



Cours 5

Version textuelle

Gestion de données dans un inventaire forestier national

La version interactive de cette cour est disponible gratuitement à l'adresse suivante :

<https://elearning.fao.org/?lang=fr>



Certains droits réservés. Ce(tte) œuvre est mise à disposition selon les termes de la licence CC BY-NC-SA 3.0 IGO (<https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.fr>).

Dans cette cours

Leçon 1: Concepts de base de la gestion de données	5
Introduction de la leçon	5
Terminologie de base pour la gestion de base de données	5
Besoins des systèmes d'information forestière	14
Résumé.....	16
Leçon 2: Collecte et gestion de données de terrain	17
Introduction de la leçon	17
Outils de données des IFN: une vue d'ensemble	17
Collecte de données pour les IFN	19
Gestion de données pour les IFN	21
Validation de données.....	24
Résumé.....	28
Leçon 3: Publication en ligne et archivage de données	30
Introduction de la leçon	30
Tableaux de bord et portails de données.....	30
Traitement analytique en ligne (OLAP)	31
Archivage de données	33
Résumé.....	36

Ce cours propose une vue d'ensemble de la collecte d'information et la gestion de données pour les inventaires forestiers nationaux (IFN).

À qui ce cours s'adresse-t-il?

Ce cours s'adresse principalement aux personnes impliquées dans les IFN, particulièrement celles intéressées par les méthodes et les outils de gestion, traitement et partage de données des IFN. Ce cours vise particulièrement:

1. Les décideurs qui planifient des investissements pour la mise en œuvre des IFN.
2. Les gestionnaires de données et les chefs d'équipe des IFN.
3. Les techniciens de données responsables de l'analyse de données.
4. Les étudiants, en tant que matériel programmatique dans les écoles forestières.
5. Les jeunes et les nouvelles générations d'agents forestiers.

Structure du cours

Ce cours est composé de trois leçons.

Leçon 1: Concepts de base de la gestion de données

Cette leçon souligne le rôle technologies de l'information et de la communication (TIC) dans la gestion des données forestières. Cette leçon visant aussi à définir une terminologie de la foresterie liée aux TIC, elle vous servira aussi d'introduction aux leçons plus complexes dans ce cours.

Leçon 2: Collecte et gestion de données de terrain


Cette leçon aborde la gestion de données de terrain et les processus de préparation des données pour le stockage à long terme et les analyses.

Leçon 3: Publication en ligne et archivage de données

Cette leçon explore la technologie OLAP (traitement analytique en ligne) et décrit l'importance des métadonnées et des microdonnées dans la collecte et le stockage de l'information sur les forêts.

À propos de la série

Ce cours conclut une série de huit cours individualisés couvrant divers aspects d'un IFN. Voici un aperçu de la série complète.

Cours	Apprentissages
Cours 1: Pourquoi un inventaire forestier national (IFN)?	Objectifs et but d'un IFN, et comment les IFN informent la conception de politiques et la prise de décisions dans le secteur forestier
Cours 2: Préparation d'un inventaire forestier national	La planification et le travail nécessaire pour mettre en place un IFN efficace ou un système national de suivi des forêts (SNSF).
Cours 3: Introduction à l'échantillonnage	Aspects généraux de l'échantillonnage dans les inventaires forestiers.
Cours 4: Introduction au travail de terrain	Considérations pour le travail de terrain, les variables au niveau parcellaire et les mesures au niveau de l'arbre.
 Cours 5: Gestion de données dans un inventaire forestier national	(Vous suivez actuellement ce cours).
Cours 6: Assurance qualité et contrôle qualité dans un inventaire forestier national	Procédures d'AQ et de CQ dans la collecte et la gestion de données d'un inventaire forestier.
Cours 7: Éléments de l'analyse de données	Approches/calculs typiques dans les analyses de données et questions connexes.
Cours 8: Résultats de l'inventaire forestier national: notification et diffusion	Publication des résultats de l'IFN et importance de la notification dans le contexte des actions REDD+.

Leçon 1: Concepts de base de la gestion de données

Introduction de la leçon

Dans cette leçon, nous comprendrons le rôle technologies de l'information et de la communication (TIC) dans la gestion des données forestières.

Pour cela, nous aborderons les termes, les techniques et les méthodes liés aux TIC qui sont utilisés pour gérer les données des IFN. Nous nous intéresserons aussi aux besoins d'un système d'information forestière.

Objectifs

A la fin de cette leçon, vous serez en mesure de:

1. Présenter la terminologie et les concepts de base pour la gestion de données dans une technologie de base de données.
2. Expliquer les principaux besoins d'un système d'information sur le suivi national des forêts.

Terminologie de base pour la gestion de base de données

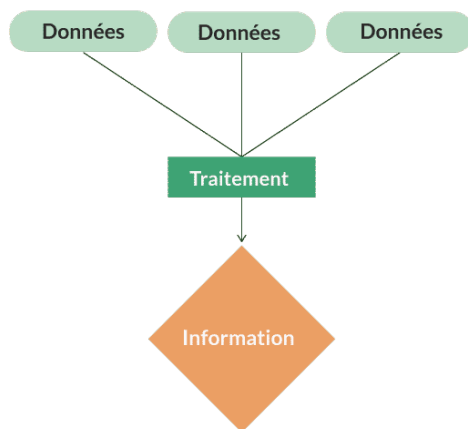
Comprendre la terminologie clé associée à la gestion de base de données vous aidera à saisir des concepts plus complexes à mesure que nous progresserons dans ce cours. Commençons par regarder la terminologie des TIC qui est pertinente pour la collecte et l'organisation de données des IFN.

Technologies de l'information et de la communication (TIC)

Les technologies de l'information et de la communication (TIC) sont une **expression générique** qui inclut la conception, la mise au point, la mise en œuvre, l'appui et la gestion de systèmes d'information informatisés. Essentiellement, les TIC utilisent les ordinateurs et les logiciels pour convertir, stocker, protéger, traiter, transmettre et extraire de l'information.

Données

Un ensemble de faits, comme des nombres, des mots, des mesures, des images, des observations ou juste des descriptions de choses. **Information** se réfère au résultat significatif obtenu après le traitement des données. Nos chiffres ne feront aucune différence si l'on ne les traduit pas en récits significatifs.



Intégrité des données

Le terme intégrité des données se réfère à l'exactitude et la cohérence des données. Lorsque l'on crée des bases de données, l'attention doit être portée sur l'intégrité des données et la manière de la maintenir. Une bonne base de données forcera l'intégrité des données autant que possible.

Par exemple, un utilisateur peut accidentellement essayer de saisir un code de type de végétation dans un champ de date. Si le système force l'intégrité des données, il empêchera l'utilisateur de commettre cette erreur, à travers un avertissement ou un message d'erreur lorsque les mauvaises données sont saisies dans un champ spécifique. C'est l'un des avantages de l'utilisation de formulaires de terrain en ligne pendant la collecte de données de terrain. Bien entendu, toutes les entrées erronées ne peuvent pas être identifiées par cette vérification formelle.

Gestion de données

Un **processus administratif** qui inclut l'acquisition, la validation, le stockage, la protection et le traitement des données pour en assurer l'accessibilité, la fiabilité et l'opportunité aux utilisateurs. Un principe fondamental de la gestion de données forestières consiste à stocker toutes les données telles qu'elles ont été relevées dans leur forme originale. Cela offre une flexibilité dans la manière dont les données peuvent être traitées et assure que tous les calculs sont reproduits à partir des données originales. Les données originales servent de «données de base» et peuvent être importantes pour contre-vérifier les erreurs détectées pendant la validation et/ou l'analyse de données.

Base de données et système de gestion de base de données (SGBD)

Une base de données est un **fonds de données placé dans un système informatique sous une forme organisée** pour un accès et une gestion plus faciles. Les bases de données sont simplement des conteneurs pour données. Dans les IFN, les bases de données sont utilisées pour gérer et archiver les

données d'inventaire de terrain, les photographies de terrain, les cartes et les données de télédétection, et les documents connexes (comme les manuels de terrain, directives, rapports d'inventaire).

Les systèmes de gestion de base de données, couramment désignés comme SGBD, sont **des logiciels qui nous permettent de réaliser diverses opérations sur les bases de données**. Les SGBD permettent aux utilisateurs d'accéder aux bases de données, et de manipuler, rapporter et représenter les données. Ils peuvent aussi aider à contrôler l'accès à la base de données. Parmi les exemples connus de logiciels de base de données ou SGBD, on trouve MySQL, Microsoft Access, Microsoft SQL Server, PostgreSQL, FileMaker Pro, Oracle Database, et InterBase.

Langage d'accès à la base de données

Un langage d'accès à la base de données est utilisé pour **accéder, saisir, actualiser ou extraire des données** du SGBD. C'est un outil informatisé central pour organiser les données, créer des bases de données et contrôler les données à l'aide de langages de requête. Plusieurs langages peuvent être utilisés dans ce but, y compris SQL (Structured Query Language).



Ressources vidéo

Les vidéos sont en langue originale (anglais).

Qu'est-ce qu'une base de données et SQL

Ce tutoriel de base de données explique le concept de SGBD (système de gestion de base de données). Pour aider les débutants, il cite des exemples de système de gestion de base de données de la vie courante. Il explique les types de SGBD pour les débutants. Il explique comment SQL fonctionne. Ce cours en vidéo est une introduction complète aux bases de données. Cliquez sur les marques temporelles pour voir les différentes sections!

Apprendre les bases de SQL en 10 minutes

[\[https://www.youtube.com/watch?v=bEtnYWuo2Bw&t=1s\]](https://www.youtube.com/watch?v=bEtnYWuo2Bw&t=1s)

Schéma

La **structure de la base de données** est appelée schéma, et celui-ci définit quel type de données est

stocké. On peut penser à une structure squelettique qui représente la vision logique de la base de données entière. Il définit la manière dont les données sont organisées et dont les composants sont reliés. Il formule toutes les contraintes qui doivent être appliquées aux données.

Les facteurs clés à considérer concernant les données lorsque l'on conçoit une base de données incluent les éléments suivants:

1. **Types de données**, qui sont utilisés dans chaque champ et classifient les valeurs des données qui ont des propriétés communes.
2. **Validation**, qui sont les règles d'acceptation des données.
3. **Champ clé, ou clé primaire** (nous en saurons plus sur les champs plus tard dans cette leçon).

Les développeurs planifient un schéma de base de données à l'avance, de sorte qu'ils savent quels composants sont nécessaires et comment ils vont les connecter les uns aux autres. Cela est aussi important dans le contexte des IFN: concevoir un schéma de base de données intelligent facilite l'analyse.

Entité et attribut

Une **entité** de données est un **objet dans un répertoire de données**. Les entités de données sont les objets du modèle de données comme «arbre» ou «sol». Les entités ne représentent aucune donnée en soi mais sont les conteneurs des attributs et des relations entre les objets.

L'**attribut** des données est une **unité d'information à l'intérieur de l'entité de données**, autrement dit, c'est un descripteur de valeur singulière pour un point de données ou un objet de données. Ce sont en quelque sorte les propriétés des entités de données: par exemple, l'entité «arbre» peut contenir les attributs suivants: espèces, taille et âge.

Modèles de données

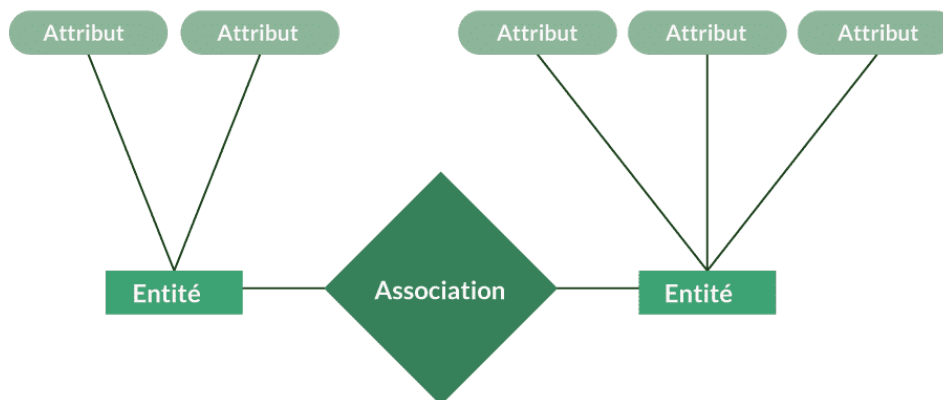
Les modèles de données définissent la **structure logique d'une base de données**. Ils définissent comment les données sont connectées entre elles et comment elles sont traitées et stockées dans le SGBD.

Un modèle de données aide à percevoir, organiser et décrire les données dans un schéma conceptuel qui inclut à la fois les données et les opérations pour manipuler l'ensemble de données. Il est important de voir maintenant des deux principaux modèles de données: le **modèle entité-association** et le **modèle de données relationnel**.

Modèle entité-association

Comme son nom l'indique, le modèle entité-association (ER, en anglais) **s'appuie sur la notion d'entités du monde réel et d'associations entre elles**. En formulant des scénarios du monde réel dans le modèle de base de données, le modèle ER crée un lot d'entités, un lot d'associations, des attributs généraux et des contraintes.

Le modèle ER est plutôt utilisé pour la schématisation conceptuelle de la base de données. La schématisation conceptuelle est la première étape du processus de conception de la base de données. L'objectif à cette étape est de concevoir une base de données qui est indépendante du logiciel de base de données et des détails physiques. Le résultat de ce processus est un modèle de données conceptuel qui décrit les principales entités de données, les attributs, les associations et les contraintes du domaine d'un problème donnée. Voyons une représentation visuelle du modèle entité-association. Diagramme illustrant le modèle entité-association.



Modèle de données relationnel

Les bases de données relationnelles, qui sont **le modèle de données le plus largement utilisé dans les SGBD**, sont devenues la principale technologie de gestion de bases de données, applicables sur les téléphones intelligents et ordinateurs personnels jusqu'aux ordinateurs centraux.

Le modèle de données relationnel **représente les données sous la forme de tables**. Une relation est une table bidimensionnelle de données nommée. Chaque relation (ou table) consiste en un ensemble de colonnes nommées et un nombre arbitraire de lignes. Un attribut est une colonne nommée d'une relation. Chaque ligne d'une relation correspond à l'enregistrement qui contient les valeurs des données

(attributs) pour une entité spécifique. Voyons une représentation visuelle d’une base de données relationnelle.

id	species_name	dbh	tree_height_calc	tree_basal_area	tree_volume_stem	tree_volume_bole	tree_biomass_ag	tree_biomass_bg
1001	Trichila prieuriana	28	12	0.061575216	0.038792386	0.024630086	0.041559922	0.011221179
1002	Acacia nilotica	40	18	0.125663706	0.118752202	0.100530965	0.123853538	0.033440455
1003	Acacia nigrescens	22	14	0.038013271	0.027939754	0.025145779	0.030169727	0.008145826
1004	Trichila emetica	24	14	0.045238934	0.033250617	0.029925555	0.035754825	0.009653803
1005	Toona ciliata	28	10	0.061575216	0.032326988	0.024630086	0.034785146	0.009391989
1006	Trichila prieuriana	32	10	0.080424772	0.042223005	0.032169909	0.045143384	0.012188714
1007	Trichila dregeana	32	14	0.080424772	0.059112207	0.048254863	0.062692425	0.016926955
1008	Acacia nilotica	28	14	0.061575216	0.045257784	0.024630086	0.048307525	0.013043032
1009	Trichila emetica	68	18	0.363168111	0.343193865	0.217900866	0.348935144	0.094212489
1010	Trichila emetica	20	16	0.031415927	0.026389378	0.025132741	0.02853468	0.007704364

■ Attributs ■ Colonne ■ Enregistrement (ligne)

Un enregistrement est une ligne complète d’éléments de données dans une table. Voyons l’exemple présenté ici. L’arbre numéro 1005 dans la table a un enregistrement – surligné en orange dans la table ci-dessus – où les données d’entrée sont toutes enregistrées en une table. Cela diffère d’un tuple (qui apparaît typiquement à travers une requête), qui est un lot d’enregistrements qui incorpore des données à partir de tables multiples. Les données dans un tuple ont une valeur de référence particulière en commun, par ex. tous les enregistrements de variables (volume, biomasse, valeur de tige, etc.), dans différentes tables, appartiennent à l’arbre numéro 1 005. Ainsi, les données, de plusieurs tables, qui forment un tuple, sont toutes associées à un arbre.



Le saviez-vous?

Quelles sont les principales caractéristiques des modèles de base de données relationnels?

Les caractéristiques majeures de ce sont modèle sont:

- Les données sont stockées dans des tables (appelées relations).
- Les relations peuvent être «normalisées».
- Chaque relation dans une base de données doit avoir un nom distinct ou unique.
- Une relation ne peut pas avoir deux attributs avec le même nom. Chaque attribut doit avoir un nom distinct.
- Au sein d'une relation, chaque ligne doit être unique.
- Chaque colonne dans une relation contient des valeurs du même domaine. Un domaine est un ensemble unique de valeurs permises pour un attribut dans une table. Par exemple, un domaine «mois de l'année» peut accepter janvier, février, jusqu'à décembre comme valeurs possibles, et un domaine de nombres entiers peut accepter des nombres entiers qui sont négatifs, positifs et zéro.



Ressources vidéo

Regardons une vidéo qui explique la normalisation de base de données pour les débutants (en anglais).

[Concept de base de la normalisation de base de données – explication simple pour débutants](#)

[Schème de base de données : diagramme entité-association](#)

[Convertir les diagrammes ER en schémas | SQL | Tutoriel 23](#)

Types de données

Typifier les données est une manière de **classifier les valeurs de données qui ont des propriétés communes**.

Lorsque l'on crée une base de données, on doit établir des types de données pour chaque champ. Par exemple, dans une base de données d'inventaire forestier, on peut avoir besoin de champs de texte pour «nom de l'enquêteur», mais de chiffres pour «âge de l'arbre». Les champs sont généralement restreints à un type de données spécifique.

Différentes sortes de valeurs de données nécessitent aussi différents volumes de mémoire pour être stockées et permettre différentes opérations avec elles. Par exemple, si le type de données est une chaîne de caractères, on peut saisir des nombres, mais on ne pourra pas faire de calculs car les chiffres seront compris par le logiciel comme des caractères. Les types de données les plus utilisés sont:

- **Nombres entiers** comme 5; 27; 65575.
- **Nombres à virgule flottante** (avec des virgules décimales, parfois appelés nombres réels, ou floteurs) Par ex. 5,2; 37,4; 196,247.
- **Caractères** comme c; F; 3; \$; #.
- **Chaînes de caractères** comme abc; deu496; 3erf08!@.
- **Valeurs booléennes**, notamment Vrai/Faux, ou Oui/Non.

Une base de données d'inventaire forestier contient typiquement tous ces types de données. En outre, elle peut aussi contenir des fichiers (photos, vidéos, fichiers audio, documents numériques), des dates, des temps et des types complexes (des attributs combinés).

Les exemples de types de données complexes dans une base de données d'inventaire forestier entrent dans deux catégories principales: les **données de coordonnées** et les **données taxonomiques**.

Les données de coordonnées (points) peut consister en un système de coordonnées de référence (SCR) horizontal, coordonné X et coordonnée Y. Un SCR vertical contient la troisième dimension, soit Z, la hauteur ou l'altitude. Dans les SCR composés, les données des systèmes horizontal et verticale peuvent être rassemblées.

Note: Système de coordonnées de référence (SCR) et système de référence spatiale (SRS) sont des synonymes et sont couramment interchangeables. Il est extrêmement important non seulement de spécifier les coordonnées, mais aussi de donner l'information du système de référence qui a été utilisé. Sans

cette information, les valeurs des coordonnées sont quasiment insignifiantes.

Les données taxonomiques peuvent contenir des données sur la famille, le genre, l'espèce, la sous-espèce, la variété. Les données taxonomiques peuvent aussi contenir une citation de l'auteur qui consiste à référencer la personne (ou l'équipe) qui a présenté pour la première fois le nom scientifique d'un taxon.



Astuces rapides!

En outre, une base de données d'IFN peut être étendue en une **base de données spatiale** qui stocke des données d'objets cartographiés (sous forme de points, lignes et polygones). Ces bases de données permettent l'utilisation de requêtes spatiales et d'analyse spatiale pour traiter les données des IFN.

Identifiant et clé primaire

Chaque entité nécessite une étiquette qui l'identifie de manière unique. Dans une base de données relationnelle, cela s'appelle une **clé primaire**. Une clé primaire est donc **une colonne de table de base de données relationnelle spéciale** (ou une combinaison de colonnes) **désignée pour identifier de manière unique chaque enregistrement de la table**. Sans les concepts étroitement liés de clé primaire et de clé étrangère, les bases de données relationnelles ne fonctionneraient pas. C'est pourquoi on doit définir une clé primaire pour chaque entité lorsque l'on construit le schéma de base de données. Rappelons qu'il peut y avoir plus d'une clé primaire pour une entité.

Base de données à fichier plat et entrepôt de données

Une base de données à fichier plat **stocke les données dans un fichier texte en clair**. Chaque ligne du fichier texte contient un enregistrement, avec des champs séparés par des délimiteurs, comme des virgules ou des tabulations.

Comme elle utilise une structure simple, une base de données à fichier plat ne peut pas contenir de tables multiples comme une base de données relationnelle. Les bases de données à fichier plat ont été développées et mises en œuvre au début des années 1970 par IBM. Les projets d'entrepôt de données utilisent des fichiers plats pour importer des données.

Les entrepôts de données **collectent le contenu de plusieurs bases de données opérationnelles**, y compris les bases de données personnelles, de groupes de travail, de services et de planification des ressources de l'entreprise (PRE). Ce sont des systèmes utilisés pour rapporter et analyser des données et sont considérés comme le composant central de l'intelligence économique.

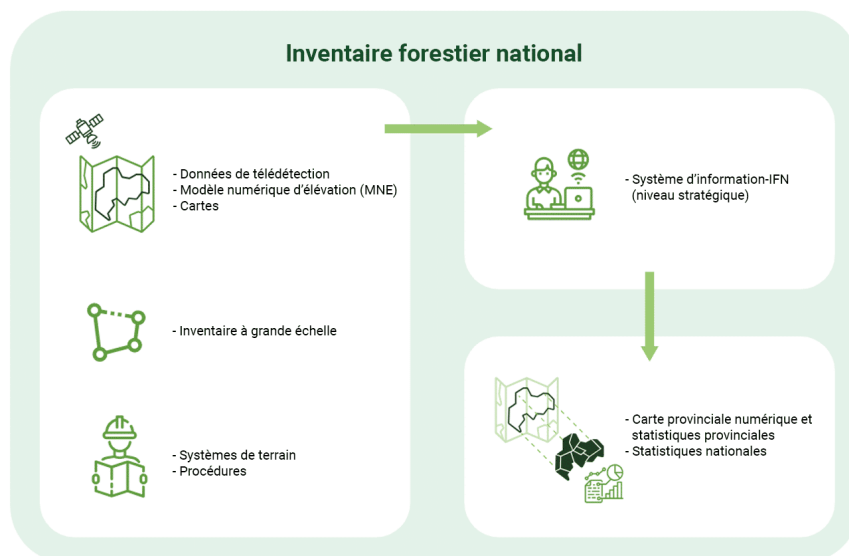
Exemple national: Institut de recherche forestière de la Corée

Cet exemple montre comment concevoir un entrepôt de données pour mettre en place une infrastructure de partage pour les données spatiales et d'enquête de terrain d'un système d'information forestière.

Comme on le voit dans cet exemple, l'architecture d'un entrepôt de données peut être complexe. Actuellement, les systèmes de gestion de données des IFN importants dans des pays avec un SNSF institutionnalisé (par ex. NAFIDAS pour l'IFN de la Suisse et TaxWebb en Suède) sont comparables à une architecture d'entrepôt de données. Aucun de ces systèmes n'est meilleur ou pire qu'un autre, et la décision du système adapté à la situation locale dépend entièrement des ressources disponibles et des besoins.

Besoins des systèmes d'information forestière

Les systèmes d'information forestière fournissent une plateforme aux décideurs avec une information de haute qualité et exhaustive qui est passée par des processus de contrôle qualité. Les données des IFN et d'autres inventaires forestiers peuvent être considérées comme les éléments essentiels dans ce concept.



Un système moderne de gestion de l'information des ressources forestières nécessite les éléments suivants:

1. **Adaptabilité aux différents environnements matériels et logiciels**, différentes conditions, et différentes échelles géographiques d'inventaire.
2. **Capacité d'utilisation des données existantes** dans toutes les phases de l'inventaire.
3. **Variabilité d'échelle** pour être transférable à un système opérationnel plus grand où il peut être tiré parti pleinement des performances du système opérationnel plus grand.
4. **Flexibilité pour s'adapter** aux divers modèles de données.
5. **Adaptabilité pour héberger les données d'IFN répétés** avec possiblement des listes de variable changeantes et d'autres adaptations.

En outre, le système doit être sûr, facile d'utilisation, accessible en ligne et offrir des outils de publication flexibles. La documentation des systèmes et de leurs procédures doit être transparente et permettre d'adapter de futures améliorations.

Compétences requises pour la gestion de données

La gestion et le traitement de données forestières nécessitent du personnel avec une connaissance solide des inventaires forestiers, de l'information concernant les caractéristiques des forêts (et de la nature) et une connaissance des outils et des méthodes nécessaires pour réaliser toutes les tâches liées à ces domaines. Fondamentalement, une expertise est nécessaire dans les trois domaines suivants:

Connaissance en statistiques	Une connaissance statistique, en particulier de l'échantillonnage statistique et de la modélisation statistique, est nécessaire pour définir les procédures et les formules mathématiques dans les étapes de planification de l'analyse de données. L'expert doit comprendre les objectifs transversaux de l'IFN, être impliqué dans les étapes de planification initiale et avoir une compréhension claire des variables utilisées.
Conception de base de données	Afin de concevoir, construire et maintenir la base de données, le concepteur de la base de données doit non seulement programmer et mettre en œuvre le système au départ, mais aussi maintenir et améliorer le système de base de données – cela requiert des qualifications informatiques spécifiques à l'emploi.

Administration de base de données

L'administration de base de données nécessite de recevoir des données du terrain, superviser leur entrée dans le système, réaliser la vérification d'erreur et correction, et traiter les requêtes *ad hoc* des utilisateurs.

L'administrateur de base de données doit avoir une expertise informatique et en base de données, et il est aussi important qu'il ait participé au travail de terrain et soit familiarisé avec les procédures utilisées, ainsi qu'avec les objectifs transversaux de l'inventaire.

Résumé

Avant de conclure, voici les principaux points d'apprentissage de cette leçon:

- Un système de gestion de base de données (SGBD) est un ensemble de programmes qui permet aux utilisateurs d'accéder à des bases de données, de manipuler, rapporter et représenter les données. Le SGBD est la technologie dominante pour toute gestion de données.
- Le modèle entité-association est plutôt utilisé pour la schématisation conceptuelle d'une base de données. Les bases de données relationnelles sont devenues la technologie dominante pour la gestion de base de données.
- Lorsque l'on planifie une base de données, on doit concevoir un schéma qui contient des entités, des attributs, des relations, des champs clés et des règles de validation.
- La gestion et le traitement de données forestières nécessitent des personnes avec une bonne connaissance des inventaires forestiers, des caractéristiques de l'information sur les forêts, et une connaissance des outils et des méthodes. Les rôles associés avec les compétences nécessaires sont entre autres: biométricien ou statisticien forestier, concepteur et technicien de maintenance de base de données, et analyste de données/administrateur de base de données.

Leçon 2: Collecte et gestion de données de terrain

Introduction de la leçon

Cette leçon aborde la gestion de données de terrain et décrit les processus nécessaires de préparation des données pour le stockage à long terme et les analyses.

Elle explique aussi comment construire une base de données fonctionnelle pour héberger des données d'inventaire forestier. Parce que la qualité des données est toujours une priorité, cette leçon examine aussi les processus de validation de données.

Objectifs

A la fin de cette leçon, vous serez en mesure de:

1. Décrire la manière dont les données de terrain sont collectées, stockées et gérées.
2. Identifier les besoins spécifiques des systèmes TIC pour l'inventaire et le suivi des forêts.
3. Expliquer comment construire une base de données fonctionnelle pour héberger des données d'inventaire forestier.
4. Identifier des méthodes pour améliorer la qualité générale d'une base de données d'IFN.

Outils de données des IFN: une vue d'ensemble

Choisir les outils TIC pour la gestion de données d'inventaire forestier

D'un point de vue technologique, il est maintenant facile de se fournir en équipement informatique de bureau, car les ordinateurs de bureau et portable courants sont capables de gérer des données des IFN.

Au niveau de l'organisation, des serveurs locaux ou des services en nuage – hébergés par un fournisseur de services en nuage – sont nécessaires. Dans la plupart des cas, l'élan majeur pour acquérir un serveur est une hausse du nombre de postes de travail pour le personnel qui utilise le réseau de manière extensive.

Supposons que l'on prévoit une croissance de plus de sept travailleurs dans un bureau, chacun utilisant un ordinateur. Dans ce cas, il conviendra de considérer un serveur pour mieux gérer l'équipe de travail et les données qu'elle produit.

Les équipes de terrain d'inventaire forestier réalisent la plupart de leurs tâches dehors, exposant leur équipement aux conditions climatiques variables et à une manipulation éprouvante. C'est pourquoi **travailler avec des ordinateurs robustes est un avantage**. Cependant, les dispositifs robustes sont généralement plus chers que les dispositifs courants.

Utilisation des technologies à écran tactile

Il existe actuellement deux variétés de technologies à écran tactile – l'**écran tactile résistif**, qui utilise la **pression du corps humain comme entrée**, et l'**écran tactile capacitif**, qui utilise les **propriétés électriques du corps humain comme entrée**. Pour choisir entre ces deux technologies, il faut tenir compte de l'environnement dans lequel le dispositif sera utilisé.

Si le dispositif sera utilisé dans des conditions plus rudes et pluvieuses, qui peuvent nécessiter d'essayer l'écran très souvent, **une dalle tactile résistive fonctionnera mieux**, car on peut l'utiliser avec des gants et un stylet. Si, par ailleurs, **on utilisera le dispositif pour des applications plus sophistiquées**, **il faut choisir une dalle tactile capacitive**. Avec les dalles tactiles capacitatives, le toucher multipoint n'est pas un problème, on peut faire défiler facilement et avoir une sensibilité excellente.

Applications de collecte mobile de données

Il existe diverses applications pour la collecte mobile de données, y compris:

- ➔ [Open Data Kit](#) (en anglais)
- ➔ [Survey Solutions](#)
- ➔ [Open Foris Collect Mobile](#) (en anglais)
- ➔ [ArcGIS Survey123](#) (en anglais)

Certaines applications sont plus adaptées aux entretiens et aux enquêtes socioéconomiques (comme ODK et Survey Solutions), tandis que d'autres fonctionnent mieux pour l'arpentage (par ex. [ArcGIS AppStudio](#) (en anglais) et [QField](#) (en anglais)) – il faut donc évaluer les applications en fonction des besoins. Intéressons-nous de plus près à l'une de ces applications, Open Foris Collect (OF Collect), qui est un outil gratuit conçu spécifiquement pour manipuler les données d'inventaire forestier.

Présentation: Open Foris Collect et Collect Mobile

Open Foris (OF) Collect (en anglais) est une application mobile sous Android pour la collecte de données pour les enquêtes de terrain. Elle fonctionne comme solution de gestion de données de terrain et permet une personnalisation complète de la structure d'inventaire de données, des variables et des règles de validation de données. L'interface utilisateur (IU) de saisie de données d'OF est générée automatiquement.

Le système peut être utilisé dans un environnement autonome – sans besoin de connexion à Internet – ou installé dans un serveur où des utilisateurs multiples peuvent travailler en simultané avec accès aux données de la même enquête. Dans un projet d'envergure, comme les IFN, une installation sur serveur est nécessaire, le serveur fournissant l'espace de stockage pour héberger les données.

OF permet aussi une validation immédiate pour améliorer la qualité des données et s'intègre avec OF Collect pour la gestion et l'analyse de données.

Collecte de données pour les IFN

Les données de terrain peuvent être collectées et enregistrées à l'aide de dispositifs électroniques ou de formulaires de terrain sur papier. Les enregistreurs de données mobiles sont utilisés dans les IFN depuis la fin des années 1980, et actuellement, les tablettes ou les téléphones intelligents robustes sont couramment utilisés pour collecter les données de terrain. L'entrée de données – sur un ordinateur après l'évaluation de terrain – peut être réalisée manuellement (avec un clavier), par câble ou sans fil à partir d'un dispositif électronique.

Si la communication mobile est disponible sur le terrain, les données peuvent être directement transférées à la base de données centrale ou au nuage. En tout cas, **il est avantageux de réaliser la saisie de données le plus tôt possible (en termes de temps et d'espace) là où les données sont générées**. Des méthodes pour vérifier les données directement sur le terrain permettent aux équipes de terrain de corriger des erreurs et des incohérences immédiatement, ce qui élimine des sources d'erreur et améliore la qualité des données.

La communication par téléphone mobile augmente la sûreté des équipes de terrain car elles peuvent communiquer plus facilement dans des situations d'urgence, et permet en outre la saisie de données en ligne vers la base de données centrale. Néanmoins, la couverture des réseaux de téléphonie mobile est

mauvaise dans de nombreuses régions, spécialement dans les zones rurales faiblement peuplées où les parcelles de terrain peuvent se trouver, rendant la communication mobile directe sur le terrain techniquement impossible.

Dans ces situations, il est crucial que le chef d'équipe vérifie les données relevées et les sauvegarde à l'aide de dispositifs de stockage portables, de sorte qu'une fois que l'équipe atteint une zone avec connexion mobile, les données peuvent être téléchargées ou envoyées à la base de données centrale.



Astuces rapides!

Les inventaires forestiers requièrent généralement la collecte de données d'attributs contenant de l'information taxonomique et de position spatiale (point). Dans la base de données de l'application, il peut être avantageux de prévoir la hiérarchie des données et les relations (par ex. cluster-parcelle-sous-parcelle-entité) au lieu de relever une base de données à fichier plat consistant uniquement en un fichier/une table. SQLite est une base de données multiplateforme populaire sur les dispositifs mobiles.

Voyons maintenant le processus de flux de données depuis le terrain jusqu'à une base de données nettoyée prête pour l'analyse.

Données de formulaire de terrain

Pour le traitement des données de formulaire de terrain, il est recommandé de conserver des photocopies des formulaires de terrain dans un bureau local, si possible. Les formulaires de terrain sont transférés au siège, et les données sont saisies manuellement dans la base de données.

Données d'ANP/téléphone intelligent

Il existe deux options pour transporter les données vers la base de données centrale.

Option 1: Après une journée de travail, les données sont exportées depuis la tablette et copiées sur un ordinateur. Une «copie de sûreté» des données est stockée. Le fichier de données est alors envoyé au siège par courriel, via stockage en nuage, ou envoyé sur une clé USB.

Option 2: Les données peuvent aussi être envoyées au serveur directement depuis la tablette.

Données principales

Les données principales comprennent:

Données de saisie: Elles sont validées, la «liste d'erreur» est retirée, et toutes les mauvaises saisies sont vérifiées.

Données de validation: Toutes les erreurs sont corrigées, et les données sont nettoyées. Elles sont prêtes pour passer à l'étape suivante.

Données d'analyse: Les données nettoyées sont prêtes pour l'analyse.

Transfert et entrée de données

Le transfert de données depuis des tablettes ou des téléphones intelligents peut être organisé de diverses manières.

1. Dans les applications en nuage, les données sont automatiquement stockées sur le serveur en nuage.
2. Supposons que les données hors ligne sont d'abord collectées, et le dispositif dispose d'une connexion à internet par données mobiles. Dans ce cas, les données peuvent être stockées sur le serveur en nuage (Google Drive, OneDrive, DropBox etc.) ou transférées par courriel au siège.
3. Dans certains systèmes, les données relevées peuvent aussi être d'abord exportées sur le stockage de la tablette puis copiées et envoyées par connexion sans fil (par ex. par Bluetooth) à un ordinateur portable et transférées plus tard à la base de données centrale.
4. Dans certains systèmes, les données peuvent aussi être envoyées directement au serveur du réseau à travers un réseau de téléphonie mobile.
5. L'utilisation conventionnelle de formulaires de terrain sur papier nécessite que les agents de terrain livrent les formulaires au siège, où ils seront alors entrés manuellement dans la base de données centrale.

Gestion de données pour les IFN

Besoins spécifiques des systèmes TIC pour l'inventaire et le suivi des forêts

Voici quelques considérations typiques pour un système d'information pratique pour les IFN:

1. Les exigences en matière de données, analyse et notification des résultats sont déterminées par les besoins en information variables qui existent pour différentes situations. Les forêts sont

complexes, et les SGBD peuvent donc contenir des données sur les arbres, d'autre végétation, la faune, les caractéristiques du sol, les flux hydriques, etc. Par ailleurs, l'information spatiale doit aussi être relevée (comme les coordonnées des attributs).

2. Les résultats d'une analyse des exigences en matière de données peuvent généralement inclure les questions suivantes: objets, données, relations, processus, voies d'accès, intégrité des données, plan d'information, partage de données et sécurité des données.
3. Les réglementations gouvernementales, les normes et les engagements de notification des résultats doivent être pris en compte pour la construction de ces systèmes.
4. Ces systèmes contiennent en général des données sur les changements dans l'environnement et des ensembles de données multitemporels.

Bases de données dans les IFN

Les bases de données sont des outils efficaces d'analyse, de stockage et de gestion de grands ensembles de données des IFN, alors qu'une feuille de calcul est plus utile pour une analyse *ad hoc* que pour héberger et analyser les données des IFN. Cependant, construire une base de données pour un enquête requiert plus de temps que l'utilisation de feuilles de calcul.

Parce qu'il y a de nombreuses façons de stocker les données collectées dans les IFN, il n'existe pas de conclusion simple concernant une forme optimale. Les exigences d'efficacité et de qualité demande un système, durable et de faible maintenance. Typiquement, les données des IFN sont stockées dans une base de données centralisée qui sert de point d'entrée principal pour toutes les données collectées.

Une base de données d'IFN est hébergée par un SGBD pour créer et gérer plusieurs bases de données. Un SGBD rend possible la création, la lecture, l'actualisation et la suppression des données dans une base de données. Le SGBD sert d'interface sûre entre les bases de données et les utilisateurs finaux ou les programmes d'application, garantissant que les données sont organisées de manière cohérente et restent facilement accessibles. Il existe de nombreux logiciels de gestion de base de données sur le marché, à la fois commerciaux et à code source ouvert.

Données catégorielles

Les SGBD peuvent héberger beaucoup de types de données différents, les données catégorielles étant parmi les plus courantes.

Une variable catégorielle est une variable qui contient un nombre fixe de valeurs possibles divisées en

groupes ou catégories. Une base de données d'IFN contient typiquement plusieurs variables catégorielles (comme la classe d'utilisation des terres, le type de végétation ou la couleur du sol) ou d'autres variables qui ont été converties sous cette forme, par exemple comme données groupées (par ex. groupement d'espèces et classe de couvert arboré).

En gestion de données, les données catégorielles peuvent être stockées sous forme de liste, ou plus souvent de table de recherche. En outre, les données catégorielles peuvent être organisées sous forme de table(s) **plate(s)** ou **hiérarchique(s)**. Voyons des exemples des deux cas.

Construire un schéma pour une base de données

Lorsque l'on commence à construire un schéma pour une base de données d'inventaire forestier, on doit définir chaque objet (entité) et décrire les variables (attributs) qui doivent être mesurés. Avant de commencer à travailler au schéma de l'IFN, il faut avoir une idée claire de la structure logique de l'enquête, une liste détaillée des variables à mesurer durant le travail de terrain et décider de la manière optimale de collecter les données pour chaque variable.

Pendant qu'on construit le schéma, on doit examiner le plan d'échantillonnage et le plan parcellaire (document) ainsi que le manuel de terrain. Les documents correspondants peuvent nous indiquer quelles entités sont observées, comment elles sont définies, quelles sont leurs relations entre elles, et quelles propriétés (attributs) elles contiennent. Généralement, les concepteurs de la base de données se réunissent avec le concepteur du plan d'échantillonnage et du plan parcellaire et repassent la liste des variables que doit contenir la base de données.



Astuces rapides!

Il est recommandé de réaliser ce schéma conceptuel du modèle de données parallèlement à la création des formulaires de terrain afin d'éviter des difficultés à l'étape du traitement de données. Trop souvent, la conception de base de données est réalisée après la création des formulaires de terrain, et la structure de données qui fonctionne bien pour le terrain (formulaires) peut être inefficace dans un système de gestion de données..

Il est recommandé d'étudier comment des bases de données sont construites dans d'autres inventaires forestiers ou d'utiliser des modèles prêts à l'emploi de schémas de base de données.

Pour garantir que le schéma de base de données permet de dériver les estimations ou les résultats désirés, il est recommandé de tester l'ensemble du flux de travail avec des données de test (fictives) avant la mise en œuvre du schéma final et son utilisation par les équipes de terrain.

Les données d'évaluation des forêts sont souvent hiérarchiques. Le modèle de base de données hiérarchique ressemble à un organigramme ou un arbre généalogique – il possède un segment racine singulier (niveau 1) connecté à des segments de niveau inférieur. Parfois, la hiérarchie comprend des niveaux multiples.



Dans une base de données d'inventaire forestier réel, cependant, la structure du schéma est souvent beaucoup plus complexe que dans les diagrammes simples que nous venons de voir. En effet, de grands volumes de données diverses nécessitent plus d'entités. Par exemple, si des données sur la perturbation de la forêt sur des parcelles d'échantillonnage sont collectées avec des variables multiples, une nouvelle entité est nécessaire.

De plus, le groupement d'attributs peut générer une nouvelle entité dans la base de données. Le groupement est requis pour rendre les interfaces utilisateur (IU) plus faciles d'utilisation sur des dispositifs mobiles avec de petits écrans. De cette manière, l'enquêteur peut juste voir un petit nombre d'attributs sur l'écran à la fois.

Validation de données

La validation de données signifie vérifier l'exactitude et la qualité des données saisies avant leur utilisation, leur importation et leur traitement. Divers types de validation peuvent être réalisés. La validation de données est une forme de nettoyage des données. Le but de la validation de données est d'assurer que les données enregistrées adhèrent aux définitions et aux exigences d'exactitude fixées

dans le protocole d'inventaire. Le but est de créer une base de données cohérente, exacte et complète.

On peut ajouter des types multiples de règles de validation dans un SGBD afin de contribuer à améliorer la qualité des données. Dans certains systèmes, il y a deux classes de sévérité et de messages:

les **avertissements** et les **erreurs**. Les avertissements sont pour les cas où l'utilisateur du système doit procéder à une simple vérification (par exemple, un ratio dhp-hauteur anormal, ou un très grand arbre), et les erreurs sont des entrées qui doivent être corrigées. Les types de validation typiques sont les suivants:

1. Absence de valeurs nulles, par ex. aucune donnée manquante autorisée.
2. Type de données, par ex. nombre entier, chaîne.
3. Unicité, par ex. les doublons ne sont pas autorisés.
4. Validation par comparaison, où un attribut de données est comparé à un autre attribut de données, ou une comparaison des valeurs d'attribut est réalisée à l'aide d'opérateurs (autrement dit $<$, $<=$, $=$, $!=$, $>=$, $>$).
5. Expressions cohérentes, par ex. un courriel doit être épilé correctement.

Validation des listes d'espèces végétales

Tout inventaire forestier a besoin d'une liste d'espèces d'arbre et des listes supplémentaires pour les bambous, d'autres plantes, ou même d'autres taxa comme des animaux (par ex. dans l'IFN de Papouasie-Nouvelle-Guinée). Lorsque l'on construit une liste d'espèces pour un IFN ou l'on compare ou fusionne toute liste d'espèces d'arbre (ou plante), les noms taxonomiques doivent être vérifiés pour éviter les fautes de frappe ou l'inflation due à des synonymes. La normalisation des noms de plante est une étape critique dans un IFN et dans toute étude de recherche écologique.

Il existe de **nombreuses bases de données mondiales sur Internet qui incluent des millions d'espèces**. La plupart de ces bases de données peut être consultée librement en ligne ou en téléchargeant la base de données sur un dispositif local. Néanmoins, ces bases de données ont souvent été compilées à partir de sources de données hétérogènes variables en termes de moment de publication et lieu d'origine. À ce jour, la liste de référence de noms de plante vasculaire est **WFO Plant List** (en anglais), (littéralement, la liste de plantes) établie par les Jardins botaniques royaux de Kew. Cependant, la TPL n'a pas été actualisée depuis une décennie et a été créée à une époque où l'information phylogénétique sur de nombreux genres n'existait pas encore.

Dans le cas de longues listes d'espèces, une méthode efficace consiste à utiliser R pour extraire et traiter les données taxonomiques pour valider les noms de plante. R est l'outil dominant pour les statistiques, l'analyse de données et l'apprentissage automatique. Plus qu'un simple paquet statistique, il s'agit d'un langage de programmation. Un avantage d'utiliser R, par rapport à des feuilles de calcul, est que les scripts R peuvent fonctionner comme document dans la chaîne de traitement de données.

Il existe divers paquets R (par ex., `taxonstand`, `taxize`, `RBIEN`, `rentrez`) ou d'outils en ligne (par ex. [Global Name Resolver](#) (en anglais) ou le [Taxonomic Name Resolution Service](#) (en anglais)) appuyant les chercheurs pour vérifier l'information taxonomique. Néanmoins, certains de ces outils utilisent la TPL comme liste de référence, et il faut vérifier si la base de données de soutien est actualisée.

La validation des noms d'espèce peut s'appuyer sur une correspondance directe et une correspondance floue. Une recherche floue établit une correspondance entre les termes même s'il y a des fautes de frappe ou d'orthographe, et la correspondance floue trouve des résultats fondés sur des similarités. Certaines méthodes R peuvent présenter à la fois les résultats et les taux de réussite pour sélectionner les meilleures correspondances singulières attendues. À l'aide de R, les algorithmes de recherche peuvent être enchaînés afin qu'une liste de plantes soit validée à l'aide de répertoires multiples. Il faut remarquer que certains de ces répertoires exigent une clé d'interface de programmation d'application (*IDP*) qui identifie l'utilisateur. De même, le fournisseur de service peut établir des limites pour que les utilisateurs API n'excèdent pas un certain nombre de requêtes par seconde (par ex., voir [conditions d'utilisation de Kew](#) (en anglais)).

Codification des espèces

La codification des espèces est fortement recommandée pour stocker et traiter les données de taxon dans un SGBD. L'utilisation de codes permet de créer un identifiant unique pour chaque entrée dans une table d'espèces et simplifie le traitement de données et l'agrégation de données dans la phase de notification des résultats. En outre, s'il y a une faute de frappe dans le nom d'espèce (dans la table de taxa), cette erreur peut facilement être corrigée à tout moment. Ainsi, toutes les espèces doivent avoir un code unique, et les genres peuvent être codifiés si nécessaire.

S'il y a de nombreux types de plantes (ou d'animaux) différents dans l'inventaire, chacun peut avoir sa propre table de taxa, comme dans l'image ci-dessous. Voyons maintenant différentes méthodes de codification des espèces. D'abord, voyons des espèces codifiées sous forme de chaînes de texte.

Code	Rang	Nom scientifique	Synonymes
ACACI/AURIC	Espèce	Acacia auriculiformis	
ACACI/CRASS	Espèce	Acacia crassicarpa	
ACACI/FARNE	Espèce	Acacia farnesiana	
ACACI/HOLOS	Espèce	Acacia holosericea	
ACACI/LEPTO	Espèce	Acacia leptocarpa	
ACACI/MANGI	Espèce	Acacia mangium	
ACACI/MEARN	Espèce	Acacia mearnsii	
ACACI/SIMSI	Espèce	Acacia simsii	
ACACI	Espèce	Acacia sp.	

Puis, voyons des espèces codifiées sous forme de nombres.

Code	Rang	Nom scientifique	Synonymes
33296	Espèce	Abutilon pictum	
25213	Espèce	Abutilon Sinense	
50004	Espèce	Abutilon sp.	
25214	Espèce	Abutilon theophrasti	
38000	Espèce	Abutilon andamanica	
10001	Espèce	Abutilon auriculiformis	
23003	Espèce	Acacia catechu	
10002	Espèce	Acacia comosa	
23004	Espèce	Acacia concinna	



Astuces rapides!

Supposons que le pays ne dispose pas de système de codification des espèces en place. Dans ce cas, une méthode applicable consiste à utiliser un système où le code est formé par une combinaison de deux chaînes de texte: la première chaîne identifie le genre, et la seconde identifie l'espèce. Des codes plus longs peuvent être ajoutés pour les formes, sous-espèces, variantes, etc. si nécessaire.

Avantages de l'utilisation de codes textuels

Utiliser des codes textuels plutôt que des nombres présentent plusieurs avantages:

1. La codification des espèces peut suivre le même ordre alphabétique que les noms d'espèce.
2. De nouvelles espèces peuvent facilement être ajoutées ou retirées de la liste.
3. Si de nouvelles espèces sont ajoutées, elles peuvent plus facilement être intégrées aux codes des espèces connexes dans le même genre. Dans un système numérique, cette connexion serait perdue.
4. Une codification des espèces efficace facilite les opérations de jointure (comme pour les tables de recherche ou les équations externes).
5. L'agrégation par genre est rapide dans les phases d'analyse et de notification des résultats (par exemple pour calculer les indices de biodiversité et le nombre de genres d'arbre par parcelle).
6. Les codes textuels sont plus lisibles que les codes numériques.

Une fois la liste d'espèces validée et approuvée, et la collecte de données engagée, on peut seulement éditer les noms ou ajouter des synonymes. S'il manque des espèces, elles peuvent aussi être ajoutées (avec de nouveaux codes et noms) à la liste pendant le cycle d'inventaire forestier.

Pour en savoir plus sur la codification des espèces à l'aide de scripts R, veuillez consulter le [matériel d'Open Foris](#) (en anglais).

Résumé

Avant de conclure, voici les principaux points d'apprentissage de cette leçon:

- Les équipes d'inventaire forestier réalisent leurs tâches dehors, signifiant que leurs outils sont exposés aux conditions climatiques variable et à une manipulation exigeante. Ainsi, l'utilisation de dispositifs robustes est recommandée.
- Les ordinateurs de bureaux et portables courants sont souvent assez efficaces pour le travail quotidien de bureau avec des données forestières. Des serveurs locaux ou des services en nuage hébergés par un fournisseur de services en nuage sont nécessaires au niveau de l'organisation.
- La collecte mobile de données améliore la qualité des données car on peut utiliser des méthode de vérification des données directement sur le terrain.

- Avant de commencer à travailler avec le schéma de l'IFN, il faut avoir une idée claire de la structure logique de l'enquête et une liste détaillée des variables à mesurer. Consulter le protocole d'inventaire est essentiel.
- La validation de données signifie vérifier l'exactitude et la qualité des données sources avant d'utiliser, d'importer, ou de traiter les données.
- Une liste d'espèces validée et une codification des espèces adaptée rendra le traitement et la notification des résultats plus efficaces. Cela améliorera aussi la qualité de l'information des résultats.

Leçon 3: Publication en ligne et archivage de données

Introduction de la leçon

Cette leçon présente certaines des méthodes, techniques et outils actuels pour la notification des données des IFN.

Elle décrit aussi le concept de métadonnées, et fournit des orientations sur l'importance de l'archivage et du stockage de données pour leur utilisation future.

Objectifs

A la fin de cette leçon, vous serez en mesure de:

1. Décrire le rôle de la technologie de tableau de bord de données dans la notification de l'information d'inventaire forestier.
2. Expliquer la technologie de «traitement analytique en ligne» (OLAP, en anglais) et comment elle fonctionne.
3. Définir le concept et les exigences de base des métadonnées et de l'archivage de données.

Tableaux de bord et portails de données

Qu'est-ce qu'un tableau de bord de données?

Les tableaux de bord et les portails ont la même fonctionnalité mais différentes utilisations. Les portails fournissent un répertoire centralisé pour l'information clé pour une organisation ou le public, et ils contiennent typiquement du texte enrichi, des raccourcis, des images et des cartes interactives. Certains portails peuvent fournir une analyse en temps réel des données sous-jacentes.

Les tableaux de bord, par ailleurs, offrent une vision rapide afin de faciliter la compréhension, avec un accès facile aux tableaux, graphiques et rapports les plus fréquemment nécessaires. En outre, en termes de règle, les portails produisent des représentations statiques des résultats dans des tables et des cartes prédéfinis, tandis que les tableaux de bord fournissent un contenu plus dynamique en utilisant des modèles de données et une analyse en temps réel.

Voici des exemples de tableaux de bord et de portails d'IFN à explorer (en version originale):

Évaluation des ressources forestières mondiales

Satellite Land Monitoring Systems (SLMS), empowered by UN-REDD/FAO

Bangladesh Forest Inventory Results

Germany data dashboard (various forest inventories)

Tableau de bord de données de la France

Portail IFN du Canada

Papua New Guinea Climate Change and Forest Monitoring Web portal

Portail du système national de surveillance des forêts (République démocratique du Congo)

Système de suivi national des terres du Suriname

Traitement analytique en ligne (OLAP)

Qu'est-ce que OLAP?

Le traitement analytique en ligne (OLAP) est une technologie de traitement informatisé qui permet l'exécution rapide de requêtes analytiques complexes. Il permet à l'utilisateur de trancher et filtrer des données, et de produire des résultats spécifiques sans avoir à mettre en œuvre des calculs séparés. C'est pourquoi les bases de données OLAP sont appréciées pour la publication des résultats.

OLAP appartient à la catégorie plus vaste de l'intelligence économique, comprenant les bases de données relationnelles, l'écriture de rapports et l'exploration de données. Les applications OLAP ont d'abord été appliquées à des rapports économiques, mais aujourd'hui cette technique est appliquée dans de multiples secteurs, y compris la foresterie et l'agriculture. Les rapports d'inventaire forestier et les résultats des IFN peuvent être traités et présentés avec OLAP.

Mais des **défis sont associés à OLAP – il utilise une terminologie différente de la terminologie normalisée des statistiques et des bases de données**. Nous n'explorerons cependant pas ces différences, mais nous intéresserons plutôt aux principaux avantages et applications des techniques OLAP. Nous verrons ensuite trois exemples d'utilisation d'OLAP dans le traitement de données d'IFN.

Système d'analysé et de publication des résultats de l'IFN suédois

Une base de données relationnelle permet à l'IFN de la Suède d'utiliser la technique OLAP. Cette base de données permet une publication et une analyse de données rapides pour les produits normalisés, utilisés pour la notification nationale et internationale. L'IFN suédois stocke les données les plus fréquemment utilisées dans Analysis Services de Microsoft (MS), qui est un outil OLAP et d'exploration de données. Les outils utilisés pour la publication et l'analyse de données sont MS Excel, MS PowerPivot et différents outils MS Reporting Services.

Pour les rapports ad-hoc et les demandes spécifiques, l'IFN suédois utilise les outils SAS Institute et Microsoft Power Pivot extraire directement les données de la base de données et réaliser l'analyse. Cependant, la plupart des autres paquets statistiques peuvent extraire et manipuler des données de la base de données MS SQL Server.

TaxWebb (en anglais) est l'outil d'analyse interactif de l'IFN de la Suède qui cherche à fournir un accès rapide et facile aux statistiques. Il permet aux utilisateurs d'élaborer leurs propres rapports.

IFN de la Suisse

Le système d'analyse et d'inventaire forestier national (NAFIDAS) a été mis au point pour l'IFN de la Suisse. NAFIDAS utilise une conception similaire à celle appliquée dans l'IFN suédois – il s'agit d'un stock de données opérationnelles avec des interfaces pour les systèmes de données sources (sources de données internes et externes), une **aire de stockage de métadonnées** (en anglais) et des outils de présentations pour les utilisateurs finaux.

NAFIDAS produits des tables et des cartes en utilisant les données de l'IFN de la Suisse ou des données des inventaires régionaux, et il présente trois composants principaux:

- 1) une application web pour la gestion, la documentation et la publication des données,
- 2) des bases de données pour le stockage, et
- 3) l'application d'analyse de données pour l'analyse.

Open Foris calc avec Saiku Analytics

La vidéo de démonstration d'Open Foris (en anglais) montre un exemple d'utilisation d'OLAP pour la publication des résultats dans Open Foris Calc* à l'aide du logiciel Saiku Analytics.

Saiku Server est un logiciel à code source ouvert en ligne qui facilite la visualisation des données et les requêtes de données. Bien qu'une version gratuite soit disponible sur le site web de Saiku, une version spéciale a été personnalisée pour une meilleure compatibilité avec Collect Earth. Saiku Server est inclus dans l'installateur de Collect Earth.

Remarque: La présentation de Saiku commence à la minute 3:44.

* Calc a été remplacé par Open Foris Arena, mais il est toujours disponible en tant qu'outil existant dans la suite Open Foris.

Archivage de données

Chaque organisation est responsable de la préservation des données. En général, les formulaires sur papier bien organisés et stockés sont assez durables, mais ils sont difficiles à réutiliser, ou même à sauvegarder.

Lorsque l'on utilise des médias numériques, les disques durs plantent, des fichiers sont placés au mauvais endroit, les formats de données deviennent dépassés, des mots de passe sont oubliés, les médias de stockage deviennent obsolètes (par ex. les disquettes) et même les personnes responsables peuvent déménager, prendre leur retraite ou mourir. Heureusement, les avancées technologiques et l'avènement des archives publiques rendent la préservation des données à long terme plus facile et plus fiable.

Une **archive de données** est un **fonds d'ensembles de données avec des métadonnées connexes stocké** de sorte qu'un éventail d'utilisateurs puisse situer, acquérir, comprendre et utiliser les données. Les données archivées sont sécurisées face aux catastrophes naturelles et anthropiques, et sont conservées sous une forme qui continuera à être accessible à mesure que la technologie change.

Avantages de l'archivage de données

L'archivage de données est précieux pour de nombreuses raisons, les plus importantes étant:

- 1) réduire la perte de données;
- 2) utiliser les données pour de nouvelles analyses et études. Les données anciennes peuvent aussi servir de données de référence précieuses pour analyser des tendances à long terme;
- 3) utiliser les données comme matériel de formation pour les étudiants car ce sont des moyens rentables d'améliorer la productivité scientifique d'un pays. Les données archivées peuvent

aider à développer un groupe de scientifiques très bien formés et productifs au niveau postdoctoral;

- 4) dissiper les préoccupations concernant l'exportation de propriété intellectuelle, et le manque d'inclusion de scientifiques locaux dans les efforts de collecte de données qui affectent souvent les scientifiques étrangers travaillant dans les pays tropicaux;
- 5) vérifier les résultats et corriger les erreurs; et
- 6) remplir les mandats et les exigences des donateurs et des gouvernements.

Métadonnées

Les métadonnées sont une **information structurée qui décrit, explique, situe ou rend facile à extraire, utiliser ou gérer une ressource d'information**. Les métadonnées sont souvent appelées **données sur les données**, ou information sur l'information.

Les métadonnées fournissent de l'information qui permet aux utilisateurs de comprendre pleinement **les données** (par ex. documents, images, ensembles de données), **les concepts** (par ex. modèles de classification) et **les entités du monde réel** (par ex. organisations, lieux, protocoles d'échantillonnage). Un exemple typique d'inventaire forestier serait un protocole de terrain détaillé qui décrit comment les mesures sont prises et quelles catégories (codes) sont utilisées.

Dans le contexte des IFN, les métadonnées peuvent être utilisées pour répondre à des questions comme: quelles données sont collectées, comment, pourquoi, quelle est leur fiabilité, et quelles questions doivent être prises en compte pour travailler avec elles. Les métadonnées peuvent aussi décrire comment obtenir les données, quels outils sont nécessaires pour travailler avec les données, et d'autres questions connexes.

L'objectif de documenter les données est de fournir assez d'information sur l'ensemble de données pour permettre à quelqu'un de travailler facilement avec ces données dans 20 ans et d'obtenir des résultats reproductibles. Plus il y a de propriétés de métadonnées attachées aux données, plus les données sont précieuses et intelligentes.

Les métadonnées sont importantes pour tout répertoire d'information car elles apportent la capacité de:

- déterminer la disponibilité, l'emplacement, l'âge de création, la propriété et l'accessibilité des données;
- comprendre et utiliser correctement les données;

- gérer plus efficacement les données; et
- garantir l'interopérabilité des données.

IFN et métadonnées

Le système d'information de l'IFN pour la gestion de données doit être bien documenté avec des métadonnées. Toutes les descriptions doivent être compatibles avec les manuels de terrain qui décrivent les données.

Pour une base de données d'IFN, les métadonnées peuvent contenir une documentation détaillée des terrains comprenant les ensembles de données, la définition, le type de mesures, les unités le cas échéant, et tout vocabulaire contrôlé ou liste de codes présents dans les données.

Une norme de métadonnées est un ensemble commun de termes et définitions qui décrit les données, montrant les propriétés caractéristiques à enregistrer et les valeurs que doivent avoir les propriétés. Les normes de métadonnées sont spécialement bien développées pour les données géospatiales. .



Astuces rapides!

Lorsque l'on établit des métadonnées pour de l'information sur les forêts, on peut citer des exemples mis en œuvre dans la foresterie et d'autres secteurs. Pour les IFN, on peut consulter la [documentation de base de données](#) (en anglais) du Programme d'inventaire et analyse des forêts (FIA, en anglais) des États-Unis. Le [Programme d'inventaire et analyse des forêts du Vermont](#) (en anglais) et l'[Inventaire forestier national des terres boisées en Angleterre en 2018](#) (en anglais) apportent d'autres exemples de métadonnées..

Microdonnées statistiques

Les microdonnées sont un concept qui se réfère aux petits ensembles de données ou données agrégées à des niveaux spécifiques (spatialement ou contextuellement). Cela est particulièrement appliqué dans des organisations internationales, comme dans la [Banque mondiale](#) (en anglais), [Eurostat](#) et certaines organisations des Nations Unies, y compris la [FAO](#).

Les microdonnées peuvent inclure des caractéristiques d'unités d'une population, comme les individus,

les ménages ou les établissements, relevés par un recensement, une enquête ou une expérience. Ainsi, le concept est plus appliqué dans les enquêtes socioéconomiques que dans les inventaires forestiers (biophysiques), mais c'est un nouveau concept et il peut être utile de le comprendre dans le contexte de la gestion de données.

Les catalogues de microdonnées sont des ensembles de données pour stocker et partager des ensembles de données relevés. Les données de l'ensemble de données peuvent provenir d'un fonds de données primaires (voir l'[exemple des microdonnées](#) (en anglais) des parcelles d'échantillonnage de l'IFN suédois) ou de données secondaires à travers l'agrégation ou la synthèse. Les résultats sont le plus souvent publiés sous forme d'agrégats pour des questions de confidentialité et parce que de grandes quantités de données sont impliquées. Typiquement, tous les identifiants personnels sont retirés pour garantir le respect de la vie privée et la confidentialité. Les catalogues de microdonnées peuvent très bien s'adapter à l'information sur les forêts sous forme de statistiques traitées.

Les microdonnées agrégées peuvent aussi être précieuses pour l'analyse politique, la recherche et les statistiques à un niveau de désagrégation poussé (par ex. par positions, populations, groupe d'âge). Elles permettent d'obtenir une vision claire de questions en étudiant les relations et les interactions entre des phénomènes. Les microdonnées sont alors centrales pour concevoir des projets et formuler des politiques, cibler des interventions et un suivi, et mesurer l'impact et les résultats des interventions.

[Le Catalogue de microdonnées sur l'alimentation et l'agriculture \(FAM\) de la FAO](#) propose un inventaire d'ensembles de données collectés à travers des enquêtes auprès des exploitations et des ménages qui contient de l'information relative à l'agriculture, la foresterie, la sécurité alimentaire et la nutrition.

Résumé

Avant de conclure, voici les principaux points d'apprentissage de cette leçon:

- Les portails fournissent typiquement de l'information clé pour l'équipe d'une organisation ou le public.
- Les tableaux de bord offrent une vision rapide de l'état de l'opération, fournissant un accès facile pour voir les principales mesures, tableaux, graphiques et rapports.

- Le traitement analytique en ligne (OLAP) est une technologie de traitement informatisé qui permet l'exécution rapide de requêtes analytiques complexes.
- La préservation des données et la responsabilité de toute organisation. Les archive de données rendent la préservation des données à long terme plus facile et plus fiable.
- Les métadonnées sont une information structurée qui décrit, explique, situe ou rend facile à extraire, utiliser ou gérer une ressource d'information.
- Les catalogues de microdonnées sont des ensembles de données pour stocker et partager de petits ensembles de données relevés.