

Département de géomatique appliquée
Faculté des lettres et sciences humaines
Université de Sherbrooke

Enrique Martin de Villodres Lusinchi

Développement d'une base de données géospatiales et mise en place d'un tableau de bord analytique pour la gestion et la diffusion des données de mammifères marins de la Patagonie.

Essai présenté pour l'obtention du grade de Maître ès sciences géographiques (M.Sc.),
Cheminement de type cours en géodeveloppement durable.

Avril 2021

© Enrique Martin de Villodres Lusinchi, 2021

Directeurs de l'essai :

Développement d'une base de données géospatiales pour la gestion et la diffusion des données de mammifères marins de la Patagonie

Présenté par : Enrique Martin de Villodres Lusinchi

Cet essai a été évalué par:

Mickaël Germain, Directeur

(Département de géomatique appliquée, Faculté des lettres et sciences humaines, Université De Sherbrooke)

Résumé :

Parmi les mammifères marins de la Patagonie, il y a des espèces qui ne sont pas ou très peu étudiées et par conséquent, le niveau de menace qui les guette est inconnu. Pour approfondir la connaissance de ces espèces maritimes de la Patagonie chilienne, il existe une grande quantité de données que sont issues d'études effectuées avec différentes méthodologies à différentes échelles, et avec différents objectifs. Par conséquent, les données existantes sont hétérogènes, redondantes ainsi qu'imbriquées. Pour exploiter les nombreuses données disponibles, il faut mettre en place un système de gestion de bases de données (SGBD) ainsi qu'un système d'information géographique (SIG) web.

Compte tenu l'hétérogénéité des données, et pour faire suite à une analyse préliminaire des différents jeux de données et de leurs relations, le choix du SGBD s'est porté sur un système NoSQL de type orienté document. Le logiciel libre MongoDB a été utilisé pour créer, gérer et interroger la base de données. Le SGBD ainsi mis en place a servi d'alimentation au SIG web pour l'analyse et la présentation des résultats.

L'outil d'alimentation sélectionné pour le SIG web, ainsi que pour le SGBD, est l'outil ETL du logiciel FME desktop. Il permet une gestion efficace et conviviale des données géospatiales en provenance de différentes sources, et plus particulièrement une interaction avec le SGBD en place. Cela a permis d'uniformiser l'information dans un délai raisonnable de sorte à pouvoir générer la BD ainsi que de créer des feuilles Google Sheets en tant que couches d'entités, et les héberger en infonuagique.

Au niveau de l'interface client du projet, le système ArcGIS Online a été choisi pour jouer le rôle du serveur cartographique en infonuagique. Les cartographies web générées par cet outil peuvent facilement intégrer le tableau de bord analytique proposé par ESRI : ArcGIS Dashboard. Ce dernier est une interface géodécisionnelle web et gratuite pour une utilisation non commerciale.

Le tableau de bord analytique mis en place répond au besoin d'afficher huit jeux de données contenant plus de 20 ans de données géographiques liées aux mammifères marins. Cette interface permet l'étude de la distribution, de la mobilité et de la taille des groupes de dauphins chiliens ainsi que les autres espèces maritimes de la Patagonie chilienne.

Remerciements

Je remercie Monsieur Mickaël Germain pour avoir accepté d'être mon directeur d'essai. Je lui exprime ma reconnaissance pour ses encouragements ainsi que de tout le soutien fourni. Je remercie également la codirectrice, Madame Liliana Perez, d'avoir proposé le sujet de ce travail, et de m'avoir accepté au sein de son équipe.

Mes remerciements vont aussi à Simon Demers, étudiant aux cycles supérieurs du Département de géographie de l'Université de Montréal, qui a partagé ses connaissances sur ce sujet et a fourni en permanence des informations ainsi qu'un éclairage essentiel au cours de ce travail.

Je tiens également à remercier toutes les personnes qui m'ont permis d'améliorer mes connaissances durant le parcours de ma maîtrise.

Finalement, je tiens à remercier ma conjointe Adriana qui a su être mon soutien moral et m'a constamment soutenu et encouragé durant dans le cadre de ce travail et durant mon parcours académique.

Table des matières

Table des matières.....	i
Listes des figures.....	iv
Listes des tableaux	v
1 Introduction.....	1
1.1 Mise en contexte.....	1
1.2 Problématique.....	2
1.3 Objectifs	3
2 Cadre théorique.....	3
2.1 Les systèmes de gestion de bases de données (SGBD).....	4
2.1.1 Les SGBD relationnelles ou SGBD SQL	4
2.1.2 Les SGBD non relationnelles ou SGBD NoSQL	8
2.1.3 Comparaison entre SGBD SQL et SGBD NoSQL.....	12
2.2 Les composantes logicielles d'un SIG web.....	14
2.2.1 Les types d'architecture logicielle d'un SIG web.....	14
2.2.2 La diffusion de l'information géospatiale.....	15
2.2.3 Les serveurs cartographiques	19
2.2.4 Les interfaces client des SIG web.....	20
2.3 L'extraction, la transformation et le chargement des données.....	23
3 Matériel et Méthodes	25
3.1 Site d'étude.....	25
3.2 Méthodologie	26
3.3 Les données	26
3.4 Le choix entre SQL ou NoSQL.....	27
3.5 Le processus ETL.....	27

Sélection de données	28
3.5.1 L'analyse des besoins	28
3.5.2 L'évaluation des données.....	29
3.5.3 La négociation utilité/qualité	30
3.5.4 Les modélisations conceptuelle, logique et physique	30
3.6 L'extraction et la transformation des données	30
3.7 L'intégration et le chargement des données.....	34
3.8 La mise en place du SIG web.....	35
3.9 Le développement de l'interface web	36
4 Résultats.....	37
4.1 Création d'une collection dans MongoDB.....	37
4.2 Création de couches d'entité (<i>Feature Layer</i>) sur ArcGIS Online.....	38
4.3 La création de cartes web avec ArcGIS Online	40
4.4 Création du tableau de bord à partir d'ArcGIS Online	41
4.4.1 L'utilisation des cartes avec les données	42
4.4.2 La liste des observations affichées.....	43
4.4.3 Les éléments infographiques.....	44
4.4.4 Les indicateurs numériques.....	45
4.4.5 Les filtres d'affichage	46
4.5 Mise au point du tableau de bord	47
5 Interprétation et discussion des résultats.....	49
6 Conclusion	50
7 Recommandations.....	50
Annexe 1 Analyse de besoins (requetés)	61
Annexe 2 : Évaluation séparée des sources données (exemple).....	62

Annexe 3 Modèle Logique.....	63
Annexe 4 Modèle physique	64
Annexe 5 : Tableau de synthèse des transformations requises (exemple).....	65
Annexe 6 Processus ETL de données.	66
Annexe 7 Comme utiliser des photos Google Drive dans des fenêtres contextuelles ArcGIS Online	67
Ajouter le résultat a la configuration d'image des fenêtres contextuelles	67
Annexe 8 : Métadonnée	68

Listes des figures

Figure 1 - BD NoSQL type orienté clé-valeur extrait de Behmo et Travers(2019a).....	9
Figure 2 - BD NoSQL type orienté colonne extrait de Behmo et Travers (2019a)	10
Figure 3 - BD NoSQL type orienté document (JSON) extrait de Behmo et Travers (2019a).....	11
Figure 4 - BD NoSQL type orienté <i>graph</i> extrait de Behmo et Travers (2019a)	12
Figure 5 - Théorème de CAP extrait de Behmo et Travers (2019a).....	13
Figure 6 – Schéma simplifié de l’architecture d’un SIG web (Modifié de Germain 2020)	15
Figure 7 Site d’étude (carte Enrique Martin de Villodres)	25
Figure 8 - Diagramme méthodologique	26
Figure 9 - Organigramme méthodologique pour la sélection des données. Modifié de (Crozat, 2016c)	28
Figure 10 - Processus ETL d’un des jeux de données	31
Figure 11 – Exemple du flux de travail pour la préparation des attributs.....	32
Figure 12 - Exemple du flux de travail pour le calcul de temps	32
Figure 13 - Exemple du flux de travail pour ajouter les noms des espèces	33
Figure 14 - Exemple du flux de travail pour l’adéquation de la syntaxe des coordonnées	34
Figure 15 - Flux de travail pour l’étape d’intégration et de chargement des données	35
Figure 16 - Flux des données à la suite du processus ETL.....	37
Figure 17 - Interface client « Robo3T » de MongoDB.....	38
Figure 18 - Création de couches d'entité (<i>Feature Layer</i>) à partir de Google Sheets.....	39
Figure 19 - diagramme de conception du tableau de bord.....	41
Figure 20 - Aspect final de l’interface	42
Figure 21 - Boutons ajoutés à la carte.....	43
Figure 22 - Listes ajoutées à l’interface	44
Figure 23 - Éléments infographiques du tableau de bord	45
Figure 24 - Indicateurs numériques ajoutés	46
Figure 25 - Filtres d’affichage	47
Figure 26 - Fenêtres contextuelles avec images	48
Figure 27 - Modèle conceptuel_logique UML	63
Figure 29 Modèle physique UML.....	64

Listes des tableaux

Tableau 1 – les normes principales de l'OGC.....	17
Tableau 2 - Comparaison entre MapServer et Geoserver.....	20
Tableau 3 - Échelle utilisée pour évaluer la qualité et l'utilité des données.....	29
Tableau 4 - Évaluation du jeu de données : (le nom du jeu de données).....	30
Tableau 5 - Tableau de synthèse des transformations requises	31
Tableau 6 - Évaluation du jeu de données : <i>Ballena_Jorobada_Estrecho (partie)</i>	62
Tableau 7 - Négociation « Utilité/Qualité » du jeu de données : <i>Ballena_Jorobada_Estrecho (partie)</i>	62
Tableau 8 – Matrice de transformation de données (partie)	65

Abréviations, sigles, etc.

API : de l'anglais *Application Programming Interface* – Interface de programmation (web)

BD : Base de Données

BI : de l'anglais *business intelligence*

CSV : de l'anglais *Comma Separated Values*

DMS : Degrée Minute Seconde

ETL : de l'anglais *Extract, Transform, load*

FME de l'anglais *Feature Manipulation Engine*

FTP (*File Transfer Protocol*)

GeoBI : de l'anglais *geospatial business intelligence*

GIF :de l'anglais *Graphics Interchange Format*

GML : de l'anglais *geography Markup Language*

JSON :de l'anglais *JavaScript Object Notation*

JPEG :de l'anglais *Joint Photographic Experts Group*

KML : de l'anglais *Keyhole Markup Language*

PNG : de l'anglais *Portable Network Graphics*

SGBD : Système de Gestion de Bases de Données

SGBDR : Système de Gestion de Bases de Données Relationnelles

SIG web : système d'information géographique avec une communication via Internet

SIG : Système d'information géographique.

SVG : de l'anglais *Scalable Vector Graphics*

SQL : *Structured Query Langage*

web: de l'anglais *World Wide Web*

XML : de l'anglais *eXtensible Markup Language*

1 Introduction

1.1 Mise en contexte

Les études sur les cétacés de la région naturelle de la Patagonie chilienne sont nombreuses, mais l'information disponible sur les différentes espèces ne donne pas un portrait exact sur leur distribution (Dellabianca et al., 2016). De plus, les études sur les petits cétacés sont rares, car ils sont difficiles à quantifier et à évaluer, alors qu'ils sont susceptibles d'extinction pour plusieurs raisons selon Schipper et al. (2008) :

- ils sont faciles à capturer pour l'alimentation humaine ou pour être utilisés comme appât;
- il existe une grande activité anthropique dans ses habitats;
- ils font l'objet de commercialisation auprès de nouveaux marchés.

Il existe des registres historiques qui donnent une idée claire du niveau de menace pour certaines espèces, mais quelques espèces ne sont pas ou très peu étudiées. Par conséquent, le niveau de menace qui les guette est inconnu (Würsig et al., s. d.).

Parmi les espèces les moins étudiées, le dauphin chilien (*Cephalorhynchus eutropia*) pourrait être une des espèces de petits cétacés la plus à risque. L'absence d'informations appropriées concernant le dauphin chilien représente un inconvénient important et difficile à gérer (Pérez-Alvarez et al., 2016). Actuellement, les chercheurs ne disposent pas d'informations à propos de la population totale de cette espèce de la grandeur des groupes à sa répartition géographique (Viddi & Lescrauwaet, 2005). Aussi, il n'y a pas de certitude quant à leur dynamique sociale, car le niveau d'interaction existant entre les différents groupes de cette espèce est très mal connu. De plus, leur affinité avec des groupes d'autres espèces de cétacés est aussi méconnue. Il reste aussi à établir comment l'activité anthropique, comme le transport maritime, le tourisme, et la pisciculture, en pleine croissance, peut affecter cette espèce (Heinrich et al., 2019; Ribeiro et al., 2007).

L'absence d'informations s'explique aussi par le fait que la plupart des études ont été effectuées dans des secteurs qui ne sont pas assez représentatifs de la distribution des habitats des petits cétacés (Heinrich et al., 2019; Pérez-Alvarez et al., 2016; Ribeiro et al., 2007). Comme les espaces occupés par certaines espèces n'ont pas été étudiés de manière uniforme, il n'est pas rare de trouver des disparités entre les prédictions des modélisations et les observations (Carlucci et al., 2016). Ces disparités sont à l'origine d'importantes questions concernant la distribution et la densité de

population du dauphin chilien, par exemple, notamment dans les canaux patagoniques et les canaux fuégiens qui sont la partie la plus australe de la Patagonie chilienne (Viddi et al., 2016).

Aux disparités qui existent entre les résultats des études antérieures, il faut ajouter que malgré l'existence de publications destinées à la connaissance du portrait de mammifères marins, il existe également beaucoup d'informations qui ne sont pas publiées ou pour lesquelles les données ne sont pas accessibles à travers un système de partage de l'information (Abramson & Gibbons, 2010; J. Capella et al., s. d.; J. J. Capella et al., 2014; Gibbons et al., 2003; Guzman et al., 2020). Pour répondre aux manques d'études sur la distribution, la mobilité et la grandeur des groupes de petits cétacés dans les canaux patagoniques, Jorge Gibbons, de l'Institute de la Patagonie de l'Université de Magellan (*Instituto de la Patagonia de la Universidad de Magallanes*), et Liliana Perez, du Département de géographie de l'Université de Montréal, travaillent sur le suivi de ces espèces. Ce qui nécessite le développement d'un système de partage de l'information à travers la mise en place de deux composantes importantes : un système de gestion de bases de données (SGBD) géospatiales, et un tableau de bord analytique web pour l'affichage et l'analyse des données.

1.2 Problématique

L'équipe de chercheurs issue du partenariat entre l'Université de Magellan et l'Université de Montréal a réalisé des observations de grand intérêt scientifique dans les canaux patagoniques. Bénéficier de ces observations pour la connaissance de la faune marine de la Patagonie chilienne ne sera possible qu'à condition que les données récoltées soient traitées de façon à permettre leur consolidation avec d'autres données. Cette consolidation nécessite une combinaison des données provenant d'études précédentes, ainsi que des données connexes extraites de l'étude de l'activité piscicole. En effet, il existe une importante quantité de données issues de la régulation environnementale imposée par le gouvernement chilien sur l'activité piscicole, et plus particulièrement pour la culture du saumon : salmoniculture. Ces données peuvent apporter des informations précieuses concernant la répartition et la concentration de plusieurs espèces marines de la région naturelle de la Patagonie chilienne. Elles sont toutefois, pour la plupart, non publiées. De plus, le format dans lequel les données seront présentées n'est pas encore connu.

Les données à traiter sont issues d'études effectuées avec différentes méthodologies, à différents moments, par différents acteurs (État, chercheurs, etc.). Ces derniers ont travaillé à différentes échelles avec différents objectifs. Par conséquent, une grande hétérogénéité entre les données

existantes est à prendre en compte. De plus, il n'est pas rare que ces informations soient redondantes ou imbriquées.

Dans le but d'approfondir la connaissance des petits cétacés, et plus particulièrement le dauphin chilien, de la Patagonie chilienne, il n'existe pas de base de données (BD) géospatiales pour exploiter l'ensemble des données disponibles. Les attributs géospatiaux de ces observations sont pourtant d'une grande importance étant donnée l'importance d'effectuer des analyses spatiales et de représenter et/ou visualiser les résultats sous la forme de cartographies. Comment mettre en place un SGBD pour prendre en compte l'ensemble des données? Quelle interface cartographique pourrait-on utiliser pour visualiser et analyser les informations? Ce sont les grandes questions auxquelles que nous allons essayer de répondre dans cet essai.

1.3 Objectifs

L'objectif principal du projet est de mettre en place un système d'information géographique (SIG) web pour faciliter la gestion et la diffusion des données de mammifères marins de la Patagonie. Le SIG sera composé d'un SGBD géospatiales, et d'une interface cartographique web intégrée à un tableau de bord analytique.

Pour ce qui est des objectifs secondaires, il est question de :

- produire les modèles conceptuel, logique et physique de la BD en respectant les caractéristiques principales des données disponibles;
- créer les requêtes principales qui devront être proposées et intégrées dans une interface cartographique web afin d'améliorer la communication des données et l'interaction avec les utilisateurs;
- développer un tableau de bord analytique.

2 Cadre théorique

Pour mettre en place un SIG pour la gestion et l'analyse des données géospatiales du projet, il est important de sélectionner les composantes principales du système : le SGBD, et l'interface client pour visualiser et analyser les données. De plus, il faut définir les différents outils d'interaction entre les composantes du SIG.

Nous allons parcourir dans cette section les différentes solutions proposées dans la littérature afin de nous aider à faire un choix du SGBD, et des différentes composantes d'un SIG web pour la mise en place d'une interface client.

2.1 Les systèmes de gestion de bases de données (SGBD)

Il existe deux grandes familles de systèmes qui sont dits « relationnels » ou « non relationnels » pour la gestion des données. Le choix d'une méthode de gestion va déterminer la liste des logiciels nécessaires pour créer, gérer et interroger la base de données ainsi que la manière d'alimenter le système et de générer les résultats.

Pour faire un choix entre les SGBD non relationnelles, aussi connues sous le nom SGBD NoSQL (Not Only SQL), et les SGBD relationnelles qui, par analogie, sont fréquemment dénommées SGBD SQL (Tannir, 2015), il faut comprendre qu'il n'existe pas une solution meilleure qu'une autre. Le choix découle de la nature des données, de leur volume et de la façon selon laquelle elles seront exploitées. Autrement dit, les SGBD non relationnelles sont des solutions pour les besoins spécifiques de gestion de données, et ne sont pas des solutions de remplacement des SGBD relationnelles (Crozat, s. d.).

2.1.1 Les SGBD relationnelles ou SGBD SQL

Les SGBD relationnelles sont dénommés SGBD SQL, car le SQL (*Structured Query Language*) est le langage de requêtes normalisé pour ce type de SGBD. Dans les SGBD relationnelles, le langage SQL permet d'effectuer différentes séquences de requêtes/opérations (transactions) sur l'information contenue dans différentes tables. Ces tables, aussi appelées *relations*, sont conçues conformément à des normes très strictes. Ces normes, dénommées « Formes Normales », sont les règles qui garantissent la cohérence de l'information contenue dans les différentes tables ainsi que les types d'associations qu'il est possible d'établir entre ces tables (Cornuéjols, 2009; Elmasri & Navathe, 2016).

Dans les BD relationnelles, l'information est répartie sur plusieurs tables. Si une BD était définie dans une seule table, le résultat aurait été une très grande quantité de données répétées (redondance de données). Cette hypothétique BD impliquerait de lentes recherches mêmes pour obtenir des résultats de quelques lignes. Cette situation aussi signifierait de longues et coûteuses mises à jour et de potentielles pertes d'informations. La gestion de cette BD serait très difficile, voire impossible. Pour éviter ces types de problèmes, la distribution de l'information sur plusieurs tables

s'impose. La méthode à suivre pour créer les différentes tables d'une BD est l'application des formes normales (Elmasri & Navathe, 2016; Teorey et al., 2011). Ces normes définissent les tables et ses attributs (colonne) et, par conséquent, le schéma de relation qui est l'ensemble des attributs et de leurs domaines (type de données) des différentes tables d'une BD relationnelle. La création d'une BD en accord aux formes normales est connue comme la normalisation (Cornuéjols, 2009; Coronel et al., 2020; Elmasri & Navathe, 2016; Teorey et al., 2011)

Les objectifs de la normalisation sont :

- éviter la redondance des données dans la BD;
- garantir que chaque table contient uniquement les attributs associés à l'entité représentée.

Tout processus de normalisation d'une BD implique l'application préliminaire des contraintes:

- **Contraintes d'intégrité des données :**
 - a) Contrainte d'unicité : chaque ligne d'une table doit avoir un identifiant ou clé primaire unique et cette dernière ne doit pas être *NULL* (vide).
 - b) Contrainte d'intégrité référentielle : dans les cas où il y a un lien entre deux tables, il doit y avoir dans l'une des deux tables une clé étrangère qui indique quelle ligne d'une table est liée à quelle(s) ligne(s) d'une autre table.
 - c) Contrainte de domaine : la valeur d'un attribut dans une table doit appartenir à un domaine de valeurs prédéfinies (*data type*).
- **Contrainte de non-redondance des données :**
 - a) Redondance directe : une donnée ne peut se trouver qu'une fois dans la même table.
 - b) Redondance indirecte : dans une table, il ne peut pas y avoir de données qui peuvent être calculées à partir d'autres données dans la même table.

Dans le but d'établir les critères de comparaison entre les BD SQL et le BD NoSQL, il serait suffisant décrire que les trois premières des cinq formes normales, car dans la plupart des cas, l'application de ces premières trois formes garantissent la non-redondance des données. L'application des autres formes normales est moins fréquente (Cornuéjols, 2009; Coronel et al., 2020; Garcia-Molina et al., 2002; Garmany et al., 2005).

Première forme normale (1NF) : une table est sous sa première forme normale (1NF) si, et seulement si, l'ensemble des colonnes ne contient que des valeurs uniques (atomiques), c'est-à-dire que chaque colonne ne peut avoir qu'une seule valeur par ligne de la table.

Deuxième forme normale (2NF) : une table est sous sa deuxième forme normale (2NF) si respectant la première forme normale, la clé primaire sert à identifier toutes les colonnes qui contiennent des attributs non-clé.

Troisième forme normale (3NF) : une table satisfait la troisième forme normale si elle est en deuxième forme normale et que chaque colonne (attribut) de la table est dépendante de la totalité de la clé primaire, et non dépendante d'une seule partie de la clé.

Toutes ces exigences font que les BD relationnelles sont d'une grande rigidité structurale. Cette rigidité permet

- d'éviter la corruption ou la destruction accidentelle des données lors d'une mise à jour;
- de réduire l'espace de stockage des données (moins de redondance);
- de gérer l'intégrité des données de manière consistante.

Parmi les SGBD relationnelles les plus connus, nous pouvons citer les logiciels suivants : Oracle développé par Oracle, MySQL géré par Oracle, SQL Server développé par Microsoft, PostgreSQL développé par The University of California ou encore DB2 développé par IBM (Behmo & Travers, 2019a; IBM, 2021).

Si la plupart des SGBD relationnelles permettent l'écriture de géométries en profitant des avantages du langage SQL, les SGBD relationnelles n'ont pas été développés pour gérer de l'information géographique. L'utilisation du langage SQL impose la nécessité de bien connaître la définition des géométries pour formuler de requêtes pour un système, mais ces dernières ne seront pas utiles s'il y a un changement de système ou des changements dans la structure des entités spatiales (Ponniah, 2003). En effet, un utilisateur sera dans l'impossibilité d'exécuter des analyses spatiales sans connaître la structure des géométries (Ramsey & Leslie, 2012).

Pour contourner ces limitations, l'OGC (Open Geospatial Consortium) a développé une norme internationale pour la gestion des géométries (SFS – Spatial Features SQL) (*OGC Standards / OGC*, s. d.) qui est intégrée dans les SGBD à référence spatiale (parfois appelés SGBDRS). Cette

norme permet de stocker et de manipuler des données spatiales comme n'importe quelle autre donnée d'une base de données relationnelle en créant de nouveaux types de domaines : Geometry et Geography (Ponniah, 2003; Ramsey & Leslie, 2012).

Un SGBD à référence spatiale doit respecter toutes les caractéristiques et contraintes des SGBD relationnelles, et en plus avoir la capacité de stocker et d'analyser des données référencées géographiquement. Ces SGBD sont adaptés pour la gestion de plusieurs formats de données géographiques, car ils sont dotés de certaines caractéristiques au niveau de contraintes spatiales ainsi que des fonctions et des opérateurs définis dans le langage SQL. Parmi les SGBD à référence spatiale les plus connus, nous pouvons citer les suivants : PostgreSQL/PostGIS, Oracle Spatial et SQL Server (Ramsey & Leslie, 2012).

Malgré les avantages que les règles de normalisation confèrent aux BD relationnelles, ces BD font face à un défi lié au volume de données souvent caractéristiques des données géographiques. Sachant que le fonctionnement optimal des BD relationnelles oblige la mise en place d'un serveur unique, des systèmes informatiques très performants ainsi que des opérateurs experts sont nécessaires pour maintenir l'optimisation du système (Coronel et al., 2011, 2020; Gaetan, 2017). Chaque année des dizaines de milliards de dollars sont investis pour optimiser la gestion des données dans un SGBD SQL (Bédard, 2010). Cependant, l'évolution des technologies de l'information a donné lieu à l'augmentation exponentielle de la demande d'espace de stockage de données depuis la fin des années 90 avec l'apparition du concept de Mégadonnées ou (*Big data*) (Coronel et al., 2020; Cukier & Mayer-Schoenberger, 2013).

Les mégadonnées sont, par définition, un « ensemble d'une très grande quantité de données, structurées ou non, se présentant sous différents formats et en provenance de sources multiples, qui sont collectées, stockées, traitées et analysées dans de courts délais, et qui sont impossibles à gérer avec des outils classiques de gestion de bases de données ou de gestion de l'information. » (Office québécois de la langue française, 2020). « Lorsqu'il est question de mégadonnées, il y a souvent une référence au principe des « trois V »: volume (généralement massif), variété (sources et formats divers) et vitesse (rapidité de traitement). Certains spécialistes ajoutent également la véracité (données crédibles et réelles)» (Office québécois de la langue française, 2020)

Les mégadonnées ont suscité la naissance de nouvelles techniques pour gérer l'information qui est limitée dans les SGBD relationnelles (Behmo & Travers, 2019a). Ces limites soulignent

principalement que les grands volumes de données (*Big data*) nécessitent une autre vision de la gestion de l'information pour que l'augmentation de la taille d'une base de données n'affecte pas sa fonctionnalité et sa performance (mise à l'échelle ou *Scalability*) (Tannir, 2015).

Comme il a été déjà mentionné, le SQL est un langage de haut niveau qui s'appuie sur une structure des données normalisée. Toutefois, pour résoudre le problème de la distribution des données massives, d'autres solutions ont été proposées qui ne reposent pas sur le paradigme relationnel pour stocker et pour interroger les données. Ces solutions sont appelées les SGBD non relationnelles ou SGBD NoSQL.

2.1.2 Les SGBD non relationnelles ou SGBD NoSQL

Les SGBD non relationnelles aussi appelées SGBD NoSQL de l'anglais *Not only SQL* sont des systèmes complémentaires aux SGBD relationnelles qui ont été conçus pour pallier ses limites fonctionnelles sans pour autant les remplacer. Le terme NoSQL a été créé en 2009 pour désigner les systèmes dénormalisés (Tannir, 2015). Contrairement aux SGBD SQL qui nécessitent un serveur unique pour un fonctionnement optimal, les SGBD NoSQL peuvent partitionner les données (*sharding*) sur plusieurs serveurs (Gaetan, 2017).

Grâce au principe de dénormalisation, il est possible de gérer dans l'information sous la forme de colonne, de document ou de graphe sans respecter les contraintes de redondance ou de la rigidité structurale définie précédemment (*MongoDB Manual*, 2020). Les SGBD NoSQL acceptent que des données soient redondantes et que l'information soit imbriquée (non atomique). Ils acceptent aussi la possibilité d'ajouter dans une collection unique des données provenant de différentes sources, et avec différents critères de domaine (non-respect de la contrainte de domaine). La dénormalisation permet d'utiliser les BD NoSQL sans faire des modifications complexes même si après sa mise en place l'utilisateur prend la décision d'incorporer une nouvelle information qui n'était pas prévue au moment de créer la BD (Garmany et al., 2005). Les SGBD NoSQL n'ont pas la rigueur de la modélisation des SGBD relationnelles, car l'information n'est pas distribuée entre différentes tables.

Dans la littérature, les quatre types de BD NoSQL les plus répandus sont les suivants selon (Näsholm, 2012; Tannir, 2015) :

- le type orienté clé-valeur;

- le type orienté colonne;
- le type orienté document;
- le type orienté *graph*.

Bases de données type orienté clé-valeur : ce sont les bases de données avec l'écriture la plus simple, toutes les données associées à une clé sont contenues par la valeur. La clé identifie la valeur qui contient n'importe quel type de données sans préciser le domaine. Étant donné que ce type de BD n'a pas de structure ni de schéma, les seules opérations SQL possibles sont les opérations de base (*CREATE, READ, UPDATE, DELETE*) (Behmo & Travers, 2019a; IBM, 2021; Martin, 1983).

La figure 1 illustre le schéma des BD NoSQL type orienté clé-valeur.

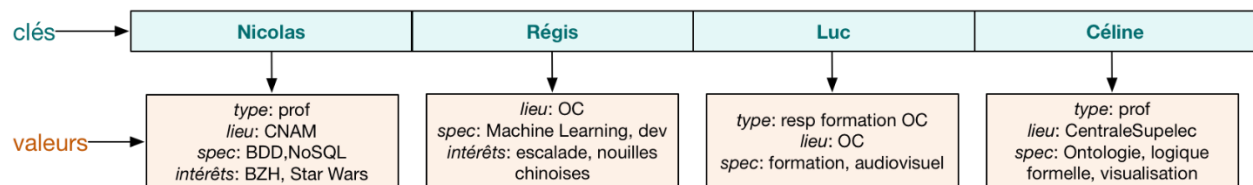


Figure 1 - BD NoSQL type orienté clé-valeur extrait de Behmo et Travers(2019a)

Parmi les logiciels de système de gestion des BD type orienté clé-valeur les plus connus se trouvent : Redis® développé par VMWare, Memcached développé par Danga, Azure Cosmos DB développé par Microsoft et SimpleDB développé par Amazon (Näsholm, 2012).

Bases de données type orienté colonne : ce sont des tables qui ne respectent pas la représentation en lignes traditionnelles des systèmes SQL, les BD du type orienté colonne distribuent les attributs comme colonnes avec la valeur de l'attribut et son identifiant (Näsholm, 2012).

La figure 2 illustre le schéma des BD NoSQL type orienté colonne.

Stockage orienté lignes					Stockage orienté colonnes							
id	type	lieu	spec	intérêts	id	type	id	lieu	id	spec	id	intérêts
Nicolas	prof	CNAM	BDD, NoSQL	BZH, Star Wars	Nicolas	prof	Céline	Centrale Supelec	Nicolas	BDD	Nicolas	BZH
Régis		OC	Machine Learning, Dev	escalade, nouilles chinoises	Céline	prof	Nicolas	CNAM	Nicolas	NoSQL	Nicolas	Star Wars
Luc	resp formation OC	OC	formation, audiovisuel		Luc	resp formation OC	Régis	OC	Régis	Machine Learning	Régis	escalade
Céline	prof	CentraleSupelec	Ontologie, logique formelle, visualisation				Luc	OC	Régis	Dev	Régis	nouilles chinoises
									Luc	formation		
									Luc	audiovisuel		
									Céline	Ontologie		
									Céline	logique formelle		
									Céline	visualisation		

Figure 2 - BD NoSQL type orienté colonne extrait de Behmo et Travers (2019a)

Parmi les systèmes de BD du type orienté colonne les plus connus se trouvent : BigTable développé par Google, HBase développé par Apache Hadoop, Spark développé par Apache, Elasticsearch développé par Elastic (Näsholm, 2012).

Bases de données type orienté document : elles se basent sur l'écriture des BD du type orienté clé-valeur, mais la valeur est remplacée par un document structuré. Ainsi, les BD de type orienté document ont un identifiant unique et l'information est stockée dans des documents sous la forme d'une structure JSON ou encore XML. JSON (*JavaScript Object Notation*) est un langage d'échange de données textuelles qui est le langage d'interaction de tous les navigateurs web. JSON est facile à générer et à analyser, et il est pratique à écrire et à lire pour les utilisateurs grâce à une syntaxe simple et à une structure en arborescence. XML est un langage qui utilise des balises pour délimiter les informations. Cependant, chaque balise peut contenir des informations, mais aussi d'autres balises, et ainsi permettre l'imbrication des informations hiérarchiques (Butler et al., 2002; *CouchDB : le guide authentique*, s. d.; *Extensible Markup Language (XML)*, s. d.).

La figure 3 illustre le schéma des BD NoSQL type orienté document.

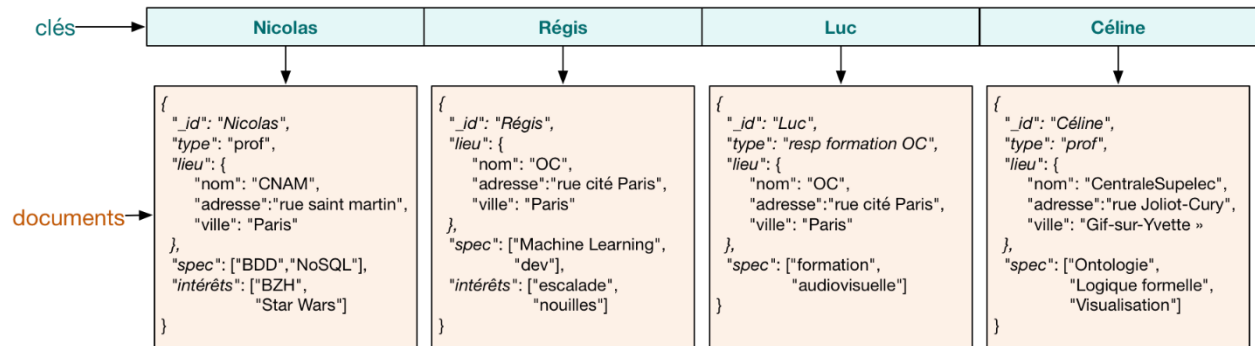


Figure 3 - BD NoSQL type orienté document (JSON) extrait de Behmo et Travers (2019a)

Dans les BD NoSQL de type orienté document, il est possible d'avoir dans une collection (équivalente à une table) de multiples documents (équivalents aux lignes d'une table) totalement différents entre eux (Näsholm, 2012). Ces documents peuvent avoir différents attributs (équivalents aux colonnes d'une table) dénommés champs (*fields*). Cette solution structurée permet de faire des requêtes complexes avec des langages d'interrogation riches comme dans le cas des SGBD relationnelles (*CouchDB : le guide authentique*, s. d.).

Parmi les systèmes de gestion de BD de type orienté document les plus connus, on peut citer : MongoDB développé par MongoDB, CouchBase développé par Apache-Hadoop, DynamoDB développé par Amazon, et Cassandra développé par Facebook – Apache (*CouchDB : le guide authentique*, s. d.; Näsholm, 2012).

Bases de données de type orienté *graph* : les BD sont dérivées de la représentation utilisée dans la théorie des graphes, les données stockées sont des éléments d'une structure interconnectée (graphe). Donc les informations sont représentées par des nœuds et des liens (Behmo & Travers, 2019a; Robinson et al., 2015).

La figure 4 illustre le schéma des BD NoSQL type orienté *graph*.

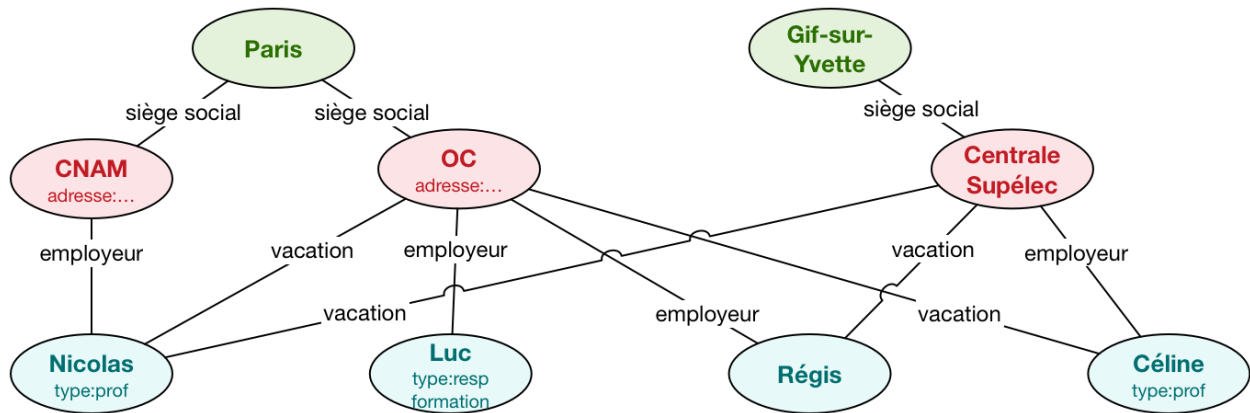


Figure 4 - BD NoSQL type orienté *graph* extrait de Behmo et Travers (2019a)

Parmi les systèmes de gestion de BD de type orienté *graph* les plus connus, il faut mentionner : Neo4j développé par Neo4j, OrientDB développé par Apache, et FlockDB développé par Twitter (Robinson et al., 2015).

2.1.3 Comparaison entre SGBD SQL et SGBD NoSQL

Plus connu comme théorème de CAP, de l'anglais (*Consistency, Availability, Partition tolerance*), le théorème de Brewer propose de définir les propriétés intrinsèques d'un SGBD (Mehra, 2019).

Les SGBD présentent trois propriétés fondamentales (Mehra, 2019) :

- **Cohérence (*Consistency*)** : cette propriété permet de montrer uniquement la valeur de la plus récente écriture d'une donnée. Plusieurs requêtes simultanées donneront le même résultat.
- **Disponibilité (*Availability*)** : cette propriété spécifie que les données doivent être disponibles en tout temps par des requêtes.
- **Distribution (*Partition Tolerance*)** : cette propriété spécifie une distribution ou un partitionnement des données sur plusieurs serveurs.

La figure 5 illustre le théorème de CAP qui classe les SGBD en trois catégories qui permettent comparer les SGBD SQL et les SGBD NoSQL.

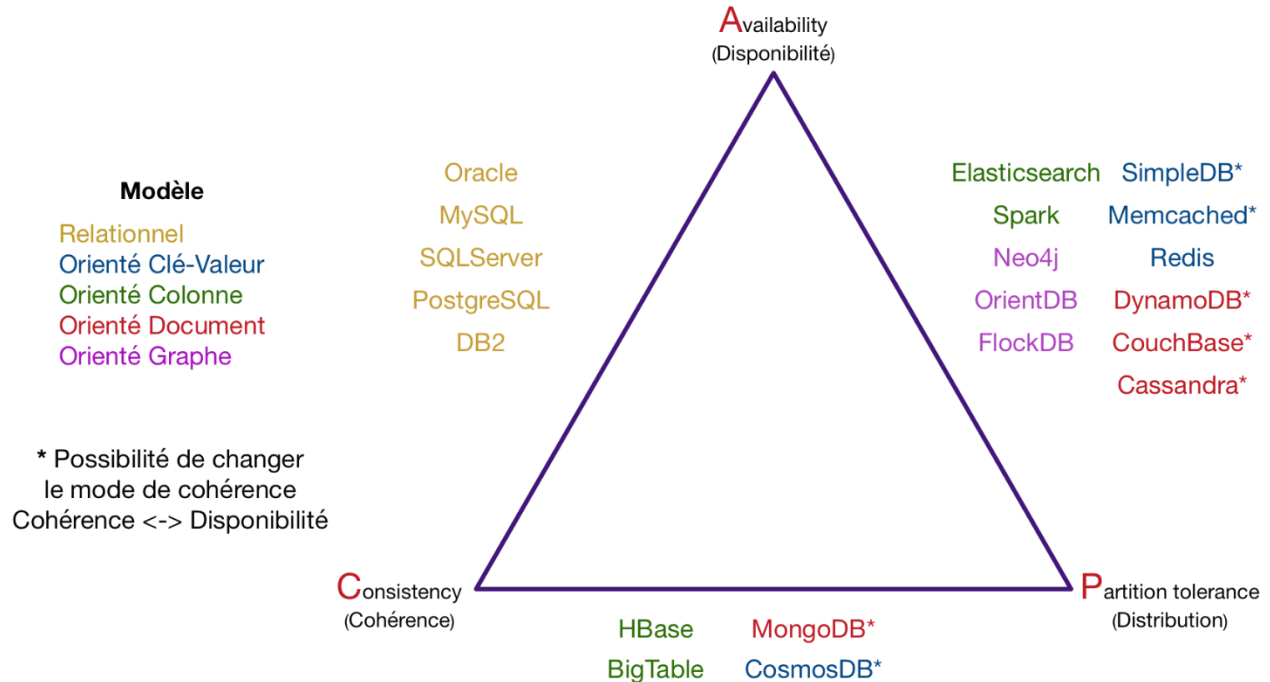


Figure 5 - Théorème de CAP extrait de Behmo et Travers (2019a)

Le théorème de CAP spécifie qu'il n'existe que deux propriétés possibles pour un SGBD.

CA (*Consistent and Available*) : en absence de partition et à la suite d'une écriture, des requêtes simultanées sur une donnée retourneront le même résultat sans temps d'attente. C'est notamment le cas des SGBD SQL (Mehra, 2019).

CP (*Consistent and Partition Tolerant*) : il s'agit de systèmes au sein desquels l'ensemble des données est répliqué et distribué sur plusieurs serveurs. Donc, à la suite d'une écriture synchronisée, des requêtes simultanées sur une donnée nécessitent la validation de la valeur sur les réplicas qui sont dans les différents serveurs. C'est le cas des BD de type orienté colonne BigTable et HBase; la BD de type orienté document MongoDB et la BD de type orienté clé-valeur Cosmos DB (IBM, 2021; Mehra, 2019).

AP (*Available and Partition Tolerant*) : à la différence des BD classées CP dans lesquelles la validation de la cohérence implique des temps d'attente dus aux synchronisations, les BD AP fournissent la réponse rapidement. Cependant, il n'est pas possible de savoir si la réponse correspond à la dernière écriture. Donc, afin de fournir un temps de réponse rapide, les réponses

aux requêtes ne sont que les plus probables, car à la suite d'une écriture, la mise à jour n'est pas synchronisée. C'est le cas des BD de type orienté colonne Spark et Elasticsearch; de type orienté document comme CouchBase, DynamoDB et Cassandra. Les BD de type orienté clé-valeur Redis, Memcached et SimpleDB ainsi que les BD de type orienté *graph* Neo4j, OrientDB et FlockDB (IBM, 2021; Mehra, 2019).

Le classement qui découle du théorème de CAP ne doit pas être interprété de manière stricte, car il peut amener à des conclusions erronées. Parmi la liste des SGBD NoSQL, il n'est pas rare de voir des solutions hybrides passées d'une propriété à une autre (Behmo & Travers, 2019b). Aussi, certains SGBD SQL connus comme NewSQL peuvent offrir la capacité de mise à l'échelle (*Scalability*) des SGBD NoSQL (Raffet, 2017).

En résumé, les SGBD NoSQL ne remplaceront pas les SGBD SQL, mais ils sont une alternative pour stocker et gérer les données différemment dans un contexte de gestion des données massives. Le choix de SGBD dépend de la nature des données et des objectifs du projet (Raffet, 2017).

2.2 Les composantes logicielles d'un SIG web

Les composantes logicielles nécessaires pour le stockage et la diffusion interactive de l'information géospatiale vers une plateforme cartographique web sont les bases de l'architecture logicielle du SIG web. En ajoutant des outils d'analyse spatiale, le système fait ressortir les relations, les modèles et les associations qui sont souvent cachés par la complexité des données (Pratt, 2020; Serre, 2016; Thierion, 2010). Cette section vise à présenter les différents éléments d'un SIG web afin de mieux définir la partie visible et importante pour l'utilisateur - parfois non expert en géomatique - : l'interface client.

2.2.1 Les types d'architecture logicielle d'un SIG web

La sélection de logiciels à utiliser dans l'architecture d'un SIG web est déterminée, entre autres, par la limitation des licences, l'accès à un SGBD, et leur utilisation finale. La figure 5 modifiée de (Germain, 2020) présente les grandes lignes de l'architecture logicielle d'un SIG web.

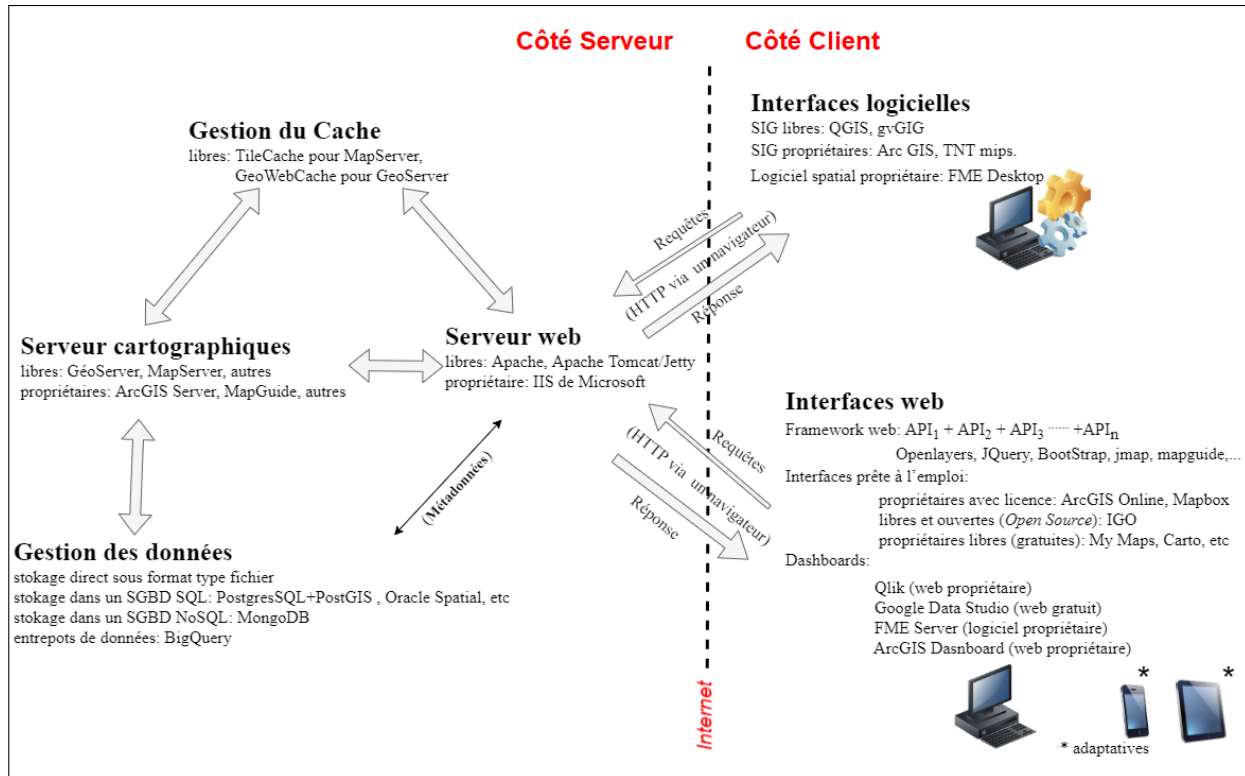


Figure 6 – Schéma simplifié de l'architecture d'un SIG web (Modifié de Germain 2020)

2.2.2 La diffusion de l'information géospatiale

Bien qu'il soit possible d'échanger de l'information sur l'Internet avec un protocole adapté comme le FTP (*File Transfer Protocol*), les formats de données géographiques ne sont pas tous adaptés à un échange simplifié pour une optimisation optimale au sein d'un SIG web. Certains formats requièrent des logiciels spécifiques pour leur création et leur lecture. D'autres formats sont composés d'un ensemble de fichiers avec des sous-extensions qui rendent leur manipulation difficile (Ramsey & Leslie, 2012).

Afin de faciliter les échanges entre les systèmes, ce qu'on appelle l'interopérabilité des systèmes, l'OGC (*Open Geospatial Consortium*) a défini plusieurs formats et services ouverts. Ces normes couvrent non seulement l'ensemble des exigences essentielles pour gérer, manipuler et visualiser les données géospatiales avec les interfaces logicielles, mais aussi avec des interfaces web (Lelong, 2016; Ramsey & Leslie, 2012; Thierion, 2010). Ces normes comprennent, notamment, des formats de stockage et de gestion de données tels que : GeoRSS, GeoJSON, GML, KML ainsi que des

services pour transporter des informations géospatiales avec le protocole HTTP (*HyperText Transfer Protocol*). Ces services permettent de contourner les limitations des formats de données géospatiales qui ne peuvent pas être accessibles sur les navigateurs web (*Web Graphics*, s. d.; *web Map Service / OGC*, s. d.). Les deux services les plus connus sont le WMS (Web Map Service) et le WFS (Web Feature Service).

Le service WMS permet la diffusion de l'information géospatiale sous la forme d'une image à travers le protocole HTTP. Une demande WMS définit principalement la ou les couches géographiques à afficher, et de la zone géographique à traiter. La réponse à la demande est une image renvoyée dans un des formats d'image web (JPEG, PNG, GIF). De plus, le service prend en charge la possibilité de spécifier le style des images avec les formats SLD (*Styled Layer Descriptor*) (*Web Map Service / OGC*, s. d.).

Les données disponibles à travers un service WMS sont accessibles à l'ensemble des personnes qui désirent les utiliser sans l'apprentissage d'un logiciel SIG (convivialité). Cependant, les géomaticiens peuvent interagir avec les données avec n'importe quel logiciel SIG (interopérabilité).

Le service WFS permet la diffusion de l'information géospatiale sous la forme vectorielle (format GML par défaut) à travers le protocole HTTP. Un tel service a l'avantage d'accéder directement et précisément aux informations géographiques : les attributs spatiaux et descriptifs. Cela permet d'effectuer des opérations de requêtes et d'analyses, et de stocker les informations en interne (*Web Feature Service / OGC*, s. d.). Une des caractéristiques intéressantes des services WFS est qu'il permet (sous autorisation) les opérations de transaction. C'est-à-dire, si le mode transactionnel est autorisé, le service WFS-T permet de créer, de modifier, de remplacer et de supprimer les informations gérées par le serveur du service WFS (*Web Feature Service / OGC*, s. d.).

Le tableau 1 illustre les principales normes des services web de l'OGC (services ouverts)

Tableau 1 – les normes principales de l'OGC

Norme internationale	Objectif	Formats	Requêtes principales Commande
Service WMS web Map Service	Transporter de données matricielles	JPEG PNG GIF	GetCapabilities : retourne les métadonnées du serveur GetMap : retourne une ou plusieurs images cartographiques géoréférencées renvoyées dans un format d'image
Service WFS Web Feature Service	Transporter de données vectorielles	GML	GetCapabilities : retourne les métadonnées du serveur WFS DescribeFeatureType : retourne la liste des attributs GetFeature : retourne les valeurs des attributs et de la composante géométrique en format GML GetGMLObject : retourne l'information GML par un identifiant
SLD* Styled Layer Descriptor	Définir le style (symbologie) d'une couche WMS	XML (Requête)	Requête SLD au serveur avec une requête de service WMS GetMap
*normes avancées			

Bien que les services de l'OGC garantissent l'accès à la dernière actualisation des données géospatiales, dans certains cas, il est plus facile d'utiliser des formats de données géospatiales dits « web » sans se soucier des échanges client-serveur. La liste des formats web est vaste, et chacun a des avantages et des défauts. Dans ce contexte, la sélection d'un format dépendra des besoins spécifiques de chaque projet.

Pour utiliser les données géographiques sur le web, des formats adaptés aux applications web ont été développés.

Parmi les principaux formats web, il est possible de citer les références suivantes :

- **GeoJSON:** format *open source* mais non géré par l'OGC. C'est un format international d'échange de données géospatiales basé sur la notation d'objets JavaScript et qui permet une écriture compacte de l'information. Il définit directement les géométries et la façon dont elles sont combinées pour représenter des données géographiques, leurs propriétés et

leur étendue spatiale. À moins que ce ne soit paramétré différemment, GeoJSON utilise le système de référence de coordonnées géographiques, *World Geodetic System 1984* (WGS84) par défaut et les unités de degrés décimaux (Butler et al., 2002). La symbologie n'est pas définie dans ce format.

- **KML:** format sous la représentation d'un langage XML basé sur la visualisation géographique, y compris l'annotation de cartes et d'images. Cela comprend non seulement la présentation des données géométriques de deux et de trois dimensions (2D et 3D) sur le globe, mais également le contrôle de la navigation de l'utilisateur (comment regarder l'information). KML utilise uniquement le système de référence de coordonnées géographiques WGS84 (Burggraf, 2015).
- **GML:** format sous la représentation d'un langage XML qui permet la description des caractéristiques géographiques 2D et 3D. GML sert de langage de modélisation pour les systèmes géographiques ainsi que de format d'échange ouvert pour les transactions géographiques sur Internet (WFS). La grammaire XML comprend non seulement les données géospatiales avec ses attributs, mais elle inclut également une description exhaustive de l'information géographique : les systèmes de coordonnées, la gestion de la topologie, la gestion du temps ainsi que des métadonnées (*Geography Markup Language / OGC*, s. d.).
- **GeoRSS:** format sous la représentation d'un langage XML qui permet un mode de diffusion des informations dans un format qui sert à publier des contenus fréquemment actualisés (temps réel) (Reed, 2017).

Les formats de bases de données embarquées

- **SQLite:** il s'agit d'un moteur de base de données embarquées qui se différencie de la plupart des bases de données SQL par le fait qu'il lit et écrit directement dans des fichiers. Donc, dans un seul fichier multiplateforme se trouve une base de données SQL complète avec des tables, index, déclencheurs et vues. SQLite n'intègre pas une gestion des systèmes de coordonnées et des fonctions d'analyse spatiale (*SQLite Home Page*, s. d.).

- **SpatiaLite:** une bibliothèque destinée à s'intégrer à SQLite pour fournir un SGBD spatiales complet et puissant. Cela permet d'intégrer les systèmes de coordonnées et des fonctions d'analyse spatiale (*SpatiaLite: SpatiaLite*, s. d.).
- **GeoPackage:** une norme de l'OGC qui décrit un ensemble de règles pour stocker la totalité de caractéristiques vectorielles, ainsi que des images dans une base de données SQLite. C'est un format qui peut gérer la symbologie.

2.2.3 Les serveurs cartographiques

Un serveur cartographique est un logiciel installé sur un serveur et qui permet au client d'accéder aux données à partir d'un serveur web (HTTP). Le logiciel « serveur cartographique » transforme l'information géospatiale pour une diffusion optimale sur le web, par exemple, dans un des services internationaux de l'OGC comme WMS ou WFS, ou bien un service propriétaire comme ArcGIS d'ESRI.

La mise en place d'un serveur cartographique (programmation et configuration) dépend d'une série de facteurs comme les formats de données en entrées, le SGBD utilisé, les systèmes de coordonnées géographiques, le style des couches, etc. Dans tous les cas, un serveur cartographique doit être capable de se connecter à plusieurs sources de données.

Les serveurs cartographiques les plus connus – dans le monde *Open Source* - sont :

- **MapServer :** disponible depuis 1997 et développé par l'Université du Minnesota en coopération avec la NASA.
- **GeoServer :** disponible depuis 2001 et développé par *The Open Planning Project (TOPP)* à New York.

Le tableau 3 montre une comparaison entre MapServer et Geoserver.

Tableau 2 - Comparaison entre MapServer et Geoserver

Entreposage des données	<i>GeoServer</i>	<i>MapServer</i>
Programmation	Langage orienté objet Java	Langage C
Logiciel multi-utilisateur	Limité	Logiciel multi-utilisateur
Gestion de Cache	Avec le système GeoWebCache	
Méthode directe (Liste non exhaustive des formats supportés)	Shapefile GeoTIFF ArcGRID JPEG2000 * GDAL formats *	Shapefile GML KML GeoTIFF ArcGRID JPEG2000 GDAL formats
Méthode de gestion (Liste non exhaustive des SGBD supportés)	PostGIS ArcSDE * Oracle Spatial and Graph (plug-in nécessaire) DB2 * SQL Server *	PostGIS ArcSDE Oracle Spatial and Graph Spatialite MySQL SQL Server
Configuration de styles, projections, etc.	Interface de configuration web	Fichiers de configuration MapFile
Gestion de la norme WFS-T	Supporte la norme WFS-T	N'est pas capable de gérer directement le mode WFS-T
Convivialité	Grande convivialité	Convivialité relative.
	*plug-in nécessaire	

Au niveau de serveurs « propriétaires », la compagnie ESRI propose des solutions largement utilisées :

- **ArcGIS server** : le logiciel serveur cartographique complet et commercial de la compagnie.
- **ArcGIS Online** : la plateforme web en infonuagique qui permet de créer des services (WFS, Tile Map, etc.), ainsi que des cartographies web et des tableaux de bord.

2.2.4 Les interfaces client des SIG web

L'interface client est la pointe de l'iceberg du SIG web avec laquelle interagit l'utilisateur pour exploiter les données. Elle peut être une interface cartographique, un tableau de bord analytique ou bien souvent, les deux interfaces intégrées ensemble.

Une interface client peut être :

- **Une interface logicielle installée sur l'ordinateur** : exemples *Open Source* comme QGIS ou propriétaire comme ArcGIS.
- **Une interface web**: elle peut être une interface programmée avec une API de l'anglais *Application Programming Interface* ou une interface web dite « prête à l'emploi ».

Une cartographie web « prête à l'emploi » est une cartographie simple destinée à la diffusion des données géographiques sur le web sans programmation. Parmi les interfaces prêtes à l'emploi les plus connues se trouvent les solutions :

- propriétaires avec licence comme ArcGIS Online, Mapbox Studio;
- libres et ouverts (*Open Source*) IGO du gouvernement de Québec;
- propriétaires, mais libres (gratuites) comme My Maps Google, Carto Studio, etc.

Mapbox peut être utilisé comme interface de programmation web (API) ainsi que comme interface web prête à l'emploi. Les APIs servent à ajouter des interactions dans l'environnement de travail tel que boutons pour sélectionner les couches à afficher, changer le fond de carte ou ajouter des barres d'outils pour accéder à des fonctionnalités.

Parmi les APIs connues, on peut citer les solutions :

- propriétaires avec licence comme ArcGIS API;
- libres et ouvertes (*Open Source*) comme OpenLayers, LeafLet;
- propriétaires, mais libres (gratuites limitées) comme Google Map, Mapbox, Here GO.

L'interface client de style « tableau de bord analytique » est une forme plus évoluée d'interfaces. Elle est largement utilisée par les systèmes géodécisionnels servant à l'affichage simultané d'indicateurs, de données brutes, d'une cartographie, d'analyses multidimensionnelles ainsi que des résultats issus de requêtes. Les composantes d'un tableau de bord (l'information, les cartes et les indicateurs) sont capables d'interagir entre eux et avec l'utilisateur à partir d'un logiciel/interface qui offre une vue d'ensemble sur un seul écran (*Définition d'un tableau de bord—ArcGIS Dashboards | Documentation, 2020; Law, 2019*). Étant donné que ces interfaces permettent à l'utilisateur de définir n'importe quelle étendue ou de filtrer les données de la manière qui lui convient le mieux, les tableaux de bord sont devenus les éléments déterminants de toute infrastructure géospatiale destinée à la prise de décisions, la visualisation de tendances, la

surveillance à temps réel ou la diffusion d'information. En résumé, c'est une puissante ressource géodécisionnelle (*ArcGIS Dashboards / Tableaux de bord de données*, s. d.).

La visualisation infographique des indicateurs paramétrés est possible à partir des langages de programmation web ou avec l'aide des APIs. Ces indicateurs affichent le comportement de données géospatiales associées à une thématique dans une interface destinée à l'analyse interactive. Parmi ces APIs se trouvent par exemple : Raphael, D3js, Google charts, Papers, etc. Cependant, il existe plusieurs options logicielles de tableaux de bord prêtes à l'emploi qui offrent la possibilité de créer des interfaces web permettant la visualisation de multiples cartes et indicateurs analytiques. Ces logiciels de tableaux de bord ont des architectures multiextensions qui ont une vaste liste de connecteurs pouvant aller chercher l'information dans les services internationaux définis par l'OGC, ainsi que dans des formats type fichiers libres et dans certains formats propriétaires (*Qlik GeoAnalytics*, s. d.; *Welcome to Data Studio! (Start here)*, s. d.).

Parmi les tableaux de bord connus, on peut citer :

- Qlik (web propriétaire) :
- Google Data Studio (web gratuit)
- ArcGIS Dashboard (web propriétaire)

Tous ces logiciels permettent entre autres de :

- mettre en valeur des phénomènes complexes avec des indicateurs simples destinés à combiner l'information;
- illustrer l'évolution des phénomènes sur une dimension temporelle;
- favoriser l'interaction entre les différentes composantes.

En résumé, les tableaux de bord sont des applications web configurables qui permettent à l'utilisateur d'exploiter sous une forme interactive l'information géographique à l'aide de graphiques simple, des cartes, et d'autres éléments visuels capables de répondre à des filtres de sélection (*ArcGIS Dashboards / Tableaux de bord de données*, s. d.; *Qlik GeoAnalytics*, s. d.; *Welcome to Data Studio! (Start here)*, s. d.; Law, 2019).

2.3 L'extraction, la transformation et le chargement des données

Pour traiter différents types de données géographiques de sorte à pouvoir alimenter une BD ou une interface client, il faut passer par un processus qui consiste à l'extraction, la transformation et le chargement des données. Ce processus est connu en géomatique décisionnelle comme un processus ETL, de l'anglais *Extract, Transform, load* (Bédard, 2010). Cependant, étant donné la diversité des sources d'informations à extraire, et compte tenu de la nécessité de consolider ces informations après des transformations requises, le processus ETL entraîne nécessairement deux autres opérations : les opérations de sélection et de consolidation (intégration) (Chafki & Desrosiers, 2011). Au final, l'ensemble des opérations présentes dans un processus ETL sont les suivantes : Sélection, Extraction, Transformation, Intégration et Chargement (Godin & Desrosiers, 2011).

Ils existent différentes façons de réaliser une opération ETL. Il est possible d'y parvenir à l'aide de langages de programmation ou à partir d'outils ETL (Crozat, 2016a). Les logiciels SIG (*open source* et propriétaires) mettent à la disposition des outils pour lire, afficher, interroger et exporter des fichiers sous plusieurs formats. Ces opérations peuvent être automatisées sous la forme d'outils de chaînes de traitement équivalentes aux outils ETL (*ModelBuilder / ArcGIS for Desktop*, s. d.; *QGIS 3.10 Manual*, s. d.). Cependant, si la source de données est une BD NoSQL, le modelleur graphique de QGIS ainsi que le *Model Builder* de ArcGIS sont limités. Il n'est pas possible de les utiliser pour travailler avec les SGBD qui ne sont pas traditionnels sans recourir à des modules d'extension (plugin) (Murphy, 2012; *QGIS MongoDB Python Plugins Repository*, s. d.).

Les logiciels propriétaires dédiés à l'ETL des données géospatiales sont sur le marché depuis 1993 (FME Desktop, 2020). Des solutions *Open Source* existent depuis peu de temps (Bédard, 2010). Ces logiciels permettent généralement les opérations ETL à partir de quatre types de sources de données : les fichiers, les répertoires, les données web ainsi que les BD relationnelles ou non (Crozat, 2016c) (Techroba, 2015).

Parmi les logiciels ETL, plusieurs choix sont possibles selon le type de licence (*Compare Data Transformation Tools / Top 10 Platforms in 2020*, s. d.; Techroba, 2015) :

- **Open Source:** Talend Open Source Data Integrator, Scriptella, KETL, Pentaho Data Integrator-Kettle, Jaspersoft ETL, GeoKettle CloverETL, HPCC Systems, Jedox, Apatar, etc.

- **Propriétaire** : Domo, Alteryx, FIVETRAN, Pentaho, Yuno, FME, etc.

Parmi tous ces outils, certains sont capables de gérer des données géospatiales. C'est le logiciel FME de l'anglais (*Feature Manipulation Engine*) qui propose le plus d'options pour l'ensemble des opérations.

3 Matériel et Méthodes

3.1 Site d'étude

La zone d'étude du projet englobe les espaces occupés par les eaux maritimes dans la région naturelle de la Patagonie chilienne. Cette région est très vaste et variée, il est possible de trouver dans ce milieu presque toutes les formes de côte. Des côtes qui font directement face à l'océan, des côtes intérieures, des golfes, des îles, des îlots, des plaines côtières, des Fjords, etc. Entouré par ces énormes extensions côtières qui peuvent être – ou non – influencées par des champs de glaces, se trouve le complexe de canaux patagoniques et canaux fuégiens avec ses très grands nombres de golfes et de baies.

La figure 7 illustre le site d'étude du présent projet.

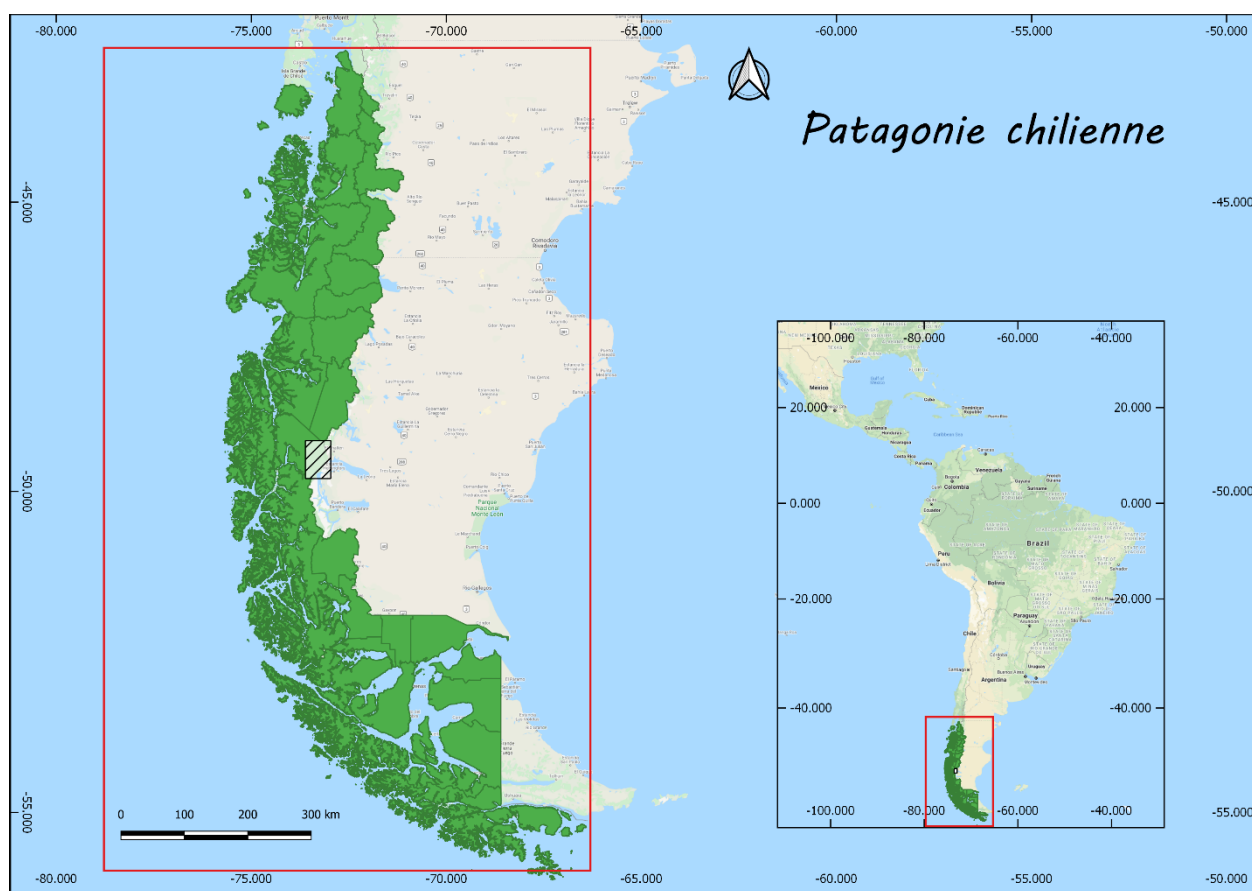


Figure 7 Site d'étude (carte Enrique Martin de Villodres)

3.2 Méthodologie

Le développement de notre projet suit plusieurs étapes méthodologiques empruntées pour le développement d'un système décisionnel. Ces étapes sont représentées à la figure 8 (modifié de Godin & Desrosiers (2011)).

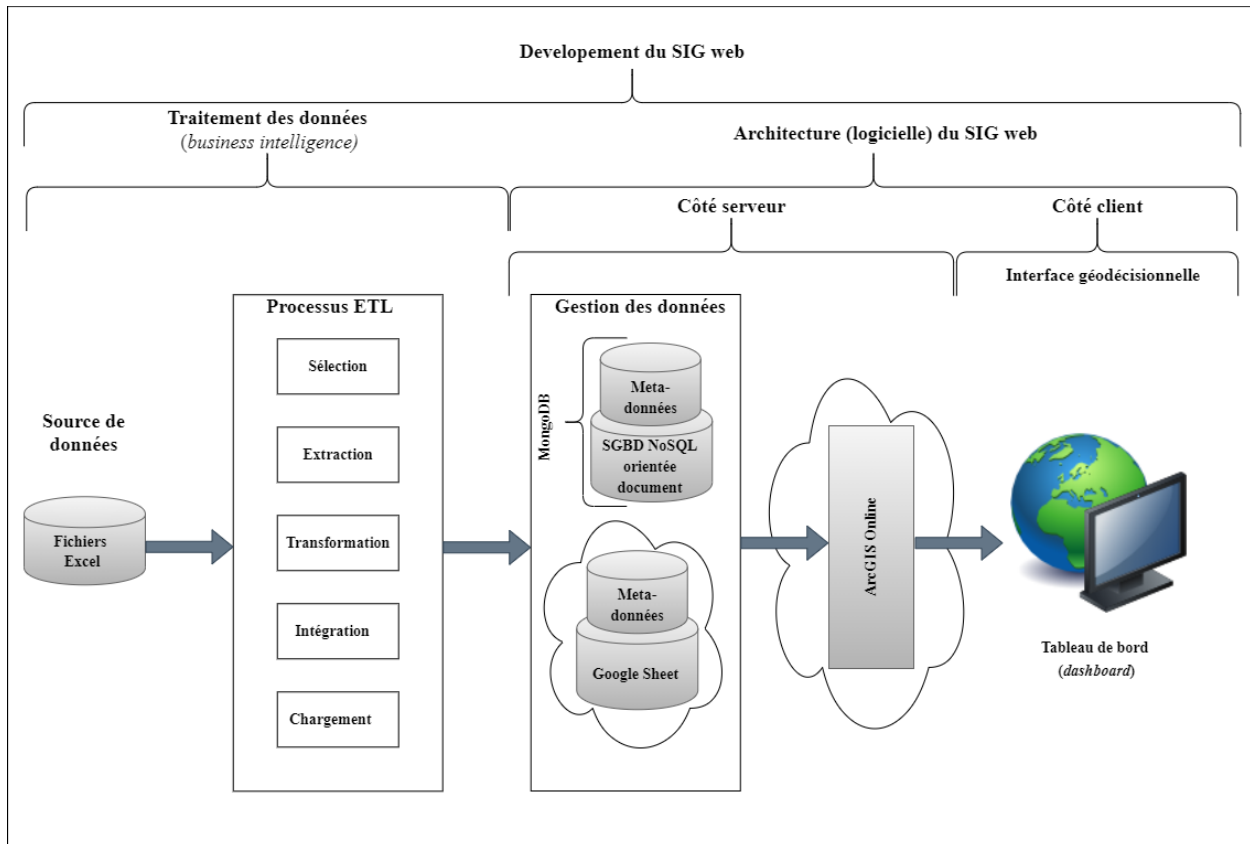


Figure 8 - Diagramme méthodologique

3.3 Les données

Les données destinées à alimenter le SGBD géospatiales, et l'interface cartographique web ont été fournies par l'équipe de recherche issue de la collaboration entre l'Université de Magellan et l'Université de Montréal.

Les données sont réparties en huit jeux de données sous la forme de fichiers Excel qui correspondent aux résultats des suivis effectués par Simon Demers, l'étudiant à la maîtrise qui travaille sur l'analyse des données à l'Université de Montréal. Les données sont extraites de l'application mobile ICETACE, ainsi que des résumés d'observations. L'ensemble de données à

traiter est issu d'études effectuées sur plus 20 ans. Plusieurs acteurs ont travaillé avec différentes méthodologies d'acquisition, à différentes échelles, et avec différents objectifs. Par conséquent, il existe une très grande hétérogénéité entre les noms des attributs et le format des attributs. Les données redondantes ou imbriquées sont fréquentes.

Des tentatives d'uniformisation ont eu lieu à propos de certains jeux de données, car des données étaient manquantes ou mal distribuées dans les fichiers. De plus, il n'y avait pas d'uniformité en ce qui concerne la projection cartographique ou le système de coordonnées utilisé pour les observations. Enfin, certaines informations étaient inutiles, comme des observations de mammifères dénommés pinnipèdes, ou encore des observations de fermes piscicoles, etc.

3.4 Le choix entre SQL ou NoSQL

Avant d'entamer le traitement des données, il a fallu faire un choix méthodologique quant à la gestion des données. À la suite de l'analyse préliminaire des différents jeux de données, et compte tenu des différences entre les BD SQL et les BD NoSQL, ce sont ces dernières de type orienté document qui ont été retenues comme les mieux adaptées à la gestion des données du projet. Cette décision reflète la difficulté de la normalisation des données compte tenu de la nature hétérogène des jeux de données. L'autre aspect qui a été considéré est de faciliter l'exploitation des données, l'analyse, et l'aide à la prise de décisions.

Parmi les BD NoSQL de type orienté document, nous avons choisi le SGBD MongoDB. Il s'agit d'un logiciel libre qui a une grande flexibilité pour gérer des données d'une grande hétérogénéité avec un langage de requête riche, et qui permet surtout la gestion des données géographiques (*MongoDB Manual*, 2020; *Query Documents — MongoDB Manual*, s. d.). MongoDB peut stocker toute l'information du projet dans une collection de documents hétérogènes. De plus, les valeurs des champs des documents peuvent inclure d'autres documents, ce qui permet de gérer des données imbriquées (*MongoDB Manual*, 2020; Tannir, 2015).

3.5 Le processus ETL

FME est le logiciel ETL retenu pour le projet afin d'effectuer le chargement et le traitement des données, mais aussi le processus de sélection des données

Sélection de données

La méthodologie utilisée pour effectuer la sélection de données est représentée à la figure 9.

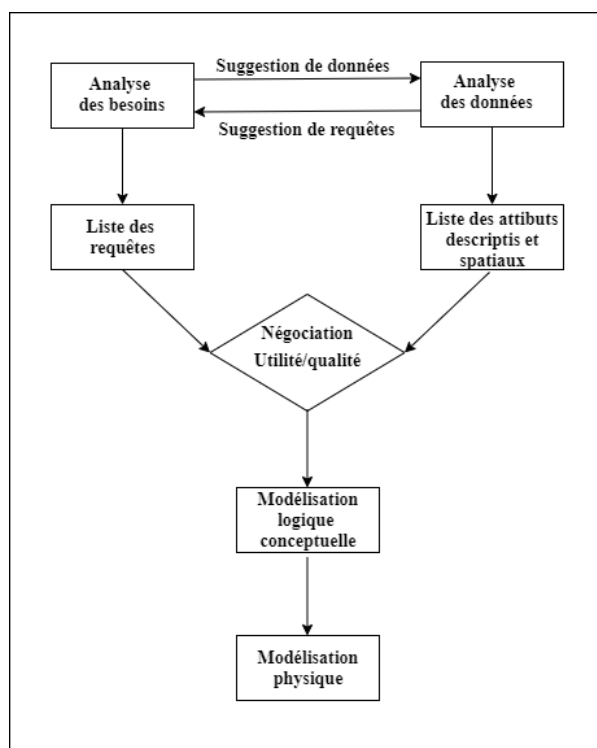


Figure 9 - Organigramme méthodologique pour la sélection des données. Modifié de (Crozat, 2016c)

3.5.1 L'analyse des besoins

Les principaux besoins des utilisateurs pour afficher et pour analyser les données géospatiales du projet ont été identifiés à la suite de consultations avec l'équipe de recherche formée de Jorge Gibbons de l'Institute de la Patagonie de l'université de Magellan (*Instituto de la Patagonia de la Universidad de Magallanes*), et de Liliana Perez et de Simon Demers du Département de géographie de l'Université de Montréal. Leurs principaux intérêts ont porté sur la création d'une interface web pour accéder à l'information stockée dans un tableur en ligne afin de faciliter le partage de données et le travail collaboratif. La solution infonuagique choisie a été l'application gratuite Google Sheets.

En plus de répondre aux attentes de stockage et de visualisation des données spatio-temporelles liées aux animaux marins observés, les consultations ont permis d'identifier les besoins de l'équipe

de chercheurs en question. Les besoins identifiés ont été séparés en deux groupes : les requêtes et les traitements.

Au niveau des requêtes, une liste de questions a été mise en place. Les requêtes ainsi programmées permettront aux utilisateurs d'accéder à des indicateurs ou réponses nécessaires pour leurs études. Cette analyse a permis de mettre en valeur les attributs descriptifs nécessaires pour répondre aux questions. L'ensemble de ces données (entièrement ou partiellement) sont les champs (*fields*) des documents qui constituent la BD NoSQL. La liste potentielle des questions à formuler (requêtes) est disponible à l'annexe 1

Au niveau des traitements, une liste de manipulations a été mise en place pour permettre une exploitation optimale des données. Chaque jeu de données a été évalué afin d'en extraire les principales caractéristiques : les attributs, les domaines et les descriptions. Cette évaluation a mis en évidence des transformations nécessaires appliquées aux jeux de données pour leur exploitation.

3.5.2 L'évaluation des données

Une méthode d'évaluation des données a été mise en place. Elle emprunte la méthode définie par Oracle et adaptée par Crozat (2016b). Dans cette méthode, l'auteur propose une brève description de la source de données en spécifiant le nom de la source, le type de source, le contenu, son origine ainsi que le nombre de lignes des tables « source ». Aussi, une échelle de classement est suggérée pour évaluer la qualité ainsi que l'utilité des données assignées à chaque source : une valeur entre 0 et 3 assignée selon les critères du tableau 6.

Tableau 3 - Échelle utilisée pour évaluer la qualité et l'utilité des données

Valeur	Qualité des données	Utilité des données
0	Données inexploitable	Données sans intérêt
1	Données peu exploitables (traitements incertains)	Données utiles pour la documentation uniquement
2	Données exploitables après traitements	Données utiles a priori
3	Données exploitables sans traitements	Données utiles avec certitudes

Modifié de (Crozat, 2016b)

Le tableau 7 résume l'ensemble des informations pris en compte pour l'évaluation des données.

Tableau 4 - Évaluation du jeu de données : (le nom du jeu de données)

Nom	Description	Type	Qualité	Utilité	Commentaires
Nom de l'attribut	Description de l'attribut	Type du domaine de l'attribut	Qualité de l'attribut (Valeur entre 0 et 3)	Utilité de l'attribut (Valeur entre 0 et 3)	Commentaires sur l'attribut

Modifié de (Crozat, 2016b)

3.5.3 La négociation utilité/qualité

Afin de déterminer les données à conserver, un arbitrage est nécessaire pour évaluer le rapport entre l'utilité et la qualité des données de chaque jeu de donnée (Crozat, 2016b). En raison des multiples noms (dans plusieurs langages) assignés aux attributs dans les différentes sources de données, il a fallu retenir une liste de noms à utiliser dans les champs de la BD NoSQL. Donc, l'arbitrage utilité/qualité a été effectué en signalant le champ cible de l'information. Un exemple de l'analyse d'un des huit jeux de données ainsi que celle de la négociation utilité/qualité est disponible à l'annexe 2.

Il faut noter que les noms des attributs conservés pour remplir les champs de la BD ont été notés en anglais pour respecter la collaboration internationale du projet.

3.5.4 Les modélisations conceptuelle, logique et physique

Afin de faciliter la création des requêtes, les modélisations conceptuelle et logique ont été réalisées comme pour une BD relationnelle sans être contraintes aux différentes normalisations (Crozat, 2016b). Ces modèles sont disponibles à l'annexe 3. Il en découle un modèle physique qui sera utilisé sous la forme d'objets JSON. Le modèle physique est disponible à l'annexe 4.

3.6 L'extraction et la transformation des données

Après l'évaluation des différents jeux de données, certaines informations ont été conservées et d'autres abandonnées. Le tableau de synthèse, tableau 8, planifie l'extraction, la transformation et le nettoyage des données. Il présente pour les différents champs de la BD (champs cible), le nom de l'attribut (attribut source) où est contenue l'information dans les différents jeux de données (*dataset* source) ainsi que les transformations requises pour les rendre exploitables.

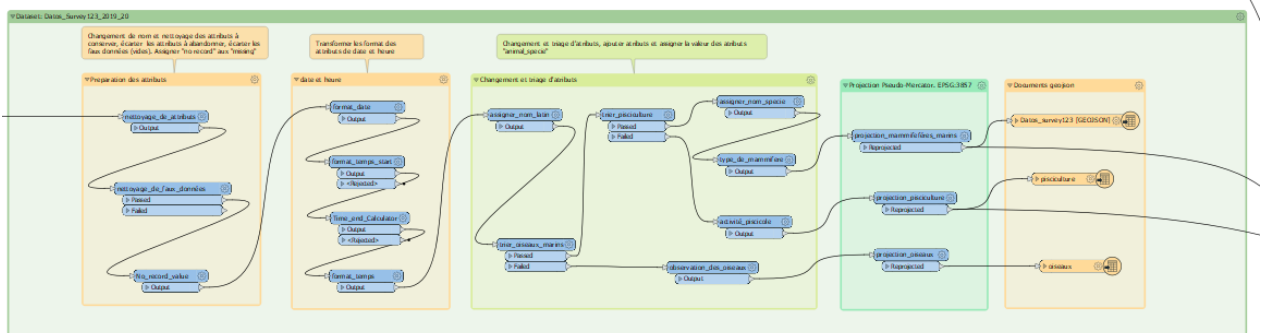
Tableau 5 - Tableau de synthèse des transformations requises

Champs cible	Dataset source	Attribut source	Règle de transformation
Nom du champ de la BD NoSQL	Jeu de données source	Nom de l'attribut dans le jeu de données source	Règles d'extraction, transformation et de nettoyage des données. En expliquant comment gérer les exceptions.

Modifié de (Chafki & Desrosiers, 2011)

Une partie d'un des tableaux de synthèse des transformations est disponible à l'annexe 5

Une fois les opérations d'extraction et de transformation de données définies, un espace de travail dans le logiciel FME (*Workspace*) contenant les transformateurs (*transformers*) requis a été créé. L'espace de travail du projet a été conçu de manière que chaque jeu de données utilisé possède une séquence de transformateurs regroupés pour différentes tâches. La figure 10 illustre le processus ETL d'un des jeux de données.

**Figure 10 - Processus ETL d'un des jeux de données**

Une vue globale de l'espace de travail contenant le flux de données à partir des huit jeux de données est disponible à l'annexe 6

Chaque jeu de données suit la séquence de transformations suivante :

- **La préparation des attributs** : la figure 11 illustre un des flux de travail créé pour renommer les attributs à conserver, et écarter les attributs non désirables et vides.

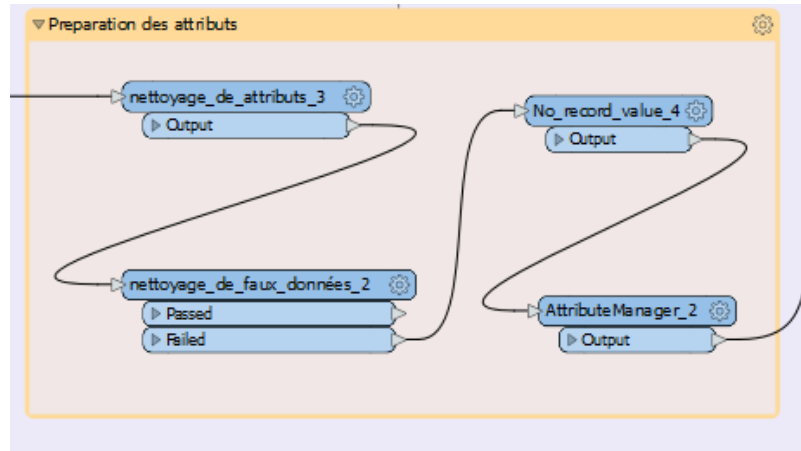


Figure 11 – Exemple du flux de travail pour la préparation des attributs

- **Transformer le format des attributs de « date et heure »** : la figure 12 illustre un des flux de travail créé pour effectuer des opérations de calculs de temps (*temporal values*), car la syntaxe qu'utilise FME est différente de celle utilisée dans les jeux de données, ainsi que par MongoDB ou par l'interface web

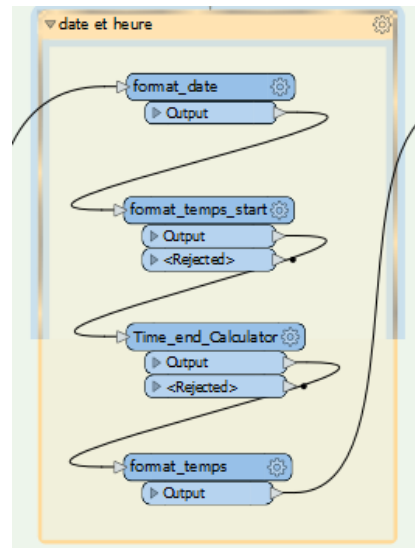


Figure 12 - Exemple du flux de travail pour le calcul de temps

- **Ajouter des attributs et définir les valeurs:** la figure 13 illustre un des flux de travail créé pour ajouter les noms des espèces animales "*animal_specie*" dans quatre langues (latin, français, anglais et espagnol)

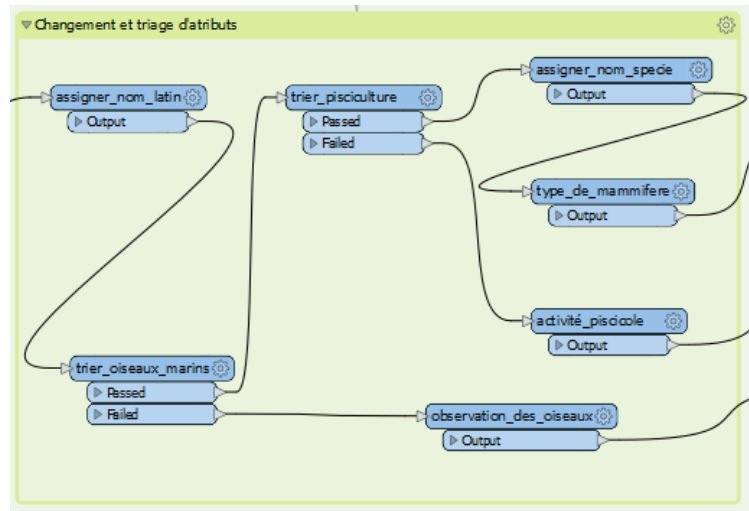


Figure 13 - Exemple du flux de travail pour ajouter les noms des espèces

- **Adéquation de la syntaxe des coordonnées :** la figure 14 illustre un des flux de travail créé pour remplacer les "S" et "W" des coordonnées par le symbole "-"; faire la correction de symbole de degré "°"; faire la transformation DMS a degré décimal; créer des géométries ainsi que pour la définition de la projection cartographique : Projection Pseudo-Mercator. EPSG:3857.

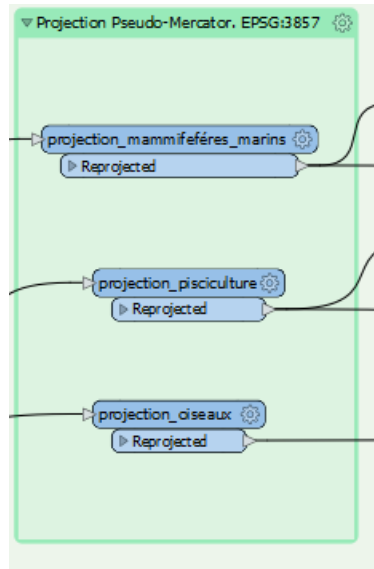


Figure 14 - Exemple du flux de travail pour l'adéquation de la syntaxe des coordonnées

3.7 L'intégration et le chargement des données

À la suite des transformations appliquées à l'ensemble de valeurs conservées des différents jeux de données, toute l'information a été consolidée sans prendre en compte les contraintes de redondance ou de rigidité structurale d'une BD SQL. Avec ces données, une collection dans la BD NoSQL de type orientée document MongoDB a été créée. L'information des documents a été enregistrée sous le format MongoDB qui est une forme d'objets de type GeoJSON. En sortie du système NoSQL, un format Google Sheets a été généré avec FME pour faciliter le partage de données et le travail collaboratif entre les membres de l'équipe. La feuille de calcul a été traitée pour ajouter les liens des photos reliées aux observations. Cette liaison permet de faciliter la création de fenêtres contextuelles (*pop-ups*). La procédure est décrite à l'annexe 7.

La figure 15 présente l'étape d'intégration et le chargement des données à la fin du processus ETL.

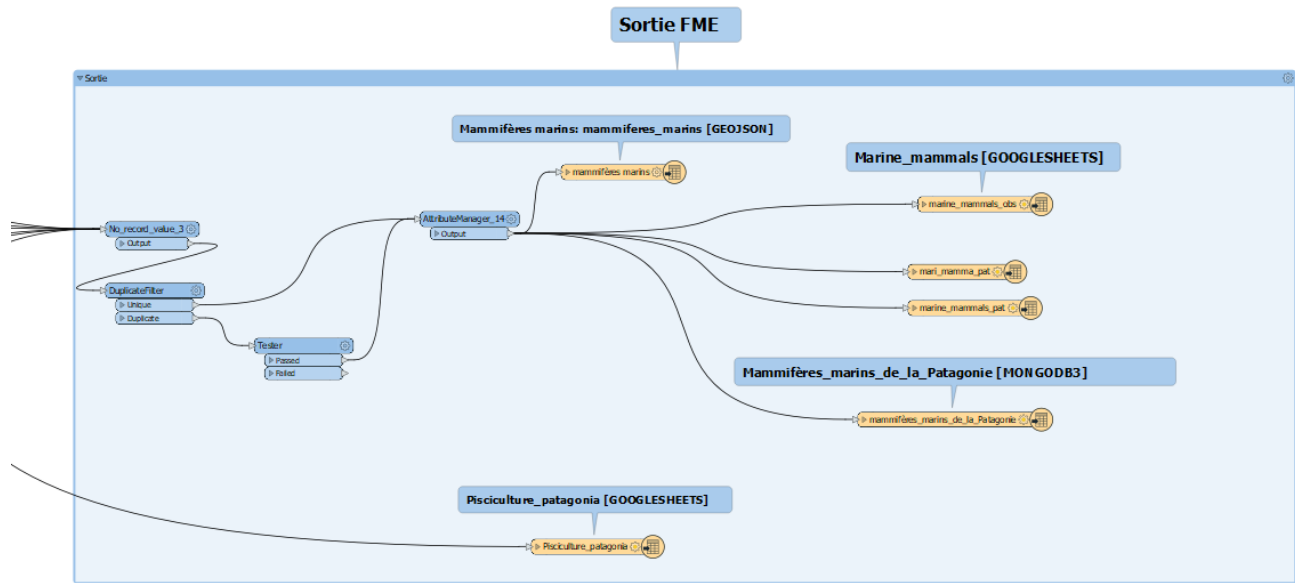


Figure 15 - Flux de travail pour l'étape d'intégration et de chargement des données

Afin de lister et décrire avec clarté la signification et le domaine des champs de la BD NoSQL, des métadonnées ont été mises en place et sont décrites à l'annexe 8

3.8 La mise en place du SIG web

En considérant la nature des données ainsi que les besoins des utilisateurs, l'architecture logicielle proposée est une solution hybride, c'est-à-dire une solution avec des composants *Open Source* et des composants propriétaires.

Au niveau de la gestion des données, l'information est stockée de deux manières :

- avec MongoDB disponible sur un serveur ouvert (igeomedia.com);
- avec des feuilles Google Sheets disponibles sur un compte Google Drive (Service infonuagique propriétaire).

Au niveau de l'interface client, l'information est affichée de deux manières :

- avec ArcGIS Online pour la réalisation des cartographies web;
- avec ArcGIS Dashboard pour la réalisation d'un tableau de bord analytique.

L'interface finale proposée pour ce projet sera la forme d'un tableau de bord analytique réalisé avec ArcGIS Dashboard, aussi appelé *Dashboard*. C'est une application (*App*) incorporée à la plateforme infonuagique ArcGIS Online. Cette interface a gagné une grande notoriété suite à l'apparition du tableau de bord déployé par l'Université Johns Hopkins pour suivre l'évolution de l'épidémie de Covid-19 (Pratt, 2020). Il s'agit d'un tableau de bord prêt à l'emploi capable de gérer de grandes quantités de données. Dashboard offre des gabarits avec fonctionnalités de bases pour mettre en place des présentations visuelles conviviales pour faciliter la visualisation et l'interprétation des données (*ArcGIS Dashboards / Tableaux de bord de données*, s. d.).

Dashboard peut assembler toutes les données d'un projet dans une seule vue afin de suivre l'évolution spatio-temporelle des événements, et synthétiser l'information pour aider la prise de décisions liées à la gestion des événements. Dashboard permet de visualiser des indicateurs ou KPIs (*key performance indicators*) de type jauge, graphiques linéaires, graphiques circulaires, etc. qui peuvent être configurés pour interagir avec l'utilisateur via des boutons sélecteurs (*Real-Time Decision Making with Operations Dashboard for ArcGIS*, 2020). Cet outil web est configurable sans passer par des étapes de programmation, et il permet un affichage adaptatif à la taille de l'écran (*responsive*). Toutefois, pour l'affichage optimal sur des smartphones, il faut créer le tableau de bord en priorisant l'affichage dans un format réduit (*Configuration requise de ArcGIS Dashboards—ArcGIS Dashboards / Documentation*, s. d.).

3.9 Le développement de l'interface web

L'interface web est composée d'une cartographie web intégrée à un tableau de bord. La figure 17 présente le schéma le flux de données à la suite du processus ETL.

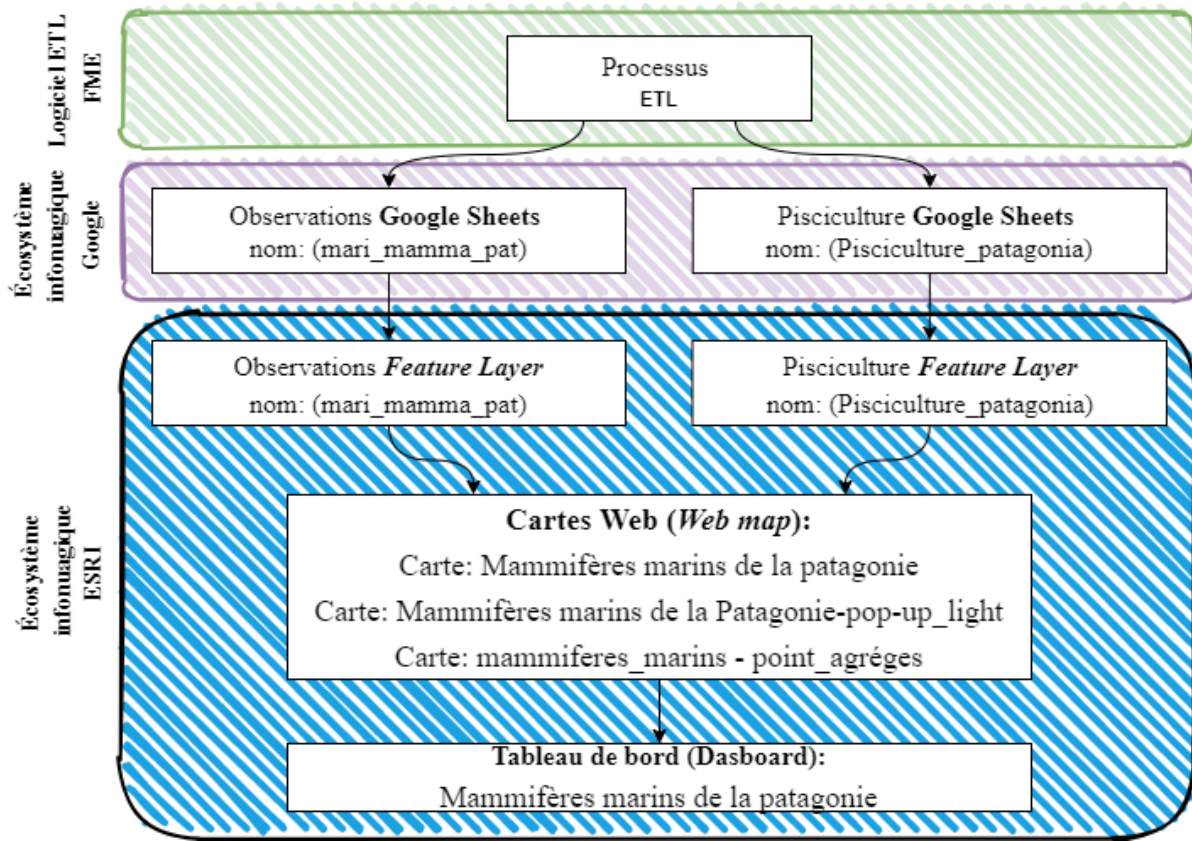


Figure 16 - Flux des données à la suite du processus ETL

4 Résultats

En suivant les étapes méthodologiques établies précédemment, nos résultats portent sur la création d'une base de données MongoDB, ainsi que la création du tableau de bord analytique.

4.1 Création d'une collection dans MongoDB.

Comme mentionné précédemment, toute l'information a été consolidée dans une collection de documents sous le format MongoDB hébergée sur le serveur du projet (igeomedia.com)

La figure 18 illustre l'aspect de l'interface client « Robo3T » qui permet un accès au SGBD NoSQL MongoDB, et l'affichage des documents créés à la suite du processus ETL.

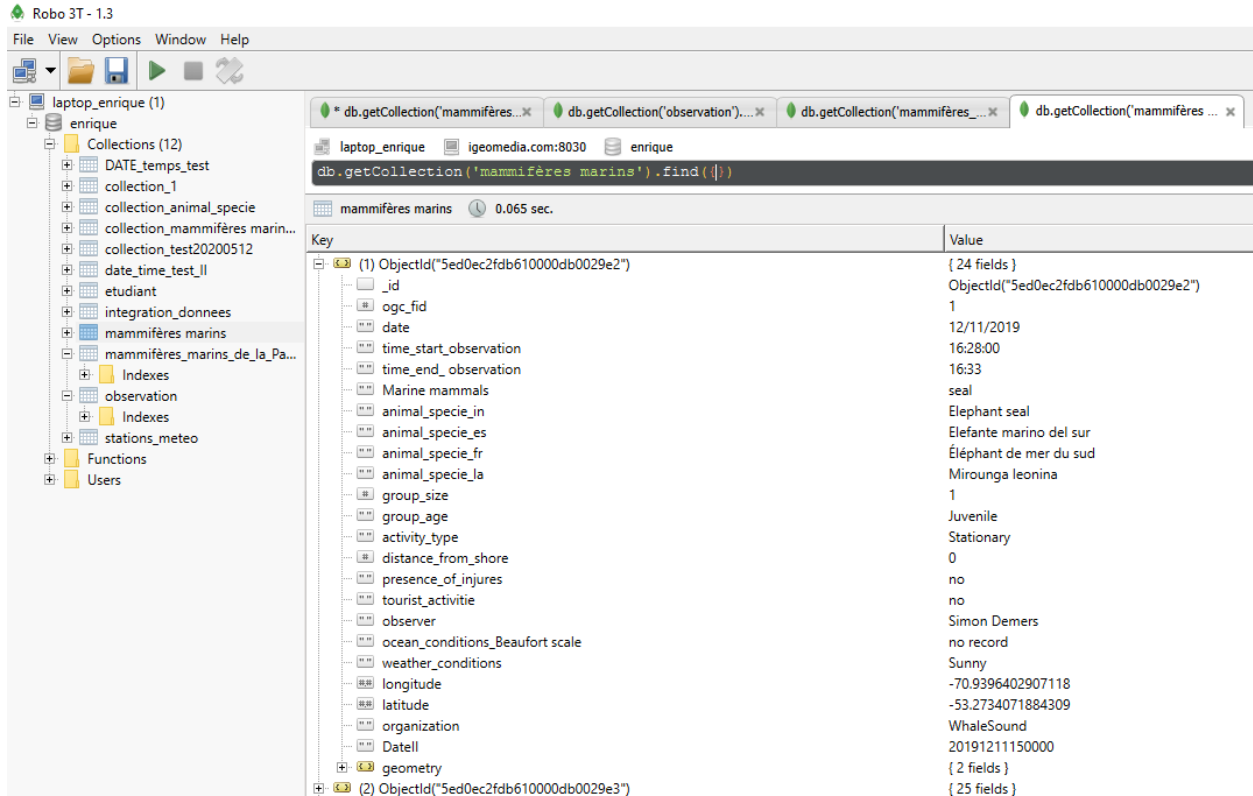


Figure 17 - Interface client « Robo3T » de MongoDB

4.2 Création de couches d'entité (*Feature Layer*) sur ArcGIS Online

La plateforme ArcGIS Online permet d'ajouter – et d'héberger – en tant que couche d'entité (*Feature Layer*) les feuilles de calcul Google Sheets créées lors du processus ETL. Cela permet aux utilisateurs autorisés de mettre à jour ou modifier ces documents. Pour ajouter Google Sheets en tant que couches d'entités et les héberger sur ArcGIS Online, il a fallu suivre la procédure dûment expliquée sur le site d'assistance technique d'ESRI pour accéder au Google Drive du projet <https://support.esri.com/en/technical-article/000022290> (*How To: Add and update Google Sheets data in ArcGIS Online*, s. d.) .

Deux couches d'entité ont été créées:

- **Mari_mamma_pat** : les données des observations sur les mammifères marins.
- **Pisciculture_patagonia** : les données des observations sur les fermes piscicoles.

Il faut mentionner que pour avoir accès aux feuilles de calcul sur Google Drive, il faut avoir le droit d'accès. De plus, pour ajouter ou mettre à jour des données ou des entités hébergées sur ArcGIS Online, il faut avoir le privilège d'utilisateur (accès) (*Types d'utilisateurs, rôles et privilèges—Aide ArcGIS Online | Documentation, s. d.*). Étant donné que les entités créées sont destinées à la création de cartes web et à leur diffusion via un tableau de bord, le niveau de partage de la couche a été configuré comme « public » (*public Sharing level*)

La figure 19 présente les étapes suivies dans la création de couches d'entité (*Feature Layer*) à partir de Google Sheets.

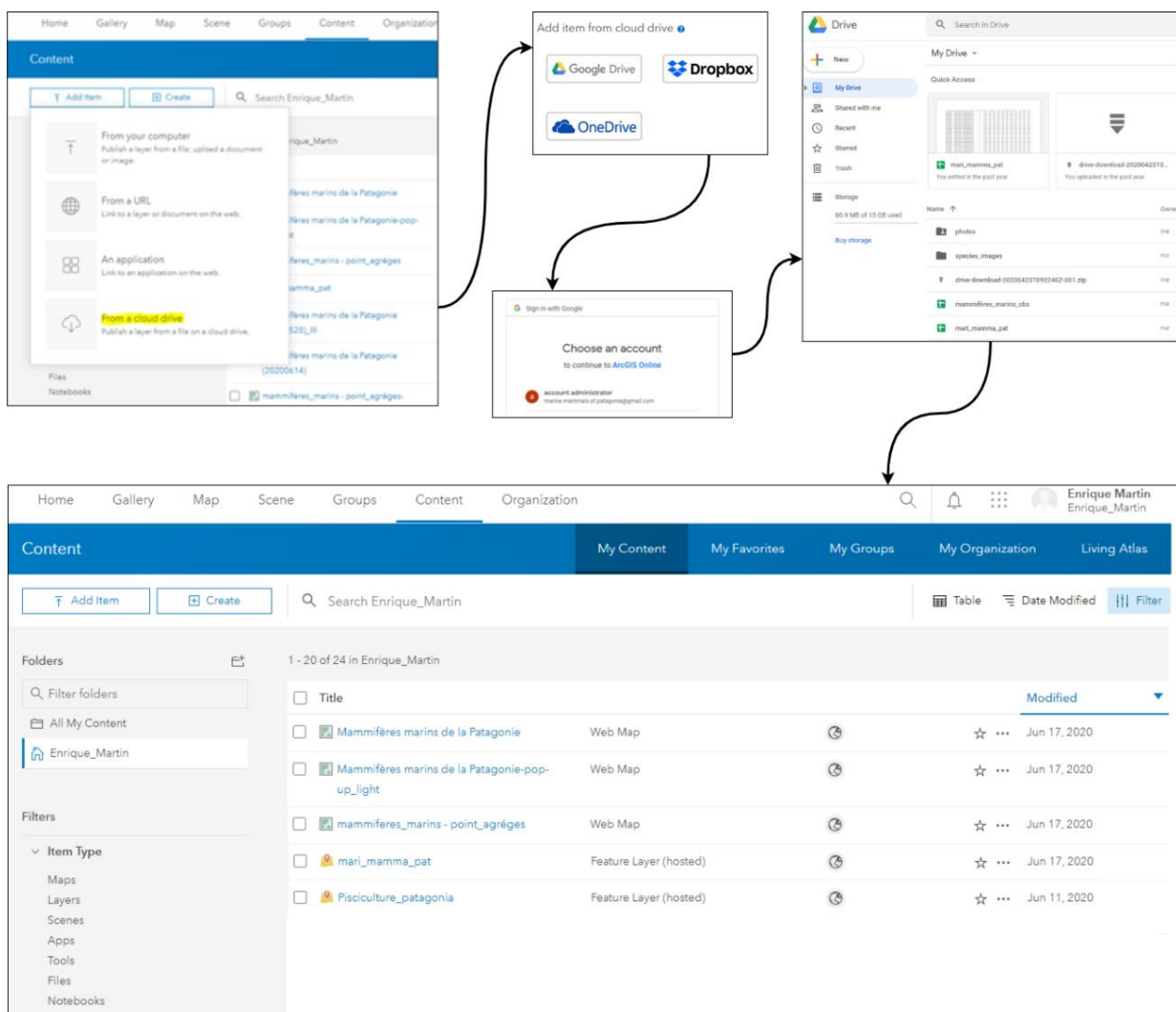


Figure 18 - Création de couches d'entité (*Feature Layer*) à partir de Google Sheets

4.3 La création de cartes web avec ArcGIS Online

La création des cartes web a suivi le protocole suivant :

- (1) Sélection d'un fond de carte adéquat pour donner un cadre référentiel aux données du projet. Cela implique la sélection d'une cartographie suffisamment riche et claire, mais qui ne pénalise pas l'importance des données affichées ni le repérage des données thématiques. Parmi les options disponibles, la thématique *World Dark Gray Canvas Base* a été privilégiée.
- (2) Ajout de couches avec les données. Cela implique l'utilisation des entités créées à partir des Google Sheets. C'est-à-dire à partir des couches **mari_mamma_pat** et **Pisciculture_patagonia**,
- (3) Configuration des styles pour mettre en valeur les données avec le fond de carte, mais, en même temps, représenter les données pour faciliter leur interprétation. Cela passe pour la configuration d'échelles d'affichage ainsi que l'agrégation des données.
- (4) Configuration de fenêtres contextuelles (*pop-ups*) pour les légendes et les infographies afin d'enrichir l'affichage des données descriptives.

Trois cartes ont été créées afin de faciliter la visualisation des données :

- **Mammifères marins de la Patagonie** : carte avec l'information des fenêtres contextuelles complète.
- **Mammifères marins de la Patagonie-pop-up_light** : carte avec l'information des fenêtres contextuelles abrégée.
- **mammiferes_marins - points_agrégés** : carte avec les observations en grappes selon l'échelle de visualisation.

La symbologie des trois cartes est la même pour permettre la configuration d'une légende unique pour le projet. Une fois les cartes créées, elles ont été enregistrées comme élément de contenu sur le compte ArcGIS Online du projet. Ces cartes destinées à la création d'une application Dashboard ont été partagées comme « public ».

4.4 Création du tableau de bord à partir d’ArcGIS Online

La figure 21 montre le gabarit conceptuel utilisé pour la configuration du tableau de bord du projet.

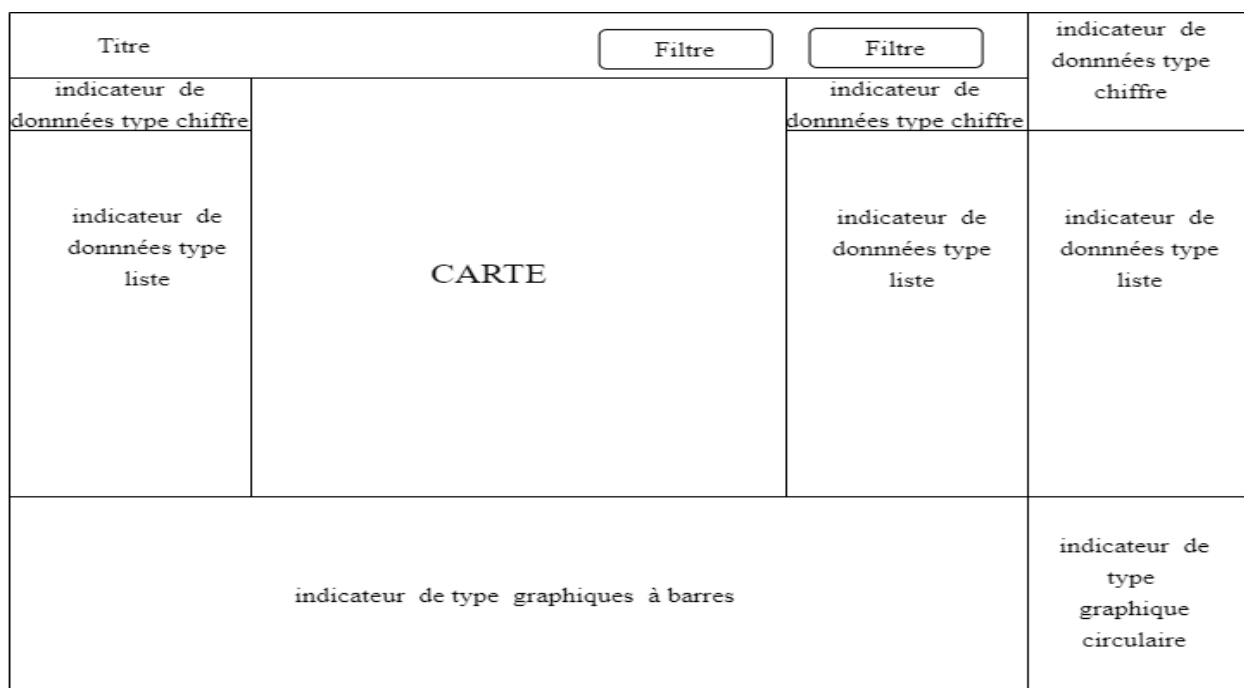


Figure 19 - diagramme de conception du tableau de bord

Pour chaque élément, l’apparence et la source qui y est reliée ont été configurées en détail. L’aspect de l’affichage a joué un rôle déterminant dans ces configurations afin d’avoir un produit agréable à utiliser et esthétique. La figure 22 présente l’aspect final de l’interface qui se trouve sur le site : <https://usherbrookegeo.maps.arcgis.com/apps/dashboards/1e3f163c5e1843ff98037657993130c2>

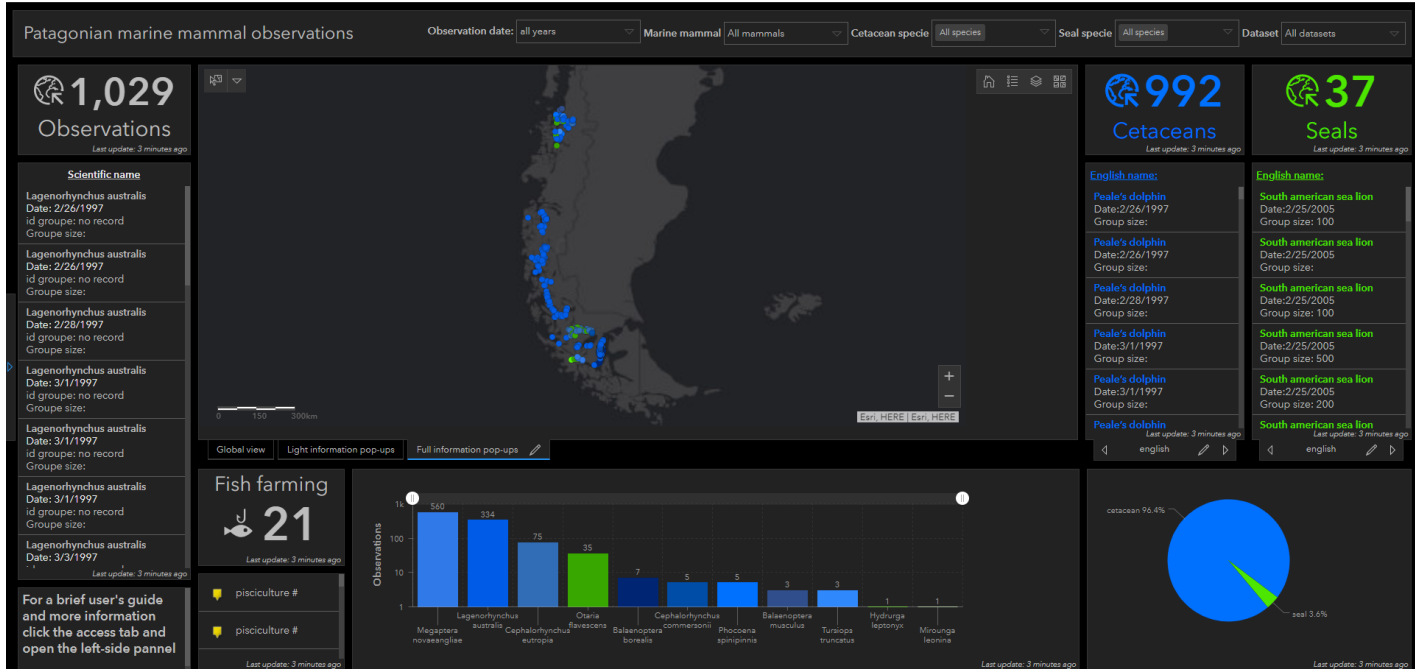


Figure 20 - Aspect final de l'interface

4.4.1 L'utilisation des cartes avec les données

Les cartes intégrées dans le tableau de bord permettent de passer d'une couche à une autre pour la même étendue géographique. Une interface a trois boutons, avec un texte identifiant, permet de sélectionner la carte à afficher (voir figure 23).

Bien que la symbologie des cartes web a été respectée par l'application Dashboard, il a fallu incorporer certains outils d'interaction pour bonifier la convivialité de l'interface. Parmi ces outils d'interaction, les éléments suivants ont été ajoutés :

- Dans la marge supérieure gauche : bouton pictogramme (icône) pour déployer un outil d'agrandissement des secteurs sélectionnés dans la carte (agrandissement par sélection).
- Dans la marge inférieure gauche : boutons avec texte identifiant pour choisir la carte à afficher.
- Dans la marge supérieure droite:
 - bouton pictogramme (icône) qui permet de revenir à la vue initiale sur la carte;
 - bouton pictogramme (icône) pour déployer la légende de la carte;
 - bouton pictogramme (icône) pour déployer un sélecteur de visibilité des couches;

- bouton pictogramme (icône) pour changer la couche de base. Option à l'utilisateur d'ajouter d'autres font de cartes (16 options);
- bouton pour affichage plein écran.
- Dans la marge inférieure droite: bouton de zoom pour effectuer des zooms avant ou arrière sur l'écran. *Zoom in/Zoom out*

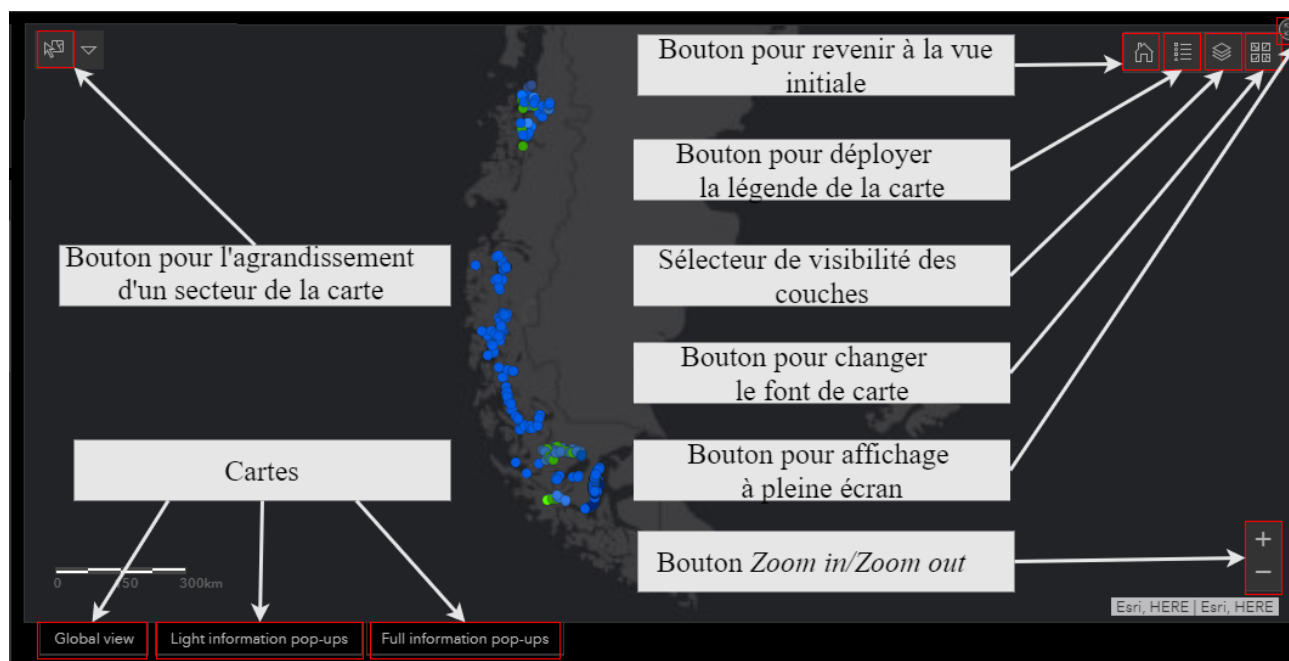


Figure 21 - Boutons ajoutés à la carte

Les coordonnées affichent des valeurs métriques, car Dashboard utilise la projection cartographique EPSG 3857 (Mercator) par défaut. En ce qui concerne les autres actions de base, elles sont classiques des interfaces cartographiques web (zooms, déplacements, etc.)

4.4.2 La liste des observations affichées

La liste des observations affichées est un élément qui sert à identifier les données. Il existe plusieurs listes reliées aux entités, et configurées pour interagir avec les cartes. Une sélection sur la liste filtre l'information sur la carte, et inversement.

Les éléments de la liste ont été configurés de manière à montrer les attributs les plus pertinents afin de fournir une information précise. Étant donné que l'interface a été conçue pour être utilisée par des personnes de différentes langues, huit listes ont été ajoutées :

- liste des noms **latins** des observations affichées;

- liste en langue **anglaise** des noms communs des **cétacés** observés;
- liste en langue **française** des noms communs des **cétacés** observés;
- liste en langue **espagnole** des noms communs des **cétacés** observés;
- liste en langue **anglaise** des noms communs des **pinnipèdes** observés;
- liste en langue **française** des noms communs des **pinnipèdes** observés;
- liste en langue **espagnole** des noms communs des **pinnipèdes** observés;
- liste des fermes piscicoles observées.

La figure 24 illustre les listes ajoutées à l'interface



Figure 22 - Listes ajoutées à l'interface

Des boutons de sélection des listes ont été ajoutés. Les dimensions, la couleur de fond, la couleur du texte, ainsi que la police du texte des listes et des boutons ont été configurées pour garder l'harmonie de l'affichage.

4.4.3 Les éléments infographiques

Deux types d'indicateurs analytiques prévus dans le diagramme de conception du Dashboard ont été ajoutés. Ces indicateurs ont été paramétrés afin de faciliter l'analyse interactive. Les éléments de visualisation infographique qui ont été configurés sont les suivants :

- un graphique à barres pour indiquer la quantité des observations par espèce;
- un graphique circulaire pour indiquer le pourcentage de chaque type de mammifère parmi les observations affichées.

La figure 25 présente le placement des éléments de visualisation sur le tableau de bord.

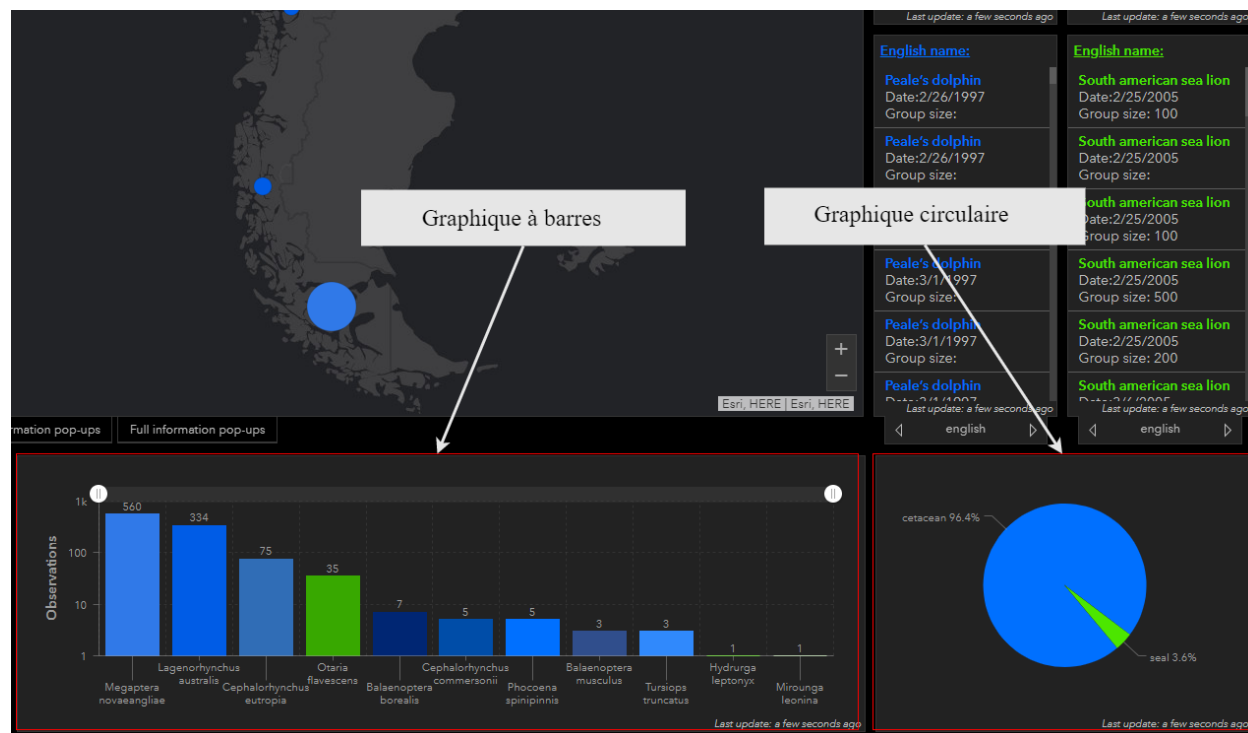


Figure 23 - Éléments infographiques du tableau de bord

4.4.4 Les indicateurs numériques

Les indicateurs numériques ajoutés au tableau de bord servent à afficher les sommes de chaque liste d'observations. Ils fonctionnent suivant l'étendue géographique de la carte. Quatre indicateurs numériques ont été ajoutés:

- un indicateur numérique de la quantité d'observations de cétacés;
- un indicateur numérique de la quantité totale d'observations;
- un indicateur numérique de la quantité d'observations de fermes piscicoles;
- un indicateur numérique de la quantité d'observations de pinnipèdes.

La figure 26 illustre le placement des indicateurs numériques ajoutés.



Figure 24 - Indicateurs numériques ajoutés

4.4.5 Les filtres d'affichage

Afin d'affiner la recherche de données ainsi que leur affichage, des boutons de filtre ont été incorporés. Ces boutons servent à déployer des outils type « liste déroulante » qui permettent lancer de requêtes destinées à trier les données. L'ensemble de ces filtres constitue un groupe de contrôle (*control group*) afin de garantir un affichage cohérent. Les filtres définis sont les suivants (figure 27) :

- un bouton de filtre **Observation date** qui ouvre un menu déroulant pour choisir les années des observations à afficher;
- un bouton de filtre **Marine mammal** qui ouvre un menu déroulant pour choisir l'affichage des observations de cétacés ou l'affichage des observations de pinnipèdes;
- un bouton de filtre **Cetacean specie** qui ouvre un menu déroulant pour choisir l'affichage des observations parmi les espèces de cétacés;
- un bouton de filtre **Seal specie** qui ouvre un menu déroulant pour choisir l'affichage des observations parmi les espèces de pinnipèdes;

- un bouton de filtre **Dataset** qui ouvre un menu déroulant pour choisir l'affichage des observations selon le jeu de données source.

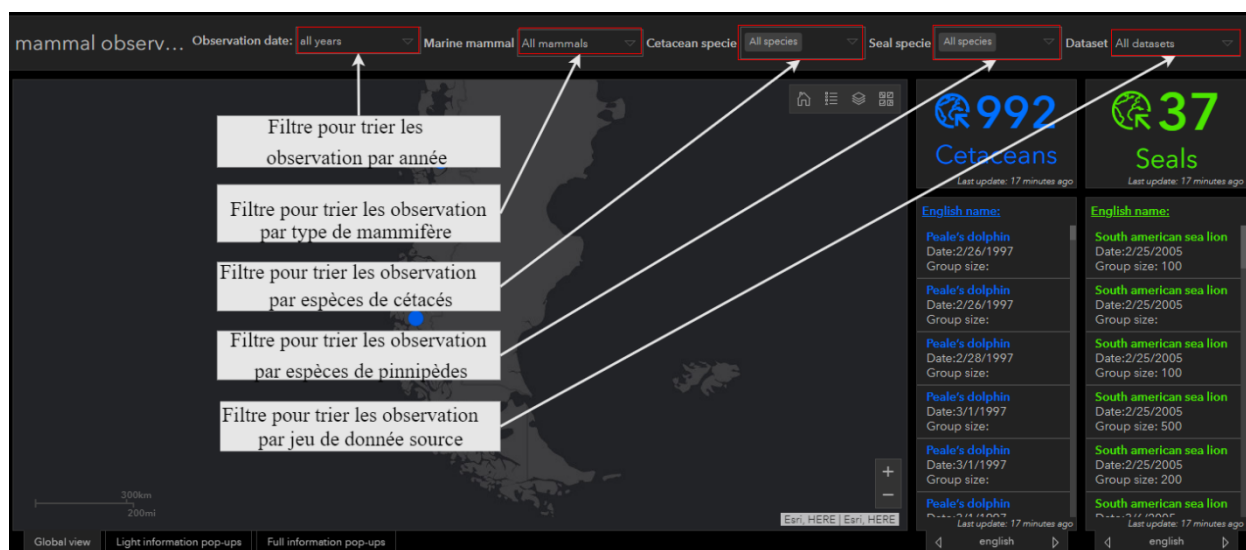


Figure 25 - Filtres d'affichage

4.5 Mise au point du tableau de bord

Tous les éléments du tableau de bord ont été configurés avec un bouton d'extension pour l'affichage plein écran. Le bouton pour activer cette fonctionnalité devient visible quand le curseur de la souris est sur un des éléments du Dashboard. Aussi, pour déployer l'information de la thématique affichée, à la place d'un bouton d'information, les fenêtres contextuelles ont été configurées avec des liens pour accéder au fichier avec des images génériques des espèces ou des photographies prises lors des observations (figure 28).



Figure 26 - Fenêtres contextuelles avec images

Il a été incorporé un panneau déroulant dans la marge gauche du Tableau de bord. Ce panneau contient une brève description des fonctionnalités, les coordonnées pour contacter l'administrateur du tableau de bord ainsi que le logo de l'Université de Montréal.

5 Interprétation et discussion des résultats

Sur le plan de la gestion des données, le SGBD NoSQL sélectionné permet de gérer les données hétérogènes du projet, et l'ensemble des informations géographiques. Il permet une très grande flexibilité pour l'ajout de documents dans la collection du système pour de futures données d'observations. De plus, les données extraites du SGBD NoSQL sont interoperables en passant par la plateforme ArcGIS Online après une simple transformation ETL. ArcGIS Online permet la création d'un service WFS ou encore de distribuer l'information dans un format web ou traditionnel : shapefile, CSV, KML, Excel, GeoJSON, etc.

L'interface web choisie pour le SIG web du projet est l'outil géodécisionnel ArcGIS Dashboard. Il s'agit d'un tableau de bord qui répond au besoin de nos partenaires pour afficher les données correspondant à huit jeux de données et à plus de 20 ans de données géographiques liées aux mammifères marins. Cette interface permet l'étude de la distribution, de la mobilité et de la taille des groupes de dauphins chiliens dans les canaux patagoniques. L'interface répond aussi à la demande de flexibilité de l'interface pour l'intégration de futures données, notamment des données de l'application de cartographie participative et collaborative ICETACE.

L'élément cartographique web permet de filtrer les données afin d'aider à l'étude des interactions existantes entre les différents groupes de dauphins chiliens ainsi que leur affinité avec des groupes d'autres espèces de cétacés. Les listes de données affichées, les indicateurs numériques ainsi que les indicateurs analytiques (le graphique à barres et le graphique circulaire) qui ont été ajoutés au tableau de bord viennent enrichir l'analyse spatiale et constituent un ensemble qui permet une analyse intuitive et simple.

Concernant la portabilité et l'accessibilité de l'interface web, ce sont des éléments qui restent à définir. Même si l'interface fournit un environnement de travail qui s'adapte à la taille et aux formats des écrans d'ordinateurs et des tablettes, l'affichage du tableau de bord n'a été pas optimisée pour la taille plus petite d'affichage (téléphone intelligent). Cependant, la visualisation du site avec un téléphone intelligent n'est pas impossible. Sur le point de l'accessibilité, l'interface est partagée en mode « public ». Il serait bien de mettre en place un système multi-utilisateur et des privilèges relatifs pour les administrateurs, les fournisseurs de données et les utilisateurs.

En résumé, Le SGBD et l'interface web proposés répondent aux demandes de nos partenaires pour la gestion, la visualisation, l'analyse et l'aide à la décision des différentes observations liées aux mammifères marins.

6 Conclusion

Le projet a été une occasion de réfléchir aux critères de sélection d'un SGBD géospatiales et d'une interface client facilitant la gestion, l'affichage et l'analyse des données, ainsi que de leurs utilisations. Nous avons également utilisé des outils de l'intelligence géodécisionnelle tels que les processus ETL pour simplifier l'analyse, l'extraction, et le traitement des données qui proviennent des différentes sources. Finalement, nous avons mis en place une base de données géospatiales ainsi qu'un tableau de bord analytique afin d'améliorer le partage, et la connaissance du Dauphin chilien ainsi que des autres espèces maritimes de la Patagonie chilienne. Les objectifs du projet ont été atteints.

Ce travail a permis de prendre en main plusieurs nouveaux outils comme MongoDB pour le SGBD et ArcGIS Dashboard pour le tableau de bord. Ils permettent de mettre en place rapidement une solution sans perdre du temps sur la normalisation des données pour le SGBD, et sur la programmation pour le tableau de bord. Ce sont des outils flexibles pour l'intégration de nouvelles données et qui sont accessibles à travers des processus ETL. La solution proposée est une solution alternative aux SIG web qui sont développés à partir de SGBD relationnelles et avec des interfaces web programmées.

7 Recommandations

Le système proposé est une première version et des améliorations peuvent être apportées. Ces améliorations n'ont pas été effectuées par manque de temps dans le cadre d'un essai de maîtrise type cours.

Au niveau du SGBD, la solution MongoDB est un système de gestion très performant. Cependant, les données sont « statiques » dans le serveur et les analyses sont limitées. Il serait intéressant de tirer profit des données contenues dans la BD pour mettre en place des outils d'analyse basés sur l'apprentissage machine comme on retrouve sur des systèmes d'entrepôt de données tels que BigQuery SIG, et ainsi générer des requêtes plus complexes.

Au niveau de données, une amélioration serait l'inclusion de données plus complètes sur la pisciculture, car au-delà du fait d'augmenter la quantité d'observation des fermes piscicoles, il est prévu dans la configuration des fenêtres contextuelles l'amélioration de la qualité des observations en ajoutant des attributs descriptifs de l'activité. Cela doit être une action conjointe de différents organismes et de groupes de recherche qui tirent profit du Dashboard du projet.

Afin d'améliorer l'expérience des utilisateurs, d'autres fonctionnalités sont prévues sur l'interface web. Ces fonctionnalités seront développées en suivant les différentes thématiques des couches à ajouter qui contiennent de l'information pertinente pour les administrateurs. Cependant, il faut retenir qu'un outil géodécisionnel doit afficher des informations pertinentes pour permettre une lecture analytique facile et intuitive.

Enfin, il faudrait également envisager de bonifier l'interface avec des sources de données en temps réel étant donné qu'ArcGIS Dashboard offre la possibilité d'importer des données provenant, notamment des flux en temps réel, avec l'extension ArcGIS GeoEvents (*ArcGIS Dashboards / Tableaux de bord de données*, s. d.) . ArcGIS Dashboard offre aussi la possibilité d'importer des données provenant de sources différentes de l'écosystème ESRI tel que les services WMS, WFS ou des formats de type fichiers. Il serait intéressant d'ajouter une liste de couches contextuelles en provenance d'autres serveurs afin d'améliorer l'analyse spatiale.

Références

- Abramson, J. Z., & Gibbons, J. (2010). NEW RECORDS OF BLUE WHALES BALAENOPTERA MUSCULUS (LINNAEUS, 1758) IN WINTER SEASON IN THE INLET WATERS OF CHILOÉ CONTINENTAL-CHILE. *Anales Del Instituto de La Patagonia*, 38(2), 107-109. <https://doi.org/10.4067/S0718-686X2010000200010>
- ArcGIS Dashboards | Tableaux de bord de données : Opérationnels, stratégiques, tactiques, informatifs.* (s. d.). Consulté 15 juillet 2020, à l'adresse <https://www.esri.com/fr-fr/arcgis/products/arcgis-dashboards/overview>
- Bédard, Y. (2010). *Le géodécisionnel : Origine, évolution, état de l'art, enjeux, R&D.* 105.
- Behmo, R., & Travers, N. (2019a). *Choisissez votre famille NoSQL.* OpenClassrooms. <https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql/4462433-choisissez-votre-famille-nosql>
- Behmo, R., & Travers, N. (2019b, novembre 28). *Maitrisez le théorème de CAP.* OpenClassrooms. <https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql/4462471-maitrisez-le-theoreme-de-cap>
- Burggraf, D. (2015, août 4). *OGC KML 2.3 [Implementation Standard].* OGC KML SWG; Open Geospatial Consortium. <http://docs.opengeospatial.org/is/12-007r2/12-007r2.html#1>
- Butler, H., Daly, M., Doyle, S., Gillies, S., Hagen, S., & Shaub, T. (2002). The GeoJSON Format. *Internet Engineering Task Force (IETF)*, 10(05), 557-570. <https://doi.org/10.1142/S0218488502001648>
- Capella, J. J., Abramson, J. Z., Vilina, Y. A., & Gibbons, J. (2014). Observations of killer whales (*Orcinus orca*) in the fjords of Chilean Patagonia. *Polar Biology*, 37(10), 1533-1539. <https://doi.org/10.1007/s00300-014-1535-5>

Capella, J., Vilina, Y., & Aa, F. Y. (s. d.). *OBSERVACIÓN DE CETÁCEOS EN ISLA CHAÑARAL Y NUEVOS REGISTROS PARA EL ÁREA DE LA RESERVA NACIONAL PINGÜINO DE HUMBOLDT, NORTE DE CHILE*. 9.

Carlucci, R., Fanizza, C., Cipriano, G., Paoli, C., Russo, T., & Vassallo, P. (2016). Modeling the spatial distribution of the striped dolphin (*Stenella coeruleoalba*) and common bottlenose dolphin (*Tursiops truncatus*) in the Gulf of Taranto (Northern Ionian Sea, Central-eastern Mediterranean Sea). *Ecological Indicators*, 69, 707-721.
<https://doi.org/10.1016/j.ecolind.2016.05.035>

Chafki, S., & Desrosiers, C. (2011). *Intégration des données et ETL*.
file:///C:/Users/Enrique/Documents/USherbrooke/Essai_2020/materiel%20d_interet/MTI820-Acetates-ETL_1pp.pdf

Compare Data Transformation Tools | Top 10 Platforms in 2020. (s. d.). Consulté 8 août 2020, à l'adresse https://www.businessintelligencemarket.com/compare-data-transformation-reviews-business-intelligence-software?gclid=CjwKCAjwmrn5BRB2EiwAZgL9okNzSaESpYJWi1x2U0gEkaQp83hHKPzHXPxileFpWuJ-CnNTs7RdDhoCru0QAvD_BwE

Configuration requise de ArcGIS Dashboards—ArcGIS Dashboards | Documentation. (s. d.). Consulté 13 août 2020, à l'adresse https://doc.arcgis.com/fr/dashboards/reference/requirements.htm#ESRI_SECTION1_9D6458CE33F543078226DC514CC2EC6B

Cornuéjols, A. (2009). *BASES DE DONNÉES CONCEPTS ET PROGRAMMATION*.
<https://www.lri.fr/~antoine/Courses/AGRO/Cours-BD/Poly-BD.pdf>

- Coronel, C., Morris, S., & Croquett, K. (2020). *DATABASE PRINCIPLES fundamentals of design, implementation, and management*. CENGAGE LEARNING EMEA.
- Coronel, C., Morris, S., & Rob, P. (2011). *Database systems : Design, implementation, and management*. Course Technology.
- CouchDB : le guide authentique*. (s. d.). Consulté 15 janvier 2021, à l'adresse <http://guide.couchdb.org/editions/1/fr/index.html>
- Crozat, S. (s. d.). *Introduction Générale aux bases de données*. Consulté 28 mars 2020, à l'adresse <https://stph.scenari-community.org/idl-bd/1/co/pri1c14.html>
- Crozat, S. (2016a). *Introduction à l'ETL et application avec Oracle*. utcFormation. <https://stph.scenari-community.org/dwh/etl1.pdf>
- Crozat, S. (2016b, septembre 1). *Introduction à la modélisation dimensionnelle*. <https://stph.scenari-community.org/dwh/mod1.pdf>
- Crozat, S. (2016c, septembre 1). *Introduction au domaine du décisionnel et aux data warehouses*. <https://stph.scenari-community.org/dwh/int.pdf>
- Cukier, K. N., & Mayer-Schoenberger, V. (2013). *The Rise of Big Data*. <http://faculty.cord.edu/andersod/The%20Rise%20of%20Big%20Data.docx>
- Définition d'un tableau de bord—ArcGIS Dashboards | Documentation*. (2020). <https://doc.arcgis.com/fr/dashboards/get-started/what-is-a-dashboard.htm>
- Dellabianca, N. A., Pierce, G. J., Raya Rey, A., Scioscia, G., Miller, D. L., Torres, M. A., Paso Viola, M. N., Goodall, R. N. P., & Schiavini, A. C. M. (2016). Spatial Models of Abundance and Habitat Preferences of Commerson's and Peale's Dolphin in Southern Patagonian Waters. *PLOS ONE*, *11*(10), e0163441. <https://doi.org/10.1371/journal.pone.0163441>

- Elmasri, R., & Navathe, S. (2016). *Fundamentals of database systems* (Seventh edition). Pearson.
- Extensible Markup Language (XML)*. (s. d.). Consulté 23 avril 2021, à l'adresse <https://www.w3.org/XML/>
- FME Desktop, manuel. (2020). *Data Integration Solutions for GIS and Location Intelligence*. Safe Software. <https://www.safe.com/data-types/gis-location-intelligence/>
- Gaetan, R. (2017, juillet 9). Qu'est-ce que le Sharding ? Définition et avantages de cette méthode de distribution des données. *LeBigData.fr*. <https://www.lebigdata.fr/sharding-definition-avantage>
- Garcia-Molina, Hector, H., Ullman, D. J., Widom, & Jennifer. (2002). *Database Systems : The Complete Book*.
- Garmany, J., Walker, J., & Clark, T. (2005). *Logical Database Design Principles*. CRC Press.
- Geography Markup Language / OGC*. (s. d.). Consulté 6 août 2020, à l'adresse <https://www.ogc.org/standards/gml>
- Germain, M. (2020). *GÉOMATIQUE SUR INTERNET Cours8*.
- Gibbons, J., Capella, J. J., & Valladares, C. (2003). *Rediscovery of a humpback whale (Megaptera novaeangliae) feeding ground in the Straits of Magellan, Chile*. 7.
- Godin, R., & Desrosiers, C. (2011). *Les entrepôts de données et l'analyse de données*. https://cours.etsmtl.ca/log660/public_docs/acetates/LOG660-Acetates-EntrepotsDeDonnees_1pp.pdf
- Guzman, H. M., Capella, J. J., Valladares, C., Gibbons, J., & Condit, R. (2020). Humpback whale movements in a narrow and heavily-used shipping passage, Chile. *Marine Policy*, 118, 103990. <https://doi.org/10.1016/j.marpol.2020.103990>

Heinrich, S., Genov, T., Fuentes Riquelme, M., & Hammond, P. S. (2019). Fine-scale habitat partitioning of Chilean and Peale's dolphins and their overlap with aquaculture. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29(S1), 212-226.
<https://doi.org/10.1002/aqc.3153>

How To : Add and update Google Sheets data in ArcGIS Online. (s. d.). Consulté 13 août 2020, à l'adresse <https://support.esri.com/en/technical-article/000022290>

How To : Use Google Drive photos in ArcGIS Online pop-ups. (s. d.). Consulté 18 août 2020, à l'adresse <https://support.esri.com/en/technical-article/000015466>

IBM. (2021, mars 23). *Cap-theorem.* IBM Cloud Learn Hub.
<https://www.ibm.com/cloud/learn/cap-theorem>

Law, D. (2019, décembre 8). Dashboards in ArcGIS. *ArcGIS Blog.* <https://www.esri.com/arcgis-blog/products/ops-dashboard/decision-support/dashboards-in-arcgis/>

Lelong, C. (2016). *Bases de données spatiales PostgreSQL/PostGIS.* 28.

Martin, J. (1983). *Managing the data-base environment.* Englewood Cliffs, N.J. : Prentice-Hall.
<http://archive.org/details/managingdatabase00mart>

Mehra, A. (2019). *Understanding the CAP Theorem—DZone Database.* Dzone.Com.
<https://dzone.com/articles/understanding-the-cap-theorem>

ModelBuilder | ArcGIS for Desktop. (s. d.). Consulté 6 mai 2020, à l'adresse <https://desktop.arcgis.com/fr/arcmap/10.3/analyze/modelbuilder/what-is-modelbuilder.htm>

MongoDB Manual. (2020). <https://Github.Com/Mongodb/Docs/Blob/v4.2/Source/Core/Write-Operations-Atomicity.Txt>. <https://docs.mongodb.com/manual/core/write-operations-atomicity>

- Murphy, J. (2012, mai 2). ESRI, ArcGIS Blog, Adding a NoSQL Plug-in. *ArcGIS Blog*.
<https://www.esri.com/arcgis-blog/products/arcgis-desktop/data-management/mongodb-example-code-for-adding-a-nosql-plug-in-data-source/>
- Näsholm, P. (2012). *Extracting Data from NoSQL Databases*.
<http://publications.lib.chalmers.se/records/fulltext/155048.pdf>
- Office québécois de la langue française,. (2020). *FICHE TERMINOLOGIQUE*.
http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26507313
- OGC Standards / OGC*. (s. d.). Consulté 18 mai 2021, à l'adresse <https://www.ogc.org/docs/is>
- Pérez-Alvarez, M. J., Olavarria, C., Moraga, R., Baker, C. S., Hamner, R. M., & Poulin, E. (2016). Historical dimensions of population structure in a continuously distributed marine species : The case of the endemic Chilean dolphin. *Scientific Reports*, 6(1), 35507.
<https://doi.org/10.1038/srep35507>
- Ponniah, P. (2003). *Database Design and Development : An Essential Guide for IT Professionals*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471728993>
- Pratt, M. (2020, juin 1). GIS Systems Lead Response to COVID-19. *Esri*.
<https://www.esri.com/about/newsroom/arcuser/gis-systems-lead-response-to-covid-19/>
- QGIS 3.10 Manual*. (s. d.). Consulté 6 mai 2020, à l'adresse
https://docs.qgis.org/3.10/en/docs/training_manual/processing/modeler_twi.html?highlight=processing%20modeler
- QGIS Mongodb Python Plugins Repository*. (s. d.). Consulté 6 mai 2020, à l'adresse
<https://plugins.qgis.org/plugins/tags/mongodb/>
- Qlik GeoAnalytics*. (s. d.). Consulté 9 août 2020, à l'adresse <https://www.qlik.com/fr-fr/products/qlik-geoanalytics>

- Query Documents—MongoDB Manual.* (s. d.). <https://Github.Com/Mongodb/Docs-Bi-Connector/Blob/DOCSP-3279/Source/Index.Txt>. Consulté 24 avril 2021, à l'adresse <https://docs.mongodb.com/manual/tutorial/query-documents/>
- Raffet, B. (2017, juin 25). *SQL vs noSQL : Quelles différences, pour quels projets?* | Sourceamax. <http://www.sourceamax.com/sql-vs-nosql-quelles-differences-pour-quels-projets/>
- Ramsey, P., & Leslie, M. (2012). *Introduction à PostGIS.* <http://postgis.fr/chrome/site/docs/workshop-foss4g/doc/introduction.html>
- Real-Time Decision Making with Operations Dashboard for ArcGIS.* (2020). https://www.esri.com/training/Engine/defaultui/player/modern.html?configuration=ReturnUrl%7C&preventRightClick=False&cc=fr-FR&ieCompatibilityMode=none&cache=20.1.10.206&playerConfUrl=%2Ftraining%2FEngine%2FPlayerConfiguration.aspx®istration=InstanceId%7C0%21LearningObjectId%7C5a2f2af490f2fd5a2e37a7cb-16147%21UserId%7CEnrique_Martin&package=LearningObjectId%7C5a2f2af490f2fd5a2e37a7cb-16147%21VersionId%7C1&tracking=True&forceReview=False
- Reed, C. (2017, août 18). *OGC GeoRSS Encoding Standard* [OGC® Community Standard Approved English]. Open Geospatial Consortium. <http://docs.opengeospatial.org/cs/17-002r1/17-002r1.html>
- Ribeiro, S., Viddi, F. A., Cordeiro, J. L., & Freitas, T. R. O. (2007). Fine-scale habitat selection of Chilean dolphins (*Cephalorhynchus eutropia*) : Interactions with aquaculture activities in southern Chiloé Island, Chile. *Journal of the Marine Biological Association of the United Kingdom*, 87(1), 119-128. <https://doi.org/10.1017/S0025315407051594>

- Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph Databases* (Second Edition). O'Reilly Media, Inc. file:///C:/Users/User/Downloads/OReilly_Graph_Databases.pdf
- Serre, D. (2016). *Concevoir la résilience urbaine : Un défi face à des complexités*. 18.
- SpatiaLite : SpatiaLite*. (s. d.). Consulté 6 août 2020, à l'adresse <https://www.gaia-gis.it/fossil/libspatialite/index>
- SQLite Home Page*. (s. d.). Consulté 6 août 2020, à l'adresse <https://www.sqlite.org/index.html>
- Tannir, K. (2015). *Introduction aux bases de données NoSQL*. 55.
- Techroba. (2015). *10 Open Source ETL Tools*.
<https://www.datasciencecentral.com/profiles/blogs/10-open-source-etl-tools>
- Teorey, T. J., Lightstone, S. S., Nadeau, T., & Jagadish, H. V. (2011). *Database Modeling and Design : Logical Design*. Elsevier Science & Technology.
<http://ebookcentral.proquest.com/lib/usherbrookemgh-ebooks/detail.action?docID=667713>
- Thierion, V. (2010). *Contribution à l'amélioration de l'expertise en situation de crise par l'utilisation de l'informatique distribuée : Application aux crues à cinétique rapide*.
Vincent Thierion. Contribution à l'amélioration de l'expertise en si École Nationale Supérieure des Mines de Paris.
- Types d'utilisateurs, rôles et privilèges—Aide ArcGIS Online | Documentation*. (s. d.). Consulté 13 août 2020, à l'adresse <https://doc.arcgis.com/fr/arcgis-online/reference/roles.htm>
- Viddi, F. A., Harcourt, R. G., & Hucke-Gaete, R. (2016). Identifying key habitats for the conservation of Chilean dolphins in the fjords of southern Chile: Key habitats and conservation of Chilean dolphins in Southern Chile. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26(3), 506-516. <https://doi.org/10.1002/aqc.2553>

- Viddi, F. A., & Lescrauwaet, A.-K. (2005). Insights on Habitat Selection and Behavioural Patterns of Peale's Dolphins (<I>Lagenorhynchus australis</I>) in the Strait of Magellan, Southern Chile. *Aquatic Mammals*, 31(2), 176-183. <https://doi.org/10.1578/AM.31.2.2005.176>
- Web Feature Service / OGC*. (s. d.). Consulté 1 août 2020, à l'adresse <https://www.ogc.org/standards/wfs>
- Web Graphics*. (s. d.). Consulté 2 août 2020, à l'adresse <https://www.w3.org/Graphics/>
- Web Map Service / OGC*. (s. d.). Consulté 27 juillet 2020, à l'adresse <https://www.ogc.org/standards/wms>
- Welcome to Data Studio! (Start here)*. (s. d.). Google Data Studio. Consulté 10 août 2020, à l'adresse <http://datastudio.google.com/reporting/0B5FF6JBKbNJxOWItcWo2SVVVeGc>
- Würsig, B., Thewissen, J. G. M., & Kovacs, K. (s. d.). *Encyclopedia of Marine Mammals—3rd Edition*. Consulté 13 janvier 2021, à l'adresse <https://www.elsevier.com/books/encyclopedia-of-marine-mammals/wursig/978-0-12-804327-1>

Annexe 1 Analyse de besoins (requetés)

Hiérarchies dimensionnelles(Godin & Desrosiers, 2011)

- Visualisation des observations dans une carte
- Visualisation des informations à propos les observations
- Visualisation des observations qui ont été effectués entre deux dates différents
- Visualisation que des observations d'une espèce dans un secteur sélectionné
- Visualisation que des observations de cétacés
- Visualisation que des observations d'un jeux de donnés
- Visualisation de photos prises lors de l'observation
- Visualisation de l'informations à propos l'activité des animaux observés
- Visualisation d proportions des observations par espèces qui ont été observées dans un secteur
- Visualisation des lieux où ont été effectuées des observations de différentes espèces de manière simultanée dans une période de l'année
- Visualisation de toutes les observations effectuées à une espèce animal
- Visualisation de toutes les observations effectuées sur un groupe identifié;
- Visualisation des coordonnées de observations de pisciculture
- Somme des observations de certaines espèces qui ont été effectuées dans un certain secteur
- Somme des observations pour une période de temps déterminée
- Accès à l'information à propos la distance de la coté des observations
- Accès à l'information à propos des conditions météorologiques lors des observations
- Accès à l'information à propos des conditions de la mer
- Accès à l'information à propos le type de platform utilisé lors des lors des observations
- Accès à l'information à propos l'observateur
- Accès à l'information à propos de animaux avec blessures dans une région

Annexe 2 : Évaluation séparée des sources données (exemple)

Nom : *Ballena_Jorobada_Estrecho.xls*

Type de source : Fichier Excel

Contenu : Observations de Baleine à bosse faites dans le détroit de Magellan l'année 2007

Origine : fourni par Simon Demers. Département de géographie de l'Université de Montréal

Nombre de lignes de la source 548

Tableau 6 - Évaluation du jeu de données : *Ballena_Jorobada_Estrecho* (partie)

Nom	Description	Type	Qualité	Utilité	Commentaires
<i>latitude</i>	Latitude de l'observation	Numérique	2	3	UTM (Universal Transverse Mercator)
<i>specie</i>	Espèce de faune marine observé	Caractères	2	3	Nom latin
<i>number_adult</i>	Quantité des animaux observés	Numérique	3	3	
<i>number_juvenil</i>	Quantité des animaux jeunes observés	Aucun	0	3	Information manquante

Modifié de (Crozat, 2016b)

Tableau 7 - Négociation « Utilité/Qualité » du jeu de données : *Ballena_Jorobada_Estrecho* (partie)

Nom	Conserver ou abandonner	Traitement requis	Champs cible si conservé
<i>latitude</i>	Données conservées	Changer à WGS 84 / Pseudo-Mercator. EPSG:3857	<i>latitude</i>
<i>specie</i>	Données conservées	Ajouter le nom en anglais	<i>animal_specie_en</i>
		Ajouter le nom en espagnol	<i>animal_specie_es</i>
		Ajouter le nom en français	<i>animal_specie_fr</i>
		Ajouter le nom en latin	<i>animal_specie_la</i>
<i>number_adult</i>	Données conservées	Le traitement n'est pas nécessaire.	<i>number_of_adults</i>
<i>number_juvenil</i>	Données abandonnées	Éliminer si NULL	<i>number_juvenil</i>

Annexe 3 Modèle Logique

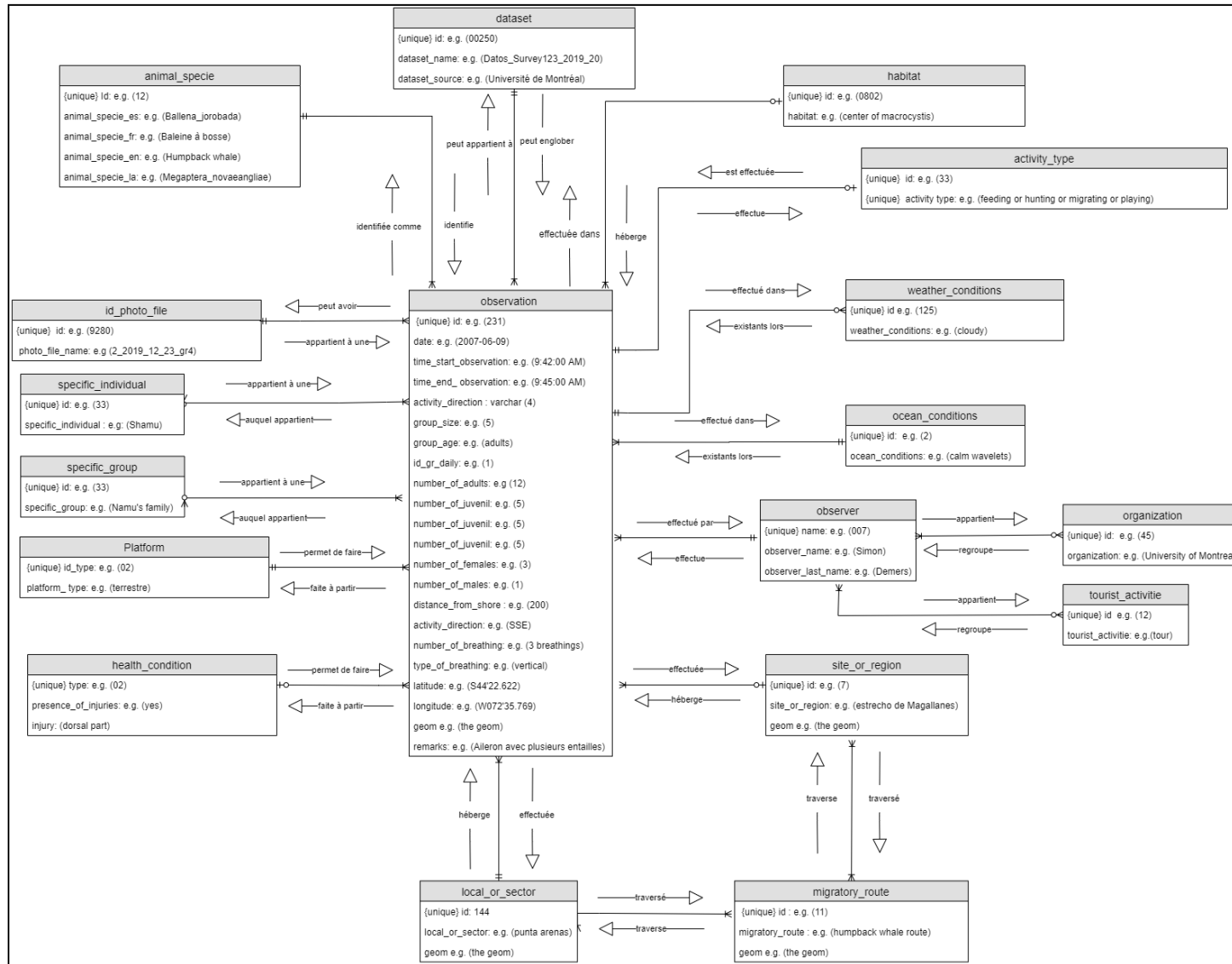


Figure 27 - Modèle conceptuel_logique UML

Annexe 4 Modèle physique

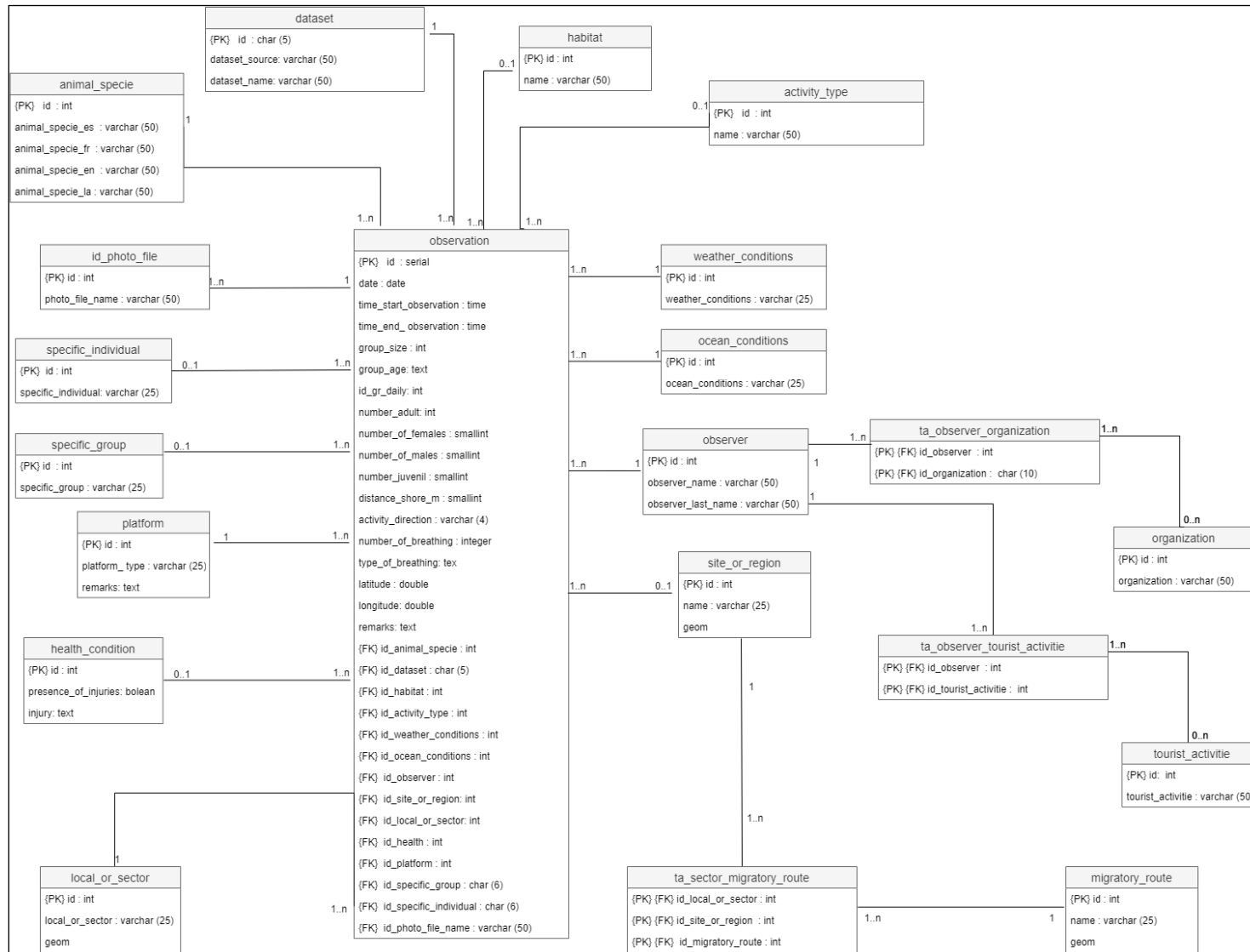


Figure 28 Modèle physique UML

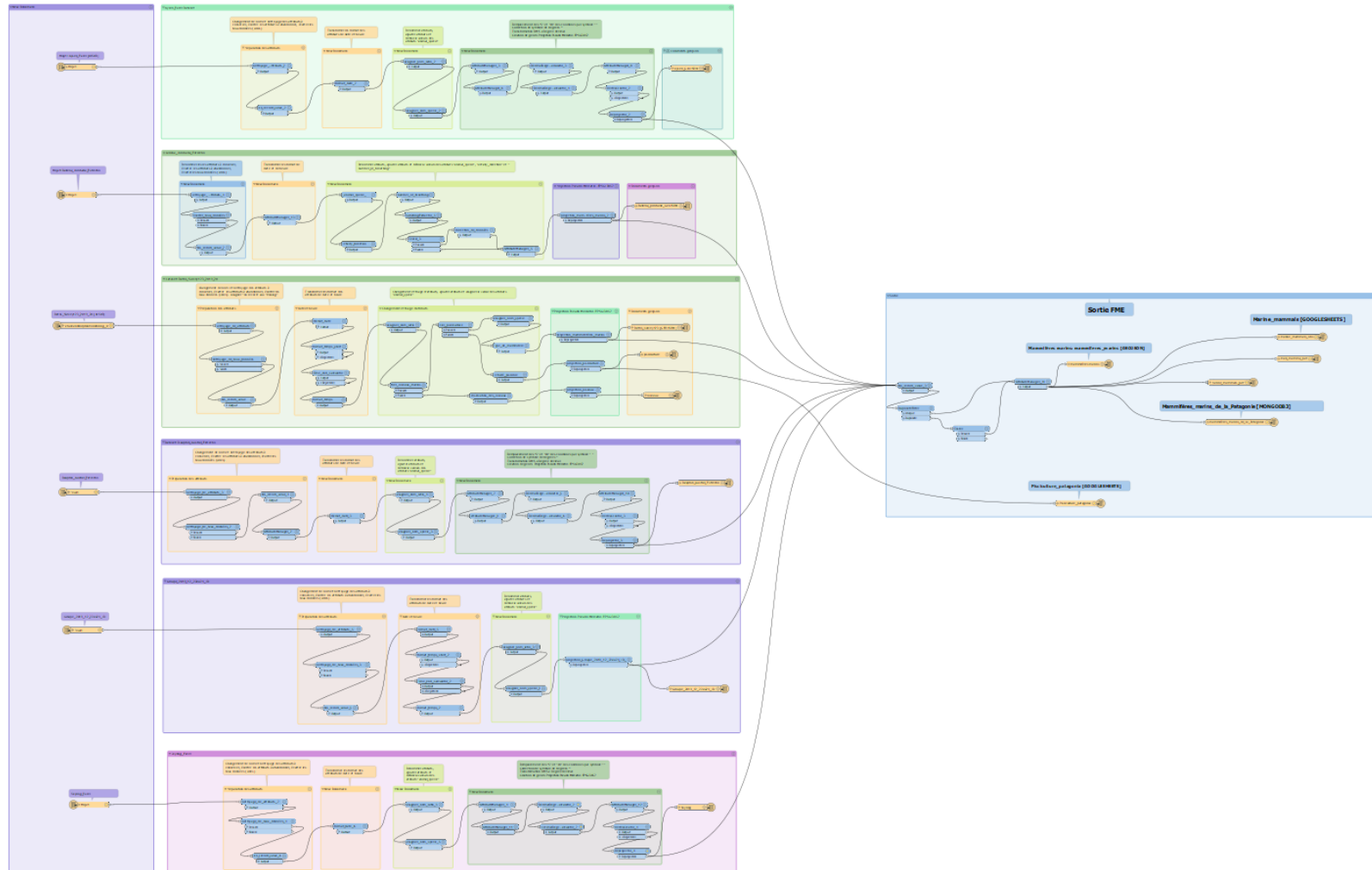
Annexe 5 : Tableau de synthèse des transformations requises (exemple)

Tableau 8 – Matrice de transformation de données (partie)

Champs cible	Dataset source	Attribut source	Règle de transformation
<i>activity_direction</i>	Ballena_Sei_Estrecho	<i>direction</i>	Nettoyer
	Groupe_2019_12_22au24_ID	<i>In what direction is the species moving?</i>	
	Ballena_Jorobada_Estrecho	<i>note</i>	
	Dauphin_Austral_Estrecho	<i>note</i>	
	Mamiferos marinos P Esperanza Enero 2020	<i>Actividad</i>	
<i>habitat</i>	Ballena_Jorobada_Estrecho	<i>habitat</i>	Le traitement n'est pas nécessaire
	Dauphin_Austral_Estrecho	<i>habitat</i>	
	Mamiferos marinos P Esperanza Enero 2020	<i>Habitat</i>	
	Skyring_Excel	<i>habitat</i>	
	Groupe_2019_12_22au24_ID	<i>Injuries</i>	
<i>latitude</i>	Aysen_Excel	<i>latitude</i>	Changer à WGS 84 / Pseudo-Mercator. EPSG:3857
	Ballena_Jorobada_Estrecho	<i>latitude</i>	
	Datos_Survey123_2019_20	<i>y</i>	
	Dauphin_Austral_Estrecho	<i>latitude</i>	
	Skyring_Excel	<i>latitude</i>	Chercher lecture et changer à WGS 84 / Pseudo-Mercator. EPSG:3857
	Ballena_Sei_Estrecho	<i>latitude</i>	
	Mamiferos marinos P Esperanza Enero 2020	<i>Coordenada</i>	
<i>time_end_observation</i>	Aysen_Excel	<i>heure_fin</i>	Nettoyer
	Ballena_Jorobada_Estrecho	<i>hour_end</i>	
	Ballena_Sei_Estrecho	<i>hour_end</i>	
	Dauphin_Austral_Estrecho	<i>hour_end</i>	
	Skyring_Excel	<i>heure_fin</i>	
	Datos_Survey123_2019_20	<i>Duration of the observation</i>	Calculer en additionnant au Hour of sighting
	Groupe_2019_12_22au24_ID	<i>Duration of the observation</i>	

Modifié de (Chafki & Desrosiers, 2011)

Annexe 6 Processus ETL de données.



Annexe 7 Comme utiliser des photos Google Drive dans des fenêtres contextuelles ArcGIS Online

« Il est possible d'inclure des images dans des fenêtres contextuelles ArcGIS Online en saisissant leur URL directe. Ces images peuvent être hébergées sur ArcGIS Online ou un site web externe, comme Google Drive et Flickr. L'URL à utiliser dans la configuration des fenêtres contextuelles doit être l'URL directe de l'image. Cette URL n'est cependant pas fournie directement par Google Drive. » (*How To: Use Google Drive photos in ArcGIS Online pop-ups, s. d.*)

Procédure pour obtenir l'URL directe d'une image dans Google Drive :

- Sélectionner l'image dans Google Drive et configurer l'partage Public
- Copier l'URL figurant dans le lien de partage.
- Cette URL contient l'ID de l'image. Cette ID est soulignée en jaune dan l'exemple:
- <https://drive.google.com/file/d/0B14bzGsfJ06JMDF5Y2VsNU9FRVk/view?usp=sharing>
- Remplacez l'ID de l'image dans l'URL suivant:
- https://drive.google.com/uc?export=view&id=<image_id>

Ajouter le résultat a la configuration d'image des fenêtres contextuelles

Annexe 8 : Métadonnée

Nom	Description	Type
<i>activity_direction</i>	Indique la direction de l'activité qui prévaut lors de l'observation	Caractères
<i>activity_type</i>	Activité qui prévaut dans le groupe observé	Caractères
<i>animal_specie_en*</i>	Nom en anglais de l'espèce de faune marine observé	Caractères
<i>animal_specie_es*</i>	Nom en espagnol de l'espèce de faune marine observé	Caractères
<i>animal_specie_fr*</i>	Nom en français de l'espèce de faune marine observé	Caractères
<i>animal_specie_la*</i>	Nom en latin (nom binomial) de l'espèce de faune marine observé	Caractères
<i>dataset_name*</i>	Nom du jeu de données d'où provient l'information	Caractères
<i>dataset_source*</i>	Origine du jeu de données d'où provient l'information	Caractères
<i>date</i>	Date de l'observation	Date
<i>distance_from_shore</i>	Distance à la côte exprimée en mètres	Numérique
<i>group_age</i>	L'âge qui prévaut dans le groupe observé	Caractères
<i>group_size</i>	Quantité des animaux observés	Numérique
<i>habitat</i>	Localisation par rapport à une forêt de <i>Macrocystis</i> (kelp) ou par rapport à la côte	Caractères
<i>id_gr_daily</i>	Identification assignée à un groupe pendant la journée	Numérique
<i>injury</i>	Type de blessures corporelles	Caractères
<i>latitude</i>	Latitude de l'observation système WGS84	Numérique
<i>local_or_sector</i>	Nom du lieu ou repère kilométrique qui indique la distance du point d'observation au point d'origine de la route	Caractères
<i>longitude</i>	Longitude de l'observation système WGS84	Numérique
<i>migratory_route</i>	Indique si l'observation a été effectuée dans une route migratoire connue.	Caractères
<i>number_juvenil</i>	Quantité des animaux jeunes observés	Numérique
<i>number_of_adults</i>	Quantité des animaux adultes observés	Numérique

<i>Number_of_breathing*</i>	Indique la quantité des souffles observés	Numérique
<i>number_of_females*</i>	Quantité des femelles observés	Numérique
<i>number_of_males*</i>	Quantité des mâles observés	Numérique
<i>observer_last_name*</i>	Nom de famille de l'observateur	Caractères
<i>observer_name*</i>	Nom de l'observateur	Caractères
<i>ocean_conditions</i>	État de la mer d'après l'échelle de Beaufort	Numérique
<i>organization*</i>	Organisation ou groupe d'organisations pour lesquelles travaille l'observateur	Caractères
<i>photo_file_name</i>	Identification unique d'une archive avec photographies de l'animal observé	Caractères
<i>platform_type</i>	Indique si la plate-forme d'observation était terrestre ou marine	Caractères
<i>presence_of_injures</i>	Présence de blessures corporelles	Booléen
<i>remarks</i>	Description des caractéristiques distinctives des animaux observés	Caractères
<i>site_or_region</i>	Emplacement	Caractères
<i>specific_group*</i>	Nom d'un groupe identifié sans équivoque	Numérique
<i>specific_individual*</i>	Nom d'un animal identifié sans équivoque.	Numérique
<i>time_end_observation</i>	Durée de l'observation	Heure
<i>time_start_observation</i>	Heure du début de l'observation	Heure
<i>tourist_activitie</i>	Participation de l'observateur dans un circuit touristique	Booléen
<i>type_of_breathing*</i>	Indique le type des souffles observés	Caractères
<i>weather_conditions</i>	Conditions météorologiques	Caractères

* Donnée découlant de l'analyse de besoins.