

Géomodélisation et Informatique :

Formalisation des représentations et représentation des formalismes

**Michel PERRIN, Jean-François RAINAUD,
Laura S. MASTELLA**

Plan de la présentation:

- **Objectifs / Contexte**
- **Formalisation des représentations métier**
 - Relations entre unités géologiques
 - Description des temps géologiques
 - Description des données pétrologiques à l'aide d'une ontologie "sur mesure"
 - Passage des concepts aux instances
- **Représenter lisiblement les formalismes**
 - Difficultés du dialogue Géologues – Informaticiens
 - Difficultés conceptuelles au niveau du langage
 - Difficultés liées aux imperfections des représentations graphiques
- **Recommandations et conclusion**

OBJECTIFS / CONTEXTE

- Nous nous proposons d'examiner les problèmes posés par l' **harmonisation des représentations formelles de la connaissance** dans le cas d'un projet d'ingénierie portant sur la modélisation géologique et associant des experts métier et des informaticiens
- Nous nous fondons pour cela sur l'expérience :
 - de **12 ans de pratique** en matière de géomodélisation associant des experts métiers et des informaticiens
 - d'une collaboration multidisciplinaire dans le cadre d'un **projet ANR** visant à la mise au point d'une plateforme multi-services pour l'étude des réservoirs géologiques (**projet e-Wok Hub, 2006-2009**)

Nous envisagerons successivement deux catégories de problèmes :

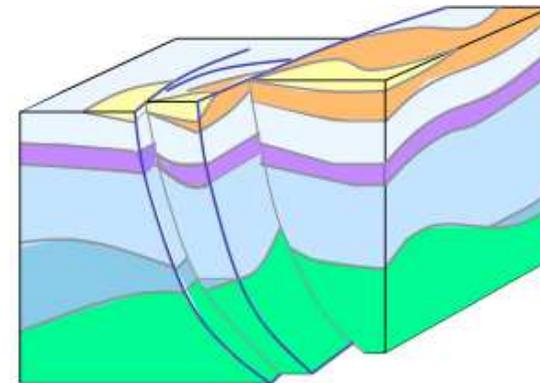
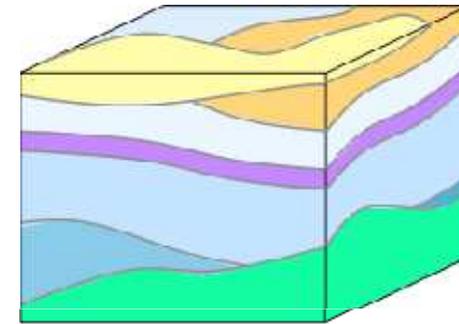
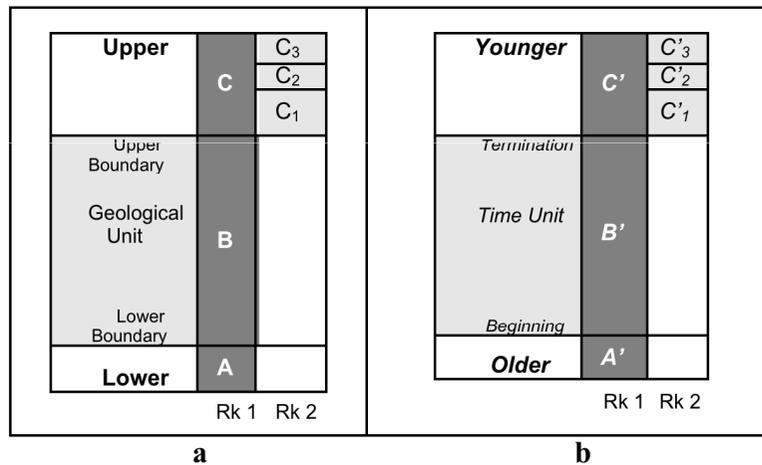
- **la formalisation des représentations** couramment utilisées par les géologues; cette formalisation est indispensable pour répondre aux exigences des constructeurs de logiciels; ce problème intéresse principalement **les informaticiens**
- **la représentation des formalismes** sous une forme qui soit lisible par les professionnels du domaine; dans notre cas, ce problème intéresse principalement **les géologues.**

FORMALISATION DES REPRÉSENTATIONS MÉTIER

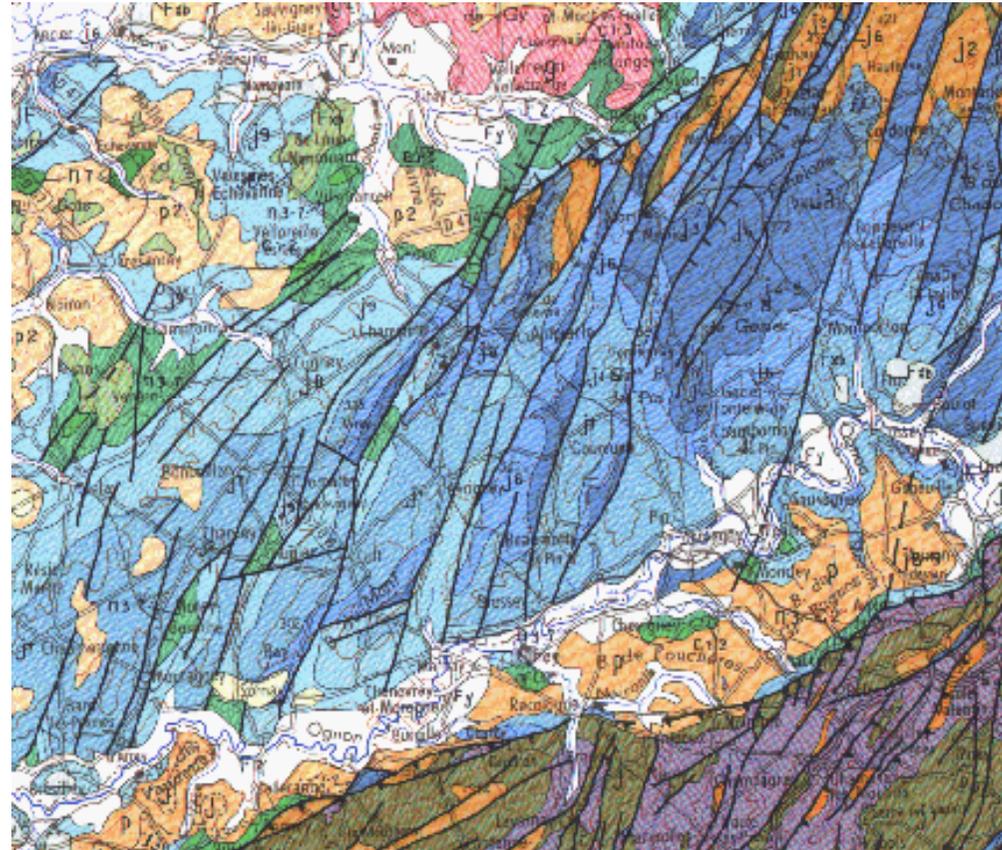
Exemple 1

Relations topologiques entre surfaces géologiques

Modèle stratigraphique:



Carte géologique



Problématique:

Si le géologue vise à réaliser l'assemblage d'un ensemble de surfaces géologiques **de manière automatique** grâce à un modelleur, il lui faut préciser les éléments d'**interprétation** suivants:

- **âges relatifs** des différentes surfaces les unes par rapport aux autres ,
- **relations d'intersection** de chacune des surfaces avec ses voisines.

Ceci peut être réalisé grâce à un **Schéma d'Evolution Géologique (SEG)** reposant un **modèle conceptuel** relatif aux objets géologiques modélisés.

Eléments complémentaires de formalisation

Modèle conceptuel

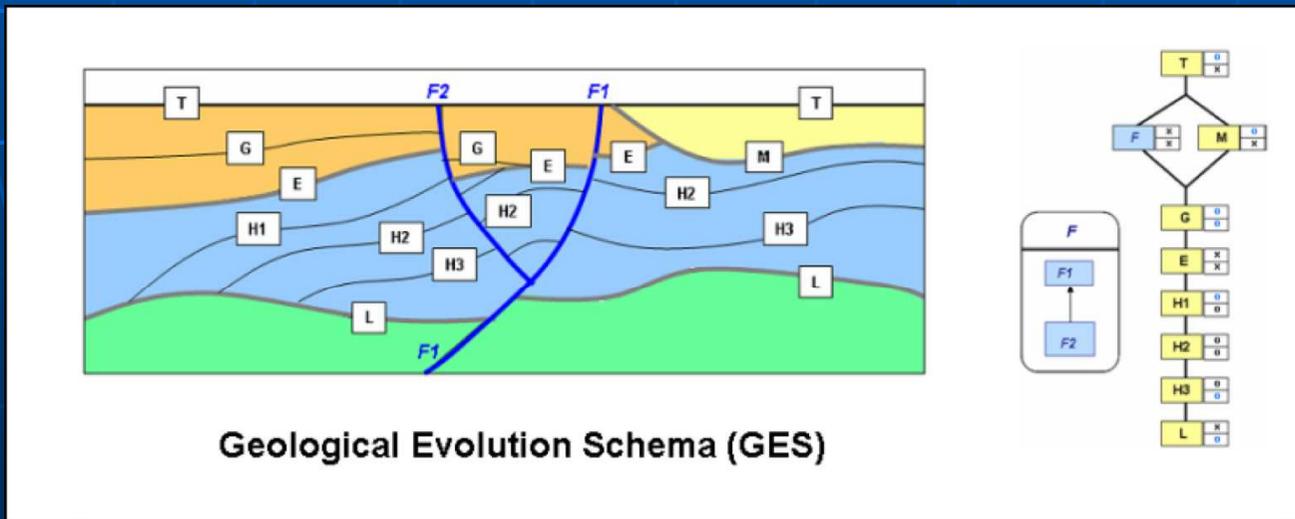
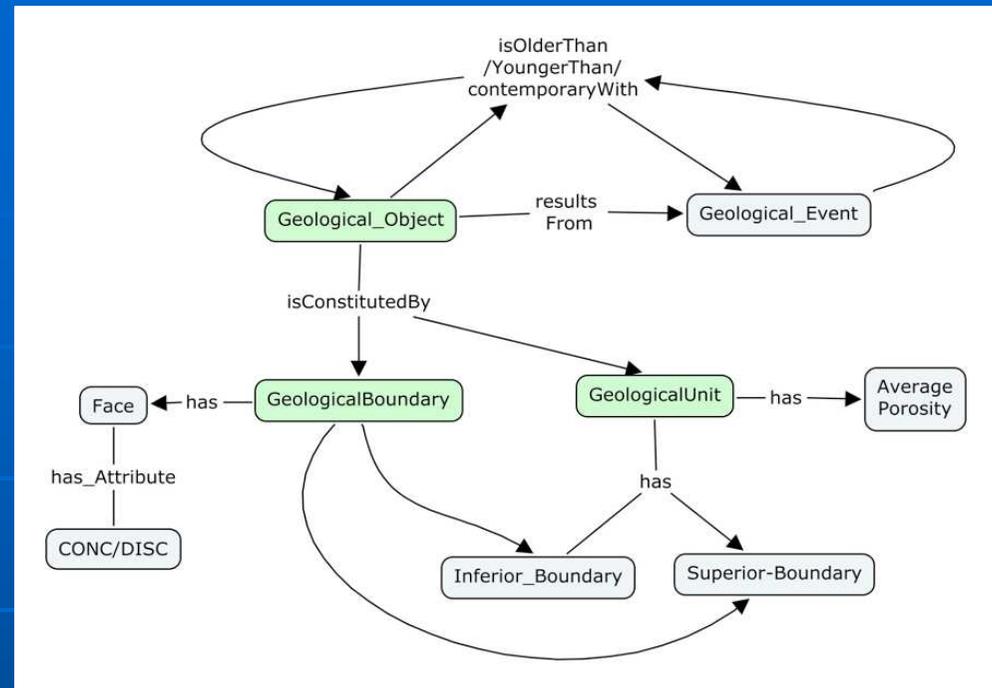
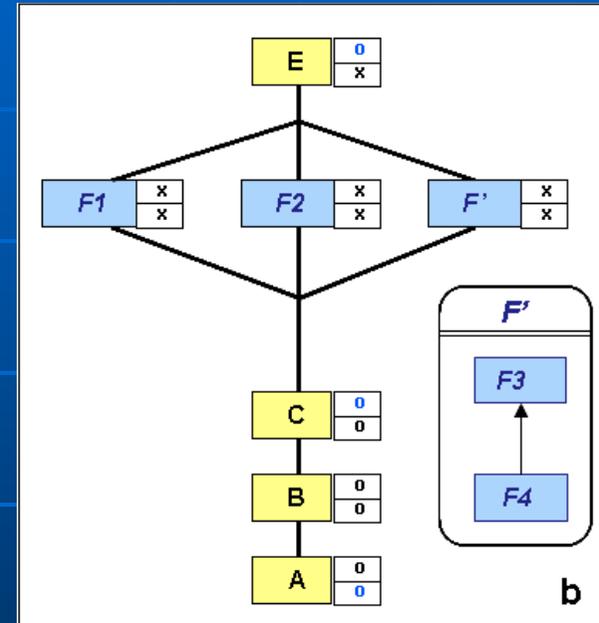
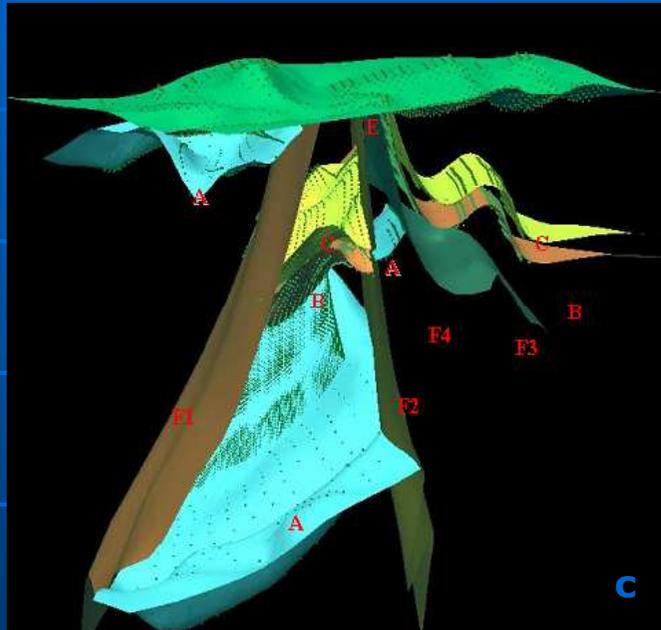


Schéma d'Evolution Géologique



Résultat

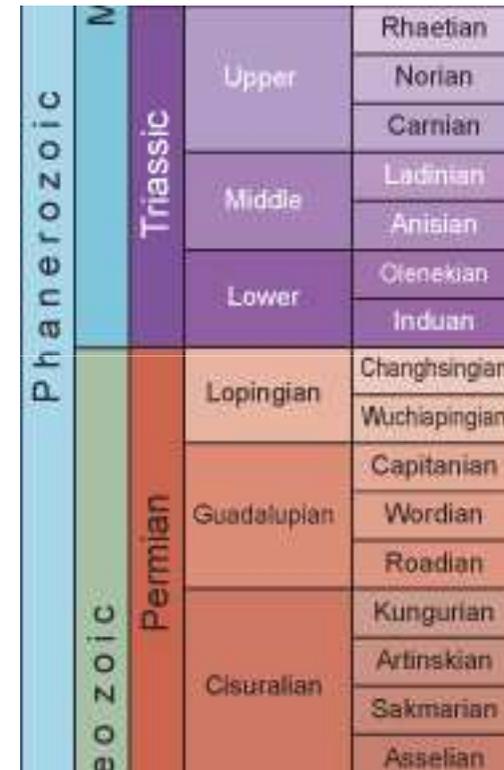


Exemple 2

Description des temps géologiques

Gestion de synonymies

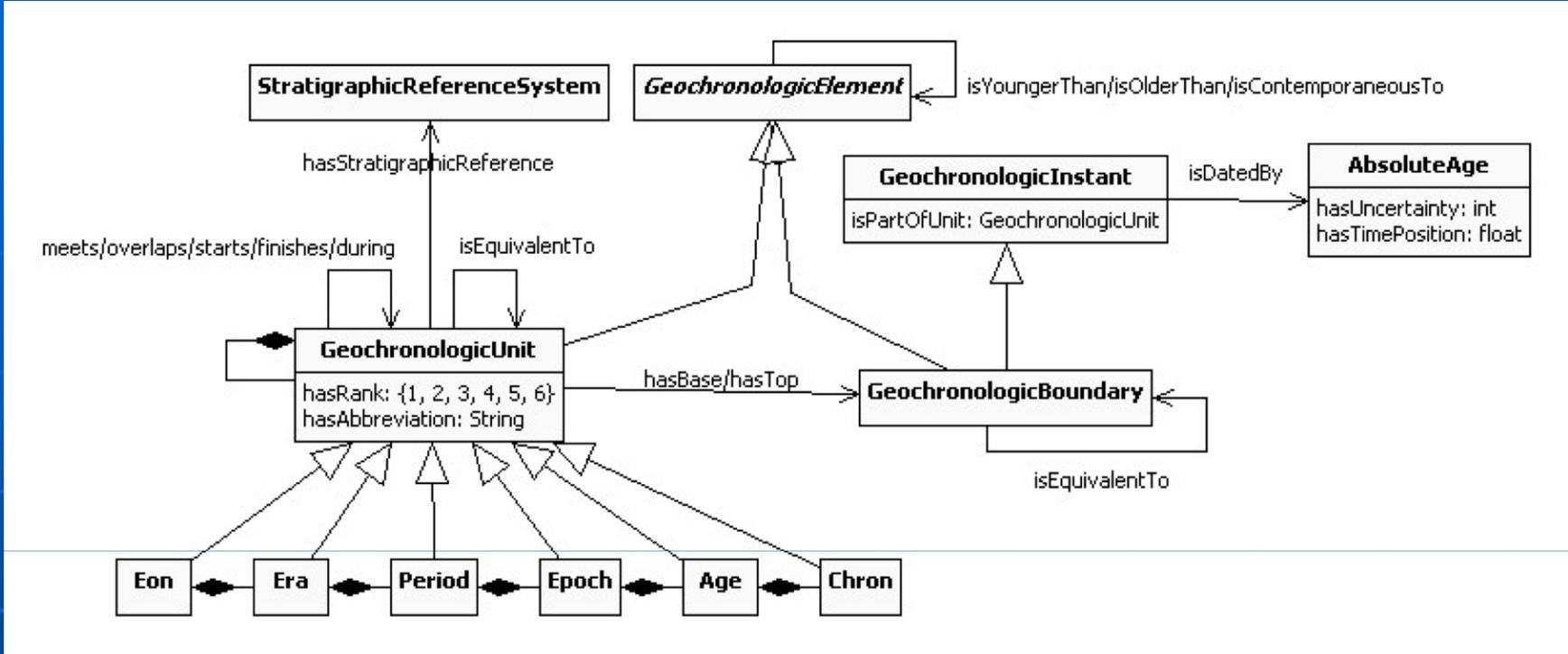
Base of	Mesozoic	Triassic	Lower Triassic	Indusian
Top of	Paleozoic	Permian	Lopingian	Changhsingian
	Rk2	Rk 3	Rk 4	Rk 5
	International Stratigraphic Scale			



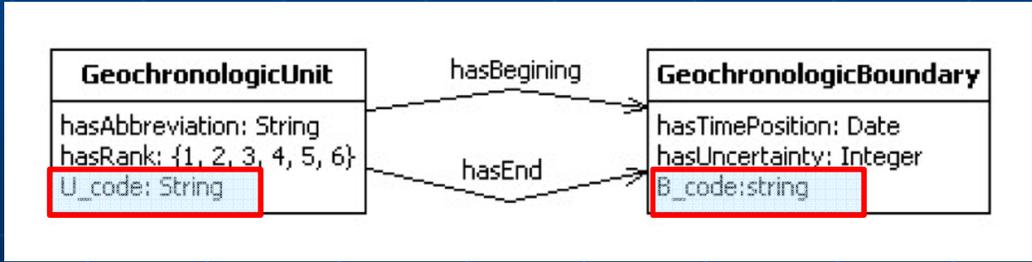
Problématique :

- Dans le cadre du projet e-Wok Hub, nous avons abordé le problème de l'analyse automatisée de **documents textes**.
- La méthodologie adoptée est celle d'une **annotation sémantique** reposant sur des **ontologies de domaine**.
- Il importe que cette méthodologie gère les **synonymies** telles que celles relatives aux limites de temps géologiques

1



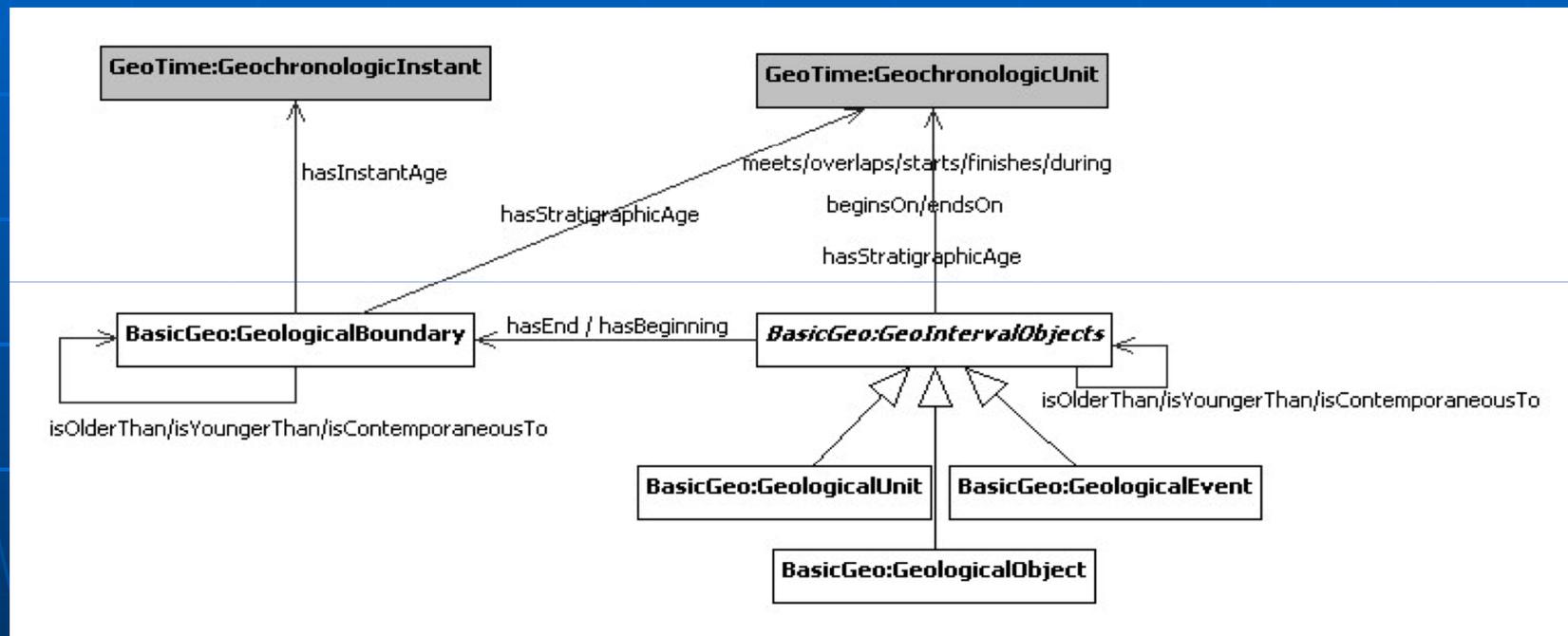
2



1- Ontologie des temps géologiques

2- Ajout d'une codification (U_code, B_code)

Ces ontologies doivent être complétées par une **ontologie de datation**, qui permet de **relier deux catégories de concepts** : les âges et les objets géologiques



La **formalisation** ainsi réalisée et la **codification** qui l'accompagnent permettent :

- d'attribuer un **âge** à chaque objet géologique (**unit** ou **boundary**)
- de **gérer** l'ensemble de **synonymies** relatives à cet âge.

The chart displays geological periods and stages. The vertical axis is labeled 'Phanerozoic' and 'eozoic'. The horizontal axis is labeled 'Triassic' and 'Permian'. The stages are listed in the following order from top to bottom: Rhaetian, Norian, Carnian, Ladinian, Anisian, Olenekian, Induan, Changhsingian, Wuchiapingian, Lopingian, Capitanian, Wordian, Roadian, Kungurian, Artinskian, Cisuralian, Sakmarian, and Asselian.

Phanerozoic	Triassic	Upper	Rhaetian	
			Norian	
			Carnian	
	Middle	Ladinian		
		Anisian		
	Lower	Olenekian		
		Induan		
	eozoic	Permian	Lopingian	Changhsingian
				Wuchiapingian
			Guadalupian	Capitanian
Wordian				
Roadian				
Cisuralian		Kungurian		
		Artinskian		
		Sakmarian		
Asselian				

Exemples :

- Une unité géologique d'âge « changhsingien » sera également reconnue unité du « Lopingien » et comme unité du « Permien ».
- La limite entre les étages « Indusien » et « Olénékien » a un âge trias inférieur (et... trias).
- Si, dans une collection des textes ayant fait l'objet d'une annotation sémantique, on recherche ceux qui font allusion à l'époque Trias inférieur, tout texte mentionnant les termes « Indusien » et « Olénékien » sera sélectionné.

Exemple 3: Description des données pétrologiques à l'aide d'une ontologie "sur mesure"

Les classifications des roches couramment utilisées sont **insatisfaisantes**:

- **Les classifications « scientifiques »** ont une valeur générale mais ne sont pas toujours opérationnelles, car on ne dispose pas des informations nécessaires pour les faire jouer (observations microscopiques, analyses géochimiques).
- **Les terminologies empiriques** utilisées pour la description des roches observées sur des affleurements ou des carottes de sondage fournissent des informations pertinentes dans un contexte défini mais sont **difficilement « exportables »**.

Dans le cadre du projet e-Wok Hub, nous avons entrepris la construction d'une ontologie « sur mesure » en vue de la prise en compte des données pétrologiques extraites de documents textuels.

Méthodologie :

- sélection d'une série de **textes de référence** (articles scientifiques) contenant le vocabulaire du domaine,
- **extraction manuelle par les experts métiers** des termes intéressants,
- **classification** de ces termes et élaboration d'un **projet d'ontologie** en collaboration avec des ingénieurs de la connaissance,
- **harmonisation** du projet d'ontologie avec la ou les ontologies existantes dans le domaine considéré (en l'occurrence la référence **GeoSciML**).

anhydritic (anhydritic clay)	magnésien (calcite faiblement magnésienne)
anhydritic (anhydritic mudstone)	magnésien (calcite hautement magnésienne)
argillites	mudstone
arkosic	oolitic grainstone
breccia	oolitic limestone
calcaire(s)	quartzite
calcaire grumeleux	quartzitic
chalk	salin (couches salines)
charbon	sand sables
Charbon (couche(s) de charbon)	sandy
conglomerate conglomérat	sandstone(s) grès
conglomeratic	sandstone (calcareous sandstone)
dolocrete	sanstone (conglomeratic sandstone)
dolomie	sandstone (gritty sandstone)
evaporite	sandy limestone
limestone(s)	silty
lydite	shales
marl marme(s)	magnésien (calcite faiblement magnésienne)
marly	magnésien (calcite hautement magnésienne)

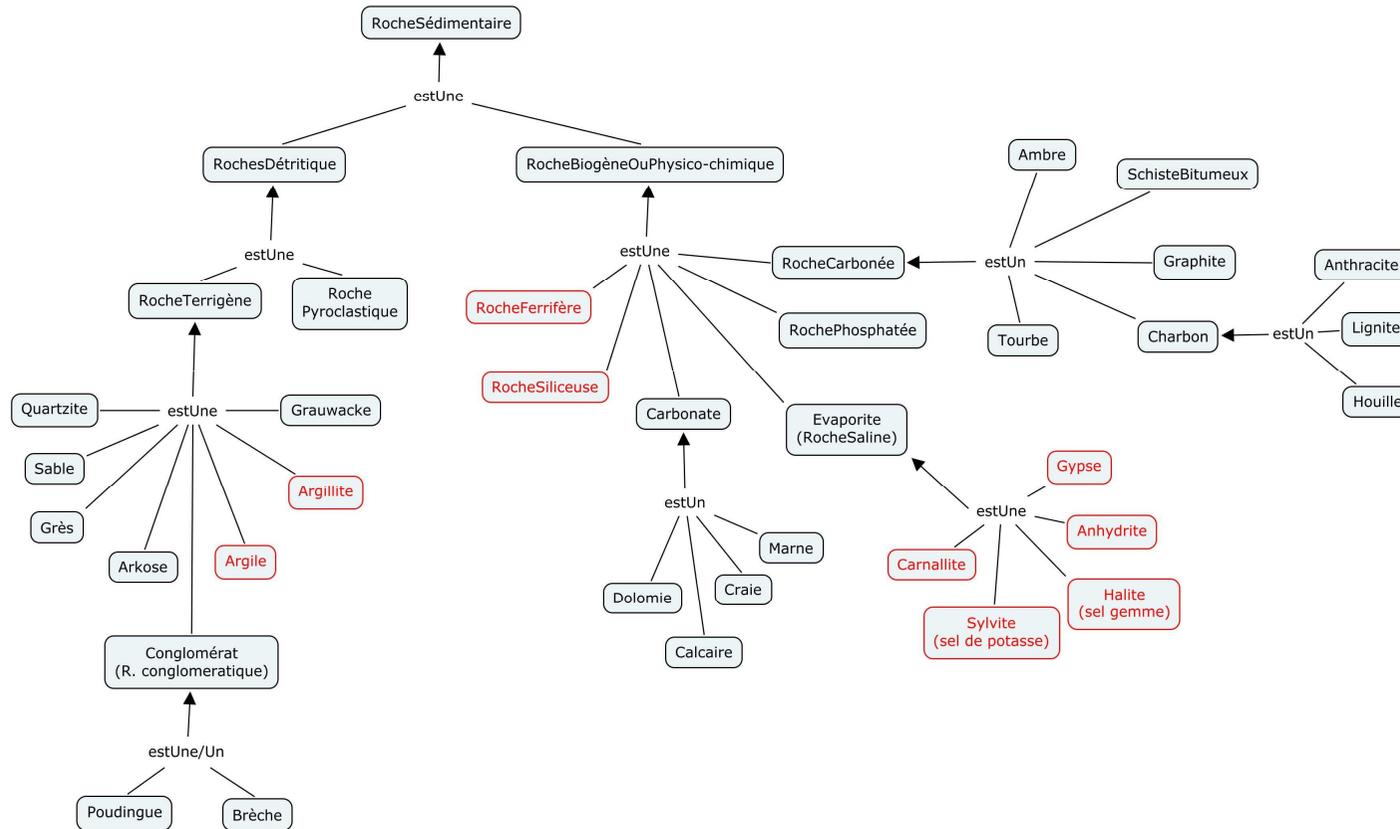
Exemples de termes extraits

Ces termes doivent être triés selon

- la langue (français / anglais)
- la forme grammaticale :
substantif, adjectif,
substantif + adjectif

en vue de définir des concepts et des instances de ces concepts dans l'ontologie à construire.

Ontologie de domaine réalisée



Fichier Édition Affichage Historique Marque-pages Outils ?

http://localhost/web/guest/ewc

Liferay Tomcat JDK-6 Tomcat JDK-5

E-Wok Client - Liferay

Welcome Géologue EWC!

e-WOK_HUB

Login E-Wok Client

Navigation portlet

Recherche

Ontologie

Cartographie

Echelle des temps

Résultat

Visualisation

Request

Executer du SPARQL

PARAMETRES :
Limestone

Requêtes disponibles

Quels sont les documents qui parlent de Limestone ?

INRIA Search Service Rechercher Effacer

Project

ANR-DEMO

Proj

- ANR-DEMO
 - Propriétés
 - Requêtes
 - surveillance
 - Classeur
 - vrac
 - Indre et Blois
 - tithonien
 - marnes OSrea
 - trias du bassin de paris
 - req-limestone
 - Zones Géographiques
 - Indre et Blois

WorkBook

Ontology

Categories	Concepts
Datation	<input type="checkbox"/> Terrigene
Localization	<input type="checkbox"/> Claystone
Geological processes	<input type="checkbox"/> Volcanic rock
Geological structures	<input type="checkbox"/> Chalk
Hydrogeology	<input type="checkbox"/> Detrital rock
Lithology	<input type="checkbox"/> Conglomeratic
Mineralogy	<input type="checkbox"/> Sand
Paleogeography	<input type="checkbox"/> Houille
Geological substances	<input type="checkbox"/> Carbonaceous rock
Basic geology	<input type="checkbox"/> Halite
	<input type="checkbox"/> Sylvite
	<input type="checkbox"/> Gypsum
	<input type="checkbox"/> Anhydrite
	<input type="checkbox"/> Siliceous rock
	<input type="checkbox"/> Sandstone
	<input type="checkbox"/> Sedimentary rock
	<input type="checkbox"/> Carbonate

Terminé

Fichier Edition Affichage Historique Marque-pages Outils ?

http://localhost/web/guest/ewc

Liferay Tomcat JDK-6 Tomcat JDK-5

E-Wok Client - Liferay

Navigation portlet

Recherche

Ontologie

Cartographie

Echelle des temps

Résultat

Visualisation

Project

ANR-DEMO

Projet

- ANR-DEMO
 - Propriétés
 - Requêtes
 - surveillance
 - classueur
 - vrac
 - indre et Blois
 - tithonien
 - marnes OSirea
 - trias du bassin de paris
 - req-limestone
 - Zones Géographiques
 - Indre et Blois

WorkBook

Icons representing different workbooks or documents.

Request

Executer du SPARQL

Quels sont les documents qui parlent de Limestone, Indre et Blois, Tithonien ?

PARAMETRES :

Limestone Indre et Blois Tithonien

Requêtes disponibles

Quels sont les documents qui parlent de Limestone, Indre et Blois, Tithonien ?

INRIA Search Service Rechercher Effacer

Result

Type	Titre des documents
	RP-53760-FR_2005_nappes-jurassiques-Cher-Indre.pdf weblab://myrepository/doc_090926_190858
	RP-51887-FR_2002_nappes-jurassiques-Cher-Indre.pdf weblab://myrepository/doc_090921_075158
	2006_Le-Hindre_Natural-Analogue-of-CO2-trapping.pdf weblab://myrepository/doc_090920_222934
	2005_Lavastre_Chlorine-transfer-out-of-a-very-low-permeability-clay-sequence-(Paris-Basin,-France)-35Cl-and-37Cl-evid weblab://D/PDF_ewok/2005_Lavastre_Chlorine-transfer-out-of-a-very-low-permeability-clay-sequence-(Paris-Basin,-France)-35Cl-and-37Cl-evid.pdf

Texte annoté:

Cette étude, financée par le ministère de l'Environnement, est programmée en deux phases réparties sur 2002 et 2003 dans le cadre des activités de service public du BRGM (opération SP 02 EAU 611) :

- phase 1 : état des connaissances et étude hydrogéologique des **réservoirs du Jurassique**, qui fait l'objet de ce rapport ;

- phase 2 : qualité des **eaux** et bilan hydraulique des **bassins**.

En partant du nord, quatre ensembles **carbonatés** ont été individualisés qui contiennent de **eau** presque exclusivement dans les zones **fracturées** :

- le **Tithonien**, au sommet du **Jurassique**, barre de **calcaire compact** qui n'affleure que dans la moitié orientale de la région étudiée et qui ne contient qu'exceptionnellement de **l'eau** sur des structures anticlinales dans le **Cher**. Dans le Pays-Fort, les nombreuses sources ne sont que des exutoires du **réservoir** sableux sus-jacent du **Crétacé** ;

- un ensemble complexe, daté du Kimméridgien inférieur à l'**Oxfordien moyen** (dans le **Jurassique** supérieur) affleure très largement. Les ressources sont localisées essentiellement dans la partie supérieure de cette **formation (Calcaire de Bourges,**

Calcaire de Morthomiers, Calcaire de Von, ...) ; les capacités se réduisent très fortement vers la base (**Calcaires** lités inférieurs, ...). Seule la frange supérieure peut fournir de **l'eau** ; cet **aquifère** a donc une capacité très faible, d'où une grande sensibilité aux épisodes de sécheresse. En domaine captif, ces **calcaires** sont **compacts** ;

- le **Jurassique moyen** (ou **Dogger**) renferme 2 à 3 barres **calcaires** qui affleurent en une bande étroite au sud des départements du **Cher** et de l'**Indre**. Elles se développent particulièrement à l'ouest du **Cher**, avec un banc oolithique poreux où la ressource est continue et peut être exploitée en profondeur. À l'est de ce cours d'**eau**, le **calcaire** est plus argileux ; les recherches en domaine profond se sont soldées par des échecs ;

- le **Jurassique** inférieur (ou **Lias**) comporte à sa partie inférieure un fin banc **calcaire compact**, très peu exploité dans sa partie affleurante, et très mal connu.

L

Étude des nappes du **Jurassique** dans les départements du **Cher** et de l'**Indre**
4 BRGM/RP-51887-FR

Exploitation des ontologies : passage des concepts aux instances

- Ces résultats ne supposent pas seulement que soient définis des **outils de description formelle de la connaissance** (modèles conceptuels et ontologies) car les utilisateurs sont au final intéressés non pas par les concepts mais par les instances de ces concepts.
- Il faut donc adjoindre aux ontologies définies, des **outils de management de la connaissance**:
 - base de données à base ontologique (exemple: **OntoDB**)
 - langage d'interrogation (exemple: **OntoQL**)

Les partenaires métiers du projet n'ont pas à connaître le détail de ces tâches d'ingénierie logicielle mais ils doivent savoir qu'elles sont fortement consommatrices de temps de développement

En conclusion:

Même dans un domaine tel que la géologie, dans lequel les connaissances sont fortement structurées, il est nécessaire d'adjoindre au vocabulaire technique des **outils spécifique de description et de gestion de la connaissance.**

Il devient alors possible de :

- mettre en oeuvre des méthodes d'**annotation sémantique** permettant l'analyse de documents textes,
- **associer différentes catégories d'objets** (par exemple des objets et des âges géologiques),
- appliquer de **règles** sur des objets_et donc d'effectuer de manière automatique des raisonnements en faisant appel à des **moteurs d'inférences**

REPRÉSENTER LISIBLEMENT LES FORMALISMES

Partenaires associés dans un projet d'ingénierie à composante informatique:

Géologues

Informaticiens

Utilisateurs
finaux

Experts
métier

Ingénieurs de
la connaissance

Développeurs

Partenaires associés dans un projet d'ingénierie à composante informatique:

Géologues



- Définition des besoins
- Contrôle des résultats (étape par étape)

Partenaires associés dans un projet d'ingénierie à composante informatique:

Géologues

Informaticiens



- Définition des formalismes (ontologies)

Partenaires associés dans un projet d'ingénierie à composante informatique:

Informaticiens



- Définition de l'architecture et des procédures

Imperfections des représentations graphiques

C map

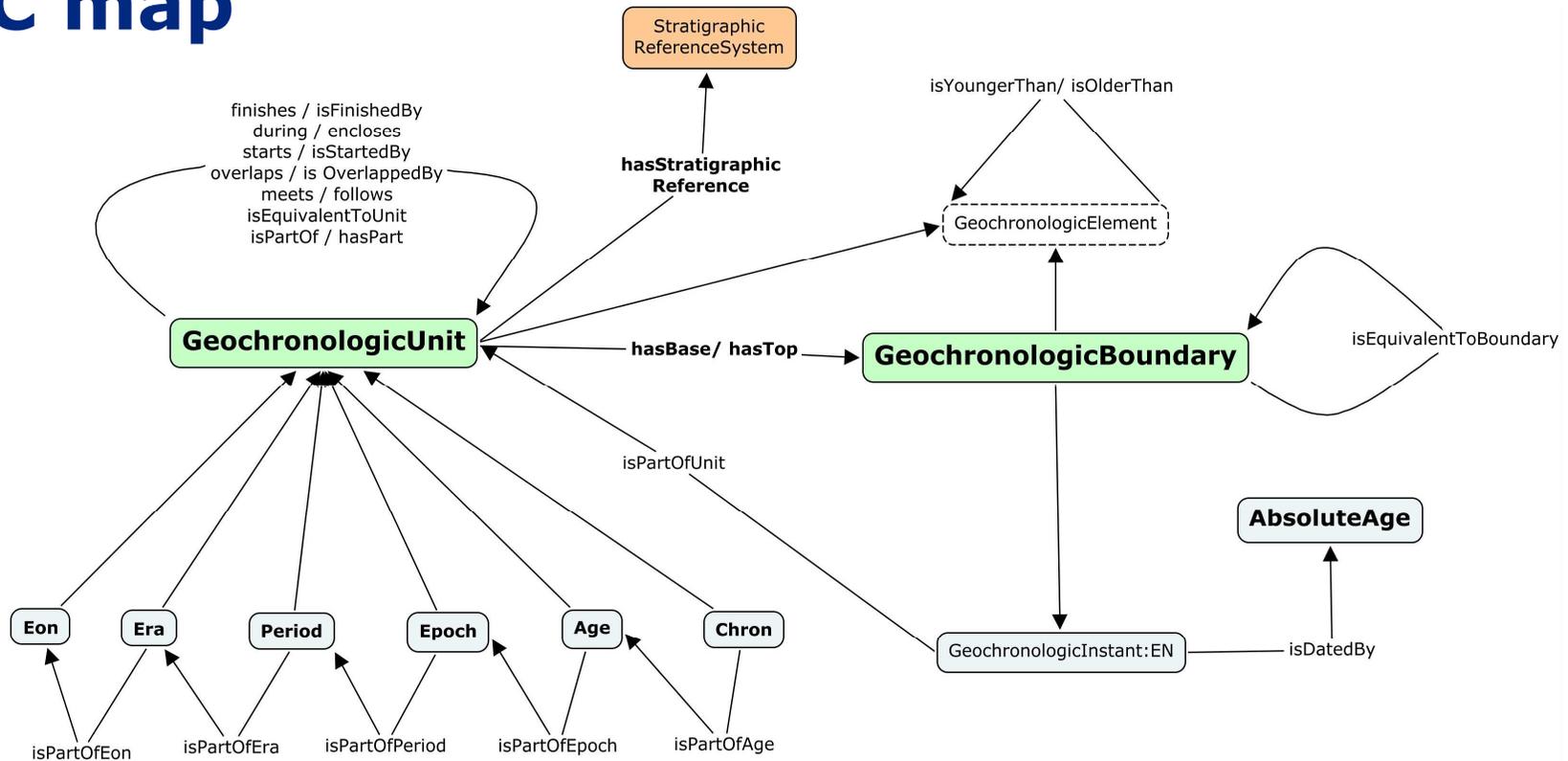
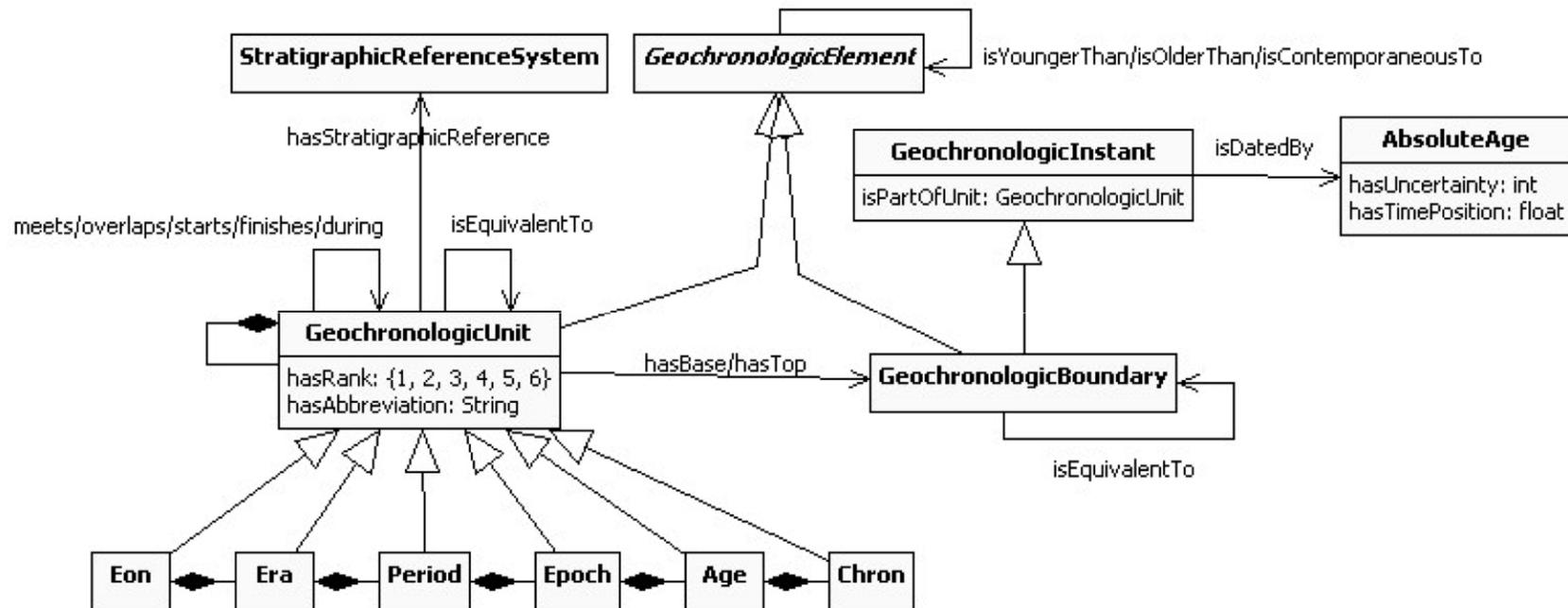
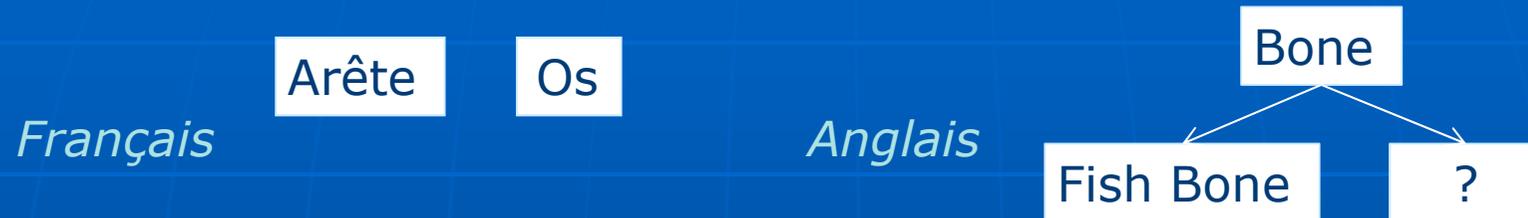


Schéma UML



Difficultés conceptuelles au niveau du langage:

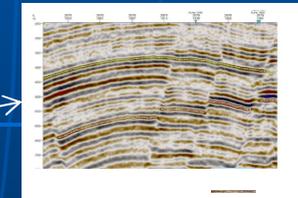
- **Polysémies et incomplétudes**

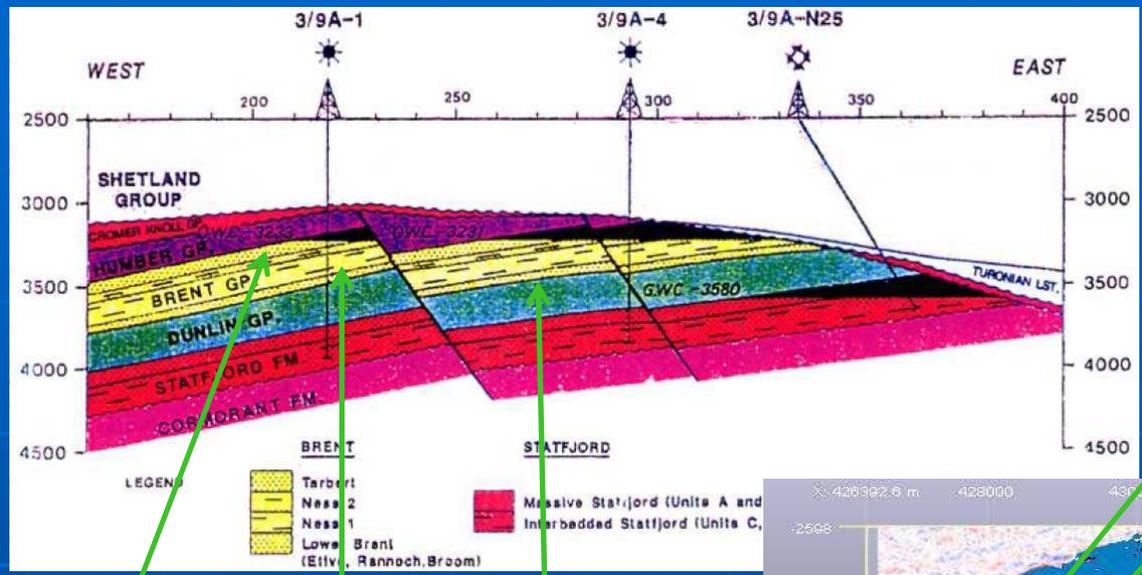


- **Ambiguïté de désignation des objets**

Un même nom peut désigner les instances de 2 objets différents, par exemple:

- un horizon sismique
- un horizon géologique





Top Brent

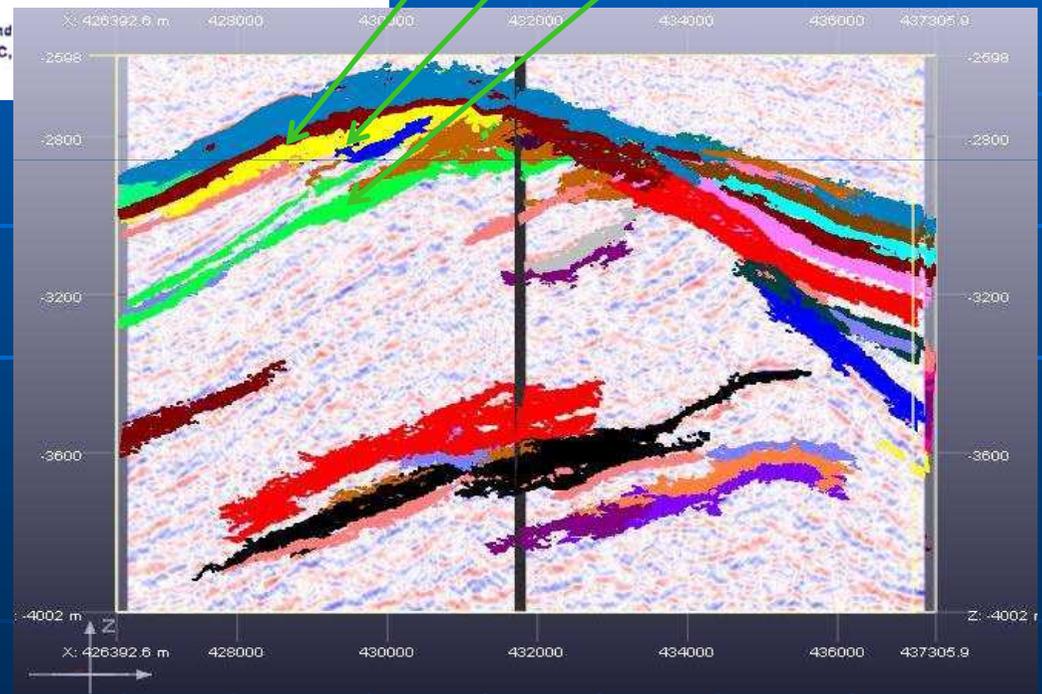
Top Etive

Top Dunlin

Top Brent

Top Etive

Top Dunlin



Requête:

*Parmi tous les horizons identifiés grâce à l'interprétation sismique, extraire ceux qui sont plus jeunes que **Top Dunlin**.*

- S'agissant d'identifier des **horizons géologiques** plus récents que Top Dunlin, la question est formulée d' **âges relatifs** (plus récent)
- Grâce à un mapping réalisé entre 2 ontologies relatives aux objets géologiques d'une part et aux horizons sismiques d'autre part et grâce à une méthodologie d'annotation sémantique (Mastella, 2010), la question peut être traduite automatiquement par le logiciel dans le domaine de la **sismique** où elle s'exprime en termes de **position relative** (plus haut).

Réponse automatique :

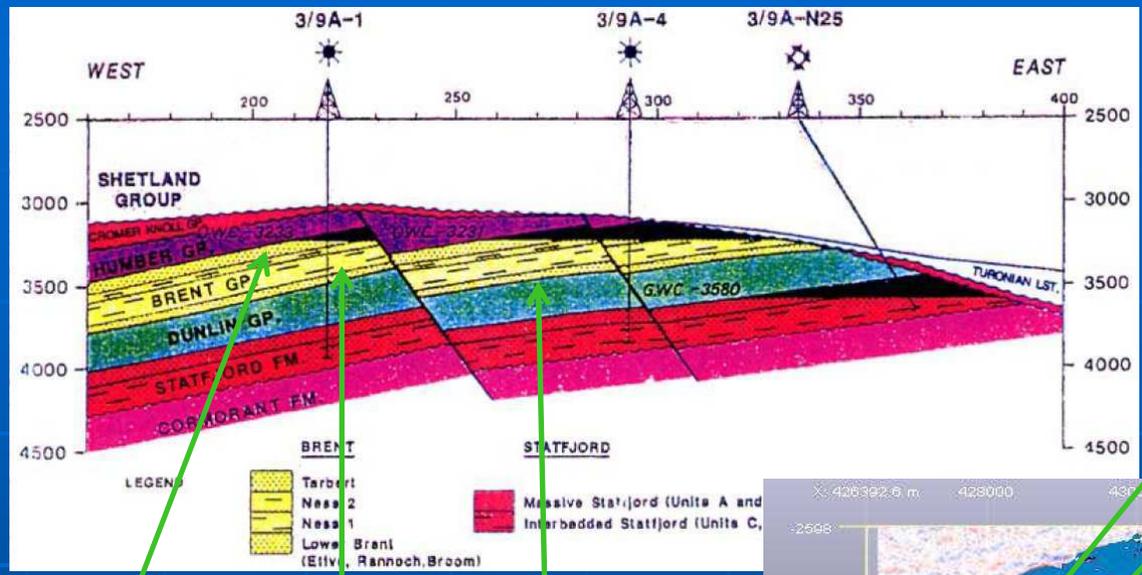
The screenshot shows the OntoQL WorkBench interface. The main window displays a SQL query:

```
SELECT DISTINCT younger.name, older.name
FROM StratigraphicBoundary AS younger,
StratigraphicBoundary AS older,
unnest (younger.isYoungerThan) as h
WHERE older.oid IN (h.oid)
AND older.name = 'Top_Dunlin'
WITH APOSTERIORI
```

Below the query, the Session Command History shows the executed command: `select distinct younger.name, older.na...`. The results are displayed in a table with two columns, both labeled 'name':

name	name
BCU	Top_...
Top_Brent	Top_Dunlin
Top_Etive	Top_Dunlin
Top_Ness1	Top_Dunlin
horizon_12	Top_Dunlin
horizon_121	Top_Dunlin
horizon_122	Top_Dunlin

Two rows are highlighted with red boxes and labeled with red text: 'Top Brent' and 'Top Etive'. The status bar at the bottom indicates: **22 rows returned in 4359.0 milliseconds**.



Top Brent

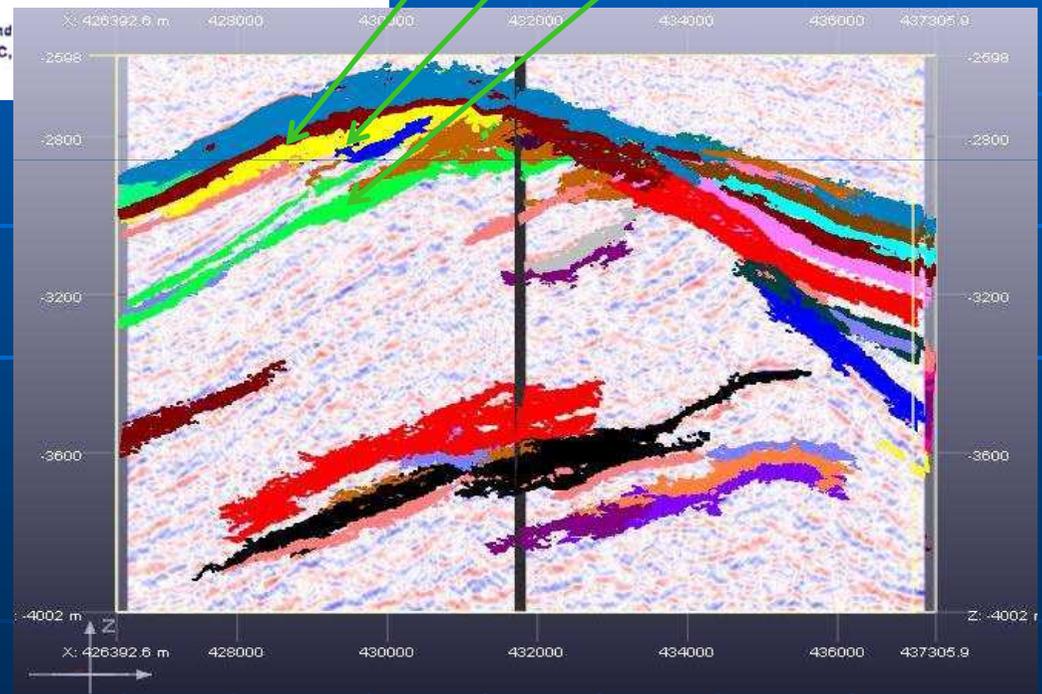
Top Etive

Top Dunlin

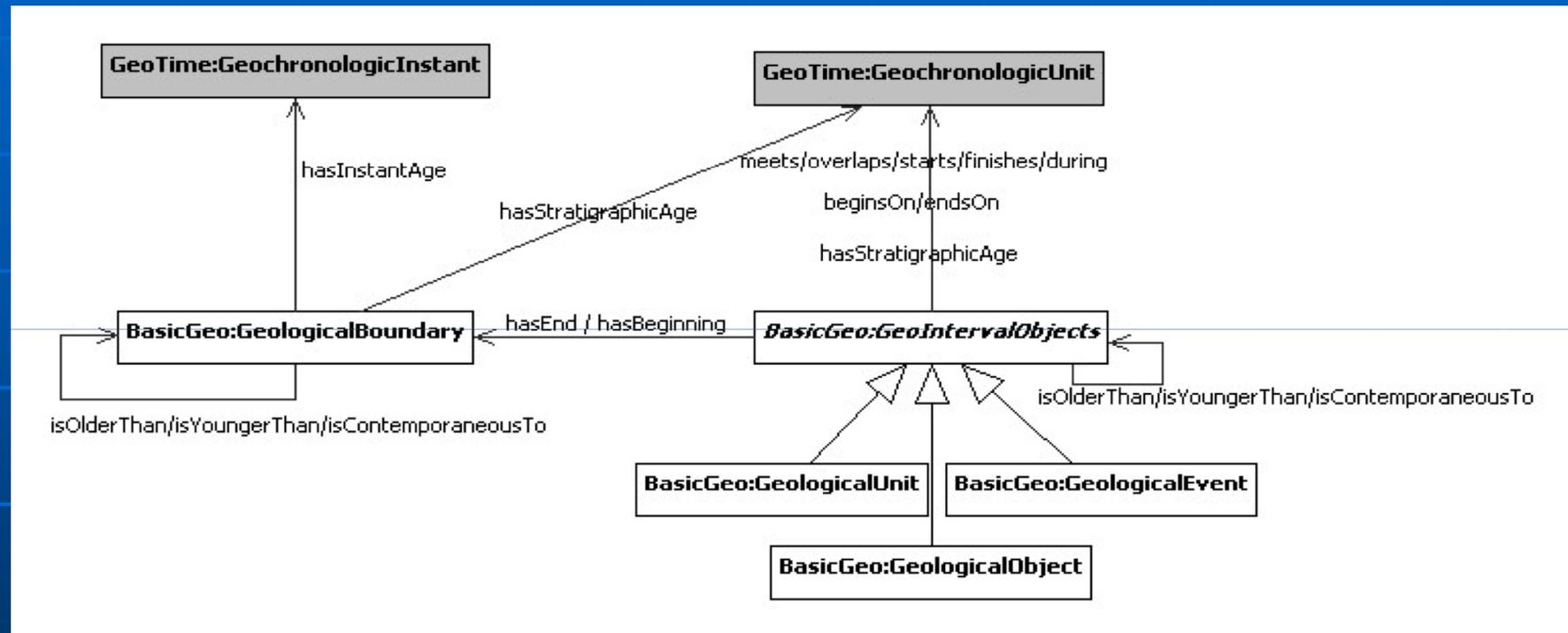
Top Brent

Top Etive

Top Dunlin

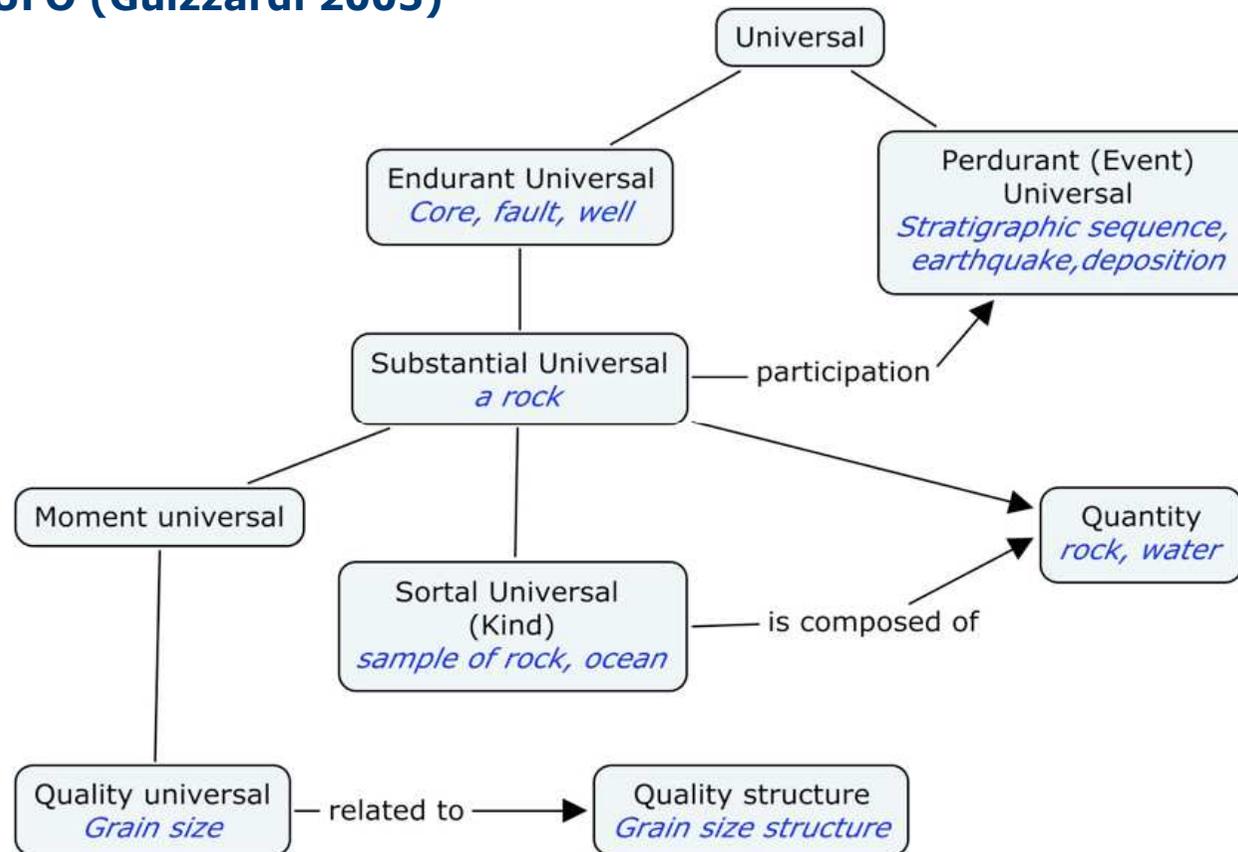


Pour réduire les risques d'ambiguïté, on peut introduire des termes spécialisés, mais ils ne sont pas toujours parlants



Exemple de l'« United Foundational Ontology » :

UFO (Guizzardi 2005)



RECOMMANDATIONS & CONCLUSION

Chacun doit faire son métier :

- Les **experts métier** sont les seuls compétents pour juger de l'**opérabilité des ontologies construites**, eu égard aux objectifs qu'ils se fixent.
- Les **ingénieurs de la connaissance** seuls compétents pour décider de l'**architecture des ontologies** elles-mêmes, dès lors que celles-ci répondent aux spécifications posées par les experts métier.
- Les **contraintes** qui résultent de cette division du travail doivent être comprises et acceptées par les **utilisateurs finaux** d'une part et **développeurs** d'autre part

- Le dialogue experts-métier / ingénieurs de la connaissance exige **que chacun se mette « dans la peau de l'autre »**
- Il faut **anticiper les difficultés à chaque stade du projet**, en faisant aussi souvent que nécessaire des retours vers les experts métiers donneurs d'ordre à l'aide de **tests** judicieusement choisis.

Le dialogue indispensable entre géologues et informaticiens s'avère riche mais long et coûteux en temps. Ces délais longs doivent être pris en compte dans la planification du projet.