

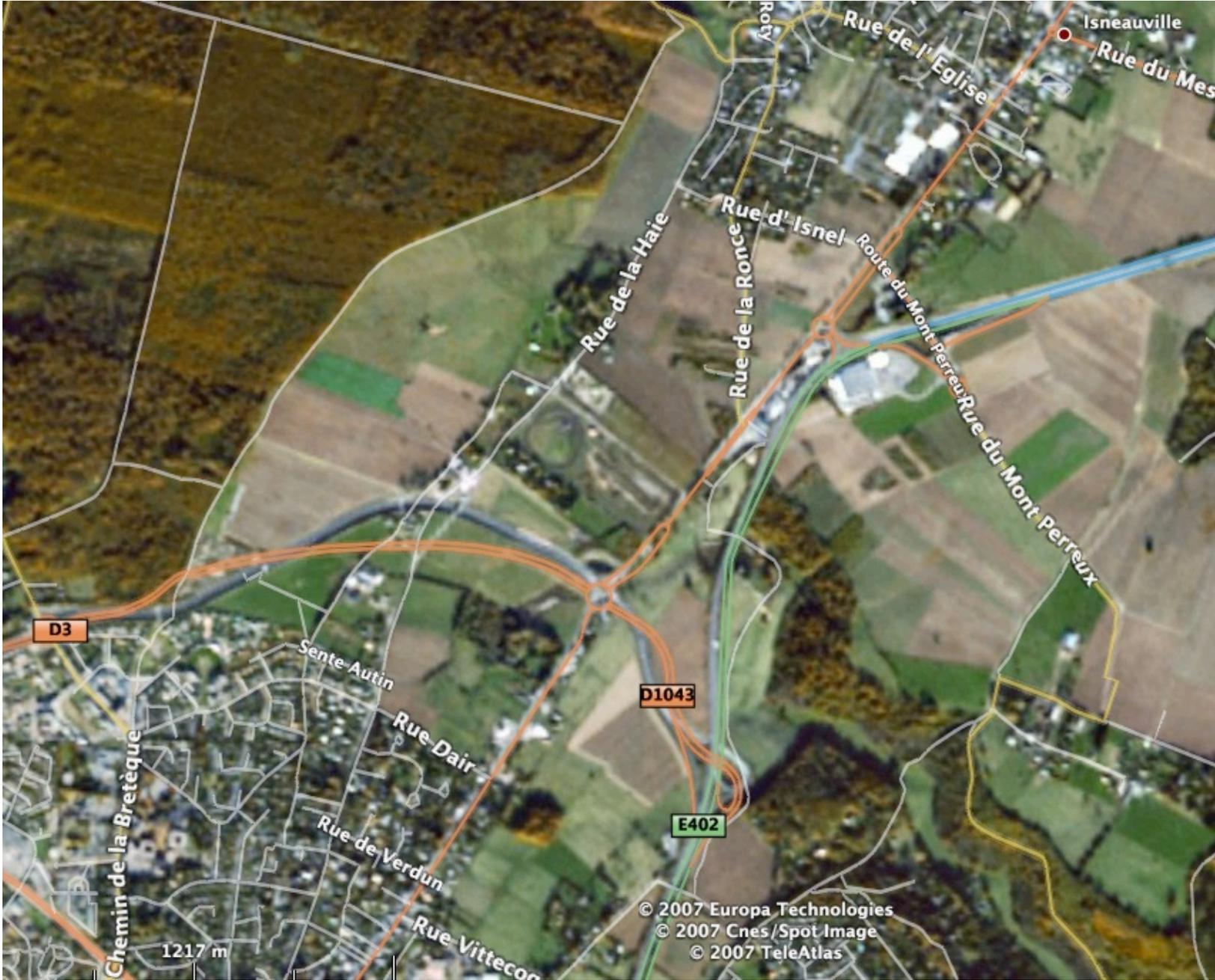
Systemes d'Informations Géographiques - Le multi-sources

Institut National des Sciences Appliquées - Rouen
Département Architecture des Systèmes d'Information
michel.mainguenaud@insa-rouen.fr

Positionnement du problème

- Schéma :
 - Bases de données géographiques sont non-exhaustives
 - Description d'un paysage vs. nombre de clients d'une entreprise, nombre de produits
 - Ambiguïtés d'interprétation
 - Spécifications détaillées d'acquisition vs. remplissage des instances
- Objectif :
 - Différentes sources doivent coopérer/s'intégrer

Bois-Guillaume (76)



Brienne-Le-Chateau (10) : décalage : 70m



Exemples de bases IGN

- BDCarto : 1 / 50 000, tout le territoire (décamètre)
- BDTopo : ne couvre pas tout le territoire (mètre)
 - Standard : ensemble de bases indépendantes couvrant un territoire rectangulaire au 1 / 10 000 (n'est plus mise à jour). Les MAJ antérieures ne respectent pas toutes les mêmes spécifications.
 - Actuelle : le découpage correspond aux départements avec prise en compte des raccords
- BDGéoroute : pour les applications routières (mètre)

Problèmes de l'indépendance des bases

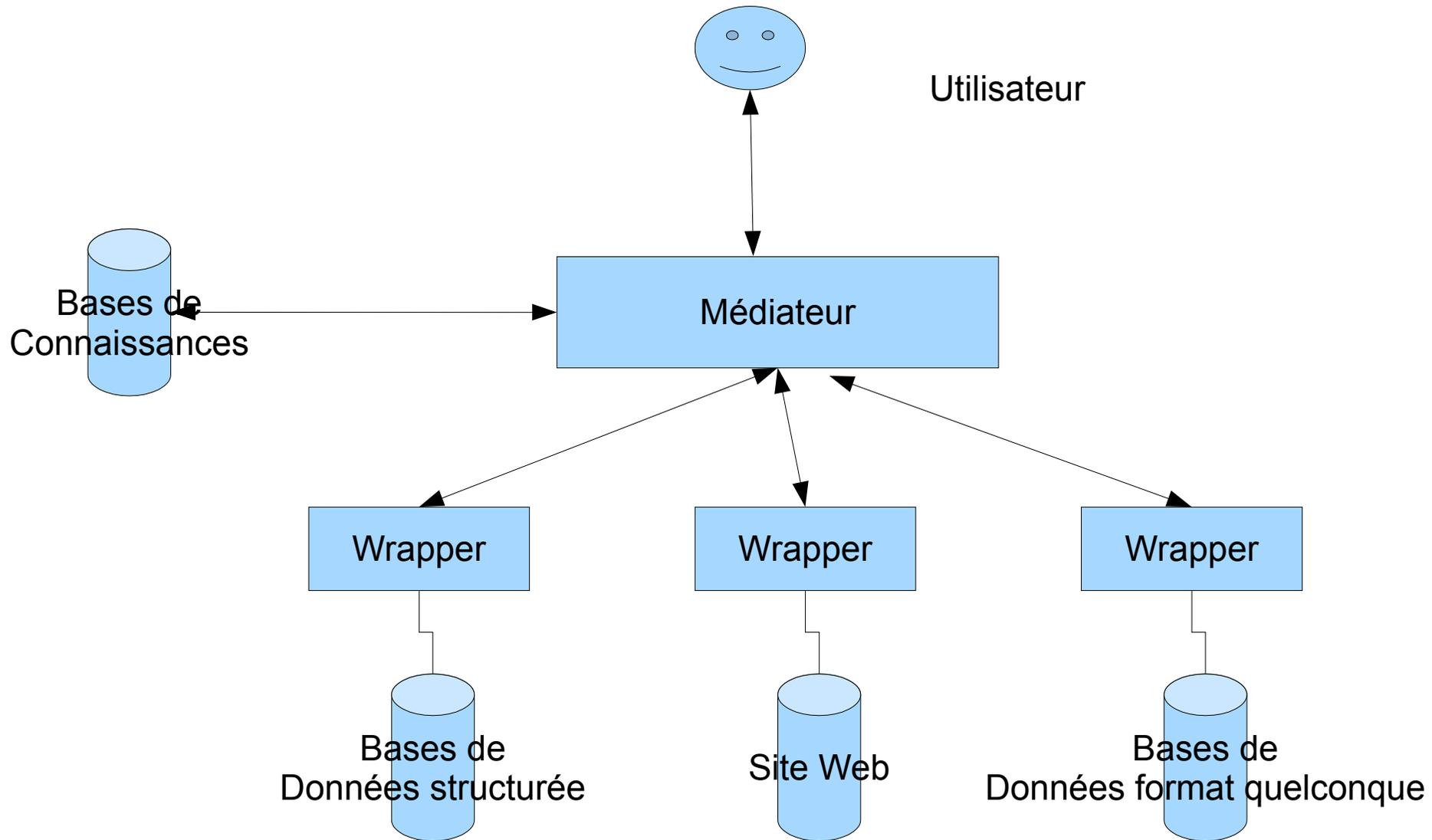
- « Basique » : cartes scannées ne nécessitant pas d'analyse mais une présentation – pouvoir d'expression faible
- Partage d'information : problème de fusion des modèles (appariements) au niveau schémas et au niveau instances
- Problème de la propagation des mises à jour sur chacune des bases indépendantes

Vocabulaire

- Méta-donnée : désigne de manière générale toute donnée relative à d'autres données
- Interopérabilité : capacité des systèmes informatiques à s'échanger de façon cohérente des données
- Intégration : consiste à constituer un tout cohérent à partir de données provenant de diverses sources

Architectures ⁽¹⁾

- Médiation
 - Données peu structurées et on dispose de peu de méta-données
 - Exemple Web
 - Nécessite des outils d'IA
 - Intervention d'un médiateur
 - Données : accessibles directement à partir des sources
 - Information : données interprétées
 - Logiciel muni d'une base de connaissance pour passer de données à information grâce à une description formelle des contenus et des propriétés (logique de description, clauses de Horn, ...)
 - Description globale, fixe et unique du domaine d'application

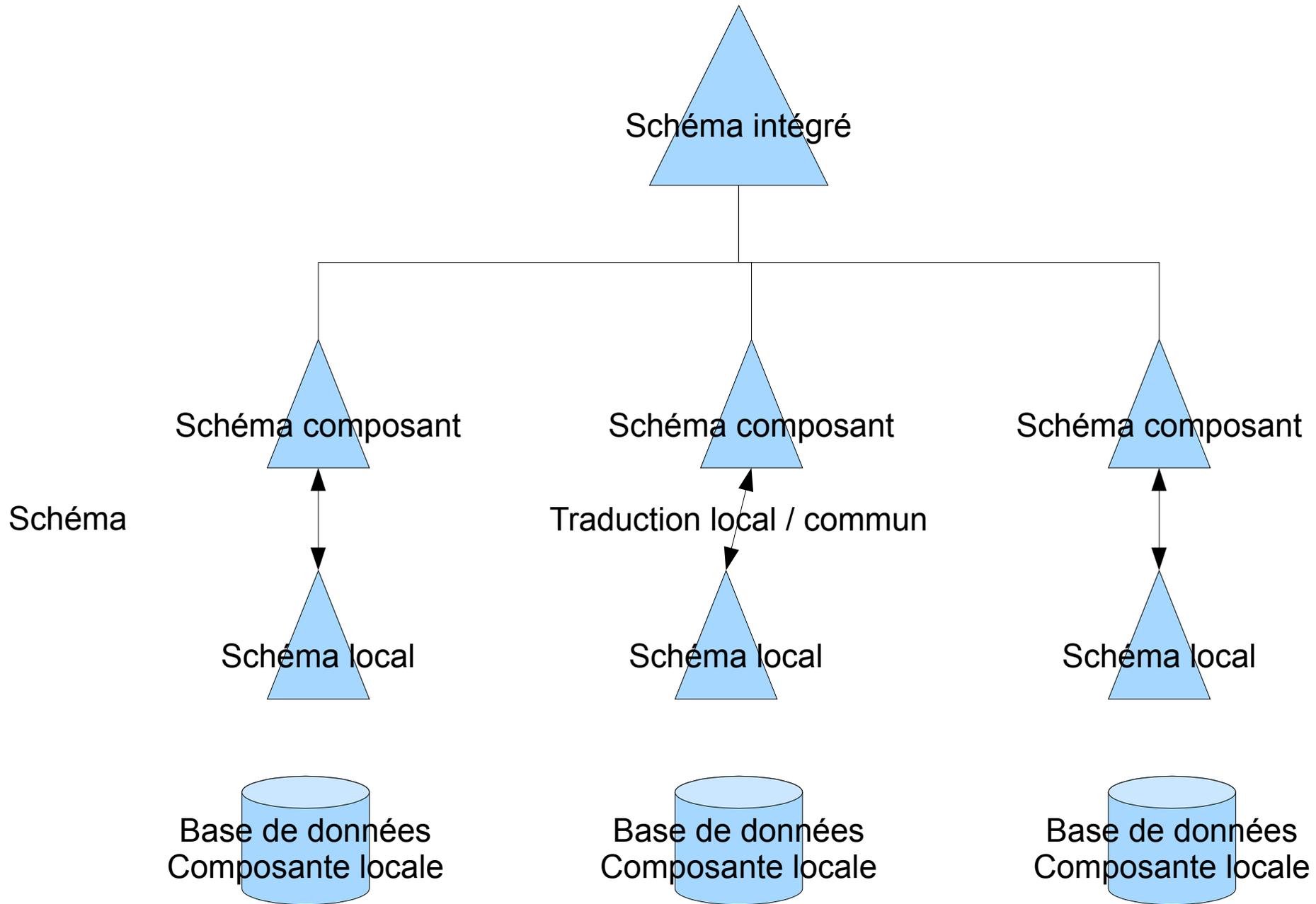


Techniques

- Local as views : les schémas locaux sont des vues sur un schéma global
- Global as views : processus d'intégration de vues (définies par les wrappers)
- Critères :
 - Complexité des sources
 - Variabilité du nombre de sources
- Objectif : Recherche et interprétation des données provenant de sources nombreuses et peu structurées

Architecture (2)

- Fédération :
 - Objectif : garder l'autonomie d'accès aux bases tout en proposant une intégration
 - Fortement couplée vs. Faiblement couplée
 - Fortement :
 - Choix d'un modèle commun à tous si possible plus riche que les modèles locaux (-> schémas composants)
 - Chaque base de données dispose de son schéma local dans son modèle de référence
 - Objectif : interrogation et gestion intégrée d'un nombre fixe de bases structurées (MAJ)



Fortement couplé

- Eventuellement participation d'une base réduite à un sous-ensemble de son schéma (schéma d'exportation)
- Schéma fédéré :
 - Intégration des différents schémas d'exportation
 - Liens entre les données des différentes bases concernant la même entité
 - Règles de contrôle de cohérence en cas de présence d'une information sur deux bases

Faiblement couplé (multibases)

- Faire communiquer deux bases :
 - Outil de communication de type ODBC (Open DataBase Connectivity)
 - Langage commun d'interface (exemple SQL dans les BD distribuées homogènes)
- On ne se soucie pas de l'hétérogénéité du contenu (seulement du point de vue langage d'interrogation) : il n'existe pas de schéma global
- L'utilisateur relie les réponses obtenues lors d'envoi de plusieurs requêtes

Conflits d'intégration

- Syntaxique :
 - Concerne les modèles de données (objet, relationnel)
- Structure :
 - Différences d'organisation dans les schémas (classe vs. Attribut, classe vs. deux classes)
- Sémantiques :
 - Problèmes de vocabulaire : homonymie, synonymie, ...
- Instanciations :
 - Problèmes de saisies incorrectes, oublis, MAJ non reportées, ...

Extension au spatial ⁽¹⁾

- Méta-données géométriques :
 - Précision,
 - Exactitudes susceptibles de provoquer des conflits de données
- Conflits de stockage de l'information :
 - Déduction : pont lors d'un croisement route/rivière stocké dans une base mais pas dans l'autre
 - Géométrie : représentation d'une route sous la forme d'un trait (1D) ou zone (2D)

Extension au spatial (2)

- Définition de classe:
 - Fragmentation : découpage des entités retenues
 - Segmentation : basée sur des attributs différents
 - Découpage d'un cours d'eau selon le régime d'écoulement vs. selon la largeur
 - Granularité : même attribut mais avec un critère d'homogénéité différent
 - Saisie si Longueur > 50 m vs. Longueur > 80 m
 - Décomposition : une entité correspond à plusieurs autres entités dans l'autre schéma
 - Terrain/Tribune vs. Terrain de football

Hypothèses de travail

- Ontologie du domaine disponible : Concepts dont l'existence est admise et la définition partagée par l'ensemble des experts du domaine
- Création d'une ontologie ad-hoc (partielle) étendue par un processus incrémental
 - Impose l'indépendance de l'implémentation des schémas et de la structure des spécifications papiers

Problèmes classiques

- Imprécision (caractère vague ou approximatif de la sémantique) → Problème ouvert
 - Outil de mesure
 - Connaissances subjectives
- Incertitude (liée à la validité d'une connaissance)
 - Dimensions :
 - Spatiale, Temporelle, Fonctionnelle
 - Evaluation : booléenne → doute
 - Fiabilité de la source, Erreur
- Ambiguïté
- Incomplétude

Imprécision

- Vague :
 - Définie à l'aide de connaissances subjectives ou flexibles
 - Insuffisance des outils de mesure
 - Exemple : « proche de » → dépend du contexte
- Approximation
 - Limites mal définies
 - Exemple : « Baie du Mont Saint Michel »
- Au sens large : vague et approximatif

Incertitude

- Mesure (localisation)
- Classement (mauvaise affectation dans une classe d'objet à cause d'une erreur dans l'expertise)
- Dans le regroupement de classes d'objets (les propriétés sont dès lors différentes)
- De généralisation spatiale (déplacement, simplification, problème d'échelle)
- Temporelle : non validité de l'objet au moment de son utilisation (couplée ou non à une erreur de classement)
- ... un bug

Ambiguïté

- Doute sur la manière de définir un objet
 - Classes disjointes / échelles différentes
 - Plusieurs sens possibles
- Conflit (désaccord)
 - Deux classifications contradictoires pour un objet sont possibles
 - Exemple : zone tampon spatial / classe A vs. B
- Non-spécificité
 - Sémantique non disjointe
 - Exemple : A est au nord de B, objet appartient à l'intersection spatiale des deux classes

Incomplétude

- Liée à l'absence de connaissance (valeurs nulles) ou des connaissances lacunaires
 - Absence : sur la représentation spatiale ou attributaire
 - Lacune : description uniquement partielle de la structure englobante (exemple en archéologie)

Outils formels de modélisation

- Incertitude
 - Probabilités (objectives et subjectives)
- Imprécis
 - Ensembles flous : Gradation sur $[0,1]$ sur l'adéquation à une notion (gradation de l'appartenance à une classe)
 - Ensembles approximatifs : haute et basse
- Mixte
 - Théories des possibilités : possibilité et nécessité
 - Théorie de l'évidence (Fonctions de croyance)

Probabilités

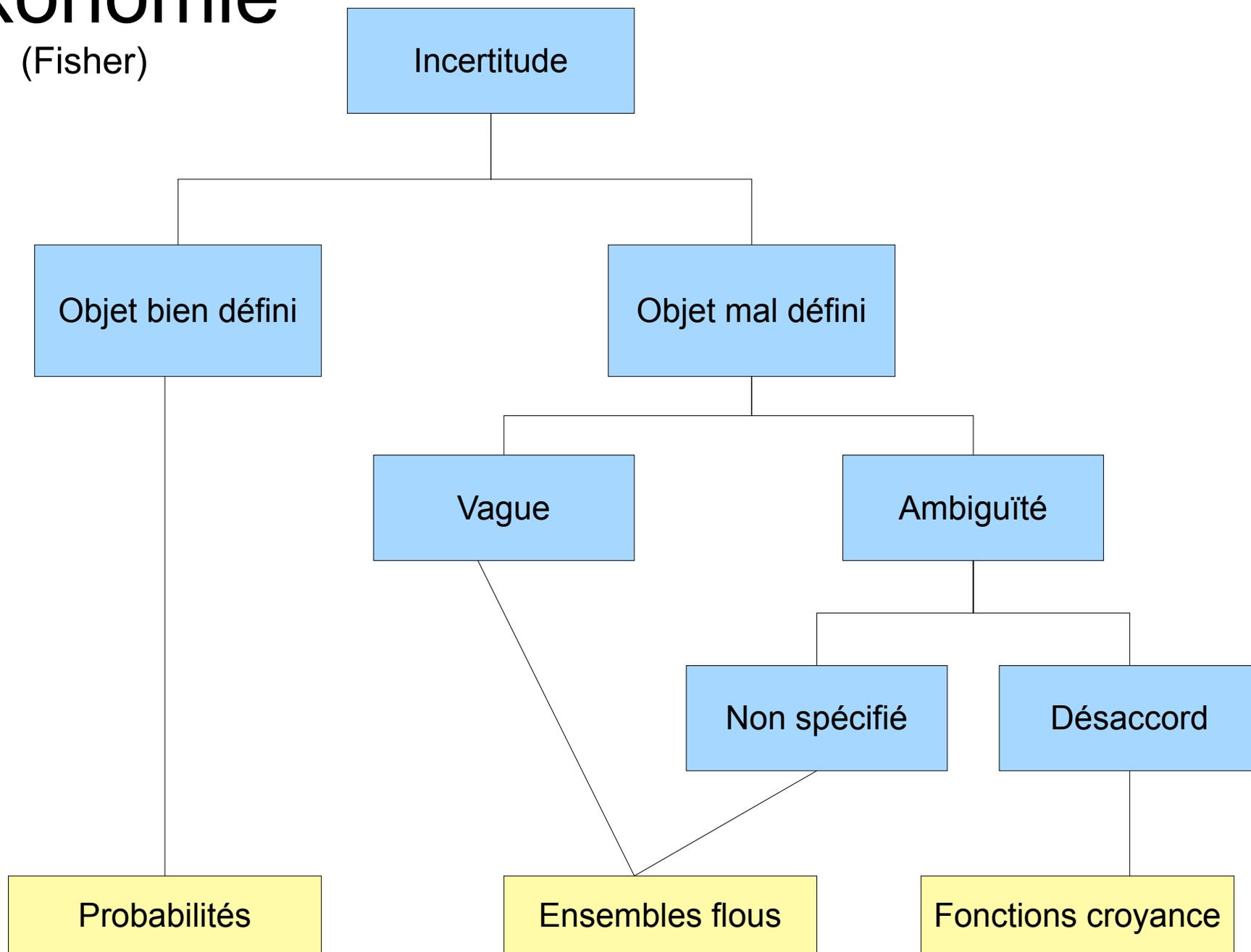
- Objectives :
 - Fonctionnement aléatoire. Probabilité = fréquence de réalisation lors d'une répétition
- Subjectives (bayésiennes) :
 - Affecté à n'importe quel fait (même non aléatoire). Probabilité = la confiance que l'on donne à la validité du fait (degré de croyance)
 - Généralisées par Dempster-Shafer (Evidence)

Théorie de l'Evidence

- Raisonnement à partir de deux mesures
 - Crédibilité
 - Plausibilité
- Approche ensembliste de l'imprécis
 - Passage de $x \in A$ à une distribution de probabilité sur les énoncés de la forme $x \in A$
 - Fonction de masse m , distribution de probabilités sur l'ensemble des parties des événements élémentaires

Taxonomie

(Fisher)



Conclusion

- Problème important
 - Hétérogénéité des fournisseurs
 - Applications déjà existantes
 - Double problème : Schéma / Données
- Se rapproche des problématiques
 - Système d'information au sens large
 - Web « sémantique »
- Encore ouvert