

# Géomodélisation et Informatique : Formalisation des représentations et représentation des formalismes

Michel Perrin  
Ecole des Mines de Paris  
60,Bd St Michel  
75006 Paris  
michel.perrin@ensmp.fr

Jean-François Rainaud  
Institut Français du Pétrole  
4 avenue de Bois-Préau  
92500 Rueil Malmaison  
01 47 52 57 25  
j-francois.rainaud@ifp.fr

Laura S. Mastella  
Petrobras  
330, Republica do Chile  
20031 Rio de Janeiro Brésil  
laura.mastella@  
petrobras.com.br

## RESUME

Dans les projets d'ingénierie associant des experts métier et des informaticiens, ces deux catégories d'acteurs doivent manipuler un même corps de connaissances qu'ils formalisent selon des modalités différentes. Considérant le cas particulier de la modélisation géologique, l'article examine les problèmes posés par ces différences de représentation et formule quelques recommandations.

## Mots clés

Formalisation de la connaissance, ontologies, modélisation géologique.

## 1. INTRODUCTION

Nous examinerons dans cet article, 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 (ingénieurs de la connaissance et développeurs).

Du point de vue de la représentation des objets, il s'agit de trouver des formalisations qui répondent aux exigences élevées des constructeurs de logiciel tout en respectant, autant qu'il est possible, les contenus sémantiques définis par les experts métier. Ce dilemme classique entre opérabilité et expressivité se ramène à un problème de *formalisation des représentations*. Dans la première partie de l'exposé, nous monterons en nous appuyant sur 3 exemples concrets que des solutions peuvent être trouvées au travers de la construction d'*ontologies de domaine* adéquates.

Du point de vue des acteurs, il convient de faciliter un dialogue que les formalismes employés rendent souvent difficile en raison de l'opacité ou de l'ambiguïté des vocabulaires et des

représentations graphiques qui leur sont attachés. Il s'agit donc de réfléchir à de possibles améliorations de la *représentation des formalismes*. Nous essaierons dans la deuxième partie de l'exposé de cerner certains des problèmes posés et de suggérer quelques pistes de solutions.

L'examen que nous ferons des deux catégories de problèmes mentionnés s'appuie sur l'expérience acquise au travers des études que nous avons menées depuis de nombreuses années sur l'assistance informatique à la construction de modèles géologiques 3D (*Géomodélisation*). Nous ferons par ailleurs allusion à notre expérience particulière au sein du projet ANR e-Wok Hub, qui a réuni dans les années 2006-2009, plusieurs équipes d'informaticiens et de chercheurs en Géosciences en vue de la réalisation d'une plateforme de service pour l'étude des réservoirs géologiques. (Projet e-Wok hub :

<http://www.inria.fr/sophia/edelweiss/projects/ewok/>)

## 2. FORMALISATION DES REPRESENTATIONS METIER

La formalisation des représentations métier est une exigence des informaticiens.

Lors de la construction de géomodèles, les experts métier basent leurs descriptions des connaissances sur diverses représentations couramment utilisées en géologie. Nous allons montrer au travers de 3 exemples, que ces représentations ne répondent pas complètement aux exigences posées par les informaticiens pour la description et le management de la connaissance scientifique sous-jacente. Dans chacun de ces cas, il convient donc de pousser plus loin la formalisation grâce à la construction de modèles conceptuels et d'ontologies de domaine.

### 2.1. Relations entre unités géologiques

L'agencement des terrains sédimentaires, dans lesquels sont localisés la plupart des réservoirs géologiques, est décrit par les géologues à l'aide du *modèle stratigraphique*, conformément auquel des unités géologiques de plus en plus récentes se superposent les unes aux autres. Par ailleurs, un assemblage d'unités sédimentaires est susceptible d'être recoupé par des failles plus récentes que lui.

Les relations topologiques entre les différentes surfaces géologiques sont déterminées par leurs caractéristiques géologiques (limites d'unités, failles) et par leurs âges relatifs. Ces relations entre surfaces se traduisent dans le tracé des cartes géologiques. A la lecture d'une carte, un géologue est ainsi capable de déterminer les caractéristiques des surfaces représentées et leurs relations d'âge.

Les relations topologiques entre surfaces géologiques doivent également être traduites dans les géomodèles 3D. Dans le cas d'une construction de géomodèle assistée par ordinateur, il faut ainsi être en mesure de préciser au cas par cas quelle surface interrompt telle autre. Pour cela, un modèle conceptuel doit être construit qui distingue les différentes catégories de surfaces (limites d'unités, failles etc.) et attribue à chacune d'elles des caractéristiques permettant de préciser ses relations topologiques avec les autres surfaces du modèle. La figure 1 montre le modèle conceptuel que nous avons réalisé, il y a quelques années, en vue de permettre l'assemblage automatisé des surfaces géologiques dans un géomodèle [1]

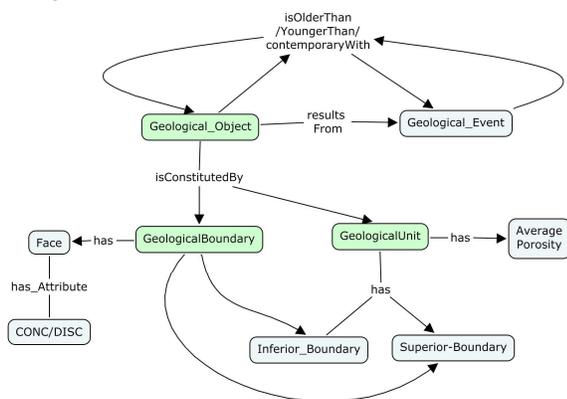


Figure 1. Modèle conceptuel relatif aux objets géologiques

## 2.2. Description des temps géologiques

Les échelles de temps couramment utilisées en géologie sont basées sur une chronologie événementielle dans laquelle les différents intervalles de temps sont chronologiquement ordonnés et hiérarchiquement organisés par rangs : *éon* (rang 1), *ère* (rang 2), *période* (rang 3) etc... Il en résulte qu'une limite correspondant à un même instant dans l'échelle des temps géologique peut être nommée de différentes manières selon qu'on se réfère à des intervalles de temps de rang plus ou moins élevé. La figure 2 illustre les différents noms sous lesquels peut être désignée une même limite de temps (en l'occurrence le début de la période Trias) dans l'échelle stratigraphique internationale.

Rang	Début du :	Fin du :
2	Mésozoïque	Paléozoïque
3	Trias	Permien
4	Trias inférieur	Lopingien
5	Indusien	Changshingien

Figure 2. Exemple de synonymies.

Les différents noms indiquent tous une seule et même limite correspondant à la base du Trias. (Données extraites de [2])

Lors d'une recherche documentaire, une telle limite de temps devra pouvoir être identifiée comme unique quel que soit le terme sous lequel elle est désignée. Dans le cadre du projet e-Wok Hub nous avons montré que ceci peut être réalisé par le biais d'un outil de requête sémantique tel que le logiciel CORESE [3]. Mais, dans ce cas encore, le formalisme utilisé par les géologues n'est pas en lui-même suffisant. Il faut en effet adjoindre aux échelles de temps géologiques une ontologie permettant de les décrire et un système de codification [4].

## 2.3. Description des données pétrologiques à l'aide d'une ontologie "sur mesure"

La nature des roches présentes dans le sous-sol peut être décrite à l'aide de classifications scientifiques ou empiriques. Les premières ont une valeur générale mais ne sont pas toujours opérationnelles, car il arrive qu'on ne dispose pas des informations nécessaires pour les faire jouer (observations microscopiques, analyses géochimiques). Les géologues utilisent parallèlement diverses terminologies pour la description des roches observées sur des affleurements ou des carottes de sondage. Ces terminologies sont basées sur des critères d'observation plus grossiers : recensement des éléments de la roche visible à l'œil nu ou à la loupe, texture, contenu en macrofossiles... Elles fournissent des informations, qui sont pertinentes dans un contexte défini mais difficilement « exportables ». Compte tenu de tout cela, on trouve dans la littérature géologique relative à une région donnée (articles, thèses, rapports divers), des termes d'origine variés, ayant un contenu sémantique hétérogènes, à partir desquels il est difficile de tirer une information exploitable par des moyens informatiques.

Dans le cadre d'un projet ciblé mettant en œuvre l'informatique, le recours à des ontologies constitue une solution intéressante, qui peut permettre de résoudre au moins en partie ce problème. Les ontologies présentent en effet plusieurs types d'avantages :

- elles combinent les avantages des classifications scientifiques et empiriques dans la mesure où elles visent à formaliser des savoirs pratiques,
- une ontologie est construite "sur mesure" en vue d'un but défini, sachant que « *there is no one correct way to model a domain but that there are always viable alternatives* » [5]
- il existe une certaine souplesse relativement au niveau de formalisation réalisé par telle ou telle ontologie depuis des « catalogues » jusqu'à des ontologies à haut niveau de structuration [6]

Dans le cadre du projet e-Wok Hub, nous avons fait l'expérience de la construction d'une ontologie « sur mesure » en vue de la prise en compte des données pétrologiques extraites de documents textuels. Celle-ci a été élaborée selon la méthodologie suivante :

- 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 ce la connaissance,
- harmonisation du projet d'ontologie avec la ou les ontologies existantes dans le domaine considéré (dans notre cas le modèle GeoSciML (<http://www.geosciml.org/>), qui a vocation à constituer à plus ou moins court terme une norme de fait).

L'ontologie ainsi constituée peut alors devenir le support à une analyse de textes basée sur une annotation sémantique informatisée [7].

## 2.4. Problèmes liés à l'exploitation des ontologies : passage des concepts aux instances

Si la création d'ontologies permet de formaliser un savoir en définissant des concepts et en précisant leur organisation logique et leurs relations avec un ensemble de propriétés, elle ne constitue pas à elle seule une solution complète pour la gestion des connaissances requises pour la modélisation géologique. En effet, les utilisateurs finaux d'outils logiciels basés sur la connaissance sont intéressés non pas par les concepts ou par les propriétés figurant dans l'ontologie mais par des objets concrets qui sont des instances des concepts décrits et par les valeurs que prennent les propriétés attachées à ces concepts.

En faisant appel à une base de données à base ontologique de type *OntoDB* accédée par l'intermédiaire d'un langage d'interrogation tel qu'*OntoQL* [8], il est possible de permettre aux utilisateurs finaux :

- d'entrer des données en attachant des instances et des valeurs aux concepts et propriétés définis dans une ontologie donnée,
- de récupérer des données de sortie, qui résulteront de la mise en œuvre des relations exprimées dans l'ontologie et, le cas échéant, de règles fondées sur cette même ontologie.

Ainsi le développement d'ontologies doit être complété par la mise en place

- d'outils de type *OntoDB* et *OntoQL*, qui permettent d'exploiter concrètement les formalismes définis,
- d'interfaces conviviales permettant aux utilisateurs d'entrer dans le système des instances et des valeurs de propriétés et de récupérer des résultats de requêtes.

Les partenaires métiers du projet n'ont pas à connaître le détail de ces tâches d'ingénierie logicielle. Il est bon qu'ils sachent cependant que cette étape est nécessaire et qu'elle peut être, en termes de développement, fortement consommatrice de temps.

## 2.5. En conclusion

Nous avons montré au travers des exemples présentés ci-dessus que dans un domaine d'intérêt à fort contenu scientifique tel que la géologie, il convient d'adjoindre au vocabulaire technique et aux connaissances pratiques des scientifiques et des experts, des outils qui permettent d'exploiter leur savoir à l'aide de moyens informatiques. Il convient ainsi de construire des formalisations spécifiques (modèles conceptuels ou ontologies de domaine) ainsi que des outils adaptés de gestion de la connaissance. Il devient alors possible :

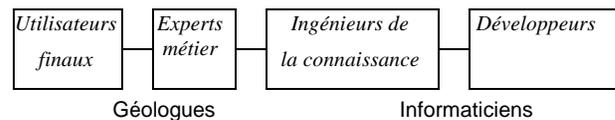
- de mettre en œuvre des méthodes d'annotation sémantique ouvrant la voie à l'analyse de documents textes,
- d'appliquer de règles sur des objets et donc d'effectuer de manière automatique des raisonnements en faisant appel à des moteurs d'inférences,
- d'associer différentes catégories d'objets, par exemple des objets et des âges géologiques.

## 3. REPRESENTER LISIBLEMENT LES FORMALISMES

Comprendre la signification des formalismes utilisés par les informaticiens pour représenter les connaissances relatives à leur domaine afin de pouvoir en contrôler la pertinence, est une nécessité pour les géologues. C'est cette deuxième catégorie de problèmes que nous allons maintenant examiner.

### 3.1. Exposé des problèmes : difficultés du dialogue Géologues – Informaticiens

Dans un projet d'ingénierie faisant appel des méthodes informatiques de gestion de la connaissance, les interlocuteurs concernés sont les suivants :



Les experts métiers et les ingénieurs de la connaissance sont ceux qui sont en charge d'élaborer les ontologies qui constitueront le support du dialogue homme <-> machine. Compte tenu des rôles qu'ils ont à jouer, on peut gratifier ces acteurs d'un minimum d'aptitudes au dialogue interdisciplinaire (curiosité intellectuelle, souci de comprendre des partenaires de culture différente). A l'inverse, ces aptitudes ne seront pas forcément celles des acteurs placés aux deux bouts de la chaîne : programmeurs d'une part et utilisateurs finaux d'autre part. En se plaçant dans le pire des cas, on peut donc considérer que les programmeurs n'auront à manipuler que des entités abstraites, dont ils pourront ignorer le contenu sémantique et que, pour leur part, les utilisateurs finaux se verront offerte la possibilité de travailler dans un environnement et avec un langage qui leur sont familiers, en s'abstrayant complètement, s'ils le souhaitent, du formalisme introduit par les ontologies.

### 3.2. Difficultés conceptuelles au niveau du langage

Même restreint aux seuls acteurs pour lequel il est nécessaire, le dialogue entre géologues et informaticiens est difficile car chacun ignore le langage de l'autre. A cela s'ajoutent des difficultés conceptuelles propres aux langages naturels d'une manière générale (le langage des géologues ne représentant qu'un cas particulier). Au nombre de celles-ci, on peut citer :

- les polysémies et les incomplétudes (en anglais, par exemple, le terme de « *bone* » englobe à la fois les concepts d'« *os* » et d'« *arête* ». Une arête pourra être désignée par le terme « *fish bone* » mais comment désigner un os sinon par le terme « *non fish bone* »? ).
- les ambiguïtés dues au fait que des objets différents sont parfois associés sous un même nom (par exemples, le mot « *roche* » désigne à la fois une catégorie de roche et un échantillon)

Pour surmonter certaines de ces difficultés, les ingénieurs de la connaissance sont amenés, à l'instar des philosophes, à introduire des éléments de vocabulaire spécialisés. Des outils ont également été mis au point pour tester la qualité logique des formalisations réalisées tels que par exemple *OntoClean* [9] ou *UFO* [10]. Une

difficulté tient cependant au fait que le vocabulaire utilisé par les ingénieurs de la connaissance est souvent peu parlant ou trompeur pour le profane (cf figure 3).

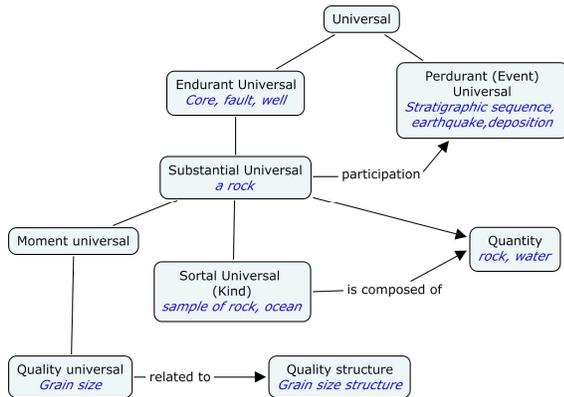


Figure 3 Propriétés définies dans UFO [10]

Les liens non renseignés correspondent à des relations *is\_ComposedOf*. Les informations figurant en italiques sont des exemples de concepts géologiques relatif aux différentes catégories de l'UFO.

### 3.3 Difficultés liées aux imperfections des représentations graphiques

Il n'existe pas de représentations des ontologies susceptibles de répondre complètement aux attentes des diverses catégories d'acteurs qui les utilisent. Le langage OWL (<http://www.w3.org/TR/owl-ref/>), qui est le plus utilisé par les informaticiens est réservé à eux seuls et est totalement illisible par des non spécialistes. Par ailleurs, aucune représentation graphique d'ontologies n'est complètement satisfaisante. Si l'on compare une représentation orientée « experts métiers » telle que les C-maps (<http://cmap.ihmc.us/>) que nous avons largement utilisée dans le projet e-Wok Hub et une représentation orientée « informaticiens » telle que les schémas UML, on constate que :

- *l'expressivité* est :

excellente dans le cas des C-maps, qui offrent la possibilité de jouer librement sur la disposition des objets, la taille, et la couleur des cartouches, le style des lignes de liaison, les polices de caractères,

moins bonne dans le cas des schémas UML : on peut certes jouer sur la plupart des paramètres précédents mais le style de chaque type d'objet (concept, propriété) est défini une fois pour toutes. Impossible donc pour les experts métiers de distinguer différents types de concepts plus ou moins importants à leurs yeux

- *l'opérabilité* est :

excellente dans le cas des schémas UML : tous les paramètres utiles à la caractérisation informatique des objets sont affichés et repérés par un style unique,

insuffisante dans le cas des C-maps, car seuls certains paramètres sont affichables.

## 4. RECOMMANDATIONS ET CONCLUSION

Dans le cas de projets d'ingénierie, une collaboration étroite et prolongée entre informaticiens (ingénieurs de la connaissance) et professionnels (experts métier) s'avère indispensable. Pour que celle-ci soit productive, il faut, à notre avis, veiller :

a/ à ce que chacun fasse son métier :

- Les experts métier doivent être considérés comme les seuls compétents pour juger de l'opérabilité des ontologies construites, eu égard aux objectifs qu'ils se fixent.
- Les ingénieurs de la connaissance sont, à l'inverse, seuls compétents pour décider de l'architecture des ontologies en question, dès lors que celles-ci répondent à l'ensemble des spécifications posées par les experts métier. De même, c'est à eux et à eux seuls qu'il revient de définir les modalités selon lesquelles les connaissances doivent être gérées (choix d'une base de données, d'un langage de requête, d'un moteur d'inférence).
- Par ailleurs, les contraintes qui résultent de cette division du travail doivent être comprises et acceptées par les acteurs situés aux deux bouts de la chaîne : utilisateurs métier d'une part et développeurs d'autre part, ce qui est parfois loin d'être facile.

b/ à ce que chacun se mette « dans la peau de l'autre » ; chacune des deux catégories d'acteurs (experts métiers et ingénieurs de la connaissance) doit expliquer à l'autre les exigences auxquelles ils doivent faire face et les contraintes que cela impose.

c/ à anticiper les difficultés afin ne pas faire des choix irréfléchis. Cet objectif est de loin celui qui est le plus difficile à mettre en œuvre. Dès lors qu'ils pensent avoir compris ce que les utilisateurs attendent du système, les constructeurs de logiciels seront toujours amenés à choisir les outils et les solutions qui sont les plus opérantes à leur niveau (pour minimiser le volume des données, pour assurer une bonne rapidité d'exécution). On court alors le risque que ces choix basés sur des critères informatiques, altèrent éventuellement les performances du système eu égard aux attentes des utilisateurs. Il est souhaitable d'anticiper dans toute la mesure du possible ces difficultés à chaque stade du projet, en faisant aussi souvent que nécessaire des retours vers les experts métiers donneurs d'ordre notamment par le moyen de tests judicieusement choisis. Un choix d'architecture logicielle ne peut être fait qu'après que les concepteurs auront envisagé en commun avec les experts métiers quelles peuvent en être toutes les conséquences possibles. Cela n'évitera pas toujours le risque de mauvaises surprises mais le réduira sensiblement.

Il faut enfin avoir conscience que la collaboration entre experts métier et ingénieurs de la connaissance est un processus itératif, coûteux en temps. Ainsi, dans le cadre du projet e-Wok Hub la finalisation d'un ensemble d'ontologies métier relatives au domaine de la modélisation de réservoirs géologiques a exigé la tenue d'une dizaine de réunions étalées sur plus de 6 mois. Ce dialogue prolongé entre acteurs est extrêmement enrichissant. Toutefois, lors de la planification d'un projet, il convient de soigneusement prendre en compte les délais qu'il impose.

## Références

- [1] Brandel, S., Schneider S., Perrin M., Guiard N., Rainaud J.F., Lienhardt P., and Bertrand Y., 2004, Automatic building of structured geological model. in *ACM Symposium on Solid Modelling and Applications. Genova (Italy)*.
- [2] Callec, Y., Janjou D., Baudin T., Luquet C., Pellé J. M., Laville P., 2006, Échelle des Temps Géologiques, *Bureau de recherches géologiques et minières (BRGM)*.
- [3] Khelif, K., Dieng-Kuntz, R., Barbry, P., 2006, An ontology-based approach to support text mining and information retrieval in the biological domain, of the *Journal of Universal Computer Science (JUCS), Special Issue on Ontologies and their Applications*, 13(12) 1881-1907
- [4] Perrin M., Mastella L., Morel O., Lorenzatti A., 2011, Geological time formalization: an improved formal model for describing time successions and their correlation , *Earth Science Informatics*, volume 4, number 2, pp. 81-96. 2011
- [5] Noy N. F., and D. L. McGuinness, 2001, *Ontology development 101: A guide to creating your first ontology*, Stanford, CA, *Stanford University - Knowledge Systems Laboratory*.
- [6] Uschold M., Gruninger M. ., 2004, Ontologies and semantics for seamless connectivity. 33(4):58–64, URL: <http://www.sigmod.org/sigmod/record/issues/0412/12.uschold-9.pdf>.
- [7] Perrin M. , Durville P., Grataloup S., Mastella, S. Lions J., Morel O., Rainaud J.F., 2008, Knowledge issues for automatic identification of CO<sub>2</sub> storage sites by means of Semantic Web. Technology, *CO<sub>2</sub> Geological Storage Workshop, Budapest*, September 2008
- [8] Jean, S, 2007, OntoQL, un langage d'exploitation des bases de données à base ontologique, *thèse Université de Poitiers*, Décembre 2007
- [9] Guarino N. , Welty C.A., 2004, An overview of OntoClean. In: S. Staab e R. Studer (Ed.). *Handbook of Ontologies*. Berlin: Springer 2004. An overview of OntoClean, p.151-171. (*International Handbook on Information Systems*)
- [10] Guizzardi, G., 2005, *Ontological Foundations for Structural Conceptual Models*. Enschede, *The Netherlands: Universal Press*, v.05-74. 2005. 410 p. (CTIT PhD Thesis Series)