | **Après :** Un polygone contenant 11 sommets et dont les deux connexions sont conservés. |
|  |  |

## Longueur minimale d’une droite

Ce traitement est assez simple puisque le but est de valider et corriger la surabondance de droites contenue dans les géométries d’éléments. Toutes les droites dont la longueur est inférieure ou égale à cette valeur sont détruites. Cette valeur peut être confondue et jugée inutile par rapport à la valeur de proximité et de distance latérale si et seulement si la longueur minimale d’une droite est inférieure ou égale à celle de la proximité et de distance latérale. Cependant, la longueur minimale d’une droite devrait toujours être supérieure aux deux autres ce qui fait qu’elles sont donc considérées complémentaires et nécessaires. Afin de ne pas déconnecter les éléments en relation aux éléments traités, la topologie doit être absolument utilisée lors de ce traitement. En fait, le traitement de filtrage des droites doit être effectué sur les **Edges** de la topologie des géométries d’éléments et non sur la géométrie des éléments afin de conserver la connexion entre les géométries des éléments.

|  |  |
| --- | --- |
| **Avant :** Un polygone contenant 24 droites et trois connexions. | **Après :** Un polygone contenant 15 droites dont les trois connexions sont conservées. |
|  |  |

## Longueur maximale d’une droite (Distance pour densifier)

*Ce traitement n’est pas considéré comme une contrainte de dimension minimale mais plutôt comme un sous traitement d’un autre traitement de contrainte de dimension minimale. Ce sous traitement est d’ailleurs utilisé dans celui de largeur et longueur de généralisation afin d’améliorer le résultat de ce dernier.*

|  |  |
| --- | --- |
| Avant la densification : 14 sommets. | Après la densification : 27 sommets. |
|  |  |

## Longueur minimale d’une ligne

Le but de ce traitement est de valider et corriger la surabondance de lignes contenues dans une classe d’éléments. Il semble assez simple mais en réalité, il est assez complexe car ce ne sont pas toutes les lignes dont la longueur est inférieure à la longueur minimale qui sont détruites. Afin de ne pas déconnecter les éléments en relation aux éléments traités, la topologie doit être absolument utilisée lors de ce traitement. En fait, le traitement de filtrage des lignes doit être effectué sur les **Edges non segmentés (Polyline)** de la topologie des géométries d’éléments et non sur la géométrie des éléments afin de conserver la connexion entre les géométries des éléments. Lorsqu’il y a deux **Edges** adjacents consécutifs, ces **Edges** sont fusionnés pour être traités comme **Edges non segmentés (Polyline)**. Les **Polylignes** connectées à ses deux extrémités ou dont la longueur est supérieure à la longueur minimale ne seront jamais détruites. Les **Polylignes** dont la longueur est inférieure à la longueur minimale et qui sont non connectées seront tous détruites sans exception. Les **Polylignes** dont la longueur est inférieure à la longueur minimale, qui sont connectées seulement à une extrémité, qu’il y a plus d’une **Polyligne** adjacente et dont la longueur est la plus petite des **Polylignes** adjacentes seront détruites. La **Polyligne** adjacente et dont la longueur est la plus grande des **Polylignes** adjacentes ne seront pas détruites mais pourrait l’être dans un traitement itératif suivant si sa longueur est inférieure à la longueur minimale. Ce traitement nécessite une reconstruction itérative de la topologie et un calcul continu pour identifier les **Edges non segmentés (Polyline)**.

|  |  |
| --- | --- |
| **Avant :** 18 lignes, 18 Edges non segmentés (Polyline). | **Après :** 11 lignes, 8 Edges non segmentés (Polyline). |
|  |  |

## Longueur minimale d’une Polyligne fermée

Ce traitement est complémentaire au traitement de longueur minimale des lignes puisqu’il est appliqué seulement pour les lignes qui sont connectées à leurs deux extrémités et dont la fusion avec ses **Edges** adjacents forment une Polyligne fermée dont la longueur est inférieure ou égale à la longueur minimale d’une Polyligne fermée. Le but est de valider et corriger une certaine densité de lignes contenues dans une classe d’éléments. La complexité de ce traitement se retrouve surtout dans l’identification des Polylignes fermées. Afin de ne pas déconnecter les éléments en relation aux éléments traités et permettre l’identification des Polylignes fermées, la topologie doit être absolument utilisée lors de ce traitement. Pour chaque Polyligne fermée dont la longueur est inférieure ou égale à la longueur minimale d’une Polyligne fermée et qui peut être formée d’un seul ou plusieurs **Edges**, le **Edge** le plus long est détruit.

|  |  |
| --- | --- |
| **Avant :** 9 lignes, 9 Edges, 3 Polylignes fermées. | **Après :** 5 lignes, 5 Edges, 0 Polyligne fermée. |
|  |  |

## Largeur minimale de généralisation

Le traitement pour valider et corriger la largeur minimale de généralisation n’est jamais utilisée de façon indépendante. Il est toujours utilisé comme sous traitement du traitement de largeur et longueur de généralisation (Intérieure, extérieure, gauche ou droite). Ce traitement est très complexe à réaliser, à expliquer et à comprendre. Pour cette raison, plusieurs pages et dessins seront nécessaires. Ces explications plus détaillées seront effectuées un peu plus loin dans le document. En résumé, le traitement pour valider et corriger la largeur minimale de généralisation nécessite l’utilisation de plusieurs sous traitements, comme la densification des sommets, la topologie, l’utilisation des relations spatiales et des opérateurs spatiaux, la triangulation de Delaunay. En fait, l’idée générale est de trouver les parties de surface dont la largeur est inférieure ou égale la largeur minimale de généralisation. Le traitement de largeur et longueur de généralisation implique plusieurs autres sous traitements comme la correction des parties de surface identifiées, la superficie minimale, la connexion des points de connexion au squelette, etc.

|  |  |
| --- | --- |
| Dans cet exemple on utilise les droites intérieures des triangles de Delaunay pour identifier les parties de surfaces dont la largeur est inférieure ou égale à la largeur minimale de généralisation (voir en rouge les droites plus foncées). | Dans cet exemple on utilise les droites extérieures des triangles de Delaunay pour identifier les parties de surfaces dont la largeur est inférieure ou égale à la largeur minimale de généralisation (voir en rouge les droites plus foncées). |
|  |  |

## Longueur minimale de généralisation

Le traitement pour valider et corriger la longueur minimale de généralisation n’est jamais utilisée de façon indépendante. Il est toujours utilisé comme sous traitement du traitement de largeur et longueur de généralisation (Intérieure, extérieure, gauche ou droite). Ce traitement est très complexe à réaliser, à expliquer et à comprendre. Pour cette raison, plusieurs pages et dessins seront nécessaires. Ces explications plus détaillées seront effectuées un peu plus loin dans le document. En résumé, le traitement pour valider et corriger la longueur minimale de généralisation nécessite l’utilisation de plusieurs sous traitements, comme la densification des sommets, la topologie, l’utilisation des relations spatiales et des opérateurs spatiaux, la triangulation de Delaunay, la squelettisation et la largeur de généralisation. En fait, l’idée générale est de trouver les parties de surface dont la largeur est inférieure ou égale la largeur minimale de généralisation et dont la longueur de ces parties est supérieure à la longueur minimale de généralisation. Le traitement de largeur et longueur de généralisation implique plusieurs autres sous traitements comme la correction des parties de surface identifiées, la superficie minimale, la connexion des points de connexion au squelette, etc.

|  |  |
| --- | --- |
| Dans cet exemple on utilise les droites intérieures des triangles de Delaunay pour identifier les parties de surfaces dont la largeur est inférieure ou égale à la largeur minimale de généralisation et dont la longueur est supérieure à la longueur minimale de généralisation (voir en rouge les lignes plus foncées correspondant à des parties du squelette intérieur). | Dans cet exemple on utilise les droites extérieures des triangles de Delaunay pour identifier les parties de surfaces dont la largeur est inférieure ou égale à la largeur minimale de généralisation et dont la longueur est supérieure à la longueur minimale de généralisation (voir en rouge les lignes plus foncées correspondant à des parties du squelette extérieur). |
|  |  |

## Superficie minimale d’un anneau extérieur

Ce traitement est assez simple en soit puisqu’il consiste à valider et corriger les anneaux extérieurs superflus/en trop pour les éléments de type surface. Un élément de type surface peut contenir un ou plusieurs anneaux extérieurs, Chaque anneau extérieur peut contenir zéro ou plusieurs anneaux intérieurs. Les anneaux extérieurs d’un élément de type surface dont la superficie est inférieure ou égale à la superficie minimale d’un anneau extérieur sont détruits incluant les anneaux intérieurs puisque par définition, la superficie d’un anneau intérieur est toujours inférieure à celle de son anneau extérieur. Ce traitement est utilisé de façon indépendante ou comme sous traitement d’un autre traitement comme c’est le cas pour le traitement de largeur et longueur de généralisation.

|  |  |
| --- | --- |
| **Avant** : 3 anneaux extérieurs, 5 anneaux intérieurs. | **Après** : 1 anneau extérieur, 2 anneaux intérieurs. |
|  |  |

## Superficie minimale d’un anneau intérieur

Ce traitement est assez simple en soit puisqu’il consiste à valider et corriger les anneaux intérieurs superflus/en trop pour les éléments de type surface. Un élément de type surface peut contenir un ou plusieurs anneaux extérieurs, Chaque anneau extérieur peut contenir zéro ou plusieurs anneaux intérieurs. Les anneaux intérieurs d’un élément de type surface dont la superficie est inférieure ou égale à la superficie minimale d’un anneau intérieur sont détruits. Dans ce traitement, même si la superficie des anneaux extérieurs sont inférieures à la superficie minimale d’un anneau intérieur, les anneaux extérieurs ne sont jamais détruits. Ce traitement est utilisé de façon indépendante ou comme sous traitement d’un autre traitement comme c’est le cas pour le traitement de largeur et longueur de généralisation.

|  |  |
| --- | --- |
| **Avant** : 3 anneaux extérieurs, 5 anneaux intérieurs. | **Après** : 3 anneaux extérieurs, 2 anneaux intérieurs. |
|  |  |

# Squelettisation

Le traitement de squelettisation est un traitement assez complexe et très technique à réaliser. Il demande beaucoup de calcul et doit tenir compte de plusieurs facteurs afin d’obtenir un bon résultat. Il permet d’identifier une ou plusieurs lignes de centre pour l’intérieur d’une ou plusieurs géométries ou pour l’extérieur d’une ou plusieurs géométries avec l’enveloppe de ces dernières.

On utilise ce traitement pour au moins deux raisons connues, soit pour l’identification d’un phénomène cartographique virtuelle correspondant aux lignes de centre de la géométrie de ce dernier (Exemple : Lignes de centre des étendues d’eau) ou soit pour valider ou corriger la généralisation (dimension minimale) des géométries d’éléments d’une échelle de représentation (Exemple: 1/50000) vers une autre plus petite (Exemple: 1/250000).

Puisque ce document fait référence à des cours d’eau et à des étendues d’eau et que ces derniers sont représentés par des géométries de type ligne et surface, les explications du traitement de squelettisation se feront par rapport à ces deux types de géométrie.

La technologie ESRI ne possède pas les fonctionnalités pour créer des squelettes ou pour généraliser des lignes ou des surfaces via les squelettes. Cependant, elle possède l’infrastructure nécessaire à la réalisation de ce traitement. Certains outils (comme le programme Skeletton développé en Java) et techniques (comme celle via le Diagramme de Voronoi) existent déjà pour créer des squelettes mais me semble incomplet et ne donne pas toujours un bon résultat. La technique proposée (via la Triangulation de Delaunay) me semble beaucoup plus solide en terme de résultat. Il est aussi plus complet puisqu’il utilise différents paramètres pour retourner un résultat plus ou moins détaillé selon les besoins des usagers.

## Géométries de type ligne (Polyline) et Surface (Polygon)

Dans la technologie ESRI, les géométries utilisées pour représenter les éléments de type ligne sont des Polylignes (Polyline) et celles utilisées pour représenter les surfaces sont des Polygones (Polygon). La particularité de ces dernières et quelles peuvent contenir plusieurs lignes et plusieurs surfaces.

Dans ESRI, la Polyligne (Polyline) est utilisée pour représenter les cours d’eau. Une Polyligne (Polyline) peut contenir aucune ou plusieurs Lignes (Path). Une Ligne (Path) peut contenir une ou plusieurs Droites (Segment). Une Droite (Segment) contient deux Sommets (Point). Un Sommet (Point) contient une coordonnée (X et Y), peut contenir une élévation (Z) et peut contenir une Mesure (M). Il n’y a pas de sens de numérisation à respecter pour une Polyligne à part celle que l’on veut bien y donner (Exemple : sens d’écoulement des eaux).

Dans ESRI, le Polygone (Polygon) est utilisé pour représenter les étendues d’eau. Un Polygone (Polygon) peut contenir aucun ou plusieurs Anneaux extérieurs (Exterior Ring). Un Anneau extérieur (Exterior Ring) peut contenir aucun ou plusieurs Anneaux intérieurs (Interior Ring). Un Anneau extérieur ou intérieur (Ring) est une Ligne (Path) qui est toujours fermée. Un Anneau (Ring) peut contenir trois au moins trois Droites (Segment). Une Droite (Segment) contient deux Sommets (Point). Un Sommet (Point) contient une coordonnée (X et Y), peut contenir une élévation (Z) et peut contenir une Mesure (M). Le sens de numérisation est une condition à respecter dans la définition d’un Polygone. Les anneaux extérieurs sont numérisés dans le sens horaire et les anneaux intérieurs sont numérisés dans le sens antihoraire.

Afin de bien comprendre le traitement de squelettisation, un parallèle sera effectué entre une même géométrie de type Polyligne et une même géométrie de type Polygon. On verra donc ci-dessous les différents sous traitements et facteurs pouvant influencer le résultat de la squelettisation pour chacun des deux types de géométrie.

|  |  |
| --- | --- |
| Polyligne contenant une ligne avec 16 droites (17 sommets). | Polygone contenant un anneau extérieur avec 43 droites (44 sommets) et un anneau intérieur avec 14 droites (15 sommets). |
|  |  |

## Points de connexion avec les éléments en relation

L’utilisation des points de connexion est un facteur important dans le traitement de squelettisation car le résultat obtenu pourrait être différent pour une même géométrie qui contient ou ne contient pas de points de connexion. Afin de s’assurer que la connexion entre le squelette et les éléments en relation soit respectée, il est obligatoire de fournir les points de connexion en paramètre d’entrée du traitement.

Le traitement de squelettisation utilise la topologie pour extraire les points de connexion avec les éléments en relation. Les classes d’éléments en relation sont donc exigées pour construire cette topologie. Le traitement extrait les géométries des **Nodes** de la topologie et calcul le centre de la géométrie des **Edges** de la topologie pour définir les points de connexion des éléments en relation.

|  |  |
| --- | --- |
| Polyligne avec quatre éléments en relation -  Deux points de connexion. | Polygone avec deux éléments en relation -  Deux points de connexion. |
|  |  |

## Enveloppe des géométries et densification

Lorsque le traitement de squelettisation est extérieur, gauche ou droite aux géométries à traiter, une enveloppe de ces dernières et un traitement de densification est exigé. L’enveloppe est calculée à partir des géométries à traiter et une distance d’agrandissement de cette dernière. L’enveloppe est fusionnée aux géométries à traiter et une densification selon une distance minimale est exigée afin d’obtenir un bon résultat.

Pour un traitement de squelettisation intérieur pour les surfaces, l’ajout d’une enveloppe aux géométries à traiter n’est pas réalisé car il est inutile. Seul le traitement de densification est effectué afin d’obtenir un meilleur résultat mais n’est pas obligatoire.

|  |  |
| --- | --- |
| Polyligne contenant une ligne extérieure (49 sommets) et une ligne intérieure (22 sommets) avec densification au 20 mètres et une distance d’agrandissement de 40 mètres pour la ligne de l’enveloppe. | Polygone contenant deux anneaux extérieurs (57 sommets et 19 sommets) et un anneau intérieur (65 sommets) avec densification au 20 mètres et une distance d’agrandissement de 40 mètres pour l’anneau de l’enveloppe. |
|  |  |

## Triangulation de Delaunay

Le traitement de triangulation selon Delaunay est très exigeant en terme de calcul. Le calcul de la triangulation ne sera pas expliqué ici puisqu’il est difficile à expliquer et à comprendre mais plusieurs algorithmes et fonctions de ce dernier sont présents sur le Web gratuitement.

|  |  |
| --- | --- |
| Triangles de Delaunay de la Polyligne. | Triangles de Delaunay du Polygone. |
|  |  |

La fonction utilisée nécessitait cependant plusieurs adaptations. Ces adaptations sont les suivantes :

1. Extraire tous les sommets des géométries à traiter.
2. Éliminer la duplication de ces sommets selon une précision.
3. Transformer les triangles résultants en droites pour chaque côté d’un triangle.
4. Fusionner toutes les droites dans une Polyligne.
5. Éliminer les droites dupliquées dans la Polyligne.
6. Éliminer les droites à l’extérieur de l’enveloppe des géométries à traiter.
7. Éliminer les droites qui superposent la Polyligne ou la limite du Polygone.
8. Transformer chaque droite en Polyligne.
9. Ajouter ces Polylignes dans une géométrie globale nommée « GeometryBag ».

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay selon l’enveloppe de la Polyligne. | GeometryBag contenant les droites des triangles de Delaunay selon l’enveloppe du Polygone. |
|  |  |

L’utilisation du Diagramme de Voronoi est une autre technique pour créer des squelettes. Son traitement est un dérivé du traitement de la Triangulation de Delaunay. On sait que des outils existent mais on ne connait pas tous les détails. C’est d’ailleurs ces lectures qui m’ont inspiré pour trouver les différentes méthodes démontrées dans ce document. Ma première action a été de développer un outil avec la technologie ESRI qui utilise le diagramme de Voronoi pour créer un squelette. Le résultat obtenu est bon mais une autre idée qui me semblait plus viable et qui me permettait d’effectuer de la généralisation m’est apparue. Une méthode de généralisation pourrait être développée avec l’utilisation du Diagramme de Voronoi mais tout est à faire et à penser. Les explications de ce traitement sont cependant absentes de ce document.

## Squelettisation à droite d’une Polyligne et à l’intérieur d’un Polygone

Le traitement de squelettisation pour une Polyligne est semblable à celui d’un Polygone. Quelques petites différentes sont cependant présentent comme par exemple dans le nom des traitements. On parle de squelettisation intérieure et extérieure pour un Polygone et d’une squelettisation à droite et à gauche pour une Polyligne. De plus, on ajoute toujours l’enveloppe pour effectuer le traitement de squelettisation d’une Polyligne.

### Sommets manquants

La toute première étape du traitement de squelettisation est de s’assurer que chaque droite des triangles de Delaunay est connectée à un sommet de la Polyligne ou du Polygone. S’il n’y a pas de connexion, il est primordial d’insérer un sommet à la Polyligne ou du Polygone afin d’avoir une connexion parfaite.

|  |  |
| --- | --- |
| Un sommet est manquant sur la Polyligne. Le sommet doit être inséré sur la Polyligne afin d’avoir une connexion parfaite entre la Polyligne et les droites des Triangles de Delaunay. | Un sommet est manquant sur le Polygone. Le sommet doit être inséré sur le Polygone afin d’avoir une connexion parfaite entre le Polygone et les droites des Triangles de Delaunay. |
|  |  |

### Droites des triangles de Delaunay

Comme son nom le dit, la deuxième étape du traitement de squelettisation à droite d’une Polyligne est d’extraire les droites des triangles de Delaunay qui sont à droite de la Polyligne et de les ajouter dans un GeometryBag. Pour les droites situées aux points d’extrémité de la Polyligne, un angle de 180 degrés de l’angle des extrémités de droites de la Polyligne est utilisé pour discriminer celles à droites de celles à gauche.

Comme son nom le dit aussi, la deuxième étape du traitement de squelettisation intérieure d’un Polygone est d’extraire les droites de Delaunay qui sont à l’intérieure du Polygone et de les ajouter dans un GeometryBag.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay qui sont à la **droite** de la Polyligne. | GeometryBag des droites des triangles de Delaunay qui sont à **l’intérieure** du Polygone. |
|  |  |

### Droites significatives des triangles de Delaunay

La troisième étape du traitement de squelettisation à droite est d’extraire seulement les droites significatives des triangles de Delaunay à droite de la Polyligne et de les ajouter dans un GeometryBag. Les droites significatives sont celles dont les deux extrémités touchent la Polyligne.

La troisième étape pour le traitement de squelettisation intérieure d’un Polygone est inutile car toutes les droites seront toujours significatives.

|  |  |
| --- | --- |
| GeometryBag contenant les droites significatives des triangles de Delaunay à **droite** de la Polyligne. | GeometryBag contenant les droites significatives des triangles de Delaunay à **l’intérieur** du Polygone. |
|  |  |

### Squelette primaire

La quatrième étape de la squelettisation est de créer le squelette primaire. Normalement on devrait partir du GeometryBag contenant les droites significatives des triangles de Delaunay mais afin de bien comprendre la différence entre le squelette primaire et celui de base, on va partir du GeometryBag contenant toutes les droites des triangles de Delaunay qui sont à la droite de la Polyligne. Comme mentionné précédemment, pour les Polygones cela ne fait aucune différence.

La création du squelette primaire se fait en traitant chaque sommet de la Polyligne ou du Polygone en ordre séquentiel. Pour chaque sommet, il faut aussi traiter séquentiellement chaque droite connectée au sommet traité. Pour chaque droite traitée, un sommet correspondant au centre de la droite est ajouté à une Polyligne. Ce traitement doit se faire de façon séparée pour chaque ligne d’une Polyligne et pour chaque anneau d’un Polygone.

|  |  |
| --- | --- |
| Polyligne contenant le squelette primaire à **droite** de la Polyligne. | Polyligne contenant le squelette primaire à **l’intérieur** du Polygone. |
|  |  |

### Squelette de base

Comme on peut le constater dans les dessins précédents, des triangles sont présents dans le squelette primaire. L’idée est de remplacer ces triangles par des lignes et conserver les connexions existantes dans le squelette. Plusieurs tests ont été réalisés et le résultat qui me semblait le plus proche de la réalité était d’éliminer la plus longue des droites de chaque triangle présent dans le squelette primaire. Le résultat permet d’obtenir le squelette de base tel que montré ci-dessous.

|  |  |
| --- | --- |
| Polyligne contenant le squelette de base à **droite** de la Polyligne. | Polyligne contenant le squelette de base à **l’intérieur** du Polygone. |
|  |  |

### Squelette de base significatif

Si au départ on avait créé le squelette primaire à partir du GeometryBag contenant les droites significatives des triangles de Delaunay qui sont à la droite de la Polyligne, dans ce cas-ci, aucun triangle dans le squelette primaire n’aurait été présent. Le squelette de base aurait été le même que le squelette primaire. Le squelette significatif est différent du complet car seuls les parties significatives du squelette sont présentes. Les autres parties du squelette n’ont aucun impacte dans un traitement de généralisation.

|  |  |
| --- | --- |
| Polyligne contenant le squelette de base significatif à **droite** de la Polyligne. | Polyligne contenant le squelette de base significatif à **l’intérieur** du Polygone. |
|  |  |

### Squelette minimal

Comme mentionné précédemment, afin de satisfaire les besoins du plus grand nombre d’usagers, un autre facteur a été considéré, soit celui de la quantité des lignes à conserver dans le squelette. Est-ce que toutes les lignes du squelette sont nécessaires pour les besoins de tous les usagers ? La réponse est probablement NON. Alors la solution retenue est sensiblement la même que celle utilisé pour traiter les dimensions minimales des longueurs de ligne. Une longueur de ligne est donc requise en paramètre d’entrée dans le traitement de squelettisation.

Pour le squelette d’une Polyligne, la grande différence c’est que la longueur est vérifiée par rapport à la longueur totale de chaque sous ensemble de lignes connectées.

Pour le squelette d’un Polygone, la grande différence avec le traitement de longueur des lignes c’est que seules les lignes qui touchent la limite du Polygone sont traitées et que la longueur de la première droite de cette ligne est soustraite avant d’effectuer la vérification. Cela a pour but de compenser pour les Polygones dont la largeur des surfaces est très grande.

|  |  |
| --- | --- |
| Polyligne contenant le squelette minimal à **droite** de la Polyligne selon une longueur minimale de 20 mètres. | Polyligne contenant le squelette minimal à **l’intérieur** du Polygone selon une longueur minimale de 20 mètres. |
|  |  |

### Squelette minimal avec points de connexion

Comme mentionné précédemment, l’identification des points de connexion des éléments en relation est un autre facteur important dans le résultat du traitement de la squelettisation.

Pour une Polyligne, comme on le constate dans le dessin précédent, même si aucun élément en relation n’est présent, des points de connexion correspondants aux extrémités de la ligne sont toujours identifiés. En fait, il n’y a pas de traitement de connexion entre les points de connexion et le squelette. Les points de connexion sont utilisés seulement lors du traitement de la généralisation. Le traitement de connexion des extrémités non connectées du squelette à la Polyligne est toujours réalisé.

Pour un Polygone, l’identification des points de connexions pour le traitement de squelettisation intérieure est vraiment importante car elle permet de conserver la connexion du squelette avec les éléments en relation. Deux traitements de connexion sont réalisés, soit celui de la connexion des points de connexion au squelette et celui des extrémités non connectées du squelette au limite du Polygone.

Plusieurs tests ont été effectués pour réussir correctement à connecter les points de connexion au squelette. La méthode retenue se fait en deux temps, soit dans un premier temps à partir du calcul de la droite moyenne des deux droites du Polygone du point connecté et de son extension au squelette et dans un deuxième temps, s’il n’y a pas de connexion possible, une droite est créée à partir du point de connexion et du plus proche point calculé au squelette.

Pour le traitement de connexion d’une extrémité non connectée du squelette, une droite est créée à partir de l’extrémité jusqu’au point le plus proche de la limite du polygone.

|  |  |
| --- | --- |
| Polyligne contenant le squelette minimal à **droite** de la Polyligne selon une longueur minimale de 20 mètres et des points de connexion des éléments en relation. | Polyligne contenant le squelette minimal à **l’intérieur** du Polygone selon une longueur minimale de 20 mètres et des points de connexion des éléments en relation. |
|  |  |

## Squelettisation à gauche d’une Polyligne et à l’extérieure d’un Polygone

Le traitement de squelettisation à gauche pour une Polyligne est identique à celui à droite puisque c’est exactement le même traitement. La seule différence c’est qu’avant d’effectuer le traitement de squelettisation, le sens de numérisation de cette Polyligne est inversé. Tout comme la squelettisation à droite, on ajoute toujours l’enveloppe pour effectuer le traitement de squelettisation d’une Polyligne.

Le traitement de squelettisation extérieure pour un Polygone est un peu différent de celui de squelettisation intérieure. L’enveloppe est ajoutée au Polygone de façon à ce que les anneaux de ce dernier soient inversés. Ainsi, l’enveloppe sera un anneau extérieur et les anneaux du Polygone seront tous inversés. Les anneaux extérieurs deviendront intérieurs et les anneaux intérieurs deviendront extérieurs.

### Sommets manquants

Comme déjà mentionné, la toute première étape du traitement de squelettisation est de s’assurer que chaque droite des triangles de Delaunay est connectée à un sommet de la Polyligne ou du Polygone. S’il n’y a pas de connexion, il est primordial d’insérer un sommet à la Polyligne ou du Polygone afin d’avoir une connexion parfaite.

|  |  |
| --- | --- |
| Un sommet est manquant sur la Polyligne. Le sommet doit être inséré sur la Polyligne afin d’avoir une connexion parfaite entre la Polyligne et les droites des Triangles de Delaunay. | Un sommet est manquant sur le Polygone. Le sommet doit être inséré sur le Polygone afin d’avoir une connexion parfaite entre le Polygone et les droites des Triangles de Delaunay. |
|  |  |

### Droites des triangles de Delaunay

Comme son nom le dit, la deuxième étape du traitement de squelettisation à gauche d’une Polyligne est d’extraire les droites des triangles de Delaunay qui sont à gauche de la Polyligne et de les ajouter dans un GeometryBag. Pour les droites situées aux points d’extrémité de la Polyligne, un angle de 180 degrés de l’angle des extrémités de droites de la Polyligne est utilisé pour discriminer celles à droites de celles à gauche.

Comme son nom le dit aussi, la deuxième étape du traitement de squelettisation extérieure d’un Polygone est d’extraire les droites de Delaunay qui sont à l’extérieure du Polygone et de les ajouter dans un GeometryBag.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay à **gauche** de la Polyligne. | GeometryBag contenant les droites des triangles de Delaunay à **l’extérieur** du Polygone. |
|  |  |

### Droites significatives des triangles de Delaunay

La troisième étape du traitement de squelettisation à gauche est d’extraire seulement les droites significatives des triangles de Delaunay à gauche de la Polyligne et de les ajouter dans un GeometryBag. Les droites significatives sont celles dont les deux extrémités touchent la Polyligne.

La troisième étape pour le traitement de squelettisation extérieure est d’extraire seulement les droites significatives des triangles de Delaunay à l’extérieur du Polygone et de les ajouter dans un GeometryBag. Les droites significatives sont celles dont les deux extrémités touchent le Polygone.

|  |  |
| --- | --- |
| GeometryBag contenant les droites significatives des triangles de Delaunay à **gauche** de la Polyligne. | GeometryBag contenant les droites significatives des triangles de Delaunay à **l’extérieur** du Polygone. |
|  |  |

### Squelette primaire

Comme mentionné auparavant, la quatrième étape de la squelettisation est de créer le squelette primaire. Normalement on devrait partir du GeometryBag contenant les droites significatives des triangles de Delaunay mais afin de bien comprendre la différence entre le squelette primaire et celui de base, on va partir du GeometryBag contenant toutes les droites des triangles de Delaunay qui sont à la gauche de la Polyligne ou à l’extérieur du Polygone.

La création du squelette primaire se fait en traitant chaque sommet de la Polyligne ou du Polygone en ordre séquentiel. Pour chaque sommet, il faut aussi traiter séquentiellement chaque droite connectée au sommet traité. Pour chaque droite traitée, un sommet correspondant au centre de la droite est ajouté à une Polyligne. Ce traitement doit se faire de façon séparée pour chaque ligne d’une Polyligne et pour chaque anneau d’un Polygone.

|  |  |
| --- | --- |
| Polyligne contenant le squelette primaire à **gauche** de la Polyligne. | Polyligne contenant le squelette primaire à **l’extérieur** du Polygone. |
|  |  |

### Squelette de base

Comme on peut le constater dans les dessins précédents, des triangles sont présents dans le squelette primaire. L’idée est de remplacer ces triangles par des lignes et conserver les connexions existantes dans le squelette. Comme déjà mentionné, le résultat qui me semblait le plus proche de la réalité était d’éliminer la plus longue des droites de chaque triangle présent dans le squelette primaire. Le résultat permet d’obtenir le squelette de base tel que montré ci-dessous.

|  |  |
| --- | --- |
| Polyligne contenant le squelette de base à gauche de la Polyligne. | Polyligne contenant le squelette de base à l’extérieur du Polygone. |
|  |  |

### Squelette de base significatif

Si au départ on avait créé le squelette primaire à partir va partir du GeometryBag contenant seulement les droites significatives des triangles de Delaunay qui sont à la gauche de la Polyligne ou à l’extérieur du Polygone, dans ce cas-ci, aucun triangle dans le squelette primaire n’aurait été présent dans le squelette primaire de la Polyligne. Par contre, pour le Polygone, il y aurait eu des triangles dans le squelette primaire. Le squelette de base aurait été le même que le squelette primaire de la Polyligne mais pas pour le Polygone. Le squelette significatif est différent du complet car seuls les parties significatives du squelette sont présentes. Les autres parties du squelette n’ont aucun impacte dans un traitement de généralisation. Rappelons-nous que les squelettes à droite, à gauche et extérieurs sont utiles seulement pour effectuer la généralisation. Les squelettes intérieurs d’un Polygone sont utiles aussi pour la généralisation mais le sont également pour définir les lignes de centre d’un phénomène cartographique.

|  |  |
| --- | --- |
| Polyligne contenant le squelette de base significatif à **gauche** de la Polyligne. | Polyligne contenant le squelette de base significatif à **l’extérieur** du Polygone. |
|  |  |

### Squelette minimal

Comme déjà mentionné, afin de satisfaire les besoins du plus grand nombre d’usagers, un autre facteur a été considéré, soit celui de la quantité des lignes à conserver dans le squelette. Est-ce que toutes les lignes du squelette sont nécessaires pour les besoins de tous les usagers ? La réponse est probablement NON. Alors la solution retenue est sensiblement la même que celle utilisé pour traiter les dimensions minimales des longueurs de ligne. Une longueur de ligne est donc requise en paramètre d’entrée dans le traitement de squelettisation.

Pour le squelette d’une Polyligne, la grande différence c’est que la longueur est vérifiée par rapport à la longueur totale de chaque sous ensemble de lignes connectées.

Pour le squelette d’un Polygone, la grande différence avec le traitement de longueur des lignes c’est que seules les lignes qui touchent la limite du Polygone sont traitées et que la longueur de la première droite de cette ligne est soustraite avant d’effectuer la vérification. Cela a pour but de compenser pour les Polygones dont la largeur des surfaces est très grande.

|  |  |
| --- | --- |
| Polyligne contenant le squelette minimal à **gauche** de la Polyligne selon une longueur minimale de 20 mètres. | Polyligne contenant le squelette minimal à **l’extérieur** du Polygone selon une longueur minimale de 20 mètres. |
|  |  |

Il n’y a pas de traitement de connexion des points de connexion au squelette pour la squelettisation d’une Polyligne ou la squelettisation extérieure d’un Polygone. Par contre, il y a toujours un traitement de connexion entre les lignes non connectées du squelette à la Polyligne ou au limite du Polygone.

# Généralisation

Pour bien des personnes, le terme « généralisation » est très vague et n’a pas nécessairement la même signification pour tous. Cependant, ce qui semble ressortir c’est que ce sont des traitements qui permettent de diminuer le niveau de détails des données spatiales. Plusieurs méthodes et techniques peuvent être utilisées pour réaliser la généralisation cartographique de façon automatique. Encore aujourd’hui, pour la plupart des organisations, il est présentement impossible d’effectuer cette généralisation complètement automatique. Les outils nécessaires pour le faire sont absents ou incomplets. Un nouvel outil est maintenant disponible pour traiter la largeur et la longueur de généralisation de Polygones et de Polylignes. Le traitement est très complexe et coûteux en temps d’exécution mais donne de bons résultats. En utilisant l’ensemble des différents outils, méthodes et techniques présentent dans ce document, deux jeux de données au 1:250 000 contenant les étendues et les cours d’eau ont été généralisés presque complètement automatique à partir des données de la BDG (+/- 1:50 000). Il ne reste que l’outil pour exécuter l’ensemble des traitements selon une séquence établie qui reste à développer.

## Largeur et longueur de généralisation d’un Polygone

Comme déjà expliqué auparavant dans ce document, les traitements de généralisation d’un Polygone sont des traitements de validation et de correction des dimensions minimales d’un Polygone selon une échelle de représentation. Le traitement de largeur et longueur de généralisation d’un Polygone est un traitement assez complexe à expliquer et à comprendre. Pour ces raisons, nous verrons ci-dessous les différentes étapes qui y sont reliées.

### Prétraitement

Dans un premier temps, il est bien important de comprendre qu’un traitement de généralisation consiste en un ensemble de plusieurs traitements de dimension minimale. La séquence de ces traitements est très importante car le résultat obtenu peut être assez différent selon la séquence choisie. Comme on peut le constater dans le dessin ci-dessous, plus la géométrie d’un élément contient d’information, plus d’information sera traitée, plus de temps de traitement sera nécessaire et plus d’information sera présent dans le résultat. Le but ultime de la généralisation est d’éliminer l’information superflue ou inutile en fonction de l’échelle de représentation désirée. Avant d’effectuer le traitement de largeur et longueur de généralisation, il est donc fortement conseiller d’éliminer le maximum d’information superflue ou inutile car ce traitement demande énormément de calcul. Il est conseiller d’éliminer le maximum de sommets, de droites et d’anneaux.

|  |  |
| --- | --- |
| **Avant** : Aucun filtre des sommets, aucun filtre des droites et aucun filtre des anneaux n’ont été effectué. | **Après** : Un filtre des sommets (1.5 mètre), un filtre des droites (3 mètres) et un filtre des anneaux (1000 mètres) ont été effectué. |
|  |  |

### Traitement pour l’Intérieur et l’extérieur d’un Polygone

Il y a deux traitements de largeur et longueur de généralisation qui sont possibles et qui sont exigés pour obtenir un résultat adéquat. Il s’agit du traitement pour **l’intérieur** et pour **l’extérieur** du Polygone. Afin de bien comprendre la différence entre les deux et de voir leur impacte, un parallèle entre les deux sera utilisé dans les différentes explications présentées ci-dessous.

Comme déjà mentionné, le traitement de largeur et de longueur de généralisation est inspiré de la même technique pour créer les squelettes. Ce traitement a donc besoin d’identifier les points de connexion, les droites des triangles de Delaunay, le squelette primaire, de base et minimale et les deux types de droites de connexion.

En fait, en raison de la caractéristique pour un Polygone d’être une géométrie fermée, le traitement de largeur et longueur de généralisation **intérieure** permet de **minimiser** le Polygone tandis que le traitement de largeur et longueur de généralisation **extérieure** permet **d’exagérer** le Polygone.

### Points de connexion

Comme présenté dans les dessins ci-dessous, le traitement de largeur et longueur de généralisation **intérieure** permet d’identifier les points de connexion des éléments en relation. Il utilise la topologie pour identifier ces derniers. Le traitement de largeur et longueur de généralisation **extérieure** ne permet pas d’identifier les points de connexion car le Polygone peut être exagéré et est inutile dans ce contexte.

|  |  |
| --- | --- |
| Deux points de connexion sont utilisés pour le traitement de largeur et longueur de généralisation **intérieure.** | Aucun point de connexion n’est utilisé pour le traitement de largeur et longueur de généralisation **extérieure.** |
|  |  |

### Droites des triangles de Delaunay

Comme pour la création des squelettes, le traitement de largeur et longueur de généralisation **intérieure** a besoin d’identifier dans un GeometryBag les droites des triangles de Delaunay qui sont à l’intérieure du Polygone tandis que le traitement de largeur et longueur de généralisation **extérieure** a besoin d’identifier dans un GeometryBag les droites des triangles de Delaunay qui sont à l’extérieure du Polygone.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay qui sont à l’**intérieure** du Polygone. | GeometryBag contenant les droites des triangles de Delaunay qui sont à l’**extérieure** du Polygone. |
|  |  |

### Largeur de généralisation

La largeur de généralisation est utilisée pour identifier toutes les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à cette dernière. Ces droites sont ajoutées dans un GeometryBag pour chacun des traitements de largeur et longueur de généralisation **intérieur** ou **extérieur**.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation (Longueur de chaque droite <= 25 mètres) et qui sont à l’**intérieure** du Polygone. | GeometryBag contenant les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation (Longueur de chaque droite <= 25 mètres) et qui sont à l’**extérieure** du Polygone. |
|  |  |

### Longueur de généralisation

À cette étape, il faut trouver les parties dont la largeur de l’intérieur ou l’extérieur d’un Polygone est inférieure ou égale la largeur minimale de généralisation et dont la longueur de ces parties est supérieure à la longueur minimale de généralisation.

À partir du GeometryBag des droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation et qui sont à l’**intérieure** ou à **l’extérieure** du Polygone, pour chaque sommet du Polygone en lien avec une droite, une ligne est créée à partir du centre de chaque série de droites consécutives et si cette ligne est plus grande ou égale à la longueur minimale de généralisation, une erreur de largeur et longueur de généralisation est détectée. Ces lignes calculées correspondent à des parties du squelette de base de l’**intérieure** ou **l’extérieure** du Polygone.

|  |  |
| --- | --- |
| Polyline correspondant à des parties du squelette dont la largeur de **l’intérieur** d’un Polygone est inférieure ou égale la largeur minimale de généralisation **intérieure** et dont la longueur de ces parties est supérieure ou égale à la longueur minimale de généralisation **intérieure**.  (Longueur chaque ligne d’erreur > 50 mètres). | Polyline correspondant à des parties du squelette dont la largeur de **l’extérieur** d’un Polygone est inférieure ou égale la largeur minimale de généralisation **extérieure** et dont la longueur de ces parties est supérieure ou égale à la longueur minimale de généralisation **extérieure**.  (Longueur chaque ligne d’erreur > 50 mètres). |
|  |  |

### Détection des erreurs

Lorsqu’une erreur de largeur et longueur de généralisation est détectée, tous les sommets du Polygone en lien avec les droites de Delaunay qui intersectent les lignes en erreur seront déplacés ou ajoutés selon le centre de la droite en relation. Lorsque plusieurs droites en erreur touchent un même sommet du Polygone, le premier sommet est déplacé et les suivants sont ajoutés. Le nouveau Polygone ainsi remodelé contiendra des parties invalides (les parties où les limites se superposent). Comme on peut le constater dans les dessins ci-dessous, lorsque ces parties invalides sont éliminées du Polygone (Opérateur spatial : Simplifié), cela peut faire en sorte que le Polygone résultant peut contenir plusieurs plus petits Polygones ou anneaux dérivés de l’original.

|  |  |
| --- | --- |
| Correction du Polygone selon les lignes d’erreurs de largeur et longueur de généralisation détectées.    Avant le traitement de filtrage des anneaux selon la superficie minimale d’un anneau de la généralisation intérieure (Anneaux > 0 mètres). | Correction du Polygone selon les lignes d’erreurs de largeur et longueur de généralisation détectées.  Avant le traitement de filtrage des anneaux selon la superficie minimale d’un anneau de la généralisation extérieure (Anneaux > 0 mètres). |
|  |  |

### Filtrage des anneaux

Après la correction des erreurs de largeur et longueur de généralisation, un filtrage des anneaux selon la superficie minimale des anneaux **intérieurs** ou **extérieurs** est effectué afin d’éliminer les anneaux superflus ou trop petits.

Le squelette **intérieur** du Polygone initial pourra être utilisé pour calculer le squelette du nouveau Polygone généralisé et les lignes d’erreurs de largeur et longueur de généralisation.

Le squelette **extérieur** du Polygone initial pourra être utilisé pour calculer les lignes d’erreurs de largeur et longueur de généralisation seulement.

|  |  |
| --- | --- |
| Après le traitement de filtrage des anneaux selon la superficie minimale d’un anneau de la généralisation intérieure (Anneaux > 3 600 mètres). | Après le traitement de filtrage des anneaux selon la superficie minimale d’un anneau de la généralisation extérieure (Anneaux > 3600 mètres). |
|  |  |

### Polygone généralisé

Le résultat du Polygone généralisé selon la largeur et longueur de généralisation **intérieur** peut faire en sorte que ce dernier sera déconnecté de ses éléments en relation. Dans ce cas-ci, un traitement de reconnexion sera nécessaire.

Pour le résultat du Polygone généralisé selon la largeur et longueur de généralisation **extérieur,** il n’y a pas de déconnexion avec ses éléments en relation. Cependant certaines parties de ces éléments en relation pourraient se retrouver à l’intérieure du Polygone. Ces parties devront être éliminées mais ce n’est pas le cas ici.

|  |  |
| --- | --- |
| Résultat du Polygone généralisé selon la largeur et longueur de généralisation **intérieure** et selon la superficie minimale des anneaux **intérieurs** et **extérieurs**. | Résultat du Polygone généralisé selon la largeur et longueur de généralisation **extérieure** et selon la superficie minimale des anneaux **intérieurs** et **extérieurs**. |
|  |  |

### Squelette du Polygone généralisé

Pour les étendues d’eau et puisque qu’elle est exigée, la création du nouveau squelette pour le Polygone généralisé devra se faire.

Pour le Polygone généralisé selon la largeur et longueur de généralisation **intérieure**, le squelette sera créé à partir de l’intersection (opération spatiale) entre le squelette **intérieur** du Polygone initial et le Polygone généralisé.

Pour le Polygone généralisé selon la largeur et longueur de généralisation **extérieure**, le squelette sera créé à partir du traitement de squelettisation **intérieure** du Polygone généralisé en tenant compte de ses éléments en relation.

|  |  |
| --- | --- |
| Création du nouveau squelette **intérieur** du Polygone généralisé selon la largeur et longueur de généralisation **intérieure.** | Création du nouveau squelette **intérieur** du Polygone généralisé selon la largeur et longueur de généralisation **extérieure.** |
|  |  |

### Lignes d’erreurs de largeur et longueur de généralisation

Si les lignes d’erreurs de largeur et longueur de généralisation doivent être identifiées et puisque certains anneaux trop petits ont été éliminés, le calcul de ces nouvelles lignes doivent être réalisées.

Pour le Polygone généralisé selon la largeur et longueur de généralisation **intérieure**, les lignes d’erreurs seront créées à partir de la différence (opération spatiale) entre le squelette **intérieur** du Polygone initial et le Polygone généralisé. Afin de conserver la connexion entre le Polygone généralisé et ses éléments en relation. Les lignes d’erreurs seront utilisées pour créer les nouveaux cours d’eau de connexion.

Pour le Polygone généralisé selon la largeur et longueur de généralisation **extérieure**, les lignes d’erreurs seront créées à partir de l’intersection (opération spatiale) entre le squelette **extérieur** du Polygone initial et le Polygone généralisé.

|  |  |
| --- | --- |
| Création des lignes d’erreurs de largeur et longueur de généralisation **intérieure.**  Remplacement d’une partie de l’étendue d’eau en cours d’eau à partir des lignes d’erreurs de largeur et longueur de généralisation **intérieure.** | Création des lignes d’erreurs de largeur et longueur de généralisation **extérieure.** |
|  |  |

### Traitements complémentaires

Les différences obtenues entre les résultats du traitement de largeur et longueur de généralisation sont assez marquantes mais dans les deux cas, la géométrie est simplifiée. Selon les besoins des usagers, on peut soit vouloir soit minimiser ou exagérer une géométrie. Dans les deux cas, ces deux traitements sont complémentaires, c’est-à-dire que pour effectuer un traitement de largeur et longueur de généralisation complet, il faut effectuer les deux **Intérieur/extérieur** ou **extérieur/intérieur**.

Comme on peut le constater dans les dessins ci-dessous, il n’y a pas beaucoup de différence dans ce cas-ci après le deuxième traitement mais dans certains cas, il peut y en avoir beaucoup. Il ne faut donc pas négliger ce deuxième traitement.

Il est aussi important de noter que la création du squelette intérieur devrait se faire après l’exécution du deuxième traitement afin que le squelette du résultat final représente bien le centre du Polygone.

|  |  |
| --- | --- |
| Résultat du traitement de largeur et longueur de généralisation **extérieure** sur le résultat du traitement de largeur et longueur de généralisation **intérieure** :  **Trois petites modifications effectuées.** | Résultat du traitement de largeur et longueur de généralisation **intérieure** sur le résultat du traitement de largeur et longueur de généralisation **extérieure** :  **Aucune modification effectuée.** |
|  |  |

### Largeur et longueur de généralisation entre les Polygones

Selon les besoins des usagers, on peut aussi vouloir effectuer le traitement de largeur et longueur de généralisation **extérieure** entre les éléments. C’est un autre facteur important à considérer dans la séquence des divers traitements de généralisation à effectuer.

|  |
| --- |
| Résultat du traitement de largeur et longueur de généralisation **extérieure** entre plusieurs éléments distincts. Ce traitement est facilement réalisable pour les Polygones mais très difficilement réalisable pour les Polylignes. Le sens de numérisation des Polygones est connus et constants ce qui n’est pas le cas pour les Polylignes. |
|  |

## Largeur et longueur de généralisation d’une Polyligne

Les traitements de généralisation d’un Polyligne sont des traitements de validation et de correction des dimensions minimales d’une Polyligne selon une échelle de représentation. Le traitement de largeur et longueur de généralisation d’une Polyligne est un traitement assez complexe à expliquer et à comprendre. Pour ces raisons, nous verrons ci-dessous les différentes étapes qui y sont reliées.

### Prétraitement

Pour les mêmes raisons que les Polygones, dans un premier temps, il est bien important de comprendre qu’un traitement de généralisation consiste en un ensemble de plusieurs traitements de dimension minimale. La séquence de ces traitements est très importante car le résultat obtenu peut être assez différent selon la séquence choisie. Comme on peut le constater dans le dessin ci-dessous, plus la géométrie d’un élément contient d’information, plus d’information sera traitée, plus de temps de traitement sera nécessaire et plus d’information sera présent dans le résultat. Le but ultime de la généralisation est d’éliminer l’information superflue ou inutile en fonction de l’échelle de représentation désirée. Avant d’effectuer le traitement de largeur et longueur de généralisation, il est donc fortement conseiller d’éliminer le maximum d’information superflue ou inutile car ce traitement demande énormément de calcul. Il est conseiller d’éliminer le maximum de sommets, de droites et de lignes.

|  |  |
| --- | --- |
| **Avant :** Aucun filtre des sommets, aucun filtre des droites et aucun filtre des lignes n’ont été effectué. | **Après :** Un filtre des sommets (1.5 mètre), un filtre des droites (3 mètres) et un filtre des lignes (250 mètres) ont été effectué. |
|  |  |

### Traitement à droite et à gauche d’une Polyligne

Il y a deux traitements de largeur et longueur de généralisation qui sont possibles et qui sont exigés pour obtenir un résultat adéquat. Il s’agit du traitement à **droite** et à **gauche** d’une Polyligne. Afin de bien comprendre la différence entre les deux et de voir leur impacte, un parallèle entre les deux sera utilisé dans les différentes explications présentées ci-dessous.

Comme déjà mentionné, le traitement de largeur et de longueur de généralisation est inspiré de la même technique pour créer les squelettes. Ce traitement a donc besoin d’identifier les points de connexion, les droites des triangles de Delaunay et le squelette primaire, de base et minimale.

Contrairement au Polygone, puisque la Polyligne n’est pas une géométrie fermée, on ne peut pas exagérer une Polyligne. On peut seulement minimiser cette dernière.

### Points de connexion

Comme présenté dans les dessins ci-dessous, le traitement de largeur et longueur de généralisation à **droite** et à **gauche** de la Polyligne permet d’identifier les points de connexion des éléments en relation. Ces points de connexion permettent de s’assurer que le résultat de la Polyligne généralisée sera encore connecté à ses éléments en relation.

|  |  |
| --- | --- |
| Deux points de connexion sont utilisés pour le traitement de largeur et longueur de généralisation à **droite** de la Polyligne. | Deux points de connexion sont utilisés pour le traitement de largeur et longueur de généralisation à **gauche** de la Polyligne. |
|  |  |

### Droites des triangles de Delaunay

Comme pour la création des squelettes, le traitement de largeur et longueur de généralisation à **droite** de la Polyligne a besoin d’identifier dans un GeometryBag les droites des triangles de Delaunay qui sont à **droite** de la Polyligne tandis que le traitement de largeur et longueur de généralisation à **gauche** de la Polyligne a besoin d’identifier dans un GeometryBag les droites des triangles de Delaunay qui sont à **gauche** de la Polyligne.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay qui sont à **droite** de la Polyligne. | GeometryBag contenant les droites des triangles de Delaunay qui sont à **gauche** de la Polyligne. |
|  |  |

### Largeur de généralisation

La largeur de généralisation est utilisée pour identifier toutes les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à cette dernière. Ces droites sont ajoutées dans un GeometryBag pour chacun des traitements de largeur et longueur de généralisation à **droite** ou à **gauche**.

|  |  |
| --- | --- |
| GeometryBag contenant les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation (Longueur de chaque droite <= 50 mètres) et qui sont à **droite** de la Polyligne. | GeometryBag contenant les droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation (Longueur de chaque droite <= 50 mètres) et qui sont à **gauche** de la Polyligne. |
|  |  |

### Longueur de généralisation

À cette étape, il faut trouver les parties dont la largeur à **droite** ou à **gauche** d’une Polyligne est inférieure ou égale la largeur minimale de généralisation et dont la longueur de ces parties est supérieure à la longueur minimale de généralisation.

À partir du GeometryBag des droites des triangles de Delaunay dont leur longueur est inférieure ou égale à la largeur de généralisation et qui sont à **droite** ou à **gauche** de la Polyligne, pour chaque sommet de la Polyligne en lien avec une droite, une ligne est créée à partir du centre de chaque série de droites consécutives et si cette ligne est plus grande ou égale à la longueur minimale de généralisation, une erreur de largeur et longueur de généralisation est détectée. Ces lignes calculées correspondent à des parties du squelette de base **à gauche** ou **à droite** de la Polyligne.

|  |  |
| --- | --- |
| Polyline correspondant à des parties du squelette dont la largeur à droite d’une Polyligne est inférieure ou égale la largeur minimale de généralisation à **droite** et dont la longueur de ces parties est supérieure ou égale à la longueur minimale de généralisation à **droite**. | Polyline correspondant à des parties du squelette dont la largeur à droite d’une Polyligne est inférieure ou égale la largeur minimale de généralisation à **gauche** et dont la longueur de ces parties est supérieure ou égale à la longueur minimale de généralisation à **gauche**. |
|  |  |

### Détection des erreurs

Lorsqu’une erreur de largeur et longueur de généralisation à **droite** ou à **gauche** d’une Polyligne est détectée, tous les sommets de la Polyligne en lien avec les droites de Delaunay qui intersectent les lignes en erreur seront déplacés ou ajoutés selon le centre de la droite en relation. Lorsque plusieurs droites en erreur touchent un même sommet de la Polyligne, le premier sommet est déplacé et les suivants sont ajoutés. La nouvelle Polyligne ainsi remodelée contiendra des parties invalides (les parties où les lignes déplacées ou ajoutées se superposent). Comme on peut le constater dans les dessins ci-dessous, lorsque ces parties invalides sont corrigées de la Polyligne (Opérateur spatial : Simplifié), cela peut faire en sorte que la Polyligne résultante peut contenir plusieurs plus petites lignes dérivées de l’original.

|  |  |
| --- | --- |
| Correction de la Polyligne selon les lignes d’erreurs de largeur et longueur de généralisation à **droite** détectées.    Avant le traitement de filtrage des lignes selon la longueur minimale d’une ligne de la généralisation (Lignes > 0 mètres). | Correction de la Polyligne selon les lignes d’erreurs de largeur et longueur de généralisation à **gauche** détectées.    Avant le traitement de filtrage des lignes selon la longueur minimale d’une ligne de la généralisation (Lignes > 0 mètres). |
|  |  |

### Filtrage des lignes

Après la correction des erreurs de largeur et longueur de généralisation d’une Polyligne, un filtrage des lignes selon la longueur minimale des lignes à **droite** ou à **gauche** est effectué afin d’éliminer les lignes superflues ou trop petites.

Les lignes éliminées suite au filtrage des lignes correspondent toujours à l’ensemble ou à une partie des lignes d’erreurs de largeur et longueur de généralisation de la Polyligne.

Les lignes d’erreurs de largeur et longueur de généralisation de Polyligne correspondent toujours à des parties du squelette.

|  |  |
| --- | --- |
| Après le traitement de filtrage des lignes selon la longueur minimale d’une ligne de la généralisation à **droite** d’une Polyligne (Lignes > 250 mètres). | Après le traitement de filtrage des lignes selon la longueur minimale d’une ligne de la généralisation à **gauche** d’une Polyligne (Lignes > 250 mètres). |
|  |  |

### Polyligne généralisée

Le résultat de la Polyligne généralisé selon la largeur et longueur de généralisation à **droite** ou à **gauche** ne peut jamais faire en sorte que la Polyligne sera déconnectée de ses éléments en relation.

Les extrémités originales de la Polyligne seront aussi toujours conservées.

|  |  |
| --- | --- |
| Résultat de la Polyligne généralisée selon la largeur et longueur de généralisation à **droite** et selon la longueur minimale des lignes. | Résultat de la Polyligne généralisée selon la largeur et longueur de généralisation à **gauche** et selon la longueur minimale des lignes. |
|  |  |

### Squelette de la Polyligne généralisée

Pour les cours d’eau, aucun squelette n’est exigé dans le résultat de largeur et longueur de généralisation à gauche ou à droite. Le seul besoin connu est au niveau de la compréhension de ce traitement.

|  |  |
| --- | --- |
| Polyligne contenant le squelette significatif de la Polyligne généralisée selon la largeur et la longueur de généralisation à **droite** de la Polyligne. | Polyligne contenant le squelette significatif de la Polyligne généralisée selon la largeur et la longueur de généralisation à **gauche** de la Polyligne. |
|  |  |

### Lignes d’erreurs de largeur et longueur de généralisation

Contrairement au Polygone, les lignes d’erreurs de largeur et longueur de généralisation à **droite** ou à **gauche** de la Polyligne identifiées lors du traitement restent les mêmes tout au long du traitement. Ces dernières sont utilisées pour identifiées les endroits où la Polyligne a été modifiée.

Les lignes d’erreurs de largeur et longueur de généralisation à **droite** ou à **gauche** de la Polyligne correspondent à l’ensemble ou à une partie du squelette significatif à **droite** ou à **gauche** de la Polyligne.

|  |  |
| --- | --- |
| Création des lignes d’erreurs de largeur et longueur de généralisation à **droite** de la Polyligne. | Création des lignes d’erreurs de largeur et longueur de généralisation à **gauche** de la Polyligne. |
|  |  |

### Traitements complémentaires

Les différences obtenues entre les résultats du traitement de largeur et longueur de généralisation à **droite** et à **gauche** de la Polyligne sont assez marquantes mais dans les deux cas, la géométrie est simplifiée et minimisée mais jamais exagérée comme pour les Polygones. Dans les deux cas, ces deux traitements sont complémentaires, c’est-à-dire que pour effectuer un traitement de largeur et longueur de généralisation complet, il faut effectuer les deux **droite/gauche** ou **gauche/droite**.

Comme on peut le constater dans les dessins ci-dessous, il n’y a pas beaucoup de différence dans ce cas-ci après le deuxième traitement mais dans certains cas, il peut y en avoir beaucoup. Il ne faut donc pas négliger ce deuxième traitement.

|  |  |
| --- | --- |
| Résultat du traitement de largeur et longueur de généralisation à **gauche** sans le filtre selon la longueur minimale des lignes sur le résultat du traitement de largeur et longueur de généralisation à **droite** :  **Deux petites modifications effectuées.** | Résultat du traitement de largeur et longueur de généralisation à **droite** sans le filtre selon la longueur minimale des lignes sur le résultat du traitement de largeur et longueur de généralisation à **gauche** :  **Deux petites modifications effectuées.** |
|  |  |

### Largeur et longueur de généralisation entre les Polylignes

Selon les besoins des usagers et comme pour les Polygones, on peut aussi vouloir effectuer le traitement de largeur et longueur de généralisation entre les éléments de type ligne. C’est un autre facteur important qui serait normalement à considérer dans la séquence des divers traitements de généralisation à effectuer. Cependant, ce traitement est beaucoup plus difficile à réaliser puisqu’il n’y a pas de règle pour le sens de numérisation des lignes. Cette fonctionnalité n’a pas été développée et ne sera pas expliquée dans ce document puisqu’aucun cas semblable n’a été trouvé dans les données de la BDG.

Même si un sens de numérisation des lignes est demandé, comme selon le sens d’écoulement des eaux, il peut être difficile de faire les liens entre les éléments pour les traitements à **droite** ou à **gauche** de la Polyligne et de traiter plus de deux éléments à la fois.

|  |
| --- |
| Problématique du traitement de largeur et longueur de généralisation à **droite** ou à **gauche** entre plusieurs d’éléments distincts de type ligne. Sachant qu’il n’y a pas de sens de numérisation connu : Quel est le lien entre les traitements à droite et à gauche ? Quelles sont les priorités entre les lignes ? Peut-on traiter plus de lignes à la fois ? Est-ce que ces cas sont plausibles dans nos données ? |
|  |

### Différences dans les résultats

Les différences obtenues entre les traitements de largeur et longueur de généralisation **droite/gauche** et **gauche/droite** incluant le filtre selon la longueur minimale des lignes (Lignes > 250 mètres) sont assez marquantes. Selon la séquence choisie, la Polyligne généralisée obtenue se situe soit plus à droite ou plus à gauche de la Polyligne originale.

Pour contourner ce problème et comme expliqué un peu plus bas, le traitement de largeur et longueur de généralisation selon une Polyligne fractionnée permet d’obtenir une Polyligne généralisée plus au centre de la Polyligne originale.

|  |  |
| --- | --- |
| Résultat du traitement de largeur et longueur de généralisation à **droite/gauche** de la Polyligne avec un filtre selon la longueur minimale des lignes (Lignes > 250 mètres) par rapport à la Polyligne originale. | Résultat du traitement de largeur et longueur de généralisation à **gauche/droite** de la Polyligneavec un filtre selon la longueur minimale des lignes (Lignes > 250 mètres) par rapport à la Polyligne originale. |
|  |  |

## Largeur et longueur de généralisation d’une Polyligne fractionnée

Le traitement de largeur et longueur de généralisation d’une Polyligne fractionnée permet d’obtenir une Polyligne généralisée plus au centre de la Polyligne originale. L’idée est d’effectuer le traitement à **droite** et à **gauche** de la Polyligne en même temps pour chaque partie significative de droite et de gauche.

### Fractionner une Polyligne

Comme déjà expliqué, il n’y a pas de différence entre le traitement à droite ou à gauche d’une Polyligne, ce n’est que le sens de numérisation de la Polyligne qui change. Le fractionnement d’une Polyligne permet de séparer la Polyligne originale en une Polyligne contenant plusieurs petites lignes correspondant chacune à la partie significative du traitement **à droite** et **à gauche** de la Polyligne. Il y a donc alternance entre le sens de numérisation pour chaque petite ligne consécutive. Le traitement de largeur et longueur de généralisation est effectué pour chaque petite ligne en utilisant leurs extrémités comme points de connexion.

|  |
| --- |
| Polyligne fractionnée contenant plusieurs petites lignes. Les lignes verte sont utilisées pour effectuer le traitement à droite et les lignes rouge sont utilisées pour le traitement à gauche. |
|  |

### Droites des triangles de Delaunay

Dans un premier temps, un GeometryBag contenant les droites des triangles de Delaunay à droite et à gauche de chaque petite ligne de la Polyligne fractionnée est créé.

|  |
| --- |
| GeometryBag contenant les droites des triangles de Delaunay à droite et à gauche de la Polyligne fractionnée. |
|  |

### Lignes d’erreurs

Dans un deuxième temps, à partir du GeometryBag des droites de Delaunay à droite et à gauche de la Polyligne fractionnée, une Polyligne contenant les squelettes significatifs de chaque petite ligne est créée. À partir de la Polyligne des squelettes significatifs, une Polyligne contenant les lignes d’erreurs détectées dans le traitement de largeur et longueur de généralisation de la Polyligne fractionnée est créée.

|  |
| --- |
| Polyligne contenant les lignes d’erreurs du traitement de largeur et longueur de généralisation de la Polyligne fractionnée. |
|  |

### Polyligne fractionnée et généralisée

On obtient ainsi le résultat du traitement de largeur et longueur de généralisation de la Polyligne fractionnée incluant le filtre selon la longueur des lignes. Comme on le constate dans le dernier dessin, la Polyligne fractionnée et généralisée est beaucoup plus au centre de la Polyligne originale. Afin d’obtenir un bon résultat, plusieurs traitements itératifs peuvent être nécessaires.

|  |
| --- |
| Résultat du traitement de largeur et longueur de généralisation de la Polyligne fractionnée incluant le filtre selon les longueur des lignes (Lignes > 250 mètres). |
|  |

# Processus de généralisation des étendues et des cours d’eau

Présentement, tous les outils de généralisation sont accessibles via une barre de menu (Barre de généralisation) dans ArcMap. Tous les outils de cette barre de menu peuvent traiter un ou plusieurs éléments. On ne peut pas qualifier ces outils comme des outils automatiques puisqu’il faut sélectionner les éléments à traiter, définir les paramètres et lancer la commande. Cependant, on peut plutôt les qualifier d’outils semi-automatiques. Ces outils ne peuvent pas être utilisés dans des Géotraitements ou via le gestionnaire de processus « ModelBuilder » de ESRI car ils n’ont pas été conçus pour ça. Il existe d’autres méthodes plus performante que les Géotraitements pour traiter les processus de généralisation. L’outil de validation des contraintes spatiales est un bon exemple de ce que cette prétention. Le prochain développement majeur sera justement consacré à développer ce genre d’outil automatique de gestion des processus et des traitements.

Comme déjà mentionné dans le document, la généralisation de deux classes d’entités (étendue et cours d’eau) pour deux jeux de données (021L et 031P) à l’échelle du 1:250000 à partir des données de la BDG qui sont à l’échelle d’environ 1:50000 a été réalisé avec succès. Plusieurs séquences de traitement d’outils différents ont été testés pour en arriver à un résultat acceptable. Cette séquence de traitement est itérative à certains endroits et elle peut être différente selon le résultat désiré. Cette séquence est donc seulement à titre indicative. À noter également que la séquence peut être différente de celle proposée dépendamment si on tient compte ou non de l’état des données d’origine.

|  |
| --- |
| Barre de menu contenant les outils de squelettisation et de généralisation.  Voir la documentation « **BarreGeneralisation.doc** » de cette barre de menu pour plus de détails. |
|  |

## Extraction des données spatiales

Cette étape consiste à extraire les étendues et les cours d’eau à partir des données de la BDG. À partir d’ArcMap, il faut utiliser l’outil « **Extract data** » de la barre de menu ESRI « **Distributed Geodatabase** ». Cet outil permet d’extraire les données exactes de la BDG et de les insérer dans une « **File Geodatabase** » de façon très performante. De plus elle permet de créer un projet contenant les données extraites. Il est important de mentionner qu’il faut extraire une partie plus grande de la partie à généraliser afin d’obtenir un bon résultat. Il est conseiller d’extraire au moins deux à trois fois la largeur de généralisation. Si la largeur de généralisation est 125 mètres, on pourrait par exemple extraire les données selon un tampon de 300 mètres autour du jeu de données à l’échelle du 1:250000.

## Projection des données spatiales

Cette étape consiste à transformer les données géographiques « **4617 : GCS\_North\_American\_1983\_CSRS**» des étendues et des cours d’eau extraites de la BDG dans un système de coordonnées projeté. La projection à utiliser dans ESRI est « **3979 : NAD\_1983\_CSRS\_Canada\_Atlas\_Lambert**». L’outil utilisé pour projeter les données spatiales dans ESRI est le géotraitement « **FeatureClassToFeatureClass** ». On utilise ce système de coordonnées projeté puisque les valeurs des dimensions minimales à utiliser pour effectuer la généralisation sont en mètre et que l’imprécision de cette référence spatiale est négligeable par rapport à la précision des données généralisées.

## Corriger la proximité des sommets

Cette étape consiste à s’assurer que tous les sommets de toutes les classes d’entité spécifiées qui se retrouvent à l’intérieur de la tolérance minimale de proximité soient déplacés afin d’être superposés. Toutes les droites ou lignes ou surfaces d’une longueur de zéro mètre sont détruites. Le but étant de rétablir une connexion parfaite entre les éléments à condition que toutes les classes d’entités possèdent la même référence spatiale et la même précision. Elle permet également de corriger l’ajustement des éléments à la limite des découpages d’entité du 1:50000 si la tolérance spécifiée est assez grande.

## Éliminer les lignes superposées

Cette étape consiste à détruire tous les cours d’eau superposés à d’autres cours d’eau ou aux limites des étendues d’eau. Cette situation peut se produire lors de la correction de la proximité des sommets des étendues et des cours d’eau.

## Filtrer les sommets des lignes et des surfaces

Cette étape consiste à détruire tous les sommets superflus de toutes les classes d’entité spécifiées selon une distance latérale minimale entre les sommets. Ce traitement conserve la connectivité entre les éléments en relation. En éliminant les sommets superflus, on diminue la grosseur des classes d’entités ce qui a normalement un impact important sur les temps des traitements restants.

## Filtrer les droites des lignes et des surfaces

Cette étape consiste à détruire toutes les droites superflues de toutes les classes d’entité spécifiées selon une longueur minimale des droites. Ce traitement conserve la connectivité entre les éléments en relation. En éliminant les droites superflues, on diminue la grosseur des classes d’entités ce qui a normalement un impact important sur les temps des traitements restants.

## Éliminer les squelettes existants dans les surfaces

Cette étape consiste à détruire les cours d’eau qui sont à l’intérieur des étendues d’eau et qui correspondent à des squelettes de surfaces. Dans la BDG, les cours d’eau possédant la valeur « **2:Inféré** » de l’attribut « **NETWORK\_FLOW\_TYPE** » sont considérés comme des squelettes. Les squelettes des étendues d’eau seront créés à nouveau vers la fin du processus selon le résultat obtenu de la généralisation.

## Fusion des lignes et des surfaces adjacentes

Cette étape consiste à éliminer la segmentation causée par le découpage d’entité entre les éléments adjacents d’une même classe. Les algorithmes utilisés pour généraliser donnent de meilleurs résultats lorsque la géométrie complète est fournie. Le résultat obtenu de la généralisation sur une partie de géométrie peut quand même être bon mais ce n’est pas une garantie. Le problème lié à la fusion des éléments adjacents est que plus un élément est gros, plus il prend du temps à traiter.

## Filtrer les anneaux intérieurs

Cette étape consiste à détruire les anneaux intérieurs qui sont superflus dans les étendues d’eau selon une superficie minimale pour les anneaux intérieurs. Plus les éléments sont petits et moins détaillés, plus les temps de traitement sont courts. On veut seulement conserver les anneaux intérieurs pour lesquels il est important de conserver ce détail.

## Adoucir les lignes et les limites des surfaces

Cette étape consiste à détruire des détails trop prononcés contenus dans une géométrie. C’est une forme de pré-généralisation des lignes et des limites des surfaces. Pour certains algorithmes de généralisation, plus il y a de détails prononcés, plus le temps de traitement est long. Donc, en diminuant ces détails, on diminue les temps de traitements.

## Généraliser l’intérieure et l’extérieure des surfaces

Cette étape consiste à généraliser l’intérieur et l’extérieur des étendues d’eau selon une largeur et une longueur de généralisation des surfaces et selon une superficie minimale des anneaux intérieurs et extérieurs. Ce traitement permet de transformer les parties ou la totalité des étendues d’eau en cours d’eau tout en conservant la connectivité entre les éléments en relation.

## Filtrer les lignes

Cette étape consiste à détruire les cours d’eau superflus selon une longueur minimale des lignes. Moins il y a d’éléments à traiter, moins les temps des traitements restants sont longs. Seuls les cours d’eau non-connectés à au moins une extrémité peuvent être détruits.

## Filtrer les lignes fermées

Cette étape consiste à détruire les cours d’eau superflues connectés à ses deux extrémités selon une longueur minimale des lignes fermées. Moins il y a d’éléments à traiter, moins les temps des traitements restants sont long. Seuls les cours d’eau connectés à ses deux extrémités avec d’autres cours d’eau ou des étendues d’eau peuvent être détruits. Seule la partie la plus longue des cours d’eau formant une ligne fermée sera détruite.

## Généraliser la partie gauche et la partie droite des lignes

Cette étape consiste à généraliser la partie gauche et la partie droite des cours d’eau selon une largeur et une longueur de généralisation des lignes et selon une longueur minimale des lignes. Ce traitement permet de détruire les parties des cours d’eau trop prononcées tout en conservant la connectivité entre les éléments en relation. Selon les besoins, on pourrait choisir de généraliser selon les lignes fragmentées.

## Créer les squelettes des surfaces

Cette étape consiste à créer les cours d’eau qui sont à l’intérieur des étendues d’eau généralisés et qui correspondent à des squelettes de surfaces. Dans la BDG, les cours d’eau possédant la valeur « **2:Inféré** » de l’attribut « **NETWORK\_FLOW\_TYPE** » sont considérés comme des squelettes. Les cours d’eau créés respectent la connexion avec les autres cours d’eau connectés aux étendues d’eau généralisés et respectent la longueur minimale des lignes spécifiée. Les valeurs par défaut des attributs contenus dans les cours d’eau créés dépendent des valeurs par défaut de chaque attribut de la classe d’entité des cours d’eau.

## Séparer les géométries multiples

Cette étape consiste à s’assurer que chaque élément des étendues et des cours d’eau contienne seulement une géométrie. Ainsi lorsqu’une étendue d’eau possède plusieurs anneaux extérieurs, le même élément sera dupliqué pour chaque polygone correspondant aux anneaux extérieurs. Pour les cours d’eau possédants plusieurs lignes, l’élément correspondant sera dupliqué pour chaque ligne.

## Valider les géométries des éléments

Cette étape consiste à s’assurer que les géométries des étendues et des cours d’eau soient valides. Les géométries ne doivent pas être vides et ne doivent pas contenir de superposition et de croisements. En fait, ils doivent respecter les normes standard de l’OGC.

## Transfert et destruction des attributs d’origine

Cette étape consiste à transférer certaines valeurs d’attributs d’un ancien élément à un nouvel élément ou à détruire certains autres attributs selon différents paramètres de traitement. Cette façon de faire dépend des décisions à prendre pour chaque valeur d’attribut des classes d’entité d’étendues et de cours d’eau. Dans le prototype testé, aucune décision n’avait été prise à ce sujet. De plus, aucun outil n’est présent pour traiter ces cas.

|  |
| --- |
| **Important** : Le processus peut nécessité une itération entre les étapes 3 à 14 inclusivement. |
|  |

# Conclusion

Beaucoup de sujets ont été discutés dans ce document. On a débuté par une certaine mise en contexte dans laquelle on retrouvait une description générale des différentes données présentes à RNCAN et une description générale des données contenues dans la BDG. Une description plus détaillée des étendues et des cours d’eau de la BDG a été présenté puisque ces dernières ont été utilisé pour décrire les différents traitements de généralisation. Une description générale des différentes contraintes d’intégrité spatiales et attributives utilisées pour les données de la BDG a aussi été présentée. L’importance de connaître l’état des données avant de pouvoir effectuer des traitements de généralisation a également été expliqué.

Par la suite, on a commencé à entrer dans le cœur du sujet. Une description générale des différents concepts ou sous-traitements utilisées dans les différents traitements de généralisation a été présentée. On a décrit brièvement la topologie, les relations spatiales, les opérateurs spatiaux et on a terminé avec les dimensions minimales. On a expliqué que le résultat des traitements de généralisation était en fait le résultat de la correction des contraintes spatiales des différentes valeurs de dimension minimale.

On a poursuivi en expliquant de façon assez détaillée les deux principaux traitements de généralisation soit la squelettisation et la généralisation des géométries de type ligne et surface. On a expliqué que les traitements de squelettisation pouvaient être utilisés pour plusieurs circonstances. On peut les utiliser pour identifier le centre d’une géométrie de type surface, pour identifier les erreurs des contraintes spatiales de dimension minimale et pour généraliser des géométries de type ligne et surface. On a aussi expliqué que certains traitements de généralisation utilisaient les traitements de squelettisation pour pouvoir changer la forme des géométries de type ligne et surface afin de respecter les valeurs de dimension minimale d’une nouvelle échelle de représentation visée. Plusieurs exemples sous forme d’illustration de la généralisation des étendues et des cours d’eau d’une échelle (1:50000) vers une autre plus petite (1:250000) ont été montrés.

On a terminé par une description générale d’une proposition de processus pour pouvoir généraliser des étendues et des cours d’eau d’une échelle (1:50000) vers une autre plus petite (1:250000). On a expliqué que le processus proposé pouvait facilement être différent selon l’état des données au départ selon le résultat qu’on voulait obtenir. On a expliqué que la séquence des traitements était très importante car le résultat pouvait grandement être différent. On se souvient par exemple que le résultat du traitement de largeur et longueur de généralisation **intérieur/extérieur** est très différent de celui **extérieur/intérieur**.

En résumé, on peut conclure en disant que les traitements de généralisation sont complexes et difficiles à comprendre et à expliquer. Ils exigent de grand calcul ce qui demande beaucoup de temps de traitement. Cependant, il a été clairement démontré par la production de deux jeux de données au 1:250000 que les résultats et les temps de traitement obtenus sont tout à fait acceptable et réalisable. En mettant un peu d’effort et d’énergie, on pourrait raisonnablement mentionner qu’on pourrait améliorer les résultats et les temps de traitements de façon modérée.